SCHOLAR COACH
Boutique Academic Coaching & Tutoring

AP STATISTICS · COURSE, PRACTICE, AND EXAMS

AP Statistics Mastery.

A complete course: original lessons, worked examples, practice problems, unit tests, and three full-length practice exams in the digital format, with every answer explained and every free-response question scored.

INSIDE

Exploring Data · Normal Distributions · Regression · Sampling and Experiments · Probability · Sampling Distributions · Inference for Proportions and Means · Chi-Square and Slopes · Decision Guide · Unit Tests · Three Full-Length Practice Exams with Scoring Guides · Formulas, Conditions, Glossary, and Tables

WEB www.scholarcoachinc.com

BOOK NOW →

CALL (862) 754-3562

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

SCHOLAR COACH
Boutique Academic Coaching & Tutoring

SCHOLAR COACH AP STATISTICS MASTERY

Table of Contents.

Course, practice, unit tests, and three full-length practice exams.

Contents

Start with the exam overview and writing guide, work the ten units in order or jump to the one you need, then finish with the capstone, three full-length practice exams, and the reference section. Click any entry to jump to that page.

PART 1: START HERE

Start Here: The Exam, Your Plan, and Your Tools 5

F.1 How the AP Statistics Exam Works and How to Use This Book 5
F.2 Writing Statistical Answers That Earn Full Credit 12
F.3 Calculator Skills for Statistics 19
Answers and Explanations 27

PART 2: THE UNITS

Unit 1: Exploring One-Variable Data 30

1.1 Data, Variables, and Categorical Graphs 30
1.2 Graphs for Quantitative Data: Dotplots, Stemplots, and Histograms 37
1.3 Describing Shape, Center, and Spread 46
1.4 The Five-Number Summary, Boxplots, and Outliers 52
1.5 Comparing Distributions in Context 57
1.6 Percentiles, z-Scores, and Transforming Data 63
T Unit 1 Test 69
Exploring One-Variable Data: Answers and Explanations 71

Unit 2: Density Curves and Normal Distributions 77

2.1 Density Curves and the Empirical Rule 77
2.2 Normal Calculations: Finding Areas 83
2.3 Inverse Normal and Finding Cutoffs 87
2.4 Assessing Normality and Using Normal Models 92
T Unit 2 Test 99
Density Curves and Normal Distributions: Answers and Explanations 100

Unit 3: Exploring Two-Variable Data 104

3.1 Two-Way Tables and Categorical Association 104
3.2 Scatterplots: Form, Direction, Strength, Outliers 110
3.3 Correlation 119
3.4 Least-Squares Regression Lines 125
3.5 Residuals, r-squared, and Standard Deviation of Residuals 131
3.6 Transforming Data to Achieve Linearity 138
T Unit 3 Test 145
Exploring Two-Variable Data: Answers and Explanations 147

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 4: Collecting Data: Sampling, Experiments, and Simulation 153
4.1 Populations, Samples, and Sources of Bias 153
4.2 Sampling Methods: SRS, Stratified, Cluster, Systematic 160
4.3 Observational Studies, Experiments, and Confounding 165
4.4 Designing Experiments: Control, Randomization, Replication, Blocking 171
4.5 Scope of Inference and Writing About Study Design 178
4.6 Simulation and Random Digits 184
T Unit 4 Test 190
Collecting Data: Sampling, Experiments, and Simulation: Answers and Explanations 192

Unit 5: Probability and Random Variables 198
5.1 Probability Rules and Sample Spaces 198
5.2 Conditional Probability, Independence, and Tree Diagrams 204
5.3 Discrete Random Variables: Mean and Standard Deviation 210
5.4 Combining Random Variables and Transforming Variables 216
5.5 Binomial Distributions 222
5.6 Geometric Distributions 229
5.7 Continuous Random Variables and Uniform Models 235
T Unit 5 Test 242
Probability and Random Variables: Answers and Explanations 243

Unit 6: Sampling Distributions 249
6.1 What Is a Sampling Distribution? 249
6.2 Sampling Distribution of a Sample Proportion 256
6.3 Sampling Distribution of a Sample Mean and the Central Limit Theorem 262
6.4 Sampling Distributions for Differences 269
T Unit 6 Test 274
Sampling Distributions: Answers and Explanations 276

Unit 7: Inference for Proportions 280
7.1 The Logic of Confidence Intervals 280
7.2 One-Proportion z-Intervals, Margin of Error, and Sample Size 286
7.3 The Logic of Significance Tests and P-Values 292
7.4 One-Proportion z-Tests 299
7.5 Errors, Significance, and Power 306
7.6 Two-Proportion z-Intervals and Tests 312
T Unit 7 Test 320
Inference for Proportions: Answers and Explanations 322

Unit 8: Inference for Means 328
8.1 The t-Distributions and Conditions for Means 328
8.2 One-Sample t-Intervals 337
8.3 One-Sample t-Tests 343
8.4 Two-Sample t-Intervals 350
8.5 Two-Sample t-Tests 356
8.6 Paired Data and Matched Pairs 363
T Unit 8 Test 369
Inference for Means: Answers and Explanations 371

Unit 9: Inference for Categorical Data: Chi-Square 376
9.1 The Chi-Square Statistic and Goodness of Fit 376
9.2 Chi-Square Test for Homogeneity 382
9.3 Chi-Square Test for Independence 388
9.4 Choosing and Interpreting Chi-Square Tests 394
T Unit 9 Test 399
Inference for Categorical Data: Chi-Square: Answers and Explanations 400

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 10: Inference for Regression Slopes 405
10.1 The Regression Model, Conditions, and Computer Output 405
10.2 Confidence Intervals for the Slope 413
10.3 Hypothesis Tests for the Slope 420
T Unit 10 Test 428
Inference for Regression Slopes: Answers and Explanations 430

PART 3: PUTTING IT TOGETHER

Capstone: Choosing Procedures and Investigative Tasks 433
C.1 Which Procedure Do I Use? The Decision Guide 433
C.2 Investigative Tasks: Combining Skills Across Units 440
Answers and Explanations 447

PART 4: FULL-LENGTH PRACTICE EXAMS

Practice Exam 1 449
Exam 1 Section I: Multiple Choice 449
Exam 1 Section II: Free Response 457
Practice Exam 1: Answers, Explanations, and Scoring 458

Practice Exam 2 466
Exam 2 Section I: Multiple Choice 466
Exam 2 Section II: Free Response 475
Practice Exam 2: Answers, Explanations, and Scoring 476

Practice Exam 3 484
Exam 3 Section I: Multiple Choice 484
Exam 3 Section II: Free Response 491
Practice Exam 3: Answers, Explanations, and Scoring 493

PART 5: REFERENCE

Reference: Formulas, Conditions, Glossary, and Tables 501
R.1 Formula Reference 501
R.2 Conditions Checklist for Every Procedure 505
R.3 Glossary 509
Tables Statistical Tables 515

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Start Here: The Exam, Your Plan, and Your Tools

The 2027 exam format, how to write full-credit answers, and the calculator skills used throughout the book.

F.1 How the AP Statistics Exam Works and How to Use This Book

Learn the 2027 exam format, the five exam units, how this book treats topics beyond the core, how to score practice exams, and two study calendars so every hour of preparation counts.

Why This Matters

Every strong exam performance starts with knowing the game. Before you can pace a test, choose which lessons to study first, or decide what to do with a topic that may no longer be tested, you need the facts about how the exam is built. This lesson gives you those facts, a way to score your own practice, two study calendars, and a map of this book.

Students often lose points for reasons that have nothing to do with statistics: spending four minutes on one multiple-choice question, writing a free-response answer with no context, or skipping review because the schedule was never planned. The format described here is first administered in May 2027 and is fully digital. Everything in this lesson touches all five exam units, because it is about how you take the whole test.

Core Ideas

The 2027 exam format

The exam has two sections of equal length and equal weight. Both are taken on a screen, you may use a graphing calculator, and a formula sheet and tables are provided for you. You do not need to memorize every formula, but you do need to know which formula applies and how to use it in context.

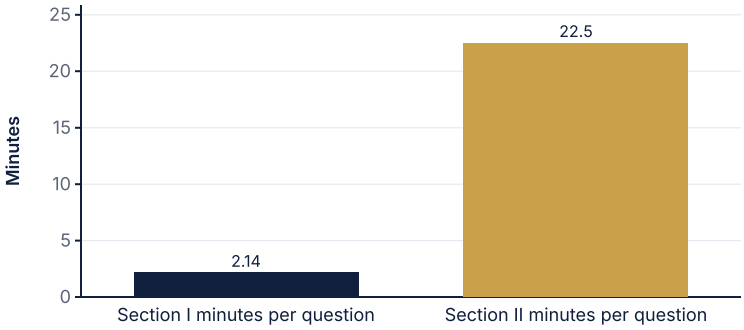
Exam format at a glance

Section I: 42 multiple-choice questions, 4 answer choices each, 90 minutes, 50% of the exam score.
Section II: 4 free-response questions, worth 10 points each (40 points in all), 90 minutes, 50% of the exam score.
Tools: fully digital exam, graphing calculator allowed, formulas and tables provided.
First administered: May 2027.

Feature	Section I	Section II
Question type	Multiple choice	Free response
Number of questions	42	4
Time	90 minutes	90 minutes
Weight	50%	50%
Average time per question	about 2.14 minutes	22.5 minutes
What earns credit	The correct choice	Method, conditions, numbers, and conclusion in context

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Average time available per question. Free-response questions give you about ten times as long, and they expect about ten times as much writing.

The five exam units

The exam is organized into five units. This book is organized into ten units because some exam units are large and are easier to learn in smaller pieces. The table shows where each exam unit lives in the book.

Exam unit	Title	Where it is in this book
1	Exploring One-Variable Data and Collecting Data	Book Units 1, 2 and 4
2	Probability, Random Variables, and Probability Distributions	Book Units 5 and 6
3	Inference for Categorical Data: Proportions	Book Unit 7
4	Inference for Quantitative Data: Means	Book Unit 8
5	Regression Analysis	Book Unit 3 (and extension Unit 10)

Topics this book marks as beyond the core

Beyond the 2027 Exam Core

The College Board revision page lists these topics as removed for 2027: analyzing departures from linearity, combining random variables, the geometric distribution, the chi-square goodness-of-fit test, and inference for regression slopes. This book still teaches them, because college courses and older exams use them, but each lesson opens with a warning label. Treat them as extension, and check your teacher's current course description before deciding how much time to give them.

Topic	Where in this book	How to treat it
Departures from linearity (residual plots)	Lesson 3.5	Learn residual plots for checking a line; treat deep analysis of curved patterns as extension
Combining random variables	Lesson 5.4	Extension
Geometric distribution	Lesson 5.6	Extension
Chi-square goodness of fit	Lesson 9.1	Extension
Inference for regression slopes	Unit 10	Extension
Chi-square homogeneity and independence	Lessons 9.2 to 9.4	Not named on the revision page, but the 2027 unit titles have no chi-square unit, so confirm with your teacher

How to score yourself

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

The two sections count equally. That means one multiple-choice question is worth a little more than one percentage point of the exam, while each point on a free-response question is worth a bit more than one percentage point too. The conversion below is a **Scholar Coach practice approach**, not an official scoring rule. It lets you track progress between practice exams.

Practice composite (out of 100)

$$\text{Composite} = \frac{50}{42}\,(\text{number of multiple-choice correct}) + \frac{50}{40}\,(\text{free-response points out of } 40).$$

Use it to compare your own practice exams over time. The Scholar Coach free-response rubrics in this book are practice rubrics, not official ones, so treat a gap of a few points as noise and look for patterns in what you lose.

Two study plans

Pick the 12-week plan if you are starting in the fall or early winter, and the 4-week plan if you are reviewing after a full course. In both plans, extension topics come last and only if your teacher includes them.

Week	Focus (lessons)	Weekly check
1	F.1 to F.3, then Unit 1 lessons 1 to 3	Describe a distribution in words
2	Unit 1 lessons 4 to 6	Compare two distributions with SOCS
3	Unit 2 (normal distributions)	Use normal calculations and cutoffs
4	Unit 3 (two-variable data)	Interpret slope, r and r-squared
5	Unit 4 (collecting data)	Name a bias and write a design
6	Unit 5 lessons 1 to 3 and 5	Use probability rules and binomial models
7	Unit 5 lesson 7, Unit 6	Explain a sampling distribution
8	Unit 7 lessons 1 to 3	Interpret an interval and a p-value
9	Unit 7 lessons 4 to 6	Run a full z-test in four steps
10	Unit 8 lessons 1 to 4	Run one-sample and two-sample t procedures
11	Unit 8 lessons 5 to 6, then C.1	Choose the correct procedure
12	C.2 and practice exams 1 to 3	Score yourself with the composite

Week	Focus	Weekly check
1	Units 1, 2 and 4 (summaries, normal model, design)	Complete each lesson's 15 practice items
2	Units 3, 5 and 6 (regression, probability, sampling distributions)	Interpret every number in context
3	Units 7 and 8 (z and t inference)	Write four-step answers under time
4	C.1, C.2, practice exams 1 to 3, mistake log review	Composite score trend

How to use this book

- Read the Why This Matters section first, so you know the exam unit and the typical trap.
- Cover the solution to each worked example and try it before reading on.
- Study the Write It Like a 5 model response, then write your own version of a similar prompt and compare it to the scoring points.
- Do the 15 practice items; items 1 to 9 are multiple choice, items 10 to 15 are free response.
- Keep a mistake log: write the error, the correct idea, and a lesson number to revisit.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: PACING THE MULTIPLE-CHOICE SECTION

The multiple-choice section has 42 questions in 90 minutes. Find the average time per question. If you want 10 minutes left at the end to review flagged questions, how long can you average per question?

SOLUTION

Average time: $90/42 \approx 2.14$ minutes, about 2 minutes 9 seconds.

Time for answering after reserving review time: $90 - 10 = 80$ minutes. Then $80/42 \approx 1.90$ minutes, about 1 minute 54 seconds per question.

Conclusion: plan on a little under two minutes per question, and flag any question that has taken more than about three minutes.

EXAMPLE 2: PACING THE FREE-RESPONSE SECTION

Section II has 4 questions in 90 minutes. Build a timing plan for one question that averages the available time and includes reading, writing and checking.

SOLUTION

Average time per question: $90/4 = 22.5$ minutes.

A sample split: about 3 minutes to read and mark the prompt, 16 minutes to compute and write, and 3.5 minutes to check context and units. Total: $3 + 16 + 3.5 = 22.5$.

Because every question is worth the same 10 points, do not spend 35 minutes on one hard part. A partly complete answer to all four questions usually earns more than a perfect answer to two.

EXAMPLE 3: COMPUTING A PRACTICE COMPOSITE

On a practice exam you answer 30 of 42 multiple-choice questions correctly and score 7, 5, 6 and 4 on the four free-response questions. Find your practice composite.

SOLUTION

Free-response total: $7 + 5 + 6 + 4 = 22$ out of 40.

Multiple-choice part: $30 \cdot \frac{50}{42} \approx 35.71$.

Free-response part: $22 \cdot \frac{50}{40} = 27.5$.

Composite: $35.71 + 27.5 \approx 63.2$ out of 100. This is your practice number for tracking growth, not an official score.

EXAMPLE 4: BASIC VOCABULARY, VARIABLE TYPES

A teacher surveys students and records the variables below. Classify each as categorical or quantitative.

Variable	Example value
Homeroom number	214
Minutes of sleep last night	455
Favorite subject	Science
Number of siblings	2
Preferred lunch line	B

SOLUTION

The **individual** is each student. A **categorical variable** places individuals into groups, and a **quantitative variable** takes numerical values where arithmetic such as averaging makes sense.

Homeroom number: categorical (it is a label; the average of two room numbers means nothing).

Minutes of sleep last night: quantitative.

Favorite subject: categorical.

Number of siblings: quantitative (discrete counts).

Preferred lunch line (A, B, C): categorical.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: PARAMETER VERSUS STATISTIC

A school has 1,800 students. A counselor takes a random sample of 120 students and finds that 42 of them slept fewer than 7 hours last night. The dotplot shows the sample percentages from 15 repeated random samples of size 120 that a simulation produced. Identify the population, the sample, the parameter, and the statistic, and explain why the dotplot is not a single number.

Sample percent sleeping fewer than 7 hours (15 simulated samples)

SOLUTION

Population: all 1,800 students at the school. Sample: the 120 students selected. Variable: whether a student slept fewer than 7 hours (categorical).

Statistic: $\hat{p} = 42/120 = 0.35$, a number computed from a sample. Parameter: p, the proportion of all 1,800 students who slept fewer than 7 hours, an unknown fixed number.

The dotplot shows that statistics change from sample to sample: the 15 sample percentages run from 29 to 42 with a mean of about 35.5. The parameter does not change; the statistic does. That variation is the reason the whole inference half of the course exists.

EXAMPLE 6: ALLOCATING STUDY HOURS

You can study 5 hours a week for 12 weeks. Reserve 4 hours total for two practice exams, and split the remaining hours among the five exam units in the proportions shown. Fill in the hours for each unit.

Exam unit	Share of study time
1	32%
2	21%
3	18%
4	18%
5	11%

SOLUTION

Total hours: $12 \cdot 5 = 60$. After the practice exams: $60 - 4 = 56$ hours.

Using the shares in the table, Unit 1 gets $0.32 \cdot 56 \approx 18$ hours, Unit 2 gets $0.21 \cdot 56 \approx 12$, Unit 3 and Unit 4 each get $0.18 \cdot 56 \approx 10$, and Unit 5 gets $0.11 \cdot 56 \approx 6$.

Check: $18 + 12 + 10 + 10 + 6 = 56$. These shares are Scholar Coach study suggestions based on how much the book teaches in each area. They are not official exam weights.

Write It Like a 5

Free-response questions begin with a context and ask for a decision. Here is a prompt about planning, written so you can practice the style of full-credit communication even in a non-statistical setting.

Prompt: A student took a practice exam and answered 33 of the 42 multiple-choice questions correctly. The free-response scores were 8, 6, 5 and 7. The student's goal is a composite of at least 75. Does the student reach the goal? If not, state one specific, measurable change that would reach it.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Free-response total: $8 + 6 + 5 + 7 = 26$ out of 40. The practice composite is $33 \cdot \frac{50}{42} + 26 \cdot \frac{50}{40} = 39.29 + 32.5 = 71.79$, which is about 71.8. Since 71.8 is below 75, the student does not reach the goal. The student needs about $75 - 71.8 = 3.2$ more composite points. Each additional multiple-choice question correct adds $50/42 \approx 1.19$ points, so three more correct answers add about 3.57 points and bring the composite to about 75.4, which reaches the goal. So the student should aim for 36 correct multiple-choice questions, for example by using the 12-week plan's weekly checks on the weakest unit.

What the grader looks for

(1) Correct free-response total of 26. (2) Correct use of the weights, composite about 71.8. (3) A clear decision that compares the composite to 75 and says no. (4) A specific numerical target (36 correct or equivalent) justified by arithmetic. (5) Conclusion stated in the language of the context, not just a number.

Coach's Playbook

1. **Know the clock.** Write 2 minutes per multiple-choice question and 22 minutes per free-response question at the top of your scratch notes.
2. **Answer easy items first.** Flag and return to anything that needs more than three minutes; every multiple-choice question is worth the same.
3. **Use the tools.** The calculator and the provided formulas and tables are there to be used. Lesson F.3 shows how.
4. **Write for a reader.** On free response, each answer should state the method, show the values, and conclude in context. Lesson F.2 gives sentence frames.
5. **Log your errors.** Sort each mistake as concept, calculation, communication, or timing. Work on the biggest category first.

Common traps

1. Treating the study calendar as optional and cramming inference in the last week. 2. Spending half the free-response time on part (a) of one question. 3. Studying extension topics deeply before the core units are secure. 4. Memorizing formulas and ignoring that the exam provides them, while skipping the reading of conditions that it does not provide for you.

Mistake Clinic

Spot the error

A student computes a practice composite for 30 multiple-choice correct and 22 free-response points as 30 + 22 = 52. Where does this fail? It adds raw counts from two sections with different numbers of items. The sections count equally, so each must be scaled to 50: $30 \cdot 50/42 + 22 \cdot 50/40 \approx 63.2$.

Spot the error

A student says, "The free-response section has 4 questions, so each question is worth 4% of the exam." Where does this fail? Section II is 50% of the exam, so each of the 4 questions is worth $50/4 = 12.5\%$, and each point on a 10-point question is worth 1.25%.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student says, "The sample proportion is 0.35, so the proportion of all students is 0.35." Where does this fail? A sample proportion is a statistic, and statistics vary. The correct statement is that 0.35 is an estimate of the unknown parameter p, and another sample would give a different value.

Real-World Connection

Every pilot, surgeon and engineer uses a checklist and a time budget, because under pressure people forget good habits. Your pacing plan and mistake log are the same tool. The skills of reading a prompt, choosing a method and communicating a decision are also what employers who handle data look for.

Challenge Corner

A student wants a composite of at least 80 and expects to score 26 total on free response. What is the minimum number of multiple-choice questions correct? **Solution:** Free-response part: $26 \cdot \frac{50}{40} = 32.5$. Need $\frac{50}{42}m \geq 80 - 32.5 = 47.5$, so $m \geq 47.5 \cdot \frac{42}{50} = 39.9$. Since m must be a whole number, $m = 40$ correct. Check: $40 \cdot 50/42 = 47.62$, and $47.62 + 32.5 = 80.12 \geq 80$. With 39 correct the composite is $46.43 + 32.5 = 78.93 < 80$.

Practice

PRACTICE

How the AP Statistics Exam Works and How to Use This Book

Show your work. A calculator is allowed. Justify conclusions in context.

1. How many multiple-choice questions are on the 2027 exam, and how many answer choices does each have?
   (A) 36 questions, 5 choices **(B)** 40 questions, 4 choices **(C)** 42 questions, 4 choices
   (D) 45 questions, 5 choices
2. Section II has 4 free-response questions in 90 minutes. What is the average time available per question?
   (A) 22.5 minutes **(B)** 20 minutes **(C)** 18 minutes
   (D) 30 minutes
3. What is the weight of the free-response section in the exam score?
   (A) 40% **(B)** 25% per question **(C)** 60%
   (D) 50%
4. Which statement about tools on the 2027 exam is correct?
   (A) The exam is on paper and calculators are not allowed
   (B) The exam is fully digital, a graphing calculator is allowed, and formulas and tables are provided
   (C) The exam is fully digital, but you must memorize all formulas
   (D) A graphing calculator is allowed only on the free-response section
5. Which of these topics does the College Board revision page list as removed for 2027?
   (A) Confidence interval for a proportion **(B)** The normal distribution
   (C) The geometric distribution **(D)** The binomial distribution
6. A poll of 200 randomly chosen voters finds that 54% support a proposal. In this context, 54% is a
   (A) statistic **(B)** parameter **(C)** population
   (D) variable
7. Which variable is quantitative?
   (A) Zip code of a student's home **(B)** Number of books checked out this month
   (C) Jersey number **(D)** Eye color
8. A student answers 36 of 42 multiple-choice questions correctly and earns 24 of 40 free-response points. Using the practice composite from this lesson, the composite is closest to
   (A) 60.0 **(B)** 66.5 **(C)** 85.7 **(D)** 72.9

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. A question asks for a confidence interval for a population mean. Which exam unit is it in?
   (A) Exam Unit 4 **(B)** Exam Unit 3 **(C)** Exam Unit 2
   (D) Exam Unit 5
10. A student plans to spend 55 minutes on 42 multiple-choice questions before moving to review. (a) Find the average time per question. (b) How many minutes remain for review if the student uses the full 90 minutes for this section?
11. A survey of a class records: favorite sport, hours of homework last night, number of pets, grade level (9, 10, 11 or 12), and whether the student owns a bike (yes or no).
    (a) Classify each variable as categorical or quantitative.
    (b) Which of these is a poor candidate for computing a mean, and why?
12. On a practice exam a student answers 28 multiple-choice questions correctly and scores 6, 8, 5 and 6 on free response.
    (a) Find the practice composite.
    (b) Which section contributes more of the composite, and by how much?
13. Design a 4-week review calendar for a student who has finished the course but is weak on Unit 4 (inference for means). Give each week a focus and a weekly check, and explain where Unit 4 time is placed.
14. List the five topics the College Board revision page lists as removed for 2027. For one of them, explain how you would decide whether to study it.
15. A school has 900 students. A student council member surveys 90 randomly chosen students, and 27 say they walk to school.
    (a) Identify the population, the sample, the variable, and its type.
    (b) Compute the statistic and write the symbol for it.
    (c) Explain why the parameter is not 0.30 for certain.

F.2 Writing Statistical Answers That Earn Full Credit

Learn the four ingredients of a full-credit statistics answer and practice sentence frames for context, conditions, interpretations, conclusions, and comparing distributions, with model, weak and strong contrasts.

Why This Matters

On the free-response section, a correct number with no words earns little. Graders award points for four things: naming the right procedure, checking conditions, getting the numbers right, and drawing a conclusion in the language of the problem. Most lost points come from the last two verbs: students compute correctly but never say what the number means about the situation.

This lesson is a toolkit of sentence frames and contrasts. The frames apply across all five exam units: describing distributions (Unit 1), interpreting probabilities (Unit 2), and writing intervals and tests (Units 3 and 4), and regression interpretations (Unit 5). Learn them now, then reuse them in every lesson that follows.

Core Ideas

The four ingredients of a full-credit answer

Four ingredients

1. **Context:** use the variable names, units and population from the problem, never just "the data" or "x and y."
2. **Conditions:** state each condition and check it with numbers from the problem (for example, $n\hat{p} \geq 10$).
3. **Calculation:** show the formula or the calculator function with its inputs, so a reader can follow the values.
4. **Conclusion:** answer the question that was asked, linked to the evidence (an interval, a p-value, a number), in context.

A graded answer is also **hedged correctly**. Samples give estimates, not certainties, so say "estimate," "predicted," "suggests," and "is convincing evidence" rather than "proves."

Interpretation frames

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Quantity	Sentence frame
Mean or median	The typical [variable] for [individuals] is about [value] [units].
Slope	For each additional [unit of x], the predicted [y] changes by about [slope] [units of y].
y-intercept	When [x] is 0 [units], the predicted [y] is [value] [units]. (Say so if x = 0 is outside the data.)
Confidence interval	We are [C]% confident that the true [parameter in words] is between [low] and [high] [units].
Confidence level	If we repeated the sampling process many times and built an interval each time, about [C]% of the intervals would capture the true [parameter].
p-value	If [H0 in words] were true, the probability of getting a result at least as [extreme direction] as [observed statistic] by random chance alone is [p-value].
r-squared	About [r-squared]% of the variation in [y] is accounted for by the linear relationship with [x].
Standard deviation of residuals	Predictions of [y] from the line are typically off by about [s] [units].
Conclusion of a test	Because the p-value of [p] is [less/greater] than alpha = [alpha], we [reject/fail to reject] H0. There [is/is not] convincing evidence that [Ha in words].

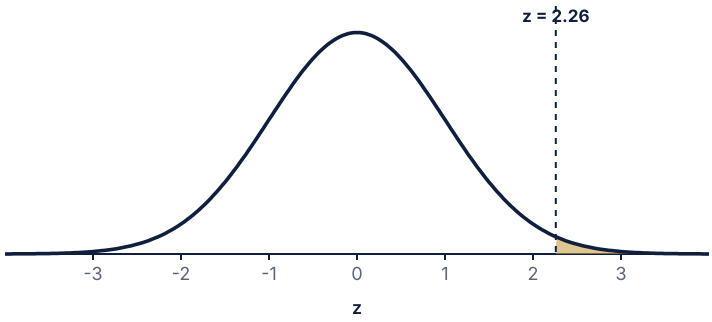
Hypothesis-test skeleton

State: define the parameter in context and write H_0 and H_a.

Plan: name the procedure and check conditions (random, 10% rule, large counts or normal/large sample).

Do: give the statistic and the p-value.

Conclude: compare p to alpha, decide, and write the sentence in context.



A p-value is an area in the tail of the sampling distribution computed assuming H0 is true. Here the upper tail beyond z = 2.26 has area about 0.012.

Comparing distributions: SOCS

Use **SOCS** to compare two distributions in a complete sentence for each: **Shape**, **Outliers**, **Center**, **Spread**. A comparison uses comparison words such as "greater," "more variable" or "both," and always includes numbers with units. Listing one group's center and then the other's, without a comparing word, is the most common reason for lost credit.

Weak answer	Strong answer
Group A has a median of 18 and group B has a median of 14.	The median commute for group A (18 minutes) is 4 minutes longer than for group B (14 minutes).
Group A has a bigger range.	Group A is more variable: its IQR is 7 minutes compared with 5 minutes for group B.
There is an outlier.	Group A has one high outlier at 41 minutes, far above the upper fence of 32.5; group B has none.
Both are skewed.	Group A is skewed right, with a long tail to the high commute times; group B is roughly symmetric.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: CONTEXT AND CONDITIONS

A student government officer surveys a simple random sample of 240 students from a school of 3,000 students; 156 say they support a later start time. Check the conditions and construct a 95% confidence interval for the proportion of all students who support it. Interpret the interval.

SOLUTION

Parameter: p, the proportion of all students at the school who support a later start time.

Conditions: Random: the sample is a simple random sample. 10% rule: 240 is less than 10% of 3,000 (which is 300). Large counts: $\hat{p} = 156/240 = 0.65$, $n\hat{p} = 156 \geq 10$ and $n(1-\hat{p}) = 84 \geq 10$.

Procedure: one-proportion z-interval. $SE = \sqrt{\frac{0.65(0.35)}{240}} \approx 0.0308$. Interval: $0.65 \pm 1.96(0.0308) = (0.590,\ 0.710)$.

Interpretation: We are 95% confident that the true proportion of all students who support a later start time is between 0.590 and 0.710.

EXAMPLE 2: INTERPRETING SLOPE AND INTERCEPT

A regression of quiz score (points out of 100) on weekly hours of study for a class gives $\hat{y} = 54.2 + 4.6x$, where the data have x between 1 and 9 hours. Interpret the slope and intercept in context.

SOLUTION

Slope: For each additional hour of weekly study, the predicted quiz score increases by about 4.6 points.

Intercept: For a student who studies 0 hours a week, the predicted quiz score is 54.2 points. Since the data only cover 1 to 9 hours, this is an extrapolation and should be treated with caution.

Note on wording: say \"predicted,\" because the line gives an average trend, not an exact result, and avoid saying that studying causes the increase unless the study was an experiment.

EXAMPLE 3: INTERVAL AND CONFIDENCE LEVEL

A random sample of 25 batteries from a production run has a mean life of 41.2 hours with a standard deviation of 6.5 hours, and a dotplot shows no strong skewness or outliers. Construct a 95% confidence interval for the mean life and explain what 95% confidence means.

SOLUTION

Conditions: random sample, $n = 25 < 10\%$ of all batteries in the run, and the sample distribution shows no strong skewness or outliers, so using t is reasonable.

Procedure: one-sample t-interval with $df = 24$, $t^* = 2.064$. $\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 41.2 \pm 2.064 \cdot \frac{6.5}{5} = 41.2 \pm 2.683$, giving $(38.52,\ 43.88)$ hours.

Interval: We are 95% confident that the true mean life of batteries in this production run is between 38.52 and 43.88 hours.

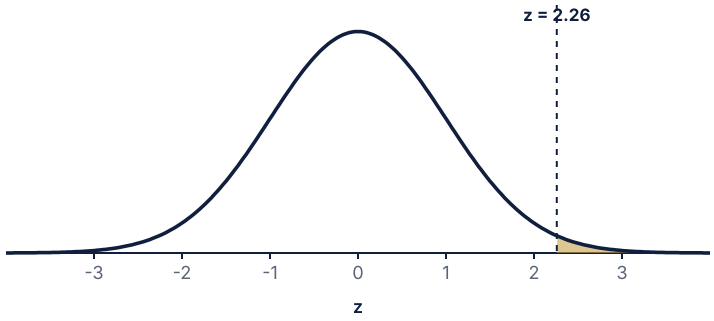
Confidence level: If we took many random samples of size 25 from this run and built a 95% interval from each, about 95% of those intervals would capture the true mean life. It does not mean there is a 95% chance this particular interval contains the mean, and it does not mean that 95% of batteries last between 38.52 and 43.88 hours.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 4: INTERPRETING A P-VALUE

A cafe owner claims 50% of customers prefer iced drinks. In a random sample of 200 customers, 116 (58%) prefer iced drinks. For $H_0 : p = 0.5$ versus $H_a : p > 0.5$, the z-statistic is 2.26 and the p-value is 0.012. Interpret the p-value and state a conclusion at $\alpha = 0.05$.



Upper-tail p-value

SOLUTION

Check: $z = \frac{0.58 - 0.5}{\sqrt{0.5(0.5)/200}} = 2.26$, and the upper-tail area is about 0.0118.

Interpretation: If the true proportion who prefer iced drinks were 0.50, the probability of a sample proportion of 0.58 or greater, by random sampling variability alone, is about 0.012.

Conclusion: Because 0.012 is less than 0.05, we reject H_0. There is convincing evidence that more than 50% of this cafe's customers prefer iced drinks.

EXAMPLE 5: R-SQUARED AND RESIDUAL STANDARD DEVIATION

A regression of daily iced-drink sales (cups) on high temperature (degrees F) gives $r = 0.82$ and a standard deviation of residuals $s = 14.3$ cups. Interpret r^2 and s.

SOLUTION

$r^2 = 0.82^2 = 0.6724$, about 67.2%.

r-squared: About 67% of the variation in daily iced-drink sales is accounted for by the linear relationship with the day's high temperature. The other 33% is due to other factors and random variation.

s: The predictions of daily sales from the least-squares line are typically off by about 14.3 cups.

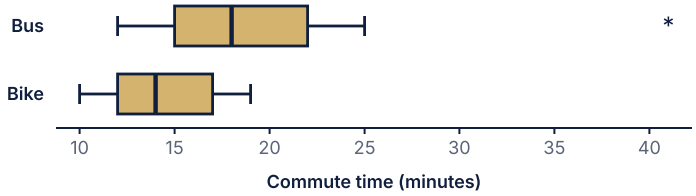
Be careful: r^2 is not the correlation and does not say the line is a good model by itself. You still need a residual plot with no pattern.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 6: COMPARING DISTRIBUTIONS WITH SOCS

Commute times in minutes for 15 students who ride the bus and 15 who ride a bike are summarized in the boxplots. Bus: 12, 14, 15, 15, 16, 17, 18, 18, 19, 20, 21, 22, 23, 25, 41. Bike: 10, 11, 12, 12, 13, 13, 14, 14, 15, 15, 16, 17, 17, 18, 19. Compare the two distributions.



Commute times for 15 bus riders and 15 bike riders

SOLUTION

Bus: Q1 = 15, median = 18, Q3 = 22, IQR = 7, mean 19.7. Fences: 4.5 and 32.5, so 41 is an outlier. Bike: Q1 = 12, median = 14, Q3 = 17, IQR = 5, mean 14.4, no outliers.

Shape: The bus times are skewed right, with a long tail; the bike times are roughly symmetric.

Outliers: The bus distribution has one high outlier at 41 minutes; the bike distribution has none.

Center: The median bus commute (18 minutes) is 4 minutes longer than the median bike commute (14 minutes).

Spread: The bus times are more variable, with an IQR of 7 minutes compared with 5 minutes for bikes.

Write It Like a 5

Prompt: A cafe claims that 40% of its customers order a seasonal drink. In a random sample of 150 customers, 69 ordered it. Is there convincing evidence that the true proportion differs from 40%? Use alpha = 0.05.

Model Response

State: Let p be the true proportion of all customers who order the seasonal drink. $H_0 : p = 0.40$ and $H_a : p \neq 0.40$, with $\alpha = 0.05$. **Plan:** one-proportion z-test. Conditions: the customers are a random sample; 150 is less than 10% of all the cafe's customers (assumed to be more than 1,500); and the expected counts under H_0 are $150(0.40) = 60 \geq 10$ and $150(0.60) = 90 \geq 10$. **Do:** $\hat{p} = 69/150 = 0.46$, $z = \frac{0.46 - 0.40}{\sqrt{0.40(0.60)/150}} = 1.50$, and the two-sided p-value is $2P(Z > 1.50) \approx 0.134$.

Conclude: Because the p-value of 0.134 is greater than 0.05, we fail to reject H_0. There is not convincing evidence that the true proportion of customers who order the seasonal drink differs from 40%.

What the grader looks for

(1) Parameter defined in context and correct hypotheses. (2) Correct procedure named. (3) Random, 10% and large-counts conditions checked with numbers. (4) Correct z of 1.50 and p-value of about 0.134. (5) Comparison of p to alpha, and a conclusion in context that uses "not convincing evidence" rather than "proves H0 is true."

Coach's Playbook

1. **Underline the question.** Circle what the prompt wants: describe, compare, estimate, or test. Your final sentence must answer that exact verb.
2. **Name the parameter.** Write the symbol and what it means in words before any number.
3. **Write numbers with units.** Every interpretation sentence should contain a value and its unit.
4. **Use comparison words.** Words like "higher," "more variable" and "both" show that you are comparing.
5. **Hedge correctly.** Use "estimate," "predicted," "convincing evidence," never "proves."
6. **Reread for context.** Spend the last minute checking that your conclusion mentions the situation, not just "reject H0."

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Common traps

1. Writing "reject H0" without a sentence about the situation. 2. Interpreting a confidence interval as the percent of data in the interval. 3. Saying the p-value is the probability that H0 is true. 4. Describing correlation as causation when the data come from an observational study.

Mistake Clinic

Spot the error

A student writes, "There is a 95% chance that the true mean is in (38.52, 43.88)." Where does this fail? The true mean is a fixed number, so a particular interval either contains it or does not. The 95% describes the method. Correct: "We are 95% confident that the true mean life is between 38.52 and 43.88 hours."

Spot the error

A student writes, "The p-value of 0.012 means there is a 1.2% chance that the null hypothesis is true." Where does this fail? A p-value assumes H0 is true and gives the chance of the data. Correct: "If the true proportion were 0.50, the probability of a sample proportion of 0.58 or higher is about 0.012."

Spot the error

A student writes, "The slope is 4.6, so studying more makes quiz scores go up by 4.6 points." Where does this fail? It claims causation from observational data and omits units and the word "predicted." Correct: "For each additional hour of weekly study, the predicted quiz score increases by about 4.6 points."

Real-World Connection

Analysts present results to managers who do not read formulas. A sentence like "We are 95% confident that between 59% and 71% of customers want the new hours" is the form that appears in reports and news articles. Learning the frames now is learning to communicate evidence.

Challenge Corner

A planner wants a 95% confidence interval for a proportion with a margin of error of at most 0.03. (a) With no prior estimate, how large should the sample be? (b) A pilot study suggested $p \approx 0.65$; how large with that? **Solution:** (a) $n \geq 0.25\left(\frac{1.96}{0.03}\right)^2 = 1067.1$, so $n = 1068$. (b) $n \geq 0.65(0.35)\left(\frac{1.96}{0.03}\right)^2 = 971.1$, so $n = 972$. The guess of 0.5 is conservative because $p(1-p)$ is largest at 0.5; the pilot estimate allows a smaller sample.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE **Writing Statistical Answers That Earn Full Credit**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A 95% confidence interval for the mean time to complete an assembly task is (12.4, 15.8) minutes. Which is the best interpretation?
   - **(A)** 95% of workers finish between 12.4 and 15.8 minutes
   - **(B)** We are 95% confident the true mean completion time is between 12.4 and 15.8 minutes
   - **(C)** There is a 95% probability that the sample mean is between 12.4 and 15.8 minutes
   - **(D)** The mean completion time is 14.1 minutes 95% of the time
2. A line relating a car's age in years (x) to its price in thousands of dollars (y) is $\hat{y} = 21.5 - 1.8x$. Which is the best interpretation of the slope?
   - **(A)** Each year of age decreases the price by exactly 1.8 thousand dollars
   - **(B)** A new car costs 1.8 thousand dollars
   - **(C)** For each additional year of age, the predicted price decreases by about 1.8 thousand dollars
   - **(D)** The price is 1.8 times the age
3. A test of $H_0 : p = 0.3$ against $H_a : p > 0.3$ gives a p-value of 0.04. Which is the best interpretation?
   - **(A)** The probability that H_0 is true is 0.04
   - **(B)** The probability that $p > 0.3$ is 0.96
   - **(C)** There is a 4% chance the sample was random
   - **(D)** If $p = 0.3$, the probability of a sample proportion at least as large as the observed one is 0.04
4. What does a 90% confidence level mean?
   - **(A)** In the long run about 90% of intervals built this way capture the parameter
   - **(B)** The interval contains the true parameter with probability 0.9 after it is computed
   - **(C)** 90% of the data are inside the interval
   - **(D)** The sample proportion is 0.90
5. For a regression, $r^2 = 0.49$. Which is the best statement?
   - **(A)** 49% of the points lie on the line
   - **(B)** About 49% of the variation in y is accounted for by the linear relationship with x
   - **(C)** The correlation is 0.49
   - **(D)** The line is correct 49% of the time
6. Which is the strongest comparison of two distributions?
   - **(A)** Class A has a median of 74 and class B has a median of 78
   - **(B)** Class A has more spread
   - **(C)** The median for class B (78) is 4 points higher than for class A (74), and class A is more variable (IQR 14 versus 8)
   - **(D)** Both classes have outliers
7. Before a one-sample t-interval for a mean with n = 14, which condition must be checked with the data?
   - **(A)** That $n\hat{p} \geq 10$
   - **(B)** That the sample size is at least 30
   - **(C)** That the population is normally distributed with known standard deviation
   - **(D)** That a graph of the sample shows no strong skewness or outliers
8. A teacher wants to test whether the mean time students take on an assignment is more than 40 minutes. Which pair of hypotheses is written correctly?
   - **(A)** $H_0 : \bar{x} = 40$, $H_a : \bar{x} > 40$
   - **(B)** $H_0 : \mu = 40$, $H_a : \mu > 40$ where μ is the true mean time for all students
   - **(C)** $H_0 : \mu > 40$, $H_a : \mu = 40$
   - **(D)** $H_0 : \mu = 40$, $H_a : \mu /= 40$ only if μ is unknown

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. A test has p-value 0.18 and $\alpha = 0.05$. Which conclusion is best?
   (A) Fail to reject H_0 ; there is not convincing evidence for the alternative in context
   (B) Reject H_0 because the p-value is small
   (C) Accept H_0 because the p-value is large
   (D) The null hypothesis is true with probability 0.82
10. A 95% t-interval for the mean number of hours of sleep of teens at a school is (6.8, 7.6).
    (a) Interpret the interval in context.
    (b) Does the interval give convincing evidence that the mean differs from 8 hours? Explain.
    (c) State what 95% confidence means.
11. A random sample of 150 customers shows 45 who use the mobile app.
    (a) Check the conditions for a one-proportion z-interval.
    (b) Construct a 95% confidence interval.
    (c) Interpret the interval in context.
12. A regression of the number of clicks on a web page (y) on the number of ads shown (x) gives $\hat{y} = 18.4 + 2.5x$ with $r = 0.78$ and s = 6.1 clicks.
    (a) Interpret the slope.
    (b) Interpret the intercept and say whether it is meaningful if x ranged from 5 to 40.
    (c) Interpret r^2.
    (d) Interpret s.
13. A test of $H_0 : p = 0.4$ versus $H_a : p > 0.4$ gives a p-value of 0.034, where p is the proportion of a city's households that recycle.
    (a) Interpret the p-value in context.
    (b) State the conclusion at alpha = 0.05.
    (c) Explain why "fail to reject" would be the conclusion at alpha = 0.01.
14. Quiz scores for two sections: Section 1: 62, 70, 71, 73, 75, 78, 80, 82, 95. Section 2: 68, 70, 72, 74, 75, 76, 77, 79, 81. Compare the distributions using SOCS. Use quartiles (median of each half).
15. A student wrote: "The p-value is 0.03, so there is a 3% chance the null hypothesis is true, so I reject H0 and the claim is proven false." Identify two errors and rewrite a correct conclusion for $H_0 : \mu = 50$ and $H_a : \mu < 50$, where μ is the true mean battery life in hours.

F.3 Calculator Skills for Statistics

Learn the built-in calculator functions that answer exam questions fast and how to write the function and inputs so graders can follow: normal, binomial, t, proportion procedures, one-variable statistics and regression.

Why This Matters

A graphing calculator is allowed on the entire exam, and formulas and tables are provided. That does not mean the calculator does the thinking. It means you do not lose points to long arithmetic, and you have time to write the explanations that earn credit. Students go wrong in two ways: by never learning the key functions, so a normal-probability question eats five minutes, and by trusting a screen without writing down what was entered.

The functions in this lesson are used in every exam unit: normal calculations (Exam Unit 1), binomial probabilities (Exam Unit 2), one-proportion procedures (Exam Unit 3), t procedures (Exam Unit 4), and linear regression (Exam Unit 5). This lesson names each function generically. Menu locations and exact prompts differ among calculators, so use the function names to find them in the built-in calculator you will use on test day.

Core Ideas

Think of each function as answering one question about a distribution. The names below are the ones used across most graphing tools. Always read the prompts: some tools ask for lower bound, upper bound, mean and standard deviation in that order, while others use a form with labeled boxes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Function	Question it answers	Inputs
normalcdf	What is the area (probability) between two values of a normal curve?	lower, upper, mean, standard deviation
invNorm	What value has a given area to its left?	area to the left, mean, standard deviation
1-Var Stats	What are the mean, standard deviation, median and quartiles of a list?	the list
LinReg	What are the slope, intercept, r and r-squared for paired data?	x list, y list
tcdf	What is the area between two values of a t-distribution?	lower, upper, degrees of freedom
binompdf	What is P(X = k) for a binomial?	n, p, k
binomcdf	What is P(X at most k) for a binomial?	n, p, k
1-PropZTest and 1-PropZInt	Test or interval for one proportion	successes, n, (p0 or confidence level)
T-Test and TInterval	Test or interval for one mean (data or summary statistics)	mean, s, n, (mu0 or confidence level)

Reading a cumulative function

A "cdf" function returns the area to the **left** of its upper bound (or between the bounds). It always adds up area, never a single point.

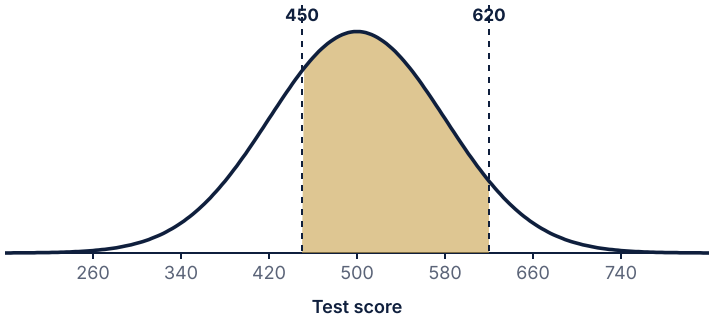
For a continuous distribution, use any very large negative number such as -1E99 for "negative infinity" and 1E99 for "positive infinity".

Upper-tail probability: $P(X > c) = 1 - \text{normalcdf}(-1E99, c, \ldots)$, or use c as lower bound and 1E99 as upper bound.

Calculator-use rule for free response

A correct answer shows the **function name and its inputs** (for example, "normalcdf(450, 620, 500, 80) = 0.667") so a grader can follow the work.

A bare number with no setup may earn no credit even if it is right. Also write what the number means in context.



normalcdf(450, 620, 500, 80) gives the shaded area, about 0.667.

Binomial reminder. For a binomial random variable X with n trials and success probability p, binompdf gives P(X = k) and binomcdf gives P(X <= k). A common inversion: P(X >= k) is **1 - binomcdf(n, p, k-1)**, because binomcdf is cumulative from the left and includes k.

t procedures. The t-distribution needs degrees of freedom. For one sample, df = n - 1. The tool may offer T-Test and TInterval with either raw data in a list or summary statistics. On the exam you will usually be given summary values. For critical values, a function like invT (area to the left, df) returns t*. For a 95% interval, the area to the left is 0.975.

Choosing the right function

Most calculator mistakes happen before any button is pressed: the student has not decided which question is being asked. Use the table below as a decision aid. Start from what you are given (a value or an area) and what you want (an area or a value), then choose the distribution.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

You know	You want	Normal	Binomial	t-distribution
A value (or two)	An area (probability)	normalcdf	binompdf or binomcdf	tcdf
An area to the left	A value (cutoff)	invNorm	Count with the smallest k such that binomcdf is at least the area	invT
Raw data in lists	Summary numbers	1-Var Stats	1-Var Stats	1-Var Stats
Paired data in lists	Line, r, r-squared	LinReg	LinReg	LinReg

Entering data and checking your work

- **Lists.** Clear old lists before entering new data. Enter x values in one list and y values in a second list of the same length; a mismatch in length produces an error or a wrong line.
- **Statistical plots.** After entering data, draw a histogram, boxplot or scatterplot to look at the shape before any calculation. A calculator will happily compute a mean for skewed data with outliers, but it will not tell you the mean is a poor summary.
- **Residuals.** After LinReg, the residual list is stored by the tool; plot the residuals against x to check for a pattern. A curved pattern means a line is not an appropriate model. This idea is developed in Lesson 3.5.
- **Mode settings.** Make sure the calculator is set to decimal display with at least four decimals, and that diagnostics such as r and r-squared are turned on, or they will not appear after a regression.
- **Reasonableness.** Estimate before you compute. For a normal question, the Empirical Rule predicts roughly 68%, 95% and 99.7%; if the calculator says 0.0002 for a value one standard deviation from the mean, an input is wrong.

Matching functions to the exam units

Exam unit	Typical calculator tasks
1: One-variable data and collecting data	1-Var Stats, boxplots and histograms, normalcdf and invNorm
2: Probability and distributions	binompdf, binomcdf, normalcdf for sampling distributions
3: Proportions	1-PropZInt, 1-PropZTest, two-proportion procedures, invNorm for z*
4: Means	TInterval, T-Test, two-sample procedures, tcdf, invT
5: Regression	LinReg, residual plots, scatterplots with the fitted line

Reading calculator output in context

Every calculator screen contains more numbers than you need, and an exam answer should quote only the ones that matter. For a t-test, report the t-statistic, the degrees of freedom and the p-value; for a regression, report the slope, intercept, r and r-squared with units. Then connect the numbers to a sentence. If the screen says p = 0.056 and alpha is 0.05, the sentence is not "p is close" but "because 0.056 is greater than 0.05, we fail to reject the null hypothesis." Treat the screen as scratch paper and your written answer as the product.

- **Rounding.** Round probabilities to three decimals and test statistics to two or three decimals, unless the problem says otherwise.
- **Labels.** Say what each number is: "t = 2.05", not just "2.05".
- **Units.** Cutoffs, means and slopes carry units; probabilities and proportions do not.
- **Assumptions.** A calculator will run a t-test on any numbers. Check conditions yourself, in words, before you press the button.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: NORMALCDF

Scores on an entrance placement test are approximately normal with mean 500 and standard deviation 80. Find the proportion of students who score between 450 and 620.

Shaded area is about 0.667.

SOLUTION

Setup: $X \sim N(500, 80)$. We want $P(450 < X < 620)$.

Calculator: normalcdf(450, 620, 500, 80) ≈ 0.667.

Conclusion: About 66.7% of students score between 450 and 620 on this test.

Check with z-scores: $z = -0.625$ and $z = 1.5$; areas 0.266 and 0.933, difference 0.667.

EXAMPLE 2: INVNORM

Using the same test, what score is needed to be in the top 10%? What score is the first quartile?

SOLUTION

Top 10%: the area to the left is 0.90. invNorm(0.90, 500, 80) ≈ 602.5.

First quartile: area to the left 0.25. invNorm(0.25, 500, 80) ≈ 446.0.

Conclusion: A score of about 602.5 or higher puts a student in the top 10%, and 25% of students score below about 446.

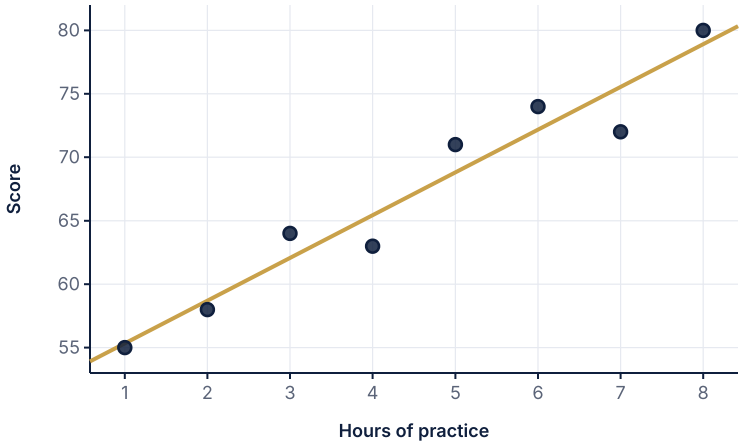
Common error: entering 0.10 for the top 10% gives the bottom 10% cutoff (about 397.5).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: 1-VAR STATS AND LINREG

The table gives hours of practice (x) and score (y) for 8 students. Use the calculator to find the mean and sample standard deviation of y, then fit a least-squares line, and report r, r^2 and the residual standard deviation. Predict the score for 5.5 hours.



Scatterplot with the least-squares line

Hours x	1	2	3	4	5	6	7	8
Score y	55	58	64	63	71	74	72	80

SOLUTION

1-Var Stats on y: $\bar{y} = 67.125$, $s_y = 8.526$ (read the sample standard deviation labeled Sx, not the population version), median 67.5.

LinReg with x list and y list: $a = 51.96$, $b = 3.369$, $r = 0.968$, $r^2 = 0.937$. The line is $\hat{y} = 51.96 + 3.369x$.

The residual standard deviation is $s = \sqrt{\sum e^2/(n-2)} = 2.31$ points.

Prediction at 5.5 hours: $51.96 + 3.369(5.5) = 70.49$, about 70.5 points. 5.5 lies inside the data range (1 to 8), so this is interpolation.

Conclusion: About 94% of the variation in scores is accounted for by the linear relationship with practice hours, and predictions are typically off by about 2.3 points.

EXAMPLE 4: BINOMCDF

A machine produces parts, and 30% of parts from a certain old mold are flawed. A box has 12 independently chosen parts. Find the probability that at most 4 are flawed, the probability that at least 5 are flawed, and the probability of exactly 3.

SOLUTION

$X \sim \text{Binomial}(12, 0.3)$.

At most 4: binomcdf(12, 0.3, 4) ≈ 0.724.

At least 5: $1 - \text{binomcdf}(12, 0.3, 4) = 1 - 0.724 = 0.276$. Do not use binomcdf(12, 0.3, 5), which includes 5 and would give the wrong tail.

Exactly 3: binompdf(12, 0.3, 3) ≈ 0.240.

Conclusion: There is about a 72.4% chance that at most 4 of the 12 parts are flawed.

EXAMPLE 5: TCDF AND A T-TEST

A sample of 18 students has a mean study time of 52.4 minutes with s = 9.1 minutes. Test $H_0 : \mu = 48$ versus $H_a : \mu \neq 48$ at alpha = 0.05, assuming conditions are met. Show how to get the p-value with tcdf and with the built-in test.

SOLUTION

$t = \frac{52.4 - 48}{9.1/\sqrt{18}} = 2.05$ with $df = 17$.

Using tcdf: upper tail = tcdf(2.05, 1E99, 17) ≈ 0.028. Two-sided p-value: $2(0.028) = 0.056$.

Using the built-in T-Test with the summary statistics (mean 52.4, s 9.1, n 18, mu0 48, not equal): $t = 2.051$, p-value $= 0.056$.

Conclusion: Because 0.056 > 0.05, we fail to reject H_0. There is not convincing evidence that the true mean study time differs from 48 minutes.

With TInterval at 95%: $(47.87, 56.93)$, which includes 48, consistent with the test.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 6: 1-PROPZTEST AND 1-PROPZINT

In a random sample of 200 students at a school, 87 say they ride the bus. Test whether the true proportion differs from 0.40 at alpha = 0.05, and build a 95% interval.

SOLUTION

Conditions: random sample; $200(0.4) = 80 \geq 10$ and $200(0.6) = 120 \geq 10$; 200 is less than 10% of the student body.

1-PropZTest with $p_0 = 0.4$, x = 87, n = 200, not equal: $z = 1.010$, p-value $= 0.312$, $\hat{p} = 0.435$. Enter the **count** 87, not the proportion.

1-PropZInt with x = 87, n = 200, C-level 0.95: $(0.366,\ 0.504)$. Note that the interval uses $\hat{p}$ in the standard error.

Conclusion: Since 0.312 > 0.05, we fail to reject H_0. There is not convincing evidence that the proportion of students who ride the bus differs from 0.40; 0.40 is in the interval.

Write It Like a 5

Prompt: The time for a technician to complete a repair is approximately normal with mean 70 minutes and standard deviation 6.5 minutes. (a) Find the probability that a repair takes between 65 and 80 minutes. (b) The slowest 15% of repairs are flagged for retraining. Find the cutoff time.

Model Response

(a) Let X be repair time, $X \sim N(70, 6.5)$. $P(65 < X < 80) = \text{normalcdf}(65, 80, 70, 6.5) = 0.717$. About 71.7% of repairs take between 65 and 80 minutes. (b) The slowest 15% lie above the 85th percentile, so the area to the left of the cutoff is 0.85. Cutoff $= \text{invNorm}(0.85, 70, 6.5) = 76.7$ minutes. Repairs that take longer than about 76.7 minutes are flagged.

What the grader looks for

(1) A named model with the mean and standard deviation. (2) The calculator function written with all four inputs. (3) The correct probability 0.717 stated in context. (4) For part (b), the correct area 0.85 (not 0.15) in invNorm. (5) A final answer with units and a sentence in context.

Coach's Playbook

1. **Name the model first.** Write the distribution (normal, binomial, t) with its parameters before touching the calculator.
2. **Sketch and shade.** A quick curve with the shaded region shows whether you need left, right or between.
3. **Enter, then say it.** Write the function with its arguments, then write the answer as a sentence in context.
4. **Check reasonableness.** A probability must be between 0 and 1. A normal cutoff for the top 10% should be above the mean.
5. **Round at the end.** Keep full calculator precision internally and round the final answer to three decimals (or as asked).
6. **Know how to store a list.** Practice entering data into lists and clearing them between problems so old data does not creep into new computations.

A short practice routine. Spend ten minutes a day for a week on a drill: pick a context, state the model, and run the matching function without notes. On day one, do only normalcdf and invNorm. Add binomial functions on day three, the t functions on day five, and the proportion and regression procedures on day seven. By the end of the week the menu locations should be automatic, which is exactly the point: on test day your attention belongs to the reasoning, the conditions and the written conclusion, not to hunting through menus.

When the tool and the formula disagree. If your calculator answer differs from a hand computation in the third decimal place, check rounding first: a z-score rounded to two decimals changes a normal area slightly, while the tool uses full precision. If the difference is large, look for a wrong input order, a wrong tail, or a count entered as a proportion. Either answer is acceptable on the exam when the setup is clear and the rounding is reasonable.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Common traps

1. Entering a proportion where the tool wants a count (x = 87, not 0.435). 2. Using the left-tail area for an upper-tail cutoff in invNorm. 3. Reading the population standard deviation instead of the sample standard deviation from 1-Var Stats. 4. Writing only the answer and not the function and inputs.

Mistake Clinic

Spot the error

A student finds P(X >= 5) for a Binomial(12, 0.3) with binomcdf(12, 0.3, 5) and writes 0.882. Where does this fail? binomcdf gives P(X <= 5). Correct: $P(X \geq 5) = 1 - \text{binomcdf}(12, 0.3, 4) = 1 - 0.724 = 0.276$.

Spot the error

To find the cutoff for the top 10% of scores in N(500, 80), a student enters invNorm(0.10, 500, 80) and reports 397.5. Where does this fail? invNorm takes the area to the left, and 0.10 is the bottom 10%. Correct: invNorm(0.90, 500, 80) = 602.5.

Spot the error

A student reads the standard deviation of a sample from 1-Var Stats as the population version and reports 7.98 for data whose sample value is 8.53. Where does this fail? A sample of 8 students estimates the population variation, so use the sample standard deviation (Sx). The population version divides by n instead of n - 1 and is too small.

Real-World Connection

Quality engineers, public-health analysts and sports scientists all run the same functions on real data every day. Fluency means a question like "what share of items are within tolerance?" becomes a single line, freeing time to ask whether the model is reasonable and what the result means.

Challenge Corner

A placement test has scores approximately normal with mean 500 and standard deviation 80. Find the middle 50% range and the interquartile range of the model. **Solution:** Q1 = invNorm(0.25, 500, 80) = 446.0 and Q3 = invNorm(0.75, 500, 80) = 554.0, so the middle 50% runs from 446.0 to 554.0 and the IQR is 107.9 (about 1.349 standard deviations). This is a helpful check: the IQR of a normal model is about 1.35 times its standard deviation, which lets you spot data that are not normal.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE **Calculator Skills for Statistics**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Scores are approximately N(500, 80). Which calculator entry finds P(X < 400)?
   (A) normalcdf(400, 1E99, 500, 80)
   (B) normalcdf(-1E99, 400, 500, 80)
   (C) invNorm(400, 500, 80)
   (D) binomcdf(400, 500, 80)
2. For N(500, 80), which entry gives the 25th percentile?
   (A) normalcdf(0, 0.25, 500, 80)
   (B) invNorm(0.75, 500, 80)
   (C) invNorm(0.25, 500, 80)
   (D) normalcdf(-1E99, 0.25, 500, 80)
3. For X ~ Binomial(12, 0.3), which expression equals P(X >= 5)?
   (A) binomcdf(12, 0.3, 5)
   (B) 1 - binomcdf(12, 0.3, 5)
   (C) binompdf(12, 0.3, 5)
   (D) 1 - binomcdf(12, 0.3, 4)
4. A least-squares line for practice hours and score is $\hat{y} = 51.96 + 3.369x$, built from data with x from 1 to 8. Which is the best prediction for x = 10?
   (A) 85.7 points, but this is extrapolation and should be treated with caution
   (B) 85.7 points, and it is reliable because r = 0.97
   (C) 33.7 points
   (D) 55.3 points
5. For a t-test with df = 17 and t = 2.05, the upper-tail area is about 0.028. Which entry gives it?
   (A) tcdf(2.05, 1E99, 17)
   (B) tcdf(-1E99, 2.05, 17)
   (C) tcdf(2.05, 1E99, 18)
   (D) normalcdf(2.05, 1E99, 0, 1)
6. A sample has 87 successes out of 200, and you test $H_0 : p = 0.4$. Which entries are correct?
   (A) p0 = 0.435, x = 87, n = 200
   (B) p0 = 0.4, x = 0.435, n = 200
   (C) p0 = 0.4, x = 87, n = 87
   (D) p0 = 0.4, x = 87, n = 200
7. 1-Var Stats shows both a sample standard deviation (Sx) and a population standard deviation. For a random sample used to estimate a population's variation, you should report
   (A) the population standard deviation
   (B) the mean
   (C) Sx, the sample standard deviation
   (D) the larger of the two
8. For a 95% one-sample t-interval with n = 18, which is the correct critical value?
   (A) 1.96
   (B) 2.110
   (C) 1.740
   (D) 2.567
9. A sample of 18 has mean 52.4 and s = 9.1. The 95% t-interval is closest to
   (A) (47.87, 56.93)
   (B) (48.20, 56.60)
   (C) (43.30, 61.50)
   (D) (52.40, 56.60)
10. Delivery times are approximately normal with mean 70 minutes and standard deviation 6.5 minutes.
    (a) Find the probability that a delivery takes between 65 and 80 minutes.
    (b) Find the probability that it takes more than 80 minutes.
    (c) Find the time that marks the slowest 15%.
11. A list of 10 weekly step counts (in hundreds) is 14, 18, 21, 22, 25, 27, 29, 33, 35, 40.
    (a) Use 1-Var Stats to report the mean, sample standard deviation, median, Q1 and Q3.
    (b) State which function output you would use for the sample standard deviation.
12. The data (x, y) are (2, 31), (3, 36), (5, 38), (6, 45), (8, 47), (9, 55), (11, 58).
    (a) Use LinReg to find the line, r and r-squared.
    (b) Predict y when x = 7.
    (c) Find the residual for the point (5, 38).
    (d) Interpret r-squared.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. A binomial random variable X has n = 20 and p = 0.15.
    (a) Find P(X <= 2).
    (b) Find P(X >= 4).
    (c) Find P(X = 3).
    (d) Find the mean and standard deviation of X.
14. A random sample of 15 batteries has mean life 12.6 hours and s = 3.1 hours. Test $H_0 : \mu = 11$ versus $H_a : \mu > 11$ at alpha = 0.05. Assume the conditions are met. Report the function used, the t-statistic, the p-value and a conclusion in context.
15. In a random sample of 150 residents, 64 favor a new bike lane.
    (a) Construct a 90% confidence interval for the true proportion in favor using 1-PropZInt.
    (b) Test $H_0 : p = 0.5$ versus $H_a : p \neq 0.5$ at alpha = 0.05 with 1-PropZTest. State the z and p-value and the conclusion in context.

Answers and Explanations

Complete solutions for the practice items in these opening chapters.

F.1 How the AP Statistics Exam Works and How to Use This Book

PRACTICE · HOW THE AP STATISTICS EXAM WORKS AND HOW TO USE THIS BOOK

1. C. The multiple-choice section has 42 questions, each with 4 answer choices. Choices with 5 options or 36, 40 or 45 questions mix up other tests' formats.
2. A. $90/4 = 22.5$ minutes. The value 18 comes from dividing 90 by 5 questions, which is a different format.
3. D. Section II is 50% of the exam score. The 25% figure is wrong because each question is 12.5% of the exam, and 40% and 60% do not match the equal weights.
4. B. The exam is fully digital, a graphing calculator is allowed, and formulas and tables are provided. The other options reverse or invent restrictions.
5. C. The geometric distribution is on the removed list. The binomial distribution is similar-sounding and tempting, but it remains in the course, as do the normal distribution and proportion intervals.
6. A. A value computed from a sample is a statistic. Choosing parameter confuses the sample result with the unknown value for all voters.
7. B. Number of books is a count with numerical meaning. Zip codes and jersey numbers are labels that look like numbers but cannot be averaged meaningfully, and eye color is categorical.
8. D. MC part: $36 \cdot 50/42 = 42.857$. FR part: $24 \cdot 50/40 = 30$. The total is 72.857, about 72.9. The value 60 comes from adding 36 and 24 without scaling.
9. A. A confidence interval for a population mean is inference for quantitative data, Exam Unit 4. Unit 3 is for proportions, which is the tempting wrong choice.
10. (a) $55/42 \approx 1.31$ minutes, about 1 minute 19 seconds per question. (b) $90 - 55 = 35$ minutes remain for review, which is more than needed; a better plan keeps about 10 minutes for review and uses 80 minutes (about 1.9 minutes per question) for answering.
11. (a) Favorite sport: categorical. Hours of homework: quantitative. Number of pets: quantitative. Grade level: categorical (labels for groups, even though they are numbers). Owns a bike: categorical.
    (b) Grade level is the poor candidate for a mean, since the labels 9 to 12 name groups and the average grade level does not describe a typical amount of anything.
12. (a) FR total: $6 + 8 + 5 + 6 = 25$. Composite: $28 \cdot 50/42 + 25 \cdot 50/40 = 33.33 + 31.25 = 64.58$, about 64.6.
    (b) The free-response part gives 31.25 points and the multiple-choice part gives 33.33 points, so the multiple-choice section contributes about 2.1 more points.
13. A full-credit answer states: Week 1 relearns t conditions, intervals and tests (Unit 8 lessons 1 to 3), weekly check is a full t-interval with conclusion; Week 2 covers two-sample and paired procedures (lessons 4 to 6), check is choosing between two-sample and paired; Week 3 mixes Units 7 and 8 with C.1 for procedure selection; Week 4 is two practice exams and the mistake log. Unit 4 gets time in every week, with the heaviest load first. A good answer gives a reason (the weak unit is placed early and revisited) and a measurable check each week.
14. The removed topics are: analyzing departures from linearity, combining random variables, the geometric distribution, the chi-square goodness-of-fit test, and inference for regression slopes. A good decision statement for one topic: "I will study the geometric distribution after finishing Units 1 to 8, because it is listed as removed and my teacher's current course description will confirm whether it is needed."
15. (a) Population: all 900 students. Sample: the 90 chosen students. Variable: how the student gets to school, or walks (yes or no), which is categorical.
    (b) $\hat{p} = 27/90 = 0.30$, a statistic.
    (c) Another random sample of 90 would likely give a different proportion, so 0.30 is only an estimate of the unknown population proportion p; the parameter could be above or below it.

F.2 Writing Statistical Answers That Earn Full Credit

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

PRACTICE · WRITING STATISTICAL ANSWERS THAT EARN FULL CREDIT

1. B. This is the standard interpretation of a confidence interval: confidence in the method capturing the true mean. Choice A confuses the interval for a mean with the spread of individual values, which is the most tempting error.
2. C. The slope is a predicted change per year, and an average trend rather than an exact value. A is wrong because it says exactly and implies each car follows the line.
3. D. A p-value is computed assuming H_0 is true and describes how surprising the data are. A reverses the conditional and is the usual misconception.
4. A. The confidence level describes the long-run capture rate of the method. B treats the interval as random after it is computed, but the parameter is fixed.
5. B. r^2 is the proportion of variation in y accounted for by the linear relationship. B confuses r^2 with r; the correlation would be 0.70.
6. C. It compares centers with a number and unit, and it compares spread with numbers. A lists medians with no comparison word, C gives no numbers, and D does not say which class.
7. D. With n = 14 < 30, the sample must show no strong skewness or outliers. A is a proportion condition and B is the wrong rule of thumb for small samples; the population standard deviation is not known.
8. B. Hypotheses are about the parameter μ, defined in context. A uses a statistic, and the null is the equal sign.
9. A. A large p-value gives no evidence for the alternative, so we fail to reject. Accepting H0 overstates what the test shows, and the 0.82 is not the probability that H0 is true.
10. (a) We are 95% confident that the true mean amount of sleep for teens at this school is between 6.8 and 7.6 hours.
    (b) Yes. Since 8 hours is not in the interval, there is convincing evidence that the mean differs from 8 hours (it appears to be less).
    (c) If we took many random samples of this size and built a 95% interval from each, about 95% of those intervals would capture the true mean.
11. (a) Random: assumed from a random sample of customers; 10% rule: 150 is less than 10% of all customers, assuming the cafe has more than 1,500 customers; large counts: $\hat{p} = 0.30$, $n\hat{p} = 45 \geq 10$, $n(1-\hat{p}) = 105 \geq 10$.
    (b) $SE = \sqrt{0.3(0.7)/150} = 0.0374$, interval $0.30 \pm 1.96(0.0374) = (0.227, 0.373)$.
    (c) We are 95% confident that the true proportion of all customers who use the mobile app is between 0.227 and 0.373.
12. (a) For each additional ad shown, the predicted number of clicks increases by about 2.5.
    (b) When 0 ads are shown, the predicted number of clicks is 18.4. Because x ranges from 5 to 40, 0 is outside the data, so the intercept is an extrapolation with no practical meaning.
    (c) $r^2 = 0.78^2 = 0.6084$: about 61% of the variation in clicks is accounted for by the linear relationship with number of ads shown.
    (d) Predictions of clicks from the line are typically off by about 6.1 clicks.
13. (a) If the true proportion of households that recycle were 0.40, the probability of getting a sample proportion at least as large as the one observed, by chance alone, is 0.034.
    (b) Since 0.034 < 0.05, we reject H_0. There is convincing evidence that more than 40% of the city's households recycle.
    (c) At alpha = 0.01, 0.034 > 0.01, so we fail to reject H_0: there is not convincing evidence that the proportion exceeds 0.40.
14. Section 1: Q1 = 70.5, median = 75, Q3 = 81, IQR = 10.5, range 33, mean 76.2, sd 9.2. Section 2: Q1 = 71, median = 75, Q3 = 78, IQR = 7, range 13, mean 74.7, sd 4.2. Fences: Section 1 (54.75, 96.75) and Section 2 (60.5, 88.5), so neither has outliers.
    Shape: Section 1 is skewed right with a high score of 95; Section 2 is roughly symmetric.
    Outliers: Neither section has outliers by the 1.5 IQR rule.
    Center: Both sections have the same median score of 75 points.
    Spread: Section 1 is more variable, with an IQR of 10.5 points compared with 7 points for Section 2, and a range of 33 points compared with 13.
15. Error 1: The p-value is not the probability that H_0 is true; it is the probability of results at least this extreme if H_0 is true. Error 2: A significance test does not prove a claim false; it gives evidence. Correct conclusion: Because the p-value of 0.03 is less than alpha = 0.05, we reject H_0. There is convincing evidence that the true mean battery life is less than 50 hours.

F.3 Calculator Skills for Statistics

PRACTICE · CALCULATOR SKILLS FOR STATISTICS

1. B. The area below 400 is the area from negative infinity to 400, which is normalcdf(-1E99, 400, 500, 80) = 0.106. Choice A gives the upper tail (0.894), which is the tempting wrong tail.
2. C. invNorm takes the area to the left, 0.25, and returns the value 446.0. Choice B would return the 75th percentile, and the normalcdf options mix area and value.
3. D. binomcdf(12, 0.3, 4) = P(X <= 4), so P(X >= 5) = 1 - 0.724 = 0.276. Choice B subtracts P(X <= 5), which also removes P(X = 5).
4. A. The line predicts 85.65 for x = 10, but 10 is outside the data range of 1 to 8, so it is extrapolation. B treats a high r as a guarantee outside the data.
5. A. The upper tail is tcdf(2.05, 1E99, 17) = 0.028. Choice A is the area to the left (0.972), and choice C uses the wrong degrees of freedom.
6. D. Enter the hypothesized proportion 0.4, the count 87, and n = 200. Choice A puts the sample proportion in p0, and choice B puts a proportion where the count belongs.
7. C. Sx is the sample standard deviation and estimates the population value; the population version divides by n and is too small for a sample.
8. B. invT(0.975, 17) = 2.110. The value 1.96 is the z critical value, which is the tempting error that ignores the extra uncertainty in s.
9. A. $52.4 \pm 2.110(9.1/\sqrt{18})$ gives (47.87, 56.93). Choice B uses z* = 1.96 and gives (48.20, 56.60), which is too narrow.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. (a) normalcdf(65, 80, 70, 6.5) = 0.717. About 71.7% of deliveries take between 65 and 80 minutes.
(b) normalcdf(80, 1E99, 70, 6.5) = 0.062. About 6.2% take longer than 80 minutes.
(c) The slowest 15% are above the 85th percentile: invNorm(0.85, 70, 6.5) = 76.7 minutes.
11. (a) Mean = 26.4, Sx = 8.06, median = 26, Q1 = 21, Q3 = 33 (hundreds of steps).
(b) Use Sx, the sample standard deviation, because these are a sample; sigma-x would divide by n and be smaller.
12. (a) LinReg gives $\hat{y} = 25.41 + 3.002x$, r = 0.980, $r^2 = 0.960$.
(b) At x = 7: 25.41 + 3.002(7) = 46.43.
(c) Predicted at x = 5: 40.42. Residual = 38 - 40.42 = -2.43 (the point is below the line).
(d) About 96% of the variation in y is accounted for by the linear relationship with x.
13. (a) binomcdf(20, 0.15, 2) = 0.405.
(b) 1 - binomcdf(20, 0.15, 3) = 1 - 0.648 = 0.352.
(c) binompdf(20, 0.15, 3) = 0.243.
(d) Mean = 20(0.15) = 3; standard deviation = sqrt(20(0.15)(0.85)) = 1.597.
14. Function: T-Test with summary statistics (mean 12.6, s 3.1, n 15, mu0 11, > mu0). $t = \frac{12.6-11}{3.1/\sqrt{15}} = 2.00$, df = 14, p-value = 0.033. Because 0.033 < 0.05, reject H_0. There is convincing evidence that the true mean battery life exceeds 11 hours.
15. (a) 1-PropZInt with x = 64, n = 150, C-level 0.90 gives (0.360, 0.493). We are 90% confident that the true proportion of residents in favor is between 0.360 and 0.493.
(b) $\hat{p} = 0.427$; 1-PropZTest with $p_0 = 0.5$: $z = -1.80$, p-value = 0.072. Since 0.072 > 0.05, we fail to reject H_0: there is not convincing evidence that the proportion in favor differs from 0.5.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 1: Exploring One-Variable Data

Describe distributions of one variable with graphs, numbers, and words. 6 lessons, 104 practice and test items, all with full explanations.

In this unit · Exam Unit 1

1.1 Data, Variables, and Categorical Graphs
1.2 Graphs for Quantitative Data: Dotplots, Stemplots, and Histograms
1.3 Describing Shape, Center, and Spread
1.4 The Five-Number Summary, Boxplots, and Outliers
1.5 Comparing Distributions in Context
1.6 Percentiles, z-Scores, and Transforming Data
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

1.1 Data, Variables, and Categorical Graphs

Learn to name individuals and variables, tell categorical from quantitative, and summarize categorical data with relative frequencies, bar graphs, and careful comparisons.

Why This Matters

Every statistical story begins with the same two questions: who or what is being measured, and what is being recorded about each one? Students who rush past these questions later pick the wrong graph, compute a mean of zip codes, or compare two groups using raw counts when the groups have different sizes. This lesson builds the vocabulary and the habits that the rest of the course depends on.

On the AP exam, this material lives in **Exam Unit 1: Exploring One-Variable Data**. Multiple-choice questions ask you to classify a variable, read a bar graph or relative frequency table, or spot a misleading display. Free-response questions reward a clear description of a categorical distribution that uses percents and names the categories in context.

Core Ideas

A **data set** is a collection of information about **individuals**. An individual can be a person, an animal, a school, a batch of cookies, or a single sensor reading. A **variable** is a characteristic that can take different values for different individuals. A spreadsheet usually has one row per individual and one column per variable.

Types of Variables

Categorical variable: places each individual into a group or category. Examples: favorite subject, type of phone, grade level, zip code.

Quantitative variable: takes numerical values for which arithmetic such as adding or averaging makes sense. Examples: commute time, number of siblings, height, battery life.

Quantitative variables are **discrete** when they take countable separate values (number of siblings) and **continuous** when they can take any value in an interval (commute time in minutes).

Test: if the average of the values would be meaningful, the variable is quantitative. A number used as a label (zip code, jersey number, student ID) is categorical.

Notice that the type of a variable depends on how it is recorded. Age in whole years is quantitative, but the category "under 13, 13 to 17, 18 and over" is categorical. Grade level 9 to 12 is a label for a category with a natural order, which some books call ordinal; for AP Statistics the key point is that you summarize it with counts and percents.

A **distribution** tells you what values a variable takes and how often it takes them. For a categorical variable, the distribution is a list of the categories with their counts or percents. A **frequency table** lists counts. A **relative frequency table** lists the proportion or percent of individuals in each category, found by dividing each count by the total.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

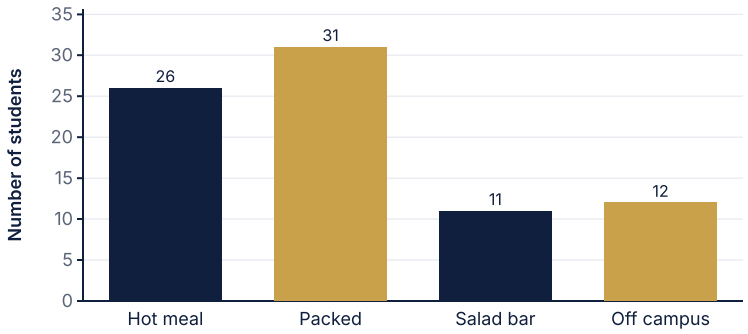
© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Relative Frequency

$$\text{relative frequency} = \frac{\text{count in the category}}{\text{total number of individuals}}$$

The relative frequencies of all categories add to 1 (or 100 percent), up to rounding. If they do not, check for a missing category or for a variable where individuals can fall in more than one category.

Two graphs display a categorical distribution. A **bar graph** puts categories on the horizontal axis and counts or percents on the vertical axis; bars are separated by gaps because the categories are separate, and they may be arranged in any order. A bar graph in which the bars are ordered from tallest to shortest is sometimes called a Pareto chart and makes the most common categories easy to find. A **pie chart** shows each category as a slice proportional to its share of the whole, so it works only when the categories together make up 100 percent of one group of individuals.

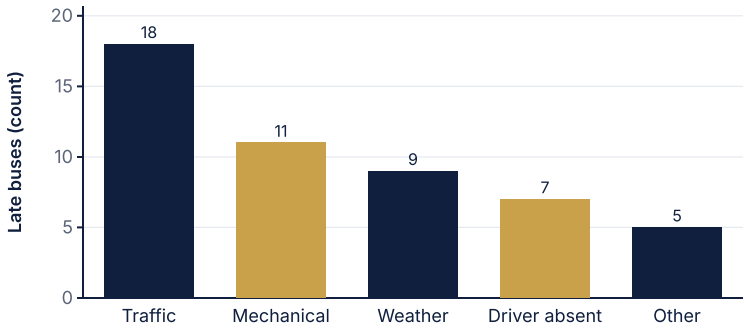


Bar graph of lunch choice for 80 students (counts).

Describing a Categorical Distribution

Name the variable and the group of individuals in context.
Give percents (or proportions), not just counts, whenever the group size matters.
Identify the most common and least common categories and comment on any notable pattern.
When comparing groups of different sizes, always compare **percents**, never raw counts.

Graphs can mislead. A bar graph whose vertical axis starts at a value other than zero makes small differences look large because the bar lengths no longer match the data. Three-dimensional bars, pictures whose area grows with the value, and pie charts with too many thin slices also distort. When you read a graph, check the axis scale and the labels before you trust your first impression.



Ordered bar graph of reasons for 50 late buses over one term.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: CLASSIFY THE VARIABLES

A school office records the following for each student: student ID number, grade level (9, 10, 11, or 12), minutes spent commuting, number of siblings, favorite subject, and zip code. Classify each variable.

SOLUTION

Student ID number: categorical. It is only a label; averaging IDs means nothing.

Grade level: categorical (ordered categories). The numbers name groups.

Minutes commuting: quantitative, continuous, because time can take any value in an interval.

Number of siblings: quantitative, discrete, because it takes whole-number values only.

Favorite subject: categorical.

Zip code: categorical, even though it is written with digits.

EXAMPLE 2: BUILD A RELATIVE FREQUENCY TABLE

Eighty students were asked how they usually get lunch: 26 buy the hot meal, 31 bring a packed lunch, 11 use the salad bar, and 12 go off campus. Make a relative frequency table and state the most common choice as a percent.

Choice	Count	Relative frequency
Hot meal	26	0.325
Packed	31	0.3875
Salad bar	11	0.1375
Off campus	12	0.15

SOLUTION

Check the total: $26 + 31 + 11 + 12 = 80$.

Divide each count by 80: hot meal $26/80 = 0.325$; packed $31/80 = 0.3875$; salad bar $11/80 = 0.1375$; off campus $12/80 = 0.15$.

The percents are 32.5, 38.75, 13.75, and 15, which add to 100.

Conclusion: The most common choice is a packed lunch, chosen by about 38.8 percent of the 80 students surveyed.

EXAMPLE 3: WHICH GRAPH IS APPROPRIATE?

In a survey of 60 students, each student was asked, "Which of these do you do at least once a week? Select all that apply." The counts were: reading 42, gaming 38, sports 29. A classmate wants a pie chart. Is that appropriate?

SOLUTION

The percents are $42/60 = 70\%$, $38/60 \approx 63.3\%$, and $29/60 \approx 48.3\%$.

These add to about 181.7 percent, far above 100 percent, because one student can appear in several categories.

A pie chart shows parts of a single whole, so it is **not appropriate**.

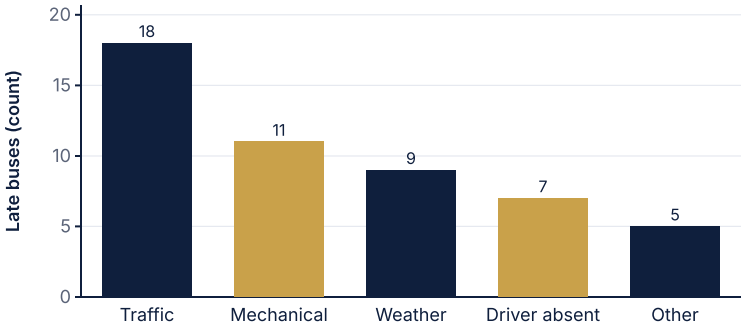
A bar graph of the percent of the 60 students who do each activity is appropriate, with each bar scaled out of 60.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 4: READ AN ORDERED BAR GRAPH

Use the bar graph of late buses above. What percent of the 50 late buses were late because of traffic or mechanical problems?



SOLUTION

From the graph, traffic accounts for 18 buses and mechanical problems for 11.

Together that is $18 + 11 = 29$ buses.

$29/50 = 0.58$.

Conclusion: About 58 percent of the late buses were late because of traffic or mechanical problems, which are the two most common reasons.

EXAMPLE 5: IS THE GRAPH MISLEADING?

A news site shows a bar graph of customer satisfaction: Brand A at 94 percent and Brand B at 90 percent. The vertical axis starts at 88 percent, so Brand A's bar is 6 units tall and Brand B's bar is 2 units tall. What impression does the graph give, and what is the truth?

SOLUTION

Bar heights are 6 and 2, so Brand A's bar looks $6/2 = 3$ times as tall as Brand B's.

The actual ratio of satisfaction is $94/90 \approx 1.044$, only about 4.4 percent higher.

The difference between the brands is 4 percentage points, which may or may not matter, but it is not a threefold difference.

Conclusion: The truncated axis exaggerates the gap. A fair bar graph starts the axis at 0 so that bar lengths are proportional to the values.

EXAMPLE 6: COMPARE GROUPS OF DIFFERENT SIZES

Northside High surveyed 120 students and Eastfield High surveyed 300 students about how they get to school. Northside: bus 54, walk 18, car 36, bike 12. Eastfield: bus 105, walk 60, car 99, bike 36. Which school has a greater share of bus riders, and how do the distributions differ?

Mode	Northside (n=120)	Eastfield (n=300)
Bus	54	105
Walk	18	60
Car	36	99
Bike	12	36

SOLUTION

Counts are misleading because the schools have different numbers of respondents, so convert to percents.

Northside: bus $54/120 = 45\%$, walk 15%, car 30%, bike 10%.

Eastfield: bus $105/300 = 35\%$, walk 20%, car 33%, bike 12%.

Conclusion: A larger share of Northside students ride the bus (45 percent versus 35 percent). Eastfield has larger shares of walkers (20 versus 15 percent), car riders (33 versus 30 percent), and bikers (12 versus 10 percent). Even though Eastfield has more bus riders in raw count (105 versus 54), the percent is lower.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Write It Like a 5

Model Response

Prompt. A school surveyed 150 students about how they mainly listen to music: streaming 87, radio 15, downloaded files 33, physical media 15. (a) Name an appropriate graph and describe how to draw it. (b) Describe the distribution in context. (c) A classmate computes the mean of the four counts, 37.5, and says a typical student listens through a source with about 37.5 students. Explain the error. **(a)** The variable, main way of listening, is categorical, so a bar graph is appropriate. Put the four methods on the horizontal axis with gaps between bars, and scale the vertical axis from 0 to at least 87 (or use percent from 0 to 60) with a label. **(b)** The percents are streaming 58 percent, downloaded files 22 percent, radio 10 percent, and physical media 10 percent, which add to 100 percent. Streaming is the most common way to listen among the 150 students surveyed, used by more than half of them. Radio and physical media are the least common and tie at 10 percent each. **(c)** The mean of the counts is the average size of a category, not a value of the variable. The variable is categorical, so there is no typical numerical value. The best summary of a typical response is the most common category, streaming.

What the grader looks for

(1) Correctly classifies the variable as categorical and chooses a bar graph. (2) Mentions labeled axes with a sensible scale starting at zero. (3) Converts counts to percents with the correct denominator of 150. (4) Names the most common and least common categories in context, not just in numbers. (5) Explains that a mean of counts has no meaning for a categorical variable.

Coach's Playbook

1. **Name the individuals first.** Say out loud who or what each row is. Then name the variable, and decide whether its values are labels or measurements.
2. **Apply the averaging test.** If the average of the recorded values would be meaningful, the variable is quantitative. If the digits only name groups, it is categorical.
3. **Find the denominator.** Before computing a percent, decide the group that the percent describes. The denominator is the size of that group.
4. **Check the total.** Relative frequencies should add to 100 percent. If they do not, a category is missing or individuals can be in several categories.
5. **Inspect the axis.** Look at where the vertical axis starts and whether the bars or slices match the numbers printed on them.
6. **Write in context.** Finish every description with the variable name, the group, and a comparison phrase such as "more common than" or "a larger share of".

Calculator tip: you can enter counts in a list and divide the list by its sum to get all relative frequencies at once. Store the result in a second list so you can check that it sums to 1. On the exam, you will not be asked to draw a pie chart by hand, but you may be asked to say whether one is appropriate.

Common traps

(1) Calling a number a quantitative variable just because it is written with digits, such as a zip code or a jersey number. (2) Comparing two groups by raw counts when the groups have different sizes. (3) Using a pie chart for a select-all-that-apply question, where percents add to more than 100. (4) Reading a bar graph without checking whether the vertical axis starts at zero.

Mistake Clinic

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Student claim: "Zip code is a quantitative variable because it is made of numbers, so I can find the average zip code of my customers." **Where it fails:** the average of zip codes has no meaning; a zip code only names a region. Adding two zip codes does not produce a location. **Correct solution:** zip code is categorical. Summarize it with counts or percents per region, shown in a bar graph.

Spot the error

Student claim: "Eastfield has 105 bus riders and Northside has only 54, so Eastfield has more students who ride the bus." **Where it fails:** the claim compares counts of two groups of different sizes (300 and 120 respondents) and so does not answer which school has a larger share. **Correct solution:** Northside has $54/120 = 45\%$ bus riders and Eastfield has $105/300 = 35\%$, so a larger share of Northside students ride the bus.

Spot the error

Student claim: "A student surveyed 40 people and recorded these percents of people who like each sport: 55 percent, 45 percent, 30 percent. I made a pie chart." **Where it fails:** the percents add to 130, so people picked more than one sport and the categories are not parts of one whole. **Correct solution:** use a bar graph with each bar showing the percent of the 40 people who like that sport; the bars do not need to sum to 100 percent.

Real-World Connection

Streaming services, schools, and health departments all start with the same step you practiced: classify the variables, then summarize categorical ones with percents. A public health team tracking flu vaccine uptake in two counties of very different populations must compare percents vaccinated, not counts, or it will draw the wrong conclusion about which county needs more outreach.

Challenge Corner

A bar graph of 200 survey responses shows three bars with heights 90, 70, and 40 on an axis that begins at 30. The tops of the bars appear to be in the ratio 60 : 40 : 10. (a) What are the true ratios of the counts? (b) Another bar for "Other" has a true count of 30 percent of the number in the tallest bar. Where would the top of its bar sit on the truncated axis, relative to the axis? **Solution.** (a) The true ratio is 90 : 70 : 40, or 9 : 7 : 4. The drawn heights are 90 - 30 = 60, 70 - 30 = 40, and 40 - 30 = 10, so the tallest bar looks 6 times as tall as the shortest while the true factor is 90/40 = 2.25. (b) 30 percent of 90 is 27, which is below the axis start of 30, so the bar would not appear at all; the graph hides the category completely. A zero-based axis avoids both problems.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE **Data, Variables, and Categorical Graphs**

Show your work. A calculator is allowed. Justify conclusions in context.

1. The number of text messages a student sends in one day is best classified as
   (A) a categorical nominal variable
   (B) a categorical ordinal variable
   (C) a quantitative discrete variable
   (D) a quantitative continuous variable

2. Which of the following is a categorical variable?
   (A) the area code of a phone number
   (B) the height of a player in centimeters
   (C) the number of goals scored in a match
   (D) the time to finish a 5 km run

3. In a survey of 60 students, 21 chose A, 27 chose B, and 12 chose C. What percent chose C?
   (A) 12 percent **(B)** 25 percent **(C)** 40 percent
   (D) 20 percent

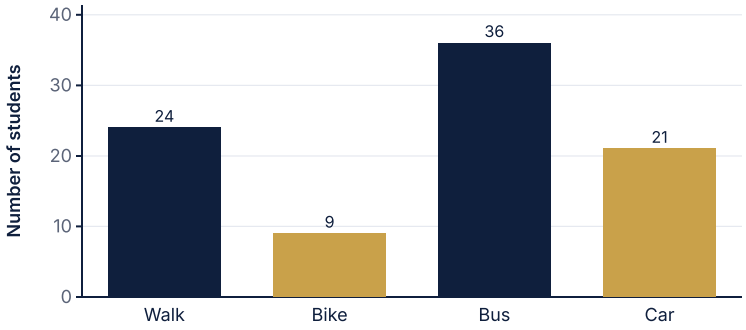
4. A survey asks students to select every after-school activity they do. The percents for sports, music, and clubs are 55, 40, and 35. Which statement is correct?
   (A) A pie chart is appropriate because there are only three categories.
   (B) A pie chart is not appropriate because the percents add to more than 100.
   (C) A pie chart is appropriate if the slices are labeled with percents.
   (D) A bar graph is not appropriate because the data are categorical.

5. A bar graph shows two teams' win percentages as 52 percent and 48 percent. The vertical axis starts at 46 percent. Which is the best description?
   (A) The bars correctly show that the first team won twice as often.
   (B) The graph is fine because both percents are between 0 and 100.
   (C) The first bar looks three times as tall as the second, exaggerating a 4 percentage point difference.
   (D) The second bar is hidden because 48 is below 50.

6. Which variable is quantitative?
   (A) a player's jersey number
   (B) a player's height in centimeters
   (C) a player's position
   (D) a player's team name

7. Club A has 18 of its 40 members who play an instrument. Club B has 27 of its 75 members who play an instrument. Which statement is correct?
   (A) Club B has a larger share because 27 is greater than 18.
   (B) The clubs have equal shares because both counts are under 30.
   (C) A larger share of Club B members play, 36 percent versus 45 percent.
   (D) A larger share of Club A members play an instrument, 45 percent versus 36 percent.

8. The table shows the main transport of 90 students. What percent of the students do NOT ride the bus?



   (A) 60 percent **(B)** 40 percent **(C)** 36 percent
   (D) 24 percent

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. Which is true of a bar graph for a categorical variable?
   (A) The bars must touch because the values are continuous.
   (B) The categories must be ordered from smallest to largest.
   (C) The bars are separated by gaps and can be reordered without changing the information.
   (D) The bar heights must add to 100 percent.
10. A fitness app stores these variables for each user: user ID, age in years, number of workouts last week, preferred workout type (run, bike, swim, other), and resting heart rate in beats per minute. (a) Classify each variable as categorical or quantitative. (b) For each quantitative variable, say whether it is discrete or continuous.
11. A cafe recorded the drink ordered by 200 customers: coffee 84, tea 46, juice 30, smoothie 40. (a) Make a relative frequency table. (b) Describe the distribution in context.
12. A graph of the number of customers for two stores uses a vertical axis that starts at 900. Store X has 960 customers and Store Y has 1,020. The bars appear to have heights 60 and 120. (a) How many times as tall does Store Y's bar appear? (b) What is the actual ratio of customers? (c) Explain why the graph is misleading.
13. Two online classes were polled on whether they prefer morning or evening live sessions. Class 1 (n = 80): morning 28, evening 52. Class 2 (n = 150): morning 66, evening 84. Which class has the larger share preferring mornings? Support your answer with percents.
14. In a poll of 75 teens, each was asked to name all the apps they use daily. Percents: video 80, messaging 72, games 44. (a) Explain why a pie chart is inappropriate. (b) Describe a better graph and compute the number of teens in each bar.
15. A factory logged 120 defective items by cause: wrong size 48, scratch 30, color 24, missing part 12, other 6. (a) Make an ordered bar graph description with percents. (b) What percent of defects come from the two most common causes? (c) Is it correct to say the typical defect count per cause is 24? Explain.

1.2 Graphs for Quantitative Data: Dotplots, Stemplots, and Histograms

Build and read dotplots, stemplots (including split stems and back-to-back displays), and histograms, and describe what each graph shows about a quantitative variable.

Why This Matters

A list of 40 numbers tells you almost nothing at a glance. A good graph turns the list into a picture of how the values are spread out, where they cluster, and which ones look unusual. The skill is twofold: build the graph correctly, and read information from it quickly and accurately.

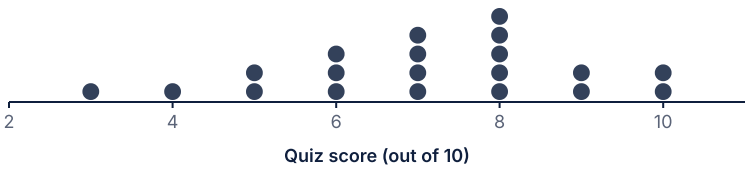
This lesson belongs to **Exam Unit 1: Exploring One-Variable Data**. Multiple-choice questions ask you to read a value, a count, or a percent from a dotplot, stemplot, or histogram, or to decide which statement about the graph is true. Free-response questions show you a graph and ask for a description in context, so the graph vocabulary here feeds directly into the next two lessons.

Core Ideas

Three graphs display a quantitative variable. They differ in how much of the original data they keep.

Dotplot

A **dotplot** puts a dot above a number line for each observation; repeated values stack.
Best for small to moderate data sets (up to a few dozen values) with values that repeat.
Every individual value can be read, so no information is lost.



Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Stemplot (Stem-and-Leaf Plot)

Split each value into a **stem** (all digits except the last) and a **leaf** (the final digit). Write stems in a column, in order, and put each leaf beside its stem in increasing order.

Always include a **key** such as "2 | 3 means 23 minutes" and keep leaves evenly spaced.

Include every stem from smallest to largest, even stems with no leaves, so that gaps stay visible.

Split stems give each stem two rows, leaves 0 to 4 on the first and 5 to 9 on the second, when one row per stem is too crowded.

A **back-to-back stemplot** shares the stems down the middle with one data set's leaves on each side, which allows a direct comparison. Leaves on the left run outward from the stem in decreasing order.

```
  1 | 2 5 7 8
  2 | 1 2 2 4 5 7 8
  3 | 1 3 4 6 8
  4 | 1 5 7
  5 |
  6 | 2
Key: 2 | 3 means 23 minutes (commute times of 20 students)
```

Stemplots keep every value and show shape at the same time, but they become unwieldy for large data sets or values with many digits. For large data sets we use a histogram.

Histogram

Divide the number line into equal-width **bins** (classes), count the observations in each bin, and draw a bar for each bin whose height is the count (or the relative frequency).

Bars **touch** because the variable is continuous or the bins cover the whole number line without gaps. Compare this with a bar graph, whose bars are separated.

Agree on an endpoint rule. This course uses the convention that a value on a boundary goes into the bin to its right, so the bin 10 to 20 includes 10 but not 20.

A histogram shows shape and rough size but hides the individual values.

Delivery times for 40 meal orders (bin width 10 minutes).

The choice of bin width matters. Very wide bins hide the shape; very narrow bins make the graph jagged and hide the overall pattern. A rule of thumb is to try between 5 and 15 bins. In every histogram, the heights of the bars add up to the number of observations, and each bar's share of that total is its relative frequency.

When you describe any of these graphs, look at the **overall pattern** (shape, center, spread) and at **deviations** from it (outliers, gaps, clusters). We develop the language of shape, center, and spread in the next lesson; here we practice reading the graphs accurately.

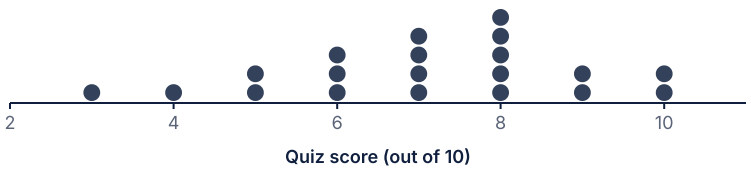
Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: READ A DOTPLOT

The dotplot shows quiz scores out of 10 for 20 students. What percent of the students scored at least 8?



SOLUTION

Count the dots at 8, 9, and 10: there are 5, 2, and 2.

The total is $5 + 2 + 2 = 9$ students.

$9/20 = 0.45$.

Conclusion: 45 percent of the 20 students scored at least 8 on the quiz.

EXAMPLE 2: BUILD A STEMPLOT

The commute times, in minutes, of 20 students are: 12, 15, 17, 18, 21, 22, 22, 24, 25, 27, 28, 31, 33, 34, 36, 38, 41, 45, 47, 62. Make a stemplot and describe the shape.

```
  1 | 2 5 7 8
  2 | 1 2 2 4 5 7 8
  3 | 1 3 4 6 8
  4 | 1 5 7
  5 |
  6 | 2
Key: 2 | 3 means 23 minutes
```

SOLUTION

Use tens digits as stems (1 to 6) and ones digits as leaves. List every stem from 1 to 6 so the gap at 5 shows.

Stem 1: 2 5 7 8. Stem 2: 1 2 2 4 5 7 8. Stem 3: 1 3 4 6 8. Stem 4: 1 5 7. Stem 5: no leaves. Stem 6: 2.

Conclusion: The commute times pile up between 12 and 47 minutes and trail off to the right, with one far-away value of 62 minutes. The shape is skewed to the right.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: READ A HISTOGRAM

Use the delivery-time histogram. What percent of the 40 orders took at least 20 minutes but less than 40 minutes?

SOLUTION

The bins 20 to 30 and 30 to 40 have heights 13 and 8.

$13 + 8 = 21$ orders.

$21/40 = 0.525$.

Conclusion: 52.5 percent of the orders took from 20 up to (but not including) 40 minutes. This is the largest share of any pair of adjacent bins, and the single most common bin is 20 to 30 minutes.

EXAMPLE 4: SPLIT STEMS

Points scored in a season by 19 players: 41, 43, 44, 46, 47, 48, 52, 53, 55, 56, 57, 58, 59, 61, 63, 64, 66, 67, 71. Draw a split-stem plot.

```
   4 | 1 3 4
   4 | 6 7 8
   5 | 2 3
   5 | 5 6 7 8 9
   6 | 1 3 4
   6 | 6 7
   7 | 1
   7 |
Key: 5 | 8 means 58 points (leaves 0-4 on the first row of a stem, 5-9 on the second)
```

SOLUTION

With stems 4, 5, 6, and 7 and so many leaves, one row per stem would hide the shape. Split each stem: the first row takes leaves 0 to 4 and the second takes leaves 5 to 9.

Row 4 (0-4): 1 3 4. Row 4 (5-9): 6 7 8. Row 5 (0-4): 2 3. Row 5 (5-9): 5 6 7 8 9. Row 6 (0-4): 1 3 4. Row 6 (5-9): 6 7. Row 7 (0-4): 1. Row 7 (5-9): none.

Count the leaves: $3 + 3 + 2 + 5 + 3 + 2 + 1 = 19$, which matches the number of players.

Conclusion: The split display shows a peak in the upper 50s and a long thin tail toward 71, which a one-row-per-stem plot would have blurred.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: BACK-TO-BACK STEMPLOT

Test scores for two classes of 15 students each: Class A: 58, 61, 64, 67, 69, 72, 73, 75, 77, 78, 81, 84, 86, 88, 92. Class B: 62, 66, 68, 71, 74, 75, 76, 79, 80, 82, 83, 85, 87, 90, 91. Build a back-to-back stemplot and compare the two classes.

Class A leaves	Stem	Class B leaves
8	5	
9 7 4 1	6	2 6 8
8 7 5 3 2	7	1 4 5 6 9
8 6 4 1	8	0 2 3 5 7
2	9	0 1

SOLUTION

Use the shared stems 5 to 9. Class A leaves go on the left in decreasing order outward; Class B leaves go on the right in increasing order.

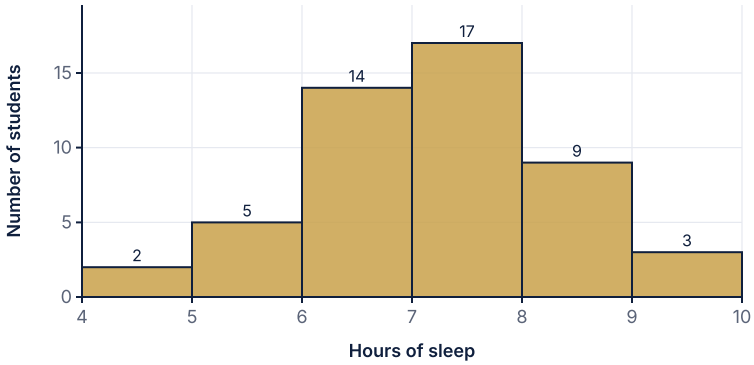
The medians are the 8th ordered value: 75 for Class A and 79 for Class B.

Class A ranges from 58 to 92 (range 34); Class B ranges from 62 to 91 (range 29).

Conclusion: Class B's typical score (median 79) is higher than Class A's (median 75), and Class B's scores are slightly less spread out. Neither class shows an outlier.

EXAMPLE 6: ESTIMATE FROM A HISTOGRAM

The histogram shows nightly sleep, in hours, for 50 students (bins of width 1 hour, from 4 to 10). (a) What percent slept fewer than 7 hours? (b) In which bin is the median? (c) Describe the shape.



SOLUTION

Bin counts are 2, 5, 14, 17, 9, 3, which add to 50.

(a) Fewer than 7 hours means the first three bins: $2 + 5 + 14 = 21$ students, and $21/50 = 0.42$, so 42 percent.

(b) The median of 50 values is halfway between the 25th and 26th. The cumulative counts are 2, 7, 21, 38, 47, 50. The 25th and 26th values are both in the bin whose cumulative count first exceeds 25, which is 7 to 8 hours.

(c) The distribution has one peak at 7 to 8 hours, a longer tail on the left (few students sleep only 4 to 5 hours), so it is slightly skewed to the left.

Conclusion: About 42 percent of students slept under 7 hours and the median sleep time is between 7 and 8 hours.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt. The stemplot shows the commute times of 20 students (stems 1 to 6, key 2 | 3 means 23 minutes). Describe the distribution of commute times. **Model.** The commute times of the 20 students are skewed to the right: most values are between 12 and 47 minutes and the values taper off toward longer times. The median commute is 27.5 minutes, the average of the 10th and 11th ordered values (27 and 28), so a typical student commutes about 27 or 28 minutes. The times range from 12 to 62 minutes. The value of 62 minutes is separated from the rest by a gap at stem 5 and is a possible unusually long commute. There is no gap or cluster elsewhere, and the largest stack of values is in the 20s.

What the grader looks for

(1) States the shape (skewed right) in context. (2) Gives a center with units, supported by a value read from the graph. (3) Describes spread, either with the range 12 to 62 minutes or with the interval where most values lie. (4) Notes the unusual value of 62 and the gap at stem 5. (5) Refers to commute times and students, not just to "the data".

Coach's Playbook

1. **Pick the graph for the job.** Small data set with repeated values: dotplot. Small to moderate data where you want exact values: stemplot. Large data set: histogram.
2. **Set up stems or bins before you plot.** Scan for the smallest and largest value, choose equal-width pieces, and write every stem or bin, including empty ones.
3. **Count before you finish.** The number of leaves, dots, or the sum of bar heights must equal the sample size.
4. **Label everything.** A stemplot needs a key, a histogram needs axis labels with units, and a dotplot needs a labeled number line.
5. **Read the graph in context.** Say "students" or "minutes", not "values". Percent answers need the total as the denominator.

Calculator tip: to make a histogram, enter the data in a list, choose a histogram plot, and set the window so that the bin width and starting point you want are used. Always check how the calculator treats a value that falls exactly on a boundary, and trace a bar to read its count.

Common traps

(1) Leaving out an empty stem, which hides a gap in the data. (2) Drawing histogram bars with gaps between them, or using unequal bin widths without adjusting heights. (3) Forgetting the key on a stemplot. (4) Reading a histogram bar as a single value; a bar from 20 to 30 shows 13 values somewhere in that interval, not the value 25.

Mistake Clinic

Spot the error

Student stemplot: stems 1, 2, 3, 4, 6 for the commute data (the student skipped stem 5). **Where it fails:** omitting the empty stem makes 62 appear to follow 47 closely and hides the gap that suggests an outlier. **Correct solution:** write stems 1 to 6 with stem 5 left empty, so the stems are evenly spaced and the gap is visible.

Spot the error

Student claim: "The tallest bar of the sleep histogram is the 7 to 8 hour bin with 17 students, so 17 students slept exactly 7.5 hours." **Where it fails:** a histogram bar summarizes all the values in a bin; it cannot tell you the exact value of any observation. **Correct solution:** 17 students slept at least 7 and fewer than 8 hours.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Student claim: "A bar graph of the delivery times would look the same as the histogram, but I will leave gaps between the bars so it looks neater." **Where it fails:** delivery time is quantitative and the bins cover the number line without gaps. Gaps would suggest separate categories or empty intervals. **Correct solution:** draw the bars touching; reserve gaps for bar graphs of categorical data.

Real-World Connection

City planners look at histograms of commute times to decide when to run extra trains, and health teams look at stemplots of small clinic samples to spot a patient whose reading is far from the others. In both cases the graph is a first look that tells the analyst what to check before any formula is used.

Challenge Corner

A histogram has bins of width 5 starting at 0 with counts 3, 7, 12, 8, 6, 4. A second histogram of the same data uses bins of width 10 starting at 0. (a) What are the new counts? (b) Why might the second histogram hide an important feature? **Solution.** (a) Combine pairs of bins: $3 + 7 = 10$, $12 + 8 = 20$, $6 + 4 = 10$. The counts are 10, 20, 10, still totaling 40. (b) The first histogram peaks at 10 to 15 with 12 values and shows that the 15 to 20 bin has fewer (8), a detail that the combined bin 10 to 20 (20 values) smooths away; with wider bins we cannot tell where in the interval the values lie.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE ### Graphs for Quantitative Data: Dotplots, Stemplots, and Histograms

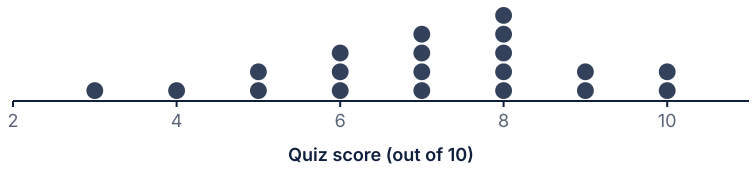
Show your work. A calculator is allowed. Justify conclusions in context.

1. Which graph displays every individual data value of a quantitative variable?

   (A) a histogram **(B)** a stemplot

   (C) a bar graph of frequencies **(D)** a pie chart

2. The dotplot shows quiz scores out of 10 for 20 students. What percent of the students scored at least 8?



   (A) 9 percent **(B)** 40 percent **(C)** 55 percent

   (D) 45 percent

3. In the delivery-time histogram (bins of width 10 from 0 to 60, counts 4, 10, 13, 8, 4, 1), what percent of the 40 orders took 40 minutes or more?

   (A) 12.5 percent **(B)** 5 percent **(C)** 10 percent

   (D) 25 percent

4. In the stemplot of 20 commute times (key 2 | 3 means 23 minutes), how many students had a commute of 30 to 39 minutes?

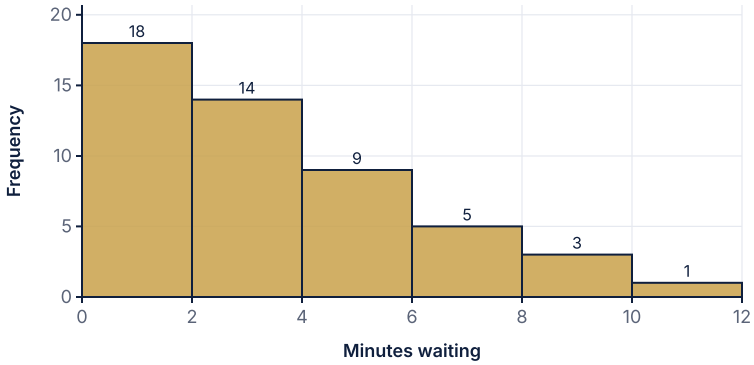
```
   1 | 2 5 7 8
   2 | 1 2 2 4 5 7 8
   3 | 1 3 4 6 8
   4 | 1 5 7
   5 |
   6 | 2
Key: 2 | 3 means 23 minutes
```

   (A) 4 **(B)** 7 **(C)** 5 **(D)** 3

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. Which describes the shape of this histogram?



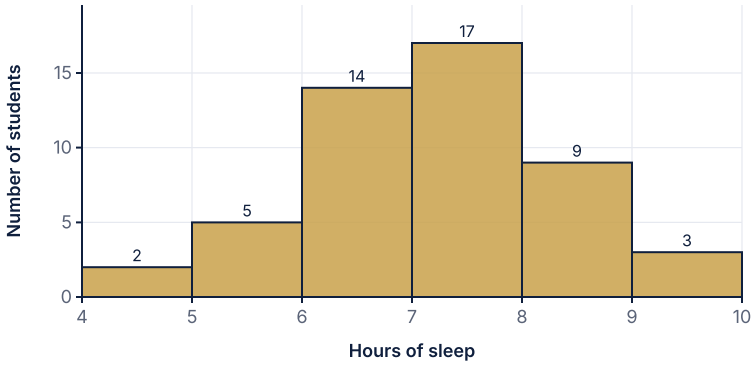
(A) skewed to the left **(B)** skewed to the right **(C)** symmetric
(D) uniform

6. What is the main difference between a histogram and a bar graph?
   (A) A histogram shows counts and a bar graph shows percents.
   (B) A histogram must have exactly five bars.
   (C) A bar graph can only show one category.
   (D) A histogram shows a quantitative variable in touching bins; a bar graph shows a categorical variable with gaps between bars.

7. A student changes the bin width of a histogram of 200 values from 10 to 1. The most likely result is
   (A) a smoother graph with the same level of detail
   (B) a graph with fewer bars that shows each value
   (C) a more jagged graph that shows more detail but may hide the overall shape
   (D) a graph whose total bar height is smaller

8. In a split-stem plot where the first row of each stem holds leaves 0 to 4 and the second row holds leaves 5 to 9, where does the value 47 appear?
   (A) on the second row of stem 4, as the leaf 7 **(B)** on the first row of stem 4, as the leaf 7
   (C) on the first row of stem 7, as the leaf 4 **(D)** on the second row of stem 5, as the leaf 7

9. The histogram shows sleep hours of 50 students (counts 2, 5, 14, 17, 9, 3 for bins 4 to 5, ..., 9 to 10). In which interval is the median?



(A) 6 to 7 hours **(B)** 7 to 8 hours **(C)** 8 to 9 hours
(D) 5 to 6 hours

10. Gym visits last month for 15 members: 8, 11, 12, 14, 15, 15, 17, 18, 19, 21, 22, 22, 26, 31, 35. (a) Make a stemplot with a key. (b) Find the median and describe the shape.

11. A histogram of 60 battery lifetimes has bins 0 to 5, 5 to 10, 10 to 15, 15 to 20, 20 to 25 hours, with counts 6, 18, 21, 11, 4. (a) Check the counts total 60. (b) Find the relative frequency of each bin. (c) What percent lasted at least 15 hours?

12. Use the back-to-back data: Class A scores 58, 61, 64, 67, 69, 72, 73, 75, 77, 78, 81, 84, 86, 88, 92 and Class B scores 62, 66, 68, 71, 74, 75, 76, 79, 80, 82, 83, 85, 87, 90, 91. (a) Find the median of each class. (b) Which class has the larger range, and what is it? (c) State one comparison supported by the stemplot.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. The values 20, 30, and 40 appear in a data set. A histogram uses bins 10 to 20, 20 to 30, 30 to 40, and 40 to 50 with the convention that a boundary value goes in the bin to its right. (a) In which bin does each of these three values go? (b) Why must a convention be stated?
14. A dotplot of the number of siblings for 25 students has dots: 0 siblings 4, 1 sibling 9, 2 siblings 7, 3 siblings 3, 4 siblings 1, 6 siblings 1. (a) What percent have at least 2 siblings? (b) Describe the shape and any unusual value.
15. A histogram has bins of width 4 starting at 0 with counts 2, 9, 14, 10, 4, 1. (a) How many observations are there? (b) Combine into bins of width 8; give the new counts. (c) Describe one feature of the original histogram that the new one hides.

1.3 Describing Shape, Center, and Spread

Describe a distribution with shape, center, and spread. Compare mean and median, understand resistance, and compute and interpret the range, IQR, variance, and standard deviation.

Why This Matters

When someone asks, "What is a typical value, and how much do values vary?", they want one number for the center and one number for the spread. Choosing the wrong number is the most common error in this topic: a mean that is dragged upward by a single huge value, or a standard deviation quoted for a distribution with two clusters.

This lesson sits in **Exam Unit 1: Exploring One-Variable Data**. Expect multiple-choice questions on the relationship between mean and median, resistance, and what a standard deviation measures, and free-response tasks that ask you to describe a distribution using **shape, center, and spread** in context.

Core Ideas

Start with shape. Look at the overall outline of a dotplot, stemplot, or histogram and name its features.

Describing Shape

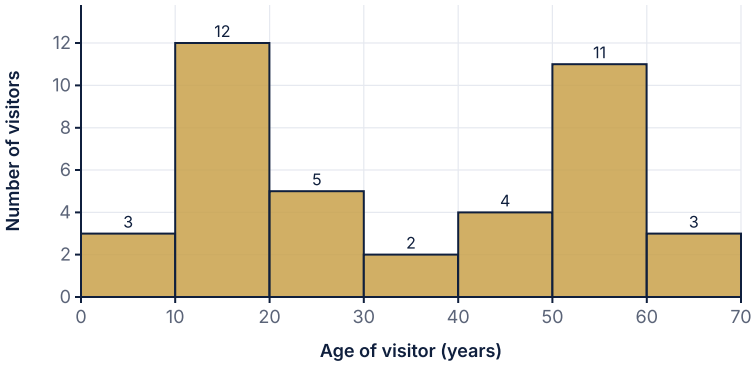
Symmetric: the left and right sides are roughly mirror images.

Skewed right: the long tail extends toward larger values. **Skewed left:** the long tail extends toward smaller values. Skew is named for the side of the tail, not the side of the peak.

Unimodal, bimodal, multimodal: one, two, or more distinct peaks (modes). Two peaks often signal two different groups mixed together.

Uniform: bars are roughly the same height, with no peak.

Also mention outliers (values far from the rest) and gaps.



A bimodal distribution: ages of 40 visitors to a science museum.

Next, center. There are three common measures.

Measures of Center

Mean: $\bar{x} = \dfrac{\sum x_i}{n}$, the balance point of the distribution.

Median: the middle value of the ordered data. For odd n it is the $\frac{n+1}{2}$th value; for even n it is the average of the two middle values.

Mode: the most frequent value (or the most common category). A data set can have no mode, one mode, or several.

The mean uses every value, so it is pulled toward a long tail. The median depends only on the middle of the ordered list.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A statistic is **resistant** if extreme values have little effect on it. The median is resistant; the mean is not. This leads to the key comparison rule below.

Mean versus Median and Shape

Symmetric: mean $\approx$ median.
Skewed right: mean $>$ median (the tail pulls the mean up).
Skewed left: mean $<$ median (the tail pulls the mean down).
For skewed distributions or distributions with outliers, the median is usually the better description of a typical value.

Now spread. The **range** is the maximum minus the minimum; it uses only two values and is not resistant. The **interquartile range** (IQR) is $Q_3 - Q_1$, the length of the interval holding the middle half of the data. It is resistant. The **standard deviation** measures the typical distance of the values from the mean.

Standard Deviation and Variance

Sample standard deviation:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

The **variance** is s_x^2.
s_x is in the same units as the data and is 0 only when all values are equal.
Neither the mean nor s_x is resistant. Use the mean and standard deviation together, and use the median and IQR together.
Interpretation sentence: "The [variable] values typically differ from the mean of [value] [units] by about [s_x] [units]."

The deviations $x_i - \bar{x}$ always add to zero, which is why we square them before averaging. The divisor $n - 1$ (rather than n) is used for a sample so that the sample variance does not underestimate the population variance. When a population standard deviation is known, it is written σ and the divisor is N; on the exam the calculator provides both.

Fifteen salaries from the worked examples: skewed right.

Worked Examples

EXAMPLE 1: MEAN, MEDIAN, AND RESISTANCE

Five students report the number of apps they opened this morning: 6, 9, 3, 12, 5. (a) Find the mean and median. (b) A sixth student reports 50. Recompute and compare.

SOLUTION

(a) Ordered: 3, 5, 6, 9, 12. The mean is $35/5 = 7$ and the median is the middle value, 6.
(b) With 50 added the sum is 85, so the mean is $85/6 \approx 14.17$. The ordered list 3, 5, 6, 9, 12, 50 has median $(6 + 9)/2 = 7.5$.
The mean more than doubled (from 7 to about 14.2); the median moved only from 6 to 7.5.
Conclusion: The median is resistant to the unusual value of 50; the mean is not.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: STANDARD DEVIATION BY HAND

Find the sample standard deviation of the values 3, 5, 6, 8, 13.

SOLUTION

The mean is $\bar{x} = 35/5 = 7$.

The deviations are $-4, -2, -1, 1, 6$ and they sum to 0. Their squares are $16, 4, 1, 1, 36$, which sum to 58.

Divide by $n - 1 = 4$: variance $s^2 = 58/4 = 14.5$.

$s = \sqrt{14.5} \approx 3.81$.

Conclusion: The values typically differ from the mean of 7 by about 3.8. The large value 13 contributes 36 of the 58, showing how much one value affects the standard deviation.

EXAMPLE 3: WHICH MEASURES SHOULD I REPORT?

Salaries (in thousands of dollars) of 15 employees at a small firm: 38, 40, 41, 42, 43, 44, 45, 46, 48, 50, 52, 55, 60, 85, 120. The mean is about 53.9, the median is 46, and the standard deviation is about 21.6. Describe the shape and recommend the best summary.

SOLUTION

Most salaries are between 38 and 60, with two much larger values (85 and 120) in a tail to the right, so the shape is skewed right.

The mean (53.9) is greater than the median (46), consistent with right skew.

Because of the skew and the high values, the median is the better description of a typical salary. The matching spread measure is the IQR: $Q_1 = 42$ and $Q_3 = 55$, so IQR $= 13$.

Conclusion: Report a median salary of 46 thousand dollars with IQR 13 thousand dollars. The mean and standard deviation are inflated by the two highest salaries.

EXAMPLE 4: SAME MEAN, DIFFERENT SPREAD

Data set A: 48, 49, 50, 51, 52. Data set B: 30, 40, 50, 60, 70. Compare the means and standard deviations.

SOLUTION

Both sets have mean 50 and median 50.

A: deviations $-2, -1, 0, 1, 2$, squares sum to 10, $s = \sqrt{10/4} \approx 1.58$.

B: deviations $-20, -10, 0, 10, 20$, squares sum to 1000, $s = \sqrt{1000/4} \approx 15.81$.

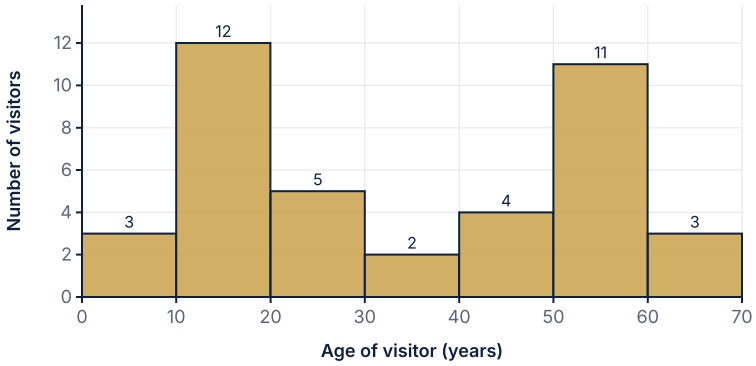
Conclusion: The two sets have the same center but B is about 10 times as spread out as A. The mean alone cannot tell them apart.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: BIMODAL SHAPE

The histogram shows the ages of 40 visitors to a science museum. Describe the distribution and explain why the mean would be a poor summary.



SOLUTION

There are two clear peaks: one at 10 to 20 years (12 visitors) and one at 50 to 60 years (11 visitors), with a dip at 30 to 40 years (only 2 visitors).

The distribution is **bimodal**, which suggests two groups: school-age visitors and adults.

The mean would land near the middle, around 30 to 40 years, where very few visitors actually are.

Conclusion: The ages are bimodal with modes near 15 and 55 years. A single center would misrepresent the data, so describe the two clusters separately.

EXAMPLE 6: EFFECT OF A CHANGE ON MEAN, MEDIAN, AND SD

Eight daily step counts (in thousands): 14, 18, 19, 21, 23, 24, 27, 31. The mean is 22.125, the median is 22, and $s \approx 5.36$. The largest value, 31, is replaced by 61 (a recording error). Predict and verify the effect on the mean, median, standard deviation, and range.

SOLUTION

The median uses only the middle two values, 21 and 23... no wait; recompute from ordered data: the middle values are 21 and 23 are not both fixed by the change.

The ordered list is 14, 18, 19, 21, 23, 24, 27, and now 61. The middle two values are 21 and 23, so the median is 22 and does not change.

The sum rises by 30, so the mean rises by $30/8 = 3.75$ to 25.875.

The standard deviation rises sharply to about 14.74 because one value is now far from the mean; the range rises from 17 to 47.

Conclusion: Median unchanged; mean up 3.75; standard deviation and range both much larger. Only the median is resistant here.

Write It Like a 5

Model Response

Prompt. A histogram of the 15 salaries above is shown, with mean 53.9 thousand dollars, median 46 thousand dollars, and standard deviation 21.6 thousand dollars. A manager says, "The typical salary is 53.9 thousand dollars." Evaluate this claim. **Model.** The distribution of salaries is skewed to the right: most of the 15 salaries are between 38 and 60 thousand dollars, but two employees earn 85 and 120 thousand dollars. Because the mean is not resistant, these two large salaries pull the mean (53.9) above the median (46). In this skewed distribution 11 of the 15 salaries are below the mean, so it is not a typical salary. The median of 46 thousand dollars is a better description of a typical employee, since half of the employees earn less than 46 thousand dollars and half earn more. The manager's claim is misleading.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

(1) Identifies the right skew and cites evidence (two large values; mean greater than median). (2) States that the mean is not resistant to the high values. (3) Recommends the median as the better center and gives its value. (4) Refers to salaries and employees in context. (5) Reaches a clear verdict on the manager's claim.

Coach's Playbook

1. **Describe in the order SOCS.** Shape, Outliers, Center, Spread. Always add context words and units.
2. **Compare mean and median first.** A large gap between them is quick evidence of skew or an outlier, in the direction of the larger value.
3. **Pair your measures.** Mean with standard deviation for roughly symmetric data without outliers; median with IQR otherwise.
4. **Interpret the standard deviation.** Use the phrase "typically differ from the mean by about" and include units.
5. **Predict before you compute.** If one value moves a long way, ask which statistics use that value: the mean, standard deviation, and range do; the median and IQR barely do.

Calculator tip: enter the data in a list and run the one-variable statistics command. Read $\bar{x}$ and the sample standard deviation s_x (not the population version σ_x) unless the data are an entire population. Scroll down for the median and the quartiles.

Common traps

(1) Describing a skewed-left distribution as skewed right because the peak is on the right; skew follows the tail. (2) Reporting the mean and standard deviation for a strongly skewed distribution. (3) Saying the standard deviation is "the average of the data" instead of the typical distance from the mean. (4) Forgetting that the mean and standard deviation of a bimodal distribution do not describe the two clusters.

Mistake Clinic

Spot the error

Student claim: "The histogram has its tallest bars on the right and the tail on the left, so it is skewed right." **Where it fails:** skewness is named by the direction of the long tail, not the location of the peak. **Correct solution:** the long tail is on the left, so the distribution is skewed left, and the mean is less than the median.

Spot the error

Student solution: "For the data 3, 5, 6, 8, 13 the standard deviation is $\sqrt{58/5} \approx 3.41$." **Where it fails:** the student divided by $n = 5$; for a sample the divisor is $n - 1 = 4$. **Correct solution:** $s = \sqrt{58/4} = \sqrt{14.5} \approx 3.81$.

Spot the error

Student claim: "After adding an extreme value, the median and the mean both increased by a lot, so neither is useful." **Where it fails:** the median is resistant and moves very little; only the mean is strongly affected. **Correct solution:** in Example 1, the median moved from 6 to 7.5 while the mean jumped from 7 to about 14.2; report the median when outliers are present.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Real-estate reports use the median home price, not the mean, because a few very expensive houses would pull the mean up in a right-skewed market. Manufacturers care about the standard deviation of a part's width, since a small spread means parts fit together reliably.

Challenge Corner

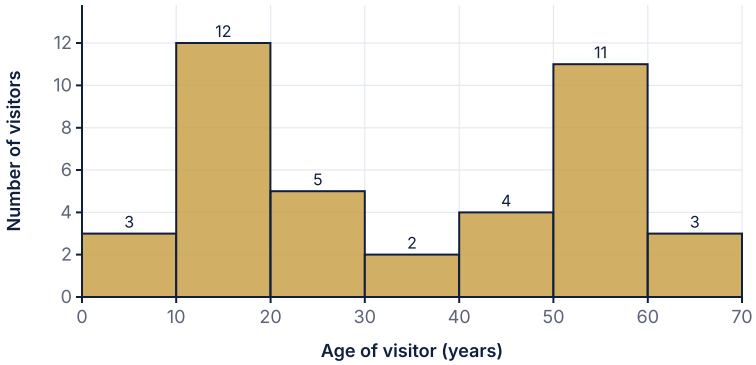
A data set of 8 values has mean 40 and standard deviation 6. One value, 52, is removed, leaving 7 values. What are the new mean and, using the fact that the sum of squared deviations from the old mean is 252, the new standard deviation? **Solution.** The original sum is $8(40) = 320$; the new sum is 268, so the new mean is $268/7 \approx 38.29$. The sum of squared deviations satisfies $\sum(x-\bar{x})^2 = \sum x^2 - n\bar{x}^2$. Original: $\sum x^2 = 252 + 8(1600) = 13052$. New: $\sum x^2 = 13052 - 2704 = 10348$, and the new sum of squared deviations is $10348 - 7(38.2857)^2 \approx 10348 - 10260.6 = 87.4$. The new variance is $87.4/6 \approx 14.6$ and $s \approx 3.82$. Removing the extreme value 52 cut the standard deviation from 6 to about 3.8.

Practice

PRACTICE **Describing Shape, Center, and Spread**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A distribution is skewed to the right. Which statement about the mean and median is most likely true?
   (A) The mean is greater than the median. **(B)** The mean is less than the median.
   (C) The mean equals the median. **(D)** The mean must equal the mode.
2. Which pair of statistics is resistant to outliers?
   (A) the mean and the standard deviation **(B)** the median and the interquartile range
   (C) the mean and the range **(D)** the median and the range
3. The values 4, 7, 9, 12, 18 have mean
   (A) 9 **(B)** 11 **(C)** 10 **(D)** 12
4. A set of times has mean 30 minutes and standard deviation 4 minutes. Which interpretation is best?
   (A) The times range from 26 to 34 minutes.
   (B) Four minutes is the median time.
   (C) The average difference between the largest and smallest times is 4 minutes.
   (D) The times typically differ from 30 minutes by about 4 minutes.
5. The histogram shows ages of visitors. Which description is best?



   (A) skewed right with one mode **(B)** bimodal, with peaks near 15 and 55 years
   (C) symmetric and unimodal **(D)** uniform
6. In a data set of 9 values, the largest value is increased from 30 to 90. Which statistic must stay the same?
   (A) the mean **(B)** the standard deviation **(C)** the median
   (D) the range

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. Set P is 10, 20, 30 and set Q is 18, 20, 22. Which statement is correct?
   (A) Both sets have the same standard deviation because the means are equal.
   (B) Q has a larger standard deviation because its values are closer together.
   (C) Set P has mean 20 and set Q has mean 22.
   (D) Both sets have mean 20, and P has a larger standard deviation.
8. Why is the IQR often preferred to the range for describing the spread of a skewed distribution?
   (A) The IQR is resistant, while the range depends only on the two most extreme values.
   (B) The IQR is always larger than the range.
   (C) The IQR is the average of the deviations.
   (D) The IQR uses every value in the data set.
9. What is the sample standard deviation of 2, 4, 6, 8?
   (A) about 2.24 **(B)** about 6.67 **(C)** about 2.58
   (D) about 5
10. Eight daily step counts (thousands): 14, 18, 19, 21, 23, 24, 27, 31. (a) Find the mean and median. (b) Find the sample standard deviation. (c) Which value would you quote as typical, and why?
11. The values 12, 15, 15, 18, 20, 22, 24, 95 are the minutes that 8 students spent on homework. (a) Find the mean and median. (b) Explain which is more appropriate and why. (c) What is the effect on the mean if the value 95 is removed?
12. A class reports a mean quiz score of 72 and a median of 81. (a) What does this suggest about the shape of the distribution? (b) Which measure of center is more representative of a typical student? (c) Describe a plausible real-world reason for this pattern.
13. The distribution of arrival times at a bus stop has two peaks, one at about 7:30 a.m. and one at about 3:30 p.m. (a) What is the shape? (b) Why would the mean arrival time describe the data poorly? (c) What should be reported instead?
14. Two machines fill bags of rice. Machine X fills have mean 500 grams and standard deviation 2 grams. Machine Y fills have mean 500 grams and standard deviation 9 grams. (a) Compare the distributions of fills. (b) Which machine is better for a customer who needs consistent bag weights? (c) Interpret the standard deviation of Machine Y in context.
15. A survey of 26 families recorded the number of pets: 0 pets for 6 families, 1 pet for 9, 2 pets for 7, 3 pets for 3, and 4 pets for 1. (a) Find the mode and the median. (b) Find the mean. (c) Is the distribution symmetric? Explain.

1.4 The Five-Number Summary, Boxplots, and Outliers

Compute quartiles and the five-number summary, draw boxplots and modified boxplots, and use the 1.5 IQR rule to decide whether a value is an outlier.

Why This Matters

The mean and standard deviation tell one story. The five-number summary tells another that is better when data are skewed or contain unusual values: it uses the median and quartiles, which resist outliers. A boxplot is a picture of those five numbers, and it is the standard tool for comparing several groups side by side.

This is part of **Exam Unit 1: Exploring One-Variable Data**. Multiple-choice items ask you to read quartiles, the IQR, or outliers from a boxplot and to apply the 1.5 IQR rule. In free-response questions you are often asked to justify whether a value is an outlier with a calculation, and a correct justification always shows the fences.

Core Ideas

Quartiles and the Five-Number Summary

Order the data. The **median** (M) splits it into a lower half and an upper half. If n is odd, leave the median out of both halves.

First quartile Q_1**:** the median of the lower half. **Third quartile** Q_3**:** the median of the upper half.

The **five-number summary** is: minimum, Q_1, median, Q_3, maximum.

About 25 percent of the values lie in each of the four sections these numbers define. Technology may use a slightly different quartile rule for small data sets; the median-of-halves method is used here and gives the same answers as the exam's rules for the examples in this course.

The **interquartile range** is $\text{IQR} = Q_3 - Q_1$. It measures the spread of the middle half of the data and, unlike the range, it is resistant to extreme values.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

The 1.5 IQR Rule for Outliers

Lower fence $= Q_1 - 1.5\,(\text{IQR})$. Upper fence $= Q_3 + 1.5\,(\text{IQR})$.

A value is flagged as an **outlier** if it is below the lower fence or above the upper fence. A value exactly on a fence is not flagged.

Flagging a value does not mean it is wrong; investigate the context before removing it.

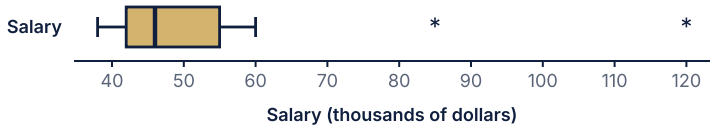
Drawing a Boxplot

Draw a box from Q_1 to Q_3 with a line at the median.

Draw whiskers from the box to the smallest and largest values that are **not** outliers.

A **modified boxplot** plots each outlier as a separate mark (an asterisk or dot) beyond the whiskers.

Use a labeled number line with units. Boxplots for comparison should share the same scale.



Modified boxplot of the 15 salaries.

A boxplot shows center (the median line), spread (the length of the box and the whiskers), skew (a median off center in the box, or one whisker longer than the other), and outliers. It does **not** show modes, gaps, or the number of observations, so two very different data sets can have the same boxplot.

Worked Examples

EXAMPLE 1: FIVE-NUMBER SUMMARY (ODD N)

Find the five-number summary of 3, 5, 7, 8, 10, 12, 13, 15, 18, 19, 25 and check for outliers.

SOLUTION

There are $n = 11$ ordered values. The median is the 6th: 12.

Remove the median from both halves. Lower half: 3, 5, 7, 8, 10, so $Q_1 = 7$. Upper half: 13, 15, 18, 19, 25, so $Q_3 = 18$.

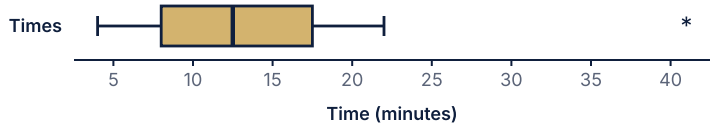
The five-number summary is 3, 7, 12, 18, 25. $\text{IQR} = 18 - 7 = 11$.

Fences: $7 - 1.5(11) = -9.5$ and $18 + 1.5(11) = 34.5$. All values lie between them.

Conclusion: There are no outliers, and the whiskers extend to 3 and 25.

EXAMPLE 2: OUTLIER AND MODIFIED BOXPLOT (EVEN N)

Twelve students timed a puzzle, in minutes: 4, 6, 7, 9, 10, 12, 13, 15, 16, 19, 22, 41. Find the five-number summary, test for outliers, and draw the modified boxplot.



SOLUTION

With $n = 12$ the median is the average of the 6th and 7th values: $(12 + 13)/2 = 12.5$.

Lower half: 4, 6, 7, 9, 10, 12, so $Q_1 = (7 + 9)/2 = 8$. Upper half: 13, 15, 16, 19, 22, 41, so $Q_3 = (16 + 19)/2 = 17.5$.

$\text{IQR} = 17.5 - 8 = 9.5$ and $1.5(9.5) = 14.25$. Lower fence $= 8 - 14.25 = -6.25$; upper fence $= 17.5 + 14.25 = 31.75$.

The value 41 exceeds 31.75, so it is an outlier. The largest value that is not an outlier is 22, where the upper whisker ends.

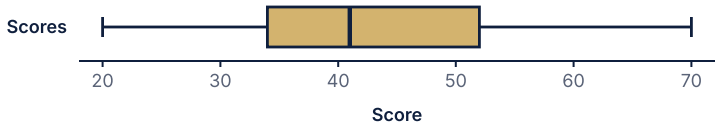
Conclusion: The time of 41 minutes is an outlier; the box runs from 8 to 17.5 with the median at 12.5.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: READ A BOXPLOT

A boxplot of scores for 80 students shows minimum 20, $Q_1 = 34$, median 41, $Q_3 = 52$, maximum 70. Find the IQR and range, and say roughly how many students scored between 34 and 52.



SOLUTION

$\text{IQR} = 52 - 34 = 18$ and range $= 70 - 20 = 50$.

The middle half of the data lies between Q_1 and Q_3, so about $0.50(80) = 40$ students scored between 34 and 52.

The median is 7 above Q_1 but 11 below Q_3, and the upper whisker (18 long) is longer than the lower (14 long).

Conclusion: About 40 students scored between 34 and 52. The distribution is mildly skewed right.

EXAMPLE 4: THE 1.5 IQR RULE FROM A SUMMARY

Song lengths in seconds for a playlist have $Q_1 = 62$, $Q_3 = 78$, minimum 35, and maximum 100. Are the minimum and maximum outliers?

SOLUTION

$\text{IQR} = 78 - 62 = 16$ and $1.5(16) = 24$.

Lower fence $= 62 - 24 = 38$; upper fence $= 78 + 24 = 102$.

The minimum 35 is below 38, so it is an outlier. The maximum 100 is below 102, so it is not.

Conclusion: Only the shortest song, 35 seconds, is flagged as an outlier. We must show the fences to justify this.

EXAMPLE 5: SAME BOXPLOT, DIFFERENT DATA

Set X: 10, 20, 22, 24, 30, 36, 38, 40, 50. Set Y: 10, 12, 30, 30, 30, 30, 38, 40, 50. Compare their boxplots.

SOLUTION

Set X: median 30, $Q_1 = (20 + 22)/2 = 21$, $Q_3 = (38 + 40)/2 = 39$.

Set Y: median 30, $Q_1 = (12 + 30)/2 = 21$, $Q_3 = (38 + 40)/2 = 39$.

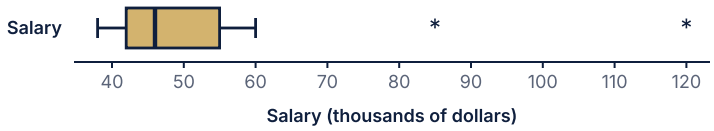
Both sets have the five-number summary 10, 21, 30, 39, 50, so the boxplots are identical.

The data differ: Y has four values equal to 30 (a spike), while X is spread evenly.

Conclusion: A boxplot cannot show clusters, gaps, or modes. Use a histogram or dotplot when those matter.

EXAMPLE 6: MODIFIED BOXPLOT OF THE SALARIES

Use the 15 salaries (thousands of dollars): 38, 40, 41, 42, 43, 44, 45, 46, 48, 50, 52, 55, 60, 85, 120. Find the outliers and describe the distribution.



Modified boxplot of the 15 salaries.

SOLUTION

The median is the 8th value, 46. Lower half (7 values): 38 to 45, so $Q_1 = 42$. Upper half: 48, 50, 52, 55, 60, 85, 120, so $Q_3 = 55$.

$\text{IQR} = 13$ and $1.5(13) = 19.5$. Fences: $42 - 19.5 = 22.5$ and $55 + 19.5 = 74.5$.

The values 85 and 120 are above 74.5, so both are outliers. The upper whisker ends at 60.

Conclusion: Salaries are skewed right with two high outliers (85 and 120 thousand dollars). The middle half of salaries lies between 42 and 55 thousand dollars, and the median is 46.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt. Times (in minutes) for 40 students to finish a puzzle have this five-number summary: minimum 9, $Q_1 = 13$, median 16, $Q_3 = 22$, maximum 41. The next largest value is 31. Determine whether there are any outliers and describe the distribution. **Model.** The IQR is $22 - 13 = 9$ minutes, so $1.5(\text{IQR}) = 13.5$. The lower fence is $13 - 13.5 = -0.5$ and the upper fence is $22 + 13.5 = 35.5$. The maximum of 41 minutes is greater than 35.5, so it is an outlier. No low value can be an outlier because the minimum of 9 is above -0.5. The next largest value, 31, is below 35.5, so the upper whisker ends at 31. The distribution of times is skewed to the right: the median (16) is closer to Q_1 (3 minutes away) than to Q_3 (6 minutes away), and the upper whisker is much longer. A typical student finished in about 16 minutes, and the middle 50 percent of the students took between 13 and 22 minutes.

What the grader looks for

(1) Computes the IQR correctly. (2) Shows both fences using the 1.5 IQR rule with numbers. (3) Compares each extreme with the fences and concludes that 41 minutes is an outlier. (4) Describes skew with evidence from the quartiles or whiskers. (5) Reports center and spread in context with units.

Coach's Playbook

1. **Order the data first.** Always sort before finding the median or quartiles; a mistake here ruins everything after it.
2. **Split at the median.** For odd n, leave the median out of both halves. For even n, split the list down the middle.
3. **Compute the fences before you decide.** Write $Q_1 - 1.5\,\text{IQR}$ and $Q_3 + 1.5\,\text{IQR}$, then compare each extreme with them.
4. **Whiskers stop at the last non-outlier.** The whisker ends at the most extreme value inside the fences, not at the fence itself.
5. **Interpret in context.** Say what a quartile means: "about 25 percent of students took less than 13 minutes".

Calculator tip: the one-variable statistics screen lists Q_1 and Q_3; check that they match your median-of-halves answer for small data sets. A box-and-whisker plot option that shows outliers separately is the modified boxplot.

Common traps

(1) Including the median in both halves when n is odd. (2) Extending the whisker to an outlier or to the fence instead of the last non-outlier value. (3) Saying a boxplot shows how many values are in each section by area; each section holds about a quarter of the values regardless of length. (4) Flagging an outlier by eye instead of using the 1.5 IQR rule.

Mistake Clinic

Spot the error

Student solution: "Data: 4, 6, 7, 9, 10, 12, 13, 15, 16, 19, 22, 41. The upper fence is 17.5 + 1.5 = 19, so 22 and 41 are outliers." **Where it fails:** the student added 1.5 instead of 1.5 times the IQR. **Correct solution:** $\text{IQR} = 9.5$, upper fence $= 17.5 + 14.25 = 31.75$; only 41 is an outlier.

Spot the error

Student claim: "The box is longer on the right, so more than half of the data lie in the right half of the box." **Where it fails:** the median splits the data in half, and each of the two sections of the box holds about 25 percent. A longer section means those values are more spread out, not more numerous. **Correct solution:** each quarter of the data lies in each section; the longer section indicates skew toward that side.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Student claim: "The two boxplots are identical, so the two classes have the same distribution of scores." **Where it fails:** identical five-number summaries do not imply identical shapes (Example 5). **Correct solution:** state that the quartiles, median, and extremes match, and use a histogram to compare shapes.

Real-World Connection

Quality-control teams chart the weight of items with boxplots and use a fence rule to flag unusual units for inspection. Sports analysts compare boxplots of points per game across teams for the same reason: the median and quartiles are not thrown off by one unusually high or low game.

Challenge Corner

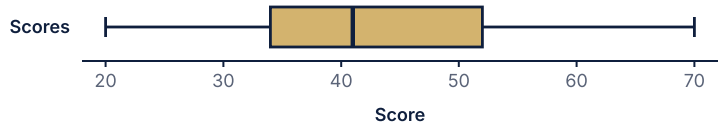
A data set has $Q_1 = 40$ and $Q_3 = 60$, a maximum of 97, and the next largest value is 74. (a) Find the interval of values that would not be flagged as outliers. (b) A student says the upper whisker ends at 90 because that is the upper fence. Explain the error and say where the whisker ends. **Solution.** (a) $\text{IQR} = 20$ and $1.5(20) = 30$, so the fences are $40 - 30 = 10$ and $60 + 30 = 90$. Values from 10 to 90 inclusive are not flagged. (b) The whisker extends to the largest data value inside the fences, not to the fence itself. The maximum 97 is above 90, so it is an outlier plotted separately, and the whisker ends at the next largest value, 74.

Practice

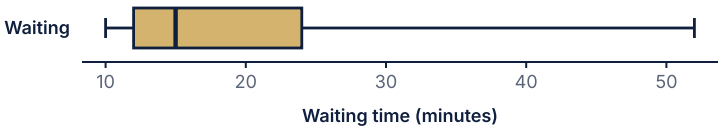
PRACTICE **The Five-Number Summary, Boxplots, and Outliers**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which list is the five-number summary?
   (A) mean, standard deviation, minimum, maximum, range
   (B) minimum, mean, median, mode, maximum
   (C) minimum, first quartile, median, third quartile, maximum
   (D) first quartile, median, third quartile, IQR, range
2. A distribution has $Q_1 = 24$ and $Q_3 = 39$. What is the IQR?
   (A) 15 **(B)** 63 **(C)** 31.5 **(D)** 7.5
3. For a data set, $Q_1 = 30$ and $Q_3 = 46$. Which value would be flagged as an outlier by the 1.5 IQR rule?
   (A) 58 **(B)** 66 **(C)** 70 **(D)** 72
4. The boxplot shows scores of 80 students. About how many scored between 34 and 52?



   (A) 20 **(B)** 40 **(C)** 80 **(D)** 60
5. Which statement about this boxplot is supported?



   (A) The distribution is skewed left.
   (B) The distribution is symmetric.
   (C) The distribution is skewed right.
   (D) The distribution is bimodal.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. The largest value in a data set of 25 values is changed from 50 to 500. Which statistic is guaranteed not to change?
   (A) the interquartile range **(B)** the range **(C)** the mean
   (D) the standard deviation
7. What is Q_1 for the data 2, 4, 6, 8, 10, 12, 14, 16?
   (A) 4 **(B)** 6 **(C)** 3 **(D)** 5
8. In a modified boxplot, the upper whisker extends to
   (A) the largest value that is not an outlier
   (B) the maximum value, whether or not it is an outlier
   (C) the upper fence
   (D) the third quartile plus the IQR
9. Two data sets have the same five-number summary. Which statement must be true?
   (A) Their histograms are identical. **(B)** Their boxplots are identical.
   (C) Their means are equal. **(D)** Their standard deviations are equal.
10. Fourteen runners' times for a 1 km run in minutes: 22, 25, 27, 28, 30, 31, 33, 35, 36, 38, 40, 43, 46, 71. (a) Find the five-number summary. (b) Test for outliers with the 1.5 IQR rule. (c) Describe the distribution.
11. The five-number summary of 60 exam scores is 45, 58, 64, 71, 99. (a) Find the IQR. (b) Determine whether the minimum or maximum is an outlier. (c) Interpret $Q_1 = 58$ in context.
12. Set X: 10, 20, 22, 24, 30, 36, 38, 40, 50 and Set Y: 10, 12, 30, 30, 30, 30, 38, 40, 50. (a) Show that their five-number summaries are equal. (b) Name one feature of the data a boxplot cannot show that distinguishes them.
13. Use the 14 running times above. The outlier 71 is a recording error and is removed. (a) Find the new median and IQR. (b) Compare the new and old IQR and range. (c) Which measure of spread changed most, and why?
14. Twenty students rated a movie from 1 to 5 stars: three 1s, five 2s, six 3s, four 4s, and two 5s. (a) Find the five-number summary. (b) Find the IQR. (c) Are there outliers?
15. A summary of daily rainfall (mm) on 30 rainy days: minimum 1, $Q_1 = 3$, median 5, $Q_3 = 9$, maximum 40, and the largest value below 40 is 19. (a) Is 40 an outlier? Show work. (b) Is the value 19 an outlier? (c) Describe the skew.

1.5 Comparing Distributions in Context

Compare two or more distributions using shape, outliers, center, and spread, with side-by-side graphs and statements that name both groups, use comparing words, and stay in context.

Why This Matters

Most real questions in statistics are comparisons: Which brand lasts longer? Did the new teaching method help? Is one city's weather more variable than another's? A comparison is only convincing when it uses numbers from both groups, addresses center, spread, and shape, and says what the numbers mean in context.

Comparison tasks appear in **Exam Unit 1** both as multiple-choice questions (which statement about the two graphs is supported?) and as free-response questions where most points are lost for forgetting to compare or for describing each group separately without connecting words such as "higher than" or "more variable than".

Core Ideas

Comparing Two or More Distributions

Address **shape**, **outliers**, **center**, and **spread** (SOCS) for every comparison.

Use **comparison words** in the same sentence as both numbers: "higher than", "less spread out than", "more skewed than".

Match your measures: use the median and IQR (or the five-number summary) for skewed distributions or distributions with outliers; use the mean and standard deviation for roughly symmetric ones.

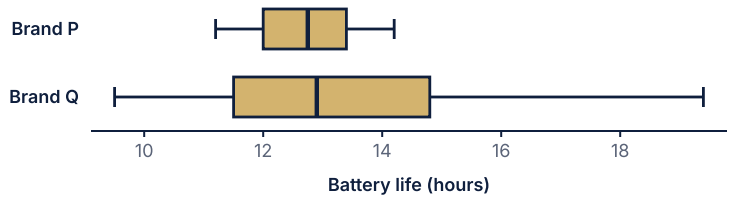
Put the context in every sentence: name the groups, the variable, and the units.

Compare **percents**, not counts, when the groups have different sizes.

The graphs that make a comparison easiest are those drawn on the **same scale**: side-by-side (parallel) boxplots, back-to-back stemplots, and stacked dotplots or histograms with identical axes. When two histograms are drawn separately, check that the horizontal scales and bin widths match before you compare.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

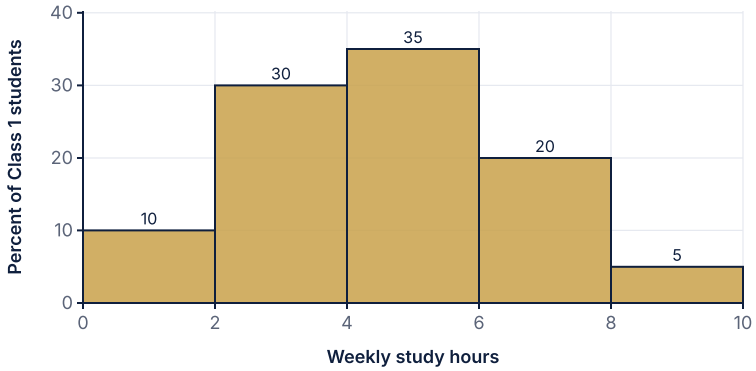


Battery life of 10 batteries from each brand.

A Complete Comparison Template

Center: "The median [variable] for [group A] ([value] [units]) is [higher/lower] than for [group B] ([value] [units])."
Spread: "[Group A] is [more/less] variable: its IQR is [value] compared with [value] for [group B]."
Shape and outliers: "Both are roughly symmetric, but [group A] has an outlier at [value]."
Never write only "A has a median of 12.9 and B has a median of 12.75" without a comparing word and a conclusion.

Use overlap with care. If the boxes of two boxplots overlap, the groups are not clearly different in their middles; if one group's third quartile is below another group's first quartile, three quarters of one group is below three quarters of the other. A difference in two samples is only a description of those samples. Whether it holds for a larger population is a question for inference, in later units.

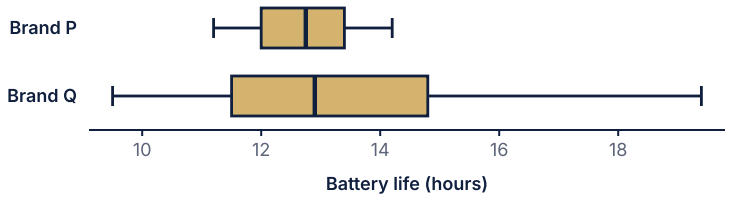


Class 1 as percents (40 students): compare with Class 2 (100 students) using percents, not counts.

Worked Examples

EXAMPLE 1: COMPARE TWO SMALL DATA SETS

Battery life in hours of 10 batteries of each of two brands. Brand P: 11.2, 11.8, 12.0, 12.4, 12.6, 12.9, 13.1, 13.4, 13.8, 14.2. Brand Q: 9.5, 10.8, 11.5, 12.2, 12.8, 13.0, 13.9, 14.8, 16.1, 19.4. Compare the distributions.



Battery life of 10 batteries from each brand.

SOLUTION

Brand P: mean 12.74, median 12.75, $s \approx 0.93$, IQR 1.4, range 3.0.

Brand Q: mean 13.4, median 12.9, $s \approx 2.85$, IQR 3.3, range 9.9.

Center: The typical battery life is similar for the two brands (medians 12.75 and 12.9 hours).

Spread: Brand Q batteries are much more variable than Brand P (IQR 3.3 hours versus 1.4 hours).

Shape and outliers: Brand P is roughly symmetric. Brand Q is skewed right: its high value, 19.4 hours, is far from the rest but is not an outlier by the 1.5 IQR rule (the upper fence is 19.75).

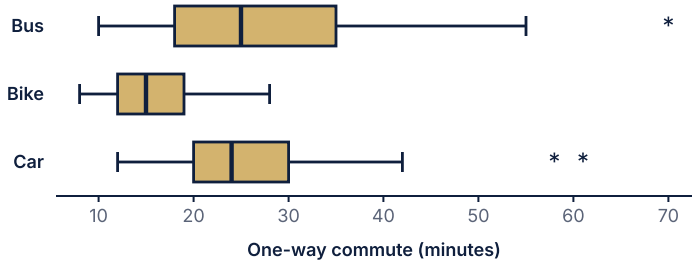
Conclusion: Both brands last about 12.8 hours typically, but Brand P is much more consistent.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: PARALLEL BOXPLOTS

The boxplots compare one-way commute times in minutes for students who ride the bus, bike, or drive. Compare the three distributions.



Commute times by mode of travel.

SOLUTION

Medians: bus 25, bike 15, car 24 minutes. Bike commutes are typically about 10 minutes shorter than bus or car commutes.

IQRs: bus $35 - 18 = 17$, bike $19 - 12 = 7$, car $30 - 20 = 10$. Bus commutes are the most variable and bike commutes the least.

Outliers: bus has one at 70 minutes; car has two, at 58 and 61 minutes. Bike has none.

Shape: all three distributions are skewed right (upper whiskers longer than lower whiskers).

Conclusion: Biking is typically fastest and most predictable. Bus and car commutes have similar medians, but the bus times vary more and a few students in each group have very long commutes.

EXAMPLE 3: DIFFERENT GROUP SIZES

Weekly study hours: Class 1 (40 students) has counts 4, 12, 14, 8, 2 in the bins 0-2, 2-4, 4-6, 6-8, 8-10. Class 2 (100 students) has counts 20, 35, 25, 15, 5. Compare the shares studying under 4 hours.

Hours	Class 1 (n=40)	Class 2 (n=100)
0-2	4	20
2-4	12	35
4-6	14	25
6-8	8	15
8-10	2	5

SOLUTION

Class 1: $(4 + 12)/40 = 0.40$, or 40 percent. Class 2: $(20 + 35)/100 = 0.55$, or 55 percent.

Class 2 has the larger count (55 versus 16), but that is only because it has more students.

Conclusion: A larger share of Class 2 students study fewer than 4 hours a week (55 percent versus 40 percent). Class 1 has a higher share in the middle bin 4 to 6 hours (35 percent versus 25 percent).

EXAMPLE 4: CHOOSING MEASURES FOR A COMPARISON

Monthly rainfall (mm) summary: City A has mean 62, median 60, standard deviation 20, IQR 28. City B has mean 58, median 41, standard deviation 41, IQR 22. Compare the typical month and the variability.

SOLUTION

City B's mean (58) is well above its median (41), which suggests a right-skewed distribution with a few very wet months. City A's mean and median are close, so it is roughly symmetric.

For City B, use the median and IQR to describe the typical month; its large standard deviation is inflated by the wet months.

Typical month: City A is wetter (median 60 mm versus 41 mm).

Variability in a typical month: City A's middle half spans 28 mm and City B's spans 22 mm, so the middle half of City B is slightly less spread out, even though City B has a much larger standard deviation because of its extreme months.

Conclusion: A typical month is wetter in City A, while City B is mostly drier with occasional very wet months.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: STATEMENTS SUPPORTED BY BOXPLOTS

The boxplots show test scores for Class A and Class B. Is each statement supported? (i) At least three quarters of Class A scored at or below Class B's median. (ii) Every student in Class B scored higher than every student in Class A.

SOLUTION

The third quartile of Class A is 70 and the median of Class B is 70.

(i) Since 75 percent of Class A scored at or below $Q_3 = 70$, at least 75 percent scored at or below Class B's median. Supported.

(ii) Class A's maximum is 88 and Class B's minimum is 45, so the score ranges overlap heavily. Not supported.

Conclusion: Class B tends to score higher, but the groups overlap a lot.

EXAMPLE 6: SAME CENTER, DIFFERENT CONCLUSION

Two bakeries record the weight in grams of loaves. Bakery X: mean 500, $s = 3$, no outliers, symmetric. Bakery Y: mean 500, $s = 12$, symmetric. A buyer needs loaves between 490 and 510 grams. Which bakery is more reliable, and about how often does each meet the target?

SOLUTION

Both bakeries have the same center, 500 grams, so center alone cannot separate them.

The target is 500 ± 10 grams. For X, 10 grams is about 3.3 standard deviations, so nearly every loaf meets the target. For Y, 10 grams is only 0.83 standard deviations, so a substantial share of loaves fall outside the range.

Conclusion: Bakery X is more reliable. The two distributions have the same center but Bakery Y varies about four times as much, so the spread decides the question.

Write It Like a 5

Model Response

Prompt. The battery lives (in hours) of 10 batteries of each of two brands are summarized by Brand P: minimum 11.2, $Q_1 = 12.0$, median 12.75, $Q_3 = 13.4$, maximum 14.2; Brand Q: minimum 9.5, $Q_1 = 11.5$, median 12.9, $Q_3 = 14.8$, maximum 19.4. Compare the distributions of battery life. **Model.** The typical battery life is nearly the same for the two brands: the median is 12.9 hours for Brand Q and 12.75 hours for Brand P. The brands differ greatly in variability. The IQR is 3.3 hours for Brand Q compared with 1.4 hours for Brand P, and the range is 9.9 hours compared with 3.0 hours. Brand P batteries are therefore much more consistent. Brand P is roughly symmetric, since its median is near the middle of the box and the whiskers are about equal. Brand Q is skewed to the right, with a long upper whisker reaching 19.4 hours. Neither brand has outliers by the 1.5 IQR rule: Brand Q's upper fence is $14.8 + 1.5(3.3) = 19.75$ hours, which is above its maximum of 19.4 hours.

What the grader looks for

(1) Compares centers using both values and a comparison word. (2) Compares spread using the IQR or range from both groups. (3) Describes shape for each group (symmetric versus skewed right). (4) Addresses outliers with the fence calculation. (5) Uses context (battery life, hours, brands) throughout and does not just list statistics.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. **Put both numbers in one sentence.** A comparison needs a connecting word between the two groups.
2. **Cover all four SOCS features.** Even if the answer is "both are symmetric with no outliers", say it.
3. **Pick measures that match the shape.** Median and IQR for skewed data or data with outliers; mean and standard deviation for symmetric data.
4. **Compare percents, not counts.** Divide by each group's own total before you compare.
5. **Do not overclaim.** Overlapping distributions mean you can say one group tends to be higher, not that every member is.

Calculator tip: store each group in its own list and plot two boxplots on the same screen with the same window. Use the trace feature to read the quartiles for both groups quickly.

Common traps

(1) Describing each group in a separate paragraph without a single comparing statement. (2) Comparing counts instead of percents when group sizes differ. (3) Stating the mean and standard deviation for a skewed data set with outliers. (4) Claiming that one group caused the difference when the data only describe the groups.

Mistake Clinic

Spot the error

Student answer: "Brand P has a median of 12.75 hours and Brand Q has a median of 12.9 hours." **Where it fails:** it lists the values but does not compare them or say what the numbers mean. **Correct solution:** "The typical battery life is similar, with a slightly higher median for Brand Q (12.9 hours) than Brand P (12.75 hours)."

Spot the error

Student answer: "Class 2 has 55 students who study under 4 hours and Class 1 has only 16, so Class 2 students study less." **Where it fails:** the classes have different sizes (100 and 40 students). **Correct solution:** compare percents: 55 percent of Class 2 versus 40 percent of Class 1 study fewer than 4 hours a week.

Spot the error

Student answer: "The boxes of the two boxplots overlap, so there is no difference between the classes." **Where it fails:** overlap does not mean no difference; the medians and quartiles of Class B are higher, so Class B tends to score higher. **Correct solution:** state the differences in medians and quartiles and note that the groups overlap.

Real-World Connection

Sports analysts compare two players using the median and spread of points per game, not just the season average, because one blowout game can inflate a mean. Doctors comparing blood-pressure readings in two clinics look at both the typical value and the variability, since a clinic with the same average but wider spread has more patients at risk.

Challenge Corner

Group 1 has five-number summary 10, 20, 30, 40, 50 and Group 2 has 15, 25, 35, 45, 55. A student says the groups have the same spread. Is that correct, and what shift would make the medians equal? **Solution.** The IQR is 20 for each group and the range is 40 for each, so the spreads are equal; the groups differ only in center, with Group 2 shifted up by exactly 5 units at every quantile. Subtracting 5 from every value in Group 2 gives medians of 30 for both groups and identical distributions.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

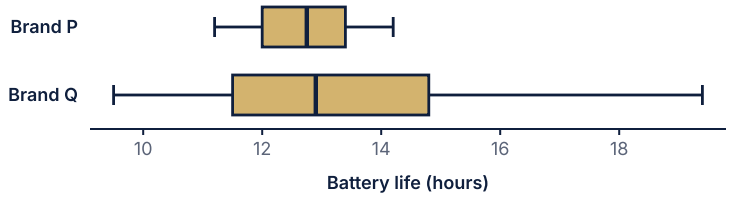
© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE Comparing Distributions in Context

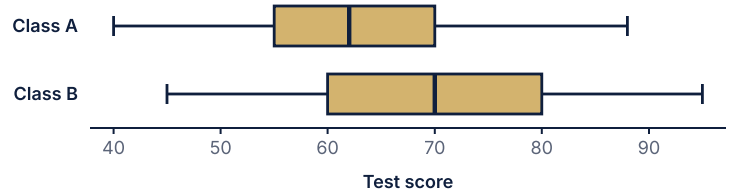
Show your work. A calculator is allowed. Justify conclusions in context.

1. Two classes have different sizes. Which comparison of the share of students who passed is most appropriate?
   (A) the number who passed in each class
   (B) the difference between the class sizes
   (C) the percent who passed in each class
   (D) the total number of students in both classes
2. Which statement is the best comparison?
   (A) The median score for Class A (82) is higher than for Class B (77).
   (B) Class A has a median of 82 and Class B has a median of 77.
   (C) Class A scored well on the test.
   (D) Class A has a larger value than Class B.
3. The boxplots compare battery life of two brands. Which statement is supported?



Battery life of 10 batteries from each brand.

   (A) Brand P is more variable than Brand Q.
   (B) Brand P has a much higher median than Brand Q.
   (C) Brand Q is more variable than Brand P.
   (D) Brand Q has no values below the median of Brand P.
4. A skewed distribution with an outlier is compared with another. Which pair of statistics is best for the comparison?
   (A) mean and standard deviation
   (B) median and IQR
   (C) mean and range
   (D) mode and range
5. In the boxplots, which statement is supported?



   (A) Every student in Class B scored higher than every student in Class A.
   (B) Class A has the larger median.
   (C) The two classes have the same range.
   (D) At least three quarters of Class A scored at or below Class B's median.
6. Two symmetric groups have the same mean but standard deviations of 3 and 12. Which conclusion is best?
   (A) They have the same distribution.
   (B) The first group has the larger median.
   (C) The second group has the larger mean.
   (D) They have the same center, but the second group is much more variable.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. Study hours: Class 1 (40 students) has 16 students under 4 hours; Class 2 (100 students) has 55 students under 4 hours. Which is correct?
   (A) A larger share of Class 1 studies under 4 hours because it has fewer students.
   (B) The shares are equal because both counts are above 10.
   (C) A larger share of Class 2 studies under 4 hours: 55 percent versus 40 percent.
   (D) Class 2 has 15 more students under 4 hours, so its share is 15 percent higher.
8. Which is NOT needed in a complete comparison of two distributions?
   (A) a comparison of centers
   (B) a statement that one group caused the other group's results
   (C) a comparison of spreads
   (D) a mention of shape and outliers
9. Team A's scores have mean 24 and median 20. Team B's scores have mean 22 and median 22. Which statement is best supported?
   (A) Team B's distribution is skewed right.
   (B) Team A's distribution is skewed left.
   (C) Neither distribution can be skewed.
   (D) Team A's distribution is more likely skewed right than Team B's.
10. Brand P batteries lasted 11.2, 11.8, 12.0, 12.4, 12.6, 12.9, 13.1, 13.4, 13.8, 14.2 hours and Brand Q batteries lasted 9.5, 10.8, 11.5, 12.2, 12.8, 13.0, 13.9, 14.8, 16.1, 19.4 hours. (a) Compare the medians. (b) Compare the ranges. (c) Write a one-sentence conclusion.
11. The boxplots of commute times (minutes) are Bus: 10, 18, 25, 35, 55 with an outlier at 70; Bike: 8, 12, 15, 19, 28; Car: 12, 20, 24, 30, 42 with outliers at 58 and 61. (a) Which mode has the smallest IQR and what is it? (b) Verify that the bus outlier of 70 is an outlier. (c) Which two modes have similar medians?
12. Class 1 (40 students) and Class 2 (100 students) have study-hour counts: Class 1 gives 4, 12, 14, 8, 2 and Class 2 gives 20, 35, 25, 15, 5 for bins 0-2, 2-4, 4-6, 6-8, 8-10. (a) Find the percent of each class studying 6 hours or more. (b) Compare the two distributions.
13. City A's monthly rainfall has mean 62 mm, median 60 mm, standard deviation 20 mm, IQR 28 mm. City B has mean 58, median 41, standard deviation 41, IQR 22. (a) What do the means and medians suggest about the shape of each city's distribution? (b) Which statistics would you use to compare typical rainfall, and why?
14. Two sections took the same quiz. Section 1: minimum 4, $Q_1 = 11$, median 15, $Q_3 = 19$, maximum 40. Section 2: minimum 6, $Q_1 = 13$, median 15, $Q_3 = 17$, maximum 25. (a) Test the maxima for outliers in each section. (b) Compare the two distributions.
15. Bakery X loaves have mean 500 g and standard deviation 3 g. Bakery Y loaves have mean 500 g and standard deviation 12 g. Both are roughly symmetric. (a) Compare the distributions. (b) Which bakery would a restaurant that needs consistent loaf weight prefer? Explain.

1.6 Percentiles, z-Scores, and Transforming Data

Find and interpret percentiles and z-scores, compare values from different distributions, and predict how adding a constant or multiplying by a constant changes center, spread, and shape.

Why This Matters

A score of 78 means little until you know what the scores look like. Was that high? Compared with whom? Percentiles and z-scores turn a raw value into a statement about position in a distribution, and they let you compare values that come from different scales, such as a time in seconds and a distance in meters.

Transformations ask a related question: if every value is converted to new units or shifted by a constant, what happens to the mean, median, standard deviation, and shape? This lesson belongs to **Exam Unit 1: Exploring One-Variable Data**; z-scores also return in the normal distribution and in regression, so they are worth mastering now.

Core Ideas

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Percentiles

The p**th percentile** of a distribution is the value with p percent of the observations at or below it. (Some books count only values strictly below; when a question depends on the difference, it states the rule.)

The first quartile is the 25th percentile, the median is the 50th, and the third quartile is the 75th.

A percentile describes **position**, not a score. Being at the 90th percentile means 90 percent of the group scored the same or lower.

A **cumulative relative frequency** table or graph shows the percent of values at or below each endpoint and is used to read percentiles.

To find the percentile of a value in a list, count the values at or below it and divide by the total. To find the value at a given percentile, find the corresponding position in the ordered list.

z-Scores (Standardized Values)

$$z = \frac{x - \text{mean}}{\text{standard deviation}}$$

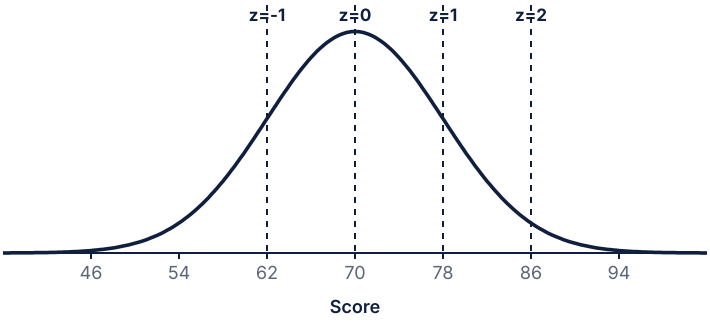
For a sample, $z = \dfrac{x - \bar{x}}{s}$; for a population, $z = \dfrac{x - \mu}{\sigma}$.

z tells how many standard deviations the value is **above** ($z > 0$) or **below** ($z < 0$) the mean.

To go backward, use $x = \text{mean} + z \cdot (\text{standard deviation})$.

Values with z near 0 are typical. Values with $|z|$ larger than about 2 are fairly unusual in a roughly symmetric, mound-shaped distribution, and 3 is very unusual; this is a guideline, not a test.

z-scores have no units, so they let you compare positions across different variables.



A z-score counts standard deviations from the mean.

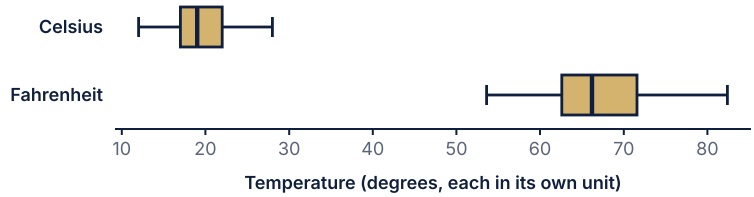
Linear Transformations

If every value is changed to $y = a + bx$ (add a, multiply by b), then:

Measures of center (mean, median, quartiles) change the same way: $\bar{y} = a + b\bar{x}$.

Measures of spread (standard deviation, IQR, range) are multiplied by $|b|$ and are **not** changed by adding a: $s_y = |b|\, s_x$.

The shape of the distribution is unchanged when $b > 0$ (a negative b flips the graph left to right). The z-score of each value is unchanged when $b > 0$.



Same data in two units: the shape is unchanged, the center shifts, and the spread is multiplied by 1.8.

Notice the logic: adding a constant slides the whole graph to the right or left (center moves, spread stays), while multiplying by a constant stretches the graph about zero (center and spread are both multiplied). Converting units such as Celsius to Fahrenheit, inches to centimeters, or dollars to cents is a linear transformation, and the rules give the new summary statistics without recomputing from the data.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Standardizing all the values of a data set with its own mean and standard deviation is itself a linear transformation. The resulting z-scores always have mean 0 and standard deviation 1, and the shape of the distribution is unchanged.

Worked Examples

EXAMPLE 1: PERCENTILE OF A VALUE

Twenty quiz scores: 55, 58, 62, 64, 66, 68, 70, 71, 73, 75, 76, 78, 80, 82, 84, 85, 88, 90, 92, 95. At what percentile is the score of 82?

SOLUTION

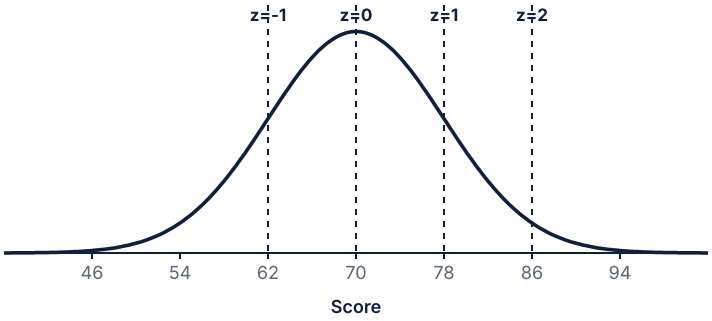
Count the scores at or below 82: 55 through 82 is 14 scores.

$14/20 = 0.70$.

Conclusion: A score of 82 is at the 70th percentile: 70 percent of the 20 students scored 82 or less.

EXAMPLE 2: COMPUTE AND INTERPRET A Z-SCORE

On a written exam the scores have mean 70 and standard deviation 8. Maya scored 78. Find her z-score and interpret it.



A z-score counts standard deviations from the mean.

SOLUTION

$z = \dfrac{78 - 70}{8} = 1$.

Conclusion: Maya's score is 1 standard deviation above the mean score of 70.

EXAMPLE 3: COMPARE ACROSS VARIABLES

Jia scored 85 on a math test (class mean 78, standard deviation 5) and 88 on a science test (class mean 80, standard deviation 8). On which test did she do better relative to her classmates?

SOLUTION

Math: $z = (85 - 78)/5 = 1.4$.

Science: $z = (88 - 80)/8 = 1.0$.

Conclusion: Jia's math score is 1.4 standard deviations above the mean, which is farther above average than her science score (1.0). She performed better relative to her class in math, even though her science score is higher.

EXAMPLE 4: WORK BACKWARD FROM A Z-SCORE

The lengths of songs in a library have mean 215 seconds and standard deviation 32 seconds. A song has a z-score of -1.5. How long is it, and is it unusual?

SOLUTION

$x = 215 + (-1.5)(32) = 215 - 48 = 167$ seconds.

A z-score of -1.5 is within 2 standard deviations of the mean, so the song is shorter than most but not unusually so.

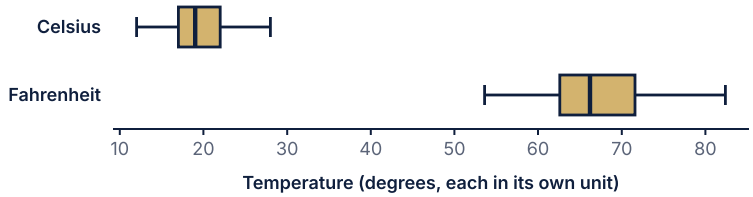
Conclusion: The song is 167 seconds long, 1.5 standard deviations below the mean.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: UNIT CONVERSION

Daily high temperatures in a city had mean 20, median 19, standard deviation 4, and IQR 5 degrees Celsius, with a five-number summary of 12, 17, 19, 22, 28. Find the corresponding statistics in degrees Fahrenheit using $F = 32 + 1.8C$.



Same data in two units: the shape is unchanged, the center shifts, and the spread is multiplied by 1.8.

SOLUTION

Mean: $32 + 1.8(20) = 68$. Median: $32 + 1.8(19) = 66.2$.

Standard deviation: $1.8(4) = 7.2$. IQR: $1.8(5) = 9$. Adding 32 does not change the spread.

The five-number summary becomes 53.6, 62.6, 66.2, 71.6, 82.4.

Conclusion: In Fahrenheit, the mean is 68 degrees, the median is 66.2 degrees, the standard deviation is 7.2 degrees, and the IQR is 9 degrees. The shape is unchanged.

EXAMPLE 6: DESIGN A TRANSFORMATION

Exam scores have mean 72 and standard deviation 10. A teacher wants new scores with mean 80 and standard deviation 12 using $y = a + bx$. Find a and b and check a score of 82.

SOLUTION

Spread: $s_y = |b|s_x$, so $12 = b(10)$ and $b = 1.2$.

Center: $\bar{y} = a + b\bar{x}$, so $80 = a + 1.2(72) = a + 86.4$ and $a = -6.4$.

Check: a score of 82 has $z = (82 - 72)/10 = 1$. The new score is $-6.4 + 1.2(82) = 92$, and its z-score is $(92 - 80)/12 = 1$.

Conclusion: Use $y = -6.4 + 1.2x$. Each student's z-score, and so relative position, is unchanged.

Write It Like a 5

Model Response

Prompt. Leo scored 83 on a written exam (mean 74, standard deviation 6) and 78 on a practical exam (mean 62, standard deviation 10). (a) On which exam did he do better relative to the group? (b) Interpret one of the z-scores. (c) The teacher adds 4 points to every written score. What happens to Leo's z-score on the written exam, and to the mean and standard deviation? **(a)** Written: $z = (83 - 74)/6 = 1.5$. Practical: $z = (78 - 62)/10 = 1.6$. Leo did slightly better relative to the group on the practical exam, because 1.6 standard deviations above the mean is farther above average than 1.5. **(b)** The practical exam z-score of 1.6 means Leo's score of 78 is 1.6 standard deviations above the mean practical score of 62. **(c)** Adding 4 points shifts every score. Leo's score becomes 87 and the mean becomes 78, so his z-score is still $(87 - 78)/6 = 1.5$. The mean increases by 4 to 78 and the standard deviation stays 6, because adding a constant does not change spread.

What the grader looks for

(1) Correct z-scores with the formula shown. (2) A comparison across exams that uses the z-scores and gives a conclusion in context. (3) A correct interpretation: standard deviations above the mean of a stated value. (4) Correct effect of adding a constant: center shifts by 4, spread unchanged, z-score unchanged. (5) Reasoning about why, not only the numbers.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. **Draw a quick number line.** Mark the mean and put the value on it; check that the sign of the z-score agrees with the picture.
2. **Label the pieces of the z formula.** Write x, the mean, and the standard deviation before you plug in.
3. **Percentile means position.** Say "X percent of the group scored at or below this value", with the group named.
4. **For transformations, split center from spread.** Center gets $a + b(\text{center})$. Spread gets $|b|(\text{spread})$.
5. **Check direction when lower is better.** For times in a race, a negative z-score is the better result.

Calculator tip: you can compute z-scores for a whole list at once by entering $(L1 - \text{mean})/s$ into a second list. The new list will have mean 0 and standard deviation 1, which is a quick check that you used the right formula.

Common traps

(1) Saying that a percentile is the percent of questions answered correctly. (2) Dividing by the mean instead of the standard deviation when computing z. (3) Changing the standard deviation when adding a constant, or forgetting to multiply it by the absolute value of b. (4) Declaring a higher raw score always better, ignoring the group mean and spread.

Mistake Clinic

Spot the error

Student claim: "I scored at the 80th percentile, so I got 80 percent correct." **Where it fails:** a percentile gives your rank within a group, not the percent of points earned. **Correct solution:** the 80th percentile means about 80 percent of the group scored the same as or lower than I did.

Spot the error

Student solution: "Adding 5 points to every test score changes the mean from 72 to 77 and the standard deviation from 10 to 15." **Where it fails:** adding a constant shifts every value by the same amount, so the distances between values do not change. **Correct solution:** the mean is 77 and the standard deviation stays 10.

Spot the error

Student solution: "Multiplying every value by -2 changes the standard deviation from 3 to -6." **Where it fails:** a standard deviation can never be negative. **Correct solution:** the new standard deviation is $|-2|(3) = 6$, and the mean is multiplied by -2 (the graph flips).

Real-World Connection

Pediatric growth charts report a child's height by percentile so that a parent can see where the child stands among children of the same age. Coaches use z-scores to compare athletes across events: a throw of 14.2 meters and a sprint of 11.8 seconds cannot be compared directly, but their z-scores can.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

A distribution of scores has mean 50 and standard deviation 10. A student finds that after a linear transformation $y = a + bx$ with $b < 0$, a score of 65 becomes 30 and a score of 40 becomes 40. Find the transformation, the new mean and standard deviation, and the new z-score of 65. **Solution.** From the two points, $30 = a + 65b$ and $40 = a + 40b$. Subtracting gives $-10 = 25b$, so $b = -0.4$, and $a = 40 - 40(-0.4) = 56$. The transformation is $y = 56 - 0.4x$. The new mean is $56 - 0.4(50) = 36$ and the new standard deviation is $|-0.4|(10) = 4$. The z-score of 65 was 1.5; the new value 30 has $z = (30 - 36)/4 = -1.5$, so the sign of each z-score has reversed because $b < 0$.

Practice

PRACTICE **Percentiles, z-Scores, and Transforming Data**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A student's score is at the 85th percentile. Which statement is correct?
   (A) The student answered 85 percent of the questions correctly.
   (B) About 85 percent of the scores in the group are at or below this score.
   (C) About 85 percent of the scores are above this score.
   (D) The student's score is 85.
2. The times of a group of runners have mean 50 seconds and standard deviation 4 seconds. A runner took 60 seconds. What is the z-score?
   (A) 10 **(B)** -2.5 **(C)** 2.5 **(D)** 0.4
3. Test 1: score 92, mean 80, standard deviation 6. Test 2: score 88, mean 70, standard deviation 8. Which statement is correct?
   (A) The score on Test 2 is relatively better, with z = 2.25 versus z = 2.
   (B) The score on Test 1 is relatively better because 92 is higher than 88.
   (C) The scores are equally good.
   (D) The score on Test 1 is relatively better because its mean is higher.
4. Salaries have mean 52 and standard deviation 9 (in thousands of dollars). Which salary has z-score -1.2?
   (A) 62.8 thousand dollars **(B)** 40.0 thousand dollars **(C)** 64.0 thousand dollars
   (D) 41.2 thousand dollars
5. Every value in a data set is increased by 10. Which statement is correct?
   (A) The mean and standard deviation both increase by 10.
   (B) The mean increases by 10 and the standard deviation is unchanged.
   (C) The mean is unchanged and the standard deviation increases by 10.
   (D) The mean increases by 10 and the standard deviation is multiplied by 10.
6. The IQR of a data set is 4. Every value is multiplied by 3. What is the new IQR?
   (A) 4 **(B)** 7 **(C)** 12 **(D)** 1.33
7. A data set has mean 12 and standard deviation 3. New values are $y = 2x - 5$. What are the new mean and standard deviation?
   (A) mean 19, standard deviation 6 **(B)** mean 19, standard deviation 1
   (C) mean 24, standard deviation 6 **(D)** mean 19, standard deviation 11
8. A data set is standardized by converting each value to a z-score. What are the mean and standard deviation of the z-scores?
   (A) mean 0 and standard deviation 0 **(B)** mean 1 and standard deviation 1
   (C) mean 0 and standard deviation 1 **(D)** mean 1 and standard deviation 0
9. The table shows cumulative relative frequencies at the end of each bin for a data set. Which is the best estimate of the percentile of the value 40?

Value up to	10	20	30	40	50	60
Cumulative relative frequency	0.08	0.30	0.55	0.72	0.90	1.00

   (A) 55th percentile **(B)** 72nd percentile **(C)** 40th percentile
   (D) 90th percentile

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. Use the 20 quiz scores 55, 58, 62, 64, 66, 68, 70, 71, 73, 75, 76, 78, 80, 82, 84, 85, 88, 90, 92, 95 (mean 75.6, $s \approx 11.46$). (a) Find the percentile of 90. (b) Find the z-score of 90 and interpret it. (c) Which student is at the 25th percentile?
11. A shot putter threw 14.2 meters (event mean 12.0, standard deviation 1.1). A sprinter ran 11.8 seconds in the 100 meter dash (event mean 12.6 seconds, standard deviation 0.5). (a) Find each z-score. (b) Which performance is better relative to the group? Remember that a lower time is better.
12. In a distribution of bag weights, a bag of 62 grams has $z = -1$ and a bag of 86 grams has $z = 2$. (a) Find the standard deviation. (b) Find the mean. (c) What is the z-score of a bag weighing 78 grams?
13. Lengths of boards in inches have mean 30, median 29, standard deviation 4, IQR 5.5. Convert to centimeters (1 inch = 2.54 cm). (a) Find the new mean, median, standard deviation, and IQR. (b) Does the shape change? (c) Does a board's z-score change?
14. Exam scores have mean 72 and standard deviation 10. The teacher applies $y = -6.4 + 1.2x$. (a) Find the new mean and standard deviation. (b) A student who scored 82 had z = 1. What is the student's new score and new z-score? (c) What does this show about transformations?
15. The ride times of a shuttle have mean 45 minutes and standard deviation 6 minutes. (a) Find the ride time with z = -2.3. (b) Is it unusual? (c) Adding a 5 minute boarding time to every ride gives a new variable. State its mean and standard deviation.

T Unit 1 Test

Check your command of one-variable data: variable types, graphs, shape, center, spread, outliers, comparisons, z-scores, and transformations.

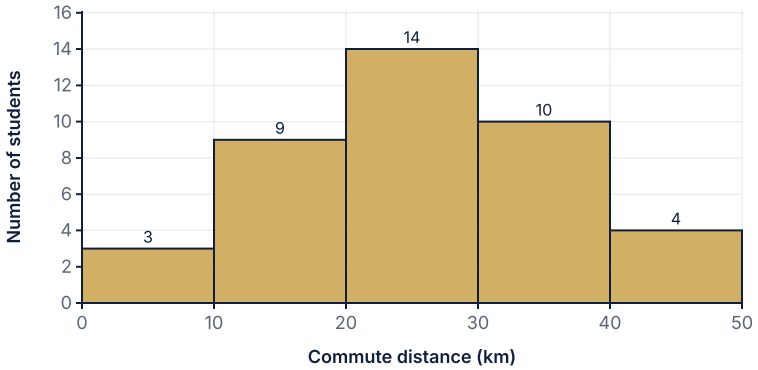
Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Part A: Multiple Choice

UNIT TEST **Unit 1 Test**

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. Which of the following is a categorical variable?
   (A) the jersey number of each player
   (B) the minutes each player was on the court
   (C) the points each player scored
   (D) the number of shots each player attempted
2. The Chess Club surveyed 60 members and 24 prefer online play. The Robotics Club surveyed 150 members and 30 prefer online play. Which statement is correct?
   (A) A larger share of Robotics Club members prefer online play because 150 is greater than 60.
   (B) The shares are the same because 24 and 30 are close.
   (C) A larger share of Chess Club members prefer online play (40 percent versus 20 percent).
   (D) The Robotics Club share is 30 percent.
3. The histogram shows the distances, in km, of 40 students' commutes (bins 0-10, 10-20, 20-30, 30-40, 40-50). What percent commute from 20 km up to (but not including) 40 km?



(A) 35 percent
(B) 60 percent
(C) 24 percent
(D) 14 percent

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

4. The stemplot shows 15 values (key 2 | 1 means 21). What is the median?

```
  1 | 2 5 7
  2 | 1 2 4 5 8 9
  3 | 1 3 6 8
  4 | 4
  5 | 2
Key: 2 | 1 means 21
```

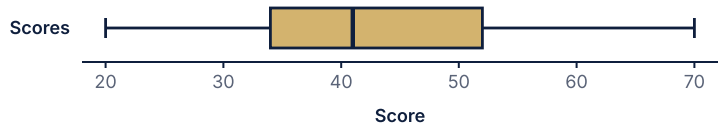
(A) 25 **(B)** 29 **(C)** 31 **(D)** 28

5. A data set has mean 38.4 and median 31. Which is the best description and choice of center?
   (A) Skewed left; the mean is the better measure of a typical value.
   (B) Symmetric; the mean and median are equally good.
   (C) Skewed right; the median is the better measure of a typical value.
   (D) Skewed right; the mean is the better measure of a typical value.

6. Delivery times have mean 24 minutes and standard deviation 5.2 minutes. Which interpretation is correct?
   (A) Delivery times typically differ from 24 minutes by about 5.2 minutes.
   (B) All delivery times are between 18.8 and 29.2 minutes.
   (C) The middle half of the delivery times differ by 5.2 minutes.
   (D) The average of the delivery times is 5.2 minutes.

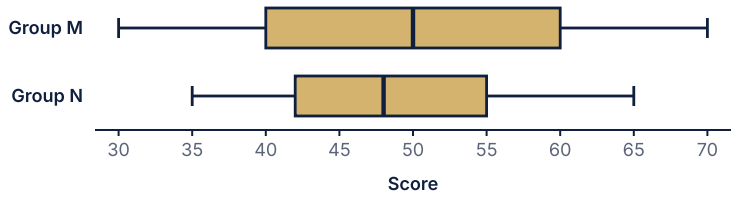
7. For the data 5, 7, 8, 10, 11, 13, 14, 16, 18, 40, which statement is correct?
   (A) 40 is not an outlier because the maximum is never an outlier.
   (B) 40 is an outlier because it is more than 2 standard deviations from the median.
   (C) 40 is an outlier because it is above the upper fence of 28.
   (D) 40 is not an outlier because the upper fence is 40.

8. A boxplot of 200 values has minimum 20, Q1 34, median 41, Q3 52, maximum 70. About how many values are greater than 52?



(A) 50 **(B)** 100 **(C)** 25 **(D)** 150

9. Which statement is supported by the boxplots of Group M and Group N?



   (A) Group N has a much higher median than Group M.
   (B) The medians are similar, but Group M is more variable than Group N.
   (C) Group N is more variable than Group M.
   (D) Every value in Group M is greater than every value in Group N.

10. A student's score is at the 80th percentile in a class of 50 students. About how many students scored at or below that score?
    (A) 10 **(B)** 80 **(C)** 45 **(D)** 40

11. On event A, Ana scored 41 (event mean 35, standard deviation 4). On event B, Ben scored 62 (event mean 50, standard deviation 10). Who did better relative to their event?
    (A) Ana, with z = 1.5 versus z = 1.2
    (B) Ben, because 62 is greater than 41
    (C) Ben, with z = 1.2 versus z = 1.5
    (D) They did equally well.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

12. A data set has mean 40 and standard deviation 12. Each value is transformed by $y = 5 + 0.5x$. What are the new mean and standard deviation?

    (A) mean 25, standard deviation 6 **(B)** mean 25, standard deviation 11

    (C) mean 45, standard deviation 6 **(D)** mean 20, standard deviation 6

13. Waiting times in minutes for 12 customers: 3, 4, 4, 5, 6, 7, 8, 9, 10, 12, 14, 29. [6 points]

    (a) Find the five-number summary. [2 points]

    (b) Use the 1.5 IQR rule to decide whether 29 is an outlier. Show the fences. [3 points]

    (c) Describe the shape of the distribution and say which measure of center is better. [1 point]

14. A teacher gave a 25-point quiz to two sections. The boxplots show Section 1 (minimum 8, Q1 12, median 15, Q3 19, maximum 24) and Section 2 (minimum 6, Q1 13, median 17, Q3 22, largest value that is not an outlier 30, and one high value of 38). Section 1 has mean 15.6 and standard deviation 3.8; Section 2 has mean 18.4 and standard deviation 7.5. [10 points]

    (a) Compare the two distributions of quiz scores. [3 points]

    (b) Show that 38 is an outlier in Section 2. [2 points]

    (c) A student in Section 1 scored 22 and a student in Section 2 scored 30. Which student did better relative to their section? Justify with z-scores. [2 points]

    (d) The teacher converts each score to a percent with $y = 4x$. Find the new mean and standard deviation of Section 1 and state what happens to the shape and to a student's z-score. [3 points]

Answers Exploring One-Variable Data: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

1.1 Data, Variables, and Categorical Graphs

PRACTICE · DATA, VARIABLES, AND CATEGORICAL GRAPHS

1. C. The count of messages takes whole-number values and can be averaged, so it is quantitative and discrete. Continuous is tempting because counts can be large, but you cannot send 3.5 messages.
2. A. An area code is a label for a region; averaging area codes is meaningless. The other three are measurements or counts with a meaningful average.
3. D. The relative frequency is 12/60 = 0.20. Using 12 as the percent confuses a count with a percent, and 25 percent wrongly divides 12 by 48, leaving out one of the other groups.
4. B. The percents total 130, so students are counted in more than one category and the categories are not parts of one whole. A bar graph, with each bar a percent of all students, is the right display.
5. C. Bar heights above the axis are 6 and 2, a 3 to 1 ratio, while the true ratio is 52/48, about 1.08. The axis must start at zero for bar lengths to be proportional.
6. B. Height is a measurement with a meaningful average. A jersey number is only a label, so it is categorical, as are position and team name.
7. D. Compare percents: 18/40 = 45 percent and 27/75 = 36 percent. Club B has more players in count but a smaller share; the last choice reverses the percents.
8. A. The non-bus riders are 24 + 9 + 21 = 54, and 54/90 = 0.60. The value 40 percent is the share who DO ride the bus (36/90), the group the question excludes.
9. C. Categories are separate, so bars have gaps and any order is allowed. Touching bars are for histograms of quantitative data. Heights add to 100 only if they are percents of one whole group.
10. (a) User ID: categorical (a label). Age: quantitative. Number of workouts: quantitative. Preferred workout type: categorical. Resting heart rate: quantitative.

    (b) Number of workouts is discrete (counts 0, 1, 2, ...). Age measured exactly and resting heart rate can take any value in an interval, so both are continuous; age recorded in whole years is often treated as discrete, so state your assumption.
11. (a) Coffee $84/200 = 0.42$; tea $46/200 = 0.23$; juice $30/200 = 0.15$; smoothie $40/200 = 0.20$. These sum to 1.00.

    (b) Coffee was the most popular drink, ordered by 42 percent of the 200 customers. Tea (23 percent) and smoothies (20 percent) were similar, and juice was the least common at 15 percent.
12. (a) $120/60 = 2$, so Y's bar looks twice as tall.

    (b) $1020/960 = 1.0625$, so Y has about 6.3 percent more customers.

    (c) The axis begins at 900 instead of 0, so bar lengths are not proportional to the counts. The graph exaggerates a modest difference; the axis should start at 0.
13. Class 1: $28/80 = 0.35$, or 35 percent. Class 2: $66/150 = 0.44$, or 44 percent. A larger share of Class 2 prefers morning sessions (44 percent versus 35 percent). The counts alone (28 versus 66) would also point to Class 2, but only because Class 2 is larger; the percents are the valid comparison.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. (a) The percents add to 196, so teens fall in several categories; a pie chart requires categories that partition one whole.
    (b) Use a bar graph with one bar per app showing the percent of the 75 teens. The counts are video $0.80(75) = 60$, messaging $0.72(75) = 54$, games $0.44(75) = 33$.
15. (a) Percents: wrong size 40, scratch 25, color 20, missing part 10, other 5 (they sum to 100). Draw bars from tallest to shortest with a vertical axis from 0 to 40 percent.
    (b) Wrong size and scratch: $40 + 25 = 65$ percent.
    (c) The mean of the five counts is $120/5 = 24$, but cause is categorical, so this average describes only the size of an average category. The most common cause is wrong size; that is the typical response.

1.2 Graphs for Quantitative Data: Dotplots, Stemplots, and Histograms

PRACTICE · GRAPHS FOR QUANTITATIVE DATA: DOTPLOTS, STEMPLOTS, AND HISTOGRAMS

1. B. A stemplot keeps each value as a leaf. A histogram shows only how many values fall in each bin, so the exact values are lost, and bar graphs and pie charts are for categorical data.
2. D. Dots at 8, 9, 10 number 5 + 2 + 2 = 9, and 9/20 = 0.45. Giving 9 percent reports the count as a percent; 55 percent is the share who scored below 8.
3. A. The last two bins hold 4 + 1 = 5 orders, and 5/40 = 0.125. Using 5 percent reports the count, and 10 percent counts only the 40 to 50 bin.
4. C. Stem 3 has the leaves 1, 3, 4, 6, 8, which is 5 values (31, 33, 34, 36, 38). Counting stem 2 would give 7, and the others miss leaves.
5. B. Most values are in the lower bins and the tail stretches toward larger values, so the distribution is skewed right. Skew is named for the side of the long tail, not for where the peak is.
6. D. Histograms use equal-width numerical bins that touch. Both graphs can show counts or percents, and neither has a required number of bars.
7. C. Narrower bins give more bars with smaller counts, so the graph looks ragged. The total of the bar heights is always the number of observations, 200.
8. A. 47 has stem 4 and leaf 7. Since 7 is in the range 5 to 9, it goes on the second row of stem 4.
9. B. Cumulative counts are 2, 7, 21, 38, so the 25th and 26th values fall in the 7 to 8 bin. The 6 to 7 bin has the second-largest count but only reaches 21.
10. (a) Stems 0 to 3. 0 | 8. 1 | 1 2 4 5 5 7 8 9. 2 | 1 2 2 6. 3 | 1 5. Key: 2 | 1 means 21 visits. There are 1+8+4+2=15 leaves.
    (b) The median is the 8th ordered value, 18 visits. Most members visited 11 to 22 times, with a few high values at 26, 31, and 35, so the shape is skewed right.
11. (a) $6 + 18 + 21 + 11 + 4 = 60$.
    (b) 0.10, 0.30, 0.35, 0.1833, 0.0667 (they sum to 1).
    (c) $11 + 4 = 15$ batteries, $15/60 = 0.25$, so 25 percent lasted at least 15 hours.
12. (a) Class A median is the 8th value, 75; Class B median is 79.
    (b) Class A range is $92 - 58 = 34$; Class B range is $91 - 62 = 29$, so Class A has the larger range.
    (c) Class B scored higher in the middle (median 79 versus 75), and its scores are a bit less variable, with no scores below 62 while Class A has scores of 58 and 61.
13. (a) 20 goes in the bin 20 to 30; 30 goes in the bin 30 to 40; 40 goes in the bin 40 to 50.
    (b) Without a rule the same data could produce different bar heights, so two readers could disagree about the counts and the apparent shape.
14. (a) $7 + 3 + 1 + 1 = 12$ students, $12/25 = 0.48$, so 48 percent.
    (b) The distribution is skewed right with a peak at 1 sibling. The value 6 siblings is separated from the others by a gap at 5 and may be an unusual observation.
15. (a) $2 + 9 + 14 + 10 + 4 + 1 = 40$.
    (b) $2 + 9 = 11$, $14 + 10 = 24$, $4 + 1 = 5$, total 40.
    (c) The original shows the peak at 8 to 12 (14 values) and a slow decline; the width-8 histogram has its peak at 8 to 16 with 24 values and cannot show where inside the interval the values cluster.

1.3 Describing Shape, Center, and Spread

PRACTICE · DESCRIBING SHAPE, CENTER, AND SPREAD

1. A. The long right tail pulls the mean toward larger values, so the mean exceeds the median. The reverse inequality holds for left skew.
2. B. The median and IQR depend only on the middle of the ordered data. The range uses the two extreme values, so it is not resistant, which rules out the last choice.
3. C. The sum is 50 and 50/5 = 10. The value 9 is the median, which is what you get if you pick the middle value instead of averaging.
4. D. Standard deviation measures the typical distance from the mean. It does not give the range of the data, which can be much wider than mean plus or minus 1 SD.
5. B. There are two separate tall bars (10 to 20 and 50 to 60) with a dip between them, so the shape is bimodal. The distribution is not unimodal even though it is roughly balanced.
6. C. The largest value does not affect the middle (5th) value, so the median is unchanged. The mean, standard deviation, and range all increase.
7. D. Both means are 20. P's values are farther from the mean (standard deviation 10) than Q's (standard deviation 2). Equal means say nothing about equal spread.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

8. A. A single extreme value changes the range but not the IQR. The IQR uses only the quartiles, so it is not based on every value, and it is never larger than the range.
9. C. The mean is 5 and the squared deviations are 9, 1, 1, 9, which sum to 20. Dividing by n - 1 = 3 gives variance 6.67 and s = 2.58. Dividing by 4 gives the population value 2.24; 6.67 is the variance, not the standard deviation.
10. (a) Mean $= 177/8 = 22.125$; median $= (21 + 23)/2 = 22$.
    (b) $s \approx 5.36$ (sum of squared deviations $= 200.9$, divided by 7, square root).
    (c) The mean and median are almost equal and the data are roughly symmetric with no outliers, so either is fine; report the mean 22.1 thousand steps with standard deviation 5.4 thousand.
11. (a) Mean $= 221/8 \approx 27.6$; median $= (18 + 20)/2 = 19$.
    (b) The value 95 is an outlier and the distribution is skewed right, so the median of 19 minutes is more representative; the mean exceeds all but one of the data values.
    (c) Without 95 the sum is 126 for 7 values, so the mean falls to 18, a drop of about 9.6 minutes. The median would change only to 18.
12. (a) The mean is below the median, which suggests the distribution is skewed left (a tail toward low scores).
    (b) The median of 81 is more representative because the low scores pull the mean down.
    (c) A few students who missed the quiz or did not study scored very low, while most scored in the 80s and 90s.
13. (a) The distribution is bimodal.
    (b) The mean would fall near midday, when few people arrive, so it does not describe a typical arrival.
    (c) Report both modes (the morning and afternoon clusters) and describe the spread within each; do not collapse them into one center.
14. (a) Both machines fill about 500 grams on average, but Y's fills vary much more than X's.
    (b) Machine X is better because its fills are more consistent.
    (c) The fills from Machine Y typically differ from the mean of 500 grams by about 9 grams.
15. (a) Mode $= 1$ pet (9 families). The median is the average of the 13th and 14th values; the cumulative counts are 6, 15, so both are 1 and the median is 1.
    (b) Mean $= (0 \cdot 6 + 1 \cdot 9 + 2 \cdot 7 + 3 \cdot 3 + 4 \cdot 1)/26 = 36/26 \approx 1.38$.
    (c) The mean (1.38) exceeds the median (1), and the tail runs toward 3 and 4 pets, so the distribution is skewed right.

1.4 The Five-Number Summary, Boxplots, and Outliers

PRACTICE · THE FIVE-NUMBER SUMMARY, BOXPLOTS, AND OUTLIERS

1. C. The five-number summary uses the two extremes, the median, and the two quartiles. It does not include the mean or standard deviation, which are not resistant.
2. A. IQR is $Q_3 - Q_1 = 39 - 24 = 15$. Adding the quartiles gives 63, and 31.5 is their average.
3. D. IQR = 16, so the upper fence is 46 + 24 = 70. Only a value above 70 is an outlier, so 72 is flagged. The value 70 is exactly on the fence and is not flagged.
4. B. The values between Q1 = 34 and Q3 = 52 are the middle half, so about 0.5(80) = 40 students. A quarter, 20, is the amount in one section of the box.
5. C. The upper whisker is much longer than the lower one and the median is near Q1, so the tail is on the right. A boxplot cannot show modes, so bimodal is unsupported.
6. A. The IQR depends on the quartiles, which do not change when only the maximum moves. The range, mean, and standard deviation all increase.
7. D. With n = 8 the lower half is 2, 4, 6, 8, whose median is (4 + 6)/2 = 5. Choosing 4 or 6 uses only one of the two middle values.
8. A. Whiskers stop at the most extreme values inside the fences. Outliers are plotted separately beyond the whiskers.
9. B. A boxplot is drawn from the five numbers alone, so the boxplots match. The data can differ in shape, mean, and standard deviation (Example 5).
10. (a) Minimum 22, $Q_1 = 28$, median $(33 + 35)/2 = 34$, $Q_3 = 40$, maximum 71.
    (b) $\text{IQR} = 12$, $1.5(12) = 18$, fences 10 and 58. The value 71 is above 58, so it is an outlier; the upper whisker ends at 46.
    (c) The times are skewed right with one high outlier (71 minutes). The median is 34 minutes and the middle half of the times lie between 28 and 40 minutes.
11. (a) $\text{IQR} = 71 - 58 = 13$.
    (b) $1.5(13) = 19.5$; lower fence 38.5, upper fence 90.5. The minimum 45 is above 38.5 (not an outlier); the maximum 99 is above 90.5, so it is an outlier.
    (c) About 25 percent of the 60 students (about 15 students) scored below 58.
12. (a) Both have median 30, $Q_1 = 21$ ($(20 + 22)/2$ and $(12 + 30)/2$), $Q_3 = 39$ ($(38 + 40)/2$ for both), minimum 10, maximum 50.
    (b) Set Y has a spike of four values at 30, while X is spread out evenly; a boxplot cannot show clusters, gaps, or modes.
13. (a) For the 13 remaining values the median is the 7th, 33. Lower half 22, 25, 27, 28, 30, 31 gives $Q_1 = 27.5$; upper half 35, 36, 38, 40, 43, 46 gives $Q_3 = 39$; $\text{IQR} = 11.5$.
    (b) The IQR changed from 12 to 11.5; the range changed from 49 to 24.
    (c) The range changed most because it depends directly on the extreme value; the IQR is resistant.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. (a) The ordered list has 20 values. The median is the average of the 10th and 11th values, both 3, so the median is 3. Q_1 is the average of the 5th and 6th values (both 2), so $Q_1 = 2$. Q_3 is the average of the 15th and 16th values (both 4), so $Q_3 = 4$. Summary: 1, 2, 3, 4, 5.
    (b) $\text{IQR} = 4 - 2 = 2$.
    (c) Fences are $2 - 3 = -1$ and $4 + 3 = 7$; no value is outside, so there are no outliers.
15. (a) $\text{IQR} = 6$, $1.5(6) = 9$, upper fence $= 9 + 9 = 18$. Since $40 > 18$, 40 is an outlier.
    (b) 19 is also greater than 18, so 19 is an outlier too (it is not inside the fence), meaning the upper whisker ends at a smaller value.
    (c) The median (5) is closer to Q_1 (2 away) than Q_3 (4 away), and there are large high values, so the distribution is skewed right.

1.5 Comparing Distributions in Context

PRACTICE · COMPARING DISTRIBUTIONS IN CONTEXT

1. C. Percents adjust for group size. Counts favor the larger class and would mislead.
2. A. A good comparison uses a comparing word and gives numbers for both groups in context. The second choice only lists values, and the others lack numbers or context.
3. C. Brand Q's box and range are longer (IQR 3.3 versus 1.4 hours), so it is more variable. The medians (12.9 and 12.75) are nearly equal, and Brand Q's minimum, 9.5, is below Brand P's median.
4. B. The median and IQR are resistant to skew and outliers. The mean, standard deviation, and range are all pulled by extreme values.
5. D. Class A's third quartile is 70, equal to Class B's median, so at least 75 percent of Class A scored at or below 70. The ranges overlap heavily, so not everyone in B beat everyone in A.
6. D. Equal means give the same center for symmetric data, but the standard deviations differ, so the spreads differ. They are not the same distribution.
7. C. Shares are 55/100 = 55 percent and 16/40 = 40 percent. The count difference of 39 reflects the larger class size, not a larger share.
8. B. Descriptive comparison of SOCS features is required. Causal claims cannot be made from a comparison of distributions alone.
9. D. Team A's mean exceeds its median, which suggests right skew. Team B's mean equals its median, which suggests symmetry. Skewed left would have mean below median.
10. (a) The median for P is $(12.6 + 12.9)/2 = 12.75$ hours and for Q is $(12.8 + 13.0)/2 = 12.9$ hours, nearly the same.
    (b) The range for P is $14.2 - 11.2 = 3.0$ hours; for Q it is $19.4 - 9.5 = 9.9$ hours, so Q is much more variable.
    (c) The brands have similar typical battery life, but Brand P is far more consistent.
11. (a) Bike: $19 - 12 = 7$ minutes (bus 17, car 10).
    (b) Bus IQR $= 35 - 18 = 17$; the upper fence $= 35 + 1.5(17) = 60.5$. Since $70 > 60.5$, 70 is an outlier.
    (c) Bus (25) and car (24).
12. (a) Class 1: $(8 + 2)/40 = 25\%$. Class 2: $(15 + 5)/100 = 20\%$.
    (b) Class 1 is centered in the 4-6 bin (35 percent), and Class 2 is centered in the 2-4 bin (35 percent). Class 1 students tend to study more hours; 25 percent study 6 or more hours versus 20 percent in Class 2. Both distributions are unimodal and roughly symmetric to slightly right-skewed.
13. (a) City A's mean and median are close, so it is roughly symmetric. City B's mean exceeds its median by 17 mm, which suggests right skew from a few very wet months.
    (b) Use the medians (60 versus 41 mm) and the IQRs (28 versus 22 mm), because City B is skewed and its mean and standard deviation are inflated by the wet months.
14. (a) Section 1: $\text{IQR} = 8$, upper fence $19 + 12 = 31$; 40 is an outlier. Section 2: $\text{IQR} = 4$, upper fence $17 + 6 = 23$; 25 is an outlier.
    (b) The sections have the same median (15), but Section 1 is more spread out (IQR 8 versus 4). Both are roughly symmetric in the middle with a high outlier (40 and 25).
15. (a) The two distributions have the same center (500 g), but Bakery Y has four times the standard deviation, so its loaf weights vary much more.
    (b) Bakery X, because its loaves typically differ from 500 g by only about 3 g, compared with 12 g for Bakery Y, so it provides more consistent weights.

1.6 Percentiles, z-Scores, and Transforming Data

PRACTICE · PERCENTILES, Z-SCORES, AND TRANSFORMING DATA

1. B. A percentile is a statement about position within a group. The 85th percentile means about 85 percent of the group scored the same or lower, not that 85 percent of the questions were correct.
2. C. $z = (60 - 50)/4 = 2.5$. The value 10 is only the deviation from the mean, not divided by the standard deviation; 0.4 inverts the ratio.
3. A. Test 1: z = (92 - 80)/6 = 2. Test 2: z = (88 - 70)/8 = 2.25. The higher z-score shows the better standing relative to the class; raw scores are not comparable.
4. D. $x = 52 + (-1.2)(9) = 52 - 10.8 = 41.2$. The value 62.8 is $z = +1.2$; 40 uses 52 - 12.
5. B. Adding a constant shifts the distribution but does not change distances between values, so the spread is unchanged.
6. C. Multiplying every value by 3 multiplies the spread by 3, so the IQR is 12. Adding 3 would give 7, and dividing by 3 gives 1.33.
7. A. Mean: 2(12) - 5 = 19. Standard deviation: $|2|(3) = 6$. The constant -5 does not affect spread, which rules out 1 and 11.
8. D. Subtracting the mean makes the new mean 0 and dividing by the standard deviation makes the new standard deviation 1.
9. B. The cumulative relative frequency at 40 is 0.72, so 72 percent of the values are at or below 40. The 55th percentile is the percentile of 30, and 90 is for 50.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. (a) 18 of 20 scores are at or below 90, so $18/20 = 90\%$: the 90th percentile.
    (b) $z = (90 - 75.6)/11.46 \approx 1.26$. The score of 90 is about 1.26 standard deviations above the mean of 75.6.
    (c) Five of the 20 scores are at or below 66 ($5/20 = 25\%$), so the student who scored 66 is at the 25th percentile.
11. (a) Shot put: $z = (14.2 - 12.0)/1.1 = 2.0$. Sprint: $z = (11.8 - 12.6)/0.5 = -1.6$.
    (b) The shot put is 2.0 standard deviations above average, which is better (farther) than average. The sprint is 1.6 standard deviations below the average time, which is faster than average, so better. The shot put performance is the more unusual and therefore relatively better (2.0 versus 1.6 standard deviations in the good direction).
12. (a) The two bags are $2 - (-1) = 3$ standard deviations apart and differ by $86 - 62 = 24$ grams, so $s = 24/3 = 8$ grams.
    (b) Mean $= 62 + 1(8) = 70$ grams (since $z = -1$ means one standard deviation below the mean).
    (c) $z = (78 - 70)/8 = 1$.
13. (a) Mean $2.54(30) = 76.2$ cm; median $2.54(29) = 73.66$ cm; standard deviation $2.54(4) = 10.16$ cm; IQR $2.54(5.5) = 13.97$ cm.
    (b) No; multiplying by a positive constant does not change the shape.
    (c) No; for $b > 0$ both the numerator and denominator of z scale by b, so z-scores are unchanged.
14. (a) New mean $-6.4 + 1.2(72) = 80$; new standard deviation $1.2(10) = 12$.
    (b) New score $-6.4 + 1.2(82) = 92$; $z = (92 - 80)/12 = 1$.
    (c) A linear transformation with $b > 0$ changes the scale but not a value's relative position, so its z-score is unchanged.
15. (a) $x = 45 - 2.3(6) = 31.2$ minutes.
    (b) It is more than 2 standard deviations below the mean, so it is a fairly unusually short ride if the distribution is roughly mound-shaped.
    (c) Mean $= 45 + 5 = 50$ minutes; standard deviation is unchanged at 6 minutes.

T Unit 1 Test

UNIT TEST · UNIT 1 TEST

1. A. A jersey number only labels a player; averaging jersey numbers is meaningless. The other three are counts or measurements.
2. C. Compare percents: 24/60 = 40 percent and 30/150 = 20 percent. Counts alone do not account for the different group sizes.
3. B. The bins 20-30 and 30-40 hold 14 + 10 = 24 students, and 24/40 = 0.60. Giving 24 reports the count, and 35 percent is only the 20-30 bin (14/40).
4. D. With n = 15 the median is the 8th value in order. Counting leaves from the top gives 12, 15, 17, 21, 22, 24, 25, 28, so the median is 28. Choosing 25 stops one value early (the 7th).
5. C. A mean above the median signals a long right tail, and the mean is pulled by that tail while the median resists it, so the median is better for a typical value.
6. A. The standard deviation is a typical distance from the mean. It does not bound the data, and the middle half is described by the IQR.
7. C. The median is 12, Q1 = 8, Q3 = 16, IQR = 8, and the upper fence is 16 + 12 = 28. Since 40 > 28 it is an outlier; the 1.5 IQR rule is the criterion, not a standard deviation count.
8. A. Q3 is the 75th percentile, so about 25 percent of 200 values lie above it: 50. The value 100 corresponds to values above the median.
9. B. The medians are 50 and 48, which are close. Group M's IQR is 20 (40 to 60) and Group N's is 13 (42 to 55), so M is more variable.
10. D. 80 percent of 50 students is 40. The value 10 is the number scoring above the student, and 80 confuses the percentile with a count.
11. A. Ana: z = (41 - 35)/4 = 1.5. Ben: z = (62 - 50)/10 = 1.2. The larger z-score indicates a higher standing within the group; raw scores cannot be compared.
12. A. Mean: 5 + 0.5(40) = 25. Standard deviation: 0.5(12) = 6; adding 5 does not change spread, so 11 is wrong.
13. Scoring guide.
    (a) 2 points: median $= (7 + 8)/2 = 7.5$ (1); minimum 3, $Q_1 = (4 + 5)/2 = 4.5$, $Q_3 = (10 + 12)/2 = 11$, maximum 29 (1).
    (b) 3 points: $\text{IQR} = 11 - 4.5 = 6.5$ (1); fences $4.5 - 1.5(6.5) = -5.25$ and $11 + 9.75 = 20.75$ (1); 29 exceeds 20.75, so it is an outlier (1).
    (c) 1 point: the distribution is skewed right with a high outlier; the median (7.5 minutes) is the better measure of center because the mean (9.25) is pulled up by the outlier.
    Model response. The five-number summary is 3, 4.5, 7.5, 11, 29. The IQR is 6.5, so the lower fence is -5.25 and the upper fence is 20.75. The waiting time of 29 minutes is above 20.75, so it is an outlier. The waiting times are skewed to the right, so the median of 7.5 minutes is a better description of a typical wait than the mean of about 9.25 minutes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. Scoring guide.

(a) 3 points: center compared with numbers and a comparing word (Section 2's median 17 is higher than Section 1's 15) (1); spread compared (Section 2 IQR 9 versus 7 for Section 1, and a larger range) (1); shape and outlier mentioned in context (both roughly symmetric to slightly right-skewed; Section 2 has a high outlier) (1).

(b) 2 points: $\text{IQR} = 22 - 13 = 9$ and upper fence $22 + 1.5(9) = 35.5$ (1); $38 > 35.5$, so it is an outlier (1).

(c) 2 points: Section 1 student $z = (22 - 15.6)/3.8 \approx 1.68$; Section 2 student $z = (30 - 18.4)/7.5 \approx 1.55$ (1); the Section 1 student did better relative to the section, being 1.68 standard deviations above the mean versus 1.55 (1).

(d) 3 points: new mean $4(15.6) = 62.4$ percent (1); new standard deviation $4(3.8) = 15.2$ percentage points (1); the shape is unchanged and each student's z-score is unchanged, since multiplying by a positive constant rescales center and spread equally (1).

Model response. (a) Section 2 has a higher median (17 versus 15) and is more variable (IQR 9 versus 7; range 32 versus 16). Both distributions are roughly symmetric, but Section 2 has a high outlier of 38. (b) The upper fence is $22 + 1.5(9) = 35.5$ and 38 is above it. (c) The Section 1 student has $z \approx 1.68$ and the Section 2 student has $z \approx 1.55$, so the Section 1 student did better relative to their section. (d) The new mean is 62.4 percent and the new standard deviation is 15.2 percentage points; the shape and z-scores do not change.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 2: Density Curves and Normal Distributions

Model data with the normal curve and answer percent and percentile questions. 4 lessons, 74 practice and test items, all with full explanations.

In this unit · Exam Unit 1

2.1 Density Curves and the Empirical Rule
2.2 Normal Calculations: Finding Areas
2.3 Inverse Normal and Finding Cutoffs
2.4 Assessing Normality and Using Normal Models
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

2.1 Density Curves and the Empirical Rule

Learn how density curves turn histograms into areas, where the mean and median sit on skewed curves, and how the 68-95-99.7 rule and z-scores describe normal distributions.

Why This Matters

Real data sets are bumpy, but many of them follow a smooth, bell-like pattern so closely that we can answer questions about them with a curve instead of a list of numbers. A **density curve** is that curve. It lets you say things like "about 16 percent of the batteries last longer than 425 minutes" without counting a single battery.

Students most often go wrong by treating the height of the curve as a probability, by placing the mean and median in the wrong order on a skewed curve, or by applying the 68-95-99.7 rule to data that are not bell-shaped. This skill sits in the exam's Unit 1 (Exploring One-Variable Data) and returns in the probability and sampling units, where the normal curve is the model for almost everything.

Core Ideas

A histogram shows how many observations fall in each bin. If you scale a histogram so its total area is 1 and then imagine bins shrinking to nothing, the outline becomes a smooth curve. That curve is a density curve.

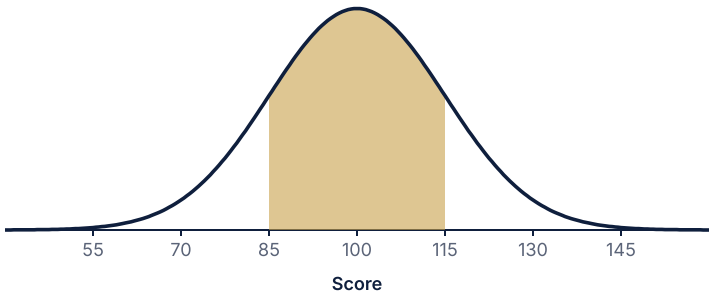
Density curve: definition

A density curve is a curve that (1) lies on or above the horizontal axis and (2) has total area exactly 1 underneath it.
The area under the curve between two values a and b is the proportion of observations that fall between a and b.
The curve describes the overall pattern of a distribution. It is a model, so it is an idealization of the real data.

Notice what this means. Height is not a probability. Only **area** is. The area above a single exact value is zero, so for a continuous model $P(X = a) = 0$, and $P(X \le a)$ equals $P(X < a)$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



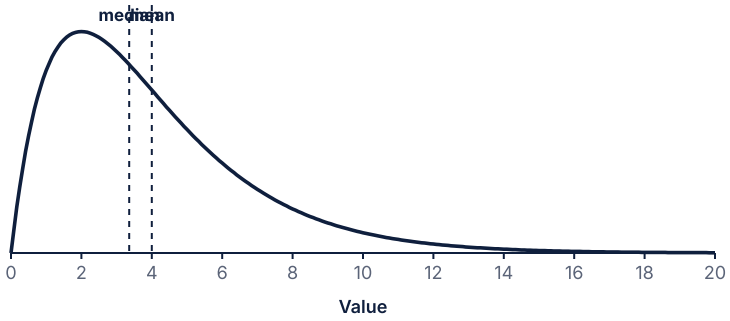
The shaded area is the proportion of scores between 85 and 115, about 0.68.

Median and mean on a density curve

The **median** is the equal-areas point: half the area lies to its left and half to its right.

The **mean** is the balance point of the curve, the place where a solid cutout of the curve would balance on a pencil.

Symmetric curve: mean = median. Right-skewed curve: the long right tail pulls the mean above the median. Left-skewed curve: the mean is below the median.



A right-skewed density curve. The tail pulls the mean (4.0) to the right of the median (about 3.36).

Use the symbols μ and σ for the mean and standard deviation of a density curve, because they describe an idealized model rather than a sample. Use $\bar{x}$ and s for a sample of actual data.

The Normal Curve and the 68-95-99.7 Rule

The **normal distributions** are a family of symmetric, single-peaked, bell-shaped density curves. Each one is completely determined by its mean μ (the center) and its standard deviation σ (the spread). We write $N(\mu, \sigma)$. Changing μ slides the bell left or right; changing σ makes it wider or narrower.

The 68-95-99.7 (empirical) rule

In a normal distribution, about **68%** of observations lie within 1 standard deviation of the mean: between $\mu - \sigma$ and $\mu + \sigma$.

About **95%** lie within 2 standard deviations, and about **99.7%** lie within 3.

By symmetry, the tails are small: about 16% above $\mu + \sigma$, 2.5% above $\mu + 2\sigma$, and 0.15% above $\mu + 3\sigma$.

Interval	Percent inside	Percent in each tail beyond
$\mu \pm 1\sigma$	68%	16%
$\mu \pm 2\sigma$	95%	2.5%
$\mu \pm 3\sigma$	99.7%	0.15%

Every one of those percents is a rounded value. Exact areas for other cutoffs need the standard normal table or a calculator, which is the topic of the next lesson.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Standardizing: the z-score

$z = \dfrac{x - \mu}{\sigma}$ tells how many standard deviations x is from the mean. Positive means above, negative means below.

If x comes from $N(\mu, \sigma)$, then z comes from the **standard normal** distribution $N(0, 1)$.

Z-scores let you compare values from different scales, such as a 640 on one test and an 87 on another.

Worked Examples

AREA AS A PROPORTION ON A RECTANGLE

A density curve for the waiting time X (in minutes) of a school bus is a flat line of height 0.1 over the interval from 0 to 10. (a) Verify that this is a valid density curve. (b) Find the proportion of waits between 2 and 5 minutes. (c) Find the proportion of waits longer than 7.5 minutes. (d) State the median and mean.

SOLUTION

(a) The curve is above the axis, and the area is width times height $= 10(0.1) = 1$. Both conditions hold.

(b) Area $= (5 - 2)(0.1) = 0.30$. About 30% of waits are between 2 and 5 minutes.

(c) Area $= (10 - 7.5)(0.1) = 0.25$, so 25% of waits exceed 7.5 minutes.

(d) The curve is symmetric, so the median and mean are both 5 minutes. The median splits the area into 0.5 and 0.5.

MEDIAN AND MEAN ON A TRIANGLE

A density curve for the distance a runner covers in a drill is the line $f(x) = x/2$ over $0 \le x \le 2$ (kilometers), and 0 elsewhere. The area of a triangle is $\frac{1}{2}bh$. (a) Show that the total area is 1. (b) Find the proportion of values below 1. (c) Find the median. (d) The mean is $4/3$. Compare it with the median and describe the skew.

SOLUTION

(a) Base 2, height $f(2) = 1$: area $= \frac{1}{2}(2)(1) = 1$.

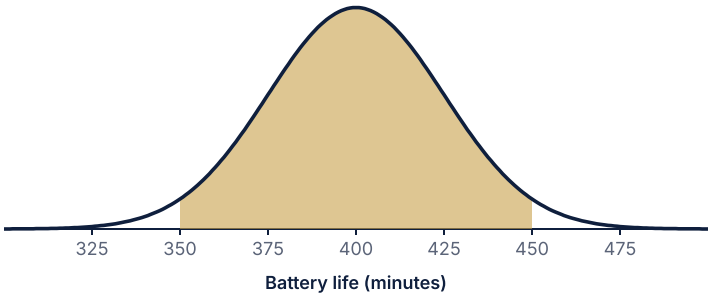
(b) Between 0 and 1 the triangle has base 1 and height $f(1) = 0.5$: area $= \frac{1}{2}(1)(0.5) = 0.25$. Only 25% of values are below 1.

(c) The area to the left of m is $\frac{1}{2}m \cdot (m/2) = m^2/4$. Setting $m^2/4 = 0.5$ gives $m = \sqrt{2} \approx 1.414$ km.

(d) The mean $4/3 \approx 1.333$ is **less** than the median 1.414, so the curve is skewed left (long tail toward 0). The balance point is pulled toward the tail.

USING THE EMPIRICAL RULE

The battery life of a phone model is approximately normal with mean 400 minutes and standard deviation 25 minutes. (a) Between what values do the middle 95% of battery lives fall? (b) What percent of phones last longer than 450 minutes? (c) What percent last less than 375 minutes?



Middle 95% of battery lives.

SOLUTION

Define X = battery life, $X \sim N(400, 25)$. The curve is bell shaped, so the rule applies.

(a) $400 \pm 2(25)$ gives 350 to 450 minutes.

(b) 450 is $\mu + 2\sigma$. The 95% inside leaves 5% outside, split equally, so 2.5% last longer than 450 minutes.

(c) 375 is $\mu - \sigma$. The 68% inside leaves 32% outside, so 16% last less than 375 minutes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

COMPARING WITH Z-SCORES

Mia scored 640 on a reading test with mean 580 and standard deviation 40. She scored 87 on a science test with mean 78 and standard deviation 6. On which test did she do better relative to the other students?

SOLUTION

Reading: $z = (640 - 580)/40 = 1.5$.

Science: $z = (87 - 78)/6 = 1.5$.

The two z-scores are equal, so Mia performed equally well relative to each group: both scores are 1.5 standard deviations above the mean. Raw scores cannot be compared directly because the scales differ.

RECOVERING THE MEAN AND STANDARD DEVIATION

The lengths of a species of fish in a lake are approximately normal. The middle 95% of lengths run from 48 cm to 56 cm. (a) Find the mean and standard deviation. (b) About what percent of fish are longer than 54 cm?

SOLUTION

(a) The center of the middle 95% is the mean: $\mu = (48 + 56)/2 = 52$ cm. The interval is $\mu \pm 2\sigma$, so $4\sigma = 8$ and $\sigma = 2$ cm.

(b) 54 is $\mu + \sigma$, so about 16% of fish are longer than 54 cm.

COMBINING PIECES OF THE RULE AND CHECKING DATA

Test scores for 400 students are approximately $N(70, 8)$. (a) What percent score between 62 and 94? (b) About how many students would you expect to score above 86? (c) A sample of 25 package weights has mean 51.2 and standard deviation 5.1 grams. If 17 of the 25 weights lie within one standard deviation of the mean and 24 within two, is the normal model reasonable?

SOLUTION

(a) 62 is $\mu - 1\sigma$ and 94 is $\mu + 3\sigma$. Area from -1σ to the mean is 34%; from the mean to $+3\sigma$ is 49.85%. Total $\approx 83.85\%$.

(b) 86 is $\mu + 2\sigma$, so 2.5% score above it: $0.025(400) = 10$ students.

(c) $17/25 = 68\%$ within 1 sd (rule says 68%) and $24/25 = 96\%$ within 2 sd (rule says 95%). Those proportions match the rule closely, so a normal model is reasonable, though a graph should also be checked.

Write It Like a 5

Model Response

Prompt: The amount of juice a machine pours into a bottle is approximately normal with mean 500 ml and standard deviation 4 ml. Bottles with less than 492 ml are rejected. (a) What percent of bottles are rejected? (b) Between what two amounts do the middle 99.7% of bottles fall? (a) Let X be the volume, $X \sim N(500, 4)$. The cutoff 492 is $(492 - 500)/4 = -2$, so it is 2 standard deviations below the mean. By the 68-95-99.7 rule, 95% of bottles are within 2 standard deviations, so 5% are outside, and by symmetry 2.5% are below 492 ml. About **2.5% of bottles are rejected**. (b) The middle 99.7% lie within 3 standard deviations: $500 \pm 3(4)$, which is **488 ml to 512 ml**.

What the grader looks for

1 point: identifies the distribution and parameters in context. 1 point: converts the cutoff to a z-score or standard deviations from the mean. 1 point: correctly uses the symmetry or the 95% to reach 2.5% in one tail. 1 point: states the answer in context with units and percent. 1 point: uses 3 standard deviations and gives both endpoints with units.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Name the variable and distribution: say what X is and write $N(\mu, \sigma)$ with units.
2. Sketch and label a bell curve with the mean at the center and tick marks every σ.
3. Shade the region the question asks about.
4. Compute area from the rule, or from z-scores if the cutoff is not a whole number of standard deviations.
5. Answer in a sentence using the context and units.

- A quick check on any normal problem: the mean is the center, so exactly 50% is on each side.
- The mean and median of a density curve differ only if the curve is skewed.
- If you can't sketch the shape, you can't use the rule: look at a graph first.
- When someone gives you the middle 95%, the endpoints are $\mu \pm 2\sigma$: use them to find μ and σ immediately.

Common traps

1. Reading the height of the curve as a probability. Only area counts. 2. Putting the mean below the median on a right-skewed curve. The tail drags the mean toward the tail. 3. Saying "68% above the mean minus one sd." 68% is the middle; the part above $\mu - \sigma$ is 84%. 4. Using the rule for data that are clearly skewed or have gaps. Check the shape first.

Mistake Clinic

Spot the error

A student says: "A density curve for a variable has height 3 at its peak, so there is a 300% chance of the peak value." Error: height is not probability; only area is. A density curve can exceed 1 in height if it is narrow, as long as the total area is 1. Correct: the probability of any exact value is 0; probabilities come from areas between two values.

Spot the error

A student is told scores are $N(60, 5)$ and says that 95% of scores are between 55 and 65. Error: 55 to 65 is $\mu \pm 1\sigma$, which is the 68% interval. Correct: 95% is $\mu \pm 2\sigma$, so 50 to 70.

Spot the error

A student sees a strongly right-skewed income histogram and says that since the mean is $\$52$ thousand, 68% of incomes lie within one standard deviation of it. Error: the 68-95-99.7 rule applies only to approximately normal data, and a skewed distribution does not satisfy it. Correct: describe the skew, and use the median and quartiles instead.

Real-World Connection

Quality-control engineers in factories track bolt widths, cereal box weights and battery lives with normal models. A rule such as "flag anything beyond 3 standard deviations" comes straight from the 99.7% part of the empirical rule, because values that far out happen only 0.3% of the time by chance.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

A density curve is a triangle over $0 \le x \le 1$ with $f(x) = 2(1-x)$. Find the median and the mean, and compare. Solution: the area to the right of m is a triangle with base $1-m$ and height $2(1-m)$, so area $= (1-m)^2$. Set $(1-m)^2 = 0.5$: $m = 1 - 1/\sqrt{2} \approx 0.293$. The mean of this triangle is $1/3 \approx 0.333$. Since mean $>$ median, the curve is skewed right, with its long tail toward 1.

Practice

PRACTICE

Density Curves and the Empirical Rule

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which condition must every density curve satisfy?
   (A) The highest point has height 1
   (B) The total area under the curve is 1
   (C) The curve is symmetric
   (D) The mean is 0
2. A density curve is skewed to the right. Which statement is true about its mean and median?
   (A) The mean is greater than the median
   (B) The mean is less than the median
   (C) The mean equals the median
   (D) They cannot be compared
3. Adult trout lengths are approximately normal with mean 70 cm and standard deviation 8 cm. About what percent of trout are between 62 and 78 cm?
   (A) 34% **(B)** 50% **(C)** 68% **(D)** 95%
4. SAT-style practice scores are approximately $N(500, 100)$. About what percent score above 700?
   (A) 0.15% **(B)** 16% **(C)** 5% **(D)** 2.5%
5. A runner's time is 85 seconds. For the group, mean 70 and standard deviation 6. What is the z-score?
   (A) 0.4 **(B)** 1.5 **(C)** 2.5 **(D)** 15
6. A density curve is a flat line of height 0.05 over the interval from 0 to 20. What proportion of values lie between 5 and 9?
   (A) 0.05 **(B)** 0.80 **(C)** 0.45 **(D)** 0.20
7. The weights of cans are approximately normal with mean 50 g and standard deviation 4 g. Which interval contains the middle 95% of cans?
   (A) 42 to 58 **(B)** 46 to 54 **(C)** 38 to 62 **(D)** 48 to 52
8. Test A: score 78, class mean 70, sd 4. Test B: score 90, class mean 80, sd 8. Relative to classmates, on which test is the student's performance stronger?
   (A) Test B, because 90 is higher than 78
   (B) Test A, because z = 2 versus 1.25
   (C) Test B, because the mean is higher
   (D) They are equal
9. A value has z-score -1.2. What does this mean?
   (A) The value is 1.2 below the mean
   (B) The value is 1.2 standard deviations above the mean
   (C) The value is 1.2% below the mean
   (D) The value is 1.2 standard deviations below the mean
10. A density curve is the line $f(x) = 2(1-x)$ over $0 \le x \le 1$. (a) Show the total area is 1. (b) Find the proportion of values below 0.5. (c) Find the median and compare it with the mean $1/3$.
11. Bottle volumes are approximately $N(355, 2)$ ml. (a) What percent of bottles hold between 351 and 359 ml? (b) What percent hold more than 361 ml? (c) What percent hold between 353 and 361 ml?
12. Ana swam 100 m in 62 seconds; team mean 66, sd 2.5. Ben ran 400 m in 58 seconds; team mean 62, sd 3. Lower times are better. Who had the stronger performance relative to teammates? Justify with z-scores.
13. Thickness of a metal sheet is approximately normal. The middle 68% of sheets lie between 62 mm and 74 mm. (a) Find the mean and standard deviation. (b) What percent of sheets are thinner than 56 mm?
14. Twenty-five package weights (grams) are 41, 44, 45, 46, 47, 47, 48, 49, 49, 50, 50, 50, 51, 51, 52, 52, 53, 53, 54, 55, 56, 57, 58, 60, 63. The mean is 51.2 and the standard deviation is 5.13. (a) How many values lie within one standard deviation of the mean, and how many within two? (b) Compare with the 68-95-99.7 rule and decide whether a normal model is reasonable.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

15. Explain why the mean of a left-skewed density curve is less than its median. Use the ideas of balance point and equal areas, and describe where each sits relative to the tail.

2.2 Normal Calculations: Finding Areas

Standardize any normal value, then find the area to the left, right or between two values using the z-table and the built-in calculator, and turn proportions into counts.

Why This Matters

Once you know a distribution is approximately normal, almost any question about "what percent" or "what proportion" becomes an area problem. The 68-95-99.7 rule handles only whole numbers of standard deviations. Real questions use cutoffs like 68 inches or 12.5 ounces, so you need a method that works for any value. That method is **standardize, then find the area**.

Students lose points by finding the wrong tail, by forgetting to standardize before using the table, and by writing a bare number with no context. On the AP exam, normal calculations appear in Unit 1 and are the foundation for the probability, sampling distribution and inference units that follow. Free-response questions expect you to define the variable, name the distribution, and show the setup, not just a calculator answer.

Core Ideas

Every normal distribution $N(\mu, \sigma)$ can be converted to the **standard normal** $N(0, 1)$ by the z-score. Areas under the standard normal curve are tabulated, and the built-in calculator can compute them directly.

Procedure for a normal calculation

1. **State** the variable and distribution in context: "Let X be ...; X is approximately $N(\mu, \sigma)$."
2. **Sketch** a bell curve, label the mean, mark the cutoff(s) and shade the region.
3. **Standardize** each cutoff: $z = \dfrac{x - \mu}{\sigma}$.
4. **Find the area** with the z-table or calculator.
5. **Conclude** in a sentence with context: proportion, percent, or expected count.

The **standard normal table** gives the area to the **left** of a value z. Write $P(Z < z)$. Your tables list z to two decimal places, with the whole number and first decimal down the side and the second decimal across the top.

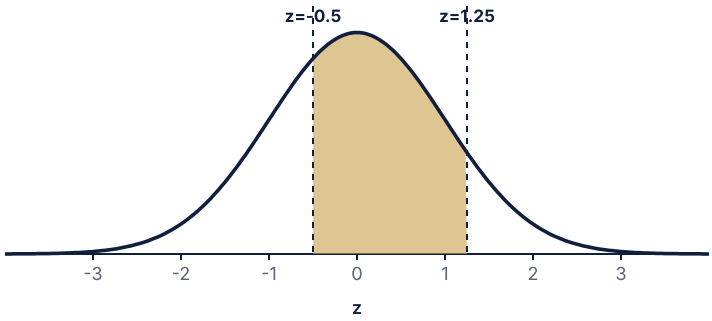
Table and calculator moves

Area to the left: $P(Z < z) =$ table value.

Area to the right: $P(Z > z) = 1 - P(Z < z)$.

Area between: $P(a < Z < b) = P(Z < b) - P(Z < a)$ (larger minus smaller).

Calculator: normalcdf(lower, upper, mean, sd). For a right tail use upper = 1E99; for a left tail use lower = -1E99. You may enter the original values and skip standardizing, but you must still show the setup on the written exam.



P(-0.5 \lt Z \lt 1.25) = 0.8944 - 0.3085 = 0.5858.

Table values are rounded, so a table answer and a calculator answer can differ in the third or fourth decimal. Both are accepted when the setup is shown. The calculator answer is the more precise one.

Reasonable-Answer Checks

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

- The area to the left of the mean is exactly 0.50. A cutoff above the mean must give a left area above 0.50.
- A tail past 2 standard deviations must be small, around 0.02 to 0.03. A tail past 3 is under 0.002.
- Areas are proportions, so they must lie between 0 and 1. A negative or greater-than-1 answer signals a subtraction in the wrong order.
- The "continuous" fact helps: $P(X \le a) = P(X < a)$, so you never worry about strict versus non-strict inequalities.

Why does the table only list left areas? Because the curve is symmetric, and the total area is 1, so any region you could want is built from left areas. A right tail is a complement. An interval between two cutoffs is the larger left area minus the smaller one. A region that is symmetric about the mean, such as $-1.5 < Z < 1.5$, can also be found as one minus two equal tails. Learning to move among these forms is more useful than memorizing any single formula.

When you move from Z back to the original variable, say what the numbers mean in the original units. A statement such as "the area to the right of $z = 1.4$ is 0.0808" is not yet an answer. The answer is "about 8.1% of the club's women are taller than 68 inches." The grader reads the conclusion line first, so make it complete and specific.

Worked Examples

READING THE STANDARD NORMAL TABLE

For the standard normal variable Z, find (a) $P(Z < 1.25)$, (b) $P(Z > 1.25)$, and (c) $P(-0.5 < Z < 1.25)$.

SOLUTION

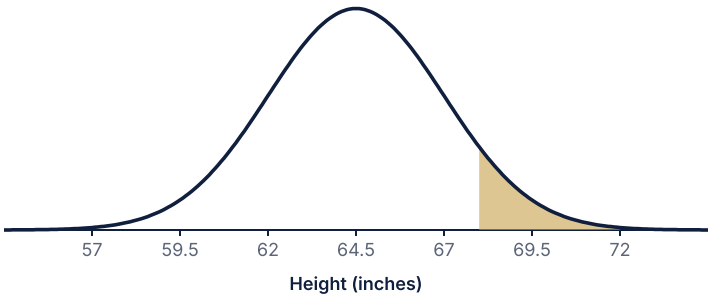
(a) Table: row 1.2, column 0.05 gives $P(Z < 1.25) = 0.8944$.

(b) Right tail is the complement: $1 - 0.8944 = 0.1056$.

(c) $P(Z < 1.25) - P(Z < -0.5) = 0.8944 - 0.3085 = 0.5859$ (the calculator gives 0.5858).

RIGHT TAIL ON A CONTEXT PROBLEM

Heights of adult women in a large club are approximately normal with mean 64.5 inches and standard deviation 2.5 inches. What proportion are taller than 68 inches?



The shaded right tail is about 0.081.

SOLUTION

Let X be height in inches; $X \sim N(64.5, 2.5)$.

Standardize: $z = (68 - 64.5)/2.5 = 1.40$.

$P(X > 68) = P(Z > 1.40) = 1 - 0.9192 = 0.0808$. Calculator: normalcdf(68, 1E99, 64.5, 2.5) = 0.0808.

About 8.1% of the club's women are taller than 68 inches.

AREA BETWEEN TWO VALUES

Scores on a placement test are approximately $N(72, 9)$. What proportion of scores are between 60 and 85?

SOLUTION

Let X be a score; $X \sim N(72, 9)$.

$z_1 = (60 - 72)/9 = -1.33$ and $z_2 = (85 - 72)/9 = 1.44$.

Table: $P(Z < 1.44) - P(Z < -1.33) = 0.9251 - 0.0918 = 0.8333$. Calculator with unrounded z: normalcdf(60, 85, 72, 9) = 0.8345.

About 83% of scores fall between 60 and 85.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

TWO TAILS AND EXPECTED COUNTS

A machine cuts bolts with diameters approximately $N(10, 0.12)$ mm. A bolt is defective if it is thinner than 9.8 mm or thicker than 10.3 mm. (a) What proportion are defective? (b) In a run of 5000 bolts, about how many are defective?

SOLUTION

Let X be the diameter, $X \sim N(10, 0.12)$.

Thin: $z = (9.8 - 10)/0.12 = -1.67$, left area 0.0475. Thick: $z = (10.3 - 10)/0.12 = 2.50$, right area $1 - 0.9938 = 0.0062$.

(a) Add the two disjoint tails: $0.0475 + 0.0062 \approx 0.054$ (calculator 0.0540). About 5.4% are defective.

(b) Expected count $= 5000(0.054) = 270$ bolts.

PERCENT INSIDE AN INTERVAL AND A COUNT

Scores for 1200 students on a reading assessment are approximately $N(500, 100)$. About how many students scored between 400 and 600? Does this agree with the empirical rule?

SOLUTION

400 and 600 are $z = -1$ and $z = 1$.

$P(-1 < Z < 1) = 0.8413 - 0.1587 = 0.6827$.

Expected count $= 1200(0.6827) \approx 819$ students.

The rule says 68%, which is 816, so the results agree up to rounding of the rule.

CHAINING A NORMAL CALCULATION

A student's bus ride to school is approximately $N(34, 6)$ minutes. The student is late if a ride is longer than 40 minutes. If the rides on two different mornings are independent, find (a) the probability of being late on a given day, (b) the probability of being late both mornings, and (c) the probability of being late at least once.

SOLUTION

(a) $z = (40 - 34)/6 = 1$ and $P(Z > 1) = 0.1587$.

(b) Independent: $0.1587^2 = 0.0252$.

(c) Use the complement of never late: $1 - (1 - 0.1587)^2 = 1 - 0.7078 = 0.2921$.

Interpretation: about 29% of two-morning stretches include at least one late ride.

Write It Like a 5

Model Response

Prompt: The weights of loaves from a bakery are approximately normal with mean 454 g and standard deviation 8 g. A loaf is labeled "underweight" if it is below 440 g. What proportion of loaves are underweight? In a day's batch of 600 loaves, how many underweight loaves would you expect? Let X be the weight of a loaf in grams, $X \sim N(454, 8)$. The cutoff has $z = (440 - 454)/8 = -1.75$. From the table, $P(Z < -1.75) = 0.0401$. So about **4.0% of loaves are underweight**. For 600 loaves, the expected count is $600(0.0401) \approx 24$, so I would expect **about 24 underweight loaves** in a typical batch.

What the grader looks for

1 point: defines the variable and states the normal model with its mean and standard deviation. 1 point: standardizes correctly, or gives the calculator setup normalcdf(-1E99, 440, 454, 8). 1 point: finds the correct left-tail area (0.0401). 1 point: states the proportion in context. 1 point: multiplies by 600 to give a sensible expected count and says it is an approximation.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Draw first. A shaded sketch prevents wrong-tail errors.
2. Decide whether you want left, right or between. Write the probability statement.
3. Use the calculator to compute, but still write the standardizing z or the normalcdf setup.
4. Check the answer against the sketch: a small shaded area should give a small number.
5. Write a full sentence with context and units.

- A shaded tail on the right needs the upper bound 1E99 on the calculator, not 99.
- Use more digits in intermediate z-values, and round only the final answer to three or four decimals.
- "Percent" means multiply the proportion by 100. "Expected number" means multiply by the group size.
- If a question gives a probability and asks for a cutoff, you are in the inverse problem, covered in the next lesson.

Common traps

1. Using a table value as a right tail. The table gives area to the left. 2. Forgetting to standardize and looking up 68 in the z-table. 3. Subtracting in the wrong order for an "between" probability and getting a negative. 4. Reporting only a number without context, or saying "normal" without naming the mean and standard deviation.

Mistake Clinic

Spot the error

A student finds $P(X > 68)$ for $N(64.5, 2.5)$ by computing $z = 1.4$ and writing the answer 0.9192. Error: 0.9192 is the area to the left of z = 1.4. The question asks for the right tail. Correct: $1 - 0.9192 = 0.0808$.

Spot the error

To find the proportion of scores between 60 and 85 in $N(72, 9)$, a student computes $0.0918 - 0.9251 = -0.8333$ and reports 83.33% with a note that the sign does not matter. Error: a proportion cannot be negative; the subtraction was in the wrong order. Correct: larger area minus smaller area, $0.9251 - 0.0918 = 0.8333$.

Spot the error

A student finds the probability that a bolt is defective for a two-sided limit by taking only the thick tail, $P(Z > 2.5) = 0.0062$, and says 0.6% are defective. Error: bolts thinner than 9.8 mm are also defective. Correct: add both tails, $0.0475 + 0.0062 \approx 0.054$.

Real-World Connection

Weather offices, public health teams and sports analysts routinely ask: what fraction of days, patients or players fall beyond a cutoff? When a variable is close to normal, the standardize-then-find-the-area method gives those answers in seconds, which is why it appears so often in later inference procedures.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

The time to assemble a toy is approximately $N(18, 3)$ minutes. Find the probability that a randomly chosen toy takes between 15 and 21 minutes, and the probability that exactly two of three independent toys take that long. Solution: 15 and 21 are $z = \pm 1$, so $p = 0.6827$. For exactly two of three, use the product of independent events: $3(0.6827)^2(1 - 0.6827) = 3(0.4661)(0.3173) \approx 0.4437$.

Practice

PRACTICE **Normal Calculations: Finding Areas**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Using the standard normal table, $P(Z < 1.25)$ is closest to:
   (A) 0.1056 **(B)** 0.8944 **(C)** 0.9500 **(D)** 1.2500
2. For Z standard normal, what is $P(Z > 1.4)$?
   (A) 0.0808 **(B)** 0.9192 **(C)** 0.8081 **(D)** 0.1400
3. Scores are approximately $N(72, 9)$. The proportion of scores between 60 and 85 is closest to:
   (A) 0.17 **(B)** 0.34 **(C)** 0.83 **(D)** 0.95
4. Heights are approximately $N(64.5, 2.5)$. Which calculator entry gives the proportion taller than 68?
   (A) normalcdf(-1E99, 68, 64.5, 2.5) **(B)** normalcdf(0, 68, 64.5, 2.5)
   (C) normalcdf(64.5, 68, 68, 2.5) **(D)** normalcdf(68, 1E99, 64.5, 2.5)
5. Heights are approximately $N(64.5, 2.5)$. The proportion shorter than 62 inches is closest to:
   (A) 0.1587 **(B)** 0.8413 **(C)** 0.0668 **(D)** 0.2500
6. Scores for 1200 students are $N(500, 100)$. About how many scored between 400 and 600?
   (A) 600 **(B)** 1200 **(C)** 1020 **(D)** 816
7. A student scores at z = 0.88. This score is at about which percentile?
   (A) 12th **(B)** 50th **(C)** 81st **(D)** 88th
8. For a standard normal variable Z, $P(-1 < Z < 0.6)$ is closest to:
   (A) 0.2743 **(B)** 0.8413 **(C)** 0.7257 **(D)** 0.5671
9. A ride time X is approximately normal. Which statement is correct for a continuous normal variable?
   (A) $P(X = 40)$ is the height of the curve at 40 **(B)** $P(X \leq 40) = P(X < 40)$
   (C) $P(X \leq 40) = 1 - P(X < 40)$ **(D)** $P(X = 40) = 0.5$
10. IQ-style scores are approximately $N(100, 15)$. (a) Find the proportion above 130. (b) Find the proportion between 85 and 120. (c) Compare (a) with the empirical rule.
11. Soup can contents are approximately $N(12, 0.3)$ oz. A can is rejected if it holds less than 11.6 oz or more than 12.5 oz. (a) Find the proportion rejected. (b) Of 8000 cans, about how many are rejected?
12. The time for a webpage to load is approximately $N(2.4, 0.5)$ seconds. (a) What proportion of loads take longer than 3.2 seconds? (b) What proportion take between 2.0 and 3.0 seconds? Show your standardizing.
13. A pump's daily output is approximately $N(300, 20)$ gallons. Find the proportion of days with output below 275 gallons. On a 365-day year, about how many such days would you expect? State the assumption you made.
14. Explain, using areas, why $P(Z > 1.2) = 1 - P(Z < 1.2)$ and why the table (area to the left) can still answer questions about right tails and intervals. Then find $P(Z > 1.2)$ and $P(0.5 < Z < 1.2)$.
15. A bus ride is approximately $N(34, 6)$ minutes. A student is late when a ride is over 40 minutes. For two independent mornings, find (a) P(late both days) and (b) P(late at least once). Interpret (b).

2.3 Inverse Normal and Finding Cutoffs

Work backward from a percent or percentile to a cutoff value using invNorm and the z-table, find quartiles and middle intervals, and solve for unknown means and standard deviations.

Why This Matters

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Last lesson you were handed a cutoff and asked for the area. Now the roles swap: you are handed an area (a percent, a percentile, a probability) and asked for the cutoff. Scholarship committees pick the top 10% of scores, factories set limits so that only 2% of items fail, and doctors flag the lowest 5% of growth measurements. All of these are **inverse normal** problems.

The most common errors are looking up the wrong area in the table, using the right-tail area where the table needs the left-tail area, and forgetting to convert the z-value back to the original units. This skill appears in Unit 1 of the exam for percentiles and the normal model, and the same "find a critical z-value" move reappears when you build confidence intervals in later units.

Core Ideas

A **percentile** is the percent of observations that fall at or below a value. The 90th percentile is the value with 90% of the distribution to its left. Because the z-table gives left-tail areas, percentiles are the natural bridge from an area to a z-score.

Inverse normal procedure

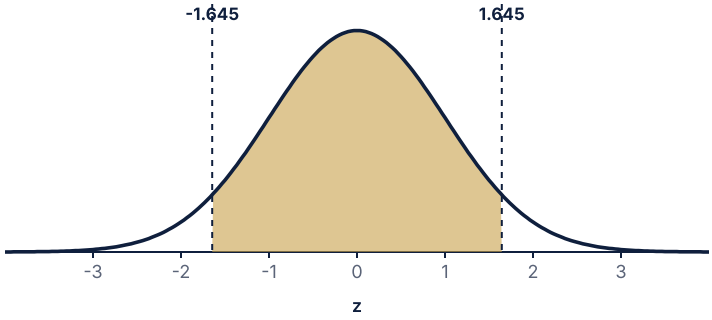
1. **Draw** the curve and shade the area you know. Label it with its value.
2. **Convert** the area to a LEFT-tail area: top 10% means left area 0.90; middle 90% means 5% in each tail, so left areas 0.05 and 0.95.
3. **Find z** in the interior of the table closest to that left area, or use invNorm(area) on the calculator.
4. **Unstandardize:** $x = \mu + z\sigma$.
5. **Conclude** in context with units.

Calculator syntax

invNorm(area to the left, mean, standard deviation) returns the cutoff x directly.

invNorm(area) with no mean or sd returns the z-score for the standard normal.

The area you enter is always the LEFT area. Top 15% of $N(70, 6)$ is invNorm(0.85, 70, 6).



The middle 90% of a normal distribution lies within 1.645 standard deviations of the mean.

Some z-values are worth knowing. They are the critical values you will meet again in confidence intervals.

Area to the left	z-score	Meaning
0.50	0.000	median
0.90	1.282	90th percentile
0.95	1.645	95th percentile, edge of middle 90%
0.975	1.960	edge of middle 95%
0.995	2.576	edge of middle 99%

Percentiles and the Quartiles of a Normal Model

The quartiles are the 25th and 75th percentiles. For any normal distribution $Q_1 = \mu - 0.6745\sigma$ and $Q_3 = \mu + 0.6745\sigma$, so the IQR of a normal distribution is about 1.35σ. This links the normal model to the boxplot ideas from earlier: if a data set's IQR is much larger or smaller than $1.35s$, the data are probably not normal.

Two warnings. First, the percentile of a value is a statement about position, not a score: the 80th percentile means 80% of the group scored at or below, not that the person got 80 points. Second, the table lists z-values in the interior; if the exact area you want is not there, take the nearest one and say so (for 0.90, 0.8997 is closest and gives $z = 1.28$). The calculator avoids this and gives 1.2816.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A useful mental model is a two-way street. Forward travel: cutoff, then z-score, then area. Backward travel: area, then z-score, then cutoff. The z-score is the middle stop in both directions, so if you are ever lost, ask yourself whether you are holding a cutoff or an area, and whether you need to find the other one. Another useful habit is a reasonableness check: a cutoff for the top 10% must be above the mean, and for the bottom 5% must be below it. If your answer falls on the wrong side of the mean, the tail was misread.

Inverse normal problems also let you test whether data look normal. If a data set has a 25th percentile and 75th percentile, you can compare them with $\mu \pm 0.6745\sigma$. A large mismatch is evidence the distribution is skewed or has unusual tails, which is the idea behind the next lesson.

Worked Examples

FROM A PERCENTILE TO A Z-SCORE AND BACK

(a) Find the z-score for the 90th percentile of a standard normal distribution. (b) Scores are approximately $N(70, 6)$. A student scores 75. What percentile is that?

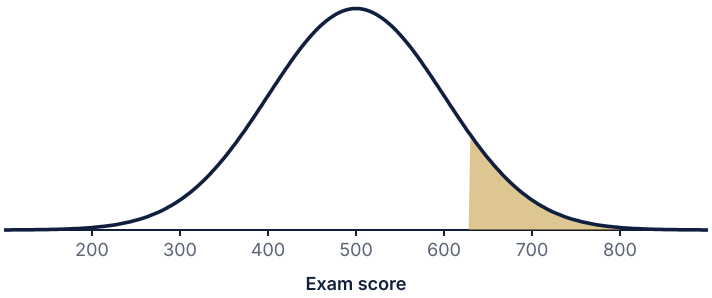
SOLUTION

(a) Find the entry in the table nearest 0.9000: it is 0.8997, in row 1.2, column 0.08. So $z \approx 1.28$. Calculator: invNorm(0.90) = 1.2816.

(b) $z = (75 - 70)/6 = 0.83$. Table: $P(Z < 0.83) = 0.7967$ (calculator 0.7977). The score is at about the 80th percentile, meaning about 80% of scores are at or below 75.

CUTOFF FOR THE TOP 10%

Scores on a placement exam are approximately $N(500, 100)$. A scholarship goes to the top 10% of scores. What is the minimum score needed?



The top 10% begins near 628.

SOLUTION

Let X be an exam score, $X \sim N(500, 100)$. The top 10% has 90% to its left.

$z = 1.28$ (invNorm(0.90) = 1.2816).

$x = 500 + 1.2816(100) = 628.2$. Calculator: invNorm(0.90, 500, 100) = 628.2.

A score of about 628 or higher puts a student in the top 10%.

CUTOFF FOR THE BOTTOM 5%

The lifetime of a phone battery is approximately $N(400, 25)$ minutes. A company wants to replace the shortest-lasting 5% of batteries under warranty. What lifetime marks the warranty limit?

SOLUTION

Left area is 0.05, so $z = -1.645$.

$x = 400 + (-1.645)(25) = 358.9$ minutes.

Calculator: invNorm(0.05, 400, 25) = 358.9.

Batteries that last less than about 359 minutes are replaced.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A SYMMETRIC MIDDLE INTERVAL

Adult trout lengths are approximately $N(68, 4)$ cm. Find the lengths that bound the middle 90% of trout.

SOLUTION

The middle 90% leaves 10% in the tails, 5% in each, so the left areas are 0.05 and 0.95.

$z = \pm 1.645$.

Lower $= 68 - 1.645(4) = 61.42$. Upper $= 68 + 1.645(4) = 74.58$.

The middle 90% of trout are between about 61.4 cm and 74.6 cm.

QUARTILES AND THE OUTLIER FENCES

Daily step counts (in thousands) are approximately $N(50, 10)$. (a) Find Q_1, Q_3 and the IQR. (b) Use the 1.5 IQR rule to find the outlier fences, and find the proportion of a normal distribution that falls outside them.

SOLUTION

(a) $z = \pm 0.6745$. $Q_1 = 50 - 0.6745(10) = 43.26$ and $Q_3 = 56.74$. IQR $= 13.49$.

(b) Lower fence $= 43.26 - 1.5(13.49) = 23.02$; upper fence $= 56.74 + 20.23 = 76.97$. Those are $z = \pm 2.70$.

$P(|Z| > 2.70) = 2(1 - 0.9965) = 0.007$. So the 1.5 IQR rule flags only about 0.7% of values in a perfectly normal distribution, as expected for a rule designed to catch unusual ones.

FINDING AN UNKNOWN MEAN AND STANDARD DEVIATION

Reaction times are approximately normal. 20% of times are below 40 ms and 10% are above 70 ms. Find μ and σ.

SOLUTION

20th percentile: $z = -0.8416$. 90th percentile: $z = 1.2816$.

Two equations: $40 = \mu - 0.8416\sigma$ and $70 = \mu + 1.2816\sigma$.

Subtract: $30 = 2.1232\sigma$, so $\sigma = 14.13$ ms.

Then $\mu = 40 + 0.8416(14.13) = 51.89$ ms. Check: $51.89 + 1.2816(14.13) = 70.0$.

Write It Like a 5

Model Response

Prompt: A bakery's loaf weights are approximately normal with mean 454 g and standard deviation 8 g. The bakery wants to label as "premium" the heaviest 15% of loaves. What is the minimum weight of a premium loaf? Let X be loaf weight in grams, $X \sim N(454, 8)$. The heaviest 15% means 85% of loaves weigh less, so I need the 85th percentile. The z-score with left area 0.85 is $z = 1.036$ (invNorm(0.85)). Then $x = 454 + 1.036(8) = 462.3$ g. Check by calculator: invNorm(0.85, 454, 8) = 462.29. **A loaf must weigh at least about 462.3 grams to be premium.**

What the grader looks for

1 point: defines the variable and the normal model with parameters. 1 point: converts "top 15%" to a left area of 0.85 (not 0.15). 1 point: finds the correct z-value. 1 point: unstandardizes with $x = \mu + z\sigma$ or gives invNorm with all three arguments. 1 point: states the conclusion with units and context.

Coach's Playbook

1. Shade the known area on a sketch and note whether it is in the left tail, right tail or middle.
2. Rewrite the known area as a left-tail area.
3. Get z from invNorm(area) or from the interior of the table.
4. Compute $x = \mu + z\sigma$. A cutoff above the mean must use a positive z.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. Sanity-check by plugging x back into normalcdf: you should recover the original area.

- Always say what the cutoff means: "scores of 628 or higher" or "lifetimes below 359 minutes."
- For a middle area, split the leftover equally between the two tails.
- The z for the 50th percentile is 0; for percentiles below 50 the z is negative.
- Round only at the end; the z-values 1.28 versus 1.2816 can move a cutoff by a full point.

Common traps

1. Entering the right-tail area into invNorm. It wants the left area. 2. Looking up the area in the table's margin instead of its interior. 3. Forgetting to convert z back: reporting 1.28 as the cutoff score. 4. For a middle 90%, using 0.90 instead of 0.95 as the upper left area.

Mistake Clinic

Spot the error

A student finds the cutoff for the top 10% of $N(500, 100)$ as invNorm(0.10, 500, 100) = 371.8. Error: 0.10 is the right-tail area, but invNorm uses the left area. A cutoff below the mean cannot be the top 10%. Correct: invNorm(0.90, 500, 100) = 628.2.

Spot the error

To find the middle 90% for $N(68, 4)$, a student uses $z = \pm 1.28$ (from 0.90). Error: 1.28 leaves only 10% in the right tail, so the interval is the middle 80%. Correct: the left areas are 0.05 and 0.95, giving $z = \pm 1.645$ and the interval 61.4 to 74.6.

Spot the error

A student writes: "Top 15% of $N(70, 6)$: $z = 1.036$, so the cutoff is 1.036." Error: z is not in the original units. Correct: $x = 70 + 1.036(6) = 76.2$.

Real-World Connection

Growth charts, credit risk and college admissions use percentiles. A child at the 25th percentile for height is shorter than 75% of children the same age. If heights are approximately normal, the 25th percentile is $\mu - 0.6745\sigma$, found with the same inverse normal steps as this lesson.

Challenge Corner

A pump should be set to a mean μ so that only 2% of 12-ounce bottles hold less than 12 ounces, when the fill sd is 0.3 oz. Find μ. Solution: the 2nd percentile has $z = -2.054$, so $12 = \mu - 2.054(0.3)$ and $\mu = 12.616$ oz. Setting the mean above the label ensures only 2% of bottles are under-filled.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE **Inverse Normal and Finding Cutoffs**

Show your work. A calculator is allowed. Justify conclusions in context.

1. What is the z-score of the 95th percentile of a standard normal distribution?
   (A) 0.95 **(B)** 1.28 **(C)** 1.645 **(D)** 1.96
2. Scores are approximately $N(500, 100)$. What is the minimum score for the top 10%?
   (A) 372 **(B)** 628 **(C)** 672 **(D)** 872
3. Battery lives are $N(400, 25)$ minutes. The shortest-lasting 5% of batteries last less than about:
   (A) 359 **(B)** 350 **(C)** 375 **(D)** 441
4. Which calculator entry gives the cutoff for the top 15% of a $N(70, 6)$ distribution?
   (A) invNorm(0.15, 70, 6) **(B)** invNorm(0.85) **(C)** normalcdf(0.85, 70, 6)
   (D) invNorm(0.85, 70, 6)
5. Trout lengths are $N(68, 4)$. The middle 90% of lengths lie between about:
   (A) 60.0 and 76.0 **(B)** 61.4 and 74.6 **(C)** 62.7 and 73.3
   (D) 64.0 and 72.0
6. Scores are $N(70, 6)$. A score of 75 is at about which percentile?
   (A) 20th **(B)** 75th **(C)** 80th **(D)** 83rd
7. For a $N(50, 10)$ distribution, the third quartile is closest to:
   (A) 54.5 **(B)** 66.4 **(C)** 60.0 **(D)** 56.7
8. Can contents are $N(12, 0.3)$ oz. A can in the lightest 3% holds less than about:
   (A) 11.4 **(B)** 11.6 **(C)** 11.9 **(D)** 12.6
9. Daily output is $N(300, 20)$. The 40th percentile is closest to:
   (A) 292.0 **(B)** 308.0 **(C)** 305.1 **(D)** 294.9
10. SAT-style scores are $N(500, 100)$. (a) Find the 75th percentile. (b) Find the minimum score needed to be in the top 5%. Show the z-scores.
11. The heights of plants are approximately $N(36, 3)$ cm. Find the heights that bound the middle 99% of plants.
12. A machine has fill standard deviation 0.3 oz. The company wants only 2% of 12-oz bottles to be under-filled (less than 12 oz). What mean setting should be used? Assume fills are normal.
13. A test has scores that are approximately normal. The 30th percentile is 62 and the 90th percentile is 80. Find the mean and standard deviation.
14. Scores are $N(85, 12)$. (a) Find Q_1, Q_3 and the IQR. (b) Find the 1.5 IQR fences, and the proportion of a normal distribution outside them.
15. Explain the difference between finding an area from a cutoff and finding a cutoff from an area. Then find the 90th percentile of a $N(100, 15)$ distribution and verify it by computing the area to its left.

2.4 Assessing Normality and Using Normal Models

Decide whether a normal model fits by combining shape, the 68-95-99.7 check and normal probability plots, and learn when a normal model is reasonable or risky.

Why This Matters

Every normal calculation in this unit rests on one assumption: that the data really are approximately normal. If that assumption is false, the percents you compute can be badly wrong. A model that predicts negative waiting times or exam scores above the maximum is a warning that the normal curve does not fit. So before you standardize anything, you must be able to look at data and judge whether a normal model is reasonable.

Students often say "the data are normal" after one glance at a graph, or reject normality because a histogram is slightly lumpy. This lesson gives you a toolkit: shape from a dotplot or histogram, a check against the 68-95-99.7 rule, and the normal probability plot. These skills belong to Unit 1 of the exam, and you will use them again in inference, where conditions such as "approximately normal" must be justified with a graph.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Core Ideas

No real data set is exactly normal. The question is whether the normal curve is a **good enough** model for the purpose. We combine three kinds of evidence, and the more of them agree, the more confident we can be.

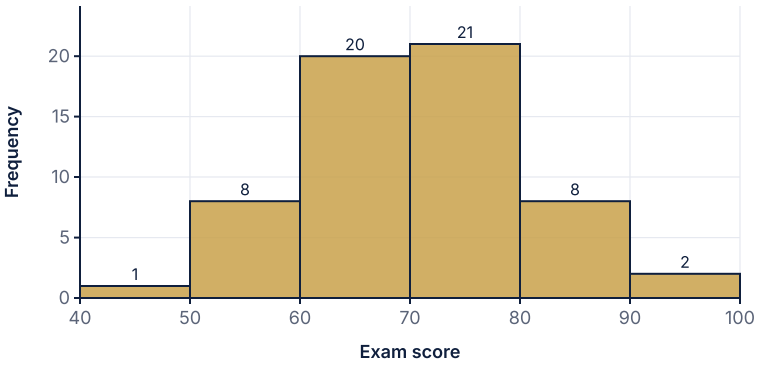
Three ways to assess normality

Shape. A histogram, dotplot or stemplot should be roughly symmetric, single-peaked and bell-like, with no strong skew, gaps or extreme outliers.

Rule check. About 68% of values should lie within 1 sample standard deviation of $\bar{x}$, about 95% within 2, and nearly all within 3.

Normal probability plot. Plot each data value against its expected z-score. Roughly a straight line means roughly normal.

A **normal probability plot** (also called a normal quantile plot) sorts the data, finds the z-score each ordered value would have in a perfect normal sample, and plots the pairs. When the data are normal, the points fall close to a straight line. Systematic bends, curves or steps are evidence of non-normality. The calculator and software produce these plots automatically.



60 exam scores: roughly symmetric and single-peaked.

Reading a normal probability plot

Points near a straight line: approximately normal.

Points curve upward on the right (convex, a smile shape): right-skewed. The largest values are bigger than a normal sample would give.

Points curve the other way: left-skewed. A clear gap or plateau suggests clusters or rounded data.

The slope of the line estimates $1/s$, and the line crosses z = 0 at about $\bar{x}$.

When Is a Normal Model Reasonable?

- The variable is quantitative and takes many values (not a handful of whole numbers).
- The distribution is not strongly skewed and has no extreme outliers. A single mild outlier does not ruin a normal model; a pile of values at a boundary does.
- The model does not predict impossible values with noticeable probability. If $\bar{x} - 3s$ is below zero for a variable that cannot be negative, be suspicious.
- Small samples (say under 15) are hard to judge. The graph can look irregular even when the population is normal, so use judgment and be modest in your claims.

It is acceptable on the exam to say "approximately normal" when the graph is mound-shaped and roughly symmetric. Always cite evidence from the graph or counts rather than saying "looks normal".

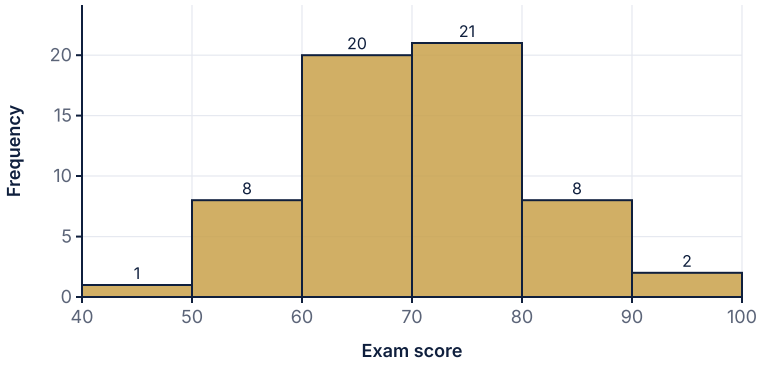
Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

HISTOGRAM SHAPE AND THE RULE

The histogram shows 60 exam scores with mean 70.0 and standard deviation 10.0. Is a normal model reasonable? Check the 68-95-99.7 rule.



60 exam scores.

SOLUTION

Shape: single peak in the 60s and 70s, roughly symmetric, no outliers or gaps.

Within 1 sd (60 to 80): $20 + 21 = 41$ of 60, or 68.3% (rule: 68%).

Within 2 sd (50 to 90): $8 + 20 + 21 + 8 = 57$ of 60, or 95% (rule: 95%).

Both the shape and the counts match the rule, so a normal model is reasonable for these scores.

CHECKING THE RULE ON A SMALL SAMPLE

Twenty runners' times (minutes) are 36, 38, 41, 42, 44, 45, 47, 47, 49, 49, 51, 51, 53, 53, 55, 55, 58, 59, 61, 65. The mean is 49.95 and the standard deviation is 7.70. How many are within 1 and 2 standard deviations of the mean?

SOLUTION

1 sd: 42.25 to 57.65. The values 44 through 55 (twelve values) fall inside: $12/20 = 60\%$.

2 sd: 34.55 to 65.35. All 20 values fall inside: 100%.

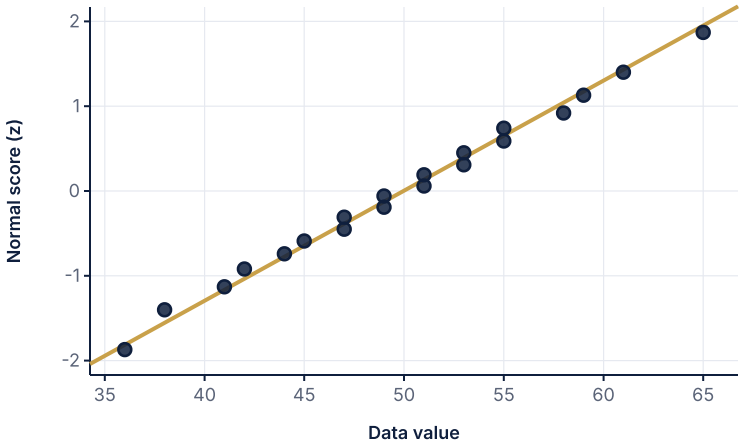
With only 20 values, 60% versus 68% and 100% versus 95% are within normal sampling variation, so the counts do not contradict normality.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A NORMAL PROBABILITY PLOT THAT IS LINEAR

The normal probability plot shows the same 20 running times. Interpret it.



Normal probability plot, data set A: the points hug the line.

SOLUTION

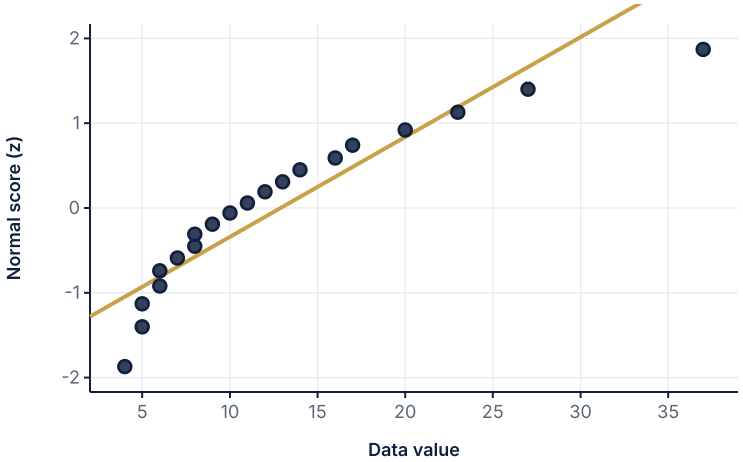
The points fall very close to the straight line across the whole range, with no systematic bend.

So the normal probability plot supports an approximately normal distribution for the running times.

Together with the symmetric dotplot and the rule check, the normal model is reasonable.

A NORMAL PROBABILITY PLOT THAT IS CURVED

Twenty waiting times (minutes) are 4, 5, 5, 6, 6, 7, 8, 8, 9, 10, 11, 12, 13, 14, 16, 17, 20, 23, 27, 37, with mean 12.9 and standard deviation 8.49. The plot shows the normal probability plot. Is a normal model reasonable?



Normal probability plot, data set B: the points bend away from the line.

SOLUTION

The points form a curve that bends upward on the right instead of following the line.

The largest values (27 and 37) lie far above where a normal sample would put them, which is the signature of right skew.

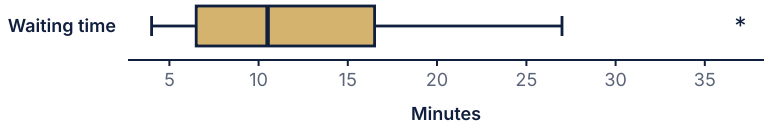
Mean 12.9 is above the median 10.5, which also suggests right skew. A normal model is not reasonable.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

BOXPLOT AND IMPOSSIBLE PREDICTIONS

For the waiting times above, the five-number summary is 4, 6.5, 10.5, 16.5, 37 (37 is an outlier by the 1.5 IQR rule). (a) Describe the shape. (b) If a normal model with mean 12.9 and sd 8.49 were used, what proportion of waiting times would it predict to be negative?



Right-skewed, with one high outlier.

SOLUTION

(a) The median is closer to Q_1 than Q_3, the upper whisker and outlier are long, so the distribution is skewed right.

(b) $z = (0 - 12.9)/8.49 = -1.52$ and $P(Z < -1.52) = 0.064$. The model predicts about 6.4% negative waiting times, which are impossible.

Conclusion: the normal model is not appropriate for these data. Use the median and IQR to summarize them.

CHOOSING AMONG VARIABLES

Which of these is most likely to be approximately normal for a large sample of adults? (i) heights, (ii) household incomes, (iii) number of siblings. Justify each decision.

SOLUTION

(i) Heights are a continuous measurement, tend to be symmetric and single-peaked, and are the standard example of a roughly normal variable.

(ii) Incomes are strongly right-skewed with a few very high values, so they are not normal.

(iii) The number of siblings is discrete with few values (0, 1, 2, ...), with most values small, so a bell curve fits poorly.

Only heights are reasonably modeled as normal.

Write It Like a 5

Model Response

Prompt: A technician records the time (in seconds) a device takes to boot on 20 trials. The normal probability plot shows points that curve upward away from the line at the right end, and the histogram has a long right tail. The mean is 12.9 s and the median is 10.5 s. Should the technician use a normal model to find the percent of boots longer than 25 s? Explain. No. The histogram has a long right tail, so the distribution is skewed right, not symmetric. The mean (12.9) is greater than the median (10.5), which also indicates right skew. In the normal probability plot, the points curve upward away from the line at the high end, which means the largest boot times are bigger than a normal model would predict. Because all three pieces of evidence show right skew, a normal model is **not appropriate**, and a normal calculation would give unreliable percents in the right tail. The technician should use the observed data or a skew-friendly model.

What the grader looks for

1 point: reaches the correct conclusion (normal model not appropriate). 1 point: cites evidence from the histogram (skew). 1 point: cites mean versus median or another numeric clue. 1 point: correctly interprets the curved normal probability plot, including direction. 1 point: mentions the consequence for the tail calculation, in context.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Graph the data first: histogram, dotplot or stemplot.
2. Describe shape, center, spread and unusual values.
3. Compare mean and median, and check the percent within 1 and 2 sd.
4. Make a normal probability plot and look for a straight line.
5. State a conclusion and name the evidence you used.

- Use the calculator's normal probability plot: enter the data in a list and select the last plot type.
- A straight line with a little wiggle is fine; a clear curve is not.
- With small samples, say "there is no strong evidence against normality" rather than "the data are normal."
- If a variable has a natural minimum or maximum near the mean, the normal model usually fails.

Common traps

1. Calling data normal just because the histogram has one peak, while ignoring a long tail. 2. Calling data non-normal because a small-sample histogram is a little lumpy. 3. Reading the normal probability plot backward: a curve up on the right is right skew, not left skew. 4. Saying "the mean and median are equal, so it must be normal." Equality is consistent with symmetry but does not prove a bell shape.

Mistake Clinic

Spot the error

A student sees 12 of 20 values within one sd (60%) and says the data are definitely not normal because 60% is not 68%. Error: with 20 values, sampling variation is large. Correct: 60% is within the normal range of variation for small samples, so the rule check does not contradict normality.

Spot the error

A student sees points curving upward on the right of a normal probability plot and writes, "left-skewed because the plot goes up." Error: the largest values are larger than normal, so the tail is on the right. Correct: right-skewed.

Spot the error

A student says, "The data are normal because the mean equals the median." Error: symmetry is only one feature. A uniform or bimodal symmetric distribution also has mean equal to median. Correct: check for a mound shape, the rule counts and a linear normal probability plot.

Real-World Connection

Before data scientists use any method that assumes normality (including many of the inference procedures later in this course), they check the data with histograms and probability plots. In manufacturing, a normal probability plot of measured part sizes is a routine quality-control picture.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

A sample of 40 values has mean 20 and standard deviation 10, and the minimum is 1. Explain whether normal is plausible, and compute how many values below 1 a normal model with those parameters would predict. Solution: $z = (1 - 20)/10 = -1.9$, and $P(Z < -1.9) = 0.029$, so about $40(0.029) \approx 1.1$ values below 1. We observe none below 1 and exactly one at 1, which is consistent with normal. The shape graph must still be checked for symmetry before concluding.

Practice

PRACTICE

Assessing Normality and Using Normal Models

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which graph gives the most direct evidence about whether data are approximately normal?
   (A) Scatterplot of the data versus time
   (B) Normal probability plot
   (C) Bar chart of frequencies
   (D) Pie chart
2. A normal probability plot has points that curve upward away from the line at the right end. The distribution is most likely:
   (A) Left-skewed
   (B) Bimodal
   (C) Symmetric
   (D) Right-skewed
3. A histogram has two clear peaks separated by a valley. A normal model is:
   (A) Reasonable, because it is symmetric
   (B) Reasonable, because the mean is in the middle
   (C) Not reasonable, because it is not single-peaked
   (D) Reasonable only if the sample is large
4. A sample has 90% of values within 1 standard deviation of the mean. Compared with the 68-95-99.7 rule, this suggests:
   (A) The data are more concentrated than normal
   (B) Perfect normality
   (C) The data are more spread out than normal
   (D) The standard deviation is too small
5. Which variable is LEAST likely to be approximately normal?
   (A) Adult heights in a large city
   (B) Scores on a long exam
   (C) Weights of 12-ounce cereal boxes
   (D) Annual incomes of 500 families
6. A boxplot has the median close to Q1, a long upper whisker and a high outlier. A normal model is:
   (A) Reasonable
   (B) Not reasonable, because the distribution is skewed right
   (C) Reasonable because the whiskers are long
   (D) Not reasonable, because it is skewed left
7. Which statement about normal models is true?
   (A) A normal model is an approximation, judged by graphs and counts
   (B) If the mean equals the median, the data are normal
   (C) Real data sets can be exactly normal
   (D) Normality can be checked only with a histogram
8. The percent of students scoring 100 on a quiz is 30%, and the mean is 88. A normal model is poor mainly because:
   (A) The mean is too high
   (B) Quizzes have too few questions
   (C) There is a pile of values at the upper limit, causing skew
   (D) The standard deviation is zero
9. For a data set with $\bar{x} = 12$ and $s = 6$ that cannot be negative, a normal model would put what proportion below zero?
   (A) About 2.3%
   (B) About 0.1%
   (C) About 5%
   (D) About 16%

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. A histogram of 80 plant heights shows counts of 2, 10, 24, 28, 12, 4 in six equal bins; the mean is about 51 and the sd is about 11. The bins are 20-30, 30-40, 40-50, 50-60, 60-70, 70-80. Check the 68-95-99.7 rule and decide whether a normal model is reasonable.
11. Describe how to read a normal probability plot. What does each of these patterns mean? (a) points close to a line, (b) points curving upward at the right, (c) points forming steps.
12. Waiting times have mean 8.1 min, median 5.0 min, sd 7.0 min, and a minimum of 0.5. (a) Explain why a normal model is a poor fit. (b) What proportion would a normal model predict to be below 0?
13. Sixteen student heights (inches) are 58, 60, 61, 62, 62, 63, 64, 64, 65, 65, 66, 67, 68, 69, 70, 72, with mean 64.8 and sd 3.8. (a) What percent are within 1 sd? within 2 sd? (b) Is a normal model reasonable?
14. For each variable, state whether a normal model is plausible and why: (a) lengths of 300 pencils from one machine, (b) the number of pets owned by students, (c) time to complete a 5 km race.
15. A quality engineer finds that the normal probability plot of bolt diameters is straight, except that the lowest two points sit far below the line. What does this suggest, and how should it affect the analysis?

T Unit 2 Test

Test your skills with density curves, normal calculations, inverse normal, percentiles and assessing normality.

Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Part A: Multiple Choice

UNIT TEST **Unit 2 Test**

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. Which of these could be a density curve?
   (A) A flat line of height 0.5 over the interval 0 to 4
   (B) A flat line of height 0.25 over the interval 0 to 4
   (C) A flat line of height 1 over the interval 0 to 4
   (D) A curve whose height is 0.25 at its peak and negative in the tails
2. A density curve is skewed to the left. Which is true?
   (A) Mean $>$ median **(B)** Mean = median **(C)** Mean $<$ median
   (D) The mean cannot be found
3. Battery lives are approximately $N(400, 25)$ minutes. About what percent of batteries last between 350 and 425 minutes?
   (A) 68% **(B)** 97.5% **(C)** 95% **(D)** 81.5%
4. A value of 55 comes from a distribution with mean 48 and standard deviation 5. Its z-score is:
   (A) 1.4 **(B)** 7 **(C)** 11.0 **(D)** -1.4
5. Lengths are approximately $N(65, 3)$. The proportion greater than 70 is closest to:
   (A) 0.095 **(B)** 0.048 **(C)** 0.952 **(D)** 0.167
6. Scores are approximately $N(100, 15)$. The proportion between 90 and 115 is closest to:
   (A) 0.25 **(B)** 0.34 **(C)** 0.59 **(D)** 0.84
7. Scores are $N(70, 10)$. What score marks the top 20%?
   (A) 78.4 **(B)** 61.6 **(C)** 81.6 **(D)** 91.6
8. Scores are $N(70, 10)$. The middle 90% of scores lie between:
   (A) 57.2 and 82.8 **(B)** 60.0 and 80.0 **(C)** 50.4 and 89.6
   (D) 53.6 and 86.4
9. Heights are $N(48, 5)$. A height of 56 is at about which percentile?
   (A) 16th **(B)** 80th **(C)** 95th **(D)** 99th
10. A normal probability plot of 40 values shows points that curve upward on the right, away from the line. The best description is:
    (A) Skewed to the right **(B)** Approximately normal **(C)** Skewed to the left
    (D) Uniform

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

11. A teacher adds 5 points to every score. The original scores had mean 70 and standard deviation 8. The new scores have:
(A) Mean 75, sd 13 **(B)** Mean 70, sd 13 **(C)** Mean 75, sd 8
(D) Mean 70, sd 8

12. Heights are $N(48, 5)$. Which calculator entry gives $P(42 < X < 55)$?
(A) normalcdf(42, 55, 0, 1) **(B)** normalcdf(48, 55, 42, 5) **(C)** invNorm(42, 55, 48, 5)
(D) normalcdf(42, 55, 48, 5)

13. [6 points] A tube-cutting machine produces lengths approximately $N(10, 0.8)$ cm. A tube is defective if it is shorter than 9 cm or longer than 11.5 cm. (a) [3 points] Find the proportion of defective tubes. (b) [1 point] In a run of 2000 tubes, about how many are defective? (c) [2 points] The company labels the longest 10% of tubes "extra long." Find the minimum length of an extra-long tube.

14. [10 points] A writing test has scores that are approximately $N(70, 12)$ in one large class. (a) [2 points] Find the percentile of a score of 81. (b) [2 points] Find the proportion of scores between 60 and 75. (c) [2 points] A prize goes to the top 15% of scores. Find the minimum prize score. (d) [2 points] In a second class, the times (minutes) 20 students took to finish were 4, 5, 5, 6, 6, 7, 8, 8, 9, 10, 11, 12, 13, 14, 16, 17, 20, 23, 27, 37, with mean 12.9 and standard deviation 8.49. Compute the percent within 1 and 2 standard deviations of the mean and compare with the rule. (e) [2 points] The normal probability plot of these times curves upward on the right. Is a normal model appropriate? What does a normal model predict about the proportion of negative times?

Answers

Density Curves and Normal Distributions: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

2.1 Density Curves and the Empirical Rule

PRACTICE · DENSITY CURVES AND THE EMPIRICAL RULE

1. B. Every density curve lies on or above the axis and encloses area exactly 1. Height 1 (A) is not required, since narrow curves can be taller; symmetry (C) and mean 0 (D) apply only to special cases.
2. A. The long right tail pulls the balance point (the mean) toward the tail, so the mean exceeds the median. Choosing B reverses the order, which fits left skew.
3. C. 62 and 78 are $\mu \pm 1\sigma$, so about 68% lie between. 34% (A) is only one half of that interval, and 95% (D) corresponds to two standard deviations.
4. D. 700 is $(700 - 500)/100 = 2$ standard deviations above the mean. 5% lie outside $\mu \pm 2\sigma$, so 2.5% are above 700. Choosing 5% (C) forgets to split into two tails; 0.15% (A) is the 3 sd value.
5. C. $z = (85 - 70)/6 = 2.5$. Dividing by the mean instead of subtracting it, or reporting 15 (the raw difference), omits the standardizing step.
6. D. The area is width times height: $(9 - 5)(0.05) = 0.20$. Choice A (0.05) is the height, which is not a probability; choice C (0.45) uses the end positions incorrectly.
7. A. 95% lies within 2 sd: $50 \pm 2(4)$ gives 42 to 58. 46 to 54 is only one sd (68%), and 38 to 62 is three sd (99.7%).
8. B. Test A: $z = (78 - 70)/4 = 2$. Test B: $z = (90 - 80)/8 = 1.25$. The higher raw score on B is misleading because B has a larger spread; relative to classmates, A is stronger.
9. D. A z-score counts standard deviations from the mean; the negative sign means below. Choice A drops the unit of standard deviations; choice B has the wrong direction.
10. (a) The graph is a triangle with base 1 and height $f(0) = 2$: area $\frac{1}{2}(1)(2) = 1$. It is above the axis, so it is valid.
(b) Area to the right of 0.5 is a triangle with base 0.5 and height $f(0.5) = 1$: $\frac{1}{2}(0.5)(1) = 0.25$. So the proportion below 0.5 is $1 - 0.25 = 0.75$.
(c) Area right of m is $(1 - m)^2 = 0.5$, so $m = 1 - 1/\sqrt{2} \approx 0.293$. The mean $1/3 \approx 0.333$ exceeds the median, which fits a right-skewed curve with its tail toward 1.
11. Let $X \sim N(355, 2)$.
(a) 351 and 359 are $\mu \pm 2\sigma$, so about 95% of bottles.
(b) 361 is $\mu + 3\sigma$, so about 0.15% hold more than 361 ml.
(c) 353 is $\mu - 1\sigma$ and 361 is $\mu + 3\sigma$: $34\% + 49.85\% = 83.85\%$, about 84%.
12. Ana: $z = (62 - 66)/2.5 = -1.6$. Ben: $z = (58 - 62)/3 = -1.33$. Lower times are better, so more negative is stronger. Ana is 1.6 standard deviations faster than her team mean; Ben is 1.33 faster. Ana's performance is stronger relative to her teammates.
13. (a) The mean is the midpoint of the middle 68%: $\mu = (62 + 74)/2 = 68$ mm. The interval is $\mu \pm \sigma$, so $2\sigma = 12$ and $\sigma = 6$ mm.
(b) 56 is $\mu - 2\sigma$ (68 - 12). 5% of sheets are outside $\mu \pm 2\sigma$, so about 2.5% are thinner than 56 mm.
14. (a) Within one sd (46.11 to 56.37) there are 17 values; within two sd (40.99 to 61.49) there are 24 values.
(b) 17/25 = 68% within 1 sd (rule: 68%) and 24/25 = 96% within 2 sd (rule: 95%). These match the rule closely, so a normal model is reasonable. A graph such as a histogram should also be checked for symmetry and a single peak.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

15. The median is the equal-areas point: half the area is on each side. The mean is the balance point, which is sensitive to how far the area extends. In a left-skewed curve the long tail is on the left, so the area far to the left exerts a large lever arm and pulls the balance point toward the tail. The median is resistant to that stretch and stays closer to the main bulk of the data. So the mean sits to the left of the median, in the direction of the tail.

2.2 Normal Calculations: Finding Areas

PRACTICE · NORMAL CALCULATIONS: FINDING AREAS

1. B. The table gives area to the left of z = 1.25, which is 0.8944. 0.1056 is the right tail, found by the complement; 1.25 is the z-value itself, not a probability.
2. A. The table value 0.9192 is the left area; the right tail is $1 - 0.9192 = 0.0808$. Choosing B forgets to take the complement.
3. C. $z = -1.33$ and $z = 1.44$; $0.9251 - 0.0918 = 0.8333$, about 0.83. Choice D would be the 2 sd interval; B would be only half of 1 sd.
4. D. A right tail needs lower bound 68 and upper bound 1E99 (a huge number). Choice A gives the left tail, so it is the complement of what is wanted.
5. A. $z = (62 - 64.5)/2.5 = -1$, so the left area is 0.1587. 0.8413 is the area to the right of -1, the wrong tail.
6. D. 400 and 600 are ± 1 sd, so the proportion is 0.6827 and $1200(0.6827) = 819$, closest to 816. 600 is the count of students above the mean, not the interval; 1020 is about 85%.
7. C. The table gives 0.8106 to the left of 0.88, so the score is at about the 81st percentile. Percentile is the area to the left; it is not the z-score or the z times 100.
8. D. $P(Z < 0.6) - P(Z < -1) = 0.7257 - 0.1587 = 0.5670$. Choice B is just $P(Z < 1)$ and choice C is $P(Z < 0.6)$, so both ignore the lower bound.
9. B. Continuous variables have zero probability at any single point, so $P(X \le 40) = P(X < 40)$. Treating the height as a probability (A) confuses density with area.
10. Let $X \sim N(100, 15)$.

    (a) $z = (130 - 100)/15 = 2$, so $P(X > 130) = 1 - 0.9772 = 0.0228$.

    (b) $z = -1$ and $z = 1.33$. $P(-1 < Z < 1.33) = 0.9082 - 0.1587 = 0.7495$, about 0.75 (calculator 0.7501).

    (c) The empirical rule says 2.5% lie above $\mu + 2\sigma$. The exact answer 2.28% is close, consistent with the rule's rounding.
11. Let $X \sim N(12, 0.3)$ ounces.

    (a) Low: $z = (11.6 - 12)/0.3 = -1.33$, area 0.0918. High: $z = (12.5 - 12)/0.3 = 1.67$, right tail $1 - 0.9525 = 0.0475$. Rejected: $0.0918 + 0.0475 = 0.1393$ (calculator 0.1390). About 13.9% are rejected.

    (b) $8000(0.139) \approx 1112$ cans.
12. Let X be load time in seconds, $X \sim N(2.4, 0.5)$.

    (a) $z = (3.2 - 2.4)/0.5 = 1.6$, so $P(X > 3.2) = 1 - 0.9452 = 0.0548$ (about 5.5%).

    (b) $z = (2.0 - 2.4)/0.5 = -0.8$ and $z = (3.0 - 2.4)/0.5 = 1.2$. $P(-0.8 < Z < 1.2) = 0.8849 - 0.2119 = 0.6730$. About 67% of loads take between 2.0 and 3.0 seconds.
13. Let $X \sim N(300, 20)$. $z = (275 - 300)/20 = -1.25$ and $P(Z < -1.25) = 0.1056$. Expected days $= 365(0.1056) \approx 38.5$, about 38 or 39 days. The assumption is that daily output is approximately normal with the same mean and sd every day.
14. The total area under the curve is 1, so the area to the right of 1.2 and the area to the left add to 1: $P(Z > 1.2) = 1 - P(Z < 1.2)$. An interval is the larger left area minus the smaller left area, so every region can be built from left areas. $P(Z > 1.2) = 1 - 0.8849 = 0.1151$. $P(0.5 < Z < 1.2) = 0.8849 - 0.6915 = 0.1934$.
15. Let $X \sim N(34, 6)$. $P(\text{late}) = P(Z > 1) = 0.1587$.

    (a) Independence gives $0.1587^2 = 0.0252$.

    (b) $1 - (1 - 0.1587)^2 = 0.2921$. Interpretation: in about 29% of two-morning stretches, the student is late at least once.

2.3 Inverse Normal and Finding Cutoffs

PRACTICE · INVERSE NORMAL AND FINDING CUTOFFS

1. C. The left area 0.95 corresponds to $z = 1.645$. The value 1.96 (D) is the 97.5th percentile, the edge of the middle 95%; 1.28 (B) is the 90th percentile.
2. B. Top 10% means left area 0.90, so $z = 1.28$ and $x = 500 + 1.28(100) = 628$. Choosing 372 uses the wrong tail (z negative); 872 uses z = 3.72.
3. A. Left area 0.05 gives $z = -1.645$, so $x = 400 - 1.645(25) = 358.9$. Choice D adds instead of subtracting, giving the top 5% cutoff.
4. D. The top 15% has 85% to its left, so enter 0.85 with mean and sd. Choice A uses the right-tail area; choice B omits the mean and sd so it returns a z-score, not a score.
5. B. The middle 90% leaves 5% in each tail: $z = \pm 1.645$, so $68 \pm 1.645(4)$ gives 61.4 to 74.6. Choice A uses z = 2, which is 95%.
6. C. $z = (75 - 70)/6 = 0.83$, and the area to the left is about 0.80. The percentile is not the score 75 (B) or the z-score times 100 (D).
7. D. $Q_3 = \mu + 0.6745\sigma = 50 + 6.745 = 56.7$. Choice B uses 1.645 and gives the 95th percentile.
8. A. The 3rd percentile has $z = -1.88$, so $x = 12 - 1.88(0.3) = 11.44$. Choice D is the 97th percentile; C uses an sd that is too small.
9. D. The 40th percentile is below the median, so $z = -0.253$ and $x = 300 - 0.253(20) = 294.9$. Choice C is the 60th percentile, a reflection error.
10. (a) The 75th percentile has $z = 0.6745$: $x = 500 + 0.6745(100) = 567.4$.

    (b) The top 5% has 0.95 to the left, $z = 1.645$: $x = 500 + 1.645(100) = 664.5$. A score of about 665 or higher.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

11. Let $X \sim N(36, 3)$. The middle 99% leaves 0.5% in each tail: left areas 0.005 and 0.995, $z = \pm 2.576$. Bounds: $36 \pm 2.576(3)$ gives 28.27 and 43.73 cm. The middle 99% of plants are between about 28.3 cm and 43.7 cm.
12. Let $X \sim N(\mu, 0.3)$. We need $P(X < 12) = 0.02$, so $z = -2.054$. Then $12 = \mu - 2.054(0.3)$ and $\mu = 12 + 0.616 = 12.616$ oz. The mean must be set to about 12.62 ounces.
13. Let μ and σ be unknown. $z_{0.30} = -0.5244$ and $z_{0.90} = 1.2816$. Equations: $62 = \mu - 0.5244\sigma$ and $80 = \mu + 1.2816\sigma$. Subtract: $18 = 1.806\sigma$, so $\sigma = 9.97$. Then $\mu = 62 + 0.5244(9.97) = 67.23$. The scores have mean about 67.2 and standard deviation about 10.0.
14. (a) $Q_1 = 85 - 0.6745(12) = 76.91$ and $Q_3 = 93.09$, so IQR $= 16.19$.
    (b) Fences: $76.91 - 1.5(16.19) = 52.62$ and $93.09 + 24.28 = 117.38$. These are at $z = \pm 2.70$. The proportion outside is $2(1 - 0.9965) \approx 0.007$, so about 0.7% of a normal distribution is flagged.
15. Area from a cutoff: standardize the cutoff, then find the area. Cutoff from an area: convert the area to a left area, find z, then unstandardize. The 90th percentile: $z = 1.2816$ and $x = 100 + 1.2816(15) = 119.2$. Verify: $z = (119.2 - 100)/15 = 1.28$, and the area to the left is 0.8997, or 0.90. Therefore 119.2 is the 90th percentile.

2.4 Assessing Normality and Using Normal Models

PRACTICE · ASSESSING NORMALITY AND USING NORMAL MODELS

1. B. A normal probability plot compares the data directly with the normal pattern. A scatterplot versus time (A) shows trend, and bar or pie charts (C, D) are for categories.
2. D. Large values are larger than a normal sample would have, so the right tail is longer: right-skewed. Left-skew (A) would show the reverse curve at the lower end.
3. C. A normal curve is single-peaked. Symmetry alone (A) and a central mean (B) do not guarantee a bell shape, and sample size (D) does not change a bimodal shape.
4. A. Far more than 68% within 1 sd means the values pile up near the mean, a narrower and peaked shape than the normal. Saying they are more spread out (C) reverses this.
5. D. Incomes are strongly right-skewed. Heights, filled-box weights and long-exam scores tend to be symmetric and mound-shaped.
6. B. A median close to Q1 with a long upper tail and outlier signals right skew, which is inconsistent with a normal curve. Long whiskers (C) do not help; they suggest skew.
7. A. Normality is judged through shape, the rule and probability plots. Exact normality (C) never occurs in real data, and equal mean and median (B) alone prove nothing.
8. C. Many values at the maximum create a boundary pile-up and left skew that a bell curve cannot match. A high mean (A) by itself is not the reason.
9. A. $z = (0 - 12)/6 = -2$, so a normal model predicts $P(Z < -2) = 0.0228$, about 2.3%. Choice D uses 1 sd instead of 2, and choice B uses 3 sd. Since the variable cannot be negative, even this small positive amount signals a poor fit.
10. The 1 sd range (about 40 to 62) is close to the bins 40-60, which contain $24 + 28 = 52$ of 80, or 65%. The 2 sd range (about 29 to 73) is close to the bins 30-70, which contain $10 + 24 + 28 + 12 = 74$ of 80, or 92.5%. Both are close to 68% and 95%, and the histogram is symmetric and single-peaked. A normal model is reasonable.
11. A normal probability plot sorts the data and plots each value against its expected z-score. (a) Points close to a straight line: approximately normal. (b) Points curving upward on the right: right-skewed, because the largest values are bigger than a normal sample would give. (c) Steps: values are clustered or rounded; the data are discrete or have limited precision, so normality is questionable.
12. (a) The mean (8.1) is much larger than the median (5.0), which signals right skew. The sd is nearly as large as the mean, and the minimum is just 0.5, so there is a lot of room on the upper side only.
    (b) $z = (0 - 8.1)/7.0 = -1.157$, $P(Z < -1.157) \approx 0.124$. A normal model would predict about 12% of waiting times to be negative, which is impossible, so the model fails.
13. (a) Interval for 1 sd: 61.0 to 68.6 contains 61, 62, 62, 63, 64, 64, 65, 65, 66, 67, 68, which is 11 of 16, or 69%. For 2 sd: 57.2 to 72.4 contains all 16 values: 100%.
    (b) The counts are close to 68% and 95%, given the small n, so there is no evidence against a normal model; confirm with a graph of shape.
14. (a) Pencil lengths from a machine are continuous, symmetric and mound-shaped: normal is plausible. (b) The number of pets is discrete with few values and most students near 0 or 1: not normal. (c) Race times are usually right-skewed, with a hard lower limit: normal is questionable unless the group is homogeneous.
15. The straight body of the plot says the bulk of the diameters are approximately normal. The two low points are outliers on the low side, perhaps measurement errors or defective bolts. The engineer should investigate those two values before using normal calculations, and report results with and without them.

T Unit 2 Test

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

UNIT TEST · UNIT 2 TEST

1. B. Area must equal 1: $0.25(4) = 1$. A gives area 2, C gives area 4, and D has negative height, which is not allowed.
2. C. A long left tail pulls the mean below the median. A (mean $>$ median) is for right skew.
3. D. 350 is -2σ and 425 is $+1\sigma$: $47.5\% + 34\% = 81.5\%$. Choice A is only the 1 sd interval, and choice C ignores that the upper bound is not 2 sd.
4. A. $z = (55 - 48)/5 = 1.4$. Choice B forgets to divide by the standard deviation; D has the wrong sign.
5. B. $z = (70 - 65)/3 = 1.67$; the right tail is $1 - 0.9525 = 0.0475$. Choice C is the left area, the wrong tail.
6. C. $z = -0.67$ and $z = 1$: $0.8413 - 0.2525 = 0.5888$. Choice D is just $P(Z < 1)$, ignoring the lower cutoff.
7. A. Top 20% has 0.80 to the left, $z = 0.8416$, and $x = 70 + 8.4 = 78.4$. Choice B is the bottom 20%, a left-right error.
8. D. The middle 90% has 5% in each tail, $z = \pm 1.645$, so 70 ± 16.45. Choice A uses z = 1.28 (the middle 80%).
9. C. $z = (56 - 48)/5 = 1.6$ and the area to the left is 0.9452, about the 95th percentile. The z-score of 1.6 does not equal the 16th percentile (A).
10. A. The upward curve on the right means the largest values are higher than normal predicts: right skew. Choice C reverses the direction.
11. C. Adding a constant shifts the center but does not change the spread. The mean becomes 75 and the sd stays 8. Choice A adds 5 to the sd too.
12. D. Enter lower, upper, mean and standard deviation of the original variable. Choice A with mean 0 and sd 1 would be for z-scores.
13. Scoring guide (6 points). (a) 3 points: 1 for the model $X \sim N(10, 0.8)$ with defined variable; 1 for both z-scores ($z = -1.25$ and $z = 1.875$); 1 for adding the tails: $P(Z < -1.25) = 0.1056$ and $P(Z > 1.875) = 0.0304$, total about 0.136. (b) 1 point: $2000(0.136) \approx 272$ tubes. (c) 2 points: 1 for $z = 1.2816$ for left area 0.90, 1 for $x = 10 + 1.2816(0.8) = 11.03$ cm.

    Model response: Let X be tube length, $X \sim N(10, 0.8)$. (a) $P(X < 9) = P(Z < -1.25) = 0.1056$ and $P(X > 11.5) = P(Z > 1.875) = 0.0304$, so about 13.6% of tubes are defective. (b) About 272 of 2000. (c) The 90th percentile has $z = 1.28$, so $x = 10 + 1.28(0.8) = 11.03$; tubes of 11.03 cm or longer are extra long.
14. Scoring guide (10 points). (a) 2 points: $z = (81 - 70)/12 = 0.917$ (1); percentile 0.8203, about the 82nd (1). (b) 2 points: $z = -0.833$ and $z = 0.417$ (1); $0.6616 - 0.2023 = 0.459$ (1). (c) 2 points: $z = 1.036$ (1); $x = 70 + 1.036(12) = 82.4$ (1). (d) 2 points: within 1 sd (4.41 to 21.39) is 16 of 20 = 80% (1); within 2 sd (-4.1 to 29.8) is 19 of 20 = 95%; the 1 sd value is well above the 68% of the rule and the data are skewed (1). (e) 2 points: not appropriate, right skew (1); $z = (0 - 12.9)/8.49 = -1.52$ so the model predicts about 6.4% negative times, which is impossible (1).

    Model response: (a) $z = 0.917$, so the score is at about the 82nd percentile. (b) $P(60 < X < 75) = P(-0.83 < Z < 0.42) = 0.459$. (c) The top 15% starts at the 85th percentile: $70 + 1.036(12) = 82.4$. (d) 16 of 20 (80%) are within one sd and 19 of 20 (95%) within two; the one-sd percent is too high and the mean (12.9) exceeds the median (10.5). (e) The curved normal probability plot shows right skew, so a normal model is not appropriate; it would predict 6.4% negative times, which cannot happen.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 3: Exploring Two-Variable Data

Describe and model relationships between two variables. 6 lessons, 104 practice and test items, all with full explanations.

In this unit · Exam Unit 5 (Regression)

3.1 Two-Way Tables and Categorical Association
3.2 Scatterplots: Form, Direction, Strength, Outliers
3.3 Correlation
3.4 Least-Squares Regression Lines
3.5 Residuals, r-squared, and Standard Deviation of Residuals
3.6 Transforming Data to Achieve Linearity
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

3.1 Two-Way Tables and Categorical Association

Learn to read two-way tables, find marginal and conditional distributions, compare groups with segmented bar graphs, and explain why association is not causation.

Why This Matters

Most interesting questions in statistics involve two variables at once: Do students who ride the bus differ from students who drive in what they prefer? Is one training method linked to a better result? When both variables are categories, the tool is the **two-way table**, and the skill is turning raw counts into the *right* percentages. Students lose points by comparing counts instead of percentages, by dividing by the grand total when the question asks for a conditional percentage, and by writing the word "cause" when the data only show a link.

Describing association between categorical variables is the first step toward the regression ideas that dominate Exam Unit 5, and the same table skills return in probability (conditional probability) and in chi-square work later in this book. Expect free-response questions that give a table and ask you to compare groups, and multiple-choice items that test whether you chose the correct denominator.

Core Ideas

A **two-way table** (contingency table) classifies each individual by two categorical variables. Rows show one variable, columns the other, and each cell counts individuals with that pair of values. Add a row of column totals and a column of row totals and you have the margins.

Vocabulary for Two-Way Tables

Marginal distribution: the distribution of ONE variable alone, found from the totals in the margin. Divide each row total (or column total) by the grand total.

Conditional distribution: the distribution of one variable among only the individuals who have a specific value of the other variable. Divide each cell by the total of the row (or column) you are conditioning on.

Joint relative frequency: a cell count divided by the grand total.

Read "given", "among", "of those who" as signals: the group named after those words is the denominator.

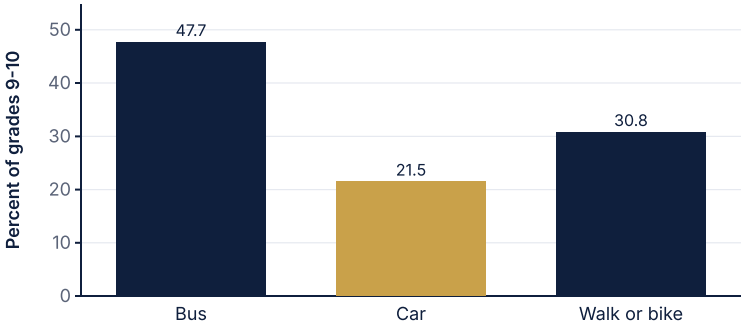
Here is a fictional survey of 260 students at one school about how they usually get to school. Study it; we will use it throughout the lesson.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

	Bus	Car	Walk or bike	Total
Grades 9-10	62	28	40	130
Grades 11-12	38	72	20	130
Total	100	100	60	260

The right-hand column and the bottom row are the margins. The marginal distribution of transportation is 100/260 = 38.5% bus, 38.5% car, 23.1% walk or bike. Conditional distributions let us compare the two grade groups fairly. Among grades 9-10, the percentages are 47.7% bus, 21.5% car, 30.8% walk or bike. Among grades 11-12 they are 29.2%, 55.4%, 15.4%.



Conditional distribution of transportation among grades 9-10 (the three percentages add to 100).

When Are Two Variables Associated?

Two categorical variables are **associated** if the conditional distributions of one variable differ for different values of the other. If the conditional distributions are (nearly) identical, there is no association.

Compare the conditional distributions with a **segmented (stacked) bar graph** where every bar is scaled to 100 percent, or with a side-by-side table of conditional percentages.

Use percentages, never raw counts, when the group sizes differ.

A **segmented bar graph** draws one bar for each value of the explanatory variable and splits each bar into segments showing the conditional percentages of the response variable. If the segment boundaries line up across bars, there is no association. In our survey, the car segment is much taller for grades 11-12 than for grades 9-10, which signals association.

Association is not causation. Finding that older students drive more does not show that being in grade 11 causes driving to school in any mechanism we can isolate: age, licensing rules, and family circumstances are all tangled together. Only a well-designed randomized experiment can support a cause-and-effect conclusion (Unit 4 of this book). In an observational study a **confounding variable** is one that is related to both the explanatory and response variables and mixes up their effects.

Sometimes combining groups reverses a comparison. This is **Simpson's paradox**: a trend that appears in every subgroup can reverse when the subgroups are pooled, because the subgroups differ greatly in size. Example 6 shows it.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: MARGINAL DISTRIBUTION

Using the transportation table above, find the marginal distribution of how students get to school.

	Bus	Car	Walk or bike	Total
Grades 9-10	62	28	40	130
Grades 11-12	38	72	20	130
Total	100	100	60	260

SOLUTION

Divide each column total by the grand total 260.

Bus: 100/260 = 38.5%. Car: 100/260 = 38.5%. Walk or bike: 60/260 = 23.1%.

The percentages sum to 100.1 only because of rounding. Conclusion: students are equally likely to take the bus or a car, and about a quarter walk or bike.

EXAMPLE 2: CHOOSING THE DENOMINATOR

(a) What percent of the grade 11-12 students drive a car? (b) What percent of the students who drive a car are in grades 11-12? Explain why the answers differ.

SOLUTION

(a) The phrase "of the grade 11-12 students" makes that row the denominator: 72/130 = 55.4%.

(b) "Of the students who drive" makes the car column the denominator: 72/100 = 72.0%.

The numerator is the same, but the groups are different. Statement (a) conditions on grade; statement (b) conditions on transportation mode. Reversing the condition changes the denominator and therefore the answer.

EXAMPLE 3: IS THERE AN ASSOCIATION?

Compare the distributions of transportation for the two grade groups. Is there evidence of association?

SOLUTION

Conditional distribution for grades 9-10: bus 62/130 = 47.7%, car 28/130 = 21.5%, walk or bike 40/130 = 30.8%.

Grades 11-12: bus 38/130 = 29.2%, car 72/130 = 55.4%, walk or bike 20/130 = 15.4%.

The two distributions are quite different (car use is 55.4% versus 21.5%). Because the conditional distributions differ, transportation mode and grade group are associated in this sample. Both groups have 130 students, so counts and percentages tell the same story here, but percentages are the safe comparison.

EXAMPLE 4: FILLING IN A TABLE

A club of 80 members is split by instrument family and by year in the program. Of the members, 35 are new and 45 are returning. 28 members play strings, and of the strings players 12 are new. Complete the two-way table and compare new and returning members.

	Strings	Other	Total
New	12	23	35
Returning	16	29	45
Total	28	52	80

SOLUTION

Strings: 12 new + 16 returning = 28 (since 28 - 12 = 16).

Other instruments: 35 - 12 = 23 new and 45 - 16 = 29 returning, total 52.

Among the 35 new members, 12 play strings: 12/35 = 34.3%. Among the 45 returning members, 16/45 = 35.6%. These are close, so there is little association between year in the program and instrument family.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: ASSOCIATION VERSUS CAUSATION

The survey in this lesson found that students in grades 11-12 are more likely to drive. A student writes: "Being older makes students choose cars, so the school should make students younger to reduce traffic." Evaluate the claim.

SOLUTION

The data come from an observational survey. No one was randomly assigned to a grade.

Grade is associated with driving, but variables such as holding a license, owning a car, and distance from home are confounders linked to both grade and transportation.

A conclusion of causation is not justified. Even without causation the association is real and useful for prediction: the school could expect more cars when upper-grade enrollment rises.

EXAMPLE 6: SIMPSON'S PARADOX

Two repair shops fix devices. Each job is easy or hard. On-time repairs: Shop A: easy 19 of 20, hard 54 of 180. Shop B: easy 160 of 180, hard 4 of 20. Compare the shops (a) within each job type and (b) overall.

	Easy on time	Hard on time
Shop A	19 of 20	54 of 180
Shop B	160 of 180	4 of 20

SOLUTION

(a) Easy jobs: A 95.0% on time versus B 88.9%. Hard jobs: A 30.0% versus B 20.0%. Shop A is better for each job type.

(b) Overall: A has 73 of 200 = 36.5% on time; B has 164 of 200 = 82.0%. Shop B looks far better.

The reversal happens because Shop A is given mostly hard jobs (180 of 200) while Shop B is given mostly easy ones. Job difficulty is a lurking variable linked to both shop and outcome. The fair comparison controls for job type, so Shop A performs better.

Write It Like a 5

Model Response

Prompt: A fictional poll of 300 listeners asked how they mainly listen to music. Among 120 listeners under 30, 66 prefer streaming, 30 prefer radio, 24 prefer owned files. Among 180 listeners 30 and over, 54 prefer streaming, 90 prefer radio, 36 prefer owned files. Is there an association between age group and listening preference? Support your answer. Response: Because the groups have different sizes, I compare conditional percentages within each age group. Under 30: streaming 66/120 = 55.0%, radio 30/120 = 25.0%, owned files 24/120 = 20.0%. Age 30 and over: streaming 54/180 = 30.0%, radio 90/180 = 50.0%, owned files 36/180 = 20.0%. The conditional distributions differ noticeably: streaming is much more common among listeners under 30 (55.0% versus 30.0%) and radio much more common among listeners 30 and over (50.0% versus 25.0%), while owned files are equal at 20.0%. Therefore, in this sample, age group and listening preference are associated. Since this is a poll and not an experiment, we cannot conclude that age causes the preference.

What the grader looks for

(1) Uses conditional percentages computed within each age group, not raw counts or grand-total percentages. (2) Computes the values correctly. (3) Compares the distributions with specific numbers. (4) States a conclusion about association in context. (5) Avoids a causal claim for observational data.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. **Circle the grouping phrase.** Words after "of", "among", or "given" name the denominator.
2. **Choose rows or columns.** Condition on the explanatory variable, so each compared group is a row (or column) of its own.
3. **Check your percentages.** Conditional percentages add to 100 within each group; marginal percentages add to 100 across the margin.
4. **Compare with numbers.** Quote at least two percentages side by side, then give a conclusion in context.
5. **Ask about confounding.** If the study is observational, name a plausible lurking variable instead of claiming cause.

- On the built-in calculator, divide cells by totals using a list or table; round to one decimal place only at the end.
- A percent difference of a few points in a small sample may be sampling variation. Say "in this sample" for descriptive comparisons.
- Mental model: association means "knowing one variable changes your guess about the other."

Common traps

(1) Using the grand total as the denominator when the question says "given" or "among." (2) Comparing counts when the groups have different sizes. (3) Concluding that association proves one variable causes the other. (4) Mixing up P(A given B) with P(B given A), as in Example 2.

Mistake Clinic

Spot the error

Student claim: "Of the 100 students who drive, 72 are in grades 11-12, so 72/260 = 27.7% of drivers are in grades 11-12." Where does this fail? The question conditions on drivers, so the denominator is the number of drivers (100), not 260. Correct: 72/100 = 72.0% of drivers are in grades 11-12. The value 27.7% is the joint relative frequency: the share of ALL students who are both drivers and upperclass.

Spot the error

Student claim: "Shop B has a higher overall on-time rate, so Shop B is better at every kind of job." This fails because the overall rate mixes job types of very different difficulty. Correct: compare within job type (A: 95.0% and 30.0%; B: 88.9% and 20.0%). Shop A is better on both easy and hard jobs; Shop B only looks better because it handles mostly easy jobs.

Spot the error

Student claim: "Teens who sleep more have higher grades, so more sleep causes better grades." The data come from a survey. Correct: the data show association only. A confounder such as study habits or time management might lead to both more sleep and better grades. Only a randomized experiment could support a causal claim.

Real-World Connection

Hospitals, schools, and sports leagues use two-way tables to compare outcomes between groups. Whenever a news headline compares two percentages, ask: what is the denominator, and could a lurking variable (such as the difficulty of the cases) explain the difference?

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

A fictional survey of 400 people has 160 who prefer cooking at home and 240 who prefer eating out. Of the 240 who prefer eating out, 60 percent are under 40. Overall, 220 people are under 40. Are cooking preference and being under 40 associated in this sample? Solution: Under 40 and eat out: 0.60 × 240 = 144. Under 40 and cook at home: 220 - 144 = 76. Of the 160 who cook at home, 76/160 = 47.5% are under 40, versus 60.0% of the eat-out group. The conditional percentages differ, so the variables are associated in this sample. The overall under-40 rate is 220/400 = 55%, which equals neither conditional rate, as expected when association exists.

Practice

PRACTICE **Two-Way Tables and Categorical Association**

Show your work. A calculator is allowed. Justify conclusions in context.

1. In the transportation table of this lesson, which expression gives the percent of bus riders who are in grades 9-10?
   (A) 62/100 **(B)** 62/130 **(C)** 62/260 **(D)** 100/260
2. A segmented bar graph shows two bars, one per class section, each split into Pass and Fail. The Pass segment is 70 percent of each bar. This suggests
   (A) a strong association between section and result
   (B) no association between section and result
   (C) that section A caused more students to pass
   (D) that the sections have equal numbers of students
3. Of 50 seniors, 30 take a language class. Of 150 juniors, 60 take a language class. Which statement is true?
   (A) Juniors are more likely because 60 is greater than 30
   (B) The groups are equally likely because 90 of 200 is 45%
   (C) Seniors are more likely than juniors to take a language class
   (D) No comparison is possible without more data
4. Which is the best description of a marginal distribution?
   (A) The distribution of one variable obtained from the totals in a table margin
   (B) The distribution of one variable among individuals with a given value of another
   (C) The count in a single cell divided by the grand total
   (D) The difference between two conditional percentages
5. In a sample, 40 students are left-handed and 360 are right-handed. 10 left-handed and 90 right-handed students play string instruments. What is the best conclusion?
   (A) Left-handed students are less likely to play strings because 10 < 90
   (B) Handedness and playing strings show no association in this sample
   (C) Right-handed students are more likely to play strings
   (D) Playing strings causes right-handedness
6. A study finds that cities with more ice cream sales have more drownings. The best explanation is
   (A) ice cream causes drowning **(B)** drowning causes ice cream sales
   (C) the study must have been an experiment **(D)** a confounding variable such as hot weather
7. Two clinics treat easy and severe cases. Clinic 1 has the higher success rate for each type of case, yet Clinic 2 has the higher overall success rate. This is called
   (A) the law of large numbers **(B)** selection bias in sampling
   (C) Simpson's paradox **(D)** the central limit theorem
8. 24 of 80 girls and 36 of 120 boys prefer a gym class to study hall. Which is correct?
   (A) Boys prefer gym because 36 > 24
   (B) Girls prefer gym because 24/80 > 36/120
   (C) Preference is associated because the group sizes differ
   (D) Preference is not associated with gender in this sample, because both rates are 30%

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. In a segmented bar graph comparing two groups, the 'agree' segment is 40 percent of Group 1's bar and 65 percent of Group 2's bar. What does this show?
   (A) The variables are independent because both bars have the same height
   (B) The variables are associated, since the conditional distributions differ
   (C) Group 2 has more people than Group 1
   (D) Group membership causes agreement
10. A fictional survey of 150 teens: of 90 who play a sport, 54 get 8 or more hours of sleep; of 60 who do not play a sport, 24 get 8 or more hours of sleep.
    (a) Make a two-way table with totals.
    (b) Compute the conditional percentages of 8+ hours of sleep for each group.
    (c) Is there an association? Can you conclude that playing a sport causes more sleep?
11. Use the transportation table from this lesson (grades 9-10: 62 bus, 28 car, 40 walk; grades 11-12: 38 bus, 72 car, 20 walk).
    (a) What percent of all students walk or bike?
    (b) Of the students who take the bus, what percent are in grades 11-12?
    (c) What percent of students are in grades 11-12 AND take a car?
12. In a fictional school, 120 of 200 students in the choir are in the music-theory class, and 60 of 100 students not in the choir are in the music-theory class.
    (a) Compute the two conditional percentages.
    (b) What does your answer say about association?
    (c) Suppose 100 more students join the choir, all of them taking music theory. Recompute and explain why percentages, not counts, are the right comparison.
13. Two tutoring centers report passing rates. Center X: first-time students 90 of 100 pass, repeat students 5 of 20 pass. Center Y: first-time students 17 of 20 pass, repeat students 30 of 100 pass.
    (a) Compare the centers within each student type.
    (b) Compare overall.
    (c) Explain the reversal.
14. A fictional survey of 500 adults found that people who eat breakfast daily have a lower rate of missing work than those who do not.
    (a) Is this an experiment or an observational study?
    (b) Name one possible confounding variable and explain how it could create this association.
    (c) Describe how a randomized experiment could investigate cause and effect.
15. A fictional survey about a school start time: Grade 9: 45 favor, 15 oppose. Grade 12: 30 favor, 50 oppose.
    (a) Build the full table with totals.
    (b) Give the marginal distribution of opinion.
    (c) Describe the association using the heights of the 'favor' segments of a segmented bar graph.

3.2 Scatterplots: Form, Direction, Strength, Outliers

Learn to draw and describe scatterplots with form, direction, strength and unusual features, to pick explanatory and response variables, and to recognize curves, clusters and outliers.

Why This Matters

A scatterplot is the first picture you draw whenever two quantitative variables are measured on the same individuals. It answers three questions before any formula is used: Is there a pattern? How strong is it? Is anything strange? Students who skip the picture and jump to a line often fit a straight line to a curve or let one odd point control the whole story.

On the AP exam (Exam Unit 5, Regression), free-response questions ask you to describe a scatterplot in context using **form, direction, strength, and unusual features**, and multiple-choice questions ask you to identify which variable is explanatory or to judge which statement fits the plot. Good descriptions use the names of the variables, not just the words "x" and "y".

Core Ideas

Each individual contributes one point (x, y). The **explanatory variable** x is the one we think helps explain or predict changes in the **response variable** y, and it goes on the horizontal axis. Calling one variable explanatory does not mean that it causes the other to change; it only says which one we are using to explain or predict.

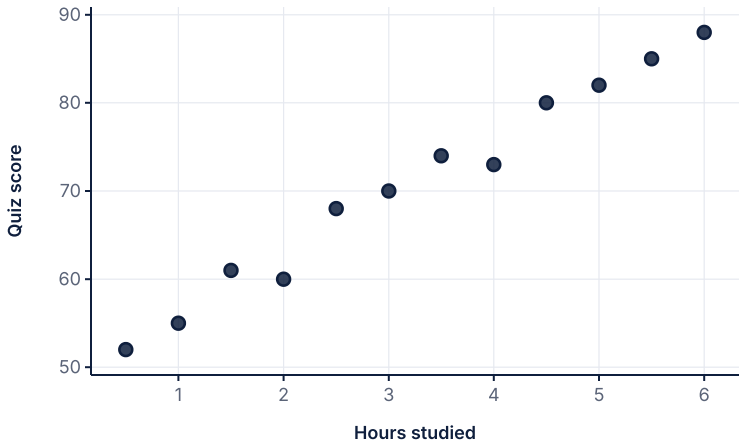
Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Describing a Scatterplot (FDSU)

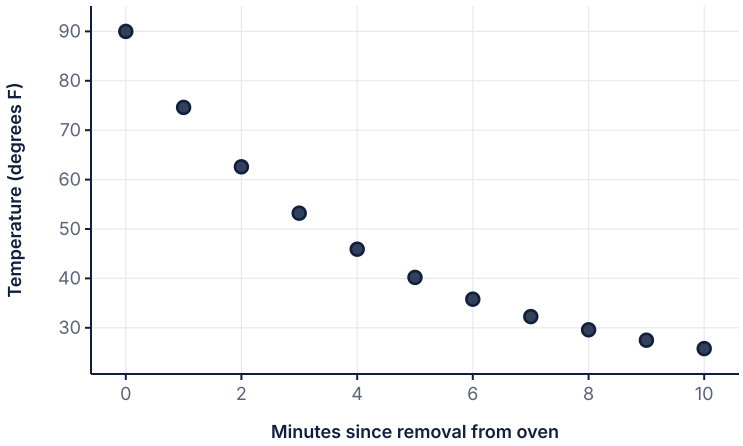
Form: linear, curved, or no clear form. Are there clusters or separate groups?
Direction: positive association (y tends to increase as x increases) or negative association (y tends to decrease as x increases).
Strength: how closely the points follow the form: weak, moderate, or strong.
Unusual features: outliers (points far from the pattern), clusters, gaps, and points that have an unusual x value.
Always describe in the context of the variables and units.

The next plot shows 12 fictional students: hours spent studying for a quiz (explanatory) versus quiz score (response, out of 100).



Study time versus quiz score for 12 fictional students: a positive, roughly linear, strong pattern.

The form is roughly linear, the direction is positive, and the strength is strong (the correlation is about 0.99, a number we study in the next lesson). There are no obvious outliers.



Cooling of a fictional dish: a negative, curved (not linear) pattern.

The cooling plot is negative and strong but **curved**: the temperature drops quickly at first and then levels off. A straight line would be a poor model, even though the points are strongly related. Do not describe such a pattern as "linear."

Reading Strength and Unusual Points

Strength is about how tightly the points hug the pattern, not about the slope. A steep, loose cloud is weaker than a gentle, tight one.
An **outlier** in a scatterplot is a point that falls outside the overall pattern: its y value is far from what the pattern predicts, or its x value is far from the other x values.
If a plot has two clusters, ask whether a third variable (such as vehicle type) separates the groups.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: EXPLANATORY OR RESPONSE?

For each pair, choose the explanatory variable and the response variable: (a) a runner's weekly training distance and her 5K time; (b) the number of rainy days in a month and the amount of umbrella sales; (c) the age of a laptop and its resale price.

SOLUTION

(a) Explanatory: weekly training distance; response: 5K time. We are asking whether more training predicts faster times.

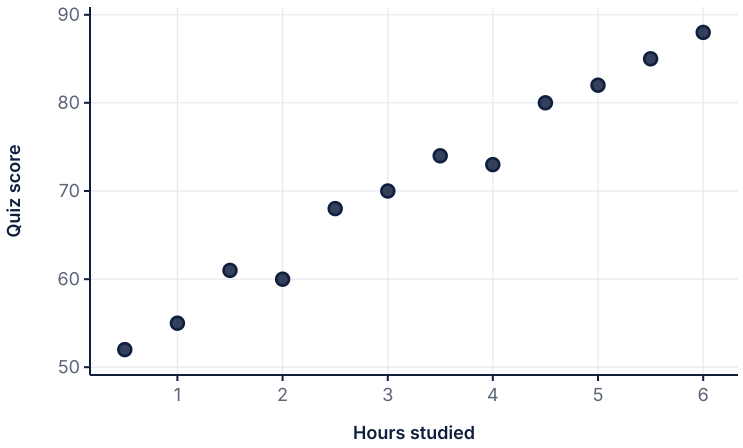
(b) Explanatory: number of rainy days; response: umbrella sales.

(c) Explanatory: age of the laptop; response: resale price.

In each case the choice depends on the question asked. If we instead wanted to predict training from 5K time, the roles would swap.

EXAMPLE 2: DESCRIBING A LINEAR PATTERN

Describe the association in the study-time plot.



SOLUTION

Form: roughly linear. Direction: positive, since students who studied more tended to earn higher quiz scores. Strength: strong, because the points lie close to a line.

Unusual features: no outliers and no clusters.

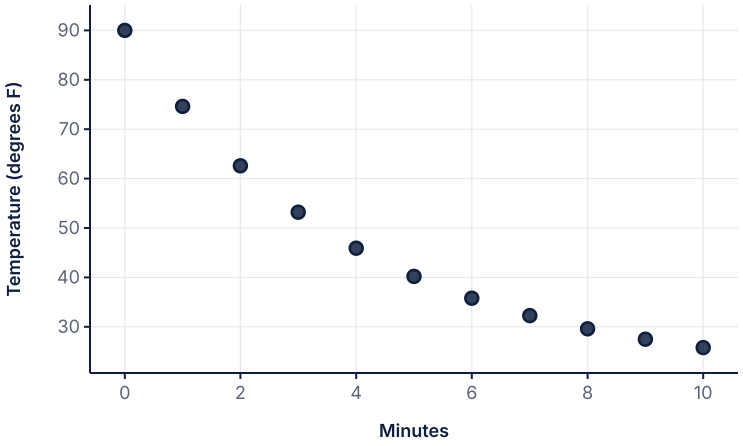
Summary in context: There is a strong, positive, roughly linear association between hours studied and quiz score for these 12 students (r about 0.99).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: A CURVED PATTERN

Describe the cooling data and explain whether a line is appropriate.



SOLUTION

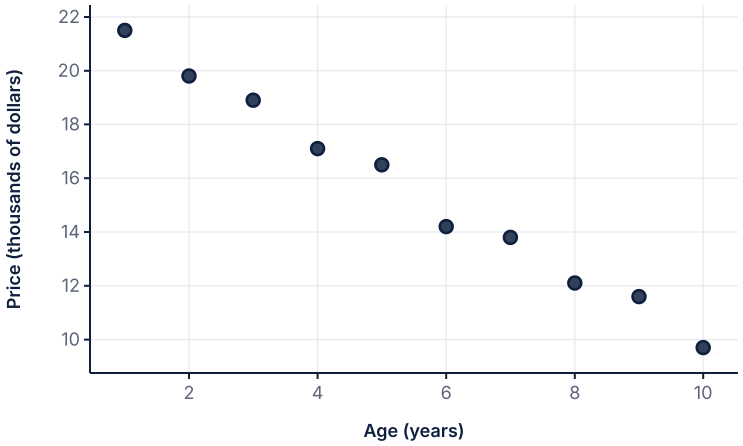
Form: curved (decreasing at a decreasing rate). Direction: negative. Strength: strong, the points follow a smooth curve very closely.

A straight line would overestimate in the middle of the data and underestimate at the ends, so a linear model is not appropriate.

Even so, a linear correlation computed on these data is about -0.95. A large |r| does not show that the form is linear.

EXAMPLE 4: NEGATIVE AND STRONG

The plot shows the age (years) and asking price (thousands of dollars) of 10 fictional used cars. Describe the association.



SOLUTION

Direction: negative (older cars cost less). Form: roughly linear. Strength: strong, the points follow a line closely.

The correlation is about -1.00. In context: there is a strong, negative, linear association between a car's age and its asking price; each extra year goes with a lower price.

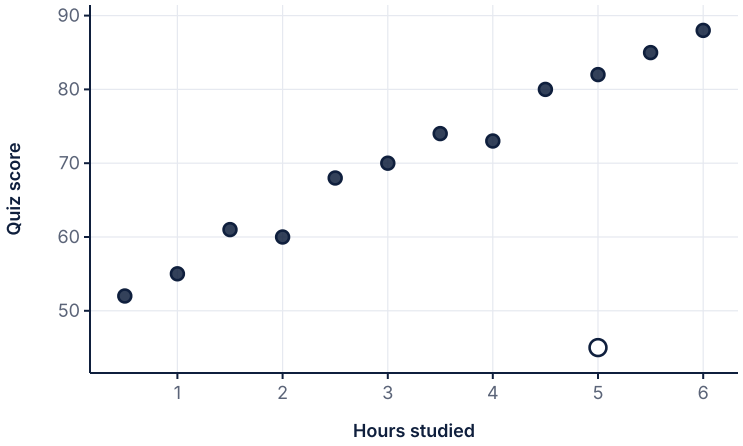
No outliers are visible.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: AN UNUSUAL POINT

A 13th student studied 5.0 hours but scored only 45. Describe how this point fits into the study-time plot and what it does to the correlation.



SOLUTION

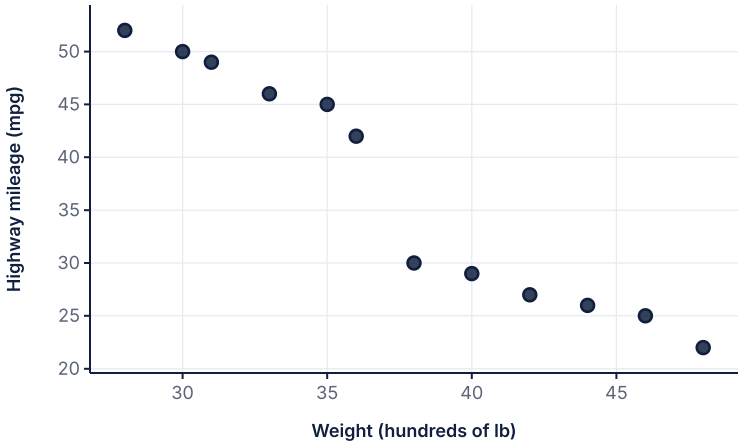
The point (5.0, 45) is far below the other students who studied about 5 hours (scores near 82). It is an outlier in the y direction: its x value is ordinary, but its response is far from the pattern.

Correlation without it: about 0.99. With it: about 0.66. The outlier weakens the apparent strength of the relationship.

The point should be investigated (illness? a data entry error?) rather than automatically removed. A good description mentions the point and, if it is removed, says so and why.

EXAMPLE 6: TWO CLUSTERS

The plot shows weight (hundreds of pounds) and highway mileage (mpg) for 12 fictional vehicles. The first six points are compact cars, the last six are light trucks. Describe the pattern.



SOLUTION

Overall form: roughly linear and negative; heavier vehicles get fewer miles per gallon.

But the points form two clusters with a gap between weights 36 and 38. The compact cars (first cluster) have much higher mileage than the trucks, and within each cluster the trend is also negative.

Pooling the groups gives r about -0.97. Because the clusters suggest a third variable (vehicle type), it may be better to analyze the groups separately.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt: A fictional gym records, for 10 members, the number of weekly visits and the member's resting heart rate (beats per minute). The scatterplot slopes downward, the points are fairly close to a line, and one member with 7 visits has a heart rate much higher than the rest. Describe the association. Response: There is a moderately strong, negative, roughly linear association between weekly gym visits and resting heart rate: members who visit more often tend to have lower resting heart rates. One member who visits 7 times per week has a resting heart rate far above the pattern predicted by the other members, so that point is an outlier in the response direction. It should be checked for a recording error or a special circumstance before drawing conclusions. This is an observational association; it does not show that more visits cause lower heart rates.

What the grader looks for

(1) Names direction in context. (2) Names form (linear) in context. (3) Describes strength with a justification. (4) Identifies and locates the unusual point. (5) Uses variable names, and avoids causal language.

Coach's Playbook

1. **Choose axes first.** Explanatory on the horizontal axis, response on the vertical.
2. **Run FDSU.** Form, Direction, Strength, Unusual features, in that order, every time.
3. **Use words from the story.** "Older cars tend to cost less" beats "y decreases."
4. **Look for curve and clusters before computing anything.** A numerical summary cannot see them.
5. **Do not overreach.** Describe association; claim cause only for randomized experiments.

- On the built-in calculator, enter x and y in two lists and plot to check form before you compute r.
- Strength words are judgment calls; support them with how tightly points cluster, and mention r only after the form is known to be linear.
- A good description is 3 to 5 sentences with the variable names in them.

Common traps

(1) Calling a curved pattern "linear" because r is large. (2) Reversing explanatory and response variables. (3) Describing strength by steepness of the slope. (4) Forgetting to mention an outlier, or removing it without justification.

Mistake Clinic

Spot the error

Student: "The cooling plot has r = -0.95, so the relationship is linear and strong." Where does this fail? A large $|r|$ does not prove linearity. Correct: the plot is curved (negative, strong, nonlinear); a straight line would systematically miss the shape.

Spot the error

Student: "The line is steep, so the association between age and price is strong." Strength depends on how close points are to the pattern, not on steepness. Correct: judge the closeness of the points to a line (or curve); the slope depends on the units chosen.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Student: "There is a positive association: weight is the response and mpg is the explanatory variable." The direction here is negative (heavier means fewer mpg), and the roles are reversed if we are predicting mileage from weight. Correct: weight is explanatory, mileage is the response, and the association is negative.

Real-World Connection

Fitness trackers, weather apps and school dashboards all start with scatterplots: steps versus sleep, humidity versus temperature, absences versus grades. Analysts check form and unusual points first, because a few bad sensor readings can change the story.

Challenge Corner

In the 12-student study-time data, a 13th student studied 12 hours (twice the largest value so far) and scored 100. Does this point look like an outlier? Solution: Its x value is far beyond the other students (maximum 6 hours), so it is unusual in the x direction. The line fit to the first 12 students is about score = 49.5 + 6.52 (hours), which would predict about 128 at 12 hours; a score is capped at 100, so the observed 100 is below the extrapolated line but only because of the ceiling. The new correlation is about 0.94, still strong. The point is not an outlier in y relative to the pattern in the way Example 5's student was, but because of its extreme x value it can pull a fitted line strongly (this is leverage, Lesson 3.5). Describe it as a point with an extreme x value that roughly follows the trend, and note that predictions at 12 hours are extrapolation.

Practice

PRACTICE **Scatterplots: Form, Direction, Strength, Outliers**

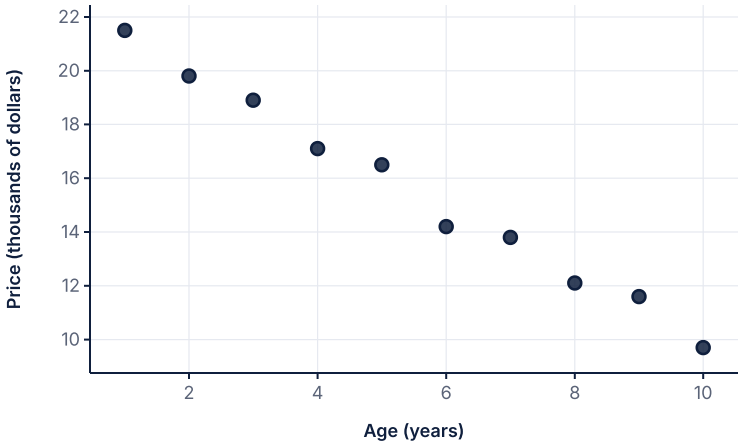
Show your work. A calculator is allowed. Justify conclusions in context.

1. In a study of whether temperature affects ice cream sales, which is the explanatory variable?
   (A) Ice cream sales **(B)** The month of the year **(C)** Temperature
   (D) The price of ice cream
2. A scatterplot shows that as a bike's weight increases, its top speed decreases. The association is
   (A) negative **(B)** positive **(C)** linear **(D)** strong
3. Which description best fits a scatterplot of the cooling dish data (temperature versus time)?
   (A) Strong, negative, linear **(B)** Strong, negative, curved
   (C) Weak, positive, curved **(D)** Moderate, negative, linear
4. In a scatterplot, a point is an outlier if
   (A) it falls well outside the overall pattern of the other points
   (B) it has the largest x value
   (C) it has the largest y value
   (D) it lies exactly on the fitted line

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

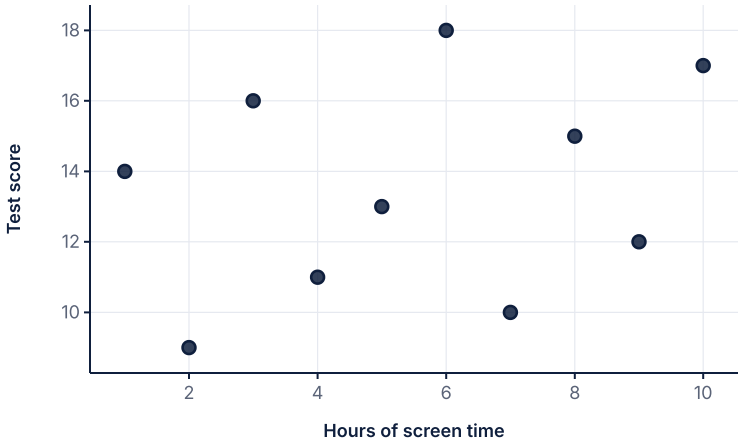
© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. Which statement best describes the scatterplot of age and price of used cars?



(A) Weak, negative, linear association
(B) Strong, positive, linear association
(C) Strong, negative, curved association
(D) Strong, negative, roughly linear association

6. The scatterplot shows screen time and a test score for 10 fictional students. Which is the best description?



(A) A strong positive linear association
(B) A strong negative linear association
(C) A weak association with no clear form
(D) A strong curved association

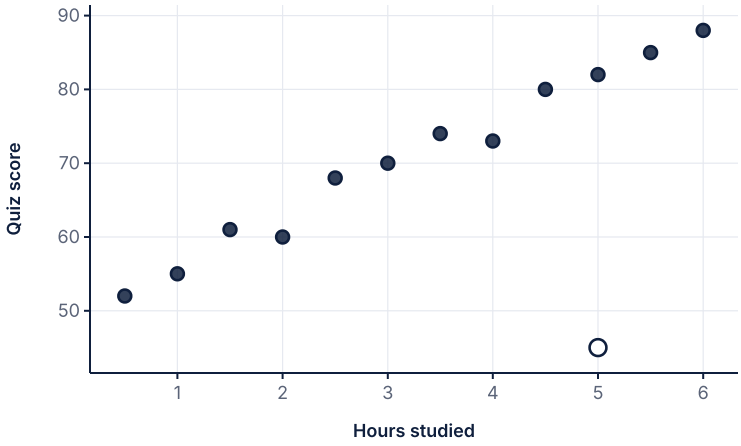
7. A scatterplot of 100 students shows a strong positive linear association between shoe size and reading score. The best explanation is

(A) bigger feet cause better reading
(B) grade level is a lurking variable linked to both
(C) better reading causes bigger feet
(D) the plot must have an error

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

8. In the scatterplot, the highlighted point (5.0, 45) is best described as



(A) an outlier in the y direction
(B) an outlier in the x direction
(C) an influential point with a high x value
(D) part of the linear pattern

9. Two clusters appear in a scatterplot of weight versus mileage for compact cars and trucks. The most appropriate response is to

(A) remove the cluster with fewer points
(B) compute one line and ignore the clusters
(C) report that there is no association
(D) consider analyzing the groups separately and look for a lurking variable like vehicle type

10. A fictional teacher records, for 8 students, the number of absences in a term (x) and the final grade (y): (0, 94), (1, 91), (2, 90), (3, 84), (4, 83), (5, 77), (6, 76), (8, 68).

(a) Which variable is explanatory?

(b) Describe the association using form, direction, strength, and unusual features.

(c) Compute r.

11. A fictional scatterplot of dog age (years) versus daily activity (hours) is curved: activity falls rapidly for ages 1 to 5, then levels off.

(a) Describe the form and direction.

(b) Why would a straight line be a poor summary?

(c) Does the pattern show that aging causes less activity? Explain.

12. For each situation choose the explanatory and response variables and state whether a positive or negative association is expected.

(a) Number of hours of sleep and number of mistakes on a typing test.

(b) Outdoor temperature and a home's heating bill.

(c) Number of pages in a book and its price.

13. A fictional dataset has 12 pairs of (hours studied, quiz score). Without the 13th student, r is about 0.99. Add a student who studied 5.0 hours and scored 45.

(a) What happens to r?

(b) Is the new point an outlier in x, in y, or both?

(c) What should an analyst do before deciding to delete the point?

14. A fictional survey compares weekly practice time and the number of errors in a piano recital for 20 students. The scatterplot shows a negative curved pattern with two students who practice 20 hours but still make many errors.

(a) Describe the association.

(b) Name the unusual features.

(c) Suggest a lurking variable that could explain the two students.

15. Weight and mileage data for 12 fictional vehicles form two clusters (compact cars and trucks), and pooling gives r about -0.97.

(a) Describe the overall pattern.

(b) Why might separate analyses be better?

(c) Within the compact cars (weights 28 to 36), is the association positive or negative?

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

3.3 Correlation

Use the correlation coefficient r to describe the direction and strength of a linear association, learn its properties, and recognize what correlation does not tell you.

Why This Matters

After you see a linear pattern in a scatterplot, you want one number that says how strong it is. That number is the **correlation coefficient** r. It is one of the most quoted and most misused statistics: students say $r = 0.8$ means "80% correlated", assume a high r proves cause, or conclude that $r = 0$ means the variables are unrelated.

In Exam Unit 5, you are expected to interpret r, state its properties, match values of r to scatterplots, and know when it is not appropriate. You will rarely compute r by hand; the built-in calculator gives it, and computer output lists it. What earns points is knowing what it means and what it does not.

Core Ideas

The correlation r measures the **direction and strength of the linear association** between two quantitative variables. Its value comes from standardizing each variable (turning values into z-scores) and averaging the products of the z-scores.

The Correlation Coefficient

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$-1 \le r \le 1$. The sign gives the direction; $|r|$ near 1 means a strong linear pattern; $|r|$ near 0 means little linear pattern. $r = 1$ or $r = -1$ only when all points lie exactly on a line.

Properties of r

No units. r does not change if you change units (inches to centimeters) or add a constant to every value of either variable.
Symmetric. Swapping x and y gives the same r.
Only linear. r measures linear association. A strong curved pattern can have r near 0.
Not resistant. One outlier can raise or lower r a lot.
Not causation. A large $|r|$ never shows that x causes y.

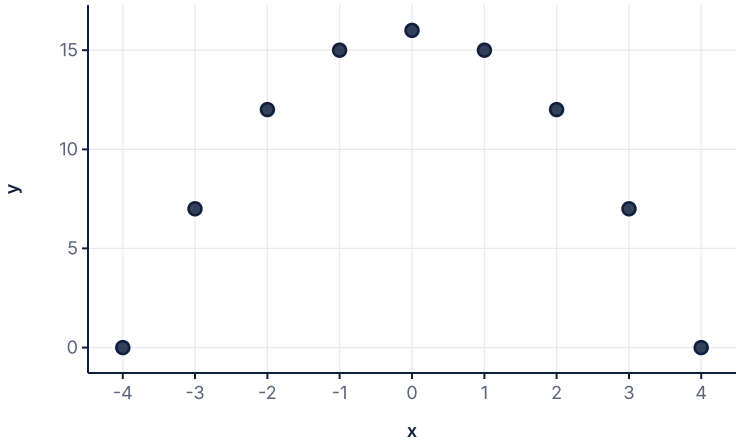
Consider the five fictional points $(1, 3), (2, 5), (3, 4), (4, 8), (5, 10)$. The means are $\bar{x} = 3$ and $\bar{y} = 6$, and the standard deviations are $s_x = 1.581$ and $s_y = 2.915$. The z-score products are listed below.

x	y	z for x	z for y	product
1	3	-1.265	-1.029	1.302
2	5	-0.632	-0.343	0.217
3	4	0.000	-0.686	0.000
4	8	0.632	0.686	0.434
5	10	1.265	1.372	1.735

The products add to 3.688; dividing by $n - 1 = 4$ gives $r = 0.922$. Points in the upper right and lower left (both z-scores with the same sign) contribute positively and push r toward 1; points in the other two quadrants push it toward -1.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



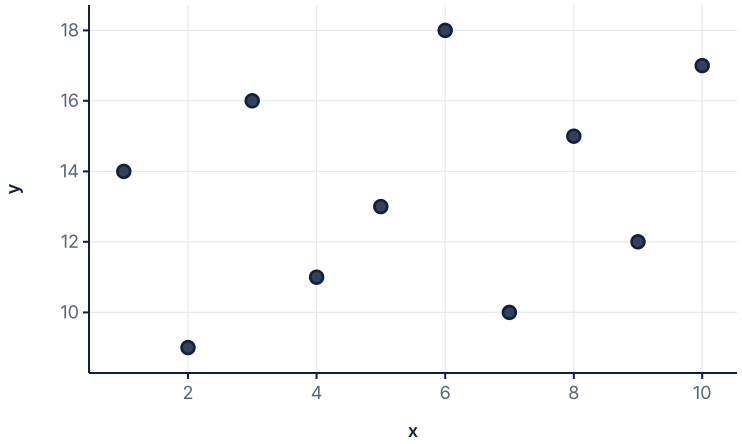
A perfect parabola: r = 0.00, yet y is completely determined by x. Correlation measures linear association only.

The parabola shows why you must always look at the scatterplot. The relationship is perfect but not linear, so r is exactly 0. The other danger is a lone outlier: moving one point can change r from near 1 to a modest value.

Worked Examples

EXAMPLE 1: MATCHING R TO A PLOT

Three fictional data sets each have 10 points. Set A has no clear pattern, Set B falls steadily with small scatter, Set C rises steadily with small scatter. Their correlations are 0.27, -0.96 and 1.00 in some order. Match each set to its r.



SOLUTION

Set C (strong increase) must have the large positive value, 1.00. Set B (strong decrease) must have the negative value, -0.96. Set A (no pattern) has the value near 0, 0.27.

The sign matches the direction and the size of |r| matches how tightly the points cluster. The plot shown is Set A, whose points wander with no trend.

EXAMPLE 2: COMPUTING R FROM Z-SCORES

Find r for the points (1,3), (2,5), (3,4), (4,8), (5,10) using the z-score products in the table above.

SOLUTION

The sum of the products of z-scores is 3.688.

$r = \frac{3.688}{5-1} = 0.922$.

Interpretation: there is a strong, positive, linear association between x and y. A graphing calculator's linear regression gives the same value and is the efficient approach on the exam.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: CHANGING UNITS

The correlation between hours studied and quiz score is r = 0.991. If study time is converted to minutes, what is the new correlation? What if we swap which variable is plotted on the x axis?

SOLUTION

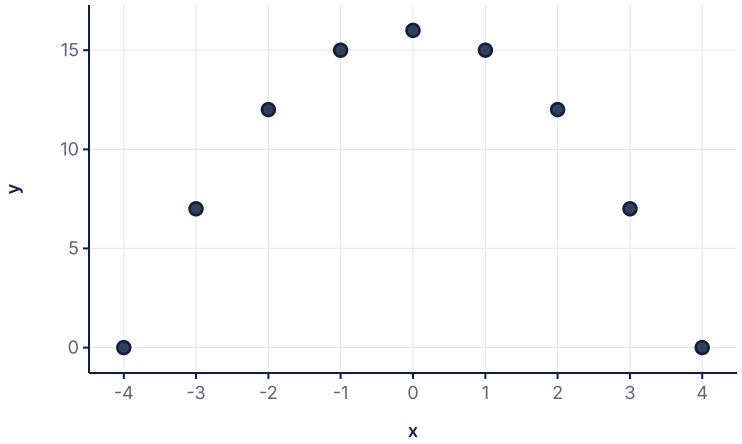
Multiplying every x value by 60 does not change the standardized values, so r is unchanged: 0.991 (computed from minutes: 0.991).

Swapping x and y leaves the products of z-scores the same, so r is still 0.991.

Neither change affects r. (A change of units would change the slope of a regression line, which is why the slope has units but r does not.)

EXAMPLE 4: R EQUALS ZERO

A fictional kite's height y (meters) is exactly 16 - x^2 at horizontal distances x = -4, -3, ..., 4 meters. Compute r and interpret it.



SOLUTION

The scatterplot is a perfect parabola opening downward.

Correlation: r = 0.00. The mean of z-score products is 0 because the left side of the curve (positive slope) cancels the right side (negative slope).

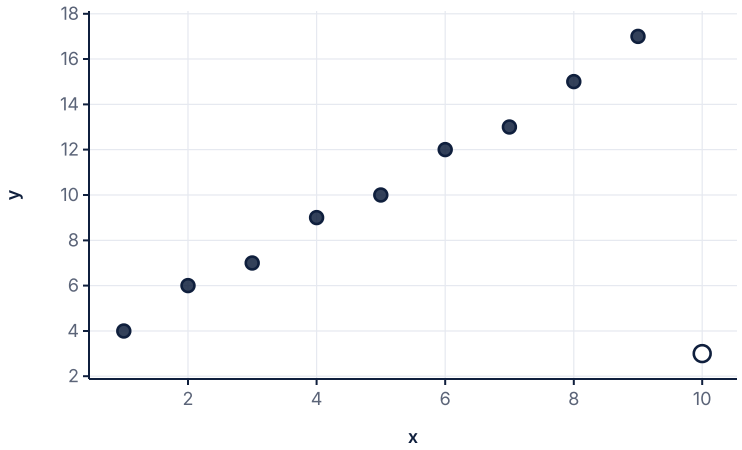
Interpretation: there is no LINEAR association, but there is a perfect curved association. r = 0 does not mean the variables are unrelated.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: ONE OUTLIER

Ten points lie close to a rising line and give r = 0.998. The last point (10, 18) is moved to (10, 3). What happens to r?



SOLUTION

Before: r = 0.998. After the move: r = 0.479.

The single point, far from the pattern and at the end of the x range, drops r from near 1 to a modest value. Correlation is not resistant to outliers.

Always check the scatterplot for outliers before relying on r.

EXAMPLE 6: RESTRICTED RANGE AND INTERPRETATION

In the 12-student study-time data, r = 0.991. If we look only at students who studied 3.5 hours or more (6 students), r = 0.969. Explain the change, and interpret the first r.

SOLUTION

With the full range of study times the line explains most of the variation. In the restricted range, the vertical scatter is larger compared with the spread of x, so r drops (the correlation depends on the range of x values included).

Interpretation of r = 0.991: there is a strong, positive, linear association between study time and quiz score in this sample.

Do not say "99% of the scores are explained" or "studying causes the score": r alone describes direction and strength of linear association only. (The proportion of variation explained by the line is r^2, Lesson 3.5.)

Write It Like a 5

Model Response

Prompt: A fictional scatterplot shows daily hours of sunshine and a town's ice cream sales for 40 days. The correlation is r = 0.91. A student writes: "Because r = 0.91, sunshine causes 91% of ice cream sales." Evaluate the statement. Response: The student misreads r. The value r = 0.91 indicates a strong, positive, linear association between hours of sunshine and sales on these days: days with more sunshine tended to have higher sales. It is not a percentage of sales caused by sunshine. Also, correlation does not prove causation; these are observational data and a variable such as temperature could be related to both sunshine and sales. Finally, r is only meaningful if the scatterplot is roughly linear and free of outliers, which should be checked first.

What the grader looks for

(1) Interprets r as direction and strength of linear association, in context. (2) Rejects the percentage interpretation. (3) States correlation does not imply causation, ideally with a confounder. (4) Mentions linearity or outliers as a condition for using r.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Coach's Playbook

1. **Plot first.** Confirm the pattern is roughly linear and look for outliers.
2. **Compute r with technology.** Use the built-in calculator's linear regression to display r; make sure diagnostics are on.
3. **Interpret with a sentence template.** "There is a [strong/moderate/weak], [positive/negative], linear association between [x in context] and [y in context]."
4. **Test your answer against the picture.** Sign matches slope direction; |r| near 1 matches tight points.
5. **Add a caution.** Mention outliers, nonlinearity, or causation as appropriate.

- r has no units, and the order of x and y does not matter.
- A 'good' r depends on the context: 0.6 may be exciting in social data and disappointing in a lab measurement.
- If every point lies on a horizontal line, r is undefined because the standard deviation of y is zero.

Common traps

(1) Treating r as a percent or as the proportion of variation explained (that is r^2). (2) Saying r = 0 means no relationship. (3) Using r for a curved pattern. (4) Believing a strong correlation shows cause and effect.

Mistake Clinic

Spot the error

Student: "The correlation between age and price of used cars is -0.97, so price is not related to age because r is negative." The sign of r gives the direction only; a value near -1 is a very strong negative linear association. Correct: older cars tend to cost less, and the association is strong.

Spot the error

Student: "I changed the heights from inches to centimeters, so r must change." Correlation is unitless and unchanged by linear changes of scale. Correct: r stays the same; only the slope and intercept of a regression line would change.

Spot the error

Student: "The scatterplot looks like a U shape, but r = 0.02 so there is no association." r only measures linear association. Correct: there is a strong curved (nonlinear) association; r is not appropriate to summarize it.

Real-World Connection

Weather analysts, school counselors and quality inspectors report correlations between measurements. A responsible report pairs the number with the scatterplot, because r can hide curves, clusters and outliers. When groups' averages are correlated (for example average school scores and average family income), the correlation among individuals is usually weaker than the correlation among the group averages.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

A fictional data set of 8 individual students has r = 0.63 between hours of practice and performance rating. When the students are grouped into 4 pairs (averaging x and y within each pair of neighbors in the list), the correlation of the 4 group means is higher. Explain why in general, without computing. Solution: Averaging smooths out individual variation. The vertical scatter around the pattern shrinks relative to the spread of the means along the pattern, so the correlation of averages is typically stronger than the correlation of individuals. Using group-level correlation to describe individuals overstates the strength of the individual relationship (the ecological fallacy). To check, compute r for the 8 points (0.63) and for the means of the neighboring pairs; the second is larger.

Practice

PRACTICE **Correlation**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which value is NOT possible for a correlation coefficient?
   (A) 1.4 **(B)** -0.62 **(C)** 0 **(D)** 1
2. A scatterplot shows a tight, downward-sloping linear pattern. Which r fits best?
   (A) -0.15 **(B)** -0.94 **(C)** 0.94 **(D)** -1.40
3. Heights (inches) and weights (pounds) of students have r = 0.62. If heights are converted to centimeters, the correlation is
   (A) 0.62 × 2.54 = 1.57 **(B)** 0.62 / 2.54 = 0.24 **(C)** 0.62
   (D) not defined
4. The correlation between x and y is 0.30. The correlation between y and x is
   (A) -0.30 **(B)** 0.70 **(C)** 1.00 - 0.30
   (D) 0.30
5. A scatterplot shows a perfect parabola, and the calculator gives r = 0. The correct interpretation is
   (A) there is no linear association, but there may be a strong nonlinear one
   (B) x and y are unrelated
   (C) the calculator made an error
   (D) the association is strong and linear
6. A newspaper says that r = 0.80 between daily exercise and a wellness score. The best interpretation is
   (A) 80 percent of people who exercise have high wellness scores
   (B) exercise causes 80 percent of the wellness score
   (C) exercise explains 80 percent of the variation in wellness
   (D) a strong positive linear association between exercise and wellness score
7. Adding one point far from the pattern at the right end of a scatterplot with r = 0.95 will most likely
   (A) leave r unchanged because r is resistant **(B)** decrease r noticeably
   (C) increase r **(D)** change r to exactly 0
8. Which statement about r = -0.7 and r = 0.7 is correct?
   (A) -0.7 is a stronger relationship
   (B) 0.7 is a stronger relationship
   (C) They indicate the same strength but opposite directions
   (D) -0.7 is impossible because it is smaller than 0
9. The correlation for the data (1,3),(2,5),(3,4),(4,8),(5,10) is r = 0.922. If every y value is increased by 10, what is the new r?
   (A) 0.922 **(B)** 0 **(C)** 9.22 **(D)** 10.922
10. A fictional teacher records (hours of sleep, quiz score) for 6 students: (5, 62), (6, 70), (6.5, 72), (7, 80), (8, 85), (9, 88).
    (a) Make a sketch or describe the scatterplot.
    (b) Compute r.
    (c) Interpret r in context.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

11. The correlation between the number of apps on a phone and its battery life (hours) is r = -0.45 for 200 fictional phones.
    (a) Describe the association in words.
    (b) Would converting battery life to minutes change r? Explain.
    (c) Can we conclude installing apps shortens battery life? Why or why not?
12. Points (1,3),(2,5),(3,4),(4,8),(5,10) have z-score products 1.30, 0.22, 0.00, 0.43, 1.74 (in order, rounded).
    (a) Which points contribute most to a positive r?
    (b) Compute r from the products.
    (c) If the last point were changed to (5, 2), would r increase or decrease? Compute the new r.
13. A scatterplot of a fictional ball's height y at times x = 0, 1, 2, 3, 4, 5, 6 seconds is parabolic and symmetric, with r about 0.
    (a) Does r = 0 mean the ball's height is unrelated to time?
    (b) What type of association is present?
    (c) What should be used instead of r to describe it?
14. Two groups of data: Group P has 30 individual students with r = 0.45 between hours studied and score. A second analysis uses the average hours and average score of 6 classes and finds r = 0.85.
    (a) Why is the correlation of averages larger?
    (b) What error would a reader make by applying the 0.85 to individual students?
15. For the study-time data (12 students), r = 0.991. Another fictional student studied 5.0 hours and scored 45; with this student r = 0.664.
    (a) Explain the change in r.
    (b) Is r resistant? Justify.
    (c) What should be reported?

3.4 Least-Squares Regression Lines

Compute the least-squares line from data or summary statistics, interpret slope and intercept in context, and make predictions while avoiding extrapolation.

Why This Matters

When a scatterplot is roughly linear, we summarize it with a line and use the line to predict. The **least-squares regression line** (LSRL) is the line chosen by a definite rule, not by eye. The big skills are computing the slope and intercept, interpreting them in context (a favorite exam point), and knowing when a prediction is safe. Students often interpret the slope as a total change instead of a *predicted change per unit*, forget the word "predicted", or predict far outside the data.

Regression is the heart of Exam Unit 5. Free-response questions routinely give a data table or computer output and ask for the equation, an interpretation of the slope, and a prediction with a comment on extrapolation.

Core Ideas

The line $\hat{y} = a + bx$ predicts the response y from the explanatory variable x. The symbol $\hat{y}$ ("y-hat") is a *predicted* value, not an observed one. A **residual** is the observed value minus the predicted value: $y - \hat{y}$. The least-squares line makes the sum of the squared residuals as small as possible.

Least-Squares Regression Line

Slope:

$$b = r\,\frac{s_y}{s_x}$$

Intercept:

$$a = \bar{y} - b\,\bar{x}$$

The line always passes through the point $(\bar{x}, \bar{y})$.
Equation: $\hat{y} = a + bx$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

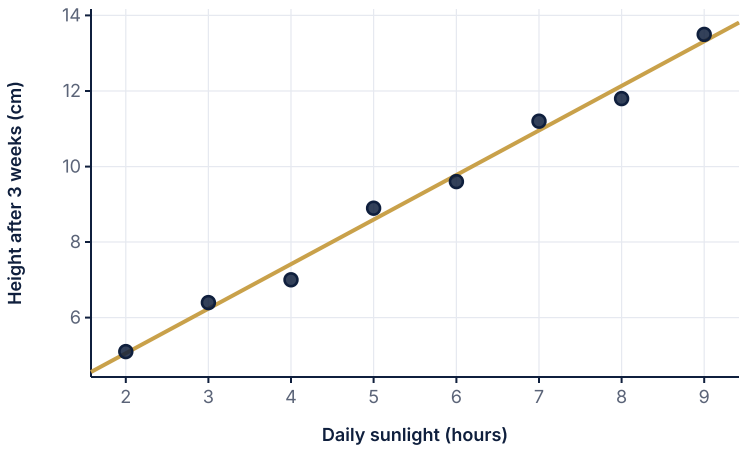
Interpreting the Slope and Intercept

Slope: "For each additional [unit of x], the predicted [y in context] increases (decreases) by [|b|] [units of y]."

Intercept: "When [x in context] is 0, the predicted [y in context] is [a] [units]." The intercept may have no practical meaning if x = 0 is outside the data or impossible.

Always say **predicted** (or "on average").

Here are 8 fictional seedlings: hours of daily sunlight (x) and height after three weeks in cm (y). The summary statistics are $\bar{x} = 5.500$, $s_x = 2.449$, $\bar{y} = 9.188$, $s_y = 2.903$, and $r = 0.995$. Then $b = rs_y/s_x = 1.180$ and $a = \bar{y} - b\bar{x} = 2.699$, so the line is $\hat{y} = 2.699 + 1.180x$.



Seedling data with the least-squares line; it passes through (mean of x, mean of y).

Prediction and Extrapolation

Interpolation: predicting for an x value inside the range of the data. Usually reasonable if the pattern is linear.

Extrapolation: predicting outside the range of the x values. Risky, because the linear pattern may not continue (a seedling cannot grow forever with more light).

A prediction is not an observation: actual values vary around the line.

Reading computer output. Regression output lists the intercept (labeled "Constant") and slope (labeled by the explanatory variable) in the Coef column. The line is read straight from that column. The output also gives S and R^2, which we study in Lesson 3.5.

Why squares? Residuals above the line are positive and below are negative, so they cancel in a plain sum. Squaring makes every miss count and penalizes big misses more. A consequence is that the LSRL is pulled toward points far from the pattern, a theme we return to in Lesson 3.5. Also notice that the slope formula links two ideas: the slope is the correlation rescaled by the ratio of spreads, so a one-standard-deviation increase in x goes with an increase of only r standard deviations in the predicted y.

Worked Examples

EXAMPLE 1: FROM SUMMARY STATISTICS

Fictional data on weekly practice minutes (x) and a performance score (y) have $\bar{x} = 62$, $s_x = 12$, $\bar{y} = 78$, $s_y = 9$, and $r = 0.8$. Find the LSRL.

SOLUTION

Slope: $b = 0.8 \cdot 9/12 = 0.6$.

Intercept: $a = 78 - 0.6(62) = 40.8$.

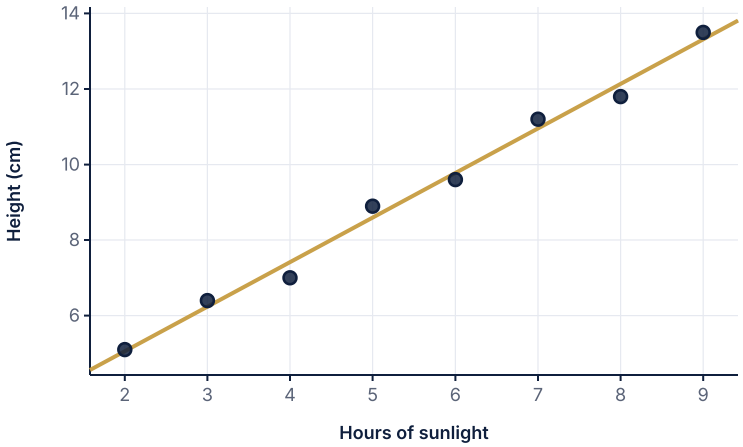
Line: $\hat{y} = 40.8 + 0.6x$. Check: at $x = 62$, $\hat{y} = 40.8 + 37.2 = 78 = \bar{y}$, so the line passes through $(\bar{x}, \bar{y})$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: INTERPRETING SLOPE AND INTERCEPT

For the seedling line $\hat{y} = 2.70 + 1.18x$, interpret the slope and the intercept.



SOLUTION

Slope: for each additional hour of daily sunlight, the predicted height after three weeks increases by about 1.18 cm.

Intercept: with 0 hours of daily sunlight, the predicted height is about 2.70 cm. Since the data have x from 2 to 9 hours, 0 hours is extrapolation, so this value has little practical meaning (though a seedling in the dark still has some height).

Both interpretations use 'predicted' and the units, and mention context.

EXAMPLE 3: PREDICTION AND EXTRAPOLATION

Use the seedling line to predict the height for (a) 6.5 hours of sunlight and (b) 20 hours. Comment on each.

SOLUTION

(a) $\hat{y} = 2.699 + 1.180(6.5) = 10.4$ cm. Since 6.5 is within the data range (2 to 9), this is interpolation and is reasonable.

(b) $\hat{y} = 2.699 + 1.180(20) = 26.3$ cm. But 20 hours of daily sunlight is impossible (a day has only 24 hours, and the data stop at 9), so this is extrapolation and the prediction is not trustworthy.

EXAMPLE 4: COMPUTER OUTPUT

Regression output for the seedling data is shown. Write the equation and predict the height for 4 hours of sunlight.

Predictor	Coef	Description
Constant	2.699	intercept a
Sunlight	1.180	slope b
S	0.298	typical residual size
R-sq	99.1%	percent of variation explained

SOLUTION

The Coef column gives intercept 2.699 and slope 1.180, so the line is $\hat{y} = 2.699 + 1.180x$.

At x = 4: $\hat{y} = 2.699 + 1.180(4) = 7.42$ cm.

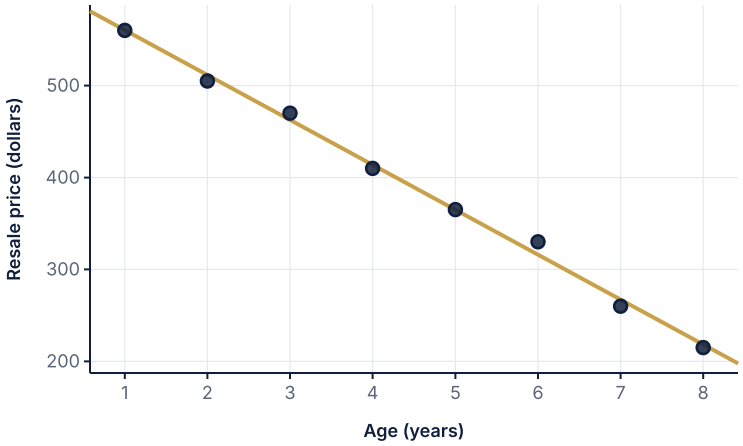
Remember that this is a predicted height, not a guarantee.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: FITTING A LINE TO DATA

A fictional tablet's resale price (dollars) is recorded by age (years): (1,560), (2,505), (3,470), (4,410), (5,365), (6,330), (7,260), (8,215). Find the LSRL and interpret the slope.



SOLUTION

Using the built-in calculator's linear regression: $\hat{y} = 609.3 - 48.9x$, with $r = -0.998$.

Slope: for each additional year of age, the predicted resale price decreases by about 48.9 dollars.

Intercept: a new (age 0) tablet is predicted to cost about 609 dollars; age 0 is just outside the data, so treat this loosely. Check that the line passes through the means: $\bar{x} = 4.5$, $\bar{y} = 389.4$.

EXAMPLE 6: REGRESSING X ON Y

Using the summary statistics of Example 1 ($\bar{x} = 62$, $s_x = 12$, $\bar{y} = 78$, $s_y = 9$, $r = 0.8$), find the LSRL for predicting practice minutes from score. Is it the algebraic inverse of $\hat{y} = 40.8 + 0.6x$?

SOLUTION

Now the response is x: $b = rs_x/s_y = 0.8(12)/9 = 1.067$.

Intercept: $a = 62 - 1.067(78) = -21.20$. Line: $\hat{x} = -21.20 + 1.067y$.

Solving the first line for x would give $x = (y - 40.8)/0.6$ with slope $1/0.6 = 1.667$, not 1.067. Because $|r| < 1$, the two regression lines differ: the LSRL depends on which variable is the response, so choose the roles carefully.

Write It Like a 5

Model Response

Prompt: A fictional coach records weekly training hours (x) and 3K run time in minutes (y) for 20 runners and finds $\hat{y} = 17.2 - 0.45x$ for training hours between 2 and 12. (a) Interpret the slope. (b) Predict the time for a runner who trains 8 hours. (c) Should we use the line for a runner who trains 30 hours? Response: (a) For each additional hour of weekly training, the predicted 3K time decreases by 0.45 minutes. (b) $\hat{y} = 17.2 - 0.45(8) = 13.6$ minutes, which is interpolation because 8 is between 2 and 12. (c) No. 30 hours is far outside the data range (extrapolation), and the linear trend cannot continue because the predicted time would be $17.2 - 0.45(30) = 3.7$ minutes, an unrealistic value for a 3K run.

What the grader looks for

(1) Slope interpretation with 'predicted', 'each additional hour' and units. (2) Correct substitution and result with units. (3) Identifies the prediction as interpolation. (4) Rejects the extrapolation with a reason tied to the context.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. **Define the variables.** Write y-hat in words if the problem allows it.
2. **Get the line.** From summary statistics use b = r sy/sx and a = ybar - b xbar; from data use the calculator's LinReg.
3. **Check with the means.** Plug xbar into the line; you should get ybar.
4. **Interpret with a template.** "For each additional ___ the predicted ___ increases/decreases by ___."
5. **Check the x value before predicting.** Inside the data range or extrapolating?

- Turn diagnostics on in the built-in calculator so r appears with the line.
- The slope has the same sign as r.
- If x is measured in different units, the slope changes (a unit rescale) but r does not.

Common traps

(1) Leaving out 'predicted' when interpreting. (2) Swapping the roles of x and y in the formula for b. (3) Interpreting the intercept when x = 0 is impossible or outside the data. (4) Predicting far outside the data range without a warning.

Mistake Clinic

Spot the error

Student: "The slope is 0.45, so the runner's time decreases by 0.45 minutes." Where does it fail? There is no 'per unit' and no 'predicted'. Correct: for each additional hour of weekly training, the predicted time decreases by 0.45 minutes (on average).

Spot the error

Student computes the slope as b = r sx/sy using x as the explanatory variable. The standard deviations are inverted. Correct: b = r (sy/sx), response over explanatory. With r = 0.8, sx = 12, sy = 9: b = 0.8(9/12) = 0.6, not 1.07.

Spot the error

Student: "For 20 hours of sunlight the seedling will be 36 cm tall, because the line says so." The data extend only to 9 hours and 20 hours is impossible in a day. Correct: this is extrapolation, so the prediction is unreliable.

Real-World Connection

Insurers, schools and sports analysts use regression lines to predict. A fair report always gives the data range, the typical size of prediction errors (Lesson 3.5), and a warning about extrapolation. Dashboards that project trends years into the future often ignore that last point.

Challenge Corner

A fictional data set has $s_x = 4$, $s_y = 10$, and the LSRL is $\hat{y} = 3 + 2.5x$. (a) Find r. (b) If $\bar{x} = 6$, find $\bar{y}$. (c) What is the slope when predicting x from y? Solution: (a) $b = rs_y/s_x$ gives $2.5 = r(10/4)$, so $r = 1$. (b) The line passes through the means: $\bar{y} = 3 + 2.5(6) = 18$. (c) With $r = 1$ the points lie exactly on the line; the reverse slope is $rs_x/s_y = 0.4 = 1/2.5$, which equals the algebraic inverse. This is the only case in which the two regression lines coincide.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE **Least-Squares Regression Lines**

Show your work. A calculator is allowed. Justify conclusions in context.

1. The least-squares regression line is the line that
   (A) minimizes the sum of the residuals
   (B) passes through the first and last data points
   (C) minimizes the sum of the squared residuals
   (D) has a slope equal to r
2. The summary statistics are xbar = 20, sx = 5, ybar = 50, sy = 15, r = 0.6. What is the slope of the LSRL?
   (A) 1.8 **(B)** 0.2 **(C)** 0.6 **(D)** 2.0
3. For the data in the previous question, the intercept is
   (A) 38 **(B)** -14 **(C)** 50 **(D)** 14
4. The LSRL for predicting a tablet's price (dollars) from age (years) is price-hat = 612 - 48 age. Which statement correctly interprets the slope?
   (A) Each tablet loses exactly 48 dollars in value per year
   (B) For each additional year of age, the predicted price decreases by 48 dollars
   (C) A new tablet costs 48 dollars
   (D) The price drops by 612 dollars over its lifetime
5. A regression line fit to data with x between 10 and 40 is used to predict y for x = 100. This prediction is
   (A) reliable if r is close to 1 **(B)** reliable because the line is least-squares
   (C) unreliable because it is extrapolation **(D)** impossible to make
6. The LSRL always passes through which point?
   (A) (0, 0) **(B)** (median of x, median of y)
   (C) (smallest x, smallest y) **(D)** (mean of x, mean of y)
7. Computer output gives Constant coefficient 3.2 and Hours coefficient 1.7 for predicting score from hours of practice. The predicted score for 5 hours is
   (A) 8.5 **(B)** 4.9 **(C)** 11.7 **(D)** 19.3
8. The sign of the slope of the LSRL is always the same as the sign of
   (A) the intercept **(B)** the correlation r **(C)** the mean of x
   (D) the residuals
9. If r = 0.5 and sy = sx, the slope of the LSRL for predicting y from x is
   (A) 0.5 **(B)** 1.0 **(C)** 0.25 **(D)** 2.0
10. For 6 fictional bike riders, x = ride length (km) and y = water used (hundreds of mL): (5,12), (10,19), (15,24), (20,33), (25,38), (30,45).
    (a) Find the LSRL.
    (b) Interpret the slope.
    (c) Predict the water use for a 17 km ride, and say whether it is interpolation.
11. Fictional data: xbar = 12, sx = 3, ybar = 80, sy = 6, r = -0.75.
    (a) Find the equation of the LSRL.
    (b) Predict y when x = 15.
    (c) Interpret the slope in the context of x = hours of video games per week and y = exam score.
12. A line y-hat = 25 + 4x predicts a plant's height (cm) from days of growth, using data for x from 5 to 20 days.
    (a) Predict the height at 12 days.
    (b) Predict the height at 100 days and critique it.
    (c) Interpret the intercept and say whether it is meaningful.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. A student states: 'The regression line for predicting y from x is y-hat = 3 + 2x. Therefore the line for predicting x from y is x-hat = -1.5 + 0.5y.'
    (a) Explain why this reasoning is faulty unless r has a special value.
    (b) Given sx = 2, sy = 5 and slope 2, find r and the correct slope for predicting x from y.
14. Seedling data: hours of sunlight x = 2,...,9 and heights 5.1, 6.4, 7.0, 8.9, 9.6, 11.2, 11.8, 13.5 cm have LSRL $\hat{y} = 2.699 + 1.180x$.
    (a) Show that the line passes through (xbar, ybar).
    (b) What is the predicted height for 8 hours, and the residual for the 8-hour seedling (observed 11.8 cm)?
    (c) Interpret that residual.
15. A coach's data give sx = 1.5 (training hours), sy = 2.4 (minutes), and the LSRL for minutes on hours has slope -0.9.
    (a) Find r.
    (b) If the times were recorded in seconds instead of minutes, what would the slope be?
    (c) What would r be?

3.5 Residuals, r-squared, and Standard Deviation of Residuals

Use residuals and residual plots to judge a linear model, interpret r-squared and the standard deviation of the residuals, and identify outliers, high-leverage points and influential points.

Why This Matters

A line is only useful if it fits well enough for the job. This lesson gives you the tools to judge a fit: **residuals** show how far each prediction misses, a **residual plot** reveals whether a line was the right kind of model at all, r^2 says how much of the variation the line accounts for, and s says how large a typical miss is. You will also learn which unusual points control the line.

Regression analysis (Exam Unit 5) tests these ideas heavily. Typical prompts ask you to compute a residual and interpret its sign, to read a residual plot and decide whether a linear model is appropriate, and to interpret r^2 and s in context. The standard error of the residuals s and r^2 are also shown in computer output, so you need to know where to find them.

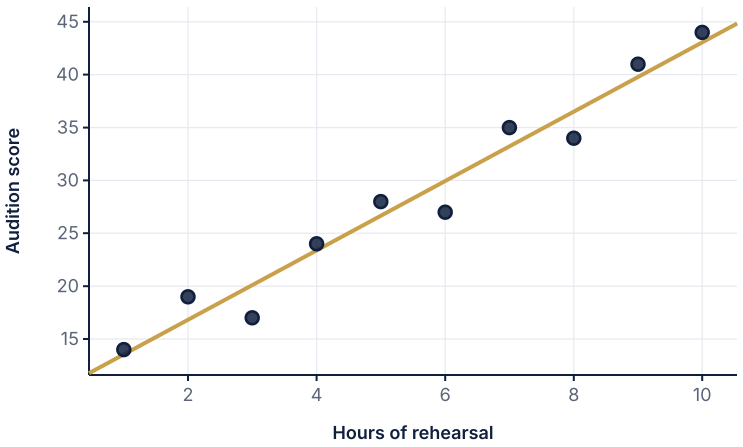
Core Ideas

Residual

$$\text{residual} = y - \hat{y} = \text{observed} - \text{predicted}$$

A positive residual means the point lies above the line (the line underpredicts); negative means below (the line overpredicts). The residuals of a least-squares line add to 0.

The next data are fictional: hours of rehearsal (x) and an audition score (y) for 10 musicians. The LSRL is $\hat{y} = 10.27 + 3.28x$. The residual for the musician with $x = 3$ and $y = 17$ is $17 - (10.27 + 3.28 \cdot 3) = -3.10$: the line overpredicts that musician's score by 3.10 points.



Rehearsal hours and audition score with the least-squares line.

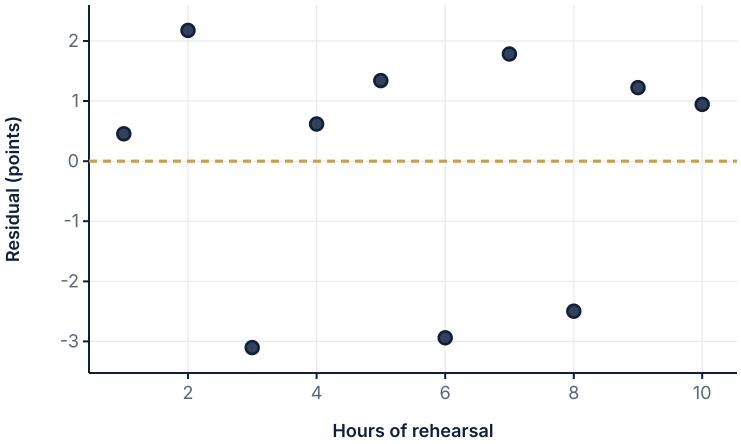
Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

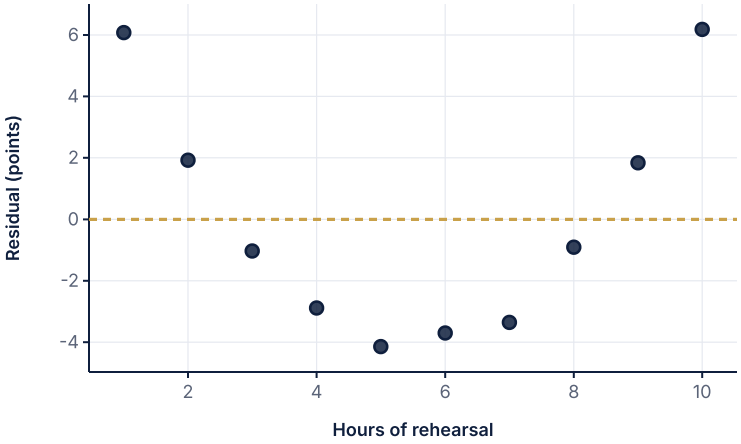
Residual Plots

A **residual plot** graphs residuals against x (or against predicted values). If a linear model is appropriate, the plot shows a random scatter with no pattern around the horizontal line at 0 and roughly constant vertical spread.

A curved pattern (for example, a U shape) means the relationship is not linear. A fan shape (spread that grows or shrinks) means the prediction errors are not constant.



This residual plot (rehearsal data) shows no clear pattern, so a linear model is appropriate. Contrast it with the next plot, from a fictional curved data set where a line was forced onto the points.



The U shape is a signal that the line is the wrong model. The correlation for the curved data is still high ($r = 0.974$), which shows why r alone cannot certify linearity.

r-squared and the Standard Deviation of the Residuals

Coefficient of determination: r^2 is the proportion of the variation in y that is accounted for by the linear relationship with x. In context: "About [r squared as percent] of the variation in [y] is explained by the linear relationship with [x]."

Standard deviation of the residuals:

$$s = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}}$$

It is the typical size of a prediction error, in the units of y: "The actual [y] typically differs from the value predicted by the line by about [s] [units]."

$r^2 = 1 - \dfrac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}$.

For the rehearsal data $r^2 = 0.960$ and $s = 2.16$ points. About 96.0% of the variation in audition scores is explained by the linear relationship with rehearsal hours, and the actual scores typically differ from the predicted scores by about 2.16 points.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unusual Points in Regression

Outlier: a point with a large residual (far from the line in the y direction).

High-leverage point: a point with an x value far from the mean of x. It has the potential to move the line.

Influential point: a point whose removal would substantially change the slope, intercept, or correlation. Influential points are usually outliers with high leverage (far in x and off the trend).

Worked Examples

EXAMPLE 1: COMPUTING AND INTERPRETING A RESIDUAL

Using the rehearsal line $\hat{y} = 10.27 + 3.28x$, find the residual for the musician with $x = 7$ and $y = 35$ and interpret it.

SOLUTION

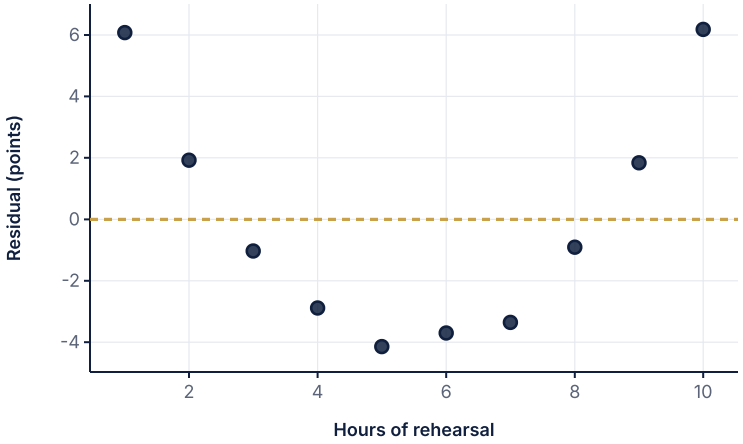
Predicted: $\hat{y} = 10.267 + 3.279(7) = 33.22$.

Residual: $35 - 33.22 = 1.78$.

Interpretation: this musician's score is 1.78 points above the score predicted by the line for 7 hours of rehearsal.

EXAMPLE 2: READING A RESIDUAL PLOT

The residual plots shown belong to the rehearsal data and the curved data. For each, is a linear model appropriate?



SOLUTION

Rehearsal data (random scatter around 0, no pattern, similar spread): a linear model is appropriate.

Curved data (negative residuals in the middle, positive at the ends): there is a clear curved pattern, so a linear model is not appropriate. Consider a nonlinear model or a transformation (Lesson 3.6).

Always argue from the residual plot, not from r.

EXAMPLE 3: INTERPRETING R-SQUARED

A fictional study finds r = 0.80 between the daily minutes a student practices typing and typing speed. Find r-squared and interpret it.

SOLUTION

$r^2 = 0.80^2 = 0.64$.

About 64% of the variation in typing speed is explained by the linear relationship with practice minutes. The remaining 36% is due to other factors and variation not explained by the line.

It does not mean that 64% of students improve, nor that the correlation is 64%.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 4: INTERPRETING S

Regression output for the rehearsal data gives S = 2.156 and R-sq = 96.0%. Interpret both values.

Term	Coef
Constant	10.267
Hours	3.279
S	2.156
R-sq	96.0%

SOLUTION

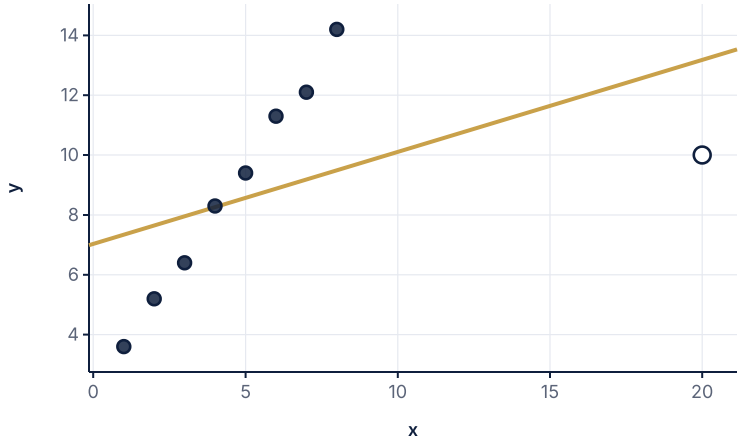
S: the actual audition scores typically differ from the scores predicted by the least-squares line by about 2.16 points.

R-sq: about 96.0% of the variation in audition scores is explained by the linear relationship with hours of rehearsal.

Together they say the line predicts well: a typical miss of about 2 points on a scale where scores range from 14 to 44.

EXAMPLE 5: OUTLIER, LEVERAGE, INFLUENCE

Eight fictional points follow the line $\hat{y} = 2.14 + 1.48x$ closely. Three different points are added one at a time: P1 = (4.5, 20), P2 = (20, 10), P3 = (20, 32). Classify each.



SOLUTION

P1 = (4.5, 20): x is near the mean of x (4.5) but y is far above the line. It is an outlier with LOW leverage. The slope stays 1.48 (the line shifts up), and r falls to 0.67.

P2 = (20, 10): x is far from the others and the point is far below the trend. High leverage and an outlier, so it is INFLUENTIAL: the slope falls from 1.48 to 0.31 and r from 1.00 to 0.51.

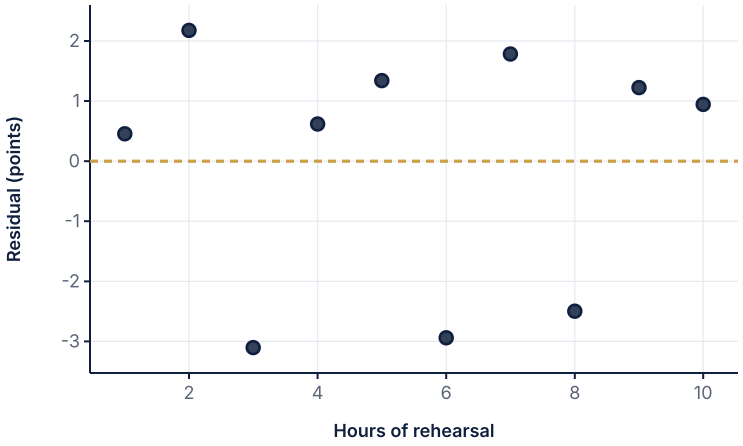
P3 = (20, 32): x is far away, so it has high leverage, but it lies on the trend. It is NOT influential; the slope stays 1.49 and r increases to 1.00.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 6: FULL ASSESSMENT OF A LINE

For the rehearsal data, decide whether the linear model is appropriate and summarize how well it predicts.



SOLUTION

Condition check: the scatterplot is roughly linear; the residual plot shows random scatter with no pattern or fan shape; no point is far from the others in x.

Strength: $r^2 = 0.960$ (about 96.0% of the variation is explained) and $s = 2.16$ points.

Conclusion: the line $\hat{y} = 10.27 + 3.28x$ is an appropriate and useful model for rehearsal hours between 1 and 10; predictions are typically off by about 2 points.

Write It Like a 5

Model Response

Prompt: A fictional analyst fits a line to predict a battery's remaining charge (percent) from hours of use. The residual plot shows a clear U shape, r-squared is 0.94 and s = 3.1. Is a linear model appropriate? Explain. Response: No. Although r-squared is high (94% of the variation in remaining charge is explained by the linear relationship with hours of use) and a typical prediction error is about 3.1 percentage points, the residual plot shows a U-shaped pattern. A random scatter is required for a linear model to be appropriate; the curved pattern means the line underpredicts at some values of hours and overpredicts at others, so the relationship is not linear and a different model (for example, a transformed or curved model) is better.

What the grader looks for

(1) Concludes that the linear model is not appropriate. (2) Cites the pattern in the residual plot as the reason. (3) Correctly interprets r-squared in context (not as a verdict on linearity). (4) Interprets s in context, or notes that high r-squared does not cancel curvature.

Coach's Playbook

1. **Compute residuals as observed minus predicted.** Positive means the point is above the line.
2. **Plot residuals against x.** On the built-in calculator, store residuals during regression and plot them against the x list.
3. **Look for patterns, not just size.** Curve means wrong model, fan means unequal spread, random scatter means line is fine.
4. **Interpret r-squared and s in context,** with the words 'variation explained' and 'typical prediction error.'

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5 **Check for unusual points.** Ask: is it far in y (outlier), far in x (leverage), and does the line change if it is removed (influence)?

- r-squared compares the line's errors with the errors from predicting the mean of y for everyone.
- Large leverage plus a point off the trend means a small number of points can control the whole line.
- Report results with and without an influential point; do not delete data just because it is inconvenient.

Common traps

(1) Computing the residual as predicted minus observed. (2) Calling r-squared 'the percent of points on the line.' (3) Judging linearity from r-squared instead of the residual plot. (4) Confusing a high-leverage point (far in x) with an outlier (far in y).

Mistake Clinic

Spot the error

Student: "A residual of -3 means the prediction was off by negative 3 points, so the line is bad." A negative residual only tells direction. Correct: the observed value is 3 below the predicted value; whether the fit is good depends on the size of residuals overall (for example s).

Spot the error

Student: "r-squared = 0.92, so 92% of the data points lie on the regression line." Correct: 92% of the variation in y is explained by the linear relationship with x; very few points lie exactly on the line.

Spot the error

Student: "The point (20, 32) is an outlier because x = 20 is far from the other x values, so it changes the line." Correct: it has high leverage, but because it lies on the trend, removing it leaves the slope essentially unchanged (1.49 versus 1.48), so it is not influential and not an outlier in the residual sense.

Real-World Connection

Engineers examine residual plots before trusting a calibration line, and economists look for fan shapes in the residuals of a forecast model. In sports analytics, a few extreme performances with unusual x values can distort a prediction line, so analysts check leverage and influence.

Challenge Corner

Using the rehearsal data, the sum of squared residuals is SSE = 37.19 and the total variation is SST = 924.1. (a) Show that r-squared = 1 - SSE/SST. (b) Show that s = 2.16. Solution: (a) $1 - 37.19/924.1 = 0.9598$, which equals r^2 from the correlation. (b) $s = \sqrt{37.19/8} = 2.156$ because there are $n - 2 = 8$ degrees of freedom. The two formulas agree with the calculator output.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE **Residuals, r-squared, and Standard Deviation of Residuals**

Show your work. A calculator is allowed. Justify conclusions in context.

1. The observed value is 42 and the line predicts 37. The residual is
   (A) 5 **(B)** -5 **(C)** 42 **(D)** 79
2. A residual plot for a fitted line shows a clear curve. The best conclusion is
   (A) a linear model is appropriate if r is large
   (B) the outliers should be removed
   (C) a linear model is not appropriate
   (D) the slope must be 0
3. The correlation is r = -0.9. The coefficient of determination is
   (A) -0.81 **(B)** 0.81 **(C)** 0.9 **(D)** 0.1
4. Which is the correct interpretation of r-squared = 0.72 for predicting weight from height?
   (A) About 72% of the variation in weight is explained by the linear relationship with height
   (B) 72% of people's weights are predicted exactly
   (C) The correlation between height and weight is 0.72
   (D) Height causes 72% of the weight
5. The standard deviation of the residuals s = 4.2 points for a line predicting exam score. The best interpretation is
   (A) the line is 4.2% accurate
   (B) the slope is 4.2
   (C) the average residual is 4.2
   (D) actual scores typically differ from predicted scores by about 4.2 points
6. A point has an x value far from the mean of x and lies far from the pattern of the other points. It is best described as
   (A) a high-leverage, influential point
   (B) a point with low leverage
   (C) a point that cannot change the slope
   (D) an average point
7. A point lies far above the line, but near the mean of x. It is most likely
   (A) influential on the slope
   (B) an outlier with low leverage that has little effect on the slope
   (C) a high-leverage point
   (D) not an outlier
8. The least-squares residuals
   (A) always sum to the largest residual
   (B) are all positive
   (C) always sum to 0
   (D) are all zero when r is high
9. Which residual plot pattern suggests that the variability of the prediction errors is not constant?
   (A) A random cloud with equal spread
   (B) Points all lying on the horizontal axis
   (C) A downward straight line
   (D) A fan shape, with spread increasing as x increases
10. The line $\hat{y} = 10.27 + 3.28x$ predicts audition score from rehearsal hours. A musician rehearsed 9 hours and scored 41.
    (a) Compute the residual.
    (b) Interpret it.
    (c) Why do residuals of the least-squares line sum to 0?
11. A regression of weight (kg) on height (cm) for fictional students gives r = 0.60, s = 6.5 kg.
    (a) Find and interpret r-squared.
    (b) Interpret s.
    (c) Would the model predict well? Justify.
12. Describe what each residual plot pattern tells you about a linear model: (a) random scatter around 0; (b) a U shape; (c) a fan shape that widens as x increases.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. Eight fictional points follow $\hat{y} = 2.14 + 1.48x$ (r = 0.998). Adding the point (20, 10) gives slope 0.31 and r = 0.51.
    (a) Is (20, 10) high leverage? Explain.
    (b) Is it influential? Explain.
    (c) Would you recommend deleting it?
14. For the rehearsal data, SSE = 37.19 and SST = 924.1 with n = 10.
    (a) Compute r-squared.
    (b) Compute s.
    (c) In one sentence, say what each value tells us.
15. A fictional scatterplot of fuel used (liters) against distance (km) has r = 0.98, but the residual plot shows a clear curve.
    (a) Explain how r can be so high with a poor linear model.
    (b) What would you do next?
    (c) What does a high r-squared fail to tell you?

3.6 Transforming Data to Achieve Linearity

Use logarithms to straighten exponential and power relationships, fit and check the model with residual plots, and back-transform to make predictions in the original units.

Beyond the 2027 Exam Core

College Board's revision page lists analyzing departures from linearity as removed from the 2027 exam. In this lesson, the use of residual plots to diagnose curvature and unequal spread, and the model comparisons in Example 6, are extension material; check your teacher's current course description. The core skills of transforming data to achieve linearity and back-transforming predictions are used throughout regression work.

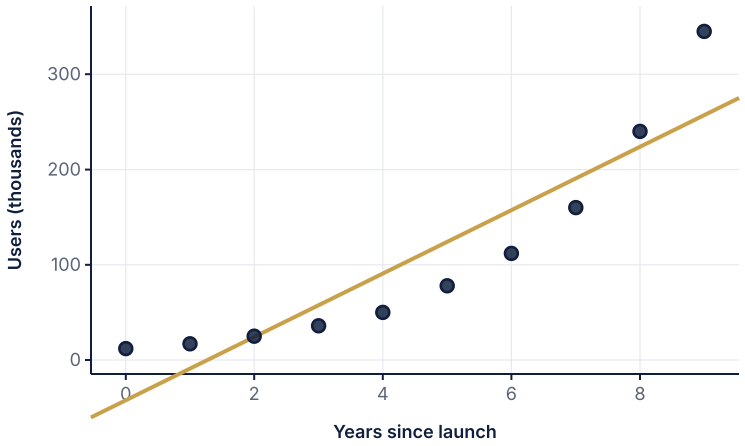
Why This Matters

Not every relationship is a straight line. Populations, prices with interest, and viral posts grow by a *constant factor* each period; areas, periods and drag forces follow *power laws*. When the scatterplot curves, you can often transform one or both variables with a logarithm so the new plot is linear, fit a least-squares line to the transformed data, and then undo the transformation to predict in the original units.

Students go wrong by fitting a line to the curve, by forgetting to back-transform (reporting a log value as if it were the prediction), and by interpreting the slope of a log model as an additive change. In Exam Unit 5 you may be asked to choose a transformation from a residual plot, interpret a model in transformed units, or compute a prediction with computer output.

Core Ideas

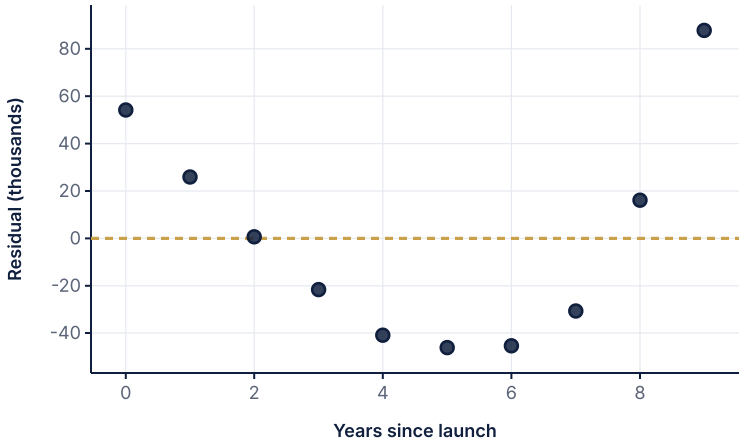
A fictional social app tracks users (thousands) by year since launch. The scatterplot of users against years curves upward, and the line fits poorly: the residual plot is a clear U shape.



Users grow faster each year: a curved pattern, with a poor linear fit.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Residuals from the straight-line fit: a U shape, so a linear model is not appropriate.

Choosing a Transformation

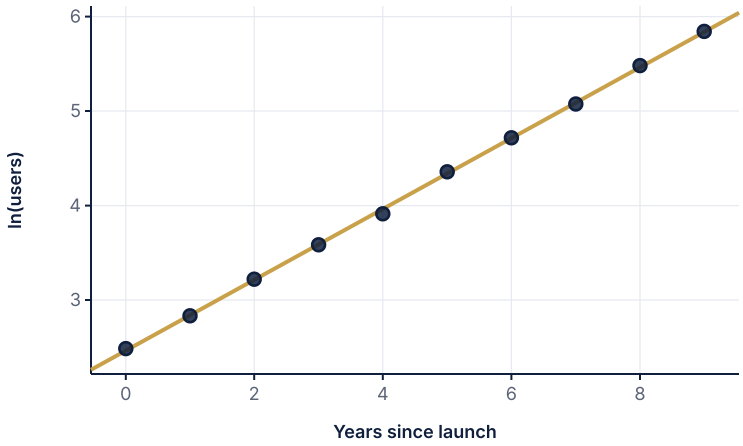
Exponential model $y = ab^x$: y changes by a constant factor per unit of x. Plot $\ln y$ against x; if this is linear, fit $\widehat{\ln y} = a + bx$.

Power model $y = ax^p$: plot $\ln y$ against $\ln x$; if this is linear, fit $\widehat{\ln y} = a + p \ln x$.

Use any logarithm (natural or base 10); the pattern and the conclusions do not depend on the base.

A transformation works when the residual plot of the transformed data shows no pattern.

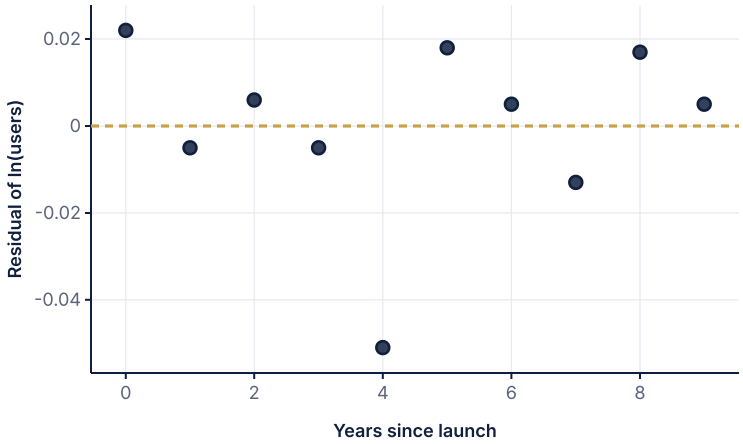
Taking the natural log of users gives a nearly perfect line: $r = 0.9998$ and $\widehat{\ln y} = 2.463 + 0.375x$.



After the log transformation of y, the pattern is linear.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Residuals for the log model: random scatter around 0, so the transformed model is appropriate.

Back-Transforming

From $\widehat{\ln y} = a + bx$, undo the log with the exponential:

$$\hat{y} = e^{a+bx} = e^{a}\left(e^{b}\right)^{x}.$$

The multiplier e^{b} is the predicted **factor** by which y changes for each additional unit of x.

From $\widehat{\ln y} = a + p\ln x$:

$$\hat{y} = e^{a}x^{p}.$$

Always report predictions in the original units of y, not in log units.

Back-transforming: $\hat{y} = e^{2.463}(e^{0.375})^{x} = 11.74\,(1.455)^{x}$. So the predicted number of users grows by a factor of about 1.455 (roughly 46% per year).

Departures from linearity. Residual plots help you diagnose problems with a linear fit: a U or arch shape signals curvature (try a transformation), a fan shape signals unequal spread (also consider a transformation of y), and a single far-off residual signals an outlier. A transformation that works should leave a patternless residual plot. Several transformations can produce a good fit, so choose using the residual plot and context, not just r.

Worked Examples

EXAMPLE 1: WHICH TRANSFORMATION?

Fictional data: a colony doubles every day: 100, 200, 400, 800, 1600 cells on days 0 to 4. A brick-shaped box with side s has volume s cubed: 1, 8, 27, 64 for s = 1 to 4. For each, name a transformation to linearize.

Day	0	1	2	3	4
Cells	100	200	400	800	1600

SOLUTION

Colony: y changes by a constant factor (x2) each unit of x, so it is exponential. Plot ln(cells) against day: $\ln y = \ln 100 + (\ln 2)x$, an exact line.

Volume: $V = s^{3}$ is a power model. Plot ln V against ln s: $\ln V = 3\ln s$, a line with slope 3.

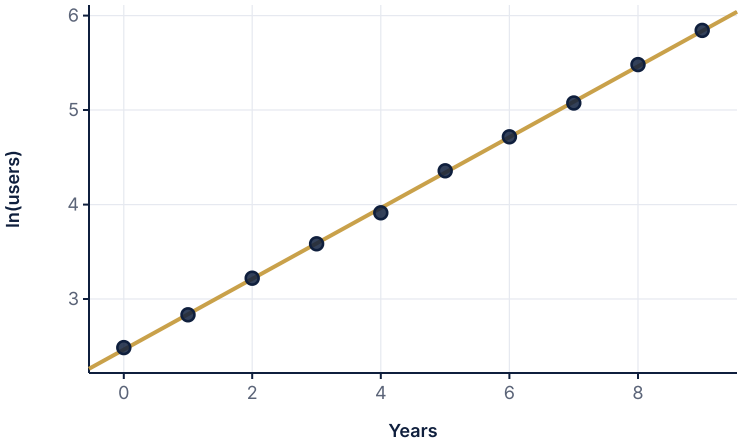
General rule: constant ratios of y for equal steps in x suggest a log of y; a straight line on log-log axes suggests a power model.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: FIT AN EXPONENTIAL MODEL

Using the app data, find the least-squares line for ln(users) on years and write the model.



SOLUTION

Compute $\ln y$ for each year (for example $\ln 12 = 2.485$ at x = 0) and use the calculator's linear regression on (x, ln y).

Result: $\widehat{\ln y} = 2.463 + 0.375x$, r = 0.9998.

Residual plot of the transformed data shows no pattern, so the model is appropriate. Back-transform: $\hat{y} = 11.74(1.455)^x$ thousand users.

EXAMPLE 3: BACK-TRANSFORMING A PREDICTION

Predict the number of users at 5.5 years and at 10 years, and comment.

SOLUTION

At x = 5.5: $\widehat{\ln y} = 2.463 + 0.375(5.5) = 4.526$, so $\hat{y} = e^{4.526} = 92.4$ thousand users.

At x = 10: $\widehat{\ln y} = 6.214$, so $\hat{y} = 500$ thousand users.

The first is interpolation. The second is extrapolation (the data end at 9 years), and growth by a constant factor cannot continue forever, so treat it cautiously. The final answer must be in thousands of users, not log units.

EXAMPLE 4: INTERPRETING THE SLOPE

Interpret the slope 0.375 of the log model $\widehat{\ln y} = 2.463 + 0.375x$ in context.

SOLUTION

In log units: for each additional year since launch, the predicted ln(users) increases by 0.375.

In original units: each additional year multiplies the predicted number of users by $e^{0.375} = 1.455$, an increase of about 45.5% per year.

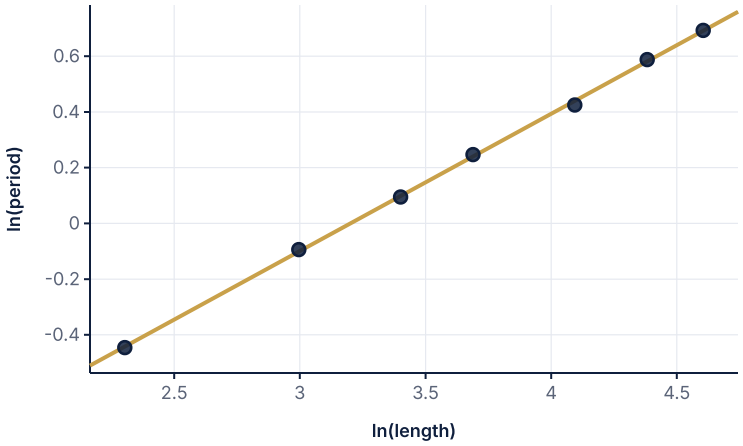
A common error is to say that users increase by 0.375 thousand per year; that would describe a linear model.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: A POWER MODEL

Fictional pendulum data: length L (cm) and period T (seconds): (10, 0.64), (20, 0.91), (30, 1.10), (40, 1.28), (60, 1.53), (80, 1.80), (100, 2.00). Fit a power model and predict the period for a 50 cm pendulum.



SOLUTION

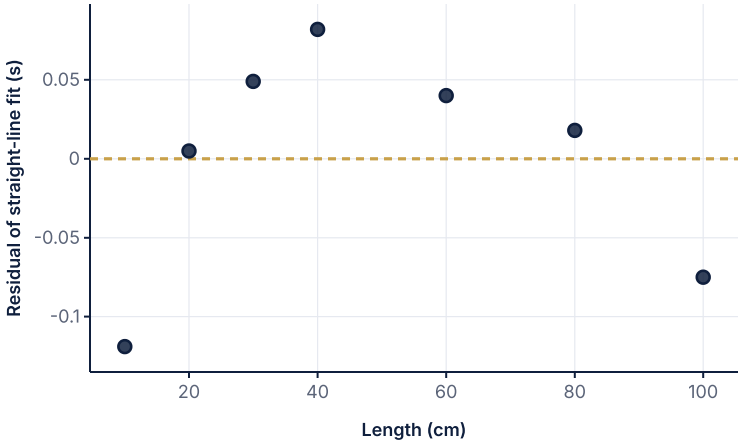
The plot of T against L curves (r = 0.989 for the line, but the residuals form an arch). Take logs of both variables.

Regression of ln T on ln L: $\widehat{\ln T} = -1.578 + 0.493 \ln L$, r = 0.9998. The exponent is about 0.5, matching the known square-root relationship.

Back-transform: $\hat{T} = e^{-1.578} L^{0.493} = 0.206\, L^{0.493}$. At L = 50: $\hat{T} = 1.42$ seconds.

EXAMPLE 6: COMPARING MODELS (EXTENSION)

For the pendulum data, compare the straight-line model with the log-log model. Which is better, and what evidence supports your choice?



SOLUTION

Straight-line fit: r = 0.989 but the residuals are -0.12, 0.005, ... (negative at the shortest length, positive in the middle, negative at the longest): an arch, so the line is systematically wrong.

Log-log fit: r = 0.9998, and the residuals of ln T are all within 0.015 of 0 with no pattern.

The log-log (power) model is better because its residual plot has no pattern; the line, despite its high r, shows curvature. Residual plots decide appropriateness, not r alone.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt: For the app data, the residual plot from a line fit of users on years is U shaped. A student computes $\widehat{\ln(\text{users})} = 2.463 + 0.375\,\text{years}$ with a patternless residual plot. Predict the number of users after 6 years and interpret the slope. Response: The log model is appropriate because its residual plot shows no pattern, unlike the U-shaped residuals of the linear fit. At 6 years, $\widehat{\ln y} = 2.463 + 0.375(6) = 4.713$, so $\hat{y} = e^{4.713} = 111$ thousand users (this is interpolation, since 6 is within the data from 0 to 9 years). The slope 0.375 means each additional year multiplies the predicted number of users by $e^{0.375} = 1.45$, about a 45% increase per year.

What the grader looks for

(1) Justifies the model with a residual plot. (2) Substitutes into the transformed equation. (3) Back-transforms to original units. (4) Interprets the slope as a multiplicative factor, in context.

Coach's Playbook

1. **Plot and diagnose.** Curved scatterplot and U-shaped residuals mean a line is the wrong model.
2. **Guess the family.** Constant ratio per step: log y vs x. Straight on log-log: power.
3. **Transform and refit.** Run linear regression on the transformed lists; check the new residual plot.
4. **Predict in transformed units, then undo.** Use e raised to the predicted log (or 10 raised to it if you used base 10).
5. **Interpret multiplicatively** for log y vs x: each additional unit of x multiplies the predicted y by e^b.

- On the built-in calculator, create a new list with the ln command, then run linear regression on the new list.
- Use the same base for the transformation and the back-transformation.
- Keep extra decimals in a, b until the final prediction to avoid rounding error after exponentiating.

Common traps

(1) Reporting a log value as the final prediction. (2) Interpreting the slope of a log model as an additive change in y. (3) Judging the transformed model only by r. (4) Taking a log of zero or negative values (impossible), which signals a different model.

Mistake Clinic

Spot the error

Student: "At 6 years the model predicts ln(users) = 4.71, so there will be 4.71 thousand users." The 4.71 is in log units. Correct: users = e^4.71 = 111 thousand.

Spot the error

Student: "The slope 0.375 means users rise by 0.375 thousand each year." The slope belongs to the model for ln y. Correct: each additional year multiplies the predicted number of users by e^0.375 = 1.455.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Student: "The line has r = 0.989, so no transformation is needed for the pendulum data." A high r can coexist with curvature. Correct: examine the residual plot; the arch shows the line is systematically wrong, and the log-log model has patternless residuals.

Real-World Connection

Scientists plot sound intensity, earthquake size and bacteria counts on logarithmic scales because those quantities grow by factors rather than amounts. Financial analysts fit models to log prices for the same reason. Recognizing constant-factor growth is a skill that connects statistics with algebra.

Challenge Corner

A fictional model for the value of a collectible is $\widehat{\ln V} = 3.2 - 0.12t$ (t in years). (a) By what factor does the predicted value change each year? (b) How long until the predicted value is half of its starting value? Solution: (a) $e^{-0.12} = 0.887$, so the value is multiplied by about 0.887 each year (an 11.3% decrease). (b) We need $e^{-0.12t} = 0.5$: $t = \ln 0.5/(-0.12) = 5.78$ years. The starting-value statement does not depend on the intercept 3.2, which only sets the starting value $e^{3.2} = 24.5$.

Practice

PRACTICE **Transforming Data to Achieve Linearity**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A scatterplot of y against x curves upward, and y roughly doubles each time x increases by 1. Which transformation is most likely to linearize the data?
   (A) Plot ln y against x **(B)** Plot ln y against ln x
   (C) Plot x against y without change **(D)** Plot y against 1/y
2. A residual plot for a linear fit shows a clear U shape, and the residual plot after transforming y with ln shows random scatter. We should
   (A) keep the original linear model since r is high
   (B) remove the points at the ends
   (C) use the transformed model
   (D) use the transformed model only if its r is lower
3. The model $\widehat{\ln y} = 1.2 + 0.05x$ is fit to transformed data. The predicted y at x = 10 is
   (A) 1.70 **(B)** 5.47 **(C)** 1.66 **(D)** 0.53
4. In the model $\widehat{\ln y} = 0.4 + 0.2x$, each additional unit of x multiplies the predicted y by about
   (A) 1.22 **(B)** 0.2 **(C)** 1.2 **(D)** 0.4
5. Data follow the power model $y = 3x^2$. Which plot will be linear?
   (A) ln y against x **(B)** y against ln x **(C)** y against x
   (D) ln y against ln x
6. After a log transformation of y, the slope of the LSRL for (x, ln y) is negative. This means the original y
   (A) decreases by a constant factor less than 1 for each unit of x
   (B) decreases by a constant amount for each unit of x
   (C) is always negative
   (D) increases by a constant factor

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. Which statement about transformations is true?
   (A) A transformation always increases r
   (B) The residual plot of the transformed data should show no pattern if the model is appropriate
   (C) Logs can be taken of zero values
   (D) The base of the logarithm changes the predicted y
8. The log-log model is $\widehat{\ln y} = -1.0 + 1.5 \ln x$. The predicted y when x = 4 is about
   (A) 1.08 **(B)** 6.0 **(C)** 2.94 **(D)** 0.37
9. Which table of (x, y) values most suggests exponential growth?
   (A) x = 1, 2, 3, 4 with y = 5, 10, 15, 20
   (B) x = 1, 2, 3, 4 with y = 5, 6, 7, 8
   (C) x = 1, 2, 3, 4 with y = 40, 30, 20, 10
   (D) x = 1, 2, 3, 4 with y = 5, 10, 20, 40
10. Fit statistics for an app: $\widehat{\ln y} = 2.463 + 0.375x$ (x in years, y in thousands of users).
    (a) Predict users at 7 years.
    (b) Interpret the slope.
    (c) Why is the prediction at 15 years risky?
11. Fictional data (x, y): (1, 3), (2, 6), (3, 12), (4, 24), (5, 48).
    (a) Describe the pattern in the ratio of consecutive y values.
    (b) Which transformation linearizes the data, and what are the slope and intercept of the line for ln y against x?
    (c) Write the model for y.
12. A scientist fits $\widehat{\ln T} = -1.578 + 0.493 \ln L$ to pendulum period T (s) and length L (cm).
    (a) Predict T for L = 25 cm.
    (b) What happens to the predicted period if the length is doubled?
    (c) Write the model for T in the form $\hat{T} = cL^p$.
13. A fictional residual plot for a line fit shows spread that fans out as x increases, while the scatterplot is roughly linear.
    (a) What does the fan shape say about prediction errors?
    (b) Name a transformation that might help.
    (c) How will you check that it worked?
14. For the collectible model $\widehat{\ln V} = 3.2 - 0.12t$ (t in years), (a) find the predicted starting value, (b) find the factor by which the value changes each year, (c) find the predicted value after 10 years.
15. A student compares two models for the same data. Model 1 (linear): r = 0.97, residual plot shows an arch. Model 2 (log-log): r = 0.995, residual plot shows random scatter.
    (a) Which model should be used and why?
    (b) Could the student rely on r alone? Explain.
    (c) What is one thing to check before trusting predictions from Model 2?

T Unit 3 Test

Check your mastery of two-variable data: tables, scatterplots, correlation, regression, residuals and transformations.

Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

UNIT TEST

Unit 3 Test

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

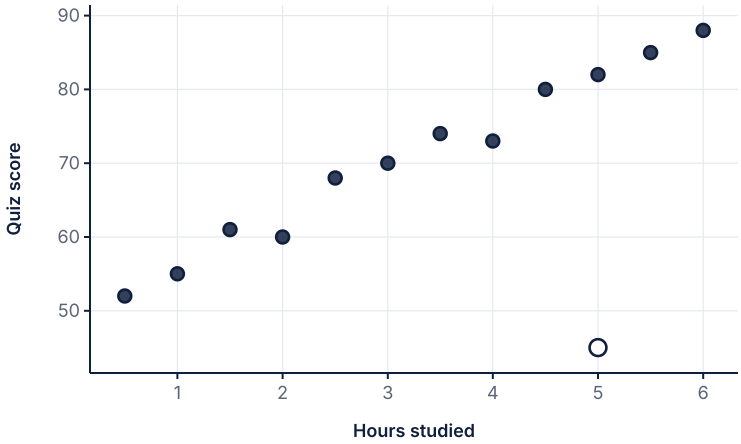
1. A fictional survey asked students about a phone policy. Of the students who favor the policy, what percent are in Grade 12?

	Favor	Oppose	Total
Grade 9	48	32	80
Grade 12	36	84	120
Total	84	116	200

(A) 30.0% **(B)** 42.9% **(C)** 18.0% **(D)** 45.0%

2. A survey of adults finds that people who use a fitness tracker have lower resting heart rates. Which conclusion is justified?
   (A) Using a tracker lowers resting heart rate
   (B) Lower heart rate makes people buy a tracker
   (C) There is no association because the data are observational
   (D) The two variables are associated, but the survey alone does not show that trackers cause lower heart rates

3. In the scatterplot of study time and quiz score, the highlighted point is best described as



   (A) an outlier in the y direction, far below the pattern
   (B) a high-leverage point because of its large x value
   (C) part of a second cluster
   (D) an outlier in the x direction

4. A scatterplot shows scores that decrease quickly at first and then level off as practice time increases. Which description is best?
   (A) Strong, positive, linear **(B)** Strong, negative, linear **(C)** Strong, negative, curved
   (D) Weak, negative, curved

5. The correlation between height in inches and arm span in inches is 0.92. If both measurements are converted to centimeters, the correlation is
   (A) 2.34 **(B)** 0.92 **(C)** 0.36 **(D)** 1.00

6. A fictional LSRL is: predicted exam score = 40 + 5 (hours of tutoring). Which interpretation of the slope is correct?
   (A) For each additional hour of tutoring, the predicted exam score increases by 5 points
   (B) Each hour of tutoring raises a student's exam score by exactly 5 points
   (C) A student with no tutoring scores 40 points
   (D) The correlation between tutoring and score is 5

7. For a data set, xbar = 40, sx = 10, ybar = 120, sy = 25, and r = -0.6. The LSRL for predicting y from x is
   (A) y-hat = 120 - 0.6x **(B)** y-hat = 180 - 0.24x **(C)** y-hat = 180 - 1.5x
   (D) y-hat = 60 - 1.5x

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

8. A regression has s = 3 (the standard deviation of the residuals). A point has a residual of 9. How many standard deviations is that residual from 0?
   (A) 9 **(B)** 27 **(C)** 0.33 **(D)** 3
9. The LSRL is y-hat = 12 + 3x. For the observation (5, 24), the residual is
   (A) -3 **(B)** 3 **(C)** 39 **(D)** 12
10. For the LSRL predicting weight from height, r-squared = 0.64. Which interpretation is correct?
   (A) 64% of the points are on the line
   (B) About 64% of the variation in weight is explained by the linear relationship with height
   (C) The correlation is 0.64
   (D) Height explains 64% of each person's weight
11. A point is far to the right of the other points and far off the line they follow. Removing it changes the slope dramatically. This point is
   (A) a low-leverage outlier **(B)** a point with a residual of exactly 0
   (C) an average point with no effect **(D)** an influential, high-leverage point
12. A log model for a growing app is ln(y-hat) = 0.5 + 0.3x (x = years, y = users in thousands). The predicted number of users at x = 4 is about
   (A) 1.7 thousand **(B)** 2.0 thousand **(C)** 5.47 thousand
   (D) 0.55 thousand
13. A fictional library surveyed 240 patrons about their preferred book format. Among 120 patrons under 25, 36 prefer print and 84 prefer e-books. Among 120 patrons age 25 and over, 72 prefer print and 48 prefer e-books. [6 points]
   (a) What percent of all patrons prefer e-books? [1 point]
   (b) Calculate the percent who prefer e-books within each age group. [2 points]
   (c) Is there an association between age group and format preference? Justify using your results. [2 points]
   (d) A librarian concludes that being under 25 causes patrons to prefer e-books. Is the conclusion justified? [1 point]
14. A fictional bike-share company records the daily high temperature x (degrees Celsius) and the number of rentals y (hundreds) on 10 days: (10, 8.2), (12, 9.6), (14, 10.1), (16, 12.4), (18, 13.0), (20, 15.2), (22, 15.9), (24, 18.3), (26, 18.8), (28, 21.0). Computer output for the regression of rentals on temperature is below. [10 points]
   Predictor Coef
   Constant 0.8176
   Temp 0.7070
   S = 0.4729 R-sq = 98.9%
   The residuals, in the order of the days, are: 0.31, 0.30, -0.62, 0.27, -0.54, 0.24, -0.47, 0.52, -0.40, 0.39. A plot of these residuals against temperature shows no pattern.
   (a) Write the equation of the least-squares line and interpret the slope in context. [2 points]
   (b) Describe the association between temperature and rentals and explain why the residuals support using a linear model. [2 points]
   (c) The residual for the day with temperature 20 is listed above. Show how it is computed and interpret it. [2 points]
   (d) Interpret R-sq and S in context. [2 points]
   (e) Predict the rentals for a day with a high of 40 degrees Celsius and comment on the reliability of your prediction. [2 points]

Answers Exploring Two-Variable Data: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

3.1 Two-Way Tables and Categorical Association

PRACTICE · TWO-WAY TABLES AND CATEGORICAL ASSOCIATION

1. A. Among bus riders (column total 100), 62 are in grades 9-10, so 62/100 = 62%. The tempting 62/130 conditions on grade 9-10 instead, which answers the reverse question.
2. B. Equal conditional percentages mean the distributions match, so there is no association. Equal percents do not tell us anything about group sizes, and association never proves cause.
3. C. Seniors: 30/50 = 60%. Juniors: 60/150 = 40%. The tempting count comparison ignores that there are three times as many juniors.
4. A. A marginal distribution ignores the other variable. The second option describes a conditional distribution and the third a joint relative frequency.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. B. Left: 10/40 = 25%. Right: 90/360 = 25%. The conditional percentages are equal, so there is no association. Raw counts differ only because the groups differ in size.
6. D. Warm weather increases both swimming and ice cream sales, so it is a lurking variable. Association does not imply cause, and nothing here says the data came from an experiment.
7. C. A reversal between subgroup and pooled comparisons caused by unequal subgroup sizes is Simpson's paradox. The other terms describe different ideas.
8. D. 24/80 = 0.30 and 36/120 = 0.30, so the conditional distributions match. Counts differ only because there are more boys.
9. B. Different conditional percentages mean association. Bars in a segmented graph are scaled to 100 percent, so their heights say nothing about group size, and association alone does not establish cause.
10. (a) Sport: 54 yes, 36 no (total 90). No sport: 24 yes, 36 no (total 60). Column totals: 78 yes, 72 no, grand total 150.
    (b) Sport: 54/90 = 60.0%. No sport: 24/60 = 40.0%.
    (c) The percentages differ (60.0% versus 40.0%), so sport and sleep are associated in this sample. Because this is an observational survey, we cannot conclude that playing a sport causes more sleep; a confounder such as general healthy habits could explain it.
11. (a) 60/260 = 23.1%.
    (b) 38/100 = 38.0%.
    (c) 72/260 = 27.7%. This is a joint relative frequency, with the grand total in the denominator.
12. (a) Choir: 120/200 = 60%. Not in choir: 60/100 = 60%.
    (b) The conditional percentages are equal, so there is no association between choir membership and taking music theory in this sample.
    (c) New choir: 220/300 = 73.3% take theory versus 60%, so an association now exists. Comparing counts (220 versus 60) would mislead because the group sizes differ; percentages account for group size.
13. (a) First-time: X 90.0%, Y 85.0%. Repeat: X 25.0%, Y 30.0%. So X is better for first-time students, Y for repeat students.
    (b) Overall: X 95/120 = 79.2%, Y 47/120 = 39.2%. X looks far better overall.
    (c) The groups differ in composition: X sees mostly first-time students (who pass more often) while Y sees mostly repeat students. This is Simpson's paradox: student type is a lurking variable. A fair comparison is made within student type.
14. (a) Observational: people chose whether to eat breakfast.
    (b) For example, an overall healthy lifestyle or a regular work schedule: people with such habits might both eat breakfast regularly and miss fewer workdays, so the association would appear even if breakfast had no effect.
    (c) Recruit volunteers, randomly assign half to eat a daily breakfast and half not, track missed workdays, and compare. Random assignment spreads confounders evenly between groups so a difference can reasonably be attributed to the treatment.
15. (a) Grade 9: 45, 15, total 60. Grade 12: 30, 50, total 80. Column totals: favor 75, oppose 65, grand total 140.
    (b) Favor 75/140 = 53.6%, oppose 65/140 = 46.4%.
    (c) Favor segment: Grade 9 is 45/60 = 75.0%; Grade 12 is 30/80 = 37.5%. The segment heights differ a lot, so opinion and grade are associated: ninth graders favor the change far more often than twelfth graders in this sample.

3.2 Scatterplots: Form, Direction, Strength, Outliers

PRACTICE · SCATTERPLOTS: FORM, DIRECTION, STRENGTH, OUTLIERS

1. C. The question asks whether temperature helps explain sales, so temperature is explanatory and sales is the response. Reversing them is the tempting error.
2. A. Direction describes whether y tends to rise or fall as x increases. 'Linear' and 'strong' describe form and strength, not direction.
3. B. The points follow a smooth decreasing curve closely. Calling it linear because it decreases ignores the visible curvature.
4. A. An outlier departs from the pattern. An extreme x or y value that follows the trend is not necessarily an outlier.
5. D. The points fall steadily and tightly along a line. A positive direction or a curve is not shown.
6. C. The points scatter with no clear trend (correlation about 0.27). Seeing some up-and-down variation does not create a pattern.
7. B. Older students have larger feet and read better. Association alone does not imply cause, and lurking variables are the classic explanation.
8. A. Its x value is typical, but its y value is far below the pattern for that x. It is not extreme in x.
9. D. Distinct groups often signal a third variable. Ignoring them or deleting one group would hide the structure.
10. (a) Absences is explanatory; final grade is the response.
    (b) The association is negative, linear and strong: more absences go with lower grades, and the points follow a line closely. There are no outliers.
    (c) r = -0.99 (computed on the built-in calculator).
11. (a) Negative and nonlinear (curved, decreasing at a decreasing rate).
    (b) A line would overestimate activity in the middle ages and underestimate near the ends, so predictions would be systematically wrong.
    (c) No. The data are observational; other factors (health, owner routine, breed) could be involved. The data show association and describe a pattern, not cause.
12. (a) Explanatory: hours of sleep; response: typing mistakes; negative association expected.
    (b) Explanatory: outdoor temperature; response: heating bill; negative.
    (c) Explanatory: number of pages; response: price; positive.
    The choice depends on the question; reversing the roles would give a different line but the same direction.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. (a) r drops to about 0.66 (the line is pulled and the points are less tight).
(b) In y only: its x value (5.0) is within the range of the x values, but its score is far below the pattern.
(c) Check whether it is a recording error or a legitimate unusual case, report results with and without it, and remove it only with a justified reason.
14. (a) Negative, curved, moderately strong: errors decrease as practice time increases, with diminishing returns.
(b) The two students with 20 practice hours and many errors lie far above the pattern: outliers in the response direction.
(c) For example, the two may be playing a much harder piece; difficulty of the piece could influence both practice time and errors.
15. (a) A strong, negative, roughly linear association with two distinct clusters; heavier vehicles have lower mileage.
(b) The groups represent a lurking variable (vehicle type) and have different intercepts; a single line may not represent either group well.
(c) Negative: within the compact cars mileage drops from 52 to 42 mpg as weight increases (r about -0.99 within this group).

3.3 Correlation

PRACTICE · CORRELATION

1. A. Correlation must lie between -1 and 1 inclusive. 1.4 is out of range; students sometimes confuse r with a slope, which can be any size.
2. B. A strong negative association has r near -1. +0.94 has the wrong sign, and -1.40 is impossible.
3. C. Linear changes of units do not change r. Multiplying by 2.54 would take r out of range.
4. D. Correlation is symmetric in x and y. Only the regression line depends on the roles of x and y.
5. A. r measures only linear association. A perfect curve can have zero correlation. Calling it 'unrelated' is the common mistake.
6. D. r gives direction and strength of linear association. The variation explained would be r squared = 0.64, and causation is not established.
7. B. Correlation is not resistant; a far outlier can lower r a lot. It would not make r exactly 0 or leave it unchanged.
8. C. Strength is |r|, and the sign gives the direction. Many students mistakenly treat negative values as 'weaker'.
9. A. Adding a constant shifts the points but not their pattern, so r is unchanged. The other values are out of range or ignore this property.
10. (a) The points rise from lower left to upper right close to a line: positive, linear, strong.
(b) r = 0.978.
(c) There is a strong, positive, linear association between hours of sleep and quiz score for these 6 students. This is an observational association and does not show that more sleep causes higher scores.
11. (a) There is a moderate, negative, linear association: phones with more apps tend to have shorter battery life.
(b) No. Correlation is unchanged by linear changes of units.
(c) No. The data are observational; other factors (phone age, screen use) could be related to both app count and battery life.
12. (a) The last point (5, 10) contributes the largest product (1.74) because it is above average in y and above average in x; (1, 3) contributes next, since it is below average in both variables.
(b) r = sum/4 = 0.922.
(c) r decreases: the point (5, 2) is below average in y but above average in x, so it contributes a negative product; the new r = 0.069.
13. (a) No. The height is completely determined by time.
(b) A strong nonlinear (curved) association; the rise and fall cancel in the correlation.
(c) A curve or other nonlinear model (for example a quadratic); r is only suitable for linear patterns.
14. (a) Averaging reduces the individual variation around the pattern, so the points for class means lie closer to a line.
(b) The reader would overstate how strongly an individual's study time predicts his or her own score (the ecological fallacy); the individual-level association is only moderate (r = 0.45).
15. (a) The added point lies far below the pattern, so the points are less tightly clustered around a line and r drops.
(b) No. One point lowered r from 0.99 to 0.66, so r is not resistant to outliers.
(c) Report the scatterplot, note the unusual student, and give r both with and without that student, explaining why any point was set aside.

3.4 Least-Squares Regression Lines

PRACTICE · LEAST-SQUARES REGRESSION LINES

1. C. The criterion is the smallest sum of squared residuals. The plain sum of residuals is zero for the LSRL (and for many other lines), so it cannot be the criterion.
2. A. b = r sy/sx = 0.6(15/5) = 1.8. The value 0.2 comes from inverting the ratio, and 0.6 forgets the standard deviations.
3. D. a = ybar - b xbar = 50 - 1.8(20) = 14. A student who uses the wrong sign or forgets to multiply gets other values.
4. B. The slope is a predicted (average) change in price per additional year. Saying 'exactly' treats the line as deterministic, and the other choices confuse the slope with the intercept.
5. C. A strong fit inside the range does not guarantee the pattern continues outside it.
6. D. From a = ybar - b xbar, putting x = xbar gives y-hat = ybar. The origin is on the line only when the intercept is 0.
7. C. 3.2 + 1.7(5) = 11.7. 8.5 is 1.7 × 5 alone, and 4.9 is 3.2 + 1.7.
8. B. Since b = r sy/sx and the standard deviations are positive, b and r share a sign.
9. A. b = r(sy/sx) = 0.5(1) = 0.5. Using 1/r gives 2.0, and ignoring r gives 1.0.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. (a) Using the built-in calculator: $\hat{y} = 5.400 + 1.320x$ with r = 0.998.
    (b) For each additional kilometer, the predicted water use increases by 1.320 hundred mL (about 132 mL).
    (c) $\hat{y} = 5.400 + 1.320(17) = 27.84$ hundred mL. Since 17 is between 5 and 30, it is interpolation.
11. (a) b = -0.75(6/3) = -1.5; a = 80 - (-1.5)(12) = 98. So y-hat = 98 - 1.5x.
    (b) y-hat = 98 - 1.5(15) = 75.5.
    (c) For each additional hour of video games per week, the predicted exam score decreases by 1.5 points.
12. (a) 25 + 4(12) = 73 cm (interpolation).
    (b) 25 + 4(100) = 425 cm. This is extrapolation far beyond 20 days; plants stop growing linearly, so 425 cm is not believable.
    (c) With 0 days of growth the predicted height is 25 cm. Day 0 is just outside the data, so interpret with care; it may represent the starting height of a seedling but the line was only fit for 5 to 20 days.
13. (a) Solving for x algebraically gives the line through the same points only if the data fall exactly on a line (r = 1 or -1). Otherwise regressing x on y minimizes horizontal squared distances, giving a different line.
    (b) b = r sy/sx so 2 = r(5/2) and r = 0.8. The slope for predicting x from y is r sx/sy = 0.8(2/5) = 0.32, not 0.5.
14. (a) xbar = 5.500, and $2.699 + 1.180(5.500) = 9.188 =$ ybar.
    (b) $\hat{y}(8) = 12.14$ cm; residual = 11.8 - 12.14 = -0.34 cm.
    (c) The negative residual means the 8-hour seedling is about 0.34 cm shorter than the line predicts.
15. (a) b = r sy/sx so -0.9 = r(2.4/1.5), r = -0.5625.
    (b) Multiplying y by 60 multiplies the slope by 60: -54 seconds per hour.
    (c) r is unchanged by units: -0.5625.

3.5 Residuals, r-squared, and Standard Deviation of Residuals

PRACTICE · RESIDUALS, R-SQUARED, AND STANDARD DEVIATION OF RESIDUALS

1. A. Residual = observed - predicted = 42 - 37 = 5. Reversing the subtraction gives -5, a very common sign error.
2. C. Curved residual patterns mean the relationship is not linear. A large r can occur with a curved relationship.
3. B. r-squared = 0.81, interpreted as 81% of the variation explained. It cannot be negative, and 0.9 is $|r|$.
4. A. r-squared is the share of the variation in y accounted for by the linear relationship. The correlation would be about 0.85, and nothing here establishes causation.
5. D. s measures the typical prediction error in units of y. The average residual of a least-squares line is exactly 0.
6. A. Far in x gives leverage; far from the trend gives influence on the slope. Such points can move the line a lot.
7. B. Its residual is large (outlier), but being near the mean of x gives it little leverage, so it mainly shifts the intercept.
8. C. For any least-squares line with an intercept, positive and negative residuals balance exactly. High r does not make residuals zero.
9. D. A fan shape indicates unequal spread. A random cloud is what we want. A straight-line residual plot is not possible for least squares because residuals are uncorrelated with x.
10. (a) $\hat{y} = 10.267 + 3.279(9) = 39.78$; residual $= 41 - 39.78 = 1.22$.
    (b) This musician's score is 1.22 points above the score predicted for 9 hours of rehearsal.
    (c) The least-squares line is the one for which positive and negative misses balance: the line passes through (xbar, ybar), which forces the sum of residuals to 0.
11. (a) r-squared = 0.36. About 36% of the variation in weight is explained by the linear relationship with height.
    (b) Actual weights typically differ from predicted weights by about 6.5 kg.
    (c) Only moderately well: most of the variation (64%) remains unexplained, and a typical miss of 6.5 kg is large relative to differences among students.
12. (a) A linear model is appropriate; the errors show no pattern.
    (b) The relationship is curved, so a linear model is not appropriate; the line underpredicts for some x and overpredicts for others. Consider a transformation or curved model.
    (c) The linear form may be fine, but the size of prediction errors grows with x, so predictions are less precise for large x; consider a transformation that stabilizes spread.
13. (a) Yes. Its x value 20 is far from the other x values (1 to 8), so it has high leverage.
    (b) Yes. It is far from the trend and the slope fell from 1.48 to 0.31 and r from 1.00 to 0.51: the line changed substantially.
    (c) Not automatically. First check whether it is an error; report the analysis with and without it and explain any decision; any prediction near x = 20 needs additional data.
14. (a) r-squared = 1 - 37.19/924.1 = 0.960.
    (b) s = sqrt(37.19/8) = 2.16.
    (c) About 96.0% of the variation in scores is explained by the line, and actual scores typically differ from predictions by about 2.2 points.
15. (a) r measures how tightly points cluster around a straight line; a gently curved pattern can still be very close to a line, so r stays high even though the line systematically misses.
    (b) Examine the form, consider a curved model or a transformation of x or y (Lesson 3.6), and recheck the residual plot.
    (c) It does not tell you whether the relationship is linear or whether the model is appropriate; the residual plot does.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

3.6 Transforming Data to Achieve Linearity

PRACTICE · TRANSFORMING DATA TO ACHIEVE LINEARITY

1. A. Constant-factor growth is exponential, so the log of y against x is linear. A log-log plot is for power models, where the ratio depends on the ratio of x values.
2. C. The patternless residuals show the transformed model is appropriate. A high r in the original fit does not repair its curved residual pattern.
3. B. ln y-hat = 1.2 + 0.5 = 1.7, so y-hat = e^1.7 = 5.47. The value 1.70 is a log value left un-exponentiated.
4. A. The multiplier is e^0.2 = 1.221, about 1.22. Reading 0.2 as an additive change in y mixes up the log scale with the original scale.
5. D. Taking logs gives ln y = ln 3 + 2 ln x, a line in (ln x, ln y). For ln y against x, the pattern remains curved.
6. A. A negative slope b gives a multiplier e^b < 1, so y shrinks by a constant percentage, not a constant amount.
7. B. Judge the transformed model by its residual plot. A transformation does not always improve r, logs of 0 do not exist, and the base cancels when you back-transform.
8. C. ln 4 = 1.386, so ln y-hat = -1.0 + 1.5(1.386) = 1.079 and y-hat = e^1.079 = 2.94. The value 1.08 is the log value, not yet back-transformed.
9. D. Equal steps in x multiply y by the same factor 2, which is exponential. The other tables change y by constant amounts, which is linear.
10. (a) $\widehat{\ln y} = 2.463 + 0.375(7) = 5.088$, so $\hat{y} = e^{5.088} = 162.1$ thousand users.
    (b) Each additional year multiplies the predicted number of users by $e^{0.375} = 1.455$ (about a 46% increase per year).
    (c) 15 years is far outside the data range of 0 to 9 years, so it is extrapolation; constant-factor growth rarely continues indefinitely (the market saturates).
11. (a) Each y is twice the previous one: constant ratio 2, so exponential growth.
    (b) Plot ln y against x. ln y = ln 3 + (x - 1) ln 2 = ln 1.5 + x ln 2, so the intercept is ln 1.5 = 0.405 and the slope is ln 2 = 0.693.
    (c) y = 1.5(2)^x. Check: x = 3 gives 12.
12. (a) $\widehat{\ln T} = -1.578 + 0.493 \ln 25 = 0.009$, so $\hat{T} = 1.01$ s.
    (b) Doubling L multiplies the prediction by $2^{0.493} = 1.407$ (about 41% longer).
    (c) $\hat{T} = 0.206 L^{0.493}$.
13. (a) Prediction errors are larger for larger x; predictions are less precise there, so a single value of s misdescribes the typical error at all x.
    (b) A log transformation of y (or a power transformation) often stabilizes spread when errors grow in proportion to y.
    (c) Refit with the transformed response and inspect the new residual plot: it should show random scatter with roughly equal spread.
14. (a) At t = 0: $e^{3.2} = 24.5$ (units of V).
    (b) $e^{-0.12} = 0.887$, so the value is multiplied by about 0.887 each year, a decrease of 11.3% per year.
    (c) $\widehat{\ln V} = 3.2 - 1.2 = 2.0$, so $\hat{V} = e^{2.0} = 7.39$.
15. (a) Model 2, because its residual plot has no pattern, indicating a power relationship is appropriate; Model 1's arch shows systematic error.
    (b) No. r measures closeness to a line, and a curved association can have a high r; the residual plot detects the curve.
    (c) Whether the x value being predicted is within the data range (extrapolation), and that the back-transformed prediction is reported in the original units.

T Unit 3 Test

UNIT TEST · UNIT 3 TEST

1. B. The phrase 'of the students who favor' makes the Favor column total (84) the denominator: 36/84 = 42.9%. The tempting 30.0% is 36/120, which conditions on Grade 12 instead.
2. D. An observational survey can show association but cannot establish cause. Both causal statements go beyond the data, and observational data can certainly show association.
3. A. Its x value (5.0 hours) is inside the range of the others, but its score is far below the pattern. Leverage requires an extreme x value.
4. C. The direction is negative and the points follow a smooth curve, so the form is curved rather than linear. Calling it linear because it decreases is the common error.
5. B. Correlation has no units and is unchanged by linear changes of scale. The value 2.34 (0.92 × 2.54) would be out of range, and 1.00 would require a perfect line.
6. A. The slope is a predicted change per additional hour. The first alternative drops 'predicted' and treats the line as exact; the others describe the intercept or confuse slope with r.
7. C. b = r sy/sx = -0.6(25/10) = -1.5 and a = ybar - b xbar = 120 + 60 = 180. Using sx/sy gives -0.24, and forgetting the sign of b gives other intercepts.
8. D. Divide: 9/3 = 3 standard deviations. Multiplying gives 27, and dividing in the wrong order gives 0.33.
9. A. y-hat = 12 + 15 = 27, so residual = 24 - 27 = -3. Reversing the subtraction (predicted minus observed) gives +3.
10. B. r-squared is the proportion of variation in y explained by the linear relationship with x. The correlation would be 0.8 in magnitude, and the other choices misread it.
11. D. A far-right x value gives leverage, and being off the trend lets it pull the line, so removing it changes the slope. A low-leverage outlier would change mainly the intercept.
12. C. ln(y-hat) = 0.5 + 0.3(4) = 1.7, so y-hat = e^1.7 = 5.47 thousand. Reporting 1.7 forgets to back-transform.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. Scoring guide (6 points): (a) 1 point: 132/240 = 55%. (b) 2 points: 1 point each for under 25: 84/120 = 70%; 25 and over: 48/120 = 40%. (c) 2 points: 1 point for comparing the conditional percentages (70% versus 40%) and 1 point for concluding that age group and format preference are associated in this sample because the conditional distributions differ. (d) 1 point: No; this is an observational survey, so the association does not establish cause, and confounders such as technology access could explain it.

    Model response: (a) E-book readers total 84 + 48 = 132, and 132/240 = 55%. (b) Under 25: 84/120 = 70%. Age 25 and over: 48/120 = 40%. (c) The percentages differ noticeably (70% versus 40%), so age group and preferred format are associated: younger patrons prefer e-books more often. (d) Not justified. The survey is observational, so it shows association only; other variables, like comfort with technology or device ownership, might be related to both age and preference.

14. Scoring guide (10 points): (a) 1 point for rentals-hat = 0.818 + 0.707 (temperature); 1 point for the slope: for each additional degree Celsius, the predicted number of rentals increases by about 0.71 hundred (about 71 rentals). (b) 1 point: strong, positive, linear association with no outliers; 1 point: the residual plot has no pattern (random scatter around 0), so a linear model is appropriate. (c) 1 point: predicted = 0.818 + 0.707(20) = 14.96 and residual = 15.2 - 14.96 = 0.24; 1 point: the actual rentals that day were about 0.24 hundred above the predicted. (d) 1 point: about 98.9% of the variation in daily rentals is explained by the linear relationship with temperature; 1 point: actual rentals typically differ from the predicted rentals by about 0.47 hundred. (e) 1 point: 0.818 + 0.707(40) = 29.1 hundred rentals; 1 point: 40 degrees is far outside the observed range of 10 to 28, so this is extrapolation and may be unreliable (rentals may level off or fall in extreme heat).

    Model response: (a) Rentals-hat = 0.818 + 0.707(temp). For each additional degree Celsius, the predicted number of rentals increases by about 0.71 hundred. (b) The association is strong, positive and linear, with no outliers; the residuals show no pattern, so a line is an appropriate model. (c) Predicted = 14.96 hundred, residual = 15.2 - 14.96 = 0.24: that day had about 0.24 hundred more rentals than predicted. (d) About 98.9% of the variation in rentals is explained by the linear relationship with temperature, and actual rentals typically differ from predictions by about 0.47 hundred. (e) Predicted rentals = 29.1 hundred, but since 40 is well outside 10 to 28 this is extrapolation and not reliable.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 4: Collecting Data: Sampling, Experiments, and Simulation

Design studies, avoid bias, and decide what conclusions are justified. 6 lessons, 104 practice and test items, all with full explanations.

In this unit · Exam Unit 1

4.1 Populations, Samples, and Sources of Bias
4.2 Sampling Methods: SRS, Stratified, Cluster, Systematic
4.3 Observational Studies, Experiments, and Confounding
4.4 Designing Experiments: Control, Randomization, Replication, Blocking
4.5 Scope of Inference and Writing About Study Design
4.6 Simulation and Random Digits
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

4.1 Populations, Samples, and Sources of Bias

Learn to separate populations from samples and parameters from statistics, and to explain exactly how a flawed data collection method pushes results in a predictable direction.

Why This Matters

Every claim you read about "what students think" or "how many households recycle" comes from data, and data are only as trustworthy as the way they were collected. A huge sample collected badly is worse than a small sample collected well. This lesson gives you the vocabulary to say exactly who was studied, what was measured, and why a method might push results in a predictable direction.

Students most often go wrong in two places: they mix up a parameter with a statistic, and they name a type of bias without explaining *which direction* it pushes the result and *why*. Both skills are tested in Exam Unit 1 (Exploring One-Variable Data and Collecting Data), usually in free-response questions where "name the bias" earns almost nothing without the explanation.

Core Ideas

Statistics begins with a question about a **population**, the entire group of individuals we want to know about. A **census** measures every individual in the population. A **sample** measures only some of them. A census sounds ideal, but it is often too expensive, too slow, or impossible. Testing every battery a factory makes until it dies would leave nothing to sell, so we sample.

Key vocabulary

Population: the entire group of individuals or objects we want information about.

Sample: the part of the population from which we actually collect data.

Parameter: a number that describes a population (usually unknown). Written with Greek letters such as μ and p.

Statistic: a number that describes a sample (known once the data are in). Written $\bar{x}$ and $\hat{p}$.

We use a statistic to estimate a parameter.

A quick memory trick: **P**opulation and **P**arameter both start with P; **S**ample and **S**tatistic both start with S. When a problem gives a percent from a survey, it is a statistic; when it describes everyone in the group, it is a parameter.

Between the population and the sample sits the **sampling frame**, the actual list or procedure from which individuals are drawn. If the frame is incomplete, some members of the population have no chance of being chosen. A phone directory is a poor frame for "all adults in a city" because it leaves out people with unlisted or mobile-only numbers.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Idea	Question it answers	Example
Population	Who do we want to learn about?	All 2,400 students at Larkspur High
Sampling frame	Who could actually be selected?	The school's attendance roster
Sample	Who was actually measured?	The 120 students who completed the survey
Parameter	What is the true value?	Proportion p of all 2,400 who bike to school
Statistic	What did the sample show?	$\hat{p} = 0.35$ in the sample

Bias means a method systematically favors certain outcomes, so the statistic will tend to be too high (or too low) in the long run, no matter how large the sample is. Bias is different from ordinary sampling variability, which is random scatter that shrinks as the sample grows. Increasing the sample size does **not** fix bias.

Common sources of bias

Undercoverage: some part of the population has no chance (or a reduced chance) of being in the sample, often because of the frame.

Voluntary response: people choose whether to respond (call-in, online poll, open survey). Those with strong opinions respond most.

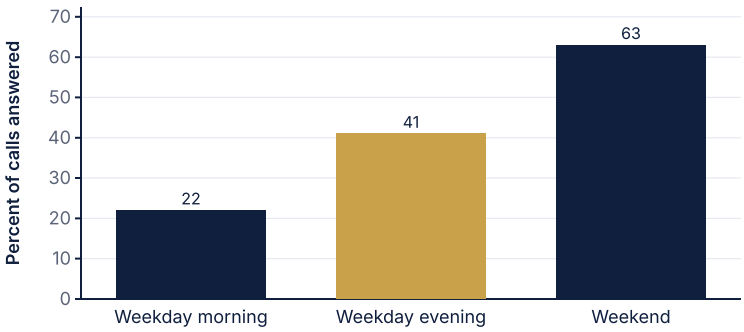
Convenience sampling: choosing individuals who are easy to reach. The easy-to-reach may differ from everyone else.

Nonresponse: individuals chosen for the sample cannot be contacted or refuse. The ones who do not answer may differ from those who do.

Response bias: people give inaccurate answers (embarrassing question, interviewer pressure, poor memory, untruthfulness).

Wording bias: a leading or confusing question steers answers.

Notice the difference between nonresponse and voluntary response. In **nonresponse**, a random or planned sample was chosen but some chosen people did not answer. In **voluntary response**, nobody was chosen; people volunteered themselves. Undercoverage is about who was left off the frame before anyone was chosen.



A made-up call log shows why nonresponse depends on when you call. A phone poll made only on weekday mornings would hear mostly from people who are home then.

When you describe bias, always give the **direction**. Write: "This method will likely overestimate the proportion who ___ because ___." A graded response needs the cause-and-effect chain: the flawed method leads to a sample that differs from the population in a way related to the variable, which pushes the statistic in a stated direction.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

PARAMETER OR STATISTIC?

A music teacher asks every student in one class of 28 about practice time and finds a mean of 4.5 hours per week. The teacher wants to know the mean for all 600 band students in the district. Identify the population, sample, parameter and statistic.

SOLUTION

Population: all 600 band students in the district (the group the teacher wants to learn about).

Sample: the 28 students in the one class.

Parameter: μ, the mean weekly practice time of all 600 band students (unknown).

Statistic: $\bar{x} = 4.5$ hours, the mean of the 28 students in the sample.

Note that this sample is also a convenience sample, so $\bar{x}$ may not represent the district well.

CENSUS OR SAMPLE?

A cafe owner wants the average amount of time customers wait in line. (a) Describe a census for this question. (b) Explain why a sample is more sensible.

SOLUTION

(a) A census would record the wait time of every customer who ever visits the cafe (or of every customer in a defined period such as a month).

(b) Timing every customer is costly and demanding on staff, while a well chosen sample of customers can estimate the mean wait quite accurately. A census also is not always possible because future customers do not exist yet.

The sample statistic $\bar{x}$ estimates the population mean wait μ.

SAMPLING FRAME AND UNDERCOVERAGE

A town has 8,000 households. The council wants to learn the proportion that use public transit and draws a sample from a list of households that have a landline telephone. Identify the frame and the bias, with direction.

Group	Households in town	In landline frame
Landline	5,200	5,200
No landline	2,800	0
Total	8,000	5,200

SOLUTION

The frame is the list of landline households, not all 8,000 households.

Undercoverage: households with no landline are never selected.

Direction: if landline-free households are younger or rent more often and are more likely to use transit, the sample will likely underestimate the proportion who use public transit.

A better frame would combine address lists, which cover essentially every household.

WHICH TYPE OF BIAS?

Name the main source of bias in each and give the likely direction. (a) A radio host asks listeners to call in about whether the city should raise parking fines. (b) Interviewers ask teenagers, "Do you ever skip the daily homework you are supposed to do?" in front of their parents. (c) Of 500 emailed surveys, 80 are returned.

SOLUTION

(a) Voluntary response. People with strong opinions, especially those angry about a fine increase, are the most likely to call, so opposition to higher fines is overstated.

(b) Response bias. With parents present, teens are likely to deny skipping homework, so the proportion who skip is underestimated.

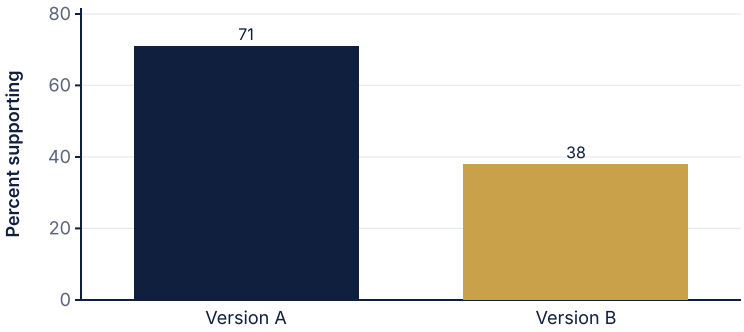
(c) Nonresponse. The response rate is $80/500 = 16\%$. People who return surveys may care more about the topic than those who do not, so the results may misrepresent the 420 non-responders.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

WORDING BIAS

A survey asks two random halves of a school the same issue in different words. Version A: "Do you support limiting phone use during class to improve learning?" Version B: "Do you support banning students from their own phones?" The reported support is 71% for A and 38% for B. Explain what this tells you.



Same issue, different wording

SOLUTION

The only thing that differs between the groups is the wording, because each half was chosen at random.

The 33 percentage point gap ($71 - 38 = 33$) is therefore evidence that wording influences answers. Version A includes a positive reason ("to improve learning"), which encourages agreement, while Version B uses the harsher word "banning."

A neutral version such as "Should phone use during class be restricted, allowed, or left to each teacher?" would reduce wording bias.

SAMPLE SIZE DOES NOT CURE BIAS

A website invites visitors to vote on whether the school should adopt a four-day week. Over a weekend 9,400 visitors vote and 82% say yes. A random sample of 150 students from the roster finds 57% yes. A classmate argues the 9,400 votes must be more accurate. Respond.

SOLUTION

The web poll is a voluntary response sample. Students who care strongly about a four-day week are the ones who bother to vote, and the poll may also be open to people outside the school, so it is biased toward yes.

The bias is systematic, so adding more respondents just gives a very precise estimate of the wrong quantity.

The random sample of 150 is drawn from the roster (a complete frame), so each student has an equal chance and the sample is much more trustworthy despite its smaller size. Its statistic $\hat{p} = 0.57$ is a reasonable estimate of the school proportion p, with some sampling variability.

A larger sample helps only when the method is unbiased.

Write It Like a 5

Prompt: A restaurant chain wants to know what proportion of its customers are satisfied. A manager leaves comment cards at the register with a box labeled "Tell us how we can do better," and 64 of the 3,200 customers that week return them. About 70% of the returned cards are complaints. Identify the sampling method, describe one source of bias and explain how it affects the estimate of the proportion of dissatisfied customers.

Model Response

The sampling method is voluntary response because customers decide for themselves whether to fill out a card. Only 64 of 3,200 customers (2%) responded, and customers who are upset are more motivated to write a comment than customers who are satisfied, especially with a prompt that asks how the chain can do better. Therefore the cards over-represent unhappy customers, and the statistic of 70% complaints will likely overestimate the proportion of all customers who are dissatisfied. To reduce bias, the manager could select customers at random from receipts or loyalty records and follow up with those chosen.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

(1) Correctly names voluntary response (not just "bias"). (2) Links the bias to the reason people respond, in the context of the restaurant. (3) States the direction: overestimates the proportion dissatisfied. (4) Optionally offers a random-selection fix. A response that names the bias but gives no direction or no reason earns only partial credit.

Coach's Playbook

1. Underline the group the question is really about. That is the population.
2. Ask who was actually measured. That is the sample. If they do not look like the population, suspect bias.
3. Decide whether a number describes the population (parameter) or the sample (statistic).
4. Ask how individuals ended up in the sample: chosen at random, volunteered, easy to reach, or left off the frame?
5. Name the bias type, then write the sentence: "This method will likely [overestimate / underestimate] ___ because ___."
6. If asked for a fix, suggest random selection from a complete frame and neutral wording.

Common traps

1) Writing "the sample is biased" without naming the type, the reason and the direction. 2) Saying a bigger sample reduces bias; it only reduces random variability. 3) Mixing up nonresponse (chosen but did not answer) and voluntary response (self-selected). 4) Calling a statistic a parameter because the number sounds exact.

Mistake Clinic

Spot the error

A student writes: "A poll of 5,000 people is large enough that bias is not a concern." Where it fails: bias comes from the method, not the size. If the 5,000 were all gathered at one shopping mall, the sample would still miss people who never shop there. Correct: "Sample size controls random variability, not bias. We must check how the 5,000 were selected."

Spot the error

A student says: "The mean of 4.5 hours for our class of 28 is the parameter because it is the exact mean." Where it fails: exact does not mean parameter. The 28 students are a sample, so 4.5 is a statistic. Correct: "4.5 hours is a statistic, $\bar{x}$. The parameter μ is the mean for the whole population of band students."

Spot the error

A student identifies a mailed survey with a 15% return as "voluntary response." Where it fails: the households were chosen by the researchers, then most failed to respond, which is nonresponse. Voluntary response applies when people volunteer without being selected. Correct: "Nonresponse bias, because the 85% who did not return the survey may differ from those who did."

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Newsrooms and city agencies publish margins of error alongside polls, but those numbers only describe random variability. A poll gathered from an app's users cannot report a trustworthy margin of error for "all residents" because the method has bias that no formula can measure. Skilled readers of data first ask how the sample was chosen, then look at the number.

Challenge Corner

A school has 1,500 students: 600 freshmen and sophomores and 900 juniors and seniors. The principal surveys students who stay for after-school tutoring, which only 10% of underclassmen and 40% of upperclassmen attend. The principal asks whether they would pay for a late bus. Suppose truly 70% of underclassmen and 30% of upperclassmen would pay. (a) Find the true school proportion. (b) Explain why the tutoring sample is biased and in which direction. Solution: (a) Underclassmen who would pay: $0.70(600) = 420$. Upperclassmen: $0.30(900) = 270$. Total 690 of 1500, so $p = 0.46$. (b) In tutoring, upperclassmen make up $0.40(900) = 360$ of $0.10(600) + 360 = 420$ students, about 86% of the pool, so the sample over-represents the upperclassmen group who are less likely to pay. The expected sample proportion is $(0.70(60) + 0.30(360))/420 = (42 + 108)/420 \approx 0.357$, an underestimate of 0.46.

Practice

PRACTICE

Populations, Samples, and Sources of Bias

Show your work. A calculator is allowed. Justify conclusions in context.

1. A county health office selects 400 residents from the county's list of registered voters and asks each about vaccination status. The proportion of vaccinated residents among all 52,000 county residents is unknown. Which statement is correct?
   - **(A)** The proportion vaccinated in the 400 selected is a statistic, and the proportion in all 52,000 residents is a parameter.
   - **(B)** The proportion vaccinated in the 400 selected is a parameter, and the proportion in all 52,000 residents is a statistic.
   - **(C)** Both proportions are statistics because they are percentages.
   - **(D)** Both proportions are parameters because they describe vaccinated people.
2. An online store emails a survey to all 2,000 customers on its list, and 130 reply. The store reports that 88% of customers are satisfied. The main concern is
   - **(A)** voluntary response, because the store did not choose which customers to email
   - **(B)** nonresponse, because only a small part of those chosen replied and they may differ from those who did not
   - **(C)** undercoverage, because 130 is less than 2,000
   - **(D)** response bias, because 88% is a high percentage
3. Which survey method is most likely to suffer from undercoverage when the population is all adults in a city?
   - **(A)** Mailing a survey to a random sample of residential addresses
   - **(B)** Selecting every 20th name from the city's complete voter and resident list
   - **(C)** Calling numbers randomly selected from the printed city telephone directory
   - **(D)** Visiting randomly chosen street blocks and interviewing everyone who lives there
4. A researcher wants to know how many hours per week teenagers spend gaming. She interviews teens at an arcade on a Saturday afternoon. This method is an example of
   - **(A)** convenience sampling that will likely overestimate gaming time
   - **(B)** voluntary response that will likely underestimate gaming time
   - **(C)** stratified sampling that will likely overestimate gaming time
   - **(D)** convenience sampling that will likely underestimate gaming time

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. Two random halves of a class are asked: "Should the school cafeteria stop selling sugary drinks?" and "Should the cafeteria stop taking away students' choice of drinks?" Support for the change is 64% and 31%. The best explanation is
   (A) undercoverage, because some students were absent
   (B) nonresponse, because both groups should have the same support
   (C) voluntary response, because the students chose which question to answer
   (D) wording bias, because the random assignment of the question wording isolates its effect
6. Increasing the size of a sample will
   (A) reduce random sampling variability but not remove bias caused by a flawed method
   (B) remove bias but not random variability
   (C) remove both bias and random variability
   (D) have no effect on either
7. A town has 3,000 residents. A pollster records the opinions of the 250 residents who answer the door during a weekday afternoon visit. Which is true?
   (A) The sample proportion is a parameter, and the method is unbiased because 250 is more than 5% of the town.
   (B) The sample proportion is a statistic, and the method is likely to be biased because people home on weekday afternoons are not representative.
   (C) The sample is a census because the pollster visited homes in person.
   (D) The sampling frame is the 250 who answered, so the sample has no bias.
8. Which is an example of response bias?
   (A) Only people who feel strongly about a topic call in to a radio poll.
   (B) A phone survey misses households that have no phone.
   (C) People who are asked about how often they exercise report higher numbers than the numbers shown on their fitness trackers.
   (D) A sample is chosen from the people who are easiest to reach.
9. A sample of 60 students from a school of 1,200 has a mean commute of 18 minutes. The symbol for 18 and the quantity it estimates are, respectively,
   (A) μ and $\bar{x}$ **(B)** $\hat{p}$ and p **(C)** s and σ **(D)** $\bar{x}$ and μ
10. A company has 12,000 employees. The human resources office emails a survey about remote work to 600 employees chosen from the employee directory, and 210 respond. Of those, 168 say they prefer remote work.
    (a) Identify the population, the sample, and the statistic (give its value as a proportion).
    (b) What is the response rate?
    (c) Explain one possible source of bias and its direction.
11. For each situation, name the most likely bias and state its direction.
    (a) A local news site asks readers, "Should the city waste money on a new skate park?" and publishes the percent who say no.
    (b) A school counselor asks students, in person, whether they have ever copied another student's homework.
    (c) A park ranger measures the lengths of fish caught by anglers at the dock to estimate the mean length of all fish in a lake.
12. A student says, "Our school survey got 1,100 responses out of 1,200 students, so it is sure to be accurate." Explain why a high response rate is helpful but does not guarantee accuracy. Give an example of a flaw that could remain.
13. A high school wants to learn the proportion of all its 1,800 students who would join a new robotics club. The principal considers three plans: (i) ask students in the cafeteria at lunch on one day, (ii) post a link on the school website, (iii) obtain a list of all students and contact a randomly chosen 150 by email, following up with those who do not reply.
    (a) Which plan is best? Explain.
    (b) Describe the weakness of each of the other two plans.
14. Write two versions of a survey question about school start times, one that could produce wording bias and one that is neutral. Explain how the first would likely change the results.
15. A town has 10,000 residents. A national survey claims that 60% of all residents in the town own a bicycle. A town club surveys 400 randomly chosen residents from a complete list and finds 228 who own a bicycle.
    (a) Compute the sample proportion and say whether it is a parameter or a statistic.
    (b) Does the result prove the claim is false? Explain using the idea of sampling variability.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

4.2 Sampling Methods: SRS, Stratified, Cluster, Systematic

Learn how simple random, stratified, cluster, systematic and multistage samples are chosen, how to carry out the randomization with digits or a calculator, and how to tell the designs apart.

Why This Matters

Once you know that bias comes from the way a sample is chosen, the next question is: how should we choose? Random selection is the answer, but "random" comes in several flavors, and the AP exam expects you to name each one, describe how to carry it out, and explain why you would choose it.

Students lose points by describing a sampling method as "random" without saying how the randomness is produced, or by calling a cluster sample a stratified sample. Both are Exam Unit 1 skills (Collecting Data), and a free-response question often asks you to "describe a method" that uses random digits, a random number generator, or a table.

Core Ideas

A **simple random sample (SRS)** of size n is a sample chosen so that every group of n individuals in the population has the same chance of being the sample. Two things follow: every individual has the same chance of being selected, and every *combination* of individuals is equally likely. Methods that give each individual an equal chance without giving every group an equal chance (such as picking every 10th name) are not SRS.

How to select an SRS (describe it this way on the exam)

1. Label every individual in the frame with a distinct number, using the same number of digits (for example 01 to 40, or 001 to 650).
2. Use a random number generator or a table of random digits. With a table, read digits in groups of the label length, starting at a stated line.
3. Ignore numbers that are not labels and ignore repeats, since an individual cannot be chosen twice.
4. Stop when you have n different labels; the individuals with those labels are the sample.

With a calculator: generate random integers between 1 and N, and discard repeats, until you have n different ones.

Here is a short line of random digits. We will use it to choose an SRS of 5 students from a roster of 40, labeled 01 to 40. We read two digits at a time.

Line 1 of random digits (grouped in fives)
65326 94846 90307 22524 99019 06431 57280 11220

Reading pairs: 65, 32, 69, 48, 46, 90, 30, 72, 25, 24, 99, 01. Discard 00 and anything above 40, and skip repeats. The first five valid labels are 32, 30, 25, 24, 01, so the sample consists of the students with those labels.

Simple random sampling is the foundation, but it is not always practical or efficient. Other probability-based designs split or organize the population first.

Probability sampling designs

Stratified random sample: divide the population into groups of similar individuals called **strata**, then take a separate SRS from every stratum and combine them. Use it when strata differ on the variable measured; it reduces variability of the estimate.

Cluster sample: divide the population into groups called **clusters** (often naturally occurring and each a mini-population), choose an SRS of clusters, and measure every individual in the chosen clusters. Use it to save time and money.

Systematic random sample: choose a random starting point among the first k individuals, then select every kth individual, where $k \approx N/n$.

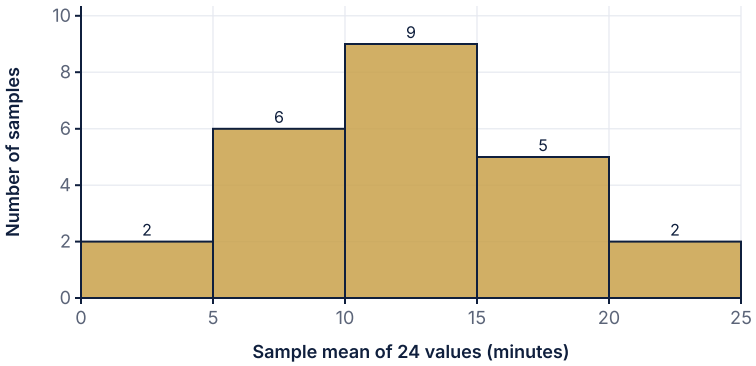
Multistage sample: use several stages of random selection, for example SRS of schools, then SRS of classes within schools, then SRS of students within classes.

The most commonly confused pair is stratified versus cluster. **Strata** are chosen to be *different from each other* and *similar inside*; you sample from every stratum. **Clusters** are chosen to *look like the whole population* inside; you sample only some of the clusters, and then take everyone in them.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Design	Divide into	What you randomly select	Who is measured
SRS	No groups	Individuals	The chosen individuals
Stratified	Strata (similar inside)	An SRS from every stratum	The chosen individuals in every stratum
Cluster	Clusters (mini-populations)	An SRS of clusters	Everyone in chosen clusters
Systematic	A list in order	One random start, then every kth	Those at positions start, start+k, ...
Multistage	Several levels	Random selection at each level	Individuals at the last stage



Made-up results of repeated random samples show sampling variability: statistics cluster around the true value but do not all equal it.

Proportional allocation is a common stratified rule: if a stratum makes up 20% of the population, it supplies 20% of the sample. A stratified design can give a more precise estimate than an SRS of the same size when the strata differ on the variable of interest.

Worked Examples

CHOOSING AN SRS WITH RANDOM DIGITS

A teacher wants an SRS of 5 students from a roster of 40 (labeled 01 to 40). Use Line 1 above: 65326 94846 90307 22524 99019 06431 57280 11220. Describe and carry out the selection.

Line 1 of random digits
65326 94846 90307 22524 99019 06431 57280 11220

SOLUTION

Label the students 01 to 40. Read the digits in pairs from the start of Line 1.

Pairs: 65, 32, 69, 48, 46, 90, 30, 72, 25, 24, 99, 01.

Skip pairs that are 00 or above 40 and any repeats. This leaves the labels 32, 30, 25, 24, 01.

The sample is the five students with these labels.

SYSTEMATIC SAMPLE

A school has 900 lockers numbered 1 to 900. The janitor wants to check 30 of them using a systematic sample. Find k and list the first four lockers if the random start is 17.

SOLUTION

$k = N/n = 900/30 = 30$.

Pick a random start between 1 and 30; here it is 17.

The sample is lockers 17, 47, 77, 107, ... up to 887, which is exactly 30 lockers.

Every locker has an equal chance of being chosen (1 in 30), but lockers 17 and 18 can never both be in the sample, so this is not an SRS.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

STRATIFIED SAMPLE WITH PROPORTIONAL ALLOCATION

A school of 1,200 students has 480 freshmen, 360 sophomores, 240 juniors and 120 seniors. The student council wants a stratified random sample of 100 students with proportional allocation. How many students come from each grade, and how should they be chosen?

Grade	Students	Share	Sample size
Freshmen	480	40%	40
Sophomores	360	30%	30
Juniors	240	20%	20
Seniors	120	10%	10
Total	1,200	100%	100

SOLUTION

Each grade's share of the sample equals its share of the school: $480/1200 = 40\%$, $360/1200 = 30\%$, $240/1200 = 20\%$, $120/1200 = 10\%$.

So 40 freshmen, 30 sophomores, 20 juniors and 10 seniors, for a total of 100.

Within each grade, label the students and use random digits to choose an SRS of that size.

CLUSTER OR STRATIFIED?

A city has 60 neighborhoods. (a) A planner randomly selects 5 neighborhoods and surveys every household in them. (b) Another planner groups the neighborhoods by income level (low, middle, high) and surveys a random sample of households within each level. Name each design and say when it works well.

SOLUTION

(a) Cluster sample. The neighborhoods are the clusters, a random sample of clusters is chosen, and everyone in them is surveyed. It is efficient because workers travel to only 5 places, and works best when each neighborhood resembles the whole city.

(b) Stratified sample. Income level defines the strata, and every stratum gets its own random sample. It works well when income is related to the variable measured, because each income group is guaranteed representation.

The difference: strata are all sampled from; only some clusters are sampled and then everyone inside is measured.

MULTISTAGE SAMPLING

A state education office wants to survey Grade 8 students. It randomly chooses 20 of the state's 300 middle schools, then randomly chooses 3 Grade 8 classes in each chosen school, then randomly chooses 10 students in each chosen class. Describe the design and compute the sample size.

SOLUTION

This is a multistage sample with three stages: schools, classes, then students.

Sample size: $20 \times 3 \times 10 = 600$ students.

It is practical because it avoids a list of every Grade 8 student in the state; lists are needed only for the chosen schools and classes.

IS IT AN SRS?

A coach has 20 players, 10 who play offense and 10 who play defense. She randomly picks 3 offense players and 3 defense players to try a new drill. (a) Is every player equally likely to be chosen? (b) Is the sample of 6 an SRS of the 20 players?

SOLUTION

(a) Yes. Each offense player has a $3/10$ chance and each defense player also has a $3/10$ chance.

(b) No. In an SRS of size 6, every group of 6 players is equally likely, including a group of 6 offense players. This design makes that group impossible, so it is a stratified sample, not an SRS.

Number of equally likely groups in an SRS of 6 from 20 is $\binom{20}{6} = 38760$. The stratified design allows only $\binom{10}{3}\binom{10}{3} = 14400$ groups.

Write It Like a 5

Prompt: A music festival expects 4,800 attendees: 3,000 with general tickets, 1,200 with VIP tickets and 600 with student tickets. The organizers want to survey 160 attendees about parking. Describe a sampling method that makes sure all three ticket types are represented and explain how you would carry it out using a random number generator.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Use a stratified random sample with ticket type as the strata, because attitudes about parking may differ among general, VIP and student attendees. Proportional allocation gives general $3000/4800 \times 160 = 100$ attendees, VIP $1200/4800 \times 160 = 40$, and student $600/4800 \times 160 = 20$, for a total of 160. Within each ticket type, number the attendees from 1 to the stratum size (1 to 3000, 1 to 1200, and 1 to 600). Use the random number generator to produce random integers in that range, ignoring repeats, until 100, 40 and 20 different numbers are selected for the three strata. The attendees with these numbers form the sample.

What the grader looks for

(1) Names stratified sampling and identifies the strata. (2) Explains why strata help or why they were chosen. (3) States the sample size from each stratum with correct calculations. (4) Describes the random procedure: labels, generator or digits, ignoring repeats. (5) Does not describe a cluster or convenience sample by mistake.

Coach's Playbook

1. Ask: is the whole population divided into groups? If no, the design is an SRS or systematic.
2. If yes, ask: will I sample from every group (stratified) or only from some groups and measure everyone in them (cluster)?
3. For any random method, state the labels, the digits or generator, and the rule for repeats and out-of-range numbers.
4. For systematic, compute $k = N/n$ and use a random start between 1 and k.
5. Calculator tip: use the random integer command with the lower limit 1 and the upper limit N; keep going until you have n different values.

Common traps

1) Calling a method "random" without describing the randomization. 2) Treating clusters as strata. 3) Forgetting to label all individuals using the same number of digits, so 5 and 05 are confused. 4) Claiming a systematic sample is an SRS even though it prevents some groups from occurring.

Mistake Clinic

Spot the error

A student writes: "I will use a stratified sample by randomly choosing 3 of the 12 classrooms and surveying every student in them." Where it fails: the classrooms are clusters, since only some are chosen and everyone in them is measured. Correct: "This is a cluster sample. A stratified sample would take an SRS of students from every classroom."

Spot the error

A student says: "To choose an SRS of 8 from 90 students, number them 1 to 90 and read digits two at a time, so 5 stands for 5." Where it fails: the labels must have the same number of digits, otherwise single and double digit labels are confused, and 100 or 00 are not labels. Correct: "Label the students 01 to 90, read two digits at a time, skip 00 and 91 to 99, and skip repeats."

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student says: "Picking every 10th name from the roster is an SRS because each name is equally likely." Where it fails: equal chance for each person is necessary but not enough; an SRS requires that every group of size n has the same chance. Names 3 and 4 cannot both be in a every-10th sample. Correct: "This is a systematic random sample."

Real-World Connection

National surveys rarely use a pure SRS. They use stratified, cluster and multistage designs: choose regions, then neighborhoods, then households, then one adult per household. Each stage is randomized and weights account for unequal chances. The same ideas guide a quality inspector who samples cartons from each shift (strata) before opening every box in a few cartons (clusters).

Challenge Corner

A factory makes 6,000 widgets per day on three machines: Machine A makes 3,000, B makes 2,000 and C makes 1,000. Defect rates are about 6%, 3% and 1%. (a) Explain why stratifying by machine could give a more precise estimate than an SRS of 300 widgets. (b) Find the stratified sample sizes with proportional allocation and the expected number of defects in the sample. Solution: (a) The strata differ strongly on the variable (defect rate), so ensuring each machine contributes its proper share removes the chance variation in how many items come from each machine. (b) Sample 150 from A, 100 from B and 50 from C. Expected defects: $150(0.06) + 100(0.03) + 50(0.01) = 9 + 3 + 0.5 = 12.5$, which is $12.5/300 \approx 4.2\%$ overall.

Practice

PRACTICE **Sampling Methods: SRS, Stratified, Cluster, Systematic**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which statement best describes a simple random sample of 30 students from a school?
   (A) Every group of 30 students has the same chance of being the sample.
   (B) Every student has the same chance of being selected, and that is all that is required.
   (C) The sample contains the same number of students from each grade.
   (D) Students are selected from the first 30 names on a random ordering of homeroom classes.
2. A county has 40 high schools. A researcher randomly selects 6 schools and surveys every senior at those schools. This is a
   (A) stratified random sample (B) cluster sample (C) simple random sample
   (D) systematic sample
3. A roster has 50 students labeled 01 to 50. Using the digits 45333147893288040723 read in pairs from the left, the first 4 students for an SRS are
   (A) not determined, since digits must be read in singles
   (B) the first four pairs regardless of whether they are valid labels
   (C) 45, 33, 31, 47
   (D) the first four distinct pairs between 00 and 99
4. A system selects every 25th customer from an ordered list of 2,000 customers after a random start from 1 to 25. How many customers are in the sample and is it an SRS?
   (A) 80 customers, and it is an SRS because each customer is equally likely
   (B) 25 customers, and it is an SRS because the start is random
   (C) 25 customers, and it is not an SRS because the start is random
   (D) 80 customers, and it is not an SRS because not every group of 80 is possible

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. A school has 800 day students and 200 boarding students. The principal selects an SRS of 40 day students and an SRS of 10 boarding students. This design is
   (A) a stratified random sample with proportional allocation
   (B) a cluster sample with 2 clusters
   (C) an SRS of 50 students
   (D) a multistage sample with 50 stages
6. Which situation is a good reason to use a stratified sample rather than an SRS?
   (A) The population has no list of individuals but has a list of neighborhoods
   (B) Opinions probably differ greatly among age groups, and we want precise estimates from each
   (C) Interviewers want to travel to as few locations as possible
   (D) We prefer not to use random numbers
7. A coach numbers 36 players 01 to 36 and wants a random group of 4 to test shoes. Which procedure is an SRS?
   (A) Read two-digit labels from a random digit table, ignoring labels above 36, 00 and repeats, until 4 labels are found
   (B) Pick the first four players who arrive at practice
   (C) Pick every 9th player after a random start from 1 to 9
   (D) Pick one player from each of four position groups at random
8. In a multistage sample, a surveyor chooses 8 of 50 towns at random, then 5 of the streets in each chosen town at random, then 6 households on each chosen street. The number of households surveyed is
   (A) 19 **(B)** 48 **(C)** 240 **(D)** 1,500
9. Which feature distinguishes a cluster sample from a stratified sample?
   (A) In a cluster sample every group is sampled, and a random subset of individuals is measured.
   (B) In a cluster sample the groups must differ greatly from each other.
   (C) In a cluster sample no random selection is used.
   (D) In a cluster sample only some groups are sampled, and everyone in them is measured.
10. A library has 650 members labeled 001 to 650. Describe how to use a table of random digits to select an SRS of 12 members, including what to do with repeats and with numbers that are not labels.
11. A high school has 500 students in each of four grades. The principal wants to survey 100 students. (a) Describe a stratified sample with proportional allocation. (b) Describe a cluster sample using homerooms (20 homerooms of 25 students). (c) Explain one advantage of each.
12. A grocery store has 40 checkout lanes in a long list and wants to inspect 8 of them with a systematic sample. (a) Find k. (b) Explain how to choose a random start. (c) List the lanes inspected if the start is 3. (d) Explain why this is not an SRS.
13. The principal says: "I chose 10 classrooms at random and surveyed 5 random students in each classroom, so my sample is a stratified sample." Name the design correctly, defend your answer, and describe the stages.
14. A park has three trail zones with 600, 300 and 100 visitors on a typical day. A ranger will sample 50 visitors. (a) Compute the proportional allocation. (b) Explain why the ranger might choose a larger share than proportional from the smallest zone.
15. A survey firm wants a random sample of 1,000 adults from a state. Propose a multistage design with at least three stages and explain why each stage uses randomization.

4.3 Observational Studies, Experiments, and Confounding

Learn to tell sample surveys, observational studies and experiments apart, to name confounding and lurking variables precisely, and to word conclusions that match the way the data were collected.

Why This Matters

"Does this study show that X causes Y?" is one of the most important questions in statistics, and the answer depends almost entirely on how the data were collected, not on the size of the numbers. Good statistical thinkers can look at a description of a study and say what kind it is, what conclusion it supports, and what could be fooling us.

Students often write "correlation is not causation" and stop. The exam wants more: name a specific **confounding variable**, explain how it is linked to both the explanatory and response variables, and say why that makes it impossible to separate their effects. This is Exam Unit 1 (Collecting Data), and it is the setup for experimental design in the next lesson.

Core Ideas

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A **sample survey** collects information from a sample of a population, typically by asking questions, and does not try to change anything. A **study** that tries to show how one variable relates to another comes in two basic types.

Observational study versus experiment

Observational study: researchers observe and measure individuals or record how they behave but do **not** impose any treatment. The people choose their own behavior or conditions.

Experiment: researchers deliberately impose a **treatment** on **experimental units** (called **subjects** if they are people) in order to observe the response.

Explanatory variable: the variable that may explain or influence changes in the response variable. In an experiment, the explanatory variable is what the researchers control.

Response variable: the outcome measured to see the effect of the treatments.

Only a well designed experiment can provide convincing evidence of cause and effect.

Why is an observational study not enough? Because the groups being compared might differ in more ways than the one we care about. Suppose a school tracks students and finds that those who eat breakfast have higher quiz scores. Students who eat breakfast may also sleep more, come from households with more routine, or be more organized, and any of these could be responsible for the higher scores.

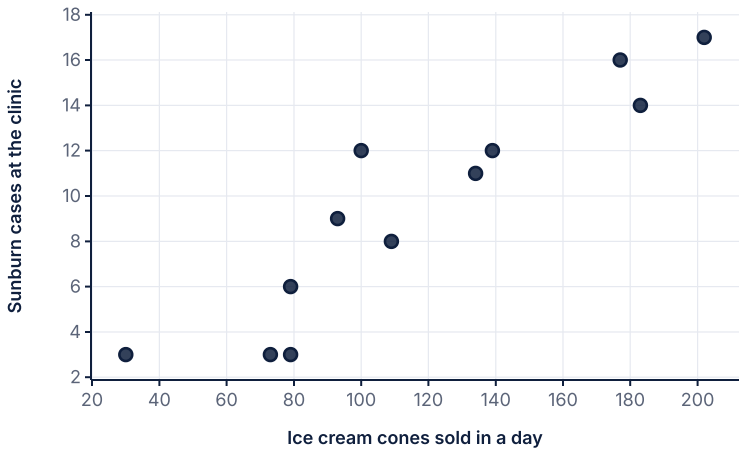
Confounding and lurking variables

A **lurking variable** is a variable that is not among the explanatory or response variables in a study but may influence the response, and may be related to the explanatory variable.

Confounding occurs when two variables (explanatory or lurking) are associated in such a way that their effects on a response variable cannot be distinguished from each other.

To describe a confounder, state: (1) the lurking variable, (2) how it is associated with the explanatory variable, and (3) how it could influence the response.

The scatterplot below uses invented daily data: the number of ice cream cones sold at a stand and the number of sunburn cases treated at a nearby clinic. The two are strongly associated, but neither causes the other. A hot sunny day (the lurking variable) pushes up both.



Made-up data: both counts rise on hot days.

An observational study can be done in two time directions. A **retrospective** study looks backward: it examines existing data or records on events that have already taken place (for example, health records from the past ten years). A **prospective** study identifies individuals now, then follows them forward in time and records outcomes as they occur. Prospective studies usually give more reliable measurements, but both types suffer from confounding.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Type of investigation	Researchers impose a treatment?	Can show cause and effect?	Typical wording
Sample survey	No	No (describes the population)	"We asked a random sample of 400 residents..."
Observational study	No	No, because of confounding	"We compared people who already do X with those who do not."
Experiment	Yes	Yes, if well designed with random assignment	"We randomly assigned 60 volunteers to..."

Be careful with the word "sample." A study can use a random sample *and* be observational, or use volunteers *and* be an experiment. Whether it is an experiment depends on whether a treatment is **imposed**, not on how the individuals were found.

Worked Examples

CLASSIFY THE STUDY

A researcher records the number of hours each of 80 high school athletes sleeps on school nights and their reaction times on a computer test. Another researcher asks 80 athletes to sleep 6 hours on nights in one week and 9 hours in another week, and measures reaction time after each week. Classify each study and name the explanatory and response variables.

SOLUTION

The first researcher only records existing sleep habits, so it is an observational study.

The second researcher assigns the sleep conditions, so it is an experiment (with 6 hours and 9 hours as the two treatments).

In both, the explanatory variable is hours of sleep and the response variable is reaction time.

SPOTTING THE LURKING VARIABLE

A study of 2,000 adults reports that those who take a daily multivitamin have fewer sick days per year than those who do not. Identify a possible confounding variable and explain how it could be responsible for the observed difference.

SOLUTION

Possible confounder: overall health-conscious habits, such as exercising regularly and eating well.

People who choose to take multivitamins are more likely to also exercise and eat well (association with the explanatory variable), and these habits themselves reduce illness (influence on the response).

Therefore we cannot tell whether the multivitamin or the healthy habits caused the lower number of sick days.

RETROSPECTIVE OR PROSPECTIVE?

(a) A team of researchers examines the school records of 600 graduates from the last ten years to compare attendance in Grade 9 with graduation outcomes. (b) A different team recruits 600 current Grade 9 students, records their attendance this year, and follows them until graduation. Classify each.

SOLUTION

(a) Retrospective: the data already exist and the researchers look back.

(b) Prospective: the individuals are identified now and followed forward in time.

Both are observational. Neither can establish that attendance causes the outcome because factors like family support are confounded with attendance.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

CONFOUNDING WITH A TABLE

A report compares two groups of 200 commuters each. Those who bike have a lower average blood pressure than those who drive. The table shows additional data about the groups.

Group	Mean age (years)	Exercise weekly	Mean systolic pressure
Bike commuters (200)	34	78%	118
Drive commuters (200)	49	41%	127

SOLUTION

Bikers are younger on average (mean age 34 versus 49) and a larger proportion exercise weekly (78% versus 41%).

Both age and exercise are associated with commuting type and are known to affect blood pressure, so they are confounded with commuting by bike.

The difference in blood pressure cannot be attributed to biking alone. An experiment, in which commuters are randomly assigned to bike or drive, would be needed to isolate the effect.

EXPERIMENT OR NOT?

A teacher gives one class a new study app for a month and gives another class of similar students the usual notes. Test scores are higher in the app class. The teacher says the app caused the improvement. Evaluate the claim.

SOLUTION

The teacher imposed a treatment (app versus notes), so it is an experiment, but there is no random assignment of students to the treatments: the classes were already formed.

Class differences (the teacher's style, time of day, student motivation) are confounded with the app. For example, the app class may meet in the morning when students are more alert.

Without random assignment, the data support an association but not a cause-and-effect conclusion. Randomly assigning individual students to the app or the notes would balance such factors.

EXPLAINING WHY A DESIGN CANNOT SHOW CAUSATION

Public records show that cities with more fire trucks sent to a fire have larger fire damage costs. A reporter concludes that sending more trucks increases damage. Explain the flaw.

SOLUTION

This is an observational study with no controlled treatment.

Lurking variable: the size of the fire. Larger fires lead the department to send more trucks (association with the explanatory variable) and also cause more damage (influence on the response).

So size of fire is confounded with the number of trucks. The data do not show that trucks cause damage.

Write It Like a 5

Prompt: A researcher surveys 1,000 randomly selected teenagers and finds that those who report playing video games for more than 3 hours per day have lower average grades than those who play less. A news headline reads, "Video games lower grades." (a) Is this an experiment or an observational study? (b) Is the headline justified? Explain, naming a possible confounding variable.

Model Response

(a) This is an observational study. The researcher did not assign gaming times; teens reported what they already do, so no treatment was imposed. (b) The headline is not justified. A possible confounding variable is the amount of time teens spend on homework or sleep. Teens who play games for many hours may have less time for homework, and less homework time is also related to lower grades, so the effects of gaming and homework time cannot be separated. The random sample lets us generalize the association to teens in the population, but without random assignment to different gaming times, we cannot conclude that gaming causes lower grades. Only a well designed experiment could do that.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

(1) Identifies observational study and justifies it with no imposed treatment. (2) States that the conclusion of causation is not justified. (3) Names a specific lurking variable. (4) Explains the link to both gaming and grades. (5) Optionally notes the random sample supports generalizing the association.

Coach's Playbook

1. Ask: did the researchers impose a treatment? Yes means experiment, no means observational study or sample survey.
2. Name the explanatory and response variables in plain language.
3. If observational, brainstorm a lurking variable that is related to both variables. Check that it makes sense in the context given.
4. Write the full chain: lurking variable is associated with the explanatory variable, and affects the response.
5. For retrospective/prospective, ask whether the data already existed (retrospective) or the people are followed from now on (prospective).
6. State your conclusion using the right verb: "associated with" for observational, "caused" only for a well designed experiment.

Common traps

1) Calling a study an experiment because it has volunteers, groups or a comparison. 2) Naming a lurking variable that is not related to the explanatory variable, so it does not explain the difference. 3) Writing "other factors" without naming any. 4) Confusing the lurking variable with the response variable itself, such as saying that the confounder is "lower grades."

Mistake Clinic

Spot the error

A student writes: "This is an experiment because the researchers compared people who exercise with people who do not." Where it fails: the researchers did not assign exercise; people chose it. Correct: "This is an observational study, because no treatment was imposed. Exercise is the explanatory variable."

Spot the error

A student writes: "A possible confounder is that some students are smarter." Where it fails: the claim is not connected to the explanatory variable and is vague. Correct: "Students who choose to take the advanced course may already have strong study skills, which are associated with taking the course and also raise test scores, so study skills are confounded with course choice."

Spot the error

A student says: "A random sample was used, so the study shows cause and effect." Where it fails: random sampling controls bias in who is selected, but cause and effect needs random assignment of imposed treatments. Correct: "A random sample lets us generalize to the population, but without random assignment of treatments we only have evidence of association."

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Health headlines often come from large observational studies that follow people for years. Researchers then adjust for factors like age and smoking, but they can never adjust for every lurking variable. That is why important medical claims are tested again with randomized trials, where chance, not choice, determines who gets the treatment.

Challenge Corner

A town finds that neighborhoods with more libraries per resident have lower crime rates. Suppose the town has three kinds of neighborhoods. In high-income neighborhoods (400 per 1,000 residents have a library card, crime rate 20 per 1,000), middle-income (250, crime rate 35) and low-income (100, crime rate 60). (a) Explain why library density might not reduce crime. (b) Describe an experiment in a different context that could test whether a program reduces crime. Solution: (a) Income is a lurking variable: higher-income neighborhoods have more libraries and, for many other reasons such as better lighting, schools and jobs, lower crime. The pattern is across income groups (library card use falls from 400 to 100 as crime rises from 20 to 60), so the libraries cannot be separated from income. (b) Randomly assign comparable neighborhoods to receive an after-school program or not, then compare crime rates after a year; random assignment balances income and other factors across the groups.

Practice

PRACTICE **Observational Studies, Experiments, and Confounding**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Researchers ask 300 adults how many cups of tea they drink per day and then compare their reported blood pressure. This is
   (A) an observational study, because no treatment is imposed
   (B) an experiment, because there are two variables
   (C) an experiment, because the sample was random
   (D) a census, because everyone was asked a question
2. Which is a retrospective study?
   (A) Following 500 healthy adults for 10 years and recording illnesses as they arise
   (B) Using hospital records from the past 15 years to compare outcomes of patients with and without a certain habit
   (C) Randomly assigning 40 adults to two diets and measuring weight after 3 months
   (D) Surveying 200 adults today about their current exercise
3. A study finds that students who use a tutor have lower grades than those who do not. Which is the most reasonable explanation that involves a lurking variable?
   (A) Tutors make students worse at the subject.
   (B) The study should have used a larger sample.
   (C) Students who struggle in class are more likely to seek a tutor.
   (D) Tutoring takes time away from other activities.
4. In an experiment, the experimental units are
   (A) the treatments that are compared
   (B) the variable that is measured at the end
   (C) the larger population of interest
   (D) the individuals or objects to which treatments are assigned
5. A clinic gives half of 100 volunteer patients a new stretching routine and half the usual care by their own choice. Patients with stronger motivation choose the routine and recover faster. Which concept best describes the difficulty?
   (A) Confounding between motivation and the treatment
   (B) Nonresponse
   (C) Undercoverage
   (D) Response bias

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. A researcher compares the test scores of students in two classes, one that uses a new textbook and one that uses the old one. The classes were not randomly formed. The strongest conclusion that can be drawn is that
   (A) the textbook caused any difference in scores
   (B) any difference may be due to the textbook or to preexisting differences between classes
   (C) there is no difference in the population
   (D) random selection removed all bias
7. Which statement about a sample survey is correct?
   (A) It describes the population but does not impose treatments.
   (B) It can establish cause and effect if the sample is large.
   (C) It always uses random assignment.
   (D) It is a type of experiment.
8. In a study of whether the temperature of a classroom affects test scores, which is the response variable?
   (A) Temperature **(B)** Classroom **(C)** Test score
   (D) Students
9. A study finds that people who own a pet dog walk more steps per day than people without dogs. Which statement is the best interpretation?
   (A) Owning a dog causes people to walk more.
   (B) Walking more causes people to buy a dog.
   (C) The result is invalid because the study compared two groups.
   (D) Dog ownership is associated with more steps, but because the study is observational, we cannot conclude that owning a dog causes more walking.
10. A coach records how many protein shakes each of 60 athletes drinks per week and their change in muscle mass over a season. The coach concludes shakes increase muscle mass.
    (a) Is this an experiment or an observational study? Explain.
    (b) Name a lurking variable and explain how it could confound the result.
    (c) Describe in one or two sentences how to change the study so it could support a cause-and-effect conclusion.
11. Identify the explanatory variable, response variable and whether the study is an experiment or observational: "Fifty volunteers were each randomly assigned to listen to classical music or to silence while solving puzzles, and the number of puzzles solved in 20 minutes was recorded." Explain what additional step would make the conclusions more convincing.
12. A researcher examines the medical records of 2,500 adults over the last twenty years and finds that those who ate fish more than twice a week had lower rates of a certain heart condition. (a) Is this retrospective or prospective? (b) Is it an experiment? (c) Describe a confounder.
13. Two school districts are compared: District A recently adopted a later start time and District B did not. After one year, District A's attendance rate is higher. Explain why we cannot conclude that the start time caused the higher attendance, naming two possible confounding variables.
14. Explain the difference between random selection and random assignment. For each, name the kind of conclusion it supports.
15. A report states: "Among 400 teens, those who use social media over 4 hours a day were more likely to report feeling anxious." Write a careful interpretation that uses the right language for this type of study, and describe one confounder that links heavy use and anxiety.

4.4 Designing Experiments: Control, Randomization, Replication, Blocking

Learn the building blocks of a well designed experiment, from control groups and placebos to random assignment, blocking and matched pairs, and how to describe a design in words and with a diagram.

Why This Matters

A well designed experiment is the one tool in statistics that can support a claim of cause and effect. But "well designed" has a precise meaning: a comparison group, random assignment, controlled outside influences and enough subjects. Each piece fixes a specific weakness, and the AP exam asks you to describe a complete design and to justify each feature.

Many students lose points by listing features without tying them to the context ("I would use random assignment") or by confusing blocking with stratifying. This lesson shows how to describe an experiment in words and with a diagram, which is a standard part of Exam Unit 1 free-response answers.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Core Ideas

Vocabulary of experiments

Experimental units: the individuals or objects to which treatments are applied. Subjects: experimental units that are human beings.
Factor (explanatory variable): the variable whose effect we want to study. Levels: the values of a factor.
Treatment: a specific condition applied to the units; it is a single level of one factor or a combination of levels of several factors.
Response variable: the outcome measured after treatments are applied.
Control group: a group that receives no treatment, a standard treatment or a placebo, used for comparison.

Placebo and the placebo effect. A placebo is an inactive treatment that looks the same as the real one. People often show improvement just because they believe they are being treated; this is the placebo effect. A placebo group lets us separate the effect of the active treatment from the effect of expecting to get better.

Blinding. In a **single-blind** experiment, either the subjects or the people who evaluate the response do not know which treatment each subject received. In a **double-blind** experiment, neither the subjects nor those who interact with them or measure the response know who received which treatment. Blinding prevents expectations from changing behavior or measurements.

The three principles of experimental design

Control: keep other variables the same for all groups (same time, same equipment, same instructions) and include a comparison group so differences can be attributed to the treatment.
Random assignment: use chance to decide which units receive which treatment, so that groups are alike before treatment on average, including on lurking variables you did not think of.
Replication: use many experimental units in each group, so the effect of the treatment can be distinguished from chance variation among individuals. (Do the same thing to many units; repeating the whole experiment is also called replication.)

Three common designs organize these principles.

Experimental designs

Completely randomized design: all units are randomly assigned to the treatments with no other grouping.
Randomized block design: units are first separated into blocks of similar units (similar in a way expected to affect the response), then units within each block are randomly assigned to treatments. Comparisons are made within blocks.
Matched pairs design: a special block design in which blocks have size two. Either two similar units are paired and randomly assigned one treatment each, or one unit receives both treatments in random order, with a comparison of the differences.

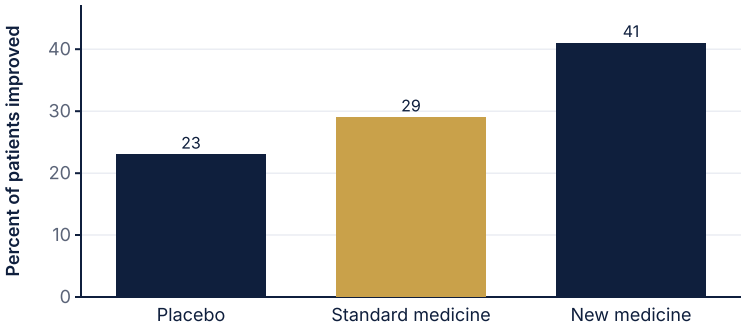
The purpose of blocking is to remove a known source of variability. Strata in sampling and blocks in experiments sound alike, but strata are groups in a population from which we sample, whereas blocks are groups of experimental units that are randomly assigned to treatments.

Design	Diagram in words	Comparison made between
Completely randomized	Units randomly assigned to Group 1 (Treatment A) and Group 2 (Treatment B), then compare responses	Group means or proportions
Randomized block	Separate units into blocks; within each block randomly assign Treatment A and Treatment B; compare within blocks	Treatments within each block
Matched pairs	Each subject receives both treatments in random order (or pairs are randomly split); compare the two responses for each subject	Differences within pairs

Statistical significance. If the differences observed between treatment groups are so large that they would rarely happen by chance alone when the treatments have no effect, we say the results are **statistically significant**. The formal test comes later in the course, but the idea is already useful: random assignment lets us ask whether a difference is larger than the chance variation we would expect from the random splitting of the units into groups.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Invented results of a three-group experiment with 100 patients per group. The placebo group still improves, which is why it is needed as a baseline.

Worked Examples

COMPONENTS OF AN EXPERIMENT

A coach wants to know whether a stretching routine before practice reduces the number of minor injuries. Twenty-four players volunteer. Half are randomly assigned to stretch for 10 minutes before every practice, and the other half do a seated rest for 10 minutes. Injuries are counted over a season. Identify the units, factor, treatments, response and control.

SOLUTION

Experimental units (subjects): the 24 players.

Factor: the pre-practice routine; levels: stretching, seated rest.

Treatments: stretching for 10 minutes; seated rest for 10 minutes.

Response variable: number of minor injuries during the season.

Control group: the seated-rest group; it spends the same time as the stretching group so time is the same for both.

RANDOM ASSIGNMENT WITH DIGITS

Ten subjects labeled 0 to 9 will be assigned to Group A and Group B, 5 each. Use the digits 74852 60574 79380 92752 and read single digits from the left. The first five distinct digits go to Group A; the rest go to Group B.

Random digits
74852 60574 79380 92752

SOLUTION

Digits read: 7 4 8 5 2 (stop when five different digits have appeared).

Skip any digit that has already been used.

Group A: 2, 4, 5, 7, 8.

Group B gets the remaining subjects: 0, 1, 3, 6, 9.

Because each subject is equally likely to land in Group A, the groups should be alike on average.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

COMPLETELY RANDOMIZED DESIGN DIAGRAM

A researcher wants to know whether fertilizer type (Brand X or Brand Y) affects tomato yield, using 40 identical plants. Describe a completely randomized design and give a diagram.

Step 1: Random assignment	Step 2: Treatments	Step 3: Observe
Group 1: 20 plants	Brand X fertilizer	Mean yield per plant
Group 2: 20 plants	Brand Y fertilizer	Mean yield per plant

SOLUTION

Label the plants 01 to 40 and use a random number generator to select 20 different plants for Brand X; the other 20 get Brand Y. Give all plants the same sunlight, water and pot size.

After the growing season, measure the total yield of each plant and compare the mean yields of the groups.

RANDOMIZED BLOCK DESIGN

A farmer has 24 plots, 12 on a sunny hillside and 12 in a shaded valley, and will compare two varieties of wheat. Explain why blocking helps and describe the design.

Block	Randomly assign	Treatment 1	Treatment 2
Hillside (12 plots)	6 and 6	Variety 1 (6 plots)	Variety 2 (6 plots)
Valley (12 plots)	6 and 6	Variety 1 (6 plots)	Variety 2 (6 plots)

SOLUTION

Sunlight strongly affects yield, so differences between hillside and valley plots would add variability to the comparison. Block the plots by location so comparisons are made within similar plots.

Within the hillside block, randomly assign 6 plots to Variety 1 and 6 to Variety 2. Do the same in the valley block.

Compare the yields of the varieties within each block, then combine the information.

MATCHED PAIRS DESIGN

A shoe company wants to compare the durability of two sole materials. Twenty runners each wear one shoe with Material P on one foot and Material Q on the other for 100 miles. Describe the design and why it is a matched pairs experiment.

SOLUTION

Each runner is a block of size two, since both materials are used on the same person, who runs the same distance on the same surface.

For each runner, flip a coin to decide which foot gets Material P (so the left and right foot are not always the same).

Measure the wear on each shoe and compute the difference (P minus Q) for each runner; analyze the 20 differences. Pairing removes runner-to-runner variability such as weight and stride.

BLINDING AND PLACEBO

A school tests whether a herbal tea improves focus. Students receive either tea or a similar-tasting flavored water. A teacher rates each student's focus. Explain how to make it double-blind and why that matters.

SOLUTION

The flavored water is the placebo; it is prepared to look and taste like the tea.

Subjects are blind if they are not told which drink they received; the teacher rating focus is blind if she does not know which students drank tea. An assistant who does not rate focus prepares and hands out coded cups.

Double blinding matters because a teacher who expects improvement could rate tea drinkers more highly, and students who think they have a focus aid may try harder.

Write It Like a 5

Prompt: A researcher believes that listening to a short guided breathing recording before a quiz may improve quiz scores. Thirty volunteer students are available. Describe a completely randomized design to test this, including a control, random assignment, blinding where possible and the comparison to be made. Explain why random assignment is important.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

The 30 students are the experimental units. The treatments are a 5-minute guided breathing recording and a 5-minute recording of neutral background sounds that serves as a control and placebo, so that all students spend the same time with headphones on. Label the students 01 to 30 and use a random number generator to choose 15 different labels for the breathing recording; the other 15 hear the neutral recording. All students take the same quiz in the same room at the same time. The person grading the quizzes is not told which recording each student heard (and students are not told the purpose of the recording), so the experiment is double blind. After grading, compare the mean quiz scores of the two groups. Random assignment is important because it tends to make the two groups alike in ability, test anxiety and other lurking variables, so a difference in mean scores can be attributed to the recording rather than to preexisting differences.

What the grader looks for

(1) Specifies two treatments and a control or placebo. (2) Describes random assignment with a method (labels and random number generator). (3) Controls other variables (same quiz, same time). (4) Mentions blinding and who is blind. (5) States the response and the comparison. (6) Explains that random assignment balances lurking variables and so supports cause and effect.

Coach's Playbook

1. Name the units, the factor, the treatments and the response variable in the context.
2. Decide whether there is a known source of variability to block on. If so, block; if units can serve as their own controls, consider matched pairs.
3. Write the random assignment: label, generator or digit table, group sizes, how repeats are handled.
4. Describe what is held constant, what is the control or placebo, and who is blinded.
5. State the comparison to be made: for example, "compare the mean change in score of the two groups."
6. Draw the diagram as a table: Random assignment, Treatments, Observe response.

Common traps

1) Writing "randomly assign" with no method or group size. 2) Using a control group that is treated differently in time or attention, so the comparison is unfair. 3) Confusing blocking (done before random assignment, within an experiment) with stratified sampling (done to select a sample). 4) Forgetting that replication means many units per treatment, not doing the experiment on one subject many times.

Mistake Clinic

Spot the error

A student writes: "To make it fair, I let the subjects choose which treatment they prefer, and then I compare the groups." Where it fails: self-selection brings back confounding because motivated subjects may choose the treatment. Correct: "Use random assignment, using random digits to put half of the subjects in each treatment group."

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student writes: "I will block on gender by assigning all males to Treatment 1 and all females to Treatment 2." Where it fails: in a block design, each treatment is used within each block. Gender would then be confounded with the treatment. Correct: "Within the male block randomly assign half to each treatment, and do the same in the female block."

Spot the error

A student says: "A placebo is a treatment that has no effect on the subjects." Where it fails: a placebo can produce real responses because subjects expect it to help. Correct: "A placebo is an inactive treatment that looks like the real one, and it controls for the placebo effect."

Real-World Connection

Software companies run "A/B tests" on their apps: users are randomly assigned to see one of two versions of a page, and the company compares click rates. This is a completely randomized experiment with thousands of units, and replication is huge. Drug trials add blinding, because both doctors and patients can be influenced by expectations.

Challenge Corner

A researcher wants to compare two cooking oils for crispness of fries. She has 6 cooks and 12 batches of potatoes, and believes the cook matters a lot. (a) Describe a randomized block design. (b) How many batches does each cook make with each oil? (c) Explain the benefit compared with a completely randomized design. Solution: (a) Block by cook. Each cook is a block that makes 2 batches, so each cook uses both oils once, and a coin flip decides which oil is used first. This is also a matched pairs design. (b) $12/6 = 2$ batches per cook, one with each oil. (c) Differences among cooks (skill, timing, equipment habits) are removed from the comparison of the oils, so the effect of the oil is easier to detect than in a completely randomized design that assigns 6 batches to each oil without regard to the cook.

Practice

PRACTICE **Designing Experiments: Control, Randomization, Replication, Blocking**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A random assignment of 40 volunteers into two groups of 20 is used mainly to
   - **(A)** make the two groups similar on average before treatments, including on variables the researcher did not measure
   - **(B)** ensure the volunteers represent the whole population
   - **(C)** guarantee the treatment will have an effect
   - **(D)** make sure the experiment is double blind
2. In a double-blind experiment,
   - **(A)** the researcher receives the treatment as well as the subjects
   - **(B)** neither the subjects nor those evaluating the response know which treatment each subject received
   - **(C)** every subject receives both treatments
   - **(D)** there are two control groups
3. A researcher separates 60 patients into two groups of 30 older and 30 younger patients, then randomly assigns half of each group to each of two drugs. This is a
   - **(A)** completely randomized design
   - **(B)** matched pairs design
   - **(C)** randomized block design with age as the blocking variable
   - **(D)** stratified random sample

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

4. A matched pairs design is appropriate when
   (A) there are more than two treatments
   (B) the experiment cannot use random assignment
   (C) the response variable is categorical
   (D) each subject can receive both treatments in random order, or subjects can be paired by similarity
5. The purpose of a placebo in a drug experiment is to
   (A) control for the effect of subjects' belief that they are being treated
   (B) make sure the drug is safe
   (C) increase the number of subjects
   (D) remove the need for random assignment
6. A gardener compares two soil mixes by growing 10 identical seedlings in each mix, randomly assigning the seedlings. What is the role of replication?
   (A) Growing seedlings in the same room makes the conditions equal.
   (B) Using 10 seedlings per mix lets differences between the mixes be distinguished from seedling-to-seedling variation.
   (C) Randomly assigning the seedlings avoids bias.
   (D) Both mixes are measured by one person.
7. In an experiment to compare two teaching methods, the researchers know that prior math ability strongly affects scores. The best way to use this information is to
   (A) block the students by prior ability and randomly assign the methods within each block
   (B) assign the higher-ability students to the new method
   (C) ignore ability because random assignment will handle it perfectly
   (D) survey the students about their ability after the experiment
8. Which statement is true about statistical significance in a randomized experiment?
   (A) A result is significant if the difference is large enough to be important in practice.
   (B) A result is significant if the sample is large.
   (C) A result is statistically significant if the observed difference would rarely occur by chance alone if the treatments had no effect.
   (D) A result is significant if the difference is exactly zero.
9. A taste test has each of 30 people drink two colas, labeled with random codes, in random order. This design is
   (A) completely randomized with two groups of 15
   (B) a sample survey
   (C) an observational study
   (D) matched pairs with each person serving as a block
10. A company wants to know whether a new keyboard layout speeds up typing. Forty employees volunteer.
    (a) Describe a completely randomized design, including how to randomly assign the employees using a random number generator.
    (b) Name the response variable and describe the comparison.
    (c) Draw the diagram as a table.
11. The same company now thinks that typing speed before the study (fast or slow typists) will strongly affect the results. Describe a randomized block design for the 40 employees, 20 fast and 20 slow, and explain the advantage over the design in the previous problem.
12. Describe a matched pairs design for comparing two brands of headphones on battery life using 15 test units from the same factory. Include how randomization is used.
13. A pharmaceutical study compares a new allergy medicine with a placebo. Explain what is meant by a double-blind experiment in this setting, who is blinded, and why blinding is important.
14. An experimenter assigns 10 subjects, labeled 0 to 9, to Treatment 1 (5 subjects) and Treatment 2 (5 subjects) by random assignment. The digit string is 07162 02183 22002 38786 39110 65335. Read single digits from the left to choose the 5 subjects for Treatment 1.
15. An observational study found students who take a daily short walk perform better on tests. Describe an experiment to test whether walking improves performance, and explain how it overcomes the confounding that the observational study cannot.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

4.5 Scope of Inference and Writing About Study Design

Learn to decide what a study allows you to conclude: generalizing from random sampling, cause and effect from random assignment, what significance means, and how to write a complete description of a design.

Why This Matters

A study design is like a recipe, and the conclusion you are allowed to write depends on the ingredients. Random sampling controls *who* is studied, so it lets you generalize to a population. Random assignment controls *which treatment each unit gets*, so it lets you conclude that a treatment caused a difference. Mixing these up is the most common error in written responses about studies.

This lesson ties Lessons 4.1 to 4.4 together. You will learn to state the scope of inference for any study, to write a clear description of a design that earns full credit, and to explain what "statistically significant" does and does not mean. These skills are assessed in Exam Unit 1 free-response questions, where a single sentence about the scope is often worth a whole point.

Core Ideas

Scope of inference: the two questions to ask

1. Were the individuals selected at random from the population of interest? If yes, we may **generalize** the results to that population.
2. Were the treatments assigned at random to the units? If yes (and the study is an experiment), we may conclude that the **treatment caused** the observed differences.

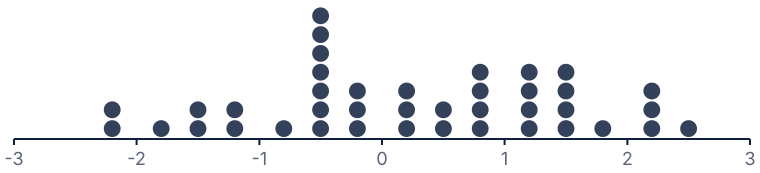
The answers to the two questions are independent. You must answer them separately, and then combine them in a sentence.

Random sampling?	Random assignment?	What you can conclude
Yes	Yes	Generalize to the population AND cause and effect
Yes	No (observational)	Generalize the association to the population, but no cause and effect
No (volunteers)	Yes	Cause and effect for units like those in the study, but no generalizing to a larger population
No	No	Neither: results describe only the individuals studied

Use careful words. In a random-sample observational study, write "is associated with." In a randomized experiment, write "caused" or "the treatment led to." When volunteers are used, write "for people similar to these volunteers."

Statistical significance versus chance. In a randomized experiment, even if the treatment does nothing, the two groups will not have identical means: random assignment itself creates differences. We ask whether the observed difference is bigger than what random assignment alone would typically produce. If it is so large that it would rarely occur by chance when the treatments are equally effective, we call it **statistically significant**, which is evidence that the treatment had an effect.

One way to see this is to **re-randomize**. Take the 12 observed scores from the experiment, shuffle them and split them into two groups of 6 at random many times, recording the difference in means each time. This shows how big a difference can be from chance alone. In the example below, 12 students took a short test after using a study app (6 students) or the usual notes (6 students). The observed difference in mean scores (app minus notes) was 2.8 points. The dotplot shows 40 re-randomized differences.



Difference in group means (points) after shuffling

None of the 40 shuffles produced a difference as large as 2.8, which suggests that 2.8 is unlikely to be just chance. (An exact count of all 924 ways to split the 12 scores shows 16 of them, about 1.7%, give a difference at least that large.) That is the idea behind a significance test, which you will study formally later.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Interpreting significance

Significant does not mean important. A tiny difference can be significant in a very large experiment.

Not significant does not mean no effect. It means the data do not provide convincing evidence of an effect; the experiment may be too small.

Significance supports a causal conclusion only when the study uses random assignment.

How to describe a design in words. Use the same flow every time: (1) identify units, factor and treatments; (2) describe the random assignment (labels and a method); (3) say what is held constant and what is blinded; (4) state the response and the comparison. Then add a diagram. A diagram is a compact picture of the design, so it should show the units, the random assignment, the treatments and the response.

Random assignment	Treatment	Measure
Group 1 (n = 6 students)	Study app	Test score
Group 2 (n = 6 students)	Usual notes	Test score
Compare	Mean score of Group 1 minus mean score of Group 2	

Worked Examples

BOTH RANDOM SAMPLING AND RANDOM ASSIGNMENT

A researcher selects an SRS of 200 adults from a city's residents. She randomly assigns 100 to use a sleep-tracking app for a month and 100 to keep a paper sleep diary. The average sleep time increases more for the app group, a statistically significant difference. What can she conclude?

SOLUTION

The adults were randomly selected, so the results can be generalized to the city's adults.

Treatments were randomly assigned, and the difference is statistically significant, so it is reasonable to conclude that the app caused a greater increase in sleep time.

Conclusion: for adults in this city, using the app (rather than the diary) leads to a greater increase in sleep time.

VOLUNTEERS AND RANDOM ASSIGNMENT

An experimenter recruits 60 volunteers at a gym and randomly assigns 30 to a new warm-up and 30 to the usual warm-up. Flexibility scores improve more with the new warm-up, and the difference is statistically significant. What can be concluded, and what cannot?

SOLUTION

Random assignment: the new warm-up caused the greater improvement for these volunteers.

The volunteers were not randomly chosen from a larger population, so we cannot generalize to all gym members or all adults; people who volunteer may be more motivated than others.

A reasonable conclusion is that the new warm-up improves flexibility for people similar to the volunteers.

RANDOM SAMPLE, NO ASSIGNMENT

A survey of 1,500 randomly selected voters finds that those who read news daily scored higher on a civics quiz than those who read news less often. The difference is statistically significant. Write a conclusion.

SOLUTION

Because the voters were randomly selected, the association can be generalized to the population of voters.

Because no treatment was assigned and people chose their own news habits, the study is observational, so we cannot conclude that reading news causes higher civics scores. Education level or interest in politics could be a confounder.

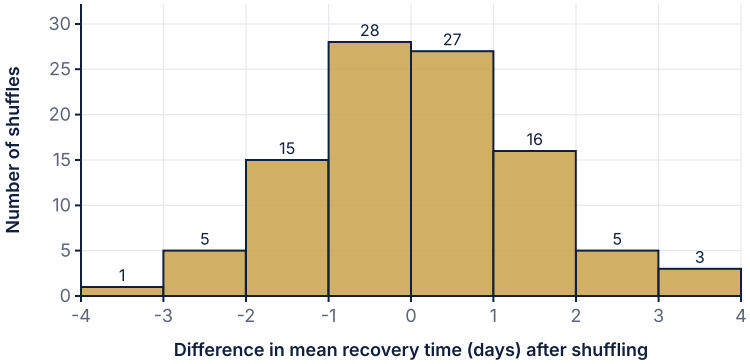
Conclusion: daily news reading is associated with higher civics scores among voters, but cause and effect is not established.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

READING A RE-RANDOMIZATION DOTPLOT

In an experiment, 20 patients were randomly split into two groups of 10; the new therapy group had a mean recovery time 3.0 days shorter than the standard group. To see how unusual this is, the 20 recovery times were shuffled and split into two groups of 10 a total of 100 times. The differences ranged from -3.4 to 3.6 days, and 3 of the 100 shuffles had a difference of at least 3.0 days in favor of the therapy. Interpret the result.



Invented summary of 100 re-randomizations. Totals 100.

SOLUTION

If the therapy had no effect, random assignment alone would produce a difference as large as 3.0 days about $3/100 = 3\%$ of the time.

That is quite unlikely, so the observed difference is statistically significant and gives convincing evidence of an effect.

Because patients were randomly assigned, we can conclude that the therapy caused shorter recovery times for these patients.

WRITE THE DESIGN AND SCOPE

A school wants to test if background music helps students complete math worksheets. Ninety volunteers from the school will take part. Describe a design with a diagram and state the scope of inference.

SOLUTION

Units: the 90 volunteers. Factor: background music with levels music/no music. Response: number of problems finished correctly in 20 minutes.

Random assignment: label the volunteers 01 to 90, use a random number generator to choose 45 different labels for the music group, and give the others silence. All students use the same worksheet, room and time limit.

Diagram: Random assignment → Group 1 (45), music → compare the mean number correct; Group 2 (45), silence → compare the mean number correct.

Scope: random assignment supports a causal conclusion about music for these volunteers. Volunteers were not randomly selected, so we should not generalize to all students.

SIGNIFICANT BUT SMALL

A very large online experiment with 400,000 randomly assigned users finds that a new button color raises the click rate from 5.00% to 5.06%, a statistically significant difference. A manager says the change is a huge success. Respond.

SOLUTION

Statistically significant means the 0.06 percentage point difference is unlikely to be due to chance. With 400,000 users even tiny differences can be detected.

Practical importance is a different question: 0.06 points means about 6 extra clicks for each 10,000 users, which may be too small to matter given the costs of making the change.

The manager should consider the size of the effect, not only the significance.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Prompt: A researcher recruits 40 volunteers from a running club and randomly assigns 20 to wear compression socks and 20 to wear regular socks during a 10 km run. The mean finishing time is 2.1 minutes faster for the compression group, and a simulation of random assignment shows a difference this large occurs about 2% of the time by chance alone. (a) What conclusion about compression socks is justified? (b) To what population can the conclusion be generalized?

Model Response

(a) Because the 40 runners were randomly assigned to the two types of socks, any differences between the groups other than the socks should be due to chance. Only about 2% of random assignments would produce a difference as large as 2.1 minutes if the socks had no effect, so the difference is statistically significant and we have convincing evidence that the compression socks caused faster finishing times for these runners. (b) The runners are volunteers from one running club, not a random sample of all runners, so the conclusion cannot be generalized to all runners. It applies to runners similar to these volunteers. A random sample of runners would be needed to generalize more widely.

What the grader looks for

(1) Links random assignment to a cause-and-effect conclusion. (2) Interprets the 2% in context as the chance of such a difference if socks had no effect. (3) States that the result is statistically significant. (4) Recognizes the volunteers were not randomly selected. (5) Limits the generalization to runners like the volunteers.

Coach's Playbook

1. Underline how the individuals were obtained. Random selection? Volunteers? Whole existing group?
2. Underline how treatments were obtained. Imposed at random? Chosen by the subjects? Not imposed at all?
3. Write two sentences: one about generalizing (sampling) and one about cause and effect (assignment).
4. Quote the context: use the real variable and real group in each sentence.
5. When a p-value or simulation is given, say how likely a difference this large is by chance alone, and use the word significant only when that chance is small.
6. Finish with a limitation: confounding, volunteers or a narrow population.

Common traps

1) Using random sampling to justify cause and effect. 2) Using random assignment to justify generalizing to everyone. 3) Treating "statistically significant" as "large" or "important." 4) Writing "the treatment has no effect" when the result is not significant, instead of "no convincing evidence of an effect."

Mistake Clinic

Spot the error

A student writes: "The sample was random, so the drug caused the improvement." Where it fails: sampling and assignment are different. If no treatment was randomly assigned, we have association only. Correct: "Because the 300 adults were randomly selected, the association can be generalized, but because they chose their own medicine we cannot say it caused the improvement."

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student writes: "We randomly assigned 50 volunteers to two diets, so the results apply to all adults." Where it fails: volunteers are not a random sample of all adults. Correct: "The diet caused the difference for these volunteers; to generalize to all adults, we would need a random sample from all adults."

Spot the error

A student writes: "The p-value is 0.20, so the therapy does not work." Where it fails: a high chance of seeing such a difference by chance means the data are consistent with no effect, not that we proved there is none. Correct: "A difference this large would happen about 20% of the time by chance, so we do not have convincing evidence the therapy works."

Real-World Connection

Medical researchers face a trade-off every day. A randomized trial at one hospital shows cause and effect but only for patients like those enrolled. A large random survey of patients can describe how a drug is used across a country but cannot prove that it works. Combining the two kinds of evidence is how treatments earn trust.

Challenge Corner

Eighteen students are randomly split into two groups of 9. The mean number of words recalled is 11.0 for Group A (new memory trick) and 8.0 for Group B. In 500 re-randomizations of the 18 scores, the difference in means (A minus B) was at least 3.0 in 20 of them. (a) Estimate the chance of a difference this large when the trick does nothing. (b) State a conclusion in context and a limitation. Solution: (a) $20/500 = 0.04$, or 4%. (b) A 4% chance is small, so the difference is statistically significant (it would be unusual by chance alone); because students were randomly assigned there is convincing evidence that the trick improves recall for these students. The students were a small group of 18, not a random sample from a larger population, so we should be cautious about generalizing.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE

Scope of Inference and Writing About Study Design

Show your work. A calculator is allowed. Justify conclusions in context.

1. A study randomly selects 500 adults from a state and randomly assigns 250 to use a new fitness app. Which conclusion is justified if the difference in fitness scores is statistically significant?
   (A) The app caused improvement, and the conclusion can be generalized to adults in the state.
   (B) The app caused improvement, but the conclusion cannot be generalized to adults in the state.
   (C) The app is associated with improvement, but no causal conclusion is possible.
   (D) The conclusion applies only to the 250 adults who used the app.
2. An experiment assigns 30 volunteers at random to two treatments and finds a statistically significant difference. Which is the most appropriate conclusion?
   (A) The treatment caused the difference for all people.
   (B) The treatment caused the difference for people like the volunteers.
   (C) The difference is due to a lurking variable.
   (D) The results are meaningless because volunteers were used.
3. A random sample of 1,200 households is surveyed. Households that recycle more also report lower energy bills, a significant difference. Which is correct?
   (A) Recycling causes lower bills in the population.
   (B) Nothing can be concluded about the population.
   (C) Recycling is associated with lower bills in the population, but this does not show recycling lowers bills.
   (D) Recycling lowers bills for the 1,200 households only.
4. In a randomized experiment, "statistically significant" means
   (A) the difference is large enough to matter in practice
   (B) the treatment is certain to work for everyone
   (C) the sample is large
   (D) the observed difference would be unlikely to occur by chance alone if the treatments had no effect
5. A re-randomization of the data from an experiment shows that a difference at least as large as the observed difference occurred in 41% of 1,000 shuffles. The best conclusion is
   (A) the observed difference is plausible by chance alone, so there is not convincing evidence of a treatment effect
   (B) the treatment has no effect
   (C) the treatment definitely caused the difference
   (D) the experiment should have used more shuffles
6. Which study allows a conclusion of cause and effect for a population of all 8,000 students at a school?
   (A) A random sample of 400 students is randomly assigned to two study methods.
   (B) 400 volunteers are randomly assigned to two study methods.
   (C) A random sample of 400 students reports their study methods and scores.
   (D) All 8,000 students are asked their study method and score.
7. A company finds that a new website design increases the mean purchase amount by 0.30 dollars, a statistically significant result with 2 million users. Which statement is most accurate?
   (A) The difference must be large and important.
   (B) The difference is unlikely to be chance, but whether 0.30 dollars matters is a separate practical question.
   (C) A statistically significant result is never small.
   (D) The result is not significant because the difference is small.
8. A researcher compares the mean scores of two groups in an observational study and finds a statistically significant difference. This provides evidence that
   (A) group membership caused the difference
   (B) the difference is due to chance
   (C) the groups differ in ways beyond chance, but not that group membership caused the difference
   (D) the study design was a randomized block design

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. Which sentence correctly describes the scope of inference for a study that uses random assignment of 24 volunteers from a community center to two exercise programs?
   (A) Cause and effect can be concluded for the entire city.
   (B) Only association can be concluded, since volunteers were used.
   (C) No conclusion can be drawn because there were only 24 subjects.
   (D) Cause and effect can be concluded for volunteers like these, but not for the entire city.
10. For each study, state whether the results can be generalized to a larger population and whether a cause-and-effect conclusion is justified. Explain each in one sentence.
    (a) 150 randomly selected teachers are randomly assigned to two training programs.
    (b) 80 volunteer students are randomly assigned to two training programs.
    (c) 500 randomly selected teachers report their training and their classroom ratings.
    (d) A school district compares two schools that chose different training programs.
11. A coach wants to know whether a new dribbling drill improves a player's accuracy. Twenty-four players on the team agree to take part. Describe an experiment, including random assignment, the response, a diagram, and state the scope of inference.
12. In an experiment, 16 plants were randomly assigned to two fertilizers (8 each). The mean growth for Fertilizer X was 3.5 cm more than for Fertilizer Y. In 200 re-randomizations of the 16 growth measurements, a difference in means of at least 3.5 cm occurred 6 times. (a) Estimate the probability of a difference this large by chance. (b) What can be concluded?
13. A news report says, "A study of 2,000 randomly selected adults found that those who eat breakfast have a lower body mass index (BMI), a statistically significant difference. Eating breakfast helps keep weight down." Evaluate the conclusion.
14. Explain the difference between statistical significance and practical importance, using an example involving a very large experiment.
15. Write a complete description of an experiment (with a diagram) that tests whether a 10-minute afternoon nap improves scores on an attention test, using 50 volunteer students. Include blinding if possible and state what conclusion would be justified if the nap group scored significantly higher.

4.6 Simulation and Random Digits

Learn to design, run, tally and interpret a simulation with random digits or a calculator, and to estimate probabilities and averages for chance processes that are hard to compute by hand.

Why This Matters

Some probabilities are too awkward to compute by hand: the chance that a quality inspector sees two defects in a sample of ten, the number of cereal boxes you need to collect a full set of prizes, the chance a team wins a best-of-seven series. A **simulation** lets us imitate the random process many times with random digits or a calculator and use the long-run results to estimate the answer.

The exam tests whether you can set up a simulation clearly enough that someone else could run it: how digits represent outcomes, what makes up one trial, what counts as success, and how many trials. Students lose credit by writing "use a random number generator" and stopping there. This is Exam Unit 1 (Collecting Data) material, and it returns when you study probability in Unit 2.

Core Ideas

A **simulation** is a model of a chance process in which we use random digits to imitate the outcomes. A **trial** is one complete run of the process (for example, one family with 4 children, one set of 5 free throws). The more trials, the closer the long-run proportion of successes tends to be to the true probability.

Four steps of a simulation (the exam expects all four)

1. **Design the model.** State how random digits (or random integers) represent the outcomes, so that the probabilities match the situation. Say which digits are skipped and whether repeats are allowed.
2. **Define a trial.** Say how many digits make up one trial, what is recorded, and when the trial stops.
3. **Run and tally.** Carry out a stated number of trials (in practice many) and record the result of each in a table.
4. **Conclude.** Estimate the probability or average with (number of trials with the event)/(total trials), and state it in context.

Assigning digits. Match the probabilities with equally likely digits. If an event has probability 0.30, use three of the ten digits 0 to 9 for it (for example 0, 1, 2 for rain, and 3 to 9 for no rain). If an event has probability 0.25, use two-digit numbers 00 to 99 and let 00 to 24 stand for it, or use digits 1 to 4 to stand for four equally likely outcomes, skipping 0 and 5 to 9.

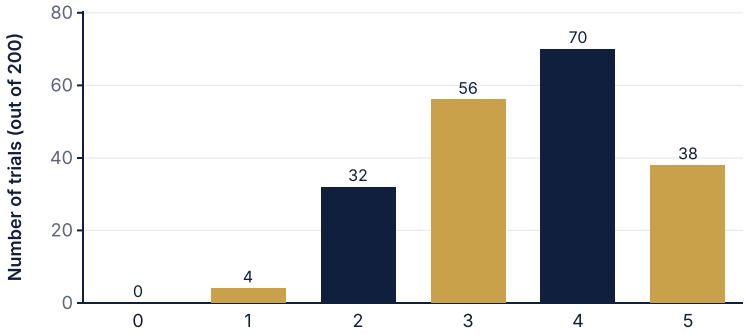
Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Situation	Digit assignment	Skip
Shooter makes 70% of shots	0, 1, 2, 3, 4, 5, 6 = make; 7, 8, 9 = miss	None
Fair six-sided die	1, 2, 3, 4, 5, 6	0, 7, 8, 9
Four equally likely prizes	1, 2, 3, 4	0, 5, 6, 7, 8, 9
35% chance of a defect	00 to 34 = defect; 35 to 99 = OK (two digits)	None
Choose 3 different people from 10 (labeled 0 to 9)	Read digits one at a time	Repeats

Sampling with and without replacement. If each trial is like independent draws (a coin flip, a die roll, a free throw), repeated digits are fine. If the process picks distinct items (a committee of 3 people, a hand of cards), ignore repeats.

The bar chart shows what 200 simulated sets of 5 free throws from a 70% shooter look like. Even with 200 trials the bars are bumpy, so the estimate of any probability from a simulation has some sampling variability. The exact probability of making at least 4 of 5 is about 0.528; the simulated proportion is 108/200 = 0.54.



Simulated number of made shots in 5 attempts for a 70% shooter.

Interpreting the results

A simulated probability is an estimate: it will vary from one run of the simulation to another.
More trials give a more stable estimate. With very few trials, results can be far from the truth.
Always give the estimate in context: "We estimate a 55% chance that ..."

Worked Examples

DESIGNING THE MODEL

A basketball player makes 70% of free throws. Design a simulation to estimate the probability that she makes at least 4 of her next 5 free throws.

SOLUTION

Digits: 0, 1, 2, 3, 4, 5, 6 represent a made shot (7 of 10 digits, or 70%); 7, 8, 9 represent a miss.
One trial: read 5 digits (repeats allowed) and count how many are 0 to 6.
A trial is a success if the count is 4 or more.
Run many trials. The estimate is the number of successes divided by the number of trials.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

RUNNING AND TALLYING 20 TRIALS

Use the 100 digits in the table (20 trials of 5 digits) to carry out the simulation from the previous example, and estimate the probability.

Trial	Digits	Made (0 to 6)	4 or more?
1	26647	4	Yes
2	37828	2	No
3	83005	4	Yes
4	96123	4	Yes
5	30669	4	Yes
6	70587	2	No
7	15021	5	Yes
8	10728	3	No
9	56722	4	Yes
10	07313	4	Yes
11	39118	3	No
12	78635	3	No
13	28645	4	Yes
14	78657	2	No
15	64012	5	Yes
16	70718	2	No
17	98399	1	No
18	58912	3	No
19	25559	4	Yes
20	43121	5	Yes

SOLUTION

Count digits 0 to 6 in each group of 5. For trial 1, 26647 has four digits from 0 to 6 (the 7 is a miss), so trial 1 gives 4 made shots; the counts for all 20 trials are in the table.

A trial succeeds when there are 4 or more made shots. Successes occur in 11 of the 20 trials.

Estimate: 11/20 = 0.55. We estimate about a 55% chance that she makes at least 4 of 5 free throws.

The exact probability is 0.528, so the estimate from 20 trials is close but not exact.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

WAITING-TIME SIMULATION

Each box of cereal contains one of 4 prizes, equally likely. Design a simulation to estimate the average number of boxes needed to get all 4 prizes, and run 5 trials with the digits 79542 20587 91589 06276 22301 28916 14336 14364 50306 06720 36190 19335 01280 81978 36163 19229 00489 24901 63.

Trial	Boxes needed
1	8
2	6
3	7
4	15
5	4

SOLUTION

Digits: 1, 2, 3, 4 represent the four prizes; skip 0 and 5 to 9. Repeats are allowed.

One trial: read digits, record each valid one as a box, and stop as soon as all four of 1, 2, 3, 4 have appeared. Record the number of boxes.

Trial results (valid digits read): 42212223 (8 boxes); 121143 (6 boxes); 3143432 (7 boxes); 311331213131224 (15 boxes); 2413 (4 boxes).

The mean number of boxes in the 5 trials is $(8+6+7+15+4)/5 = 8.0$. A computer simulation of 10,000 trials gives about 8.3 boxes, and with only 5 trials the estimate is rough.

SAMPLING WITHOUT REPLACEMENT

A club has 10 members: 4 seniors and 6 juniors. Three members are chosen at random for a committee. Design a simulation to estimate the probability that at least one senior is chosen.

SOLUTION

Label the members 0 to 9, with 0, 1, 2, 3 for the seniors and 4, 5, 6, 7, 8, 9 for the juniors.

One trial: read digits one at a time, ignoring repeats, until 3 different digits are chosen. A trial is a success if at least one of the digits is 0, 1, 2 or 3.

Run many trials. Estimate $=$ (number of successes)/(number of trials). Using 10 trials from a table, 8 contained a senior, giving an estimate of 0.80. The exact value is $1 - \binom{6}{3}/\binom{10}{3} = 1 - 20/120 \approx 0.833$.

USING A CALCULATOR

A weather service says there is a 30% chance of rain on each of three consecutive days (independent). Describe a calculator simulation to estimate the probability of rain on at least one of the three days, and conclude if 1,000 trials produce 659 successes.

SOLUTION

Model: a random integer from 1 to 10; 1, 2 or 3 means rain (30%), and 4 to 10 means no rain.

Trial: generate 3 random integers from 1 to 10 and record whether at least one is 1, 2 or 3.

Repeat for 1,000 trials and count the successes.

With 659 successes, the estimate is $659/1000 = 0.659$, so we estimate about a 66% chance of rain on at least one of the three days. (Exact: $1 - 0.7^3 = 0.657$.)

EVALUATING A SIMULATION DESIGN

A student wants to simulate the number of girls among 4 children when each child is equally likely to be a boy or a girl. She uses digits 1 to 6 to represent a girl and 7 to 9 to represent a boy. Is the design valid? Fix it.

SOLUTION

No. Digits 1 to 6 for a girl and 7 to 9 for a boy, with 0 skipped, give a girl probability of $6/9 \approx 0.67$, not 0.5.

A valid design uses 0 to 4 for a girl and 5 to 9 for a boy (each 50%), or even digits for girl and odd digits for boy.

One trial is 4 digits and records the number of digits from 0 to 4. Run many trials; the exact probability of exactly 2 girls is $6/16 = 0.375$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Write It Like a 5

Prompt: A factory machine produces a part that is defective 10% of the time, independently. An inspector tests 6 parts from a batch. Describe a simulation to estimate the probability that at least one of the 6 parts is defective. Include how to use random digits, what a trial is, how many trials, and how to estimate the probability.

Model Response

Use the digits 0 to 9, one digit per part. Let the digit 0 represent a defective part (10%) and 1 to 9 represent a good part. One trial consists of reading 6 digits, with repeats allowed since the parts are independent. A trial is a success if at least one digit is 0. Carry out 100 trials (reading 600 digits in groups of 6), keeping a tally of the trials with at least one 0. The estimated probability is the number of successful trials divided by 100. For example, if 46 of the 100 trials contain a 0, we would estimate a 46% chance that at least one of the 6 parts is defective. (Exact: $1 - 0.9^6 = 0.469$.)

What the grader looks for

(1) Assigns digits correctly so the probability is 10%. (2) Defines a trial (6 digits) and the success condition. (3) States that repeats are allowed (independent parts). (4) Specifies a number of trials and a tally. (5) Describes the estimate as successes divided by trials, with a conclusion in context.

Coach's Playbook

1. Match the probability: pick a set of digits whose share equals the probability (30% means three of the ten digits).
2. Decide how many digits make a trial, and whether repeats are allowed.
3. Define success clearly, such as "at least 4 digits from 0 to 6."
4. Run the trials and put the result of each in a table so it is auditable.
5. Compute the proportion of successes or the mean, and state it in context as an estimate.
6. Calculator tip: use the random integer command with the lower limit 1, the upper limit the number of equally likely outcomes, and the number of values per trial.

Common traps

1) Using digits that give the wrong probability, such as 1 to 4 for a 40% event while including 0 and 5 to 9 as separate outcomes. 2) Ignoring repeats when they should be ignored (choosing different people) or ignoring them when they should be kept (independent coin flips). 3) Forgetting to say what one trial is. 4) Treating the simulated estimate as the exact probability.

Mistake Clinic

Spot the error

A student writes: "To simulate a spinner with 3 equally likely colors, I use digits 0 to 9: 0 to 2 red, 3 to 5 blue, 6 to 8 green." Where it fails: digit 9 is left unassigned, so the three colors do not all have a 1/3 probability if 9 is not skipped. Correct: "Use 1, 2, 3 for the three colors and skip 0 and 4 to 9," or "1 to 3 red, 4 to 6 blue, 7 to 9 green, skipping 0."

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student says: "I ran 10 trials and got 6 successes, so the probability is exactly 0.6." Where it fails: with so few trials the estimate is unreliable and simulation results vary. Correct: "The simulation estimates the probability to be about 0.6, but with only 10 trials the estimate may be far from the true value; more trials give a more reliable estimate."

Spot the error

A student simulates a 40% free-throw shooter taking 3 shots by letting digits 1, 2, 3, 4 be a make and reading one digit for the whole trial. Where it fails: a trial for 3 shots needs 3 digits (one per shot). Correct: "Digits 0 to 3 make (40%), 4 to 9 miss; each trial reads 3 digits and counts the makes."

Real-World Connection

Engineers simulate bridges, airlines simulate delays and sports analysts simulate entire seasons. Each simulation follows the same steps you just practiced: represent each random event with random numbers, define a trial, repeat thousands of times and summarize. The more complex the system, the more valuable simulation becomes.

Challenge Corner

A game: roll a fair die until a 6 appears. Design a simulation to estimate the mean number of rolls needed and the probability that it takes more than 3 rolls. Then compute the exact values. Solution: digits 1 to 6 represent die faces (skip 0 and 7 to 9). One trial: read valid digits until a 6 appears, recording the number of rolls. Repeat many times. The estimate of the mean is the average number of rolls, and the estimate of $P(\text{more than 3 rolls})$ is the proportion of trials with a count of 4 or more. Exact values: the mean is $1/(1/6) = 6$ rolls, and $P(\text{more than 3}) = (5/6)^3 \approx 0.579$.

Practice

PRACTICE **Simulation and Random Digits**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A basketball player makes 40% of her shots. Which digit assignment correctly simulates one shot?
   - **(A)** 0, 1, 2, 3 = make; 4, 5, 6, 7, 8, 9 = miss
   - **(B)** 0, 1, 2, 3, 4 = make; 5, 6, 7, 8, 9 = miss
   - **(C)** 1, 2, 3, 4 = make; 5, 6, 7, 8, 9 = miss; ignore 0
   - **(D)** 0, 1, 2, 3, 4, 5 = make; 6, 7, 8, 9 = miss
2. In a simulation, a trial is
   - **(A)** the final estimate of the probability
   - **(B)** one complete run of the chance process, such as 5 free throws
   - **(C)** one random digit, no matter how the process is defined
   - **(D)** the number of runs that are successful
3. A spinner has 5 equally likely colors. Which digit assignment is valid for one spin?
   - **(A)** 1 and 2 = color 1, 3 = color 2, 4 = color 3, 5 = color 4, 6 = color 5, ignoring 0 and 7 to 9
   - **(B)** 1 to 4 for colors 1 to 4, and 5 or 0 for color 5, ignoring 6 to 9
   - **(C)** 1 to 5 for the five colors, ignoring 0 and 6 to 9
   - **(D)** 1 to 5 for the five colors, and 6 to 9 as extra copies of color 5
4. A simulation of 50 trials produces 18 successes. The best estimate of the probability of success is
   (A) 0.18 **(B)** 0.50 **(C)** 2.78 **(D)** 0.36

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. Which change would make a simulation estimate of a probability more reliable?
   (A) Using more trials **(B)** Using fewer digits per trial
   (C) Always skipping the digit 0 **(D)** Stopping after the first success
6. A committee of 2 is chosen at random from 5 people labeled 1 to 5. Using single digits, which procedure is correct?
   (A) Read digits, ignore 0 and 6 to 9, and allow repeats
   (B) Read digits, ignore 0 and 6 to 9 and repeats, until 2 different labels are found
   (C) Read two-digit numbers and use only those from 01 to 05
   (D) Read one digit and let it choose both people
7. A fair coin is flipped 3 times. Which simulation estimates the probability of at least 2 heads?
   (A) Digits 0 to 4 = heads, 5 to 9 = tails; 1 digit per trial; success if it is 0 to 4
   (B) Digits 1 to 3 = heads, 4 to 9 = tails; 3 digits per trial; success if at least two digits are 1 to 3
   (C) Digits 0 to 4 = heads, 5 to 9 = tails; 3 digits per trial; success if at least two digits are 0 to 4
   (D) Digits 0 to 4 = heads, 5 to 9 = tails; 3 digits per trial; success if all three digits are 0 to 4
8. A student runs a simulation twice with 100 trials each and gets estimates of 0.43 and 0.49. This shows that
   (A) the simulation was done incorrectly **(B)** the true probability is 0.46
   (C) the digits used were biased **(D)** simulation estimates vary from run to run
9. A raffle has 1 winning ticket among 20. To simulate whether a buyer of 1 ticket wins, which assignment of two-digit random numbers works?
   (A) 00 to 04 means win **(B)** 00 to 19 means win **(C)** 00 to 05 means win
   (D) 01 to 20 means win
10. A candy maker says 20% of its bags contain a prize. Design a simulation to estimate the probability that you must buy at least 4 bags to get your first prize. Define the digit assignment, a trial and the estimate.
11. A free-throw shooter makes 60% of her shots. Describe a simulation using random digits to estimate the probability that she makes exactly 3 of her next 4 shots, and run 6 trials using the digits 77447 16113 41490 32534 9009 (4 digits per trial, 0 to 5 = make).
12. A weather model says each of the next 4 days has a 25% chance of snow, independently. Describe how to use two-digit random numbers to simulate one trial and how to estimate the probability of snow on at least one day.
13. In a simulation of a tournament, a student ran 20 trials and found the team won the series 13 times. Explain why 13/20 is only an estimate, and what the student could do to improve it.
14. A class has 8 girls and 12 boys. A teacher picks 3 students at random for a project. Design a simulation to estimate the probability that all 3 are girls, including the labeling, trial and handling of repeats.
15. Critique this design: "To simulate a fair die, I use the digits 0 to 9 and let 0 and 1 be a 1, 2 and 3 be a 2, 4 and 5 be a 3, 6 and 7 be a 4, and 8 and 9 be a 5. A 6 never occurs, so I will re-roll."

T Unit 4 Test

Test your skills with populations, bias, sampling methods, observational studies, experimental design, scope of inference and simulation.

Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Part A: Multiple Choice

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

UNIT TEST

Unit 4 Test

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. A city has 90,000 registered drivers. The transportation office surveys 1,200 randomly selected drivers and finds that 31% commute more than 30 minutes. Which is correct?
   (A) 31% is a statistic and the percentage of all 90,000 drivers who commute that long is a parameter.
   (B) 31% is a parameter and the percentage in the 1,200 is a statistic.
   (C) Both are statistics because both are percentages.
   (D) 31% is a census value because the office surveyed many drivers.
2. A county health office wants to estimate the proportion of adults who own a smartphone. It samples at random from a list of households with a landline telephone. The estimate will most likely
   (A) be unbiased, because the households were chosen at random
   (B) underestimate the proportion, because adults without landlines tend to be younger and more likely to own smartphones
   (C) overestimate the proportion, because adults without landlines are less likely to own smartphones
   (D) be unbiased, because a large sample removes bias
3. A radio show asks listeners to text in their opinion about a proposed bike lane. Most of the 400 texts oppose it. This is an example of
   (A) undercoverage, likely understating opposition
   (B) stratified sampling, likely overstating opposition
   (C) voluntary response bias, likely overstating opposition
   (D) nonresponse bias, likely understating opposition
4. A school with 20 homerooms selects 4 homerooms at random and surveys every student in the chosen homerooms. This is a
   (A) stratified random sample (B) simple random sample (C) systematic sample
   (D) cluster sample
5. Students are labeled 01 to 60. Using the digits 63520 80572 31292 76562 42175 02373 read two at a time from the left, the SRS of size 3 consists of the students labeled
   (A) 52, 08, 05
   (B) 63, 52 and 08 (the first three pairs, with no rules applied)
   (C) the single digits 6, 3 and 5
   (D) the labels 01, 02 and 03
6. A plant has 1,800 machines on a list and plans a systematic sample of 60. After choosing the random start, the inspector will select every
   (A) 60th machine (B) 30th machine (C) 18th machine
   (D) 1,740th machine
7. Researchers compare 500 adults who report taking a daily vitamin with 500 who do not and find the first group has fewer colds. Which statement is best?
   (A) The study is an experiment because there are two groups.
   (B) The vitamin caused the difference because the groups were large.
   (C) The study is observational, and a confounder such as healthy habits could explain the difference.
   (D) The study is invalid because the groups were not the same size.
8. In a study of two fertilizers, the field has a sunny half and a shady half. The researcher randomly assigns each fertilizer to half of the plots within each half of the field. This design is a
   (A) completely randomized design
   (B) matched pairs design with 2 blocks
   (C) stratified random sample
   (D) randomized block design with sunlight as the blocking variable
9. Which is the best reason to use a matched pairs design for comparing two running shoes using 12 runners?
   (A) It guarantees that the experiment is double blind.
   (B) Each runner can use both shoes, so differences between runners do not affect the comparison.
   (C) It removes the need for random assignment.
   (D) It makes the sample size larger.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. Fifty volunteers at a gym are randomly assigned to one of two stretching routines. After a month, one routine shows significantly greater flexibility. Which conclusion is justified?
    (A) The routine caused greater flexibility for people like these volunteers.
    (B) The routine would cause greater flexibility for all adults.
    (C) The routine is associated with flexibility, but no cause can be shown.
    (D) The result is due to chance because the volunteers were not randomly sampled.
11. A quality inspector wants to simulate whether a part is defective when 15% of parts are defective. Using two-digit random numbers, which assignment is correct?
    (A) 00 to 15 = defective; 16 to 99 = good
    (B) 00 to 09 = defective; 10 to 99 = good
    (C) 00 to 14 = defective; 15 to 99 = good
    (D) 00 to 19 = defective; 20 to 99 = good
12. In an experiment, the response scores in the treatment group have mean 72 and standard deviation 6. One participant scored 84. Which statement about this participant is correct?
    (A) The score is 12 standard deviations above the group mean.
    (B) The score is 2 standard deviations below the group mean.
    (C) The score is 0.5 standard deviations above the group mean.
    (D) The score is 2 standard deviations above the group mean.
13. [6 points] A school has 1,500 students: 600 in the lower school (grades 9 and 10) and 900 in the upper school (grades 11 and 12). The principal wants to estimate the proportion of all students who support a later start time.
    (a) [3 points] Describe how to select a stratified random sample of 100 students with proportional allocation, using a random number generator.
    (b) [1 point] Explain why a stratified sample might give a more precise estimate than a simple random sample of 100 students.
    (c) [2 points] The principal emails the survey to the 100 selected students and only 38 respond. Describe one source of bias this creates and its likely direction if students who dislike early mornings respond more often.
14. [10 points] A teacher wants to know whether a 10-minute mindfulness exercise before a math quiz improves quiz scores. Thirty-six students volunteer: 18 who say they usually feel very anxious about quizzes and 18 who say they usually do not. The teacher believes anxiety affects scores strongly.
    (a) [2 points] Identify the explanatory variable, the response variable and the experimental units.
    (b) [4 points] Describe a randomized block design for this experiment, including the blocks, how random assignment is carried out with a random number generator, and a diagram of the design.
    (c) [1 point] Describe a control treatment and explain how blinding could be used.
    (d) [3 points] The students in the mindfulness treatment scored a mean of 5.0 points higher than those in the control treatment. The teacher re-randomized the treatment labels 1,000 times and found a difference of at least 5.0 points in 14 of the 1,000 shuffles. Interpret this result, state what conclusion is justified about the effect of the mindfulness exercise, and state a limitation on generalizing the conclusion.

Answers

Collecting Data: Sampling, Experiments, and Simulation: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

4.1 Populations, Samples, and Sources of Bias

PRACTICE · POPULATIONS, SAMPLES, AND SOURCES OF BIAS

1. A. A is correct: a number from a sample is a statistic and a number describing the entire population is a parameter. B reverses the two. C and D confuse the form of the number (a percent) with its source.
2. B. B is correct: the store chose everyone on its list, so the problem is that most did not answer. A is tempting but customers were chosen, not self-selected. C confuses a low response rate with a missing frame. D is not a reason to suspect inaccurate answers.
3. C. C is correct: a directory leaves out adults with unlisted or no landline numbers, so they have no chance of selection. A, B and D all use frames that include essentially every resident.
4. A. A is correct: the arcade is easy to reach, but arcade visitors play games more than teens at large, so the estimate is too high. B and C misname the method, since nobody volunteered and nothing was stratified. D has the wrong direction.
5. D. D is correct: the halves were formed at random, so the large difference is attributed to the wording of the questions. A, B and C describe problems that are not supported by the situation.
6. A. A is correct: larger random samples give statistics that cluster more tightly, but a biased method keeps missing the target. B reverses the two ideas, C overstates the benefit and D is wrong because variability does decrease.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. B. B is correct: it is a sample statistic, and people home during weekday afternoons are often older, retired or work from home. A misuses a size rule and calls the number a parameter. C is wrong because only some residents were measured. D mistakes the sample for the frame.
8. C. C is correct: respondents give inaccurate answers, which is response bias. A is voluntary response, B is undercoverage and D is convenience sampling.
9. D. D is correct: the sample mean $\bar{x}$ is a statistic estimating the population mean μ. A reverses them, B is for proportions and C is for standard deviations.
10. (a) Population: all 12,000 employees. Sample: the 210 employees who responded (of the 600 selected). Statistic: $\hat{p} = 168/210 = 0.80$. (b) Response rate $= 210/600 = 0.35$, or 35%. (c) Nonresponse bias: 65% of those chosen did not reply, and employees with strong feelings about remote work (especially those who like it) may be more likely to answer, so 0.80 likely overestimates the proportion in the whole company who prefer remote work. A grader wants the type, the reason and the direction.
11. (a) Voluntary response and wording bias: the loaded phrase "waste money" and the self-selection of readers who feel strongly both push toward answering no, so opposition is overestimated. (b) Response bias: students are unlikely to admit copying face to face, so the proportion who copied is underestimated. (c) Convenience sampling / undercoverage: anglers keep or report larger fish and small fish may not be caught by their gear, so the mean length is likely overestimated.
12. A high response rate removes the nonresponse problem, but accuracy also depends on who was surveyed and how the questions were asked. For example, if the survey was only distributed to the 1,200 students through an online class platform that some students could not access, or if the questions were worded in a leading way, the responses could still be biased. A sample is accurate only when the method is free of bias; a large number of responses cannot repair a flawed method.
13. (a) Plan (iii) is best because it uses a complete list of students, selects at random so every student has the same chance, and follows up to limit nonresponse bias. (b) Plan (i) is a convenience sample: students in the cafeteria at one lunch may be more social or have a specific schedule and not represent those who spend lunch in clubs or at the library. Plan (ii) is voluntary response: students interested in robotics are most likely to click the link, so interest would be overestimated.
14. A biased question: "Do you agree that forcing students to wake up before sunrise is harmful to their health?" A neutral question: "What start time would you prefer for school: earlier than 8:00, 8:00 to 8:30, or later than 8:30?" The loaded wording ("forcing," "before sunrise," "harmful") encourages agreement with the change, so the percentage favoring later start times would be overestimated.
15. (a) $\hat{p} = 228/400 = 0.57$. It is a statistic because it is computed from a sample. (b) No. A sample proportion varies from sample to sample, so 0.57 can differ from the true p even when $p = 0.60$ just by chance. A difference of 0.03 with $n = 400$ is within ordinary random variation (the standard deviation of $\hat{p}$ is about $\sqrt{0.6(0.4)/400} \approx 0.0245$). We cannot conclude the claim is false from this alone.

4.2 Sampling Methods: SRS, Stratified, Cluster, Systematic

PRACTICE · SAMPLING METHODS: SRS, STRATIFIED, CLUSTER, SYSTEMATIC

1. A. A is correct: an SRS requires that every possible group of size n is equally likely. B is necessary but not sufficient (a systematic sample also gives equal individual chances). C describes proportional or equal stratification. D selects whole classes and is not an SRS.
2. B. B is correct: schools are clusters and everyone in the chosen clusters is measured. A would sample from every school. C would select students individually. D would select every kth student.
3. C. C is correct: read pairs, skip 00 and anything above 50, and skip repeats, so the selected labels are given in the choice. The distractors read the digits wrongly or do not remove invalid labels.
4. D. $2000/25 = 80$ customers. Each customer has an equal 1-in-25 chance, but neighbors can never both appear, so not every group of 80 is possible and it is not an SRS. A claims equal chance is enough. B and C confuse $k = 25$ with the sample size.
5. A. A is correct: day and boarding are strata, each sampled separately, and $800/1000 = 80\%$ and $40/50 = 80\%$ match. B treats groups as clusters but both groups are sampled. C is wrong because an SRS of 50 could contain 45 day students, which this design forbids. D misreads stages.
6. B. B is correct: strata that differ on the variable lead to more precise estimates. A is a reason for cluster or multistage sampling, C is a reason for cluster sampling, and D is not a design reason.
7. A. A is correct. B is a convenience sample. C is systematic and does not allow all groups. D is stratified and does not allow all groups.
8. C. C is correct: $8 \times 5 \times 6 = 240$. A adds the numbers. B multiplies only two stages and D multiplies by the 50 towns instead of choosing 8.
9. D. D is correct. In a cluster sample you randomly choose clusters and measure all within them. A describes a stratified sample, B reverses the roles (strata differ from each other; clusters resemble the population), and C is false since clusters are chosen at random.
10. Label the members 001 to 650 using three-digit labels. Starting at a stated line of the table, read the digits in groups of three. Ignore 000 and any group from 651 to 999. Ignore a group that has already been selected. Continue until 12 different valid labels are found; the members with those labels are the sample. Three-digit labels make every member equally likely to be selected.
11. (a) Each grade is a stratum with $500/2000 = 25\%$ of students, so select 25 students from each grade by labeling the students 001 to 500 and choosing an SRS of 25 for each grade. (b) Label the homerooms 1 to 20, use a random number generator to choose $100/25 = 4$ homerooms, and survey all 25 students in each chosen homeroom. (c) Stratified guarantees every grade is represented and gives a more precise estimate if grades differ. Cluster saves time and cost because surveyors visit only 4 rooms and no list of all students is needed.
12. (a) $k = 40/8 = 5$. (b) Use a random number generator or digit table to choose an integer from 1 to 5 as the starting lane. (c) With start 3: lanes 3, 8, 13, 18, 23, 28, 33, 38. (d) Lanes 3 and 4 can never both be in the sample, so not every group of 8 lanes is possible; an SRS gives every group of 8 an equal chance.
13. It is a two-stage (multistage) cluster-type sample, not a stratified sample. Stage 1 is an SRS of 10 classrooms from all classrooms (only some groups are sampled), and stage 2 is an SRS of 5 students within each chosen classroom. In a stratified sample every classroom would have students selected. Because 10 classrooms are left out entirely, the groups act as clusters at the first stage.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. (a) $600/1000 \times 50 = 30$, $300/1000 \times 50 = 15$, $100/1000 \times 50 = 5$. (b) With only 5 visitors from the smallest zone, the estimate for that zone would be imprecise. A larger share (for example 10) gives a better estimate of that zone, though weights must then be used when combining zones.
15. Stage 1: random sample of counties (clusters of the state) to reduce travel. Stage 2: random sample of census blocks or neighborhoods within each chosen county. Stage 3: random sample of addresses within each block from a complete address list. Stage 4: choose one adult per household at random, for example by labeling adults. Randomizing at every stage gives each adult a known chance of selection and prevents interviewers from choosing convenient people, which avoids selection bias.

4.3 Observational Studies, Experiments, and Confounding

PRACTICE · OBSERVATIONAL STUDIES, EXPERIMENTS, AND CONFOUNDING

1. A. A is correct: tea drinking is not assigned. B treats having two variables as an experiment, C confuses random sampling with random assignment, and D is wrong because only 300 adults were asked.
2. B. B is correct: the data already exist and researchers look back. A is prospective, C is an experiment, and D is a current survey.
3. C. C is correct: prior struggle is related to using a tutor and to lower grades, which creates confounding. A claims causation from observational data, B is about sample size not confounding, and D does not explain a link between tutoring and grades.
4. D. D is correct. A is wrong because units receive the treatments. B describes the response variable. C describes the population.
5. A. A is correct: motivation is associated with the choice of routine and affects recovery. B, C and D concern sampling and measurement, not the treatment assignment.
6. B. B is correct: without random assignment, class differences are confounded with the textbook. A is the usual error of claiming causation, C is unsupported and D confuses random selection with random assignment.
7. A. A is correct. B is false because size does not remove confounding, C describes experiments and D is false because no treatment is imposed.
8. C. C is correct: the score is measured after the room temperature (explanatory) varies. A is the explanatory variable and B and D are units or conditions, not outcomes.
9. D. D is correct, using the right language for an observational study. A and B claim causation in one direction or the other. C is wrong: comparing two groups is fine.
10. (a) Observational: the coach recorded existing shake use and did not assign it. (b) Training effort: athletes who work out harder may drink more shakes (associated with the explanatory variable) and gain more muscle because of the workouts (influences the response), so the effects of shakes and effort cannot be separated. (c) Randomly assign athletes to a shake group and a no-shake group (or placebo) with the same training program, then compare changes in muscle mass.
11. Explanatory: the type of background condition (classical music versus silence). Response: the number of puzzles solved in 20 minutes. It is an experiment because the researchers assigned the conditions, and random assignment balances other factors. Conclusions would be more convincing with a larger number of volunteers, replication, and ideally blinding the person scoring the puzzles to which condition each participant had. Note that volunteers restrict generalization to a larger population.
12. (a) Retrospective: existing records are examined from the past. (b) No, it is an observational study because no treatment was imposed. (c) Overall healthy lifestyle: people who eat more fish may also exercise more, smoke less and eat more vegetables, which also lower the rate of heart conditions, so the benefits of fish cannot be separated from these habits.
13. Districts were not randomly assigned to start times, so the groups may differ in other ways that are associated with the start time and attendance. Possible confounders: (1) community income, because wealthier districts may have been able to adopt the change and have better attendance anyway; (2) district size or transportation, because better bus services may be associated with both later start times and higher attendance. The data show association but not cause.
14. Random selection (choosing a sample at random from a population) reduces bias in who is studied and supports generalizing results to the population. Random assignment (using chance to assign units to treatments in an experiment) balances other variables across groups and supports cause-and-effect conclusions about the treatments. A study can have either, both or neither.
15. A careful interpretation: "In this sample, heavy social media use was associated with higher reported anxiety. Because this is an observational study, we cannot conclude that heavy use causes anxiety." A possible confounder: loneliness or sleep problems, which may lead teens to use social media for longer and independently contribute to anxiety. Alternatively, anxious teens may turn to social media more (reverse direction), which also prevents a causal conclusion.

4.4 Designing Experiments: Control, Randomization, Replication, Blocking

PRACTICE · DESIGNING EXPERIMENTS: CONTROL, RANDOMIZATION, REPLICATION, BLOCKING

1. A. A is correct: chance balances lurking variables across groups. B describes random sampling. C is not guaranteed by any design. D concerns blinding, a separate feature.
2. B. B is correct. A, C and D are not part of the definition: double blind refers to who is unaware of the assignments.
3. C. C is correct: age is a blocking variable and random assignment is done within each block. A has no grouping, B pairs only two units, and D selects a sample, not assign treatments.
4. D. D is correct. A is not required (matched pairs compares two conditions), B would eliminate the main advantage, and C is not a requirement.
5. A. A is correct. B is a safety issue, C is not a function of the placebo and D is wrong because random assignment is still required.
6. B. B is correct: replication means many units per treatment. A is control, C is random assignment and D is not a design principle.
7. A. A is correct: blocking removes a known source of variability. B creates confounding, C is tempting but wastes a chance to reduce variability, and D collects data too late.
8. C. C is correct: significance compares the observed difference with chance variation from random assignment. A confuses statistical and practical significance, B depends on size only and D is the opposite idea.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. D. Each person tastes both colas in random order, so each person is a block of size two, which is a matched pairs experiment. A would put each person on only one cola. B and C are wrong because the researchers impose the treatment.
10. (a) Units: the 40 employees. Label them 01 to 40 and generate random integers from 1 to 40, ignoring repeats, until 20 different labels are found; those employees use the new layout. The other 20 use the standard layout. All employees type the same text under the same conditions. (b) Response: typing speed in words per minute on the same passage. Compare the mean speeds of the two groups. (c) Diagram: Random assignment → Group 1 (20 employees) → New layout → Mean speed; Group 2 (20 employees) → Standard layout → Mean speed.
11. Block by typing speed: fast typists (20) and slow typists (20). Within the fast block, randomly assign 10 to the new layout and 10 to the standard layout; do the same in the slow block. Compare the layouts within each block. The advantage is that the large differences between fast and slow typists are removed from the comparison, making a difference due to the layout easier to detect, and ensuring both layouts are tested equally on each type of typist.
12. Each test unit is a pair: units are matched, or each pair of units from the same batch is matched. Randomly assign one unit of each pair to Brand 1 and the other to Brand 2 (flip a coin or use a random digit). Run both until the battery dies under the same conditions. Record the difference in battery life for each pair and analyze the differences. Pairing removes variation between units. If the same unit cannot be used with both brands, matching on production batch is an alternative.
13. Double blind means that neither the subjects nor the researchers who give the medicine or measure symptoms know who received the new medicine and who received the placebo. For example, pills are coded by a third party. Blinding matters because subject expectations (placebo effect) and researcher expectations could influence reported symptoms and the judgment of improvement, which would bias the comparison.
14. Digits read: 0 7 1 6 2. Skipping repeats, Treatment 1 gets subjects 0, 7, 1, 6, 2. The other five subjects (3, 4, 5, 8, 9) get Treatment 2.
15. Randomly assign a set of volunteer students either to take a 15 minute walk before a test or to sit quietly for 15 minutes, with all other conditions (test, room, time of day) the same. Compare mean test scores. Random assignment makes the groups alike in motivation, study habits and fitness, so those factors cannot explain a difference. If the students who walked do significantly better, we have evidence that walking caused the improvement, at least for students like the volunteers.

4.5 Scope of Inference and Writing About Study Design

PRACTICE · SCOPE OF INFERENCE AND WRITING ABOUT STUDY DESIGN

1. A. A is correct: random assignment supports cause and effect, and random selection supports generalizing to the state's adults. B ignores the random selection, C ignores the random assignment and D understates what both random steps provide.
2. B. B is correct: random assignment supports causation, and volunteers limit generalization. A overgeneralizes, C is unlikely because random assignment balances lurking variables, and D overreacts.
3. C. C is correct: random sampling lets us generalize the association, but there is no random assignment. A claims causation, B ignores random sampling and D limits unnecessarily and still claims cause.
4. D. D is correct. A confuses statistical significance with practical importance, B overstates certainty and C is not what the term means.
5. A. A is correct: 41% is not unusual under no effect. B turns lack of evidence into proof of no effect, C reverses the logic and D misunderstands what shuffles do.
6. A. A is correct: it has both random sampling from the population and random assignment. B lacks random selection, C and D are observational.
7. B. B is correct: with huge samples tiny differences become significant. A and C equate significance with size, and D reverses the logic.
8. C. C is correct: significance addresses chance variation, not confounding. A requires random assignment, B contradicts significance and D is unrelated.
9. D. D is correct. A overgeneralizes, B ignores random assignment, and C is false: small experiments can still show cause and effect, though with less power.
10. (a) Generalize to the population of teachers and conclude cause and effect, since both random selection and random assignment were used. (b) No generalizing beyond students like the volunteers, but cause and effect is justified because of random assignment. (c) Generalize the association to teachers, but no cause and effect because there is no assignment. (d) Neither: the schools chose their programs and are only two schools, so we can only describe the two schools; any difference may be due to confounding.
11. Units: the 24 players. Treatments: dribbling drill, or the usual practice for the same time. Label players 01 to 24 and use a random number generator to select 12 different labels for the drill; the other 12 get usual practice. Response: shooting accuracy (percent of 50 shots made) in the same test after two weeks. Diagram: Random assignment → Group 1 (12 players), drill → accuracy; Group 2 (12 players), usual practice → accuracy; compare group means. Scope: random assignment allows cause and effect for these players, but they are one team, not a random sample, so do not generalize to all players.
12. (a) $6/200 = 0.03$, or 3%. (b) If the fertilizers were equally effective, a difference of 3.5 cm would occur by chance only about 3% of the time, so the result is statistically significant. Since plants were randomly assigned, there is convincing evidence that Fertilizer X caused more growth than Fertilizer Y for plants like these.
13. Random selection allows the association between eating breakfast and lower BMI to be generalized to adults in the population. But the study is observational because people chose whether to eat breakfast, so it cannot show that breakfast causes lower BMI. A lurking variable such as overall healthy habits could explain both. The word "helps" is not justified.
14. Statistical significance means a difference is unlikely to be due to chance alone. Practical importance asks whether the size of the difference matters in real life. For example, in an experiment with 1,000,000 randomly assigned users, an app update might raise the mean time on task by 0.2 seconds, a significant difference. But 0.2 seconds may be too small to matter. The size of the difference should be reported with the significance.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

15. Units: 50 volunteer students. Treatments: a 10-minute nap in a quiet room, or 10 minutes of quiet rest awake (a control for time and place). Label the students 01 to 50, use a random number generator to select 25 different labels for the nap group; the rest rest awake. Everyone takes the same attention test at the same time afterward, scored by someone who does not know which group each student was in (single blind, since students know whether they napped). Diagram: Random assignment → Group 1 (25), nap → test score; Group 2 (25), quiet rest → test score; compare mean scores. If the nap group scored significantly higher, we could conclude that the nap caused better attention scores for students like these volunteers, but not for all students.

4.6 Simulation and Random Digits

PRACTICE · SIMULATION AND RANDOM DIGITS

1. A. A is correct: four of the ten equally likely digits give 40%. B gives 50% and D gives 60%. C ignores 0, so the chance of a make is 4 out of 9 valid digits, about 44%, not 40%.
2. B. B is correct: a trial is one full repetition of the process. A is the output of many trials, C confuses one digit with one trial, and D is the count of successes.
3. C. C is correct: five equally likely digits stand for the five colors and the other digits are skipped. A gives color 1 twice the chance, B gives color 5 twice the chance, and D gives color 5 five times the chance of any other color.
4. D. D is correct: $18/50 = 0.36$. A uses the count of successes as if it were out of 100, B ignores the data and C inverts the ratio.
5. A. A is correct: more trials reduce the variability of the estimate. B, C and D change the process being modeled without improving reliability.
6. B. B is correct: the two members must be different, so repeats are ignored. A could select the same person twice, C throws away most numbers and is not a way to choose two people, and D chooses only one person.
7. C. C is correct: it uses a 50% digit assignment, 3 digits per trial and the success condition of at least two heads. A uses one digit per trial, B gives heads only a 30% chance and D requires all three heads.
8. D. D is correct: random variation between runs is normal. A and C treat ordinary variability as a mistake, and B claims a precise answer from two noisy estimates.
9. A. A is correct: the chance is $1/20 = 5\%$, which is 5 of the 100 two-digit numbers (00 to 04). B and D give 20 numbers (20%), and C gives 6 numbers (6%).
10. Digits: 0 and 1 represent a bag with a prize (2 of 10 digits is 20%); 2 to 9 represent no prize. One trial: read digits one at a time (repeats allowed) and count how many bags are needed until the first 0 or 1 appears. A trial is a success if the count is 4 or more (at least 3 failures in a row at the start means the first three digits are all 2 to 9). Run many trials, such as 100. Estimate: number of successes divided by number of trials. (Exact: $0.8^3 = 0.512$.)
11. Digits 0 to 5 = make (60%), 6 to 9 = miss. A trial is 4 digits; success if exactly 3 digits are 0 to 5. The six trials are 7744, 7161, 1341, 4903, 2534, 9009 with makes 2, 2, 4, 3, 4, 2. Successes (exactly 3): 1 of 6, so the estimate is 1/6 = 0.17. With only six trials this is rough; the exact probability is 0.346.
12. Use pairs of digits from 00 to 99 for each day, with 00 to 24 meaning snow (25 of 100 numbers) and 25 to 99 no snow. One trial reads 4 pairs (one per day, repeats allowed) and is a success if at least one pair is from 00 to 24. Repeat for many trials and divide the number of successes by the number of trials. (Exact: $1 - 0.75^4 \approx 0.684$.)
13. With only 20 trials the proportion has a lot of chance variation, so another set of 20 trials might give 9 or 16 wins. The estimate $13/20 = 0.65$ is only one value of a variable estimate. The student could run many more trials, such as 1,000 with a calculator, which makes the estimate more stable, and report it as an estimate.
14. Label the students 01 to 20 with 01 to 08 for the girls and 09 to 20 for the boys. One trial: read two-digit numbers, ignoring 00, 21 to 99 and repeats, until 3 different labels are chosen. Success if all three labels are 01 to 08. Run many trials and estimate the probability as successes divided by trials. (Exact: $\binom{8}{3}/\binom{20}{3} = 56/1140 \approx 0.049$.)
15. The design is invalid for a fair die. Digits 0 to 9 give faces 1 to 5 each a probability of 0.2 and face 6 a probability of 0, so the faces are not equally likely 1/6 each. A valid design uses digits 1 to 6 to represent the faces and ignores 0 and 7 to 9 (or re-reads them), so each face has probability 1/6. Re-rolling does not fix the problem because 6 can never occur.

T Unit 4 Test

UNIT TEST · UNIT 4 TEST

1. A. A is correct: the 31% comes from the sample of 1,200, so it is a statistic that estimates the population parameter. B swaps the terms, C labels both as statistics and D is false because 1,200 of 90,000 is not a census.
2. B. B is correct: households without landlines have no chance of selection (undercoverage), and they tend to be younger, so smartphone ownership is underestimated. A and D treat random selection from a flawed frame, or a large sample, as a cure for bias. C gives the wrong direction.
3. C. C is correct: listeners chose to respond, and those with strong views (often against a change) respond most. A names the wrong bias and direction, B names a random design and D confuses self-selection with nonresponse.
4. D. D is correct: homerooms are clusters, some are sampled and all students in them are measured. A would sample students from every homeroom, B would select students individually, and C would select every kth student.
5. A. A is correct: ignore 00 and numbers above 60, and skip repeats. The distractors ignore these rules or read the digits one at a time.
6. B. B is correct: $k = 1800/60 = 30$. A uses the sample size as k, C divides incorrectly and D subtracts instead of dividing.
7. C. C is correct: no treatment was imposed, and a variable such as exercise could be associated with both vitamin use and fewer colds. A mistakes comparison for experimentation, B ignores confounding and D is false since the groups had the same size.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

8. D. D is correct: the field is separated into blocks (sunny, shady) and treatments are randomly assigned within each block. A has no blocks, B would need blocks of size two, and C describes sampling, not assignment.
9. B. B is correct: using each runner as a block removes runner-to-runner variability. A is about blinding, C is false since the order should still be randomized, and D is not a result of pairing.
10. A. A is correct: random assignment supports causation, volunteers limit generalization. B ignores random assignment, C overgeneralizes and D confuses random sampling with the source of chance in the comparison.
11. C. C is correct: 15% of the 100 two-digit numbers is 15 numbers (00 to 14). A uses 16 numbers (16%), B uses 10 numbers (10%) and D uses 20 numbers (20%).
12. D. D is correct: $z = (84 - 72)/6 = 2$. A forgets to divide by the standard deviation, B has the wrong direction and C divides the standard deviation by the difference.
13. Scoring guide (6 points). (a) 3 points: 1 point for identifying the strata (lower school and upper school); 1 point for correct sample sizes from proportional allocation ($600/1500 \times 100 = 40$ and $900/1500 \times 100 = 60$); 1 point for describing random selection within each stratum (label the students in each stratum, generate random integers without repeats until 40 and 60 distinct labels are found). (b) 1 point: students in the lower and upper school may differ in views on start time (upper school students may be more affected by early mornings), so stratification ensures both groups are represented in the right proportions and reduces variability of the estimate. (c) 2 points: 1 point for naming nonresponse bias with the reason (62% of the chosen students did not respond and may differ from those who did); 1 point for the direction (the proportion supporting a later start time is likely overestimated). Model response: (a) Strata are the lower school (600) and upper school (900). Proportional allocation gives 40 and 60 students. Number the lower-school students 1 to 600 and the upper-school students 1 to 900; use a random integer generator to select 40 different numbers from 1 to 600 and 60 different numbers from 1 to 900, ignoring repeats. (b) The groups may differ in how they feel about start times, so stratified sampling guarantees the correct share of each group and gives a more precise estimate. (c) Nonresponse bias: only 38 of 100 chosen students answered, and students who dislike early mornings are more likely to answer, so the sample proportion would overestimate the proportion of all students who support a later start time.
14. Scoring guide (10 points). (a) 2 points: explanatory variable is the pre-quiz activity (mindfulness exercise or not) [1]; response is the quiz score, and the units are the 36 student volunteers [1]. (b) 4 points: blocks are high-anxiety and low-anxiety students [1]; random assignment within each block, half of each block to each treatment [1]; method stated, such as numbering the 18 students in a block 01 to 18 and using a random integer generator to choose 9 different labels for the mindfulness exercise [1]; a diagram or description showing blocks, random assignment, treatments and measuring the quiz score [1]. (c) 1 point: control group does a neutral 10-minute activity such as quiet reading; the person scoring the quiz does not know which treatment each student received (and students are not told the hypothesis). (d) 3 points: the chance of a difference as large as 5.0 by chance alone is about $14/1000 = 0.014$, so it is unlikely under no effect [1]; because treatments were randomly assigned, there is convincing evidence that the mindfulness exercise caused higher scores for students like these [1]; the students were volunteers from one school and not randomly selected, so we cannot generalize to all students [1]. Model response: (a) The explanatory variable is whether the student does the mindfulness exercise; the response is the quiz score; the units are the 36 volunteer students. (b) Block by anxiety: 18 high-anxiety and 18 low-anxiety students. In each block, number the students 01 to 18 and use a random number generator to choose 9 different numbers for the mindfulness exercise; the other 9 do the control activity. Diagram: Block 1 (18 high anxiety) → random assignment → 9 mindfulness / 9 control → quiz score. Block 2 (18 low anxiety) → random assignment → 9 mindfulness / 9 control → quiz score. Compare scores within blocks. (c) The control group sits quietly with a neutral reading for 10 minutes. The teacher scoring the quizzes is not told who did which activity. (d) Only about 1.4% of random shuffles produce a difference of at least 5.0 points, so the difference is statistically significant and is unlikely to be due to chance. Because of random assignment, the mindfulness exercise caused the higher scores in these students. However, the 36 students were volunteers and not a random sample, so the conclusion should not be generalized to all students.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 5: Probability and Random Variables

Model chance with rules, tables, trees, and probability distributions. 7 lessons, 119 practice and test items, all with full explanations.

In this unit · Exam Unit 2

5.1 Probability Rules and Sample Spaces
5.2 Conditional Probability, Independence, and Tree Diagrams
5.3 Discrete Random Variables: Mean and Standard Deviation
5.4 Combining Random Variables and Transforming Variables
5.5 Binomial Distributions
5.6 Geometric Distributions
5.7 Continuous Random Variables and Uniform Models
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

5.1 Probability Rules and Sample Spaces

Learn to list sample spaces, use the complement, addition and multiplication rules, read Venn diagrams and two-way tables, and tell mutually exclusive events apart from independent ones.

Why This Matters

Every probability question starts with the same two moves: decide exactly what can happen, then decide how the pieces combine. Students lose points not because the arithmetic is hard but because they add when they should multiply, forget to subtract an overlap, or treat "mutually exclusive" and "independent" as the same idea.

These ideas anchor Exam Unit 2 (Probability, Random Variables, and Probability Distributions). Expect them in multiple-choice questions about two-way tables and Venn diagrams, and as the first part of a free-response question that later asks about conditional probability or a random variable.

Core Ideas

A **random process** has outcomes that are individually unpredictable but follow a pattern in the long run. The **sample space** is the list of all possible outcomes, and an **event** is any collection of outcomes. The probability of an event is the long-run relative frequency of that event; by the law of large numbers, the proportion observed settles near the true probability as the number of repetitions grows. A short run of results does not "owe" a correction: a coin that has landed heads five times is still 50% likely to land heads next.

Basic probability facts

Every probability satisfies $0 \le P(A) \le 1$, and the probabilities of all outcomes in the sample space add to 1.

Equally likely outcomes: $P(A) = \dfrac{\text{number of outcomes in } A}{\text{number of outcomes in the sample space}}$.

Complement rule: $P(A^c) = 1 - P(A)$. Use it for \"at least one\" questions.

Addition rule (always true): $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

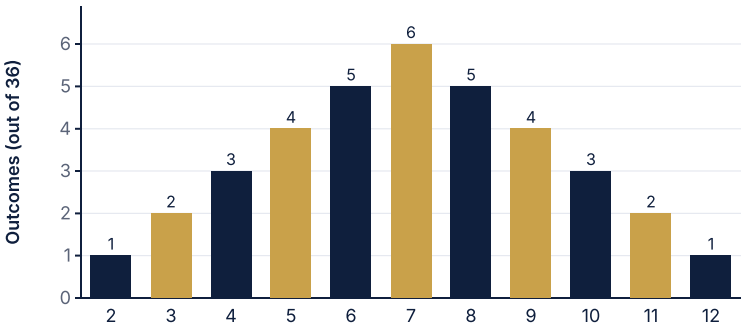
Mutually exclusive events cannot occur together: $P(A \text{ and } B) = 0$, so $P(A \text{ or } B) = P(A) + P(B)$.

When two stages of a process do not influence each other, listing outcomes in a table is a reliable way to build the sample space. Here are all 36 equally likely outcomes for the sum of two fair dice.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Die 1 \ Die 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12



Number of the 36 equally likely outcomes giving each sum of two fair dice.

The bar chart shows why a sum of 7 is the most likely: six of the 36 cells give 7, while only one cell gives 2. Counting cells is the whole method, so be systematic.

Independence and the multiplication rule

Events A and B are **independent** if knowing one occurred does not change the probability of the other.

If A and B are independent: $P(A \text{ and } B) = P(A) \cdot P(B)$. This also works as a test: if $P(A \text{ and } B) \neq P(A)P(B)$, the events are dependent.

Mutually exclusive events with positive probabilities are never independent: if one happens, the other is impossible, so knowing one occurred changes the other's probability to 0.

\"At least one\" of n independent trials: $1 - P(\text{none})$.

A **Venn diagram** draws each event as a circle, with the overlap showing "A and B." The four regions are: only A, both, only B, and neither. Their probabilities add to 1, which is the quickest way to check your work. A **two-way table** carries the same information with counts: cells are "and" events (joint), and row or column totals are single events (marginal).

	Underclass (grades 9-10)	Upperclass (grades 11-12)	Total
Walk	30	20	50
Bus	50	30	80
Car	20	50	70
Total	100	100	200

From the table, $P(\text{Car}) = 70/200 = 0.35$ and $P(\text{Car and Upperclass}) = 50/200 = 0.25$. Divide by the grand total, 200, not by a row or column total (that is a different question, taken up in the next lesson).

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: BUILD A SAMPLE SPACE

A fair coin is flipped and a spinner with three equal sectors (Red, Green, Blue) is spun. List the sample space and find P(heads and not Red).

SOLUTION

The outcomes are H-R, H-G, H-B, T-R, T-G, T-B: 6 equally likely outcomes.

Heads and not Red: H-G and H-B, so 2 outcomes.

$P = 2/6 = 1/3 \approx 0.333$.

EXAMPLE 2: COUNTING EQUALLY LIKELY OUTCOMES

Two fair dice are rolled. Find P(sum is 8), P(sum is at least 10), and P(sum is at most 11).

Die 1 \ Die 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

SOLUTION

Use the 36-cell table above. Sum 8 occurs for (2,6), (3,5), (4,4), (5,3), (6,2): $5/36 \approx 0.139$.

Sum at least 10: sums 10, 11, 12 occur 3 + 2 + 1 = 6 times, so $6/36 = 1/6 \approx 0.167$.

Sum at most 11 is the complement of sum 12: $1 - 1/36 = 35/36 \approx 0.972$.

EXAMPLE 3: AT LEAST ONE, USING THE COMPLEMENT

A device has 4 independent sensors, and each fails during a test with probability 0.05. What is the probability that at least one sensor fails?

SOLUTION

\"At least one fails\" is the complement of \"none fails.\"

Because the sensors are independent, $P(\text{none fail}) = 0.95^4 \approx 0.8145$.

$P(\text{at least one fails}) = 1 - 0.8145 = 0.1855$.

Interpretation: about 18.6% of such tests would see at least one sensor failure.

EXAMPLE 4: THE ADDITION RULE WITH A VENN DIAGRAM

In a school of students, 30% belong to the band, 25% belong to the chess club, and 10% belong to both. A student is chosen at random. Find P(band or chess) and P(neither).

Region	Probability
Only band	0.20
Both	0.10
Only chess	0.15
Neither	0.55
Total	1.00

SOLUTION

Regions: only band $= 0.30 - 0.10 = 0.20$; both $= 0.10$; only chess $= 0.25 - 0.10 = 0.15$.

$P(\text{band or chess}) = 0.20 + 0.10 + 0.15 = 0.45$, which matches $0.30 + 0.25 - 0.10$.

$P(\text{neither}) = 1 - 0.45 = 0.55$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: TWO-WAY TABLE PROBABILITIES AND A PRODUCT CHECK

Use the transportation table above (200 students). Find P(Car or Upperclass), and decide whether riding in a car and being an upperclass student are independent.

	Underclass (grades 9-10)	Upperclass (grades 11-12)	Total
Walk	30	20	50
Bus	50	30	80
Car	20	50	70
Total	100	100	200

SOLUTION

$P(\text{Car}) = 70/200 = 0.35$, $P(\text{Upperclass}) = 100/200 = 0.50$, $P(\text{Car and Upperclass}) = 50/200 = 0.25$.

$P(\text{Car or Upperclass}) = 0.35 + 0.50 - 0.25 = 0.60$. Check by counting: all 100 upperclass students plus the 20 underclass car riders give $120/200 = 0.60$.

Independence test: $P(\text{Car})P(\text{Upperclass}) = 0.35(0.50) = 0.175$, but $P(\text{Car and Upperclass}) = 0.25$.

Since $0.25 \neq 0.175$, the events are not independent: upperclass students are more likely to ride in a car.

EXAMPLE 6: MUTUALLY EXCLUSIVE VERSUS INDEPENDENT

Events A and B have P(A) = 0.40, P(B) = 0.50, and P(A or B) = 0.70. Find P(A and B), then decide whether A and B are mutually exclusive and whether they are independent. Also find P(exactly one occurs).

SOLUTION

Rearrange the addition rule: $P(A \text{ and } B) = 0.40 + 0.50 - 0.70 = 0.20$.

Since $0.20 \neq 0$, the events are not mutually exclusive.

Since $P(A)P(B) = 0.40(0.50) = 0.20 = P(A \text{ and } B)$, the events are independent.

Exactly one occurs: $P(A \text{ or } B) - P(A \text{ and } B) = 0.70 - 0.20 = 0.50$.

Lesson: these events overlap and are independent. If they had been mutually exclusive, $P(A \text{ or } B)$ would have been 0.90.

Write It Like a 5

Prompt: At a school, 40% of students play an instrument (event A), 25% sing in the choir (event B), and 15% do both. A student is selected at random. (a) Find the probability the student plays an instrument or sings in the choir. (b) Find the probability the student does neither. (c) Are A and B independent? Justify.

Model Response

(a) By the addition rule, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.40 + 0.25 - 0.15 = 0.50$. (b) \"Neither\" is the complement of \"A or B,\" so $P(\text{neither}) = 1 - 0.50 = 0.50$. (c) If A and B were independent, $P(A \text{ and } B)$ would equal $P(A)P(B) = 0.40(0.25) = 0.10$. The actual value is 0.15, which is not equal to 0.10, so A and B are not independent. In context, students who play an instrument are more likely than other students to sing in the choir.

What the grader looks for

1 point: correct use of the addition rule with the overlap subtracted. 1 point: correct value 0.50 for (a). 1 point: complement used for (b) with answer 0.50. 1 point: independence tested with a product comparison, giving both numbers. 1 point: conclusion of not independent, stated in the context of instruments and choir, not just "no."

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Name the events in words and with letters before computing anything.
2. Decide whether the question is an "or" (addition), an "and" (multiplication or a table cell), or "at least one" (complement).
3. Draw a two-way table or Venn diagram if two events overlap. Fill the overlap first.
4. Check that all regions or outcomes add to 1 (or to the grand total).
5. To test independence, compare $P(A \text{ and } B)$ with $P(A)P(B)$, using exact values.

- Calculator tip: for dice and coin spaces, count outcomes by hand from a table; for powers such as 0.95^4, key it in directly and keep at least four decimals until the final answer.
- A probability above 1 or below 0 is always an error. Use that as a quick sanity check on every answer.
- The phrase "at least" almost always signals the complement rule.

Common traps

1. Adding P(A) and P(B) for "or" without subtracting the overlap. 2. Saying that mutually exclusive events are independent (they are the opposite, unless one has probability 0). 3. Multiplying probabilities for events that are not independent. 4. Computing "at least one" as the number of trials times the single-trial probability.

Mistake Clinic

Spot the error

A student finds P(A) = 0.5, P(B) = 0.4, P(A and B) = 0.2 and writes P(A or B) = 0.5 + 0.4 = 0.9. Where it fails: Adding both probabilities counts the overlap twice. Correct solution: $P(A \text{ or } B) = 0.5 + 0.4 - 0.2 = 0.7$.

Spot the error

Two events cannot happen at the same time, with P(A) = 0.3 and P(B) = 0.2. The student says they are independent because neither affects the other, so P(A and B) = 0.06. Where it fails: Events that cannot occur together affect each other completely: if A occurs, B is impossible. For these events P(A and B) = 0, not 0.06. Correct solution: The events are mutually exclusive and therefore dependent. $P(A \text{ and } B) = 0$ and $P(A \text{ or } B) = 0.5$.

Spot the error

Probability of at least one six in three rolls of a fair die: 3 times 1/6 = 0.5. Where it fails: Adding single-trial probabilities double counts rolls where more than one six appears. Correct solution: $1 - (5/6)^3 = 1 - 125/216 = 91/216 \approx 0.421$.

Real-World Connection

Engineers use exactly this reasoning for reliability. A backup system built from several independent parts works if at least one part works, so its failure probability is the product of the part failure probabilities. Adding a second part with failure probability 0.10 turns a 0.10 system failure chance into 0.10 times 0.10, or 0.01. Streaming services, hospitals and power grids add redundancy for this reason.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

Four friends each have a birthday month that is equally likely to be any of 12 months, independently. Find the probability that at least two share a birth month. Solution: the complement is that all four months differ. $P(\text{all differ}) = \frac{12}{12} \cdot \frac{11}{12} \cdot \frac{10}{12} \cdot \frac{9}{12} = \frac{11880}{20736} \approx 0.573$, so $P(\text{at least two share}) = 1 - 0.573 = 0.427$.

Practice

PRACTICE **Probability Rules and Sample Spaces**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A fair coin is flipped three times. How many outcomes are in the sample space?
   (A) 6 **(B)** 8 **(C)** 9 **(D)** 3
2. A fair coin is flipped four times. What is the probability of at least one head?
   (A) 1/16 **(B)** 1/4 **(C)** 15/16 **(D)** 3/4
3. For events A and B, P(A) = 0.60, P(B) = 0.30, and P(A or B) = 0.70. What is P(A and B)?
   (A) 0.10 **(B)** 0.18 **(C)** 0.90 **(D)** 0.20
4. Events A and B are mutually exclusive with P(A) = 0.50 and P(B) = 0.30. What is P(A or B)?
   (A) 0.15 **(B)** 0.20 **(C)** 0.80 **(D)** 0.65
5. A subscription service surveyed 100 customers (table). A customer is selected at random. What is P(Basic or Under 30)?

	Under 30	30 and over	Total
Basic	25	35	60
Premium	15	25	40
Total	40	60	100

   (A) 0.75 **(B)** 1.00 **(C)** 0.25 **(D)** 0.60
6. For events A and B, P(A) = 0.40, P(B) = 0.25, and P(A and B) = 0.10. Which statement is true?
   (A) A and B are independent but not mutually exclusive.
   (B) A and B are mutually exclusive but not independent.
   (C) A and B are both independent and mutually exclusive.
   (D) A and B are neither independent nor mutually exclusive.
7. Two fair dice are rolled. What is the probability that the sum is 7 or that the dice show doubles?
   (A) 1/6 **(B)** 11/36 **(C)** 1/3 **(D)** 2/3
8. Each of 3 independent components fails with probability 0.10. What is the probability that at least one component fails?
   (A) 0.300 **(B)** 0.271 **(C)** 0.729 **(D)** 0.001
9. Events A and B are independent with P(A) = 0.30 and P(B) = 0.50. What is P(A or B)?
   (A) 0.80 **(B)** 0.15 **(C)** 0.35 **(D)** 0.65
10. Two of four students (Ana, Ben, Cy, Dee) are chosen as president and vice president, and the order matters.
    (a) List the sample space (12 equally likely outcomes).
    (b) Find P(Ana holds one of the two offices).
    (c) Find P(Ben is president).
11. Among 100 students, 52 take Spanish, 38 take French, and 14 take both.
    (a) Give the number of students in each region of a Venn diagram: only Spanish, both, only French, neither.
    (b) Find P(Spanish or French).
    (c) Find P(exactly one of the two languages).
    (d) Find P(neither).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

12. The table shows 300 residents by housing and pet preference.
Rent: Cats 30, Dogs 50, Neither 20 (total 100). Own: Cats 60, Dogs 100, Neither 40 (total 200).
(a) Find P(Own and Dogs).
(b) Find P(Rent or Cats).
(c) Decide whether owning a home and preferring dogs are independent. Show the numbers.
13. On each of 5 days, rain falls with probability 0.20, independently.
(a) Find the probability of no rain on all 5 days.
(b) Find the probability of at least one rainy day.
(c) Find the probability that it rains on the first day only.
14. Events A and B have P(A) = 0.30 and P(B) = 0.40.
(a) If A and B are mutually exclusive, find P(A or B).
(b) If A and B are independent, find P(A and B) and P(A or B).
(c) Explain why A and B cannot be both mutually exclusive and independent here.
15. A game: roll a fair die and flip a fair coin. You win if the die is even or the coin shows heads.
(a) List which of the 12 outcomes win.
(b) Find P(win) by counting.
(c) Find P(win) again using the complement rule.

5.2 Conditional Probability, Independence, and Tree Diagrams

Compute conditional probabilities from tables and trees, test independence with P(A given B), and use Bayes's rule as a path calculation.

Why This Matters

Real probabilities rarely stand alone. The chance that a part is defective depends on which machine made it; the chance that a test is correct depends on who is being tested. Conditional probability is the tool for updating a probability when you know something extra, and it is where the most frequent probability mistake lives: flipping P(A given B) into P(B given A).

This lesson supports Exam Unit 2 (Probability, Random Variables, and Probability Distributions). Two-way tables, tree diagrams, and "given that" questions appear in multiple choice, and a free-response question may ask you to compute a probability from a tree and then interpret it.

Core Ideas

The **conditional probability** of A given B, written $P(A \mid B)$, is the probability of A when we restrict attention to the cases where B occurred. In a two-way table, that means using the row or column of B as the new "whole."

Conditional probability and the general multiplication rule

$P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$, provided $P(B) > 0$.

General multiplication rule: $P(A \text{ and } B) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$.

Independence: A and B are independent exactly when $P(A \mid B) = P(A)$ (equivalently $P(B \mid A) = P(B)$, or $P(A \text{ and } B) = P(A)P(B)$).

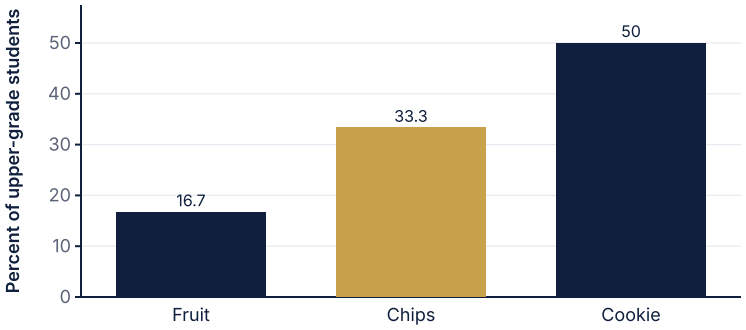
Conditioning on different events gives different answers: in general $P(A \mid B) \neq P(B \mid A)$.

Here is a two-way table of 300 students and their snack choice. To find $P(\text{Fruit} \mid \text{Upper})$ we use only the 150 upper-grade students: $25/150 \approx 0.167$. To find $P(\text{Upper} \mid \text{Fruit})$ we use only the 70 fruit choosers: $25/70 \approx 0.357$. Same cell, different denominators.

	Lower grades	Upper grades	Total
Fruit	45	25	70
Chips	60	50	110
Cookie	45	75	120
Total	150	150	300

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Conditional distribution of snack choice given Upper grades (percents add to 100).

Tree diagrams, total probability, and Bayes's rule

A **tree diagram** shows stages from left to right. First-stage branches carry unconditional probabilities; second-stage branches carry conditional probabilities. Multiply along a path to get an \"and\" probability.

Total probability: add the path probabilities that end in the event: $P(B) = P(A)P(B \mid A) + P(A^c)P(B \mid A^c)$.

Bayes's rule (as a tree computation): $P(A \mid B) = \dfrac{\text{path through A ending in B}}{\text{all paths ending in B}} = \dfrac{P(A)P(B \mid A)}{P(B)}$.

Branches leaving any one node must add to 1.

Bayes's rule sounds advanced, but it is only a conditional probability read from a tree: the numerator is one path, the denominator is the sum of every path that ends in the known outcome. A second route is a **natural-frequency table**: imagine 10,000 people, fill in counts, and divide.

Worked Examples

EXAMPLE 1: CONDITIONAL PROBABILITY FROM A TABLE

Using the snack table, find P(Fruit | Upper) and P(Upper | Fruit), and explain why they differ.

	Lower grades	Upper grades	Total
Fruit	45	25	70
Chips	60	50	110
Cookie	45	75	120
Total	150	150	300

SOLUTION

Given Upper grades, the whole is the 150 upper-grade students, 25 of whom chose fruit: $P(\text{Fruit} \mid \text{Upper}) = 25/150 \approx 0.167$.

Given Fruit, the whole is the 70 fruit choosers, 25 of whom are upper-grade: $P(\text{Upper} \mid \text{Fruit}) = 25/70 \approx 0.357$.

They differ because the conditioning group changed: the denominators are 150 and 70.

EXAMPLE 2: MULTIPLICATION RULE WITHOUT REPLACEMENT

A bag holds 5 red chips and 3 blue chips. Two chips are drawn without replacement. Find P(both red) and P(second chip is blue).

SOLUTION

Both red: $P(R_1)P(R_2 \mid R_1) = \dfrac{5}{8} \cdot \dfrac{4}{7} = \dfrac{20}{56} = \dfrac{5}{14} \approx 0.357$.

Second blue splits by the first chip: $\dfrac{5}{8} \cdot \dfrac{3}{7} + \dfrac{3}{8} \cdot \dfrac{2}{7} = \dfrac{15}{56} + \dfrac{6}{56} = \dfrac{21}{56} = 0.375$.

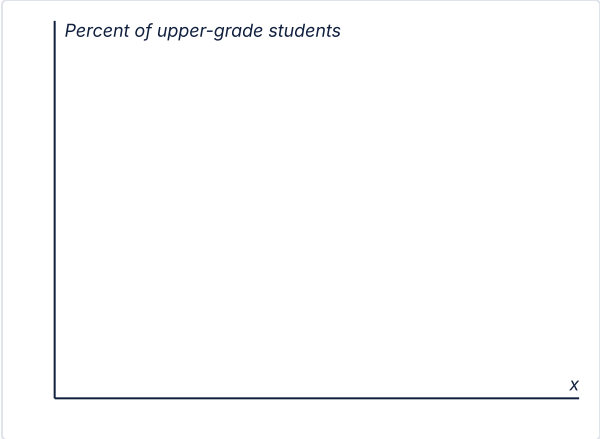
Notice $0.375 = 3/8$, the same as the chance the first chip is blue; without information about the first draw, each position is equally likely to hold a blue chip.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: IS SNACK CHOICE INDEPENDENT OF GRADE GROUP?

Use the snack table to test whether choosing fruit is independent of being in the upper grades.



Conditional distribution of snack choice given Upper grades (percents add to 100).

SOLUTION

$P(\text{Fruit}) = 70/300 \approx 0.233$, but $P(\text{Fruit} \mid \text{Upper}) = 25/150 \approx 0.167$.

Because $P(\text{Fruit} \mid \text{Upper}) \neq P(\text{Fruit})$, the events are not independent.

Context: upper-grade students are less likely to choose fruit than students overall.

EXAMPLE 4: TOTAL PROBABILITY WITH A TREE

Machine A makes 60% of a factory's parts and 3% of its parts are defective. Machine B makes 40% of the parts and 5% of its parts are defective. A part is selected at random. Find P(defective).

Path	Branch probabilities	Probability of path
Machine A, defective	0.60 × 0.03	0.018
Machine A, good	0.60 × 0.97	0.582
Machine B, defective	0.40 × 0.05	0.020
Machine B, good	0.40 × 0.95	0.380
Total		1.000

SOLUTION

Draw a tree: first stage Machine A (0.60) or B (0.40); second stage defective or good.

Defective paths: A then defective $= 0.60(0.03) = 0.018$; B then defective $= 0.40(0.05) = 0.020$.

$P(\text{defective}) = 0.018 + 0.020 = 0.038$.

About 3.8% of all parts are defective.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: BAYES'S RULE FROM THE SAME TREE

A selected part is found to be defective. What is the probability it came from Machine B?

Path	Branch probabilities	Probability of path
Machine A, defective	0.60 × 0.03	0.018
Machine A, good	0.60 × 0.97	0.582
Machine B, defective	0.40 × 0.05	0.020
Machine B, good	0.40 × 0.95	0.380
Total		1.000

SOLUTION

We want $P(B \mid \text{defective})$. The numerator is the path through B ending in defective: 0.020.

The denominator is every defective path: $0.018 + 0.020 = 0.038$.

$P(B \mid \text{defective}) = 0.020/0.038 \approx 0.526$.

Although B makes only 40% of parts, its higher defect rate means it accounts for about 52.6% of defective parts.

EXAMPLE 6: A SCREENING TEST AND NATURAL FREQUENCIES

In a fictional health survey, 8% of people have a condition. A screening test is positive for 90% of people who have it and for 7% of people who do not. Find P(has the condition | positive test) and P(no condition | negative test).

	Positive test	Negative test	Total
Has condition	720	80	800
No condition	644	8,556	9,200
Total	1,364	8,636	10,000

SOLUTION

Imagine 10,000 people: 800 have the condition and 9,200 do not.

Positives: $0.90(800) = 720$ among those with it; $0.07(9{,}200) = 644$ among those without.

$P(\text{condition} \mid \text{positive}) = 720/(720 + 644) = 720/1364 \approx 0.528$.

Negatives: 80 with the condition and $8{,}556$ without, so $P(\text{no condition} \mid \text{negative}) = 8556/8636 \approx 0.991$.

Interpretation: even with a good test, only about half of positives truly have the condition because the condition is uncommon.

Write It Like a 5

Prompt: In a music app, 70% of users have the free plan and 30% have the premium plan. Among free users, 20% listen to podcasts; among premium users, 55% do. A user is chosen at random. (a) Find the probability the user listens to podcasts. (b) Given that the user listens to podcasts, find the probability the user has the premium plan. (c) Are plan type and listening to podcasts independent?

Model Response

(a) By the tree, $P(\text{podcast}) = P(\text{free})P(\text{podcast} \mid \text{free}) + P(\text{premium})P(\text{podcast} \mid \text{premium}) = 0.70(0.20) + 0.30(0.55) = 0.14 + 0.165 = 0.305$. (b) $P(\text{premium} \mid \text{podcast}) = \dfrac{0.30(0.55)}{0.305} = \dfrac{0.165}{0.305} \approx 0.541$. (c) If independent, $P(\text{podcast} \mid \text{premium})$ would equal $P(\text{podcast}) = 0.305$. It is 0.55, so plan type and podcast listening are not independent. In context, premium users are more likely to listen to podcasts than users overall.

What the grader looks for

1 point: both podcast paths identified with correct products. 1 point: sum 0.305. 1 point: Bayes numerator (premium and podcast, 0.165) and the denominator 0.305 shown. 1 point: answer 0.541. 1 point: independence test comparing a conditional probability with the unconditional probability, and a conclusion in context.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Coach's Playbook

1. Read the phrase after "given that" or "if." That event defines the new, smaller whole.
2. Draw a tree whenever a problem has stages, or a table of 1,000 or 10,000 imagined people when a test or rate is described.
3. Write the answer as a fraction of the correct denominator before converting to a decimal.
4. For Bayes, always write numerator = one path and denominator = all paths with the known result.
5. Check: the branch probabilities leaving each node add to 1, and the final path probabilities add to 1.

- Without replacement, the denominators shrink and the conditional probabilities change on the second stage. With replacement, the stages are independent.
- Calculator tip: store intermediate path products and sum them rather than rounding each branch.
- On the exam, label tree branches with the event and its probability so partial credit is easy to give.

Common traps

1. Dividing by the grand total instead of the row or column of the given event. 2. Reporting P(A | B) when the question asks for P(B | A). 3. Using the same probability on both draws when sampling without replacement. 4. Answering a Bayes question with the test's sensitivity (for example, 0.90) instead of computing the probability of the condition given a positive result.

Mistake Clinic

Spot the error

From the snack table a student computes P(Fruit | Upper) = 25/70 = 0.357. Where it fails: 25/70 conditions on Fruit, not Upper. The given event, Upper grades, has 150 students. Correct solution: $P(\text{Fruit} \mid \text{Upper}) = 25/150 \approx 0.167$.

Spot the error

Two chips are drawn without replacement from 5 red and 3 blue. The student writes P(both red) = (5/8)(5/8) = 0.391. Where it fails: After the first red chip is removed, only 4 red chips remain among 7, so the second draw is not 5/8. Correct solution: $P = \frac{5}{8} \cdot \frac{4}{7} = \frac{5}{14} \approx 0.357$.

Spot the error

A condition affects 2% of people. A test is positive for 95% of people with it and 4% of people without it. A student says a positive test means a 95% chance of the condition. Where it fails: 95% is P(positive | condition), not P(condition | positive). The condition is rare, so many positives are false positives. Correct solution: In 10,000 people: 200 have it and 190 test positive; 9,800 do not and 392 test positive. $P(\text{condition} \mid \text{positive}) = 190/(190 + 392) \approx 0.326$.

Real-World Connection

Email spam filters use Bayes's rule. Suppose 30% of messages are spam, the filter flags 95% of spam, and it flags 2% of legitimate mail. Then $P(\text{spam} \mid \text{flagged}) = 0.285/(0.285 + 0.014) \approx 0.953$. The same arithmetic, with different inputs, explains why a rare-disease screening result needs a follow-up test.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

A box contains two coins: one fair, one with heads on both sides. A coin is chosen at random and flipped twice, giving two heads. Find the probability the coin is the two-headed coin. Solution: path two-headed and HH: $0.5(1) = 0.5$. Path fair and HH: $0.5(0.25) = 0.125$. $P = 0.5/(0.5 + 0.125) = 0.8$.

Practice

PRACTICE

Conditional Probability, Independence, and Tree Diagrams

Show your work. A calculator is allowed. Justify conclusions in context.

1. Events A and B have P(A and B) = 0.12 and P(B) = 0.30. What is P(A | B)?
   (A) 0.40 **(B)** 0.04 **(C)** 0.36 **(D)** 0.18
2. Use the snack table (300 students). What is P(Lower grades | Chips)?

	Lower grades	Upper grades	Total
Fruit	45	25	70
Chips	60	50	110
Cookie	45	75	120
Total	150	150	300

   (A) 60/150 **(B)** 60/300 **(C)** 60/110 **(D)** 110/300
3. Events A and B are independent, with P(A) = 0.50 and P(B) = 0.40. What is P(B | A)?
   (A) 0.20 **(B)** 0.40 **(C)** 0.50 **(D)** 0.80
4. A bag holds 4 green and 6 yellow marbles. Two are drawn without replacement. What is P(both green)?
   (A) 0.160 **(B)** 2/15 **(C)** 1/6 **(D)** 0.400
5. P(A) = 0.30, P(B | A) = 0.80, and P(B | not A) = 0.20. What is P(B)?
   (A) 0.24 **(B)** 0.50 **(C)** 0.38 **(D)** 0.80
6. Using the probabilities in the previous item, what is P(A | B)?
   (A) 0.80 **(B)** 0.24 **(C)** 0.30 **(D)** 0.63
7. For events A and B, P(A) = 0.40, P(B) = 0.50, and P(A and B) = 0.10. Which statement is correct?
   (A) A and B are independent because P(A and B) is small.
   (B) P(A | B) = 0.20, so knowing B occurred makes A less likely.
   (C) P(A | B) = 0.25, so A and B are independent.
   (D) P(A | B) = 0.10, the same as P(A and B).
8. On 25% of school days traffic is heavy. On heavy days the bus is late 40% of the time; otherwise 5%. Given that the bus was late, what is the probability that traffic was heavy?
   (A) 0.40 **(B)** 0.10 **(C)** 0.25 **(D)** 0.73
9. A bag has 5 red and 3 blue chips. Two are drawn without replacement. What is P(at least one blue)?
   (A) 3/4 **(B)** 5/14 **(C)** 9/14 **(D)** 15/56
10. A table of 400 students shows sleep and athletics. Athletes: 98 sleep 8 or more hours, 42 do not. Non-athletes: 130 sleep 8 or more hours, 130 do not.
    (a) Find P(8+ hours | athlete).
    (b) Find P(athlete | 8+ hours).
    (c) Is sleeping 8+ hours independent of being an athlete? Justify with numbers.
    (d) Interpret the comparison in context.
11. A raffle has 12 tickets, 4 of them winners. Two tickets are drawn without replacement.
    (a) Find P(both win).
    (b) Find P(exactly one wins).
    (c) Find P(neither wins).
    (d) Verify that (a), (b), (c) add to 1.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

12. At a school, 55% of students ride the bus. Bus riders are late 10% of days; other students are late 4% of days.
    (a) Draw a tree diagram with all probabilities.
    (b) Find P(a random student is late on a day).
    (c) Given that a student is late, find P(bus rider).
13. In a fictional survey of 5,000 people, 4% have a condition. A test is positive for 85% of those with it, and negative for 90% of those without it.
    (a) Make a two-way table of counts.
    (b) Find P(has condition | positive).
    (c) Find P(has condition | negative).
    (d) Explain why (b) is so much smaller than 85%.
14. Events A and B have P(A) = 0.60 and P(B) = 0.50.
    (a) If P(A or B) = 0.80, find P(A and B), P(A | B), and decide on independence.
    (b) If P(A or B) = 0.90 instead, find P(A | B) and decide on independence.
15. A game uses two bags. Bag 1 has 3 red and 2 blue marbles. Bag 2 has 1 red and 4 blue marbles. A fair coin chooses the bag, then one marble is drawn.
    (a) Find P(red).
    (b) Given red, find P(Bag 1).
    (c) Given blue, find P(Bag 2).

5.3 Discrete Random Variables: Mean and Standard Deviation

Build probability distributions for discrete random variables, find and interpret the mean and standard deviation, and use expected value to judge games and prices.

Why This Matters

A random variable turns the outcome of a chance process into a number, so we can ask what to expect and how much results will vary. The mean of a random variable is the foundation for fair prices, insurance, quality control, and, later in the course, the center of every sampling distribution.

This lesson belongs to Exam Unit 2 (Probability, Random Variables, and Probability Distributions). Students most often lose points by averaging the x-values without weights, forgetting to square the deviations, or interpreting a mean as a prediction for one trial.

Core Ideas

A **random variable** is a numerical outcome of a random process, written with a capital letter such as X. A **discrete** random variable has a countable list of possible values, such as the number of pets, the number of defective parts, or the points on a spinner. Its **probability distribution** lists each value with its probability.

Valid distribution and event probabilities

Every $P(X = x)$ is between 0 and 1, and $\sum P(X = x) = 1$.

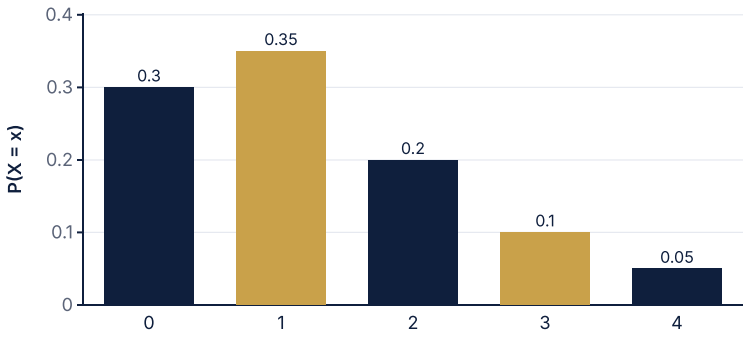
To find the probability of an event such as $X \geq 2$, add the probabilities of every value in that event.

Use the complement when it is shorter: $P(X \geq 1) = 1 - P(X = 0)$.

x (pets)	0	1	2	3	4
P(X = x)	0.30	0.35	0.20	0.10	0.05

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Probability distribution of X, the number of pets in a household (fictional).

Mean and standard deviation of a discrete random variable

Mean (expected value): $\mu_X = E(X) = \sum x_i\, P(x_i)$.

Variance: $\sigma_X^2 = \sum (x_i - \mu_X)^2\, P(x_i)$. Standard deviation: $\sigma_X = \sqrt{\sigma_X^2}$.

Interpretation of the mean: the long-run average value of X over many repetitions of the process. It need not be a possible value of X.

Interpretation of the SD: the typical distance of X from μ_X in repeated trials.

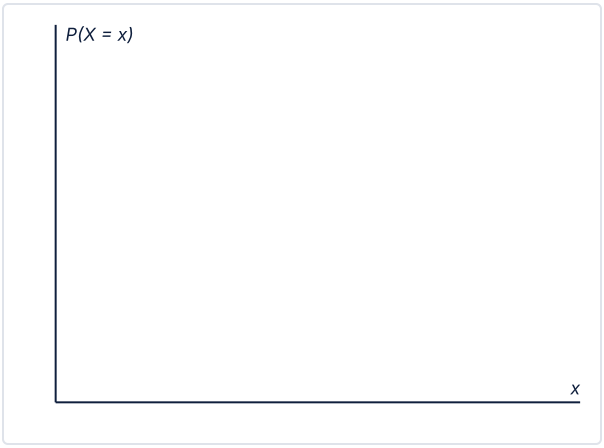
On the built-in calculator, enter the x-values in one list and the probabilities in a second list, then run one-variable statistics with the second list as frequencies. Read $\bar{x}$ as μ_X and σx (the population SD) as σ_X; do not use the sample SD s_x. When probabilities are used as frequencies, the sum of the probabilities is 1, so no extra dividing is needed.

A **fair game** has expected net gain 0. If the expected net gain is negative, the player loses on average the amount in the mean per play, even though single plays can win big.

Worked Examples

EXAMPLE 1: VALID DISTRIBUTION

The number of pets X in a fictional set of households has probabilities 0.30, 0.35, 0.20, 0.10 for x = 0, 1, 2, 3, and an unknown probability for x = 4. Find P(X = 4).



Probability distribution of X, the number of pets in a household (fictional).

SOLUTION

The probabilities must add to 1: $0.30 + 0.35 + 0.20 + 0.10 = 0.95$.

$P(X = 4) = 1 - 0.95 = 0.05$.

All values are between 0 and 1, so the distribution is valid.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: PROBABILITIES OF EVENTS

Using the pet distribution, find P(X $\geq$ 2) and P(1 $\leq$ X $<$ 3).

x (pets)	0	1	2	3	4
P(X = x)	0.30	0.35	0.20	0.10	0.05

SOLUTION

$P(X \geq 2) = 0.20 + 0.10 + 0.05 = 0.35$.

$P(1 \leq X < 3)$ includes $x = 1$ and $x = 2$: $0.35 + 0.20 = 0.55$.

Check with the complement: $P(X \geq 2) = 1 - (0.30 + 0.35) = 0.35$.

EXAMPLE 3: MEAN AND STANDARD DEVIATION

Compute the mean and SD of X, the number of pets, and interpret each in context.

x	P(x)	x * P(x)	(x - 1.25)^2 * P(x)
0	0.30	0.000	0.46875
1	0.35	0.350	0.021875
2	0.20	0.400	0.1125
3	0.10	0.300	0.30625
4	0.05	0.200	0.378125
Total	1.00	1.250	1.2875

SOLUTION

$\mu_X = 0(0.30) + 1(0.35) + 2(0.20) + 3(0.10) + 4(0.05) = 1.25$.

$\sigma_X^2 = (0 - 1.25)^2(0.30) + (1 - 1.25)^2(0.35) + (2 - 1.25)^2(0.20) + (3 - 1.25)^2(0.10) + (4 - 1.25)^2(0.05) = 1.2875$, so $\sigma_X = \sqrt{1.2875} \approx 1.13$.

Interpretation: across many households, the number of pets averages about 1.25 per household, and a household's count typically differs from 1.25 by about 1.13 pets.

EXAMPLE 4: EXPECTED NET GAIN

At a school fair, a game costs 2 tickets to play. A player wins 0 extra tickets with probability 0.60, 3 tickets with probability 0.30, or 10 tickets with probability 0.10. Let X be the net gain in tickets. Find the mean and SD, and decide whether the game is fair.

Net gain x	-2	1	8
P(X = x)	0.60	0.30	0.10

SOLUTION

Net gains: $0 - 2 = -2$, $3 - 2 = 1$, $10 - 2 = 8$ tickets, with probabilities $0.60, 0.30, 0.10$.

$\mu_X = -2(0.60) + 1(0.30) + 8(0.10) = -0.1$ tickets.

$\sigma_X^2 = (-1.9)^2(0.60) + (1.1)^2(0.30) + (8.1)^2(0.10) = 9.09$, so $\sigma_X \approx 3.02$ tickets.

The game is not exactly fair: players lose 0.1 ticket per play on average. The large SD shows single outcomes vary far more than the small mean suggests.

EXAMPLE 5: USING A GIVEN MEAN TO FIND PROBABILITIES

A random variable takes the values 10, 20, and 30 with P(10) = 0.20. Its mean is 22. Find P(20) and P(30).

SOLUTION

Let $a = P(20)$ and $b = P(30)$. Then $a + b = 1 - 0.20 = 0.80$.

Mean equation: $10(0.20) + 20a + 30b = 22$, so $20a + 30b = 20$, or $a + 1.5b = 1$.

Subtract $a + b = 0.80$: $0.5b = 0.20$, so $b = 0.40$ and $a = 0.40$.

Check: $2 + 8 + 12 = 22$. The SD is $\sqrt{(10 - 22)^2(0.2) + (20 - 22)^2(0.4) + (30 - 22)^2(0.4)} = \sqrt{56} \approx 7.48$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 6: AN EXTENDED WARRANTY

A store sells a 40-dollar extended warranty for a gadget. Over the warranty period, repair costs to the store are 0 dollars with probability 0.92, 100 dollars with probability 0.06, and 300 dollars with probability 0.02. Find the store's expected profit per warranty and the SD of the repair cost.

SOLUTION

Expected cost: $\mu = 0(0.92) + 100(0.06) + 300(0.02) = 6 + 6 = 12$ dollars.

Expected profit per warranty $= 40 - 12 = 28$ dollars.

Variance of cost: $(0-12)^2(0.92) + (100-12)^2(0.06) + (300-12)^2(0.02) = 132.48 + 464.64 + 1658.88 = 2256$.

$\sigma = \sqrt{2256} \approx 47.50$ dollars. The store profits on average, but individual repair costs vary a lot relative to the mean, so a store needs many warranties for the average to settle.

Write It Like a 5

Prompt: The number of new songs X that a student adds to a playlist in a day has the distribution below. (a) Show that this is a valid probability distribution. (b) Compute the mean of X. (c) Compute the standard deviation of X. (d) Interpret the mean in context.

x	0	1	2	3
P(X = x)	0.10	0.40	0.30	0.20

Model Response

(a) All probabilities are between 0 and 1 and $0.10 + 0.40 + 0.30 + 0.20 = 1$, so the distribution is valid. (b) $\mu_X = 0(0.10) + 1(0.40) + 2(0.30) + 3(0.20) = 1.6$ songs. (c) $\sigma_X^2 = (0-1.6)^2(0.10) + (1-1.6)^2(0.40) + (2-1.6)^2(0.30) + (3-1.6)^2(0.20) = 0.256 + 0.144 + 0.048 + 0.392 = 0.84$, so $\sigma_X = \sqrt{0.84} \approx 0.917$ songs. (d) Over many days, the student adds an average of 1.6 new songs per day. This does not mean the student adds exactly 1.6 songs on any single day.

What the grader looks for

1 point: both conditions for validity checked (range and sum). 1 point: correct mean with the weighted sum shown. 1 point: variance computed with squared deviations from the mean. 1 point: square root taken and answer 0.917 with units. 1 point: interpretation as a long-run average per day in context, not a prediction for one day.

Coach's Playbook

1. Write the distribution as a two-row table and confirm the probabilities sum to 1 before any other work.
2. For the mean, multiply each x by its probability and add. For the SD, subtract the mean first, square, weight, add, and take the square root last.
3. Enter values and probabilities into two calculator lists, run one-variable statistics with frequencies, and read the population SD.
4. For games, define X as the NET gain (winnings minus cost) so the mean is directly the average profit.
5. Always interpret in context with the words "in the long run" or "on average over many ...".

- Sanity check: the mean lies between the smallest and largest values, closer to the values with the most probability.
- Check the SD against the spread: if every value is within 2 of the mean, the SD cannot be 5.
- A probability histogram's shape (skewed right, symmetric) tells you whether the mean is above or near the center.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Common traps

1. Averaging the x-values without weights. 2. Reporting the variance as the standard deviation. 3. Using the sample SD s_x from the calculator instead of the population SD σ_x. 4. Interpreting the mean as the value you will get on the next trial.

Mistake Clinic

Spot the error

For the playlist distribution (x = 0, 1, 2, 3 with probabilities 0.10, 0.40, 0.30, 0.20), a student finds the mean as (0 + 1 + 2 + 3)/4 = 1.5. Where it fails: This treats all four values as equally likely. The values have different probabilities. Correct solution: $\mu_X = 0(0.10) + 1(0.40) + 2(0.30) + 3(0.20) = 1.6$.

Spot the error

A student finds the variance 0.84 and reports "the standard deviation is 0.84 songs." Where it fails: 0.84 is the variance, which is in squared units. The standard deviation is its square root. Correct solution: $\sigma_X = \sqrt{0.84} \approx 0.917$ songs.

Spot the error

"The mean number of calls per hour is 2.3, so the call center will receive 2.3 calls in the next hour." Where it fails: A count cannot equal 2.3, and the mean describes the long-run average, not one hour. Correct solution: The correct statement: over many hours, the call center averages 2.3 calls per hour. Any single hour might have 0, 1, 2, 3 or more.

Real-World Connection

Insurance companies and extended-warranty sellers price products using expected value. They do not know which customer will file a claim, but they can estimate the distribution of claim sizes. If the price exceeds the mean claim cost, the company profits on average, and with thousands of customers the actual average claim stays near the mean. This is also why a single player can win a carnival game even when the game's expected value is negative.

Challenge Corner

Two fair dice are rolled. Let X be the larger of the two numbers (X = 3 if the dice show 3 and 1, or 3 and 3). Find μ_X and σ_X. Solution: $P(X = k) = (2k - 1)/36$ for $k = 1, \ldots, 6$, giving probabilities $1, 3, 5, 7, 9, 11$ over 36. $\mu_X = (1 + 6 + 15 + 28 + 45 + 66)/36 = 161/36 \approx 4.472$. $E(X^2) = (1 + 12 + 45 + 112 + 225 + 396)/36 = 791/36 \approx 21.972$, so $\sigma_X^2 = 21.972 - 4.472^2 \approx 1.972$ and $\sigma_X \approx 1.404$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

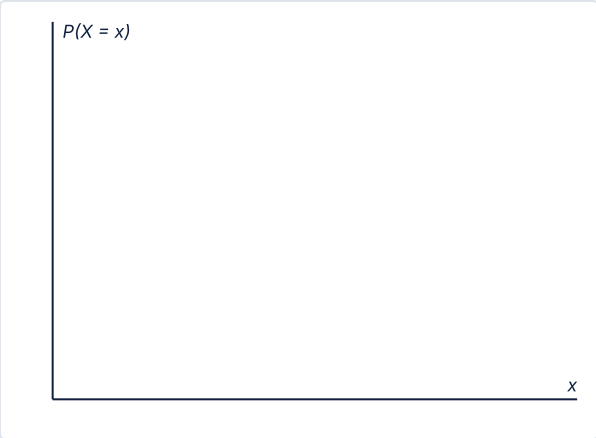
© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE Discrete Random Variables: Mean and Standard Deviation

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which table could be the probability distribution of a discrete random variable?
   (A) x = 1, 2, 3: probabilities 0.5, 0.3, 0.3
   (B) x = 1, 2, 3: probabilities 0.5, 0.3, 0.2
   (C) x = 1, 2, 3: probabilities 0.6, 0.5, -0.1
   (D) x = 1, 2, 3: probabilities 0.2, 0.2, 0.2
2. A random variable takes the values 1, 2, 3, 4 with probabilities 0.2, 0.3, p, 0.1. What is p?
   (A) 0.4 **(B)** 0.6 **(C)** 0.3 **(D)** 0.1
3. A random variable X takes the values 0, 1, 2 with probabilities 0.5, 0.3, 0.2. What is the mean of X?
   (A) 1.0 **(B)** 0.3 **(C)** 0.7 **(D)** 0.5
4. For the same random variable (0, 1, 2 with probabilities 0.5, 0.3, 0.2), what is the standard deviation?
   (A) 0.610 **(B)** 0.700 **(C)** 1.100 **(D)** 0.781
5. A call center has a mean of 2.3 calls per hour. Which interpretation is correct?
   (A) In any given hour, the center receives exactly 2.3 calls.
   (B) The center receives 2.3 calls in at least half of the hours.
   (C) Over many hours, the number of calls per hour averages about 2.3.
   (D) Every hour has between 2 and 3 calls.
6. A game costs 3 dollars to play. A player receives 10 dollars with probability 0.20 and otherwise receives nothing. What is the expected net gain per play?
   (A) -1.00 dollars **(B)** 2.00 dollars **(C)** -3.00 dollars
   (D) -2.00 dollars
7. The graph shows the distribution of X, the number of correct answers on a 4-question warm-up. What is $P(X \geq 3)$?



Distribution of X, the number of correct answers in a 4-question warm-up.

   (A) 0.1 **(B)** 0.7 **(C)** 0.9 **(D)** 0.3
8. Which variable is a discrete random variable?
   (A) The exact height of a randomly chosen student
   (B) The exact time to run a mile
   (C) The number of heads in 10 coin flips
   (D) The exact weight of a backpack
9. X takes the values 1, 2, 5 with probabilities 0.4, 0.4, 0.2. What is the mean of X?
   (A) 2.67 **(B)** 2.20 **(C)** 1.00 **(D)** 0.88

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. The number of siblings X for a randomly chosen student has probabilities 0.18, 0.37, 0.27, p, 0.06 for x = 0, 1, 2, 3, 4.
    (a) Find p.
    (b) Find P(X $\geq$ 2).
    (c) Find the mean and SD of X.
    (d) Interpret the mean in context.
11. A game costs 4 tickets to play. The player wins 0 tickets with probability 0.55, 5 tickets with probability 0.30, and 12 tickets with probability 0.15.
    (a) Define X as the net gain and make its distribution table.
    (b) Find the mean of X.
    (c) Find the SD of X.
    (d) Is the game fair? Explain, and interpret the mean.
12. A store sells a 40-dollar warranty. Repair costs are 0 dollars (probability 0.92), 100 dollars (0.06), and 300 dollars (0.02).
    (a) Find the expected repair cost.
    (b) Find the store's expected profit on 1,000 warranties.
    (c) Explain why the store is more certain of its average profit with 1,000 warranties than with 5.
13. A random variable takes the values 2, 4, 6, 8. P(2) = 0.10 and P(8) = 0.20. The mean of X is 5.6.
    (a) Find P(4) and P(6).
    (b) Find the SD of X.
    (c) Is the distribution valid? Explain.
14. The number of text alerts X a student gets in an hour has probabilities 0.20, 0.30, 0.25, 0.15, 0.07, 0.03 for x = 0, 1, 2, 3, 4, 5.
    (a) Find the mean and SD of X.
    (b) Find P(X $\geq$ 4).
    (c) Is getting 5 alerts unusual? Support your answer.
15. Two fair four-sided dice (faces 1 to 4) are rolled. Let X be the sum.
    (a) Make the probability distribution of X.
    (b) Find P(X $\geq$ 6).
    (c) Find the mean and SD of X.

5.4 Combining Random Variables and Transforming Variables

Find the mean and standard deviation of linear transformations, sums, and differences of random variables, and see why variances add while standard deviations do not.

Beyond the 2027 Exam Core

College Board's revision page lists combining random variables as removed from the 2027 exam. This lesson is extension material that appears in college statistics courses and older exams. Linear transformations of data (adding or multiplying by a constant) are still worth knowing from Unit 1. Check your teacher's current course description to see how much of the rest to study.

Why This Matters

Real quantities are built from pieces: a trip is a walk plus a ride, a profit is a revenue minus a cost, a score is a raw count converted to a scale. This lesson shows how the mean and standard deviation of the new quantity follow from those of the pieces. It also explains the single most surprising fact in the unit: variability adds, but standard deviations do not.

Students most often lose points by subtracting standard deviations for a difference, by changing the SD when only a constant is added, and by treating 2X (one variable doubled) as X1 + X2 (two independent copies).

Core Ideas

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Linear transformations: Y = a + bX

$\mu_Y = a + b\,\mu_X$ (the mean follows the same rule as the values).

$\sigma_Y = |b|\,\sigma_X$ (adding the constant a shifts everything and does not change spread; multiplying by b stretches the spread by $|b|$).

The shape of the distribution does not change, except that a negative b flips it left to right.

Example: converting Celsius to Fahrenheit, $F = 32 + 1.8C$, turns a mean of 20 degrees and an SD of 5 into a mean of 68 and an SD of 9.

Sums and differences of random variables

For any random variables: $\mu_{X+Y} = \mu_X + \mu_Y$ and $\mu_{X-Y} = \mu_X - \mu_Y$.

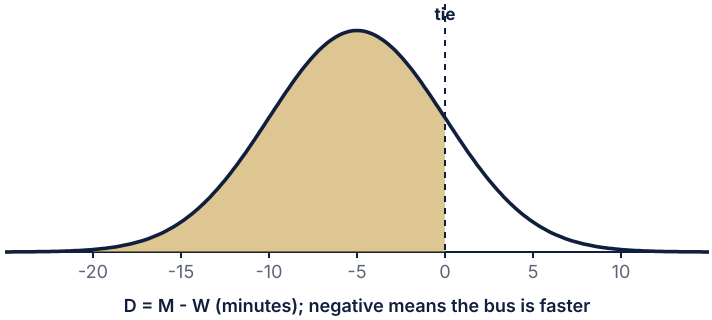
If X and Y are **independent**: $\sigma^2_{X\pm Y} = \sigma^2_X + \sigma^2_Y$, so $\sigma_{X\pm Y} = \sqrt{\sigma^2_X + \sigma^2_Y}$. Both a sum and a difference have this SD.

A sum or difference of independent normal random variables is also normal.

Without independence the variance rule can fail, so state the independence assumption.

The key reason variances add for differences is that subtracting a variable does not remove its uncertainty: the difference between a quantity that can be high or low and another that can be high or low is even harder to predict. The next table compares a transformation, a sum of independent copies, and a doubled copy for the number of songs X from the previous lesson (mean 1.6, SD 0.917).

Quantity	Mean	SD
X (songs on one day)	1.6	0.917
3X - 2	2.8	2.750
X1 + X2 (two different days)	3.2	1.296
2X (same day counted twice)	3.2	1.833



Distribution of the difference in trip times; the shaded area is P(D less than 0) = 0.841.

Worked Examples

EXAMPLE 1: LINEAR TRANSFORMATION OF SCORES

A quiz has score mean 72 and SD 9 points. The teacher curves every score with Y = 1.1X + 3. Find the mean and SD of the curved scores, and say whether the shape changes.

SOLUTION

$\mu_Y = 1.1(72) + 3 = 82.2$.

$\sigma_Y = |1.1|(9) = 9.9$. Adding 3 does not change the spread.

The shape is unchanged because the transformation has positive slope; the curved scores are a stretched and shifted copy of the originals.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: SUM OF INDEPENDENT VARIABLES

A shipping box's contents weigh X kilograms with mean 3.2 and SD 0.4, and its packaging weighs Y kilograms with mean 0.5 and SD 0.1. Assume X and Y are independent. Find the mean and SD of the total weight.

SOLUTION

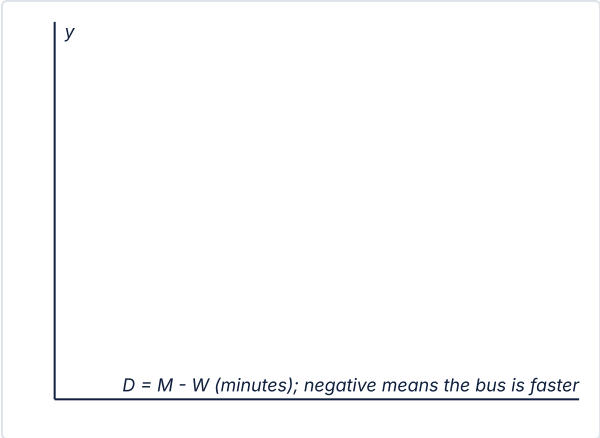
$\mu_{X+Y} = 3.2 + 0.5 = 3.7$ kg.

$\sigma_{X+Y} = \sqrt{0.4^2 + 0.1^2} = \sqrt{0.17} \approx 0.412$ kg.

The SD is not $0.4 + 0.1 = 0.5$: standard deviations of independent variables combine like the legs of a right triangle.

EXAMPLE 3: A DIFFERENCE OF NORMAL VARIABLES

A student's bus trip M has a normal distribution with mean 25 and SD 4 minutes. Walking W has mean 30 and SD 3 minutes. Assume independence. Find the probability that the bus trip is faster than the walk.



Distribution of the difference in trip times; the shaded area is P(D less than 0) = 0.841.

SOLUTION

Let $D = M - W$. $\mu_D = 25 - 30 = -5$. $\sigma_D = \sqrt{4^2 + 3^2} = 5$.

D is normal because M and W are independent normal variables.

We need $P(D < 0)$: $z = (0 - (-5))/5 = 1.00$, so $P = 0.841$.

About 84% of days the bus trip is shorter than the walk.

EXAMPLE 4: 2X VERSUS X1 + X2

In the previous lesson, the number of new songs added in a day had mean 1.6 and SD 0.917. Find the mean and SD of (a) Y = 3X - 2, (b) T = X1 + X2 for two independent days, and (c) W = 2X.

Quantity	Mean	SD
X (songs on one day)	1.6	0.917
3X - 2	2.8	2.750
X1 + X2 (two different days)	3.2	1.296
2X (same day counted twice)	3.2	1.833

SOLUTION

(a) $\mu_Y = 3(1.6) - 2 = 2.8$ and $\sigma_Y = 3(0.917) = 2.75$.

(b) $\mu_T = 1.6 + 1.6 = 3.2$ and $\sigma_T = \sqrt{0.917^2 + 0.917^2} = 0.917\sqrt{2} \approx 1.296$.

(c) $\mu_W = 2(1.6) = 3.2$ and $\sigma_W = 2(0.917) \approx 1.834$.

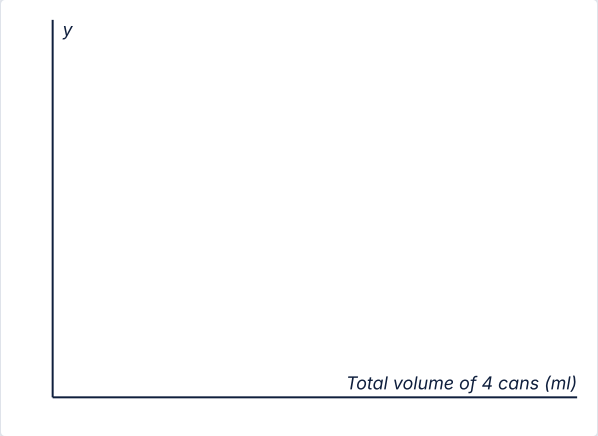
Same mean for (b) and (c), different SD: two independent days partly average out high and low days, while doubling one day doubles everything, including its extreme values.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: TOTAL OF INDEPENDENT NORMAL QUANTITIES

A cup-filling machine puts a normal amount of soup into each can, mean 250 ml and SD 4 ml. A carton holds 4 cans, with independent fills. What is the probability the carton contains less than 990 ml in total?



Total volume of four independent cans; the shaded area is P(total less than 990) = 0.106.

SOLUTION

Total $T = X_1 + X_2 + X_3 + X_4$ is normal with $\mu_T = 4(250) = 1000$ and $\sigma_T = \sqrt{4(4^2)} = 8$ ml.

$z = (990 - 1000)/8 = -1.25$.

$P(T < 990) = 0.1056$. About 10.6% of cartons are under 990 ml.

Note: this is the SD of the sum; the SD of the average per can would be $8/4 = 2$ ml.

EXAMPLE 6: PROFIT FROM A FOOD STAND

A food stand's daily sales S are normal with mean 800 dollars and SD 120 dollars. Costs are a fixed 300 dollars plus 35% of sales. So profit is P = S - (300 + 0.35S) = 0.65S - 300. Find the mean and SD of daily profit, the probability of a loss on a day, and the mean and SD of total profit over two independent days.

SOLUTION

$\mu_P = 0.65(800) - 300 = 220$ dollars. $\sigma_P = 0.65(120) = 78$ dollars.

A loss means $P < 0$, which is $S < 300/0.65 \approx 461.5$. $z = (461.5 - 800)/120 \approx -2.82$, so $P(\text{loss}) \approx 0.0024$.

Two independent days: mean $220 + 220 = 440$ dollars, SD $\sqrt{78^2 + 78^2} = 78\sqrt{2} \approx 110.3$ dollars.

The fixed cost shifts the mean but does not affect the SD.

Write It Like a 5

Prompt: On a practice assessment, the reading score R has mean 52 and SD 8, and the math score M has mean 48 and SD 6. Assume the two scores are independent. Let T = R + M and D = R - M. (a) Find the mean and SD of T. (b) Find the mean and SD of D. (c) A classmate says D has SD 8 - 6 = 2. Explain the error.

Model Response

(a) $\mu_T = 52 + 48 = 100$. Since R and M are independent, $\sigma_T = \sqrt{8^2 + 6^2} = \sqrt{100} = 10$. (b) $\mu_D = 52 - 48 = 4$. Because the scores are independent, variances add even for a difference: $\sigma_D = \sqrt{8^2 + 6^2} = 10$. (c) The classmate subtracted standard deviations. Standard deviations do not subtract: the variance of a difference is the sum of the variances, $64 + 36 = 100$, so the SD is 10. Intuitively, the difference of two unpredictable scores is at least as unpredictable as either one.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

1 point: correct mean of T. 1 point: correct SD of T using a square root of summed variances with independence stated. 1 point: correct mean of D. 1 point: correct SD of D. 1 point: explanation that variances add for differences and standard deviations cannot be subtracted.

Coach's Playbook

1. Name the new variable (Y = a + bX, T = X + Y, D = X - Y) and its parts.
2. Mean first: apply the same arithmetic to the means.
3. SD second: for a linear change use $|b|\sigma$; for combinations of independent variables add VARIANCES, then take the square root.
4. If probabilities are requested and the pieces are normal, the combination is normal: standardize and use the normal probability function.
5. State independence explicitly, because the variance rule requires it.

- Do the squares on the calculator without rounding: $\sqrt{0.4^2 + 0.1^2}$ should be typed as one expression.
- A quick check: the SD of a sum of independent variables lies between the larger SD and the sum of the SDs.
- For an average of n independent values with SD σ, the SD is $\sigma/\sqrt{n}$; this is the key idea that returns in the sampling distribution unit.

Common traps

1. Subtracting standard deviations for a difference. 2. Changing the SD when a constant is added. 3. Using $2X$ and $X_1 + X_2$ interchangeably. 4. Adding standard deviations instead of variances for a sum of independent variables.

Mistake Clinic

Spot the error

R has SD 8 and M has SD 6 (independent). A student reports SD of R - M as 8 - 6 = 2. Where it fails: Variability does not cancel when subtracting. The variance of a difference of independent variables is the sum of variances. Correct solution: $\sigma_{R-M} = \sqrt{8^2 + 6^2} = 10$.

Spot the error

Every score in a class is increased by 10 points. A student says the standard deviation of 9 becomes 19. Where it fails: Adding a constant shifts every value by the same amount; the distances between values do not change. Correct solution: The mean increases by 10 and the SD stays at 9.

Spot the error

A student writes the SD of two independent days of songs, X1 + X2, as 2(0.917) = 1.834. Where it fails: That is the SD of 2X, a single day doubled. Two independent days average out some variability. Correct solution: $\sigma_{X_1+X_2} = \sqrt{0.917^2 + 0.917^2} \approx 1.296$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Airlines and delivery services use sums of independent weights, and manufacturers use stack-up analysis: if five independent parts are stacked in a machine, the SD of the total height is the square root of the sum of the five variances, not five times a single SD. Finance uses the same idea, in a more complicated form, when diversifying a portfolio reduces total risk.

Challenge Corner

A length is measured repeatedly with independent measurements, each with SD 2 mm. How many measurements must be averaged so the average has SD at most 0.5 mm? Solution: the average $\bar{X} = (X_1 + \cdots + X_n)/n$ has variance $n\sigma^2/n^2 = \sigma^2/n$, so SD $= 2/\sqrt{n}$. Set $2/\sqrt{n} \le 0.5$: $\sqrt{n} \ge 4$, $n \ge 16$.

Practice

PRACTICE **Combining Random Variables and Transforming Variables**

Show your work. A calculator is allowed. Justify conclusions in context.

1. X has mean 5 and SD 2. Y = 4 - 3X. What is the mean of Y?
   (A) -11 **(B)** 11 **(C)** 19 **(D)** -19
2. X has mean 5 and SD 2. Y = 4 - 3X. What is the standard deviation of Y?
   (A) -6 **(B)** 2 **(C)** 6 **(D)** 10
3. X and Y are independent with SD(X) = 3 and SD(Y) = 4. What is the SD of X + Y?
   (A) 7 **(B)** 5 **(C)** 25 **(D)** 3.5
4. X and Y are independent, with SD(X) = 3 and SD(Y) = 4. Which statement about X - Y is correct?
   (A) SD(X - Y) = 1
   (B) SD(X - Y) = 7
   (C) SD(X - Y) cannot be found.
   (D) SD(X - Y) = 5, the same as SD(X + Y)
5. X is a random variable with SD 5. X1 and X2 are independent copies of X. Compare the SD of 2X and the SD of X1 + X2.
   (A) Both equal 10.
   (B) 2X has SD 10, and X1 + X2 has SD about 7.07.
   (C) 2X has SD about 7.07, and X1 + X2 has SD 10.
   (D) Both equal about 7.07.
6. Every value in a data set is increased by 5. Which statement is true?
   (A) The mean and SD both increase by 5.
   (B) The mean is unchanged and the SD increases by 5.
   (C) Both are unchanged.
   (D) The mean increases by 5 and the SD is unchanged.
7. X ~ N(100, 6) and Y ~ N(90, 8) are independent. What is P(X > Y)?
   (A) 0.762 **(B)** 0.159 **(C)** 0.841 **(D)** 0.691
8. Temperatures in Celsius have mean 20 and SD 5. With F = 32 + 1.8C, what is the SD of the Fahrenheit temperatures?
   (A) 9 **(B)** 41 **(C)** 5 **(D)** 14
9. Three independent tasks each have mean 10 minutes and SD 2 minutes. What is the SD of the total time?
   (A) 6 **(B)** 3.46 **(C)** 2 **(D)** 12
10. A set of exam scores has mean 74, SD 8, median 76, and IQR 12. Every score X is changed by Y = 1.1X + 5.
    (a) Find the mean, SD, median, and IQR of Y.
    (b) Does the shape change? Explain.
11. Machine A fills a bag with a mean of 500 g and SD 3 g. Machine B fills with a mean of 502 g and SD 4 g. Both are normal and independent. One bag is chosen from each.
    (a) Find the mean and SD of the total A + B.
    (b) Find the mean and SD of B - A.
    (c) Find the probability that the B bag is heavier than the A bag.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

12. X is the number of items a customer buys, with P(0) = 0.5, P(1) = 0.3, P(2) = 0.2 (mean 0.7, SD 0.781). The store earns 6 dollars per item and has a 4-dollar cost per customer, so profit is Y = 6X - 4.
    (a) Find the mean and SD of Y.
    (b) For two independent customers, find the mean and SD of total profit.
13. Backpack weights are normal with mean 7 kg and SD 1.2 kg, independently. A rack safely holds 40 kg.
    (a) Find the mean and SD of the total weight of 5 backpacks.
    (b) Find the probability the total exceeds 40 kg.
14. Flip a fair coin twice. X is the number of heads and Y is the number of tails; each has variance 0.5.
    (a) Explain why a student gets 1 for the variance of X + Y if independence is assumed.
    (b) Find the actual variance of X + Y.
    (c) What does this show about the rule for adding variances?
15. Scores X have mean 65 and SD 8. Let Z = (X - 65)/8 and W = 100 - 2X.
    (a) Find the mean and SD of Z.
    (b) Find the mean and SD of W.
    (c) Why is the SD of W positive even though the slope is negative?

5.5 Binomial Distributions

Check the binomial conditions (BINS), compute exact probabilities, means and standard deviations, use calculator functions, and see how the normal approximation works as an extension.

Why This Matters

Many statistics problems count successes in a fixed number of tries: how many of 20 survey respondents prefer a brand, how many of 200 parts are defective, how many quiz answers you get right by guessing. The binomial distribution handles all of them with one formula and two shortcuts for the mean and SD.

It is a core idea of Exam Unit 2 (Probability, Random Variables, and Probability Distributions) and it sits behind the inference procedures for proportions in the units that follow. The most common errors are using the model without checking conditions and mixing up the pdf with the cdf.

Core Ideas

A random variable X is **binomial** if it counts the number of successes in n trials that meet four conditions, summarized as BINS. We write $X \sim \text{Binomial}(n, p)$ or $X \sim B(n, p)$.

Letter	Condition	Check in context
B	Binary: each trial is success or failure	Define 'success' clearly
I	Independent trials	Random sampling, or sampling under 10% of the population
N	Number of trials is fixed in advance	n is known before starting
S	Same probability of success on every trial	p does not change

The binomial formulas

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, \ldots, n$, where $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ counts the arrangements of k successes among n trials.

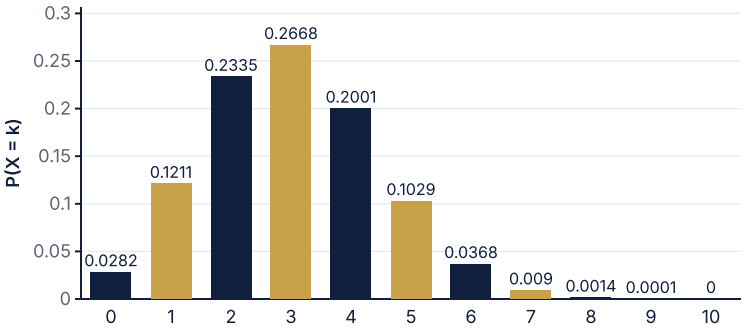
Mean: $\mu_X = np$. Standard deviation: $\sigma_X = \sqrt{np(1-p)}$.

When sampling without replacement from a population, treat the count as approximately binomial if the sample is less than 10% of the population (the 10% condition).

The distribution is skewed right when $p < 0.5$, symmetric when $p = 0.5$, and skewed left when $p > 0.5$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Binomial distribution with n = 10 and p = 0.3: skewed right, mean 3.

Calculator and probability words

Exactly k: binomial pdf, binompdf(n, p, k).

At most k ($X \leq k$): binomcdf(n, p, k).

Fewer than k: binomcdf(n, p, k-1). At least k: 1 - binomcdf(n, p, k-1). More than k: 1 - binomcdf(n, p, k).

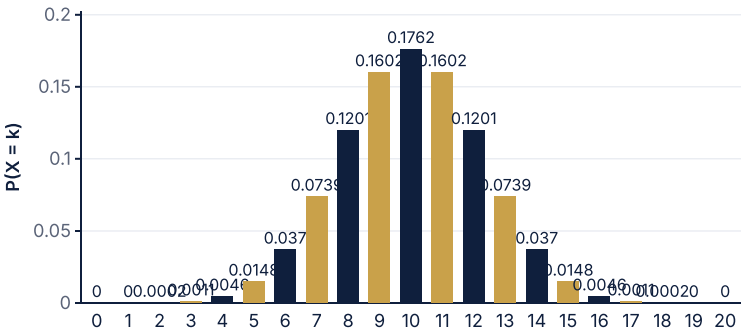
Between a and b inclusive: binomcdf(n, p, b) - binomcdf(n, p, a-1).

Wording in the problem	Event in terms of X	Calculator setup (n = 10, p = 0.3)
exactly 3	X = 3	binompdf(10, 0.3, 3) = 0.2668
at most 3; no more than 3	X $\leq$ 3	binomcdf(10, 0.3, 3) = 0.6496
fewer than 3	X $\leq$ 2	binomcdf(10, 0.3, 2) = 0.3828
at least 3; 3 or more	X $\geq$ 3	1 - binomcdf(10, 0.3, 2) = 0.6172
more than 3	X $\geq$ 4	1 - binomcdf(10, 0.3, 3) = 0.3504

Translate the words first, then the calculator. Most binomial errors are really reading errors: "fewer than 3" and "at most 3" differ by exactly one value, and that one value is worth a point.

Writing about binomial results. Always connect the number back to the situation. A probability of 0.2668 means that in about 27 of every 100 random samples of 10 students, exactly 3 would prefer online textbooks. A mean of 3 means that the count averages 3 over many such samples, not that every sample has 3. And an unusual result (a small probability, or a count more than 2 SDs from the mean) is evidence that the assumed value of p may be wrong; that reasoning is the seed of significance testing in later units.

Extension: normal approximation. When $np \geq 10$ and $n(1-p) \geq 10$, the binomial distribution is close to normal with mean np and SD $\sqrt{np(1-p)}$, and probabilities can be estimated with normal calculations. Using a continuity correction (for example, $P(X \geq 240) \approx P(Y \geq 239.5)$) improves accuracy. With a calculator available, the exact binomial answer is preferred; the approximation is useful for understanding why sampling distributions of proportions look normal.



Binomial distribution with n = 20 and p = 0.5: symmetric, mean 10, SD 2.24.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Why the formula works. Any one particular arrangement of k successes and $n-k$ failures, such as SSFSF, has probability $p^k(1-p)^{n-k}$ because the trials are independent. The number of different arrangements of k successes among n positions is $\binom{n}{k}$, and these arrangements are mutually exclusive, so we add their probabilities. That is exactly the formula: a count of arrangements times the probability of one arrangement.

Reading the shape. The mean np is the balance point. When p is small the distribution piles up near 0 with a long right tail, so the mean is larger than the most likely value. When p is large the long tail points left. When $p = 0.5$ the distribution is symmetric about $n/2$. As n grows, the shape becomes more bell-shaped, but the SD $\sqrt{np(1-p)}$ grows only like $\sqrt{n}$, so the distribution is relatively narrower in proportion to its mean.

Choosing what counts as a success. You may label either outcome a success, but be consistent. If 6% of screens are flawed, a flawed screen can be the success ($p = 0.06$) or a good screen can be ($p = 0.94$); the probabilities of events will agree as long as the question is translated correctly. Usually the outcome the question counts is the one to label a success. Also remember that X counts successes only, so a question about the number of failures needs $n - X$.

The 10% condition. Drawing without replacement from a small population changes the probability after each draw. When the sample is under 10% of the population, the change is tiny and the binomial model is an excellent approximation; when it is not (for example, 5 cards from a 52-card deck), use another method, such as the multiplication rule from the earlier lesson.

Worked Examples

EXAMPLE 1: CHECKING BINS

Decide whether each situation can be modeled with a binomial distribution.

Situation	Binomial?	Reason
Count germinating seeds among 12 planted, each with chance 0.85, independently	Yes	All four BINS conditions hold
Number of cards drawn from a deck until the first ace	No	Number of trials is not fixed
Count hearts in 5 cards from one deck, no replacement	No	Trials are dependent; 5 is more than 10% of 52

SOLUTION

The seeds: 12 trials (fixed), success is germination, independent, same probability 0.85. Binomial.

Cards until the first ace: the number of trials is not fixed. Not binomial (this is a geometric situation, covered in the next lesson).

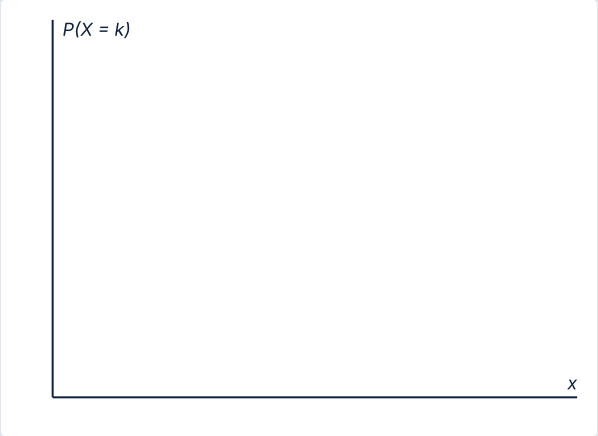
Hearts in 5 cards without replacement: probability of a heart changes after each card, and 5 is more than 10% of 52. Not binomial.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: A SINGLE BINOMIAL PROBABILITY

In a fictional large school, 30% of students prefer online textbooks. Ten students are randomly selected. Find the probability that exactly 3 prefer online textbooks.



Binomial distribution with n = 10 and p = 0.3: skewed right, mean 3.

SOLUTION

Let X be the number preferring online books. Conditions: two outcomes, independent (random sample under 10% of a large school), $n = 10$ fixed, $p = 0.3$ each. So $X \sim B(10, 0.3)$.

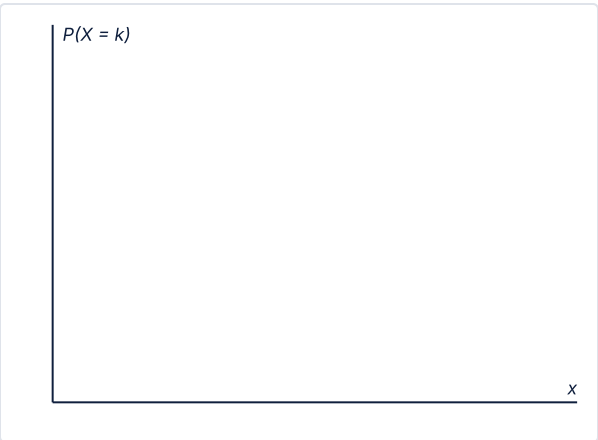
$$P(X = 3) = \binom{10}{3}(0.3)^3(0.7)^7 = 120(0.027)(0.0823543) \approx 0.2668.$$

Check on the calculator: binompdf(10, 0.3, 3) = 0.2668.

There is about a 26.7% chance that exactly 3 of the 10 prefer online textbooks.

EXAMPLE 3: CUMULATIVE PROBABILITIES

For X ~ B(10, 0.3), find P(X $\leq$ 2), P(X $\geq$ 4), and P(2 $\leq$ X $\leq$ 5).



Binomial distribution with n = 10 and p = 0.3: skewed right, mean 3.

SOLUTION

$P(X \leq 2) = \text{binomcdf}(10, 0.3, 2) = 0.3828.$

$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.6496 = 0.3504.$

$P(2 \leq X \leq 5) = P(X \leq 5) - P(X \leq 1) = 0.9527 - 0.1493 = 0.8033.$

Check against the histogram: the bars at 2 through 5 are the tall middle bars, so a probability above 0.8 makes sense.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 4: MEAN, SD, AND UNUSUAL COUNTS

A machine produces a part that is defective with probability 0.04, independently. In a batch of 200 parts, find the mean and SD of the number of defective parts. Is 15 defective parts unusual?

SOLUTION

$X \sim B(200, 0.04)$, with $\mu = 200(0.04) = 8$ defective parts.

$\sigma = \sqrt{200(0.04)(0.96)} = \sqrt{7.68} \approx 2.77$.

Fifteen defective parts is $(15 - 8)/2.77 \approx 2.53$ SDs above the mean, and the exact probability $P(X \geq 15) \approx 0.015$.

A batch with 15 or more defective parts would happen only about 1.5% of the time, so it suggests the defect rate may have risen.

EXAMPLE 5: AT LEAST, WITH CONTEXT

A seed packet claims each seed germinates with probability 0.85. A student plants 12 seeds. Find the probability that at least 10 germinate and the probability that all 12 do.

SOLUTION

$X \sim B(12, 0.85)$ (12 seeds, independent, same probability).

$P(X \geq 10) = 1 - P(X \leq 9) = 1 - \text{binomcdf}(12, 0.85, 9) = 1 - 0.2642 = 0.7358$.

$P(X = 12) = 0.85^{12} \approx 0.1422$.

Interpretation: about 74% of the time, at least 10 of 12 seeds will sprout; all 12 sprout only about 14% of the time.

EXAMPLE 6: NORMAL APPROXIMATION AS AN EXTENSION

In a fictional poll model, each of 400 randomly selected voters supports a measure with probability 0.55. Find the probability that at least 240 support it, exactly and with a normal approximation.

SOLUTION

Exact: $P(X \geq 240) = 1 - \text{binomcdf}(400, 0.55, 239) \approx 0.0247$.

Conditions for approximation: $np = 220 \geq 10$ and $n(1 - p) = 180 \geq 10$. Mean $= 220$, SD $= \sqrt{400(0.55)(0.45)} = \sqrt{99} \approx 9.95$.

With continuity correction: $z = (239.5 - 220)/9.95 \approx 1.96$, so $P \approx 0.0250$, very close to the exact value. Without the correction, $z = 2.01$ and the estimate is 0.0222.

Because the exact binomial answer is available on the calculator, use the approximation mainly to understand shape and to check reasonableness.

Write It Like a 5

Prompt: A quality inspector randomly selects 15 phone screens from a very large shipment in which 6% of screens are flawed. Let X be the number of flawed screens in the sample. (a) Explain why X is binomial. (b) Find the probability that exactly 2 screens are flawed. (c) Find the probability that at least one screen is flawed. (d) Find the mean and SD of X.

Model Response

(a) Each screen is flawed or not (binary); screens are chosen randomly from a very large shipment, so the 15 trials are independent (15 is less than 10% of the shipment); the number of trials, 15, is fixed; and each screen has the same 0.06 probability of being flawed. So $X \sim B(15, 0.06)$. (b) $P(X = 2) = \binom{15}{2}(0.06)^2(0.94)^{13} \approx 0.1691$. (c) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.94^{15} \approx 1 - 0.3953 = 0.6047$. (d) $\mu = 15(0.06) = 0.9$ flawed screens and $\sigma = \sqrt{15(0.06)(0.94)} \approx 0.92$ screens.

What the grader looks for

1 point: all four BINS conditions addressed in context, including independence or the 10% condition. 1 point: parameters n = 15 and p = 0.06 identified. 1 point: correct setup and value for exactly 2. 1 point: complement used correctly for at least one, answer 0.605. 1 point: mean and SD with correct formulas and units.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Write the BINS checklist in the margin and tie each letter to a phrase in the problem.
2. Name the variable and its distribution: "Let X = ... Then X ~ B(n, p)."
3. Translate the wording into the calculator function, using the table of probability words.
4. Write the setup before the number, for example $P(X \geq 4) = 1 - \text{binomcdf}(10, 0.3, 3)$.
5. Finish with a sentence in context, and check that the answer is between 0 and 1.

- The pdf gives one value only; the cdf adds up from 0 through k. If you need 'at least,' use the complement.
- Write the full sentence 'Let X = the number of ... Then X ~ B(n, p)' on every free-response solution; it earns the setup point by itself and organizes the rest of your work.
- For the exam, you do not need to expand the combination by hand when the calculator can supply the probability, but you must show the setup: the distribution with its parameters and the event written in terms of X.
- A quick shape check: if np is small, the distribution is skewed right and the mean is near 0.
- Checking the 10% condition protects you when sampling without replacement.

Common traps

1. Using the binomial model with a changing probability or a variable number of trials. 2. Using binomcdf(n, p, k) for P(X $\geq$ k) instead of 1 - binomcdf(n, p, k-1). 3. Reporting $np(1-p)$ as the SD and forgetting the square root. 4. Defining success as the rare event in one part and the common event in another part of the same problem.

Mistake Clinic

Spot the error

A student models the number of rolls of a fair die needed to get the first six as binomial with p = 1/6. Where it fails: The number of trials is not fixed in advance: you keep rolling until the first six. BINS fails on the N. Correct solution: The count of sixes in a fixed number of rolls is binomial. The number of rolls until the first six is geometric (next lesson).

Spot the error

For X ~ B(10, 0.3), a student computes P(X $\geq$ 4) as binomcdf(10, 0.3, 4) = 0.8497. Where it fails: binomcdf(10, 0.3, 4) is P(X $\leq$ 4), which includes 4 and everything below. Correct solution: $P(X \geq 4) = 1 - \text{binomcdf}(10, 0.3, 3) = 1 - 0.6496 = 0.3504$.

Spot the error

For n = 10 and p = 0.3, a student writes the SD as np(1 - p) = 10(0.3)(0.7) = 2.1. Where it fails: 2.1 is the variance. The standard deviation requires a square root. Correct solution: $\sigma = \sqrt{2.1} \approx 1.45$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Manufacturers use binomial models for acceptance sampling: they inspect a fixed number of items from each shipment and reject the shipment if the number of defects is too high. The rule they pick is judged by binomial probabilities, which show how often a good shipment is wrongly rejected and a bad shipment is wrongly accepted. Poll designers use the same model to understand how many supporters to expect in a sample. Even a simple classroom activity works: if every student in a class of 30 guesses on a ten-question true-false quiz, the binomial distribution predicts how many students should score 8 or more by luck alone, which is a useful benchmark when judging whether a high score reflects real knowledge.

Challenge Corner

Parts are defective with probability 0.04, independently. What is the smallest number of parts n that must be inspected so the probability of finding at least one defective part is at least 0.95? Solution: $P(\text{at least one}) = 1 - 0.96^n \geq 0.95$, so $0.96^n \leq 0.05$ and $n \geq \ln(0.05)/\ln(0.96) \approx 73.4$. So $n = 74$ (check: $1 - 0.96^{74} \approx 0.9512$ while $1 - 0.96^{73} \approx 0.9492$).

Practice

PRACTICE **Binomial Distributions**

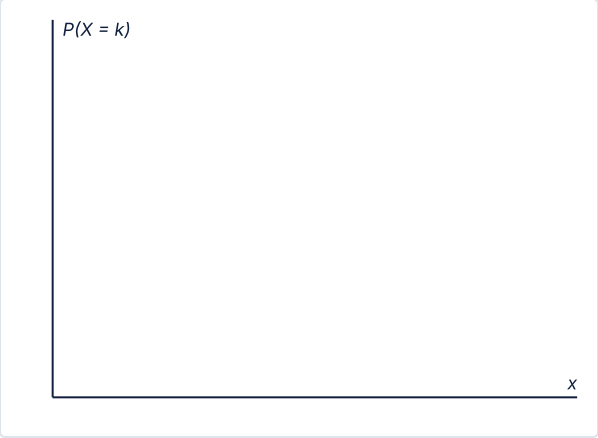
Show your work. A calculator is allowed. Justify conclusions in context.

1. Which situation can be modeled by a binomial distribution?
   - **(A)** The number of students out of 20 randomly chosen from a very large school who own a bike, if 40% of students do
   - **(B)** The number of cards drawn from a deck until the first king appears
   - **(C)** The heights of 20 randomly chosen students
   - **(D)** The number of free throws a player makes, where the player is more likely to score as warm-up progresses
2. For X ~ B(5, 0.4), what is P(X = 2)?
   (A) 0.0346 **(B)** 0.1600 **(C)** 0.3456 **(D)** 0.4000
3. For X ~ B(12, 0.5), which expression gives P(X ≥ 7)?
   (A) 1 - binomcdf(12, 0.5, 6) **(B)** binomcdf(12, 0.5, 7) **(C)** binomcdf(12, 0.5, 6)
   (D) 1 - binomcdf(12, 0.5, 7)
4. A random variable X has a binomial distribution with n = 50 and p = 0.2. What are the mean and SD?
   (A) Mean 40, SD 2.83 **(B)** Mean 10, SD 8 **(C)** Mean 10, SD 3.16
   (D) Mean 10, SD 2.83
5. Each of 30 independent pieces of equipment fails with probability 0.1. Which expression gives the probability that at most one fails?
   (A) binompdf(30, 0.1, 1) = 0.141 **(B)** binomcdf(30, 0.1, 1) = 0.184
   (C) 1 - binomcdf(30, 0.1, 1) = 0.816 **(D)** binomcdf(30, 0.1, 0) = 0.042
6. A player is shooting free throws, with probability of success rising each attempt as warm-up progresses. The count of made shots in 25 attempts is not binomial because
   - **(A)** there are more than two outcomes.
   - **(B)** the number of trials is not fixed.
   - **(C)** the probability of success is not the same on every trial.
   - **(D)** the outcomes cannot be counted.
7. X ~ B(100, 0.5) has mean 50 and SD 5. A sample gives X = 62. Which conclusion is best?
   (A) 62 is typical, about 1 SD above the mean. **(B)** 62 is unusual, about 2.4 SDs above the mean.
   (C) 62 is impossible for this distribution. **(D)** 62 is below the mean.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

8. The graph shows the distribution of X ~ B(10, 0.3). About what is $P(X \leq 2)$?



Binomial distribution with n = 10 and p = 0.3: skewed right, mean 3.

(A) 0.23 **(B)** 0.15 **(C)** 0.62 **(D)** 0.38

9. For which binomial distribution are both conditions for the normal approximation ($np \geq 10$ and $n(1-p) \geq 10$) met?

(A) n = 50, p = 0.10 **(B)** n = 20, p = 0.90 **(C)** n = 100, p = 0.30 **(D)** n = 200, p = 0.04

10. For each situation, decide whether the count is binomial. If yes, give n and p; if not, name the BINS condition that fails.

(a) The number of heads in 10 tosses of a coin that lands heads 60% of the time.

(b) The number of seniors in a random sample of 8 students from a school of 2,000 students, where 25% are seniors.

(c) The number of times you roll a die until a 4 appears.

(d) The number of red cards in 5 cards drawn without replacement from a standard 52-card deck.

11. A basketball player makes each free throw independently with probability 0.70. She takes 10 shots; X is the number made.

(a) Show X is binomial and name n and p.

(b) Find $P(X = 7)$.

(c) Find $P(X \leq 5)$.

(d) Find $P(X \geq 8)$.

(e) Find the mean and SD of X.

12. A quiz has 12 multiple-choice questions, each with 4 equally likely choices. A student guesses on every question. Let X be the number of correct answers.

(a) State the distribution of X.

(b) Find the mean and SD.

(c) Find $P(X = 3)$.

(d) Find the probability of getting at least 6 correct.

13. A machine makes a part that is defective with probability 0.04, independently. In a batch of 200 parts, what is the probability that 15 or more are defective? Interpret the probability. Would the batch be good evidence that the defect rate rose? Explain.

14. In a fictional poll model, each of 400 voters supports a measure with probability 0.55, independently. Let X be the number who support the measure.

(a) Check the conditions for the normal approximation.

(b) Use the normal approximation with continuity correction to estimate $P(X \geq 240)$.

(c) The exact value is 0.0247. Comment on the approximation.

15. Parts are defective with probability 0.04, independently. Find the smallest n such that the probability of at least one defective part in the sample is at least 0.95. Show the inequality.

5.6 Geometric Distributions

Model the number of trials until the first success, compute geometric probabilities, means and standard deviations, and contrast the geometric with the binomial model.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Beyond the 2027 Exam Core

College Board's revision page lists the geometric distribution as removed from the 2027 exam. This lesson is extension material that appears in college courses and older exams, and it deepens your understanding of the binomial model. Check your teacher's current course description to decide how much time to give it.

Why This Matters

Some questions do not count successes in a fixed number of tries. They ask how long you must wait: how many spins until a prize, how many parts inspected until the first defect, how many calls until someone answers. The geometric distribution models exactly that waiting time.

Geometric ideas build on the binomial lesson in Exam Unit 2 (Probability, Random Variables, and Probability Distributions). The common mistakes are confusing it with the binomial, misplacing the exponent, and misreading what the mean 1/p means.

Core Ideas

A random variable X is **geometric** if it counts the number of trials needed to get the first success, where the trials are Binary, Independent, and have the Same probability p of success. The number of trials is not fixed in advance, so BINS becomes "BIS" and "count until the first success."

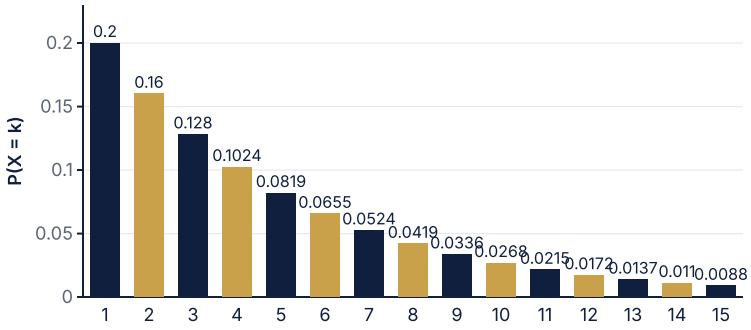
Geometric formulas

$P(X = k) = (1-p)^{k-1}p$ for $k = 1, 2, 3, \ldots$ (fail $k-1$ times, then succeed once).

$P(X > k) = (1-p)^k$ (the first k trials all fail). $P(X \le k) = 1 - (1-p)^k$.

Mean: $\mu_X = \dfrac{1}{p}$. Standard deviation: $\sigma_X = \dfrac{\sqrt{1-p}}{p}$.

Calculator: geometpdf(p, k) for $P(X = k)$ and geometcdf(p, k) for $P(X \le k)$.



Geometric distribution with p = 0.2 (k = 1 to 15): the most likely value is 1, the right tail is long, mean 5.

The graph has the signature geometric shape: the most likely single value is always 1 (an immediate success), and the probabilities then shrink by the constant factor $1-p$ each step, which creates the long right tail. Because of that tail, the mean $1/p = 5$ is larger than the most likely value.

Feature	Binomial	Geometric
Counts	Successes in n trials	Trials until the first success
Number of trials	Fixed in advance	Not fixed
Possible values	0, 1, ..., n	1, 2, 3, ... (no upper limit)
Mean	np	1/p
SD	sqrt(np(1-p))	sqrt(1-p)/p
Shape	Depends on p; bell-like for large n	Always skewed right, highest at 1

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the mean means

The mean $1/p$ is the long-run average number of trials needed to get the first success. For $p = 0.2$ the mean is 5: over many repeated sequences, it takes about 5 trials on average.

It does not say that the first success will happen exactly on the 5th trial, or that success is guaranteed by trial 5. The probability of at least one success in the first 5 trials is only $1 - 0.8^5 \approx 0.672$.

No memory: after m failures, the chance of needing more than j additional trials is still $(1-p)^j$, because the trials are independent. Past failures do not make a success \"due.\"

Worked Examples

EXAMPLE 1: IS IT GEOMETRIC?

Classify each as binomial, geometric, or neither: (a) the number of rolls of a fair die until the first six; (b) the number of sixes in 10 rolls; (c) the number of cards drawn without replacement from a deck until the first ace.

Feature	Binomial	Geometric
Counts	Successes in n trials	Trials until the first success
Number of trials	Fixed in advance	Not fixed
Possible values	0, 1, ..., n	1, 2, 3, ... (no upper limit)
Mean	np	1/p
SD	sqrt(np(1-p))	sqrt(1-p)/p
Shape	Depends on p; bell-like for large n	Always skewed right, highest at 1

SOLUTION

(a) Geometric: independent rolls, $p = 1/6$ each, counting trials until the first success.

(b) Binomial: a fixed number of trials, counting successes.

(c) Neither: sampling without replacement makes the probability of an ace change after each draw, so the trials are not independent and the probability is not constant.

EXAMPLE 2: PROBABILITY OF THE FIRST SUCCESS ON A SPECIFIC TRIAL

A fair die is rolled until a six appears. What is the probability that the first six occurs on the 4th roll?

SOLUTION

X is geometric with $p = 1/6$.

The first three rolls must fail and the fourth must succeed: $P(X = 4) = (5/6)^3(1/6) = 125/1296 \approx 0.0965$.

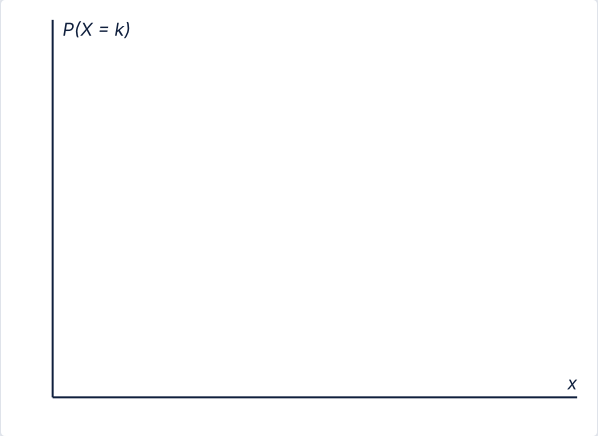
Calculator check: geometpdf(1/6, 4) = 0.0965.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: WAITING LONGER THAN K

In the same die setting, find the probability that more than 5 rolls are needed and the probability that a six appears within the first 5 rolls.



Geometric distribution with p = 0.2 (k = 1 to 15): the most likely value is 1, the right tail is long, mean 5.

SOLUTION

More than 5 rolls means the first five rolls all fail: $P(X > 5) = (5/6)^5 \approx 0.4019$.

A six within 5 rolls is the complement: $P(X \leq 5) = 1 - 0.4019 = 0.5981$.

Check: geometcdf(1/6, 5) = 0.5981. About 60% of the time you will see a six in five rolls.

EXAMPLE 4: MEAN AND SD

A student makes each basket attempt with probability 0.4, independently. Let X be the number of attempts until the first made basket. Find the mean and SD, and interpret the mean.

SOLUTION

$\mu_X = 1/0.4 = 2.5$ attempts.

$\sigma_X = \sqrt{0.6}/0.4 \approx 1.94$ attempts.

Interpretation: over many practice sessions, the student needs about 2.5 attempts on average to make the first basket, with typical variation of about 1.9 attempts. The student cannot literally take 2.5 shots in a particular session.

EXAMPLE 5: QUALITY CONTROL CONTEXT

Each part inspected is defective with probability 0.04, independently. Find the probability that the first defective part is found within the first 10 inspections, and the probability that at least 30 parts are inspected before a defect is found (that is, X $\geq$ 30).

SOLUTION

X is geometric with $p = 0.04$.

Within the first 10: $P(X \leq 10) = 1 - 0.96^{10} \approx 1 - 0.6648 = 0.3352$.

At least 30 inspections: the first 29 must all be good, $P(X \geq 30) = P(X > 29) = 0.96^{29} \approx 0.3061$.

The mean is $1/0.04 = 25$ parts, which is close to the center of those two probabilities: about a third of the time a defect appears in the first 10, and about 31% of the time it takes 30 or more.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 6: MEDIAN WAIT AND NO MEMORY

A prize appears on each spin with probability 0.2, independently. (a) What is the smallest number of spins k with at least a 50% chance of a prize within k spins? (b) A player has already had 5 spins with no prize. What is the chance that more than 3 additional spins are needed?

SOLUTION

(a) $P(X \le k) = 1 - 0.8^k \ge 0.5$, so $0.8^k \le 0.5$ and $k \ge \ln(0.5)/\ln(0.8) \approx 3.1$. Check: $P(X \le 3) = 0.488$ and $P(X \le 4) = 0.590$. So $k = 4$ spins.

(b) The spins are independent, so the past failures do not matter: $P(\text{more than 3 more}) = 0.8^3 = 0.512$.

The player is not \"due\" a prize; the chance is the same as for a brand new player.

Write It Like a 5

Prompt: A spinner at a school fair gives a prize with probability 0.10 on each spin, independently. Let X be the number of spins until the first prize. (a) Explain why X is geometric. (b) Find P(X = 5). (c) Find P(X > 10). (d) Find and interpret the mean.

Model Response

(a) Each spin is a success (prize) or failure; spins are independent; the probability of a prize is the same 0.10 on every spin; and we count trials until the first success. So X is geometric with $p = 0.10$. (b) $P(X = 5) = (0.90)^4(0.10) = 0.06561$. (c) More than 10 spins are needed when the first 10 all fail: $P(X > 10) = 0.90^{10} \approx 0.3487$. (d) $\mu_X = 1/0.10 = 10$. Over many players, the number of spins needed to get the first prize averages 10; any one player may need far fewer or far more.

What the grader looks for

1 point: conditions (binary, independent, constant p, counting until the first success) stated in context. 1 point: formula with the exponent $k - 1$ and the correct value 0.0656. 1 point: P(X > 10) recognized as ten failures, answer 0.349. 1 point: mean 1/p computed. 1 point: interpretation as a long-run average number of spins, not a guarantee.

Coach's Playbook

1. Ask: am I counting successes in a fixed number of trials (binomial) or counting trials until the first success (geometric)?
2. Define X and p in words and state the conditions.
3. Pick the form: exactly k uses $(1-p)^{k-1}p$; 'more than k' uses $(1-p)^k$; 'at most k' uses $1-(1-p)^k$.
4. Check with geometpdf and geometcdf on the calculator and keep four decimals.
5. Interpret the mean as a long-run average number of trials, not a prediction.

- The geometric distribution has no upper limit, so P(X ≤ k) is always less than 1.
- When a question asks about 'at least one success in n trials', the binomial complement and geometric cdf give the same number: $1-(1-p)^n$.
- For 'fewer than' and 'at least', convert to the equivalent 'at most' or 'more than' statement first.

Common traps

1. Using $(1-p)^k p$ instead of $(1-p)^{k-1}p$ for the first success on trial k. 2. Using 1 - (1 - p)^k for P(X > k) (it is P(X ≤ k)). 3. Interpreting the mean 1/p as the exact trial on which success will occur. 4. Believing a success is 'due' after a long run of failures.

Mistake Clinic

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student finds the probability that the first six occurs on the 4th roll as (5/6)^4(1/6). Where it fails: The exponent should count the failures before the success: three failures, not four. Correct solution: $P(X = 4) = (5/6)^3(1/6) \approx 0.0965$.

Spot the error

With p = 0.25, a student says the mean is 4 so 'you will need exactly 4 attempts to succeed.' Where it fails: The mean is a long-run average, not a prediction for one run. Many runs take 1, 2, or 10 or more attempts. Correct solution: The correct statement: over many repetitions, the number of attempts until the first success averages 4.

Spot the error

For p = 0.3, a student computes P(X $>$ 4) as 1 - 0.7^4 = 0.7599. Where it fails: 1 - 0.7^4 is P(X $\leq$ 4). More than 4 means the first four attempts all fail. Correct solution: $P(X > 4) = 0.7^4 = 0.2401$.

Real-World Connection

Geometric models describe waiting for rare events in a process with no memory: the number of calls to a help line until one is answered by a particular agent, the number of manufactured items until the first defect, the number of free throws until the first miss. In each case the mean 1/p gives planners a quick estimate of the typical wait, and the cdf tells them how likely a wait is to exceed a given limit.

Challenge Corner

Two players alternate rolling a fair die, Player 1 first. The first to roll a six wins. What is the probability Player 1 wins? Solution: Player 1 wins on roll 1, 3, 5, ... with probabilities $\frac{1}{6}$, $(\frac{5}{6})^2\frac{1}{6}$, $(\frac{5}{6})^4\frac{1}{6}$, ..., a geometric series with first term $\frac{1}{6}$ and ratio $\frac{25}{36}$. The sum is $\frac{1/6}{1 - 25/36} = \frac{1/6}{11/36} = \frac{6}{11} \approx 0.545$. Going first is an advantage.

Practice

PRACTICE **Geometric Distributions**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which random variable is geometric?
   - **(A)** The number of coin flips until the first head
   - **(B)** The number of heads in 10 coin flips
   - **(C)** The number of aces in a 5-card hand
   - **(D)** The heights of 10 students
2. A trial succeeds with probability 0.25, independently. What is the probability that the first success occurs on the 3rd trial?
   (A) 0.4219 **(B)** 0.1406 **(C)** 0.0625 **(D)** 0.2500
3. X is geometric with p = 0.2. What are the mean and standard deviation of X?
   (A) Mean 5, SD 20 **(B)** Mean 4, SD 4.47 **(C)** Mean 5, SD 4.47
   (D) Mean 0.2, SD 0.8
4. X is geometric with p = 0.3. What is P(X $>$ 4)?
   (A) 0.7599 **(B)** 0.0081 **(C)** 0.3000 **(D)** 0.2401
5. For a geometric random variable, geometcdf(p, k) computes which probability?
   (A) P(X = k) **(B)** P(X $>$ k) **(C)** P(X $\leq$ k) **(D)** P(X $\geq$ k)

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. X is geometric and has mean 4. Which interpretation is correct?
   (A) Over many repetitions, the number of trials needed averages 4.
   (B) The first success happens on the 4th trial.
   (C) The first success happens by the 4th trial for sure.
   (D) There are exactly 4 successes in the long run.
7. Each inspected item is defective with probability 0.1, independently. What is the probability that at least 10 items must be inspected to find the first defect?
   (A) 0.3487 **(B)** 0.3874 **(C)** 0.6126 **(D)** 0.1000
8. A player misses 5 shots in a row. Each shot is independent with probability of success 0.6. What is the probability the next shot is made?
   (A) Greater than 0.6, because a make is due. **(B)** Less than 0.6, because the player is cold.
   (C) 0.4 **(D)** 0.6
9. Cards are drawn without replacement from a standard deck until the first heart appears. Why is the number of cards not geometric?
   (A) The number of trials is fixed.
   (B) There are more than two outcomes.
   (C) The trials are not independent and the probability changes.
   (D) There is no first success.
10. For each, decide whether X is binomial, geometric, or neither, and explain.
    (a) X = number of flips of a fair coin until the first head.
    (b) X = number of heads in 8 flips of a fair coin.
    (c) X = number of cards drawn from a deck, without replacement, until the first ace.
    (d) X = number of customers until the first complaint, if each customer complains independently with probability 0.03.
11. A prize appears on each spin with probability 0.15, independently. X is the number of spins until the first prize.
    (a) Find $P(X = 3)$.
    (b) Find $P(X \leq 4)$.
    (c) Find $P(X > 6)$.
    (d) Find the mean and SD.
12. Parts are defective with probability 0.04, independently, and are inspected one at a time until the first defect.
    (a) Find the probability that the 10th inspected part is the first defective.
    (b) Find the probability that at least 25 parts are inspected before a defective one is found ($X \geq 25$).
    (c) Find the mean number inspected, and interpret it.
13. A student guesses on each of several 4-choice questions in turn. X is the number of questions until the first correct guess.
    (a) Find the distribution parameter p.
    (b) Find $P(X \leq 3)$.
    (c) Find the mean and SD of X.
14. Each trial succeeds with probability 0.3, independently.
    (a) With n = 5 trials, find the probability of exactly one success.
    (b) Find the probability that the first success occurs on trial 5.
    (c) Explain why these two probabilities differ.
15. Each attempt at a challenge succeeds with probability 0.10, independently.
    (a) Find the smallest number of attempts k such that the probability of at least one success within k attempts is at least 0.90.
    (b) Show that your k works and that k - 1 does not.

5.7 Continuous Random Variables and Uniform Models

Interpret probability as area under a density curve, work with uniform and normal models, find means and standard deviations, and explain why single values have probability zero.

Why This Matters

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Some quantities, such as a wait time, a height, or a measurement error, can take any value in an interval. We cannot list their values or give each one a probability. Instead we use a **density curve**: the area under the curve over an interval is the probability of landing there. This single change of thinking from heights to areas is what makes the normal distribution and every later inference procedure work.

The idea belongs to Exam Unit 2 (Probability, Random Variables, and Probability Distributions) and links to the normal model from Unit 1. Typical exam moves are finding a probability as an area, explaining why P(X = x) is 0, and using the uniform model for a mean, SD, or probability.

Core Ideas

A **continuous random variable** takes all values in an interval of numbers, such as time, length, or weight. Its distribution is described by a **density curve** that lies on or above the horizontal axis with total area exactly 1.

Probability as area

$P(a \le X \le b)$ is the area under the density curve between a and b.

For a continuous variable, $P(X = c) = 0$ for every single value c, because a single point has no width and therefore no area.

Consequently $P(X < c) = P(X \le c)$ and $P(a < X < b) = P(a \le X \le b)$. Strict versus inclusive inequalities do not matter (unlike discrete variables).

The height of a density curve is **not** a probability; it can even exceed 1 when the interval is short.

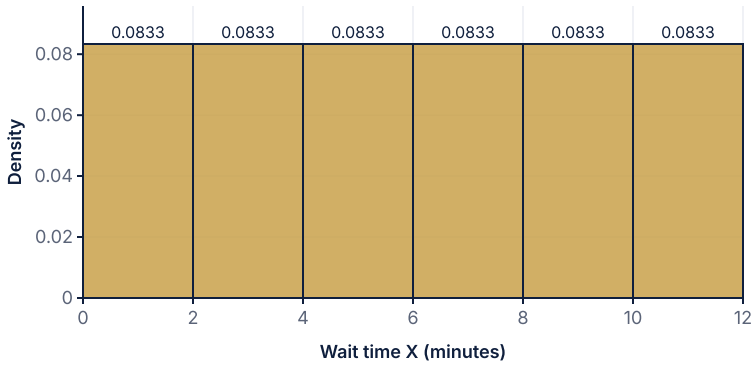
The uniform distribution on [a, b]

Density: $f(x) = \dfrac{1}{b-a}$ for $a \le x \le b$ (a rectangle of height $\frac{1}{b-a}$ and width $b - a$, so area 1).

$P(c \le X \le d) = \dfrac{d-c}{b-a}$ for $a \le c \le d \le b$: a proportion of the interval's length.

Mean (also the median): $\mu = \dfrac{a+b}{2}$. Standard deviation: $\sigma = \dfrac{b-a}{\sqrt{12}}$.

The uniform model says \"equally likely anywhere in the interval,\" which is how random number generators and some rounding errors behave.



Uniform density on [0, 12]: height 1/12, total area 12 × 1/12 = 1. The area of any slice is the probability.

Feature	Discrete random variable	Continuous random variable
Possible values	A countable list, such as 0, 1, 2, 3	Every value in an interval
How probability is described	A table or formula giving P(X = x)	A density curve; probability is area
P(X = x)	Can be positive	Always 0
P(X < c) versus P(X ≤ c)	Can differ by P(X = c)	Always equal
Mean and SD	Weighted sums of values	Balance point and spread of the curve

From histogram to density curve. Imagine measuring wait times for thousands of shuttle rides and drawing a relative frequency histogram with narrower and narrower bars. The bar tops smooth into a curve, and the area of the bars over any interval (the proportion of rides in that interval) approaches the area under the curve. The density curve is that idealized limit: it describes the whole population, not a sample. This is why proportions, percentiles, and probabilities are all areas, and why the symbols μ and σ describe the curve rather than a data set.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Percentiles and medians. The pth percentile of a continuous distribution is the value with area p to its left. The median is the 50th percentile, splitting the area into two halves. For a uniform distribution, the median equals the mean at the midpoint; for a skewed density, the mean is pulled toward the tail, so it differs from the median.

Because a density curve is just a shape with area 1, other shapes work the same way. A **skewed** density curve puts its mean on the longer-tail side of the median, just like a skewed histogram. The normal distribution is the most important density curve: probabilities are areas that we find from z-scores or from the calculator's normal functions.

Conditional probability for continuous variables works as before: restrict to the part of the interval that is given and take the new area as a proportion of the new whole. For a uniform distribution this means a ratio of lengths: $P(X > 9 \mid X > 6)$ on $[0, 12]$ is the length from 9 to 12 divided by the length from 6 to 12.

Worked Examples

EXAMPLE 1: UNIFORM PROBABILITIES AS AREAS

A student's wait X for a shuttle is uniformly distributed between 0 and 12 minutes. Find the height of the density, P(X $<$ 4), P(3 $<$ X $<$ 9), and P(X = 5).

Uniform density on [0, 12]: height 1/12, total area 12 × 1/12 = 1. The area of any slice is the probability.

SOLUTION

The width is 12, so the height is $1/12 \approx 0.0833$.

$P(X < 4) = 4(1/12) = 1/3 \approx 0.333$.

$P(3 < X < 9) = 6/12 = 0.5$.

$P(X = 5) = 0$. A single value has no width, so no area. The shuttle arriving at exactly 5.000... minutes has probability zero, even though it is possible.

EXAMPLE 2: MEAN, SD, AND A TAIL PROBABILITY

A production step takes a time uniformly distributed between 8 and 20 minutes. Find the mean, SD, and P(X $>$ 17). Interpret the mean.

SOLUTION

$\mu = (8 + 20)/2 = 14$ minutes.

$\sigma = (20 - 8)/\sqrt{12} = 12/\sqrt{12} \approx 3.46$ minutes.

$P(X > 17) = (20 - 17)/(20 - 8) = 3/12 = 0.25$.

Interpretation: across many runs of the step, the time averages 14 minutes, and a quarter of runs take longer than 17 minutes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: A NON-UNIFORM DENSITY

The lifetime X (in years) of a component has density f(x) = 0.02x for $0 \le x \le 10$ and 0 otherwise. Show f is a valid density, then find P(X < 5) and P(X > 8).

Bar approximation of the density f(x) = 0.02x on [0, 10]; the true curve is a straight line from 0 to 0.2.

SOLUTION

The graph is a right triangle with base 10 and height $0.02(10) = 0.2$. Area $= \frac{1}{2}(10)(0.2) = 1$, so f is valid.

$P(X < 5)$: a triangle with base 5 and height 0.1, so the area is $\frac{1}{2}(5)(0.1) = 0.25$.

$P(X > 8) = 1 - P(X < 8)$, where the triangle to 8 has height 0.16 and area $\frac{1}{2}(8)(0.16) = 0.64$. So $P(X > 8) = 0.36$.

Most of the probability sits at the high end, so the lifetime distribution is skewed left.

EXAMPLE 4: CONDITIONAL PROBABILITY WITH A UNIFORM VARIABLE

The shuttle wait X is uniform on [0, 12]. Given that you have already waited 6 minutes, find the probability the shuttle arrives after the 9th minute.

SOLUTION

We want $P(X > 9 \mid X > 6) = \dfrac{P(X > 9)}{P(X > 6)}$, because X > 9 is part of X > 6.

$P(X > 9) = 3/12 = 0.25$ and $P(X > 6) = 6/12 = 0.50$.

$P(X > 9 \mid X > 6) = 0.25/0.50 = 0.50$.

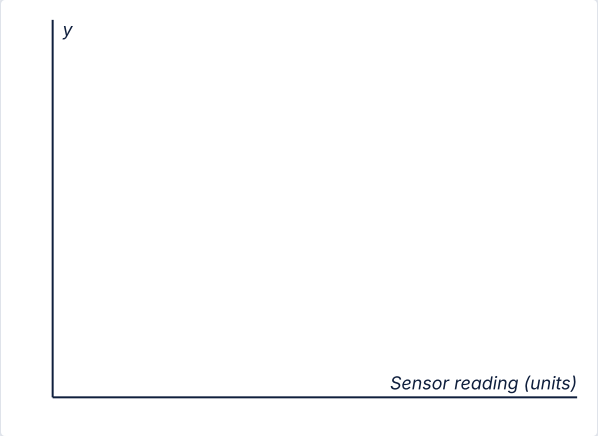
Equivalently, the remaining wait is uniform on the remaining 6 minutes, and 3 of those 6 are after minute 9.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: NORMAL PROBABILITIES AS AREAS

Sensor readings are modeled by a normal distribution with mean 50 and SD 4 units. Find P(X $>$ 56), the middle probability P(46 $<$ X $<$ 54), and the 90th percentile.



Normal model with mean 50 and SD 4; the shaded area to the right of 56 is 0.0668.

SOLUTION

$z = (56 - 50)/4 = 1.5$, so $P(X > 56) = 1 - 0.9332 = 0.0668$.

46 and 54 are 1 SD below and above the mean, so $P(46 < X < 54) \approx 0.6827$.

The 90th percentile: $z = 1.2816$, so $x = 50 + 1.2816(4) \approx 55.13$ units.

For any normal model, $P(X = 50) = 0$ exactly, so we always ask about intervals.

EXAMPLE 6: TWO INDEPENDENT UNIFORM VALUES

Two independent random-number generators each produce a value uniformly between 0 and 1, called X and Y. Find P(both are less than 0.3) and P(at least one is greater than 0.8).

SOLUTION

$P(X < 0.3) = 0.3$ and $P(Y < 0.3) = 0.3$. For independent variables, $P(\text{both}) = 0.3(0.3) = 0.09$.

\"At least one above 0.8\" is the complement of \"both at most 0.8\": $P(X \le 0.8) = 0.8$.

$P(\text{both} \le 0.8) = 0.64$, so $P(\text{at least one} > 0.8) = 1 - 0.64 = 0.36$.

The probability rules from the first lesson (independence, complement) carry over unchanged to continuous variables.

Write It Like a 5

Prompt: A tram arrives at a stop at a time that is uniformly distributed between 0 and 15 minutes after you reach the platform. Let X be your wait in minutes. (a) Describe the density curve. (b) Find P(X $<$ 5). (c) Find P(X $>$ 12). (d) Find the mean and SD of X. (e) A friend says P(X = 7) = 1/15. Respond.

Model Response

(a) X is uniform on $[0, 15]$, so the density curve is a horizontal line of height $1/15$ over that interval and 0 elsewhere; the total area is $15 \cdot \frac{1}{15} = 1$. (b) $P(X < 5) = 5/15 = 1/3 \approx 0.333$. (c) $P(X > 12) = 3/15 = 0.20$. (d) $\mu = (0 + 15)/2 = 7.5$ minutes and $\sigma = 15/\sqrt{12} \approx 4.33$ minutes. Over many days the wait averages 7.5 minutes. (e) The friend confused the height of the density, $1/15$, with a probability. For a continuous variable, a single value has no width, so no area, and $P(X = 7) = 0$. Probabilities come only from areas over intervals, such as $P(6.5 < X < 7.5) = 1/15$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

1 point: density described with height $1/15$ and area 1. 1 point: correct P(X $<$ 5) as a ratio of lengths. 1 point: correct P(X $>$ 12). 1 point: mean and SD with correct formulas and the mean interpreted in context. 1 point: explanation that the density height is not a probability and P(X = 7) = 0.

Coach's Playbook

1. Sketch the density curve and shade the region described by the question.
2. For a uniform variable, compute the shaded length divided by the total length. For other shapes, use geometry or the calculator's normal function.
3. Convert strict and inclusive inequalities freely because P(X = c) = 0.
4. For conditional questions, rewrite the given condition as the new, smaller whole.
5. Write the answer as a probability and a sentence in context.

- Check that the total area is 1 before trusting a density function.
- For a normal model, the calculator's normalcdf(lower, upper, mean, SD) returns the area; use a large number such as 1000 for infinity.
- A probability found from a density can be checked by symmetry or by the complement.

Common traps

1. Treating the density's height as a probability. 2. Claiming P(X = c) is small but positive for a continuous variable. 3. Adding or subtracting a point at an endpoint (an inclusive inequality does not change a continuous probability). 4. Using $(b-a)/2$ for the SD of a uniform variable; the correct formula is $(b-a)/\sqrt{12}$.

Mistake Clinic

Spot the error

A uniform density on [0, 12] has height 1/12. A student says P(X = 5) = 1/12. Where it fails: The height of a density curve is not a probability. A single point has no area. Correct solution: $P(X = 5) = 0$. For an interval, $P(4.5 < X < 5.5) = 1/12$.

Spot the error

For a continuous X, a student computes P(X $\leq$ 4) = 1/3 but writes P(X $<$ 4) = 1/3 - 0.001 'to remove the endpoint.' Where it fails: Removing a single point removes no area for a continuous variable. Correct solution: $P(X < 4) = P(X \leq 4) = 1/3$.

Spot the error

The student finds the SD of the uniform distribution on [8, 20] as (20 - 8)/2 = 6. Where it fails: (b - a)/2 is half the range, the largest possible deviation from the mean, not the SD. Correct solution: $\sigma = (20-8)/\sqrt{12} \approx 3.46$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Random number generators in video games and simulations draw values uniformly between 0 and 1, then convert them to other models. Rounding errors in measurements are often modeled as uniform on an interval of width equal to the rounding unit. Normal and other density models describe heights, test scores, and manufacturing dimensions. In each case, a probability is an area, and the area under the whole curve is 1. Understanding this clearly pays off later: when you read a statement such as "the p-value is the area beyond the test statistic," you will already know what an area under a curve means and why a single value never has a probability of its own.

Challenge Corner

For the lifetime density $f(x) = 0.02x$ on $[0, 10]$, find the median and compare it with the mean $\mu = \int_0^{10} x(0.02x)\,dx = 20/3 \approx 6.67$. Solution: the area from 0 to m is $\frac{1}{2}(m)(0.02m) = 0.01m^2$. Setting it equal to 0.5 gives $m = \sqrt{50} \approx 7.07$. The median (7.07) exceeds the mean (6.67), as expected for a left-skewed density.

Practice

PRACTICE **Continuous Random Variables and Uniform Models**

Show your work. A calculator is allowed. Justify conclusions in context.

1. X is a continuous random variable. What is P(X = 3)?
   (A) 0 **(B)** The height of the density at 3
   (C) 1/3 **(D)** It depends on the width of the interval.
2. X is uniform on [2, 10]. What is P(X < 5)?
   (A) 0.500 **(B)** 0.300 **(C)** 0.625 **(D)** 0.375
3. X is uniform on [0, 20]. What is the height of the density curve?
   (A) 20 **(B)** 1 **(C)** 0.05 **(D)** 0.50
4. X is uniform on [0, 12]. What are the mean and SD of X?
   (A) Mean 6, SD 3.46 **(B)** Mean 6, SD 6 **(C)** Mean 12, SD 3.46
   (D) Mean 6, SD 1
5. Which function is a valid density curve on [0, 4]?
   (A) f(x) = 0.5 **(B)** f(x) = 0.25 **(C)** f(x) = 0.5 - 0.25x
   (D) f(x) = 2
6. A normal model for sensor readings has mean 50 and SD 4. What is P(X > 56)?
   (A) 0.9332 **(B)** 0.1587 **(C)** 0.0228 **(D)** 0.0668
7. Which variable is continuous?
   (A) The number of siblings of a student **(B)** The number of cars in a lot
   (C) The time a student spends on homework **(D)** The number of text messages sent
8. X is uniform on [0, 12]. What is P(X > 9 | X > 6)?
   (A) 0.25 **(B)** 0.33 **(C)** 0.75 **(D)** 0.50
9. For a continuous random variable X, which expression equals P(2 < X < 5)?
   (A) P(X < 5) - P(X < 2) **(B)** P(X < 5) + P(X > 2) **(C)** P(X < 5) * P(X > 2)
   (D) P(X > 5) - P(X > 2)
10. A commuter's trip time X is uniform on [20, 40] minutes.
    (a) State the height of the density curve.
    (b) Find P(X < 25) and P(28 < X < 35).
    (c) Find the mean, median, and SD.
    (d) Find the 90th percentile.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

11. A component's lifetime X in years has density $f(x) = 0.02x$ for $0 \le x \le 10$.
    (a) Show that f is a valid density.
    (b) Find $P(X < 5)$ and $P(X > 8)$.
    (c) Find the median, and compare it with the mean 6.67. What does the comparison say about the shape?
12. Sensor readings are normal with mean 50 and SD 4.
    (a) Find $P(X > 56)$.
    (b) Find $P(46 < X < 54)$.
    (c) Find the 90th percentile.
    (d) Explain why $P(X = 50) = 0$ even though 50 is the most likely region.
13. Two independent random numbers X and Y are each uniform on [0, 1].
    (a) Find P(both are less than 0.3).
    (b) Find P(at least one is greater than 0.8).
    (c) Find $P(X < 0.5$ and $Y > 0.5)$.
14. A package is delivered at a time uniformly distributed between 2:00 and 2:30 pm. Let X be the number of minutes after 2:00.
    (a) Find the mean arrival time.
    (b) Given that it has not arrived by 2:10, find the probability it arrives by 2:20.
    (c) Find the probability the package arrives within 5 minutes of 2:15.
15. A random number generator gives a value X uniform on [0, 1].
    (a) Explain why $P(X = 0.5) = 0$ even though 0.5 can occur.
    (b) Find $P(0.49 < X < 0.51)$.
    (c) Explain why the height of the density, 1, does not mean the probability of a value is 1.

T Unit 5 Test

Test your command of probability rules, conditional probability, random variables, binomial and geometric models, and continuous distributions.

Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

UNIT TEST

Unit 5 Test

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. For events A and B, $P(A) = 0.45$, $P(B) = 0.30$, and $P(A \text{ or } B) = 0.60$. Which statement is correct?
   (A) A and B are mutually exclusive.
   (B) A and B are independent.
   (C) A and B are neither mutually exclusive nor independent.
   (D) A and B are both mutually exclusive and independent.
2. Each of 4 independent alarms sounds during a test with probability 0.30. What is the probability that at least one alarm sounds?
   (A) 0.7599 **(B)** 0.2401 **(C)** 0.0081 **(D)** 1.2000
3. The table shows 200 students and their activities. What is P(Choir | Grades 11-12)?

	Grades 9-10	Grades 11-12	Total
Band	40	20	60
Choir	25	35	60
Neither	35	45	80
Total	100	100	200

   (A) 0.175 **(B)** 0.35 **(C)** 0.583 **(D)** 0.30
4. 40% of messages are spam. A filter marks 90% of spam messages and 5% of non-spam messages. A message is marked. What is the probability it is spam?
   (A) 0.90 **(B)** 0.36 **(C)** 0.39 **(D)** 0.92
5. A random variable X takes values 0, 1, 2, 3 with probabilities 0.1, 0.3, 0.4, 0.2. What is the mean of X?
   (A) 1.5 **(B)** 1.0 **(C)** 1.7 **(D)** 0.4

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. A fair-ground game costs 3 tickets. A player receives 0 tickets with probability 0.5, 4 tickets with probability 0.3, and 10 tickets with probability 0.2. What is the expected net gain per play?
   (A) 0.2 tickets **(B)** 3.2 tickets **(C)** -0.2 tickets **(D)** 0 tickets
7. X and Y are independent random variables with SD(X) = 6 and SD(Y) = 8. What is the SD of X - Y?
   (A) 2 **(B)** 10 **(C)** 14 **(D)** 100
8. X ~ B(10, 0.2). What is P(X $\geq$ 2)?
   (A) 0.3758 **(B)** 0.3020 **(C)** 0.6778 **(D)** 0.6242
9. X is binomial with n = 80 and p = 0.25. What are the mean and SD of X?
   (A) Mean 20, SD 15 **(B)** Mean 20, SD 5 **(C)** Mean 20, SD 3.87 **(D)** Mean 60, SD 3.87
10. Each trial succeeds with probability 0.2, independently. What is the probability that the first success occurs on the 4th trial?
    (A) 0.1024 **(B)** 0.4096 **(C)** 0.0064 **(D)** 0.5904
11. A time X is uniform on [10, 30]. What is P(14 $<$ X $<$ 22)?
    (A) 0.30 **(B)** 0.40 **(C)** 0.20 **(D)** 0.50
12. Heights in a population are normally distributed with mean 66 and SD 3 inches. What proportion of heights is below 69 inches?
    (A) 0.1587 **(B)** 0.6827 **(C)** 0.9772 **(D)** 0.8413
13. [6 points] At a school clinic, 30% of students get a flu shot. Among students who got the shot, 4% catch the flu this season; among students who did not, 12% catch the flu. A student is chosen at random.
    (a) [2 points] Draw a tree diagram or otherwise show how to find the probability the student catches the flu, and compute it.
    (b) [2 points] Given that the student caught the flu, find the probability the student got the shot.
    (c) [2 points] Are getting the shot and catching the flu independent? Justify with probabilities.
14. [10 points] At a school fundraiser, each spin of a wheel awards X points with the distribution below.
    x = 0 (probability 0.50), x = 10 (0.30), x = 20 (0.15), x = 50 (0.05).
    (a) [2 points] Verify that this is a valid probability distribution and find the mean of X. Interpret the mean in context.
    (b) [2 points] Find the standard deviation of X.
    (c) [2 points] A spin costs 8 points. Find the expected net gain per spin and say whether the game favors the player or the organizers.
    (d) [2 points] A student plays 25 independent spins. Find the mean and standard deviation of the student's total points.
    (e) [2 points] In 10 independent spins, let Y be the number of spins that award 50 points. State the distribution of Y and find P(Y = 2).

Answers

Probability and Random Variables: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

5.1 Probability Rules and Sample Spaces

PRACTICE · PROBABILITY RULES AND SAMPLE SPACES

1. B. Each flip has 2 results, so three flips give $2^3 = 8$ outcomes. The tempting 6 comes from adding $2 + 2 + 2$ instead of multiplying.
2. C. The complement is no heads (all tails), which has probability $(1/2)^4 = 1/16$, so the answer is $1 - 1/16 = 15/16$. Choice A is only the complement.
3. D. Rearranging the addition rule: $P(A \text{ and } B) = 0.60 + 0.30 - 0.70 = 0.20$. Choice C, 0.18, is the product, which would be correct only if the events were independent, and nothing says they are.
4. C. Mutually exclusive events have no overlap, so the probabilities simply add: $0.50 + 0.30 = 0.80$. Choice A multiplies, which is the independence rule.
5. A. The overlap Basic and Under 30 is 25. $P = (60 + 40 - 25)/100 = 0.75$. Choice B adds 60 and 40 without subtracting the overlap, double counting 25 students.
6. A. $P(A)P(B) = 0.40(0.25) = 0.10 = P(A \text{ and } B)$, so the events are independent. Since $P(A \text{ and } B) = 0.10 \neq 0$, they are not mutually exclusive.
7. C. Sum 7 has 6 outcomes and doubles have 6 outcomes, and no doubles sum to 7 (doubles sums are even), so there is no overlap: $12/36 = 1/3$. Choice B wrongly subtracts an overlap of one.
8. B. $P(\text{none fail}) = 0.9^3 = 0.729$, so $P(\text{at least one}) = 1 - 0.729 = 0.271$. Choice A, 0.300, adds the three probabilities and double counts.
9. D. Independent, so $P(A \text{ and } B) = 0.30(0.50) = 0.15$ and $P(A \text{ or } B) = 0.30 + 0.50 - 0.15 = 0.65$. Choice A forgets to subtract the overlap.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. (a) Ordered pairs (president, VP): Ana-Ben, Ana-Cy, Ana-Dee, Ben-Ana, Ben-Cy, Ben-Dee, Cy-Ana, Cy-Ben, Cy-Dee, Dee-Ana, Dee-Ben, Dee-Cy, which is 12 outcomes.
    (b) Ana appears in 6 of the 12 pairs (3 as president, 3 as VP), so $P = 6/12 = 0.50$.
    (c) Ben is president in 3 outcomes: $P = 3/12 = 0.25$.
11. (a) Both = 14. Only Spanish = 52 - 14 = 38. Only French = 38 - 14 = 24. Neither = 100 - (38 + 14 + 24) = 24.
    (b) $P(\text{Spanish or French}) = (38 + 14 + 24)/100 = 0.76$. Check: $0.52 + 0.38 - 0.14 = 0.76$.
    (c) Exactly one language: $(38 + 24)/100 = 0.62$.
    (d) Neither: $24/100 = 0.24$, which equals $1 - 0.76$.
12. (a) $P(\text{Own and Dogs}) = 100/300 = 1/3 \approx 0.333$.
    (b) $P(\text{Rent or Cats}) = P(\text{Rent}) + P(\text{Cats}) - P(\text{Rent and Cats}) = 100/300 + 90/300 - 30/300 = 160/300 \approx 0.533$.
    (c) $P(\text{Own}) = 200/300 = 2/3$ and $P(\text{Dogs}) = 150/300 = 1/2$, so the product is $1/3$. This equals $P(\text{Own and Dogs}) = 1/3$, so owning and preferring dogs are independent.
13. (a) $0.8^5 \approx 0.3277$.
    (b) Complement: $1 - 0.3277 = 0.6723$.
    (c) Rain on day 1 and dry on days 2 to 5: $0.2(0.8)^4 = 0.2(0.4096) = 0.0819$.
    Independence of days is what allows multiplying here.
14. (a) Mutually exclusive: $P(A \text{ or } B) = 0.30 + 0.40 = 0.70$.
    (b) Independent: $P(A \text{ and } B) = 0.30(0.40) = 0.12$, and $P(A \text{ or } B) = 0.30 + 0.40 - 0.12 = 0.58$.
    (c) Mutually exclusive requires $P(A \text{ and } B) = 0$, but independence requires 0.12. Both cannot hold unless one event has probability 0.
15. (a) Winning outcomes: (2,H), (2,T), (4,H), (4,T), (6,H), (6,T), (1,H), (3,H), (5,H): 9 outcomes.
    (b) $P(\text{win}) = 9/12 = 0.75$.
    (c) Losing requires an odd die and tails: (1,T), (3,T), (5,T), so $P(\text{lose}) = 3/12 = 0.25$ and $P(\text{win}) = 1 - 0.25 = 0.75$, which agrees.

5.2 Conditional Probability, Independence, and Tree Diagrams

PRACTICE · CONDITIONAL PROBABILITY, INDEPENDENCE, AND TREE DIAGRAMS

1. A. $P(A \mid B) = P(A \text{ and } B)/P(B) = 0.12/0.30 = 0.40$. Choice A multiplies the two numbers instead of dividing.
2. C. Given Chips, the whole is the 110 chip choosers, and 60 of them are lower-grade, so the answer is 60/110. Choice A conditions on Lower grades instead, reversing the roles.
3. B. For independent events, knowing A does not change B, so $P(B \mid A) = P(B) = 0.40$. Choice A, 0.20, is $P(A \text{ and } B)$.
4. B. $P = (4/10)(3/9) = 12/90 = 2/15 \approx 0.133$. Choice A, 0.160, treats the draws as independent with replacement, using 0.4×0.4.
5. C. $P(B) = 0.30(0.80) + 0.70(0.20) = 0.24 + 0.14 = 0.38$. Choice A is only the path through A; it omits the path through not A.
6. D. $P(A \mid B) = 0.24/0.38 \approx 0.632$. Choice A, 0.80, is $P(B \mid A)$, the reverse conditional.
7. B. $P(A \mid B) = 0.10/0.50 = 0.20$, which is less than $P(A) = 0.40$, so B makes A less likely and the events are not independent. Choice C divides by the wrong probability.
8. D. Path heavy and late: $0.25(0.40) = 0.10$. Path not heavy and late: $0.75(0.05) = 0.0375$. $P(\text{heavy} \mid \text{late}) = 0.10/0.1375 \approx 0.73$. Choice A is $P(\text{late} \mid \text{heavy})$, the reverse.
9. C. The complement of at least one blue is both red: $5/8 \cdot 4/7 = 5/14$. So $P = 1 - 5/14 = 9/14$. Choice B is only P(both red).
10. (a) $P(8+ \text{ hours} \mid \text{athlete}) = 98/140 = 0.70$.
    (b) 228 students sleep 8+ hours, so $P(\text{athlete} \mid 8+ \text{ hours}) = 98/228 \approx 0.430$.
    (c) $P(8+ \text{ hours}) = 228/400 = 0.57$, which is not equal to 0.70, so the events are not independent.
    (d) In context, athletes in this table are more likely (70%) than students overall (57%) to sleep at least 8 hours.
11. (a) $P(\text{both win}) = \dfrac{4}{12} \cdot \dfrac{3}{11} = \dfrac{12}{132} = \dfrac{1}{11} \approx 0.0909$.
    (b) Win then lose plus lose then win: $2\left(\dfrac{4}{12} \cdot \dfrac{8}{11}\right) = \dfrac{64}{132} \approx 0.485$.
    (c) $\dfrac{8}{12} \cdot \dfrac{7}{11} = \dfrac{56}{132} \approx 0.424$.
    (d) $(12 + 64 + 56)/132 = 1$.
12. (a) First branches: Bus 0.55, Not bus 0.45. From Bus: late 0.10, not late 0.90. From Not bus: late 0.04, not late 0.96.
    (b) $P(\text{late}) = 0.55(0.10) + 0.45(0.04) = 0.055 + 0.018 = 0.073$.
    (c) $P(\text{bus} \mid \text{late}) = 0.055/0.073 \approx 0.753$.
    In context, about 75% of late students ride the bus.
13. (a) 200 have the condition and 4,800 do not. Positive: $0.85(200) = 170$ with it; false positives $0.10(4800) = 480$. Negative: 30 with it, 4,320 without. Totals: 650 positive, 4,350 negative.
    (b) $P(\text{condition} \mid \text{positive}) = 170/650 \approx 0.262$.
    (c) $P(\text{condition} \mid \text{negative}) = 30/4350 \approx 0.0069$.
    (d) The condition is rare, so the 480 false positives from the large healthy group outnumber the 170 true positives.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. (a) $P(A \text{ and } B) = 0.60 + 0.50 - 0.80 = 0.30$. $P(A \mid B) = 0.30/0.50 = 0.60 = P(A)$, so A and B are independent (also $0.60 \times 0.50 = 0.30$).
    (b) Now $P(A \text{ and } B) = 0.20$ and $P(A \mid B) = 0.20/0.50 = 0.40 \neq 0.60$, so they are not independent; B occurring makes A less likely.
15. (a) $P(\text{red}) = 0.5(3/5) + 0.5(1/5) = 0.30 + 0.10 = 0.40$.
    (b) $P(\text{Bag 1} \mid \text{red}) = 0.30/0.40 = 0.75$.
    (c) $P(\text{blue}) = 0.60$ and the Bag 2 blue path is $0.5(4/5) = 0.40$, so $P(\text{Bag 2} \mid \text{blue}) = 0.40/0.60 \approx 0.667$.

5.3 Discrete Random Variables: Mean and Standard Deviation

PRACTICE · DISCRETE RANDOM VARIABLES: MEAN AND STANDARD DEVIATION

1. B. Probabilities must be between 0 and 1 and sum to 1: 0.5 + 0.3 + 0.2 = 1. Choice A sums to 1.1, C contains a negative probability, and D sums to 0.6.
2. A. The probabilities must total 1, so $p = 1 - (0.2 + 0.3 + 0.1) = 0.4$. Choice B, 0.6, is the sum of the known probabilities rather than the remainder.
3. C. $\mu = 0(0.5) + 1(0.3) + 2(0.2) = 0.7$. Choice A averages the values 0, 1, 2 without weighting.
4. D. $E(X^2) = 1(0.3) + 4(0.2) = 1.1$ and $\sigma^2 = 1.1 - 0.49 = 0.61$, so $\sigma = \sqrt{0.61} \approx 0.781$. Choice A, 0.610, is the variance.
5. C. A mean is a long-run average over many repetitions. Choice A treats it as the result of one hour, and a count cannot be 2.3.
6. A. The expected winnings are $10(0.20) = 2$, and the cost is 3, so expected net gain is $2 - 3 = -1$ dollar. Choice B ignores the cost of playing.
7. D. $P(X \geq 3) = P(3) + P(4) = 0.2 + 0.1 = 0.3$. Choice A, 0.1, is $P(X > 3)$, and 0.9 is $P(X \leq 3)$.
8. C. The number of heads is a count with a countable list of values. Height, time, and weight can take any value in an interval, so they are continuous.
9. B. $\mu = 1(0.4) + 2(0.4) + 5(0.2) = 0.4 + 0.8 + 1.0 = 2.2$. Choice A, 2.67, is the unweighted average of 1, 2 and 5.
10. (a) $p = 1 - (0.18 + 0.37 + 0.27 + 0.06) = 0.12$.
    (b) $P(X \geq 2) = 0.27 + 0.12 + 0.06 = 0.45$.
    (c) $\mu = 0(0.18) + 1(0.37) + 2(0.27) + 3(0.12) + 4(0.06) = 1.51$. $\sigma^2 = (1.51)^2(0.18) + (0.51)^2(0.37) + (0.49)^2(0.27) + (1.49)^2(0.12) + (2.49)^2(0.06) \approx 1.21$, so $\sigma \approx 1.10$.
    (d) Over many students, the number of siblings averages about 1.51 per student.
11. (a) Net gain X: -4 (P = 0.55), 1 (P = 0.30), 8 (P = 0.15).
    (b) $\mu_X = -4(0.55) + 1(0.30) + 8(0.15) = -2.2 + 0.3 + 1.2 = -0.7$ tickets.
    (c) $\sigma^2 = (-3.3)^2(0.55) + (1.7)^2(0.30) + (8.7)^2(0.15) = 5.9895 + 0.867 + 11.3535 = 18.21$, so $\sigma \approx 4.27$ tickets.
    (d) The game is not fair, because the mean net gain is -0.7, not 0. Over many plays, a player loses an average of 0.7 ticket per play.
12. (a) $E(\text{cost}) = 0(0.92) + 100(0.06) + 300(0.02) = 6 + 6 = 12$ dollars.
    (b) Profit per warranty is $40 - 12 = 28$ dollars, so for 1,000 warranties the expected profit is $28{,}000$ dollars.
    (c) By the law of large numbers, the average cost per warranty gets closer to 12 dollars as the number of warranties grows. With only 5 warranties, a single 300-dollar claim can swing the average a lot.
13. (a) $a + b = 1 - 0.10 - 0.20 = 0.70$ and $2(0.10) + 4a + 6b + 8(0.20) = 5.6$, so $4a + 6b = 3.8$. Substituting $a = 0.70 - b$: $2.8 + 2b = 3.8$, so $b = P(6) = 0.50$ and $a = P(4) = 0.20$.
    (b) $\sigma^2 = (2 - 5.6)^2(0.1) + (4 - 5.6)^2(0.2) + (6 - 5.6)^2(0.5) + (8 - 5.6)^2(0.2) = 1.296 + 0.512 + 0.08 + 1.152 = 3.04$, so $\sigma \approx 1.74$.
    (c) Yes: all probabilities are between 0 and 1 and they sum to $0.1 + 0.2 + 0.5 + 0.2 = 1$.
14. (a) $\mu = 0(0.20) + 1(0.30) + 2(0.25) + 3(0.15) + 4(0.07) + 5(0.03) = 1.68$ alerts. $\sigma^2 \approx 1.698$, so $\sigma \approx 1.30$ alerts.
    (b) $P(X \geq 4) = 0.07 + 0.03 = 0.10$.
    (c) Five alerts is $(5 - 1.68)/1.30 \approx 2.5$ SDs above the mean and has probability only 0.03, so it is fairly unusual, though not impossible.
15. (a) Sums 2 to 8 have 1, 2, 3, 4, 3, 2, 1 favorable outcomes out of 16, so $P = 1/16, 2/16, 3/16, 4/16, 3/16, 2/16, 1/16$.
    (b) $P(X \geq 6) = (3 + 2 + 1)/16 = 6/16 = 0.375$.
    (c) By symmetry the mean is 5. $\sigma^2 = \sum(x - 5)^2 P(x)$, and the numerators over 16 are $9(1) + 4(2) + 1(3) + 0(4) + 1(3) + 4(2) + 9(1) = 40$, divided by 16 gives 2.5, so $\sigma = \sqrt{2.5} \approx 1.58$.

5.4 Combining Random Variables and Transforming Variables

PRACTICE · COMBINING RANDOM VARIABLES AND TRANSFORMING VARIABLES

1. A. $\mu_Y = 4 - 3(5) = -11$. Choice C adds 3 times the mean to 4, a sign error.
2. C. $\sigma_Y = |-3|(2) = 6$. The negative slope flips the direction but the SD cannot be negative (choice A). Choice B forgets to multiply by 3.
3. B. Independent, so $\sigma = \sqrt{3^2 + 4^2} = 5$. Choice A adds the SDs, which is the common mistake.
4. D. For independent variables, the variance of a difference also adds: $\sqrt{9 + 16} = 5$. Choice A subtracts SDs, and choice B adds them.
5. B. $\text{SD}(2X) = 2(5) = 10$, while $\text{SD}(X_1 + X_2) = \sqrt{25 + 25} \approx 7.07$. Doubling one variable is not the same as summing independent copies.
6. D. Adding a constant shifts the center only. The spread (SD) does not change, so choice A fails.
7. C. $D = X - Y$ has mean 10 and SD $\sqrt{36 + 64} = 10$. $P(D > 0) = P(Z > -1) = 0.841$. Choice A uses the SD 14, from adding the SDs.
8. A. $\sigma_F = 1.8(5) = 9$. Choice B, 41, wrongly applies the whole formula $32 + 1.8(5)$ to the SD; the 32 does not affect spread.
9. B. Variances add: $3(2^2) = 12$, so the SD is $\sqrt{12} \approx 3.46$. Choice A, 6, adds the SDs.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. (a) $\mu_Y = 1.1(74) + 5 = 86.4$. $\sigma_Y = 1.1(8) = 8.8$. Median: $1.1(76) + 5 = 88.6$. IQR: $1.1(12) = 13.2$.
    (b) The shape does not change; the transformation has positive slope, so the distribution is stretched and shifted but keeps the same form.
11. (a) $\mu = 500 + 502 = 1002$ g. $\sigma = \sqrt{3^2 + 4^2} = 5$ g.
    (b) $\mu_{B-A} = 2$ g. $\sigma = \sqrt{3^2 + 4^2} = 5$ g.
    (c) $B - A$ is normal with mean 2 and SD 5. $P(B - A > 0) = P(Z > -0.4) = 0.655$.
12. (a) $\mu_Y = 6(0.7) - 4 = 0.2$ dollars. $\sigma_Y = 6(0.781) \approx 4.69$ dollars.
    (b) Total profit: mean $0.2 + 0.2 = 0.4$ dollars; SD $\sqrt{4.69^2 + 4.69^2} \approx 6.63$ dollars.
13. (a) $\mu = 5(7) = 35$ kg. $\sigma = \sqrt{5(1.2^2)} = 1.2\sqrt{5} \approx 2.68$ kg.
    (b) $z = (40 - 35)/2.68 \approx 1.86$, so $P(\text{total} > 40) = 0.031$. About 3% of loads exceed the limit.
14. (a) If X and Y were independent, variances would add: $0.5 + 0.5 = 1$.
    (b) But $X + Y = 2$ every time, a constant, so the variance is 0.
    (c) The rule $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$ requires independence; here X and Y are strongly dependent (more heads means fewer tails), so the rule gives a wrong answer.
15. (a) $\mu_Z = (65 - 65)/8 = 0$ and $\sigma_Z = 8/8 = 1$.
    (b) $\mu_W = 100 - 2(65) = -30$. $\sigma_W = |-2|(8) = 16$.
    (c) The SD is a measure of spread (distance from the mean), which cannot be negative. The negative slope reverses the direction of the data, but the spread is multiplied by $|b|$.

5.5 Binomial Distributions

PRACTICE · BINOMIAL DISTRIBUTIONS

1. A. Twenty trials from a very large school, two outcomes, constant probability 0.40, independent. Choice A has no fixed number of trials, choice C is not a count of successes, and choice D has a changing probability.
2. C. $\binom{5}{2}(0.4)^2(0.6)^3 = 10(0.16)(0.216) = 0.3456$. Choice A, 0.0346, omits the binomial coefficient 10 that counts the arrangements.
3. A. At least 7 is the complement of at most 6: $1 - \text{binomcdf}(12, 0.5, 6) = 0.387$. Choice D subtracts the cumulative probability through 7, so it leaves out P(X = 7).
4. D. $\mu = np = 10$ and $\sigma = \sqrt{50(0.2)(0.8)} = \sqrt{8} \approx 2.83$. Choice B uses the variance 8 as the SD.
5. B. At most one means X $\leq$ 1, which is the cdf at 1: 0.184. Choice A is exactly one, which excludes zero failures.
6. C. A binomial model needs the same probability of success on every trial, and here the probability rises. The count still has two outcomes and 25 fixed trials.
7. B. $z = (62 - 50)/5 = 2.4$, so 62 is unusual (probability about 0.01 of 62 or more). Choice A confuses 12 units with 1 SD.
8. D. From the graph, $P(X \leq 2) = 0.0282 + 0.1211 + 0.2335 = 0.3828 \approx 0.38$. Choice A is only the bar at 2.
9. C. For n = 100, p = 0.3: np = 30 and n(1-p) = 70, both at least 10. The others fail: n = 50, p = 0.10 has np = 5; n = 20, p = 0.90 has n(1-p) = 2; n = 200, p = 0.04 has np = 8.
10. (a) Yes: fixed 10 trials, two outcomes, independent, same p. $n = 10$, $p = 0.6$.
    (b) Yes: 8 students is under 10% of 2,000, so approximately independent; $n = 8$, $p = 0.25$.
    (c) No: the number of trials is not fixed (it is a geometric situation).
    (d) No: sampling without replacement from a small population (5 is more than 10% of 52) makes the trials dependent and the probability changes.
11. (a) Two outcomes (make or miss), independent shots, fixed $n = 10$, constant $p = 0.70$. So $X \sim B(10, 0.70)$.
    (b) $P(X = 7) = \binom{10}{7}(0.7)^7(0.3)^3 \approx 0.2668$.
    (c) $P(X \leq 5) = \text{binomcdf}(10, 0.7, 5) \approx 0.1503$.
    (d) $P(X \geq 8) = 1 - P(X \leq 7) = 1 - \text{binomcdf}(10, 0.7, 7) \approx 0.3828$.
    (e) $\mu = 10(0.7) = 7$ shots and $\sigma = \sqrt{10(0.7)(0.3)} \approx 1.45$ shots.
12. (a) $X \sim B(12, 0.25)$: fixed 12 independent guesses, each with success probability 0.25.
    (b) $\mu = 12(0.25) = 3$ correct and $\sigma = \sqrt{12(0.25)(0.75)} = 1.5$.
    (c) $P(X = 3) = \text{binompdf}(12, 0.25, 3) \approx 0.2581$.
    (d) $P(X \geq 6) = 1 - \text{binomcdf}(12, 0.25, 5) \approx 0.0544$. Guessing rarely yields half the quiz correct.
13. $X \sim B(200, 0.04)$ with conditions met. $P(X \geq 15) = 1 - \text{binomcdf}(200, 0.04, 14) \approx 0.0152$.
    Interpretation: if the defect rate is really 0.04, only about 1.5% of batches of 200 would have 15 or more defective parts.
    Because that is a small probability, such a batch is fairly strong evidence (though not proof) that the defect rate has risen above 0.04.
14. (a) $np = 400(0.55) = 220 \geq 10$ and $n(1 - p) = 400(0.45) = 180 \geq 10$, so the normal approximation is reasonable.
    (b) Mean 220, SD $\sqrt{99} \approx 9.95$. With continuity correction, $z = (239.5 - 220)/9.95 \approx 1.96$ and $P \approx 0.025$.
    (c) The estimate 0.0250 is very close to the exact 0.0247, which shows the normal approximation works well when np and n(1-p) are large.
15. We need $1 - 0.96^n \geq 0.95$, so $0.96^n \leq 0.05$.
    Taking logarithms: $n \geq \ln(0.05)/\ln(0.96) \approx 73.4$ (the inequality reverses when dividing by the negative $\ln 0.96$).
    So $n = 74$. Check: $1 - 0.96^{74} \approx 0.951 \geq 0.95$ and $1 - 0.96^{73} \approx 0.949 < 0.95$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5.6 Geometric Distributions

PRACTICE · GEOMETRIC DISTRIBUTIONS

1. A. Counting flips until the first success is the waiting-time description of a geometric variable. Choice B is binomial (fixed number of flips), and choice C counts successes in a fixed hand.
2. B. $P(X = 3) = (0.75)^2(0.25) = 0.1406$. Choice A, 0.4219, is the binomial probability of exactly one success in three trials, which allows the success in any position.
3. C. $\mu = 1/0.2 = 5$ and $\sigma = \sqrt{0.8}/0.2 \approx 4.47$. Choice A mishandles the SD formula, and choice B uses 4 for the mean.
4. D. More than 4 trials means the first four all fail: $0.7^4 = 0.2401$. Choice A, 0.7599, is $P(X \leq 4)$.
5. C. The calculator's cumulative function adds probabilities from 1 through k, giving $P(X \leq k)$. The pdf gives the single value (choice A).
6. A. The mean of a geometric distribution is a long-run average, not a prediction or a guarantee. Choice B treats the mean as the exact trial of the first success.
7. B. At least 10 inspections means the first 9 are all good: $0.9^9 = 0.3874$. Choice A, 0.3487, is 0.9^{10} (more than 10), and choice C is its complement.
8. D. The shots are independent, so the past misses do not change the chance: 0.6. Choices A and B are the gambler's fallacy in either direction.
9. C. Without replacement, the probability of a heart changes after each card and trials depend on earlier ones, so independence and constant p fail. Choice A is wrong because geometric settings do not fix the number of trials; that is not the problem here.
10. (a) Geometric: counting flips until the first success, independent, p = 0.5.
    (b) Binomial: a fixed 8 independent flips, p = 0.5, counting heads.
    (c) Neither: drawing without replacement means the trials are dependent and the probability of an ace changes.
    (d) Geometric with p = 0.03: independent customers counted until the first complaint.
11. (a) $P(X = 3) = (0.85)^2(0.15) = 0.1084$.
    (b) $P(X \leq 4) = 1 - 0.85^4 = 1 - 0.5220 = 0.4780$.
    (c) $P(X > 6) = 0.85^6 \approx 0.3771$.
    (d) $\mu = 1/0.15 \approx 6.67$ spins and $\sigma = \sqrt{0.85}/0.15 \approx 6.15$ spins.
12. (a) $P(X = 10) = (0.96)^9(0.04) \approx 0.0277$.
    (b) At least 25 inspected means the first 24 are good: $0.96^{24} \approx 0.3754$.
    (c) $\mu = 1/0.04 = 25$ parts. Over many production runs, the number of parts inspected until the first defect averages 25; any particular run could be much shorter or longer.
13. (a) A guess is correct with probability $p = 0.25$.
    (b) $P(X \leq 3) = 1 - 0.75^3 = 1 - 0.4219 = 0.5781$.
    (c) $\mu = 1/0.25 = 4$ questions and $\sigma = \sqrt{0.75}/0.25 \approx 3.46$ questions.
14. (a) Binomial: $5(0.3)(0.7)^4 \approx 0.3601$.
    (b) Geometric: $(0.7)^4(0.3) \approx 0.0720$.
    (c) The binomial probability allows the one success to occur on any of the 5 trials (5 arrangements), while the geometric probability requires the single success to be on the 5th trial, so it is about one fifth as large.
15. (a) We need $1 - 0.9^k \geq 0.90$, so $0.9^k \leq 0.10$ and $k \geq \ln(0.1)/\ln(0.9) \approx 21.9$. So $k = 22$.
    (b) $k = 22$: $1 - 0.9^{22} \approx 0.9015 \geq 0.90$. $k = 21$: $1 - 0.9^{21} \approx 0.8906 < 0.90$. So 22 attempts are needed.

5.7 Continuous Random Variables and Uniform Models

PRACTICE · CONTINUOUS RANDOM VARIABLES AND UNIFORM MODELS

1. A. A single value of a continuous variable has no width, hence zero area and probability zero. Choice B confuses the density height with a probability.
2. D. $P(X < 5) = (5 - 2)/(10 - 2) = 3/8 = 0.375$. Choice A, 0.5, uses the interval [0, 10] or the mean, and choice C is 5/8.
3. C. The area must be 1, so the height is $1/(20 - 0) = 0.05$. Choice A confuses the width with the height.
4. A. $\mu = (0 + 12)/2 = 6$ and $\sigma = 12/\sqrt{12} \approx 3.46$. Choice B uses half the range for the SD.
5. B. A density on [0, 4] with a constant height h needs $4h = 1$, so $h = 0.25$. Choice A has area 2, choice D has area 8, and choice C is negative for x > 2, and a density cannot dip below the axis.
6. D. $z = 1.5$ so the upper tail is 0.0668. Choice A is the lower tail, and choice B is the tail beyond 1 SD.
7. C. Time can take any value in an interval. The other three are counts, which are discrete.
8. D. $P(X > 9)/P(X > 6) = 0.25/0.50 = 0.50$. Choice A, 0.25, ignores the condition and gives P(X > 9).
9. A. The probability between 2 and 5 equals the area to the left of 5 minus the area to the left of 2. Choice B adds overlapping regions and exceeds 1 in many cases.
10. (a) The width is 20, so the height is $1/20 = 0.05$.
    (b) $P(X < 25) = 5/20 = 0.25$. $P(28 < X < 35) = 7/20 = 0.35$.
    (c) Mean = median $= (20 + 40)/2 = 30$ minutes. $\sigma = 20/\sqrt{12} \approx 5.77$ minutes.
    (d) The 90th percentile satisfies $(x - 20)/20 = 0.90$, so $x = 38$ minutes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

11. (a) The graph is a triangle with base 10 and height $0.02(10) = 0.2$, so the area is $\frac{1}{2}(10)(0.2) = 1$. The function is never negative, so it is a valid density.
    (b) $P(X < 5) = \frac{1}{2}(5)(0.1) = 0.25$. $P(X < 8) = \frac{1}{2}(8)(0.16) = 0.64$, so $P(X > 8) = 0.36$.
    (c) $0.01m^2 = 0.5$ gives $m = \sqrt{50} \approx 7.07$ years. The mean (6.67) is less than the median (7.07), so the distribution is skewed left.
12. (a) $z = 1.5$, so $P(X > 56) = 0.0668$.
    (b) 46 and 54 are 1 SD from the mean, so $P(46 < X < 54) \approx 0.6827$.
    (c) $z = 1.2816$, so the 90th percentile is $50 + 1.2816(4) \approx 55.13$.
    (d) For a continuous variable, probability is area and a single point has no width, so $P(X = 50) = 0$. The curve is highest near 50, so intervals around 50 have the most probability, but no individual value has positive probability.
13. (a) $0.3 \times 0.3 = 0.09$ by independence.
    (b) $1 - P(\text{both} \le 0.8) = 1 - 0.8^2 = 1 - 0.64 = 0.36$.
    (c) $P(X < 0.5)P(Y > 0.5) = 0.5 \times 0.5 = 0.25$.
14. (a) The mean is the midpoint, 15 minutes after 2:00, that is 2:15 pm.
    (b) Given $X > 10$, the remaining time is uniform on [10, 30], so $P(X \le 20 \mid X > 10) = (20 - 10)/(30 - 10) = 0.5$.
    (c) Within 5 minutes of 2:15 means $10 < X < 20$: $P = 10/30 \approx 0.333$.
15. (a) A single point has no width, so its area under the density is 0. The value 0.5 can still occur because probabilities are assigned to intervals, and the intervals shrink toward zero as they close in on 0.5.
    (b) $P(0.49 < X < 0.51) = 0.02$.
    (c) The height of the density is not a probability; the total area (height 1 times width 1) is what equals 1. Probability is height times width.

T Unit 5 Test

UNIT TEST · UNIT 5 TEST

1. C. Rearranging the addition rule gives $P(A \text{ and } B) = 0.45 + 0.30 - 0.60 = 0.15$. It is not 0, so the events are not mutually exclusive, and $0.45 \times 0.30 = 0.135 \neq 0.15$, so they are not independent. Choice B would need the product to equal 0.15.
2. A. The complement of 'at least one' is 'none sound': $0.70^4 = 0.2401$, so the answer is $1 - 0.2401 = 0.7599$. Choice B is the probability that none sounds, and choice D adds 4×0.30, which exceeds 1.
3. B. Given Grades 11-12, the whole is the 100 students in that column, and 35 of them are in choir: 0.35. Choice C divides by the 60 choir members, which is the reverse conditional.
4. D. Spam and marked: $0.40(0.90) = 0.36$. Non-spam and marked: $0.60(0.05) = 0.03$. $P(\text{spam} \mid \text{marked}) = 0.36/0.39 \approx 0.92$. Choice A is the filter's 90% rate, which is $P(\text{marked} \mid \text{spam})$, the reverse.
5. C. $\mu = 0(0.1) + 1(0.3) + 2(0.4) + 3(0.2) = 1.7$. Choice A is the unweighted average of the four values.
6. A. Expected tickets received: $0(0.5) + 4(0.3) + 10(0.2) = 3.2$. Net: $3.2 - 3 = 0.2$ tickets. Choice B forgets the cost.
7. B. For independent variables, variances add even for a difference: $\sqrt{6^2 + 8^2} = 10$. Choice A subtracts the SDs, and choice C adds them.
8. D. $P(X \ge 2) = 1 - P(X \le 1) = 1 - 0.3758 = 0.6242$. Choice A is $P(X \le 1)$, the complement, and choice B is $P(X = 2)$.
9. C. $\mu = np = 20$ and $\sigma = \sqrt{80(0.25)(0.75)} = \sqrt{15} \approx 3.87$. Choice A uses the variance 15 as the SD.
10. A. The first three trials fail and the fourth succeeds: $(0.8)^3(0.2) = 0.1024$. Choice B uses the exponent 4, which would be four failures in a row.
11. B. For a uniform distribution, the probability is the interval length over the total length: $(22 - 14)/(30 - 10) = 8/20 = 0.40$. Choice A uses the width 6 incorrectly.
12. D. $z = (69 - 66)/3 = 1$, and the area to the left of $z = 1$ is 0.8413. Choice A is the area above $z = 1$ (the upper tail).
13. (a) 2 points: 1 point for the flu paths $0.30(0.04) = 0.012$ and $0.70(0.12) = 0.084$; 1 point for the sum 0.096.
    (b) 2 points: 1 point for the Bayes setup, the vaccinated flu path over the total, $0.012/0.096$; 1 point for the answer 0.125.
    (c) 2 points: 1 point for comparing $P(\text{flu} \mid \text{shot}) = 0.04$ with $P(\text{flu}) = 0.096$; 1 point for the conclusion of not independent, in context.
    Model response: (a) The tree has first branches Shot (0.30) and No shot (0.70), then Flu (0.04 and 0.12). $P(\text{flu}) = 0.30(0.04) + 0.70(0.12) = 0.012 + 0.084 = 0.096$. (b) $P(\text{shot} \mid \text{flu}) = 0.012/0.096 = 0.125$. (c) If independent, $P(\text{flu} \mid \text{shot})$ would equal $P(\text{flu})$. But $0.04 \neq 0.096$, so they are not independent; students who got the shot caught the flu much less often.
14. (a) 2 points: 1 point for validity (all probabilities in [0, 1] and $0.50 + 0.30 + 0.15 + 0.05 = 1$); 1 point for the mean $0(0.5) + 10(0.3) + 20(0.15) + 50(0.05) = 8.5$ points with the interpretation: over many spins, the average award is 8.5 points per spin.
    (b) 2 points: 1 point for the variance setup $(0 - 8.5)^2(0.5) + (10 - 8.5)^2(0.3) + (20 - 8.5)^2(0.15) + (50 - 8.5)^2(0.05) = 142.75$; 1 point for $\sigma = \sqrt{142.75} \approx 11.95$ points.
    (c) 2 points: 1 point for net mean $8.5 - 8 = 0.5$ points per spin; 1 point for the conclusion that the game slightly favors the player in the long run (a positive expected net gain), though single spins vary a lot.
    (d) 2 points: 1 point for the mean of the total $25(8.5) = 212.5$ points; 1 point for the SD $\sqrt{25}(11.95) \approx 59.74$ points, using independence (variances add, so the SD grows by $\sqrt{25} = 5$, not 25).
    (e) 2 points: 1 point for $Y \sim B(10, 0.05)$ with conditions (fixed 10 independent spins, constant probability); 1 point for $P(Y = 2) = \binom{10}{2}(0.05)^2(0.95)^8 \approx 0.0746$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 6: Sampling Distributions

Understand how statistics vary from sample to sample. 4 lessons, 74 practice and test items, all with full explanations.

In this unit · Exam Unit 2

6.1 What Is a Sampling Distribution?
6.2 Sampling Distribution of a Sample Proportion
6.3 Sampling Distribution of a Sample Mean and the Central Limit Theorem
6.4 Sampling Distributions for Differences
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

6.1 What Is a Sampling Distribution?

Learn the difference between parameters and statistics, see why samples vary, and describe sampling distributions by center, shape and spread, with bias and variability kept straight.

Why This Matters

A poll says 46 percent of a sample prefers a new school schedule. Would a second sample say 46 percent again? Almost certainly not. The big question of this unit is: **how much do sample results change from sample to sample, and what pattern do those changes follow?** Students who skip this idea memorize inference formulas without knowing why they work.

The most common errors are mixing up a parameter with a statistic, treating one sample result as if it were the truth, and confusing *bias* (aiming at the wrong value) with *variability* (scatter). This lesson builds the vocabulary and the simulation habit that every later inference lesson uses. On the AP exam these ideas live in Exam Unit 2 and support Exam Units 3 and 4 (confidence intervals and tests), and free-response questions often ask you to describe a simulated sampling distribution in context.

Core Ideas

A **parameter** is a number that describes a *population*. It is fixed, usually unknown, and does not change with the sample you draw. A **statistic** is a number that describes a *sample*. It is known once you collect data, and it changes from sample to sample. We use a statistic to estimate a parameter.

Quantity	Parameter (population)	Statistic (sample)
Proportion	p	$\hat{p}$
Mean	μ	$\bar{x}$
Standard deviation	σ	s
Slope of a line	β	b

Sampling variability and sampling distribution

Sampling variability: different random samples from the same population give different values of a statistic.

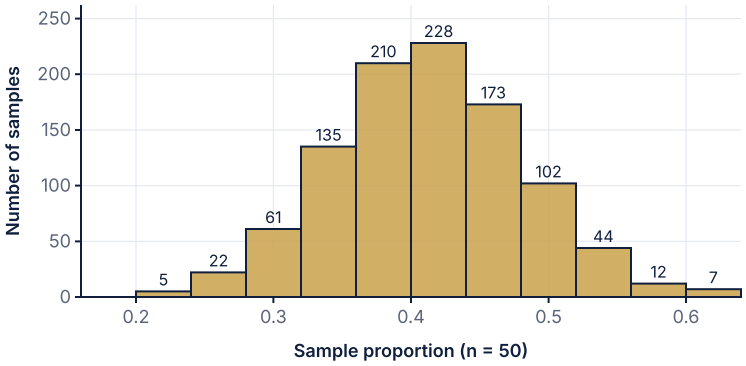
Sampling distribution: the distribution of values a statistic takes over ALL possible samples of the same size from the same population.

It is a distribution of statistics, not of individual observations. Each value in it comes from one whole sample.

We rarely can list all possible samples, so we **simulate**. Choose a population model (for example, $p = 0.40$), draw many random samples of size n, compute the statistic for each, and graph the results. The graph approximates the sampling distribution. The histogram below shows 1000 simulated sample proportions with $n = 50$ and $p = 0.40$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



1000 simulated sample proportions, true p = 0.40

Describe a sampling distribution exactly as you describe any distribution: **center, shape, spread**, and unusual features. Here the center is near 0.402 (very close to $p = 0.40$), the shape is roughly symmetric and mound-shaped, and the standard deviation is about 0.070, with values from 0.14 to 0.64.

How to describe a simulated sampling distribution

Center: where the values balance (mean or median). Compare it to the true parameter.

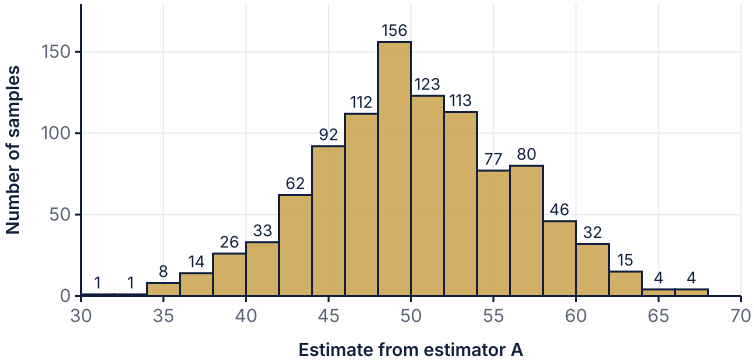
Shape: symmetric, skewed, single peak or not.

Spread: standard deviation or range. Larger samples give less spread.

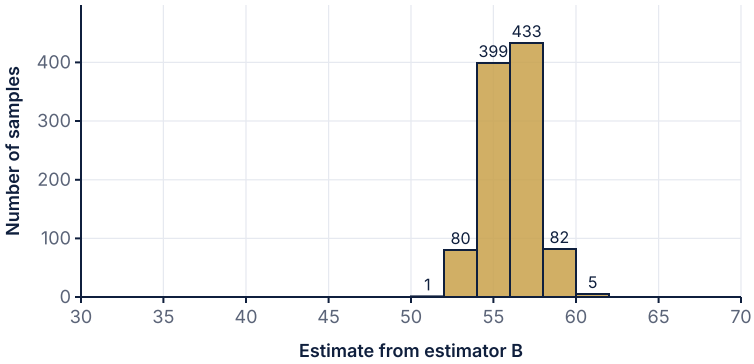
Always name the statistic and the context, for example: the sample proportion of students who walk to school.

Bias and Variability

A statistic is an **unbiased estimator** of a parameter if the center (mean) of its sampling distribution equals the parameter. Bias is a property of the *whole sampling distribution*, never of a single sample. **Variability** is the spread of the sampling distribution. Think of a target: bias means the shots are centered away from the bullseye, high variability means the shots are scattered.



Estimator A: true value 50



Estimator B: true value 50

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Estimator A is centered near the true value 50 (mean 50.2) but is widely scattered (SD 6.0). Estimator B is tightly clustered (SD 1.4) but centered near 56.0, so it is biased high. Which is better? Neither alone. The ideal estimator has *low bias and low variability*.

What controls bias and variability

Bias comes from the sampling METHOD (voluntary response, undercoverage, poor wording, a biased statistic). Random sampling avoids it; a larger sample does NOT fix it.

Variability comes mostly from sample size. For random samples, larger n gives smaller spread.

The population size barely matters as long as it is much larger than the sample (at least 10 times larger).

Unbiased estimators: $\hat{p}$ for p, $\bar{x}$ for μ, s^2 for σ^2, from random samples.

Worked Examples

PARAMETER OR STATISTIC?

A district has 4200 high school students. The principal surveys a random sample of 120 and finds 54 of them (45 percent) ride the bus. She also knows the mean commute time for all 4200 students is 23 minutes. Identify each number as a parameter or statistic and give its symbol.

SOLUTION

The 45 percent comes from a sample of 120: a statistic, $\hat{p} = 54/120 = 0.45$.

The 23 minutes is the mean of the whole population of 4200 students: a parameter, $\mu = 23$.

The true proportion of all 4200 students who ride the bus is a parameter p, unknown, which $\hat{p}$ estimates.

SEEING SAMPLING VARIABILITY

A large app has a true proportion $p = 0.40$ of users who enable dark mode. Five different random samples of 30 users give the counts below. Compute each sample proportion and explain what the table shows.

Sample	Count enabling dark mode	$\hat{p}$
1	11	0.367
2	13	0.433
3	16	0.533
4	10	0.333
5	12	0.400

SOLUTION

Divide each count by 30: 0.367, 0.433, 0.533, 0.333, 0.400.

The five statistics all differ from each other and from 0.40, even though every sample came from the same population with the same p.

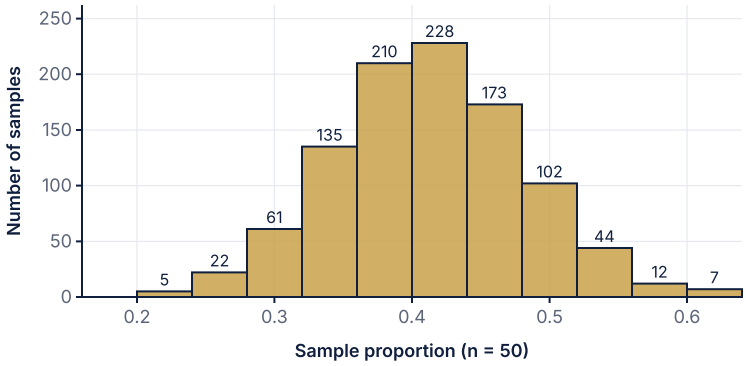
This is sampling variability. The parameter 0.40 never changed; only the sample did.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

DESCRIBE A SIMULATED SAMPLING DISTRIBUTION

Using the histogram of 1000 simulated proportions (n = 50, p = 0.40) shown earlier, a student wants to know how likely a sample of 50 would give $\hat{p} \geq 0.52$. Use a count from the histogram idea and describe the distribution.



1000 simulated sample proportions, true p = 0.40

SOLUTION

Center: about 0.402, essentially the true $p = 0.40$, so $\hat{p}$ appears unbiased.

Shape: roughly symmetric and mound-shaped. Spread: SD about 0.070.

To estimate a probability, count the simulated results at or above 0.52 and divide by 1000. Here 63 of the 1000 results were at least 0.52, so the estimate is about 0.063.

Conclusion: a sample proportion of 0.52 or more would be unusual but not impossible when $p = 0.40$ and n = 50.

BIAS OR VARIABILITY?

Researchers compare two estimators of a true value of 50. Estimator A: mean 50.2, SD 6.0. Estimator B: mean 56.0, SD 1.4. Which is biased? Which is more variable? Which would you prefer if you needed answers close to 50 on a single use?

SOLUTION

B is biased: its mean 56.0 is not 50. A is unbiased: its mean 50.2 is about 50.

A is more variable: SD 6.0 compared with 1.4.

For a single use, compare typical total error. A's error is about 6 in either direction (all scatter). B's error is about 6 too, nearly all in one direction (bias 6, scatter 1.4). So neither is clearly better here. A's scatter can be reduced with a larger sample; B's bias cannot. The best method has both small bias and small SD.

AN UNBIASED AND A BIASED ESTIMATOR, EXACTLY

A tiny population has values 1, 2, 3, 6, so $\mu = 3$ and the maximum is 6. List all 16 equally likely samples of size 2 (drawn with replacement). Find the mean of all possible $\bar{x}$ values and the mean of all possible sample maximums.

SOLUTION

The sampling distribution of $\bar{x}$ has the values (value: share of samples) 1: 1/16; 1.5: 2/16; 2: 3/16; 2.5: 2/16; 3: 1/16; 3.5: 2/16; 4: 2/16; 4.5: 2/16; 6: 1/16. Its mean is 3, equal to $\mu = 3$. So $\bar{x}$ is an unbiased estimator of μ.

The sample maximum averages 4 over all 16 samples, which is less than the population maximum of 6. The sample maximum is a biased (low) estimator of the population maximum.

A sample can never exceed the population maximum, so it tends to fall short.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A BIGGER SAMPLE CANNOT RESCUE A BIASED METHOD

A radio station wants the proportion of listeners who like a song, true p = 0.40. Its call-in line over-represents fans, so its method behaves like sampling from a group where 55 percent like the song. Simulating 5000 call-in polls: n = 100 gives mean 0.551, SD 0.049; n = 400 gives mean 0.549, SD 0.025. Explain what changed and what did not.

SOLUTION

The spread shrank from 0.049 to 0.025, about half, when n was multiplied by 4. Variability decreased.

The center stayed near 0.55, not 0.40. The bias did not shrink.

Conclusion: a larger sample reduces variability but cannot repair a flawed sampling method. Only random sampling removes the bias.

Write It Like a 5

Model Response

Prompt. A company's true proportion of defective bulbs is $p = 0.10$. A student simulates 500 random samples of 40 bulbs and records each sample proportion of defective bulbs; the graph is centered near 0.10 and roughly symmetric with most values between 0.02 and 0.18. (a) Explain what one dot in the graph represents. (b) Describe the distribution. (c) A different student claims that using n = 160 would remove any bias. Respond. **(a)** Each dot is the sample proportion
$hatp$ of defective bulbs from one random sample of 40 bulbs. The graph shows how this statistic varies from sample to sample. **(b)** The sampling distribution is approximately symmetric and mound-shaped, centered near 0.10, which equals the true parameter, so
$hatp$ appears to be an unbiased estimator of p. Nearly all values fall between 0.02 and 0.18, so the spread is about 0.05. **(c)** Increasing n to 160 would reduce the variability (the values would cluster more tightly around 0.10), but the graph is already centered at the true value, so there is no bias to remove. In general, sample size affects variability, not bias.

What the grader looks for

1 point: identifies a dot as a statistic from ONE sample of 40 bulbs, in context. 1 point: describes center AND compares it to the parameter. 1 point: describes shape and spread in context. 1 point: correctly says larger n reduces variability and does not change bias, with a reason.

Coach's Playbook

1. Ask: is this number about a population (parameter) or a sample (statistic)?
2. If a question says 'simulate,' ask what ONE dot represents: one whole sample's statistic.
3. Describe center, shape, spread, and compare the center to the target value.
4. To judge bias, compare the CENTER of the sampling distribution to the parameter. To judge variability, look at the spread.
5. Use a bigger n to shrink spread; use random selection to remove bias.

Common traps

1. Calling the sample mean a parameter because you know it. A number computed from a sample is a statistic. 2. Saying a statistic is biased because ONE sample missed the parameter. Bias is about the center of the whole distribution. 3. Thinking a larger sample fixes a biased method. 4. Writing 'the spread is large' without comparing or giving a value or context.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Mistake Clinic

Spot the error

Student: 'My sample of 80 students had a mean of 6.9 hours of sleep, so the parameter is 6.9.' **Where it fails:** 6.9 describes the sample, so it is the statistic $\bar{x}$. The parameter μ is the mean of the whole population and is unknown. **Correct:** $\bar{x} = 6.9$ is a statistic used to estimate the unknown parameter μ.

Spot the error

Student: 'The sampling distribution is the histogram of the 80 students' sleep hours.' **Where it fails:** That histogram shows individual observations in one sample, the data distribution. A sampling distribution shows values of a statistic over many samples. **Correct:** Each point in a sampling distribution of $\bar{x}$ is a sample mean from a separate random sample of 80 students.

Spot the error

Student: 'The poll missed the true value by 4 points, so the sample proportion is biased.' **Where it fails:** One miss is sampling variability. Bias means the center of the sampling distribution is not at the parameter. **Correct:** Judge bias by checking whether repeated samples average to the parameter; one error of 4 points is expected variability.

Real-World Connection

Manufacturers check batches by sampling a few items. Quality engineers simulate how a sample defect rate behaves when the true rate is acceptable, so they know whether a sample result is just normal sampling variability or a signal that the process has changed.

Challenge Corner

Problem. For the population 1, 2, 3, 6, find the exact mean of the sampling distribution of s^2 (using divisor n - 1 = 1) for samples of size 2 with replacement, and compare it with σ^2 (divisor N). **Solution.** Population variance: $\sigma^2 = 3.5$. Over the 16 samples, the sample variance with divisor 1 is $(a-b)^2/2$, averaging 3.5. That equals σ^2, so s^2 is unbiased. With divisor n = 2 the average is 1.75, smaller than σ^2: biased low. This is why we divide by n - 1.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

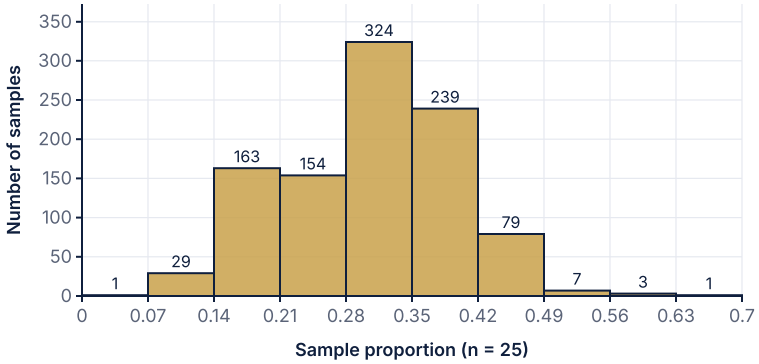
PRACTICE **What Is a Sampling Distribution?**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which of the following is a parameter?
   (A) The mean height of 40 students chosen for a study
   (B) The proportion of all registered voters in a county who support a bond
   (C) The proportion of 600 polled voters who support a bond
   (D) The standard deviation of a sample of 25 batteries

2. A coach measures the mean sprint time of 12 randomly chosen runners on two different days and gets different means. Which idea explains this?
   (A) Bias **(B)** The parameter changed **(C)** Nonresponse
   (D) Sampling variability

3. In a histogram of 1000 simulated sample proportions, each plotted value represents



1000 simulated samples, true p = 0.30

   (A) One individual person in the population **(B)** The population proportion
   (C) One sample proportion from one random sample **(D)** The average of all 1000 samples

4. A statistic is an unbiased estimator of a parameter if
   (A) the mean of its sampling distribution equals the parameter
   (B) its value in any one sample equals the parameter
   (C) its sampling distribution has small standard deviation
   (D) it is computed from a large sample

5. A pollster doubles the sample size from 500 to 1000 using the same random method. The sampling distribution of $\hat{p}$ will have
   (A) smaller spread and the bias removed **(B)** larger spread, same center
   (C) same spread, smaller bias **(D)** smaller spread, same center

6. A sampling method gives $\hat{p}$ values that average 0.62 when the true p = 0.50, with very little scatter. This method has
   (A) low bias and low variability **(B)** high bias and low variability
   (C) low bias and high variability **(D)** high bias and high variability

7. For which pair is the statistic a biased estimator of the parameter when samples are random?
   (A) sample maximum for population maximum **(B)** $\hat{p}$ for p
   (C) $\bar{x}$ for μ **(D)** s^2 (divisor n - 1) for σ^2

8. A random sample of 100 is drawn from a school of 2000 students, then another random sample of 100 from a school of 20000 students, same trait. The variability of $\hat{p}$ in the two cases is
   (A) much smaller for the larger school **(B)** much smaller for the smaller school
   (C) about the same **(D)** impossible to compare

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. To simulate the sampling distribution of $\hat{p}$ for samples of 20 when p = 0.3 using random digits 0 to 9, one dot is
   (A) one digit between 0 and 9
   (B) a group of 20 digits with 0, 1, 2 counted as successes, then successes divided by 20
   (C) a single digit being 3
   (D) 20 digits averaged
10. A town council reports that 62 percent of 150 randomly chosen residents support a new park, and that the town's true mean household size is 2.7.
    (a) Identify each number as a parameter or statistic and give its symbol.
    (b) Explain why the 62 percent might not equal the proportion for all residents.
11. A simulation of 800 random samples of size 36 from a population with p = 0.25 gives sample proportions with mean 0.251, standard deviation 0.072, roughly symmetric and mound-shaped.
    (a) What does each of the 800 values represent?
    (b) Describe the center, shape and spread.
    (c) Does $\hat{p}$ appear biased? Explain.
12. Two estimators of a true population value of 100 are tested 1000 times each. Estimator X: mean 100.2, SD 9. Estimator Y: mean 108, SD 2.
    (a) Which has more bias? Which has more variability?
    (b) Explain which you would trust more and why neither is ideal.
13. Describe how to use a table of random digits to simulate the sampling distribution of the proportion of left-handed students in samples of 25 if 12 percent of students are left-handed. State what one trial is, how many trials you would run, and what you would graph.
14. A population consists of the values 2, 4, 6. List all 9 samples of size 2 drawn with replacement, find the sampling distribution of $\bar{x}$, and show that $\bar{x}$ is unbiased for μ.
15. An online poll about a city festival gets 4000 responses and 71 percent say yes. A random phone survey of 300 residents gets 54 percent. Explain which is more trustworthy for estimating the proportion of all residents and why the larger sample does not win.

6.2 Sampling Distribution of a Sample Proportion

Find the mean and standard deviation of the sampling distribution of a sample proportion, check the Random, 10 percent and Large Counts conditions, and use the normal model for probabilities.

Why This Matters

Polls report a sample proportion and a margin of error. That margin of error comes from one fact: **how far $\hat{p}$ typically lands from the true p**. This lesson gives you the exact center and spread of the sampling distribution of $\hat{p}$ and the conditions that let you use a normal model to find probabilities.

Students often use $\hat{p}$ in place of p when finding the center, forget to check the conditions, or divide by n outside the square root. On the AP exam this is Exam Unit 2 material that feeds directly into the proportion procedures of Exam Unit 3, and it appears as both multiple-choice calculations and free-response explanations.

Core Ideas

Suppose a population has proportion p of successes and we take a random sample of size n. The sample proportion is $\hat{p} = X/n$, where X is the number of successes. Because X varies from sample to sample, so does $\hat{p}$.

Mean and standard deviation of the sampling distribution of $\hat{p}$

Mean: $\mu_{\hat{p}} = p$ (so $\hat{p}$ is an unbiased estimator of p).

Standard deviation: $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.

The SD formula is accurate when the sample is less than 10 percent of the population, that is, $n \leq 0.10N$ (the 10 percent condition, used for sampling without replacement).

Shape: approximately normal when both np and n(1 - p) are at least 10 (the Large Counts condition).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

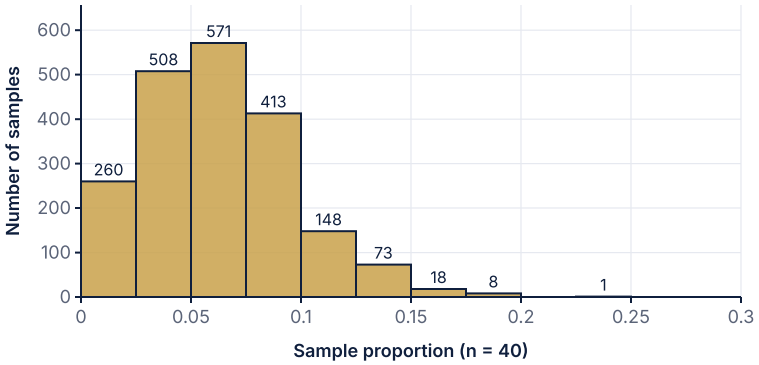
Where the formulas come from. The count of successes X in a random sample follows a binomial-type pattern with mean np and standard deviation $\sqrt{np(1-p)}$. Dividing a count by n divides both the mean and the standard deviation by n, which gives $\mu_{\hat{p}} = np/n = p$ and $\sigma_{\hat{p}} = \sqrt{np(1-p)}/n = \sqrt{p(1-p)/n}$. This is why the center equals p exactly (no bias) while the spread depends on the sample size. You can always convert between counts and proportions: a result of 63 successes in a sample of 90 is $\hat{p} = 0.70$.

- Use $\hat{p}$ when the problem talks about a sample percentage, a rate, or a fraction of the sample.
- Use the count X when the problem asks about a number of successes; convert to $\hat{p}$ by dividing by n.
- The graph of the sampling distribution is always labeled with sample proportions on the horizontal axis, and each dot is one sample.

The formula shows two things. Spread shrinks with the **square root** of n: to cut the SD in half, you must use four times as many observations. And spread is largest when p is near 0.5, because $p(1-p)$ is largest there. The table shows $\sigma_{\hat{p}}$ for $p = 0.5$.

Sample size n	$\sigma_{\hat{p}}$ when $p = 0.5$
25	0.1000
100	0.0500
400	0.0250
1600	0.0125

The shape depends on p and n. When p is small and n is modest, the distribution is right-skewed because $\hat{p}$ cannot go below 0. Below, $p = 0.05$ and $n = 40$ gives $np = 2$, which fails the Large Counts condition. The simulated distribution is skewed with mean 0.050 and SD 0.034, and the largest value is 0.225.

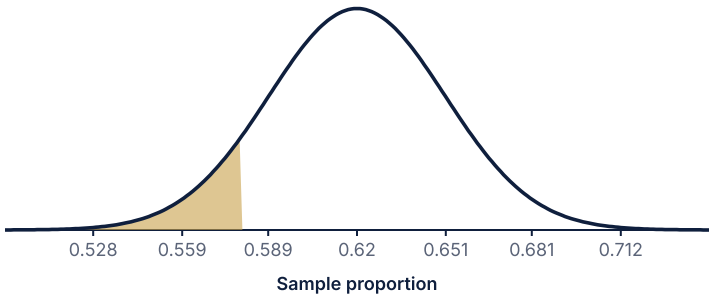


2000 simulated samples, true p = 0.05: right-skewed

Conditions checklist for using a normal model for $\hat{p}$

Random: the data come from a random sample or randomized experiment.
10 percent: $n \leq 0.10N$ when sampling without replacement.
Large Counts: $np \geq 10$ and $n(1-p) \geq 10$, using the population value p (or the claimed value p_0 in a test).

When the conditions hold, standardize with $z = \dfrac{\hat{p} - p}{\sigma_{\hat{p}}}$ and use the normal distribution (the built-in calculator's normal cdf). Here is the shape of such a model with a shaded region for a probability.



Normal model for p-hat, p = 0.62, n = 250

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Worked Examples

CENTER AND SPREAD

In a large school, 30 percent of students walk to school. A random sample of 200 students is taken (the school has thousands). Find the mean and standard deviation of the sampling distribution of $\hat{p}$ and check Large Counts.

SOLUTION

Mean: $\mu_{\hat{p}} = p = 0.30$.

SD: $\sigma_{\hat{p}} = \sqrt{0.30(0.70)/200} = 0.0324$.

Large Counts: $np = 60 \geq 10$ and $n(1-p) = 140 \geq 10$. The 10 percent condition holds since the school is much larger than 2000 students (a sample of 200 is under 10 percent). A normal model is appropriate.

FAILING A CONDITION

A factory finds that 4 percent of its parts are defective. An inspector checks a random sample of 150 parts. Can she use a normal model for $\hat{p}$?

SOLUTION

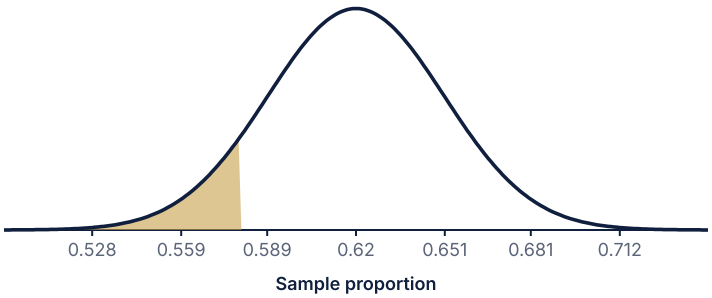
Check $np = 150(0.04) = 6$, which is less than 10.

$n(1-p) = 150(0.96) = 144$ is fine, but both counts must be at least 10.

Because np is less than 10, the sampling distribution is skewed right and a normal model should NOT be used. Use a binomial model or simulation instead.

A PROBABILITY WITH THE NORMAL MODEL

A tablet app has a true proportion $p = 0.62$ of users who finish a tutorial. In a random sample of 250 users (from millions), what is the probability that fewer than 58 percent finish it?



Normal model for p-hat, p = 0.62, n = 250

SOLUTION

Conditions: random sample; 250 is far below 10 percent of millions; $np = 155 \geq 10$ and $n(1-p) = 95 \geq 10$.

Model: $\hat{p}$ is approximately normal with mean 0.62 and SD $\sqrt{0.62(0.38)/250} = 0.0307$.

$z = (0.58 - 0.62)/0.0307 = -1.30$.

$P(\hat{p} < 0.58) = P(z < -1.30) \approx 0.0963$.

There is about a 0.0963 chance that fewer than 58 percent of the sampled users finish the tutorial.

EFFECT OF SAMPLE SIZE AND THE 10 PERCENT CONDITION

A club has 600 members. A leader wants to take a random sample of 80 members without replacement. (a) Is the 10 percent condition met? (b) If the sample were 40 members instead, how does $\sigma_{\hat{p}}$ change for p = 0.5?

SOLUTION

(a) 10 percent of 600 is 60, and 80 is greater than 60. The 10 percent condition fails, so the formula $\sqrt{p(1-p)/n}$ overstates the true spread (it is only approximately right).

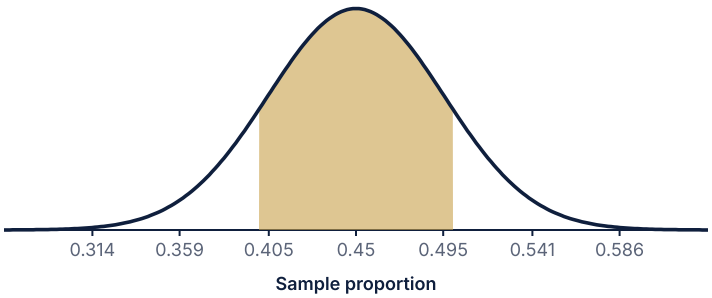
(b) With $n = 40$: $\sqrt{0.25/40} = 0.0791$; with $n = 80$: $\sqrt{0.25/80} = 0.0559$. Halving the sample multiplies the SD by $\sqrt{2} \approx 1.41$, not by 2.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

PROBABILITY OF A RANGE

In a city, $p = 0.45$ of residents support a new bike lane. A poll takes a random sample of 120 residents. Find the probability that $\hat{p}$ is between 0.40 and 0.50.



Normal model for p-hat, p = 0.45, n = 120

SOLUTION

Conditions: random sample, 120 is under 10 percent of the city, $np = 54$ and $n(1-p) = 66$, both at least 10.

$\sigma_{\hat{p}} = \sqrt{0.45(0.55)/120} = 0.0454$.

$z = (0.40 - 0.45)/0.0454 = -1.10$ and $z = (0.50 - 0.45)/0.0454 = 1.10$.

$P(-1.10 < z < 1.10) \approx 0.7291$.

About 0.7291 of such polls land within 5 percentage points of the true value.

IS THE SAMPLE RESULT SURPRISING?

A company claims $p = 0.40$ of customers renew. In a random sample of 150 customers, only 31 percent renewed. Does this make the claim doubtful? Assume 150 is under 10 percent of all customers.

SOLUTION

Conditions: $np = 60$, $n(1-p) = 90$, both at least 10; random sample.

$\sigma_{\hat{p}} = \sqrt{0.40(0.60)/150} = 0.0400$ and $z = (0.31 - 0.40)/0.0400 = -2.25$.

$P(\hat{p} \le 0.31) \approx P(z \le -2.25) = 0.0122$.

A result this low happens about 0.0122 of the time if the claim were true, so it is somewhat unusual (less than 2 percent chance, a low outcome). The result casts doubt on the claim, though formal significance tests (Exam Unit 3) are the tool for deciding.

Write It Like a 5

Model Response

Prompt. A streaming service says that 70 percent of its subscribers use the offline-download feature. A random sample of 200 subscribers (from millions) is taken. (a) Describe the sampling distribution of $\hat{p}$. (b) Find the probability that at least 75 percent of the sample uses the feature. **(a)** Because the sample is random and 200 is less than 10 percent of all subscribers, and $np = 200(0.70) = 140 \ge 10$ and $n(1-p) = 200(0.30) = 60 \ge 10$, the sampling distribution of $\hat{p}$ is approximately normal. Its mean is $\mu_{\hat{p}} = 0.70$ and its standard deviation is $\sigma_{\hat{p}} = \sqrt{0.70(0.30)/200} = 0.0324$. **(b)** $z = (0.75 - 0.70)/0.0324 = 1.54$, so $P(\hat{p} \ge 0.75) = P(z \ge 1.54) \approx 0.0614$. If the claim is true, about 0.0614 of random samples of 200 subscribers would have 75 percent or more using the feature.

What the grader looks for

1 point: states the random sample and 10 percent condition. 1 point: verifies both Large Counts conditions with numbers. 1 point: gives mean 0.70 and SD with the correct formula. 1 point: names the normal shape. 1 point: correct z and probability. 1 point: interprets the probability in context.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Coach's Playbook

1. Identify p and n. Use p (the model value), not $\hat{p}$, for the mean and SD.
2. Check Random, 10 percent and Large Counts and show the numbers.
3. Compute $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$ with the whole fraction under the square root.
4. Standardize, sketch and shade the density, then use the normal cdf on the built-in calculator.
5. Write the answer as a sentence about samples in context.

Common traps

1. Writing $\sigma_{\hat{p}} = p(1-p)/\sqrt{n}$ or $\sqrt{p(1-p)}/n$. The whole $p(1-p)/n$ goes under the root. 2. Using $\hat{p}$ instead of p when checking Large Counts in a probability problem. 3. Treating the Large Counts condition as 'n at least 30'. The rule is np and n(1 - p) at least 10. 4. Forgetting that the 10 percent condition is about the sample being at most one tenth of the population.

Mistake Clinic

Spot the error

Student: 'p = 0.2 and n = 40, so np = 8 and n(1-p) = 32. At least one is over 10, so normal is fine.' **Where it fails:** BOTH counts must be at least 10. **Correct:** np = 8 is less than 10, so the sampling distribution of $\hat{p}$ is not safely normal.

Spot the error

Student: 'The sample of 100 has $\hat{p} = 0.35$, so the SD of the sampling distribution is $\sqrt{0.35(0.65)/100}$.' **Where it fails:** The sampling distribution describes behavior if the true proportion is p. When p is known (given), use it. Using $\hat{p}$ only comes later, for estimation. **Correct:** If $p = 0.40$ is stated, use $\sqrt{0.40(0.60)/100} = 0.049$.

Spot the error

Student: 'I doubled the sample size from 100 to 200, so the standard deviation was cut in half.' **Where it fails:** SD depends on $1/\sqrt{n}$. **Correct:** Doubling n divides the SD by $\sqrt{2}$ (about 1.41). To halve the SD you need four times the sample size.

Real-World Connection

Election and market surveys pick sample sizes by working backward from the formula: a desired margin of error determines how many people to survey. Because $\sigma_{\hat{p}}$ is largest at $p = 0.5$, planners use $p = 0.5$ for the safest (largest) sample size.

Challenge Corner

Problem. A population has $p = 0.5$. What sample size makes $\sigma_{\hat{p}}$ at most 0.02? What happens to the spread at $p = 0.1$ with the same n? **Solution.** Need $\sqrt{0.25/n} \le 0.02$, so $n \ge 0.25/0.0004 = 625$. With $n = 625$ and $p = 0.1$: $\sqrt{0.09/625} = 0.0120$, which is smaller than 0.02 since $p(1-p) = 0.09 < 0.25$. Note also the Large Counts check for p = 0.1: np = 62.5 is at least 10, so the normal model still works.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

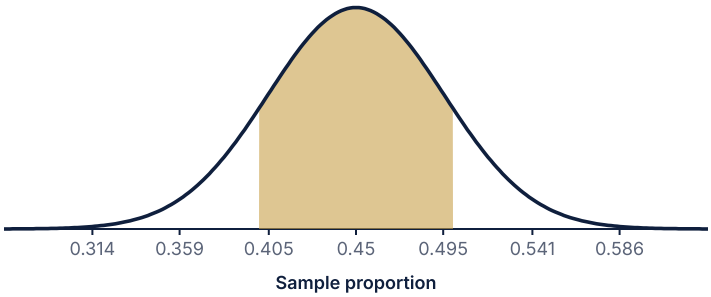
© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE Sampling Distribution of a Sample Proportion

Show your work. A calculator is allowed. Justify conclusions in context.

1. The sampling distribution of $\hat{p}$ for random samples of size 100 from a population with $p = 0.35$ has mean
   (A) 0.35 **(B)** 0.035 **(C)** 0.0477 **(D)** 0.50
2. For $p = 0.20$ and $n = 100$, the standard deviation of $\hat{p}$ is
   (A) 0.4 **(B)** 0.0016 **(C)** 0.04 **(D)** 0.0004
3. Which pair of values of p and n satisfies the Large Counts condition?
   (A) p = 0.05, n = 100 **(B)** p = 0.30, n = 25 **(C)** p = 0.90, n = 80
   (D) p = 0.20, n = 60
4. A sample is taken without replacement from a class of 150 students. The largest sample size for which the 10 percent condition holds is
   (A) 10 **(B)** 15 **(C)** 30 **(D)** 75
5. If the sample size is multiplied by 9, the standard deviation of $\hat{p}$ is multiplied by
   (A) 9 **(B)** 3 **(C)** 1/3 **(D)** 1/9
6. For $p = 0.5$ and $n = 100$, the approximate probability that $\hat{p} \le 0.45$ is
   (A) 0.023 **(B)** 0.159 **(C)** 0.841 **(D)** 0.050
7. The graph shows the normal model for $\hat{p}$ when $p = 0.45$ and $n = 120$. The shaded area corresponds to



Normal model for p-hat, p = 0.45, n = 120

(A) $P(0.40 < \hat{p} < 0.50)$ **(B)** $P(\hat{p} > 0.45)$ **(C)** $P(0.45 < \hat{p} < 0.50)$
(D) $P(\hat{p} = 0.45)$

8. Samples of size 400 are taken from a very large population with $p = 0.30$. About 95 percent of sample proportions fall between
   (A) 0.250 and 0.350 **(B)** 0.210 and 0.390 **(C)** 0.285 and 0.315
   (D) 0.255 and 0.345
9. A student writes $\sigma_{\hat{p}} = \sqrt{p(1-p)}/n$. The error is that this
   (A) is correct for n at least 30
   (B) uses p instead of the sample count
   (C) puts only the square root of p(1 - p) over n instead of p(1 - p)/n under the root
   (D) does not include the mean
10. A school claims 55 percent of its 4000 students have a part-time job. A random sample of 180 students is taken.
    (a) Check the conditions for a normal model for $\hat{p}$.
    (b) State the mean and standard deviation of the sampling distribution.
11. For $p = 0.64$ and $n = 150$ (sample from a very large population), find the probability that $\hat{p}$ is greater than 0.70. Show the conditions, the z-score, and a sentence in context.
12. A lab technician finds that 6 percent of vials are cracked. She inspects a random sample of 90 vials. Explain whether a normal model for $\hat{p}$ is appropriate and describe the likely shape.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. A food truck owner says 25 percent of customers order the spicy sauce. The owner takes a random sample of 100 customers from a very large customer base.
    (a) Find the probability the sample proportion is between 0.20 and 0.30.
    (b) Would a sample of 400 customers make this probability larger or smaller? Explain without computing.
14. A pollster wants the standard deviation of $\hat{p}$ to be at most 0.015 when p = 0.5. What is the minimum sample size? What happens to the needed size if p is known to be 0.2?
15. A manager takes a random sample of 60 members from a 500-member gym and wants to use $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$ with p = 0.4. Evaluate the 10 percent condition and explain the consequence for the actual spread.

6.3 Sampling Distribution of a Sample Mean and the Central Limit Theorem

See how sample means center at the population mean with standard deviation sigma over root n, and use the Central Limit Theorem to justify normal models for skewed populations.

Why This Matters

Averages are more predictable than single observations. A single commute can be awful; the average of 40 commutes rarely is. This lesson explains exactly how much more predictable, and why a **normal model works for sample means even when the data are not normal**: the Central Limit Theorem.

Students typically forget to divide σ by $\sqrt{n}$, mix up the distribution of individual values with the distribution of sample means, or apply the CLT to tiny samples from strongly skewed populations. On the AP exam this is Exam Unit 2 content that directly supports the t procedures in Exam Unit 4, and free-response questions often ask you to explain why a normal model is or is not appropriate.

Core Ideas

Take random samples of size n from a population with mean μ and standard deviation σ. The sample mean $\bar{x}$ changes from sample to sample. Its sampling distribution has a predictable center and spread.

Sampling distribution of $\bar{x}$

Mean: $\mu_{\bar{x}} = \mu$ (so $\bar{x}$ is an unbiased estimator of μ).

Standard deviation: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$, valid when the sample is at most 10 percent of the population ($n \le 0.10N$) for sampling without replacement.

Shape: if the population is normal, $\bar{x}$ is exactly normal for every n. If not, the Central Limit Theorem applies.

Central Limit Theorem (CLT)

For a random sample of size n from ANY population with finite mean and SD, the sampling distribution of $\bar{x}$ is approximately normal when n is large.

A common guideline is n at least 30. Strongly skewed populations or outliers may need larger n; roughly symmetric populations need smaller n.

The CLT is about the shape of the distribution of sample MEANS, not about the data themselves, which keep their original shape.

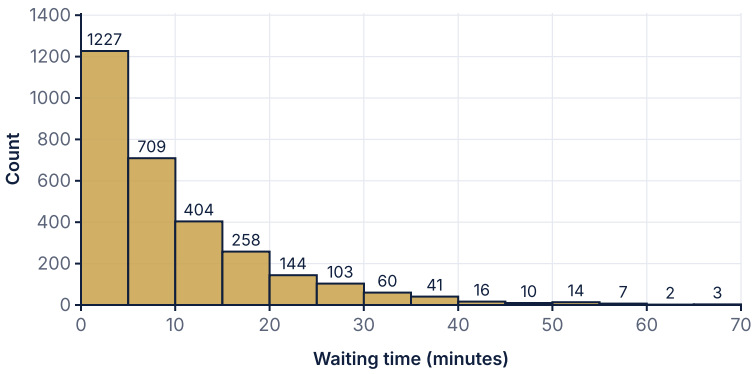
The table shows how $\sigma_{\bar{x}}$ shrinks for a population with $\sigma = 12$. Precision grows with $\sqrt{n}$: quadrupling the sample size halves the spread.

Sample size n	$\sigma_{\bar{x}} = 12/\sqrt{n}$
1	12.00
4	6.00
9	4.00
36	2.00
144	1.00

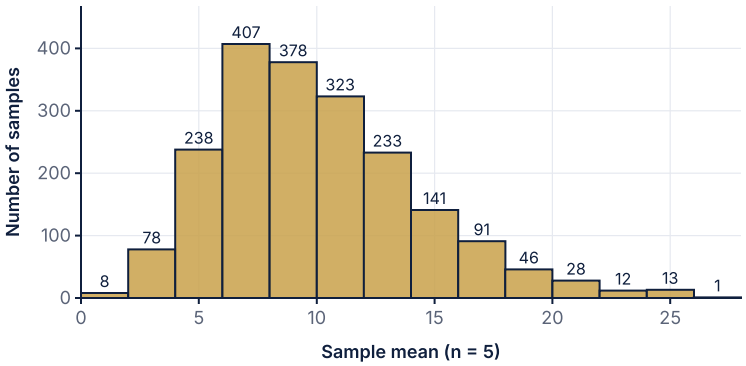
Now watch the CLT work. The population below is strongly right-skewed: waiting times with mean 10 and standard deviation 10 minutes. Samples of size 5 still show skew (mean 10.06, SD 4.35, theory 4.47). By $n = 40$ the sample means are close to symmetric with mean 10.02 and SD 1.57 (theory 1.58).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

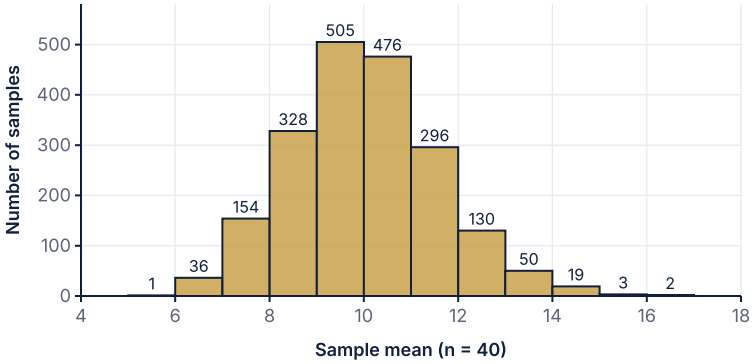
© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Population: individual waits, mean 10, SD 10 (right-skewed)



2000 simulated sample means, n = 5



2000 simulated sample means, n = 40

Checklist for a normal model for $\bar{x}$

Random: random sample or randomized experiment.

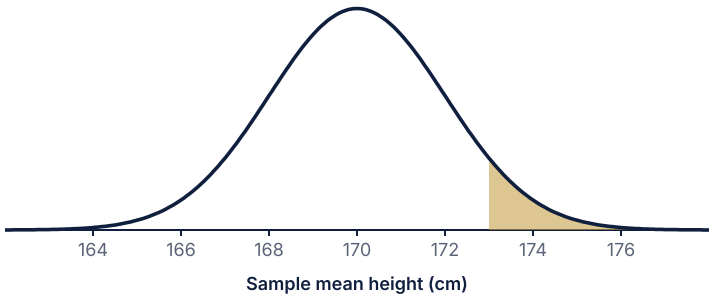
10 percent: $n \leq 0.10N$ if sampling without replacement.

Normal/Large Sample: the population is normal, or n is large enough (about 30 or more, more for strong skew or outliers), or the sample data show no strong skew or outliers.

With these conditions met, standardize with $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ and use the normal model. Compare this with the individual-value model, $z = (x - \mu)/\sigma$: the formulas differ only by the $\sqrt{n}$ in the denominator.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Sampling distribution of x-bar, n = 16

Worked Examples

CENTER AND SPREAD

A population of exam scores has $\mu = 68$ and $\sigma = 12$. Random samples of 36 students are taken from a very large group. Find the mean and standard deviation of $\bar{x}$.

SOLUTION

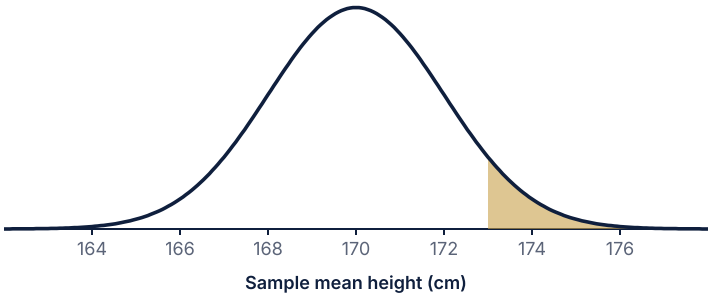
$\mu_{\bar{x}} = 68$.

$\sigma_{\bar{x}} = 12/\sqrt{36} = 12/6 = 2.00$.

The typical sample mean is within about 2 points of 68, while individual scores vary by about 12.

NORMAL POPULATION

Heights of adult members of a rowing club are normal with mean 170 cm and standard deviation 8 cm. A random sample of 16 members is chosen (the club is large). What is the probability that the sample mean height exceeds 173 cm?



Sampling distribution of x-bar, n = 16

SOLUTION

The population is normal, so $\bar{x}$ is normal for any n. Random sample, and 16 is under 10 percent of a large club.

$\mu_{\bar{x}} = 170$ and $\sigma_{\bar{x}} = 8/\sqrt{16} = 2$.

$z = (173 - 170)/2 = 1.50$.

$P(\bar{x} > 173) = P(z > 1.50) \approx 0.0668$.

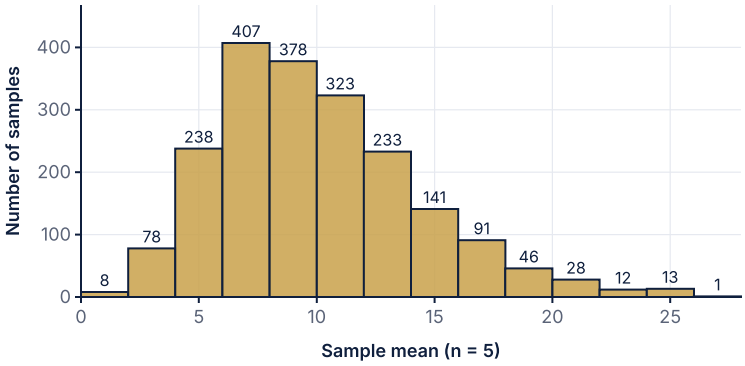
About 0.0668 of random samples of 16 members have a mean height above 173 cm.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

SKEWED POPULATION, SMALL VERSUS LARGE N

Waiting times at a service desk are strongly right-skewed with mean 10 minutes and SD 10 minutes. Describe the sampling distribution of $\bar{x}$ for n = 5 and for n = 40, using the simulated histograms.



2000 simulated sample means, n = 5

SOLUTION

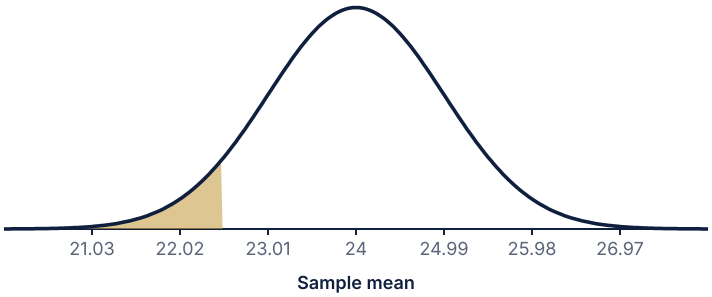
Center: both centers are close to 10 (10.06 and 10.02), as theory says: $\mu_{\bar{x}} = \mu$.

Spread: theory gives $10/\sqrt{5} = 4.47$ and $10/\sqrt{40} = 1.58$; simulations gave 4.35 and 1.57.

Shape: with n = 5 the histogram is still right-skewed, so a normal model is risky. With n = 40 it is nearly symmetric and mound-shaped, so the CLT applies and a normal model is reasonable.

USING THE CLT

A cafe's daily sales per customer have mean 24 dollars and standard deviation 7 dollars, with a right-skewed shape. In a random sample of 50 customers (a small part of all customers), what is the probability that the mean is less than 22.50 dollars?



Sampling distribution of x-bar, mean 24, SD 0.99

SOLUTION

Conditions: random; 50 is under 10 percent of all customers; n = 50 is at least 30, so by the CLT $\bar{x}$ is approximately normal even though the population is skewed.

$\sigma_{\bar{x}} = 7/\sqrt{50} = 0.9899$.

$z = (22.5 - 24)/0.9899 = -1.52$.

$P(\bar{x} < 22.5) \approx P(z < -1.52) = 0.0649$.

About 0.0649 of random samples of 50 customers would average less than 22.50 dollars.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

INDIVIDUALS VERSUS MEANS

Battery life is normal with mean 500 hours and SD 40 hours. Compare the probability that one battery lasts over 520 hours with the probability that the mean of 25 random batteries exceeds 520 hours.

SOLUTION

Individual: $z = (520 - 500)/40 = 0.50$, so $P(X > 520) = 0.3085$.

Mean: $\sigma_{\bar{x}} = 40/\sqrt{25} = 8.0$, $z = 20/8.0 = 2.50$, so $P(\bar{x} > 520) = 0.0062$.

A single battery easily exceeds 520 hours, but the average of 25 batteries rarely does. Averages vary much less than individuals.

HOW LARGE MUST THE SAMPLE BE?

A population has $\sigma = 10$. How large should a random sample be so that about 95 percent of sample means fall within 2 units of μ?

SOLUTION

About 95 percent of a normal distribution lies within 1.96 SDs. We need $1.96\sigma/\sqrt{n} \leq 2$.

$\sqrt{n} \geq 1.96(10)/2 = 9.8$, so $n \geq 96.04$.

Round UP to $n = 97$. Rounding down would not meet the requirement.

Write It Like a 5

Model Response

Prompt. The time students spend on homework nightly is right-skewed with mean 82 minutes and SD 30 minutes. A random sample of 45 students is taken from a large school. (a) Is it appropriate to use a normal model for the sampling distribution of $\bar{x}$? (b) Find the probability that the sample mean exceeds 90 minutes. **(a)** The sample is random and 45 is less than 10 percent of a large school. The population is skewed, but $n = 45 \geq 30$, so by the Central Limit Theorem the sampling distribution of $\bar{x}$ is approximately normal. **(b)** $\mu_{\bar{x}} = 82$ and $\sigma_{\bar{x}} = 30/\sqrt{45} = 4.47$. $z = (90 - 82)/4.47 = 1.79$, so $P(\bar{x} > 90) = P(z > 1.79) \approx 0.0368$. About 0.0368 of random samples of 45 students would have a mean homework time above 90 minutes.

What the grader looks for

1 point: states random sample and the 10 percent condition. 1 point: justifies normality with the CLT and n of at least 30, mentioning the skewed population. 1 point: correct mean and SD (dividing by the square root of n). 1 point: correct z. 1 point: correct probability. 1 point: conclusion in context about sample means.

Coach's Playbook

1. Decide: is the question about one value (use sigma) or a sample mean (use sigma over root n)?
2. Check Random, 10 percent, and normality or large n, and write the numbers.
3. Draw and shade a density curve labeled with the sample mean scale.
4. Use the normal cdf on the built-in calculator with mean mu and SD sigma over root n, or standardize.
5. Conclude in context using the words 'sample mean' or 'average of n'.

Common traps

1. Using σ instead of $\sigma/\sqrt{n}$ for a mean. 2. Claiming the CLT makes the DATA normal. It makes the distribution of sample MEANS approximately normal. 3. Applying 'n at least 30' automatically even when the sample shows extreme outliers or the population is very skewed. 4. Rounding a required sample size down instead of up.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Mistake Clinic

Spot the error

Student: 'Salaries are skewed, but my sample has n = 200, so salaries in the sample are normal by the CLT.' **Where it fails:** The CLT never changes the shape of the data. **Correct:** The sample salaries stay skewed. Because n = 200 is large, the sampling distribution of the sample mean is approximately normal.

Spot the error

Student: 'Scores have mean 70 and SD 10. In a sample of 25, $P(\bar{x} > 74)$ uses z = (74 - 70)/10 = 0.4.' **Where it fails:** This is the formula for an individual score. **Correct:** $\sigma_{\bar{x}} = 10/5 = 2$, so $z = (74 - 70)/2 = 2$ and the probability is about 0.0228.

Spot the error

Student: 'The population is strongly skewed and n = 6, but the CLT says it will be normal.' **Where it fails:** n = 6 is far too small for a strongly skewed population. **Correct:** Say a normal model is not justified unless the population is known to be nearly normal or n is large.

Real-World Connection

Quality control, scientific measurement, and polling all rely on averaging. Averaging 100 readings reduces the random error to one tenth of a single reading's error, which is why repeated measurements give more reliable results.

Challenge Corner

Problem. For $\mu = 68$, $\sigma = 12$ and $n = 36$, find the sample mean cutoff that is exceeded by only 10 percent of samples, and how many individual scores would need to be that extreme. **Solution.** $\sigma_{\bar{x}} = 2$. The 90th percentile of the standard normal is 1.282, so the cutoff is $68 + 1.282(2) = 70.56$. Only 10 percent of sample means of 36 are above 70.56. The same cutoff for an individual score is $z = (70.56 - 68)/12 = 0.21$, so about 0.4154 of individual scores exceed it, much more than 10 percent.

Practice

PRACTICE **Sampling Distribution of a Sample Mean and the Central Limit Theorem**

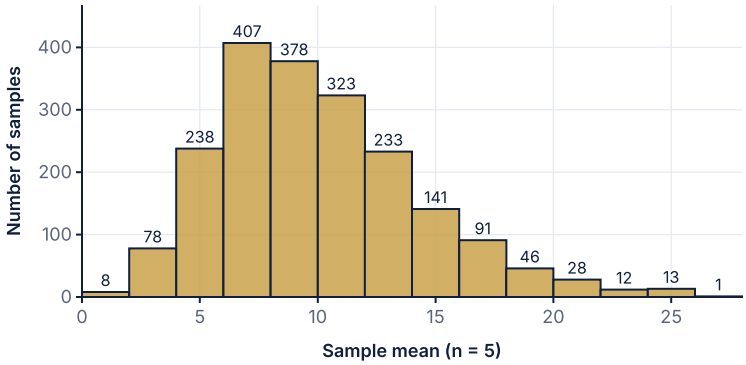
Show your work. A calculator is allowed. Justify conclusions in context.

1. A population has $\mu = 50$ and $\sigma = 20$. For random samples of size 100, the mean and SD of $\bar{x}$ are
   (A) 50 and 2 **(B)** 50 and 20 **(C)** 50 and 0.2
   (D) 5 and 2
2. The Central Limit Theorem says that for large n the sampling distribution of $\bar{x}$ is approximately normal
   (A) only if the population is normal **(B)** regardless of the population's shape
   (C) only if the population is symmetric **(D)** only if sigma is known to be small
3. To cut the standard deviation of $\bar{x}$ in half, the sample size should be
   (A) doubled **(B)** tripled **(C)** quadrupled
   (D) increased by 50 percent

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

4. The histogram shows 2000 simulated sample means from a skewed population with n = 5. Which conclusion is best?



2000 simulated sample means, n = 5

(A) The sample means are approximately normal because n is greater than 1
(B) The distribution of sample means is still noticeably skewed, so n = 5 is too small for the CLT
(C) The population must be normal
(D) The sample means have a larger SD than individuals

5. Daily rainfall (inches) in a region is strongly right-skewed. For which sample size is a normal model for the sample mean MOST justified?

(A) n = 5
(B) n = 12
(C) n = 25 with a few extreme outliers
(D) n = 100

6. Weights of bags are normal with mean 40 lb and SD 3 lb. For a random sample of 9 bags, $P(\bar{x} > 41)$ is approximately

(A) 0.159 **(B)** 0.369 **(C)** 0.841 **(D)** 0.091

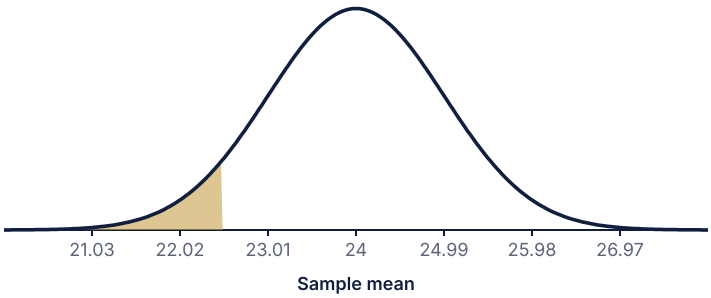
7. A student uses z = (x-bar - mu)/sigma for a sample mean with n = 16. The result will be

(A) too large in magnitude **(B)** negative **(C)** correct
(D) too small in magnitude

8. Which statement about the standard deviation of the sampling distribution of x-bar is true when sampling without replacement from a population of 400?

(A) It does not depend on n
(B) The formula is reliable only for n = 100
(C) The formula sigma/root n is reliable for n = 30
(D) It equals sigma

9. The shaded region of the sampling distribution of x-bar is the probability that the sample mean is between 20 and 22.5. The graph is centered at 24. This probability is



Sampling distribution of x-bar, mean 24, SD 0.99

(A) Small, since both values are 1.5 SD or more below the mean
(B) Large, near 0.9
(C) Exactly 0.5
(D) Equal to P(x-bar is greater than 24)

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. Scores on a reading test have mean 500 and SD 90. A random sample of 36 students is taken from a very large population.
    (a) Find the mean and SD of the sampling distribution of $\bar{x}$.
    (b) What shape does it have if the population is normal? If the population is skewed?
11. The lifetime of a type of LED is normal with mean 40000 hours and SD 5000 hours. A random sample of 25 LEDs is tested. Find the probability that the sample mean is below 38000 hours. Show conditions and a sentence in context.
12. A coach records the number of minutes players run in a game: right-skewed, mean 14, SD 9. For a random sample of 60 player-games (a small part of all games), find $P(\bar{x} > 16)$. Justify the normal model.
13. Explain in two or three sentences why the sample mean of 100 randomly selected values varies less than the mean of 10, and give the ratio of their standard deviations.
14. A manufacturer says the SD of fill volume is 4 mL. How large a random sample is needed so that the standard deviation of the sample mean is at most 0.5 mL?
15. A student says: 'The CLT says my 15 observations are normal, because 15 is above 10.' Identify two errors and give the correct statement.

6.4 Sampling Distributions for Differences

Find the mean and standard deviation of the sampling distribution of a difference of two sample proportions or means, remembering that means subtract and variances add.

Why This Matters

Many real questions compare two groups: Do students taught with videos score higher than those taught by lecture? Is the defect rate lower on the new machine? To judge such a comparison you need to know how the **difference** between two sample statistics behaves from sample to sample.

The key rule is that **means subtract but variances add**. Students often subtract the standard deviations or forget that the spread of a difference is larger than the spread of either part. On the AP exam this idea is the foundation for the two-sample procedures in Exam Units 3 and 4, and you will meet it in multiple-choice and free-response questions that ask for a mean, standard deviation, or probability for a difference.

Core Ideas

Suppose we take two **independent** random samples, one from each of two populations or treatment groups. Independence means that knowing the outcome of one sample tells nothing about the other. For any two independent random quantities, $E(X - Y) = E(X) - E(Y)$ and $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$. Subtracting two quantities never makes the result less variable, because each carries its own uncertainty.

Difference of two sample proportions, $\hat{p}_1 - \hat{p}_2$

Mean: $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$.

Standard deviation: $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$.

Conditions: independent random samples (or randomized groups); each sample at most 10 percent of its population; Large Counts for BOTH groups (n_1p_1, $n_1(1-p_1)$, n_2p_2, $n_2(1-p_2)$ all at least 10). Then the shape is approximately normal.

Difference of two sample means, $\bar{x}_1 - \bar{x}_2$

Mean: $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$.

Standard deviation: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$.

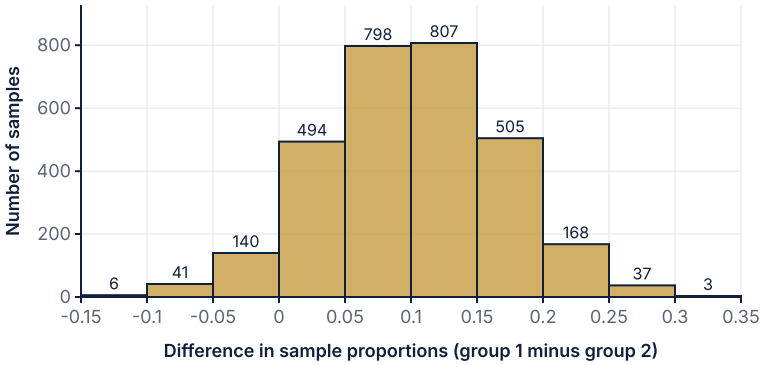
Conditions: independent random samples; 10 percent for each; each population normal, or each sample large (about 30 or more) so the CLT applies. Then the shape is approximately normal.

Notice how both standard deviation formulas are built: add the **variances** of the two sampling distributions, then take the square root. For example, if two independent sampling distributions have SDs of 3 and 4, the difference has SD $\sqrt{3^2 + 4^2} = 5$, not $4 - 3 = 1$ and not $3 + 4 = 7$.

The next graph shows 3000 simulated differences $\hat{p}_1 - \hat{p}_2$ for $p_1 = 0.60$, $n_1 = 100$ and $p_2 = 0.50$, $n_2 = 120$. The center is near the true difference of 0.10 (simulated mean 0.101) and the simulated SD, 0.068, matches the formula value 0.0670. The shape is roughly symmetric and mound-shaped.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



3000 simulated differences, p1 = 0.60 (n = 100), p2 = 0.50 (n = 120)

Notice that even though the true difference is positive, a sizable fraction (0.062 of the simulated samples) produced a negative difference. Sampling variability can reverse the order of two groups, especially for small samples.

Statistic	Mean	Standard deviation
$\hat{p}$	p	$\sqrt{p(1-p)/n}$
$\bar{x}$	μ	$\sigma/\sqrt{n}$
$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$
$\bar{x}_1 - \bar{x}_2$	$\mu_1 - \mu_2$	$\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$

Experiments count too. In a randomized experiment, the two treatment groups play the role of the two samples. Random assignment gives the independence that the formulas need, and the 10 percent condition is not needed because the subjects are not sampled without replacement from a larger population. The same sampling-distribution ideas then describe how the difference in group results would vary if the random assignment were repeated. Finally, always keep track of the order of subtraction: if you define the statistic as group 1 minus group 2, a negative value means group 2 was higher, and the sign of every z-score depends on that choice.

Worked Examples

MEAN AND SD OF A DIFFERENCE OF PROPORTIONS

In a district, 60 percent of juniors and 50 percent of seniors hold part-time jobs. Independent random samples of 100 juniors and 120 seniors are taken. Find the mean and standard deviation of $\hat{p}_J - \hat{p}_S$.

SOLUTION

Mean: $0.60 - 0.50 = 0.10$.

Variances: $0.60(0.40)/100 = 0.00240$ and $0.50(0.50)/120 = 0.00208$. They add to 0.00448.

SD: $\sqrt{0.00448} = 0.0670$.

The SD of the difference (0.0670) is larger than either part alone.

CHECKING CONDITIONS

Two lines of a factory make parts with defect rates $p_1 = 0.12$ and $p_2 = 0.03$. Inspectors sample 100 parts from each line. Can the normal model be used for $\hat{p}_1 - \hat{p}_2$?

SOLUTION

Line 1: $n_1 p_1 = 12 \geq 10$ and $n_1(1-p_1) = 88 \geq 10$. Pass.

Line 2: $n_2 p_2 = 3 < 10$. Fail.

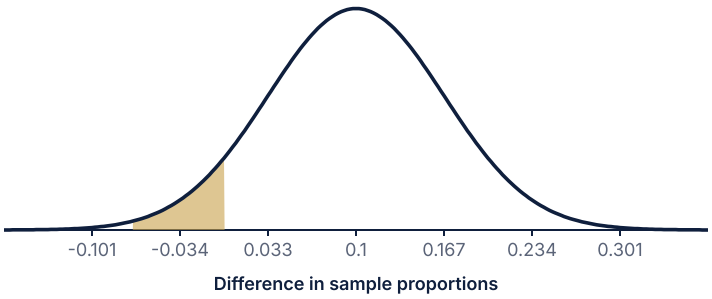
The Large Counts condition must hold for BOTH groups, so a normal model is not appropriate. Use a larger sample for line 2 (at least 334, since $334 \times 0.03 > 10$) or simulate.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

PROBABILITY OF REVERSING THE ORDER

Using the juniors and seniors above (p = 0.60, n = 100; p = 0.50, n = 120), what is the probability that a sample shows a LOWER proportion of juniors with jobs than of seniors?



Normal model with mean 0.10; shaded area is P(difference is less than 0)

SOLUTION

Conditions hold: $100(0.6) = 60$, $100(0.4) = 40$, $120(0.5) = 60$, $120(0.5) = 60$, all at least 10; independent random samples.

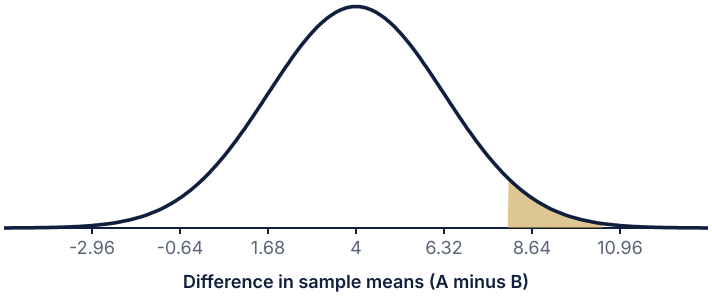
Model: normal with mean 0.10 and SD 0.0670.

We need $P(\hat{p}_J - \hat{p}_S < 0)$: $z = (0 - 0.10)/0.0670 = -1.49$, and $P(z < -1.49) \approx 0.0677$.

About 0.0677 of pairs of samples would show juniors with a lower job rate even though the true rate is higher.

DIFFERENCE OF SAMPLE MEANS

Math scores in class A have mean 72 and SD 10; class B has mean 68 and SD 12. Independent random samples of 40 students from A and 50 from B are taken (both populations large). Find the probability that the A sample mean exceeds the B sample mean by more than 8 points.



Sampling distribution of the difference in sample means

SOLUTION

Mean of the difference: $72 - 68 = 4$.

Variances: $10^2/40 = 2.500$ and $12^2/50 = 2.880$, total 5.380. SD $= \sqrt{5.380} = 2.319$.

Both samples are at least 30, so the CLT gives approximate normality. $z = (8 - 4)/2.319 = 1.72$.

$P(\bar{x}_A - \bar{x}_B > 8) = P(z > 1.72) \approx 0.0423$.

VARIANCES ADD, SDS DO NOT

Machine A fills cups with SD 3 mL and machine B with SD 4 mL, independently. For the difference (A minus B) in a single pair of cups, find the SD. Is it smaller than 4?

SOLUTION

Variances: $3^2 = 9$ and $4^2 = 16$. Adding: 25.

SD of the difference $= \sqrt{25} = 5$.

The SD of the difference is larger than 4. Subtracting does not cancel variability. The common wrong answer, $4 - 3 = 1$, would suggest the difference is less variable than either machine.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

IS A BIG DIFFERENCE SURPRISING?

A school tests a new reading app against old workbooks. Suppose the true proportion of students who reach the goal is 0.50 in both groups (no difference), with 80 students in each independent group. A sample shows a difference of 0.20 in favor of the app. How surprising is it?

SOLUTION

If $p_1 = p_2 = 0.50$, the mean difference is 0 and the SD is $\sqrt{0.25/80 + 0.25/80} = 0.0791$.

Conditions: counts $80(0.5) = 40$ in every cell, at least 10.

$z = 0.20/0.0791 = 2.53$. A difference this large in either direction has probability about 0.01141.

Such a difference would be rare by chance alone, so it is evidence that the groups really differ. This is the logic behind two-sample tests.

Write It Like a 5

Model Response

Prompt. In a large city, 40 percent of morning commuters use public transit and 30 percent of evening commuters do. Independent random samples of 150 morning and 200 evening commuters are taken. (a) Describe the sampling distribution of $\hat{p}_M - \hat{p}_E$. (b) Find the probability that the morning sample proportion exceeds the evening one by at least 0.15. **(a)** The samples are independent and random, and each is under 10 percent of its population. Large Counts: $150(0.40) = 60$, $150(0.60) = 90$, $200(0.30) = 60$, $200(0.70) = 140$, all at least 10. So the distribution is approximately normal with mean $0.40 - 0.30 = 0.10$ and SD $\sqrt{0.40(0.60)/150 + 0.30(0.70)/200} = 0.0515$. **(b)** $z = (0.15 - 0.10)/0.0515 = 0.97$, so $P(\hat{p}_M - \hat{p}_E \geq 0.15) \approx 0.1657$. About 0.1657 of such pairs of samples would show the morning transit rate at least 15 percentage points above the evening rate.

What the grader looks for

1 point: independence, random samples and the 10 percent condition. 1 point: Large Counts checked for all four counts with numbers. 1 point: mean as a difference of the parameters. 1 point: SD with variances added under a single root. 1 point: normal shape named and correct z. 1 point: probability interpreted in context.

Coach's Playbook

1. Name the two groups and the order of subtraction, and keep that order everywhere.
2. Verify independence, random selection, 10 percent for each group, and Large Counts or large samples for each group.
3. Mean: subtract the parameters. SD: add the variances of the two sampling distributions, then take the square root.
4. Standardize with the difference and its SD; shade and compute with the normal cdf.
5. Say what the probability means for pairs of samples in context.

Common traps

1. Subtracting standard deviations. Variances add, even when the statistics are subtracted. 2. Checking Large Counts for only one of the two groups. 3. Switching the order of subtraction partway through, producing a sign error in z. 4. Using a pooled proportion when the problem gives the true values of p_1 and p_2. Pooling is for tests when the null hypothesis says they are equal.

Mistake Clinic

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Student: 'The SD of $\bar{x}_1 - \bar{x}_2$ is $\sigma_1/\sqrt{n_1} - \sigma_2/\sqrt{n_2}$.' **Where it fails:** The spread of a difference is never smaller than both parts; subtracting SDs can even be zero. **Correct:** $\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$.

Spot the error

Student: 'p1 = 0.7 with n1 = 200 and p2 = 0.04 with n2 = 120. Large Counts is fine because n1 p1 = 140.' **Where it fails:** Group 2 has $n_2 p_2 = 4.8 < 10$. Every one of the four counts must pass. **Correct:** The normal model for the difference is not justified.

Spot the error

Student: 'The mean of the difference in sample means is the average of the two population means.' **Where it fails:** The mean of a difference is a difference of means. **Correct:** $\mu_1 - \mu_2$.

Real-World Connection

Clinical-style trials, A/B tests on websites, and quality comparisons between factories all compare two sample results. The sampling distribution of the difference tells analysts how large a gap can appear by chance alone.

Challenge Corner

Problem. Two independent groups have equal sizes n and equal $p = 0.5$. How large must n be so that the SD of $\hat{p}_1 - \hat{p}_2$ is at most 0.03? **Solution.** $\sqrt{0.25/n + 0.25/n} = \sqrt{0.5/n} \le 0.03$, so $n \ge 0.5/0.0009 = 555.6$, which rounds up to 556 per group. Compare with a single proportion: $\sqrt{0.25/n} \le 0.03$ needs only 278. Two groups need twice the sample size per group to get the same precision as a single sample.

Practice

PRACTICE **Sampling Distributions for Differences**

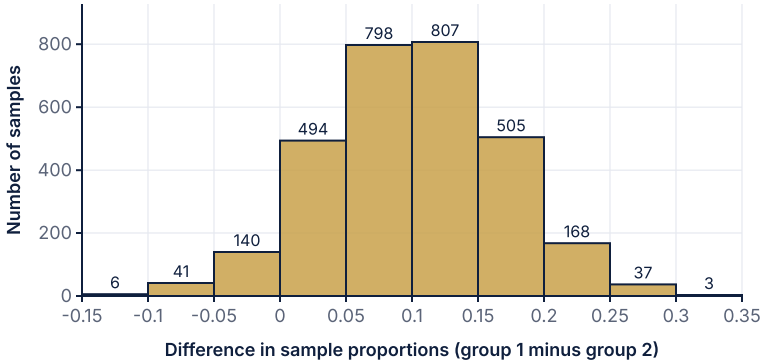
Show your work. A calculator is allowed. Justify conclusions in context.

1. For independent samples, $\hat{p}_1$ has mean 0.55 and $\hat{p}_2$ has mean 0.40. The mean of $\hat{p}_1 - \hat{p}_2$ is
   (A) 0.15 **(B)** 0.95 **(C)** 0.22 **(D)** 0.475
2. Independent sampling distributions of two statistics have SDs 6 and 8. The SD of their difference is
   (A) 2 **(B)** 14 **(C)** 10 **(D)** 100
3. Which condition is required for a normal model for $\hat{p}_1 - \hat{p}_2$?
   (A) Large Counts in either group **(B)** Large Counts in both groups
   (C) The sample sizes must be equal **(D)** Both populations must be normal
4. Group 1: $p_1 = 0.5$, $n_1 = 50$. Group 2: $p_2 = 0.1$, $n_2 = 60$. Large Counts is
   (A) met for both groups **(B)** not met for group 1 only **(C)** not met for group 2 only
   (D) not met for either group
5. For $p_1 = 0.30$, $n_1 = 100$, $p_2 = 0.20$, $n_2 = 100$, the SD of $\hat{p}_1 - \hat{p}_2$ is closest to
   (A) 0.022 **(B)** 0.0037 **(C)** 0.100 **(D)** 0.061
6. For independent samples with $\mu_1 = 50$, $\sigma_1 = 8$, $n_1 = 64$ and $\mu_2 = 45$, $\sigma_2 = 6$, $n_2 = 36$, the SD of $\bar{x}_1 - \bar{x}_2$ is
   (A) $\sqrt{2}$ **(B)** $\sqrt{14}$ **(C)** 2 **(D)** 1.5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. The histogram shows simulated differences in sample proportions (group 1 minus group 2). Which is the best estimate of the true difference $p_1 - p_2$?



3000 simulated differences, p1 = 0.60 (n = 100), p2 = 0.50 (n = 120)

(A) About 0.25 **(B)** About 0.10 **(C)** About -0.10

(D) About 0

8. Doubling BOTH sample sizes in a difference of proportions (same p values) changes the SD of the difference by a factor of about

(A) 1.41 **(B)** 0.50 **(C)** 2 **(D)** 0.71

9. Why can the two sample means NOT be treated as independent if the same students take both tests?

(A) The sample sizes are equal

(B) Sample means are never independent

(C) The samples are the same people, so the means are related and the formulas do not apply

(D) The standard deviations are equal

10. In a school, 70 percent of freshmen and 55 percent of sophomores own a laptop. Independent random samples of 90 freshmen and 110 sophomores are taken (both groups are large).

(a) Find the mean and standard deviation of $\hat{p}_F - \hat{p}_S$.

(b) Check the conditions for a normal model.

11. Using the freshman and sophomore laptop setting (70 percent with n = 90; 55 percent with n = 110), find the probability that the sample proportion of freshmen exceeds that of sophomores by more than 0.20.

12. Bakery A bakes loaves whose weights have mean 450 g and SD 12 g; bakery B has mean 440 g and SD 15 g. Independent random samples of 36 loaves from A and 49 from B are weighed (very large populations). Find the mean and SD of $\bar{x}_A - \bar{x}_B$ and the probability that $\bar{x}_A - \bar{x}_B$ is negative.

13. Explain in words why the standard deviation of a difference of independent sample means is computed by adding variances rather than subtracting. Use the numbers SD 3 and SD 4 as an example.

14. Two groups have $p_1 = 0.40$ and $p_2 = 0.40$ with 50 students each. Find the probability that the sample proportions differ by more than 0.15 in either direction. What does this suggest about drawing conclusions from small samples?

15. A student claims: 'The sampling distribution of $\bar{x}_1 - \bar{x}_2$ is normal because the CLT says everything is normal.' The samples are n1 = 8 and n2 = 10 from strongly skewed populations. Respond.

T Unit 6 Test

Check your command of parameters, statistics, and sampling distributions for proportions, means, and differences.

Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Part A: Multiple Choice

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

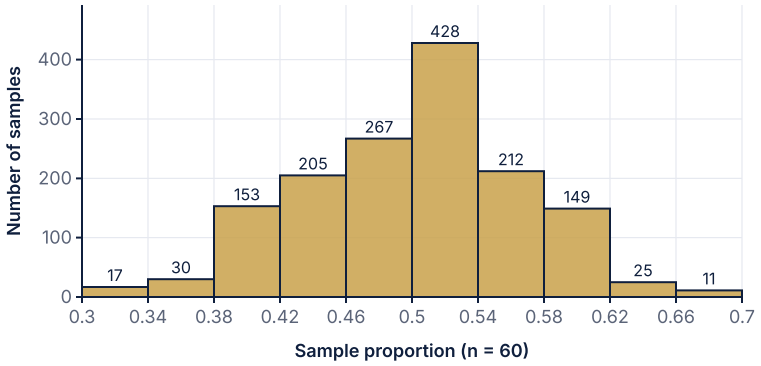
UNIT TEST

Unit 6 Test

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. A pollster reports that 58 percent of a random sample of 400 registered voters favor a measure. The 58 percent is
   (A) a parameter, because it describes voters
   (B) a statistic, because it describes a sample
   (C) the mean of the sampling distribution
   (D) the standard deviation of the population

2. The histogram shows 1500 simulated sample proportions from random samples of 60, where the true p = 0.50. Which description is best?



1500 simulated samples

   (A) Biased, centered well above 0.50
   (B) Variability decreases as more dots are added
   (C) Centered at 0.50 and strongly skewed right
   (D) Approximately unbiased and approximately normal

3. A random sample of 225 is taken from a very large population with p = 0.20. The standard deviation of $\hat{p}$ is
   (A) 0.0267 **(B)** 0.0008 **(C)** 0.0018 **(D)** 0.2667

4. For which situation is a normal model for $\hat{p}$ appropriate? (Assume random samples from very large populations.)
   (A) n = 40, p = 0.15 **(B)** n = 200, p = 0.04 **(C)** n = 100, p = 0.30
   (D) n = 50, p = 0.10

5. In a large community, 35 percent of residents recycle. In a random sample of 150, the probability that fewer than 30 percent recycle is approximately
   (A) 0.05 **(B)** 0.5 **(C)** 0.9 **(D)** 0.0996

6. A population has $\mu = 80$ and $\sigma = 24$. The SD of the sampling distribution of $\bar{x}$ for n = 64 is
   (A) 24 **(B)** 3 **(C)** 0.375 **(D)** 12

7. The Central Limit Theorem is most useful because it tells us that
   (A) the sampling distribution of $\bar{x}$ is approximately normal for large n regardless of the population shape
   (B) data from skewed populations become normal in large samples
   (C) the sample SD equals the population SD
   (D) sample means are always exactly equal to μ

8. Lengths of a part are normal with mean 100 mm and SD 20 mm. For a random sample of 25 parts, P($\bar{x}$ is greater than 106) is approximately
   (A) 0.618 **(B)** 0.382 **(C)** 0.067 **(D)** 0.933

9. A researcher takes a random sample of 100, then another of 400, from the same large population. Compared with n = 100, the sampling distribution of $\bar{x}$ for n = 400 has
   (A) one fourth the standard deviation, same mean
   (B) half the standard deviation, same mean
   (C) the same standard deviation, larger mean
   (D) twice the standard deviation, same mean

10. Independent samples: 100 from population 1 (p = 0.40) and 150 from population 2 (p = 0.30). The standard deviation of $\hat{p}_1 - \hat{p}_2$ is approximately
    (A) 0.095 **(B)** 0.0038 **(C)** 0.100 **(D)** 0.0616

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

11. Independent samples: $\bar{x}_1$ from n = 36, σ = 15; $\bar{x}_2$ from n = 25, σ = 10. The SD of $\bar{x}_1 - \bar{x}_2$ is approximately
    (A) 3.20 **(B)** 0.5 **(C)** 4.5 **(D)** 10.25
12. A radio call-in poll gets 5000 responses about a local issue. A random phone survey of 400 residents gets a very different result. Which statement is correct?
    (A) The call-in poll is more accurate because n is larger
    (B) The phone survey must be biased because it is smaller
    (C) The call-in poll likely has bias that its large size cannot fix
    (D) The two must agree exactly if both are random
13. [6 points] A school has 2400 students, and 35 percent of all students take the bus. A random sample of 150 students is selected.
    (a) [2 points] Verify the conditions needed to use a normal model for the sampling distribution of $\hat{p}$.
    (b) [2 points] Give the mean and standard deviation of the sampling distribution.
    (c) [2 points] Find the probability that fewer than 30 percent of the sampled students take the bus, and interpret it.
14. [10 points] Commute times in City A are right-skewed with mean 60 minutes and SD 12 minutes. In City B they are right-skewed with mean 57 minutes and SD 9 minutes. Independent random samples of 40 commuters from City A and 50 from City B are taken (both cities are large).
    (a) [2 points] Explain why a normal model is appropriate for the sampling distribution of the A sample mean even though the population is skewed.
    (b) [2 points] Find the mean and SD of the A sample mean, and the probability that it exceeds 62 minutes.
    (c) [3 points] Find the mean and SD of $\bar{x}_A - \bar{x}_B$ and the probability that this difference exceeds 5 minutes.
    (d) [2 points] A student says, 'Because City A's mean is only 3 minutes higher, the sample from A should be higher about half the time.' Evaluate.
    (e) [1 point] Explain how the variability of $\bar{x}_A$ would change if 160 commuters were sampled from City A instead of 40.

Answers Sampling Distributions: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

6.1 What Is a Sampling Distribution?

PRACTICE · WHAT IS A SAMPLING DISTRIBUTION?

1. B. The proportion of ALL registered voters is a population value, a parameter. The other three describe samples, so they are statistics.
2. D. Different random samples give different statistics: that is sampling variability. Bias would be a consistent miss in one direction, not a day-to-day difference.
3. C. Each value is the proportion computed from ONE whole sample of 25. It is a statistic, not an individual or the parameter, and not an average of the samples.
4. A. Unbiased means the sampling distribution is centered at the parameter. Equality in any single sample is not required, and a small SD is about variability, not bias.
5. D. More data from the same random method shrinks the spread; the center stays at p. Bias is caused by the method, so it is not removed.
6. B. The center 0.62 is far from 0.50 (biased), and little scatter means low variability. Students pick low bias because the scatter is small.
7. A. The sample maximum cannot exceed the population maximum, so it averages below it. The other three are the standard unbiased estimators.
8. C. Variability depends on the sample size, not the population size, provided the population is at least 10 times the sample (here 2000 is 20 times 100). So the two are about equal.
9. B. One trial is a group of 20 digits; digits 0, 1, 2 give probability 0.3 of success. The statistic is successes divided by 20. A single digit is not a sample proportion.
10. (a) 62 percent is from a sample of 150: statistic, $\hat{p} = 0.62$. The mean household size 2.7 describes all households: parameter, $\mu = 2.7$.
    (b) A different random sample of 150 would give a different $\hat{p}$; this sampling variability means $\hat{p}$ is an estimate that typically misses p by some amount.
11. (a) Each value is the sample proportion $\hat{p}$ from one random sample of 36 (a statistic).
    (b) Center about 0.251 (near 0.25); shape roughly symmetric, mound-shaped; spread: standard deviation about 0.072, so typical values are within 0.07 of 0.25.
    (c) No. The center 0.251 is essentially the true p = 0.25, so $\hat{p}$ appears unbiased.
12. (a) Y has more bias (center 108 is 8 above 100; X's center 100.2 is almost on target). X has more variability (SD 9 versus 2).
    (b) Neither is ideal. A single X estimate is often off by about 9 because of scatter; a single Y estimate is consistently about 8 too high. The best method has both low bias and low SD.
13. One trial: take 25 random digits from the table in which 00 to 11 (two-digit blocks 00 to 99) count as left-handed, so 12 of 100 blocks give 12 percent. Count the left-handed blocks and divide by 25 to get one $\hat{p}$. Repeat for many trials, such as 200 to 1000. Graph the sample proportions in a histogram or dotplot and describe center, shape and spread.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. The 9 samples (2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6) have means 2 (prob 1/9), 3 (prob 2/9), 4 (prob 3/9), 5 (prob 2/9), 6 (prob 1/9). The mean of the sampling distribution is (2 + 3 + 4 + 3 + 4 + 5 + 4 + 5 + 6)/9 = 36/9 = 4, and the population mean is (2 + 4 + 6)/3 = 4. They are equal, so $\bar{x}$ is unbiased for μ.
15. The random phone survey is more trustworthy. The online poll uses voluntary response, which introduces bias (people with strong opinions respond), and bias is a problem with the method, not the size. The 4000 responses make the poll consistently wrong with little scatter around the wrong value. The random sample of 300 is unbiased, with larger variability that is acceptable and can be measured.

6.2 Sampling Distribution of a Sample Proportion

PRACTICE · SAMPLING DISTRIBUTION OF A SAMPLE PROPORTION

1. A. The mean of the sampling distribution of $\hat{p}$ is the population proportion $p = 0.35$. The value 0.0477 is the standard deviation, not the center.
2. C. $\sqrt{0.20(0.80)/100} = \sqrt{0.0016} = 0.04$. The value 0.0016 is the variance: it forgets the square root.
3. D. Large Counts needs np and n(1 - p) both at least 10. For p = 0.20, n = 60: np = 12 and n(1 - p) = 48, so it passes. The others fail: np = 5, 7.5, and n(1 - p) = 8 respectively.
4. B. The sample must be at most 10 percent of 150, which is 15. Choosing 30 or 75 breaks the condition.
5. C. SD is proportional to $1/\sqrt{n}$, so multiplying n by 9 multiplies the SD by 1/3. The tempting 1/9 treats SD as proportional to 1/n.
6. B. $\sigma = \sqrt{0.25/100} = 0.05$ and $z = (0.45 - 0.50)/0.05 = -1$, so the probability is about 0.159. The value 0.841 is the area above.
7. A. The shaded region lies between 0.40 and 0.50 on the horizontal axis, so it is the probability that the sample proportion falls in that interval. It is not the area above the mean.
8. D. $\sigma = \sqrt{0.30(0.70)/400} = 0.0229$, and 1.96 SDs is about 0.045, so the interval is 0.30 plus or minus 0.045, or 0.255 to 0.345. The interval 0.250 to 0.350 is plus or minus 0.05, a round-number guess.
9. C. The whole fraction p(1 - p)/n belongs under the square root. Dividing the root by n makes the SD far too small.
10. (a) Random: the sample is a random sample. 10 percent: 180 is less than 10 percent of 4000 (400). Large Counts: $np = 180(0.55) = 99 \geq 10$ and $n(1-p) = 180(0.45) = 81 \geq 10$. So $\hat{p}$ is approximately normal.
    (b) $\mu_{\hat{p}} = 0.55$ and $\sigma_{\hat{p}} = \sqrt{0.55(0.45)/180} = 0.0371$.
11. Conditions: random sample, 150 is under 10 percent of a very large population, $np = 96 \geq 10$ and $n(1-p) = 54 \geq 10$. $\sigma_{\hat{p}} = \sqrt{0.64(0.36)/150} = 0.0392$. $z = (0.70 - 0.64)/0.0392 = 1.53$, so $P(\hat{p} > 0.70) \approx 0.0629$. About 0.0629 of random samples of 150 would have more than 70 percent in the group.
12. No. $np = 90(0.06) = 5.4 < 10$, so the Large Counts condition fails, even though $n(1-p) = 84.6$. Because p is small and n is moderate, the sampling distribution is skewed right (the proportion cannot go below 0), so a normal model would give inaccurate probabilities. Use a binomial model or simulation.
13. (a) Conditions: random, $np = 25$ and $n(1-p) = 75$ both at least 10, sample under 10 percent. $\sigma_{\hat{p}} = \sqrt{0.25(0.75)/100} = 0.0433$. $z = \pm(0.05/0.0433) = \pm 1.15$. $P(-1.15 < z < 1.15) \approx 0.7518$.
    (b) Larger. With n = 400 the SD is half as large, so more samples fall within 0.05 of 0.25.
14. We need $\sqrt{0.25/n} \leq 0.015$, so $n \geq 0.25/0.015^2 = 1111.1$, which rounds up to 1112. If p = 0.2, $n \geq 0.16/0.015^2 = 711.1$, so 712. A p farther from 0.5 means less spread, so a smaller sample suffices.
15. 10 percent of 500 is 50, and 60 exceeds 50, so the 10 percent condition fails. Sampling without replacement from a small population makes the true standard deviation smaller than the formula: it is multiplied by about $\sqrt{(N-n)/(N-1)} = 0.939$. The formula gives 0.0632, while the actual SD is about 0.0594. The formula overstates the spread, and the observations are not independent.

6.3 Sampling Distribution of a Sample Mean and the Central Limit Theorem

PRACTICE · SAMPLING DISTRIBUTION OF A SAMPLE MEAN AND THE CENTRAL LIMIT THEOREM

1. A. $\sigma_{\bar{x}} = 20/\sqrt{100} = 2$ with mean 50. The choice 0.2 divides by 100 instead of its square root.
2. B. The CLT applies to the sample mean for any population shape with finite SD once n is large. It does not require a normal population.
3. C. $\sigma/\sqrt{n}$ is halved when n is multiplied by 4. Doubling n only divides the SD by about 1.41.
4. B. At n = 5 the histogram is clearly skewed right, so the CLT has not yet taken hold. The mere fact that n exceeds 1 does not make it normal.
5. D. Strong skew requires a large sample, and outliers are a warning sign. n = 100 gives the strongest justification; n = 25 with outliers does not.
6. A. $\sigma_{\bar{x}} = 3/3 = 1$ and $z = 1$, so the probability above is about 0.159. Using 3 as the SD gives z = 0.33 and 0.369.
7. D. Dividing by sigma instead of sigma over root n gives a z-score that is root n times too small in magnitude (4 times too small here).
8. C. The formula needs only n at most 10 percent of the population: 10 percent of 400 is 40, so n = 30 qualifies and n = 100 does not. It depends on n, and equals sigma only when n = 1.
9. A. The center is 24 and the SD is about 0.99, so the region from 20 to 22.5 lies 1.5 SDs or more below the mean (z from about -4 to -1.5). The area is small, about 0.065.
10. (a) $\mu_{\bar{x}} = 500$ and $\sigma_{\bar{x}} = 90/\sqrt{36} = 15.0$.
    (b) If the population is normal, the sampling distribution is exactly normal. If it is skewed, the CLT makes it approximately normal only if n is large enough; n = 36 is usually enough for moderate skew, with more needed for extreme skew.
11. Conditions: random sample; the population is normal, so $\bar{x}$ is normal; 25 LEDs is under 10 percent of all LEDs. $\sigma_{\bar{x}} = 5000/\sqrt{25} = 1000$. $z = (38000 - 40000)/1000 = -2$. $P(\bar{x} < 38000) = P(z < -2) \approx 0.0228$. About 0.0228 of random samples of 25 LEDs would average under 38000 hours.
12. Conditions: random; 60 is under 10 percent of all player-games; n = 60 is at least 30, so by the CLT $\bar{x}$ is approximately normal despite the skewed population. $\sigma_{\bar{x}} = 9/\sqrt{60} = 1.1619$ and $z = (16 - 14)/1.1619 = 1.72$. $P(\bar{x} > 16) \approx 0.0426$. About 0.0426 of such samples have a mean above 16 minutes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. Averaging cancels some random highs and lows, and the more values averaged the more cancellation occurs. The SD of the sample mean is $\sigma/\sqrt{n}$, so the ratio of the SD for n = 100 to n = 10 is $\sqrt{10/100} = 0.316$. The mean of 100 values has about 32 percent of the SD of the mean of 10.
14. We need $4/\sqrt{n} \le 0.5$, so $\sqrt{n} \ge 8$ and $n \ge 64$. A sample of at least 64 bottles is needed. (Check: 4/8 = 0.5.)
15. Error 1: The CLT concerns the sampling distribution of the sample mean, not the individual observations, which keep their own shape. Error 2: 15 is not a general threshold; the usual guideline is n of at least 30, and skewed or outlier-prone populations need more. Correct: if the population is skewed, 15 observations may not be enough for a normal sampling distribution of $\bar{x}$; check the sample for strong skew or outliers.

6.4 Sampling Distributions for Differences

PRACTICE · SAMPLING DISTRIBUTIONS FOR DIFFERENCES

1. A. Means subtract: 0.55 - 0.40 = 0.15. The value 0.95 adds them instead of subtracting.
2. C. Add the variances: 36 + 64 = 100, and the square root is 10. Adding the SDs gives 14 and subtracting gives 2, both common errors.
3. B. Large Counts must hold for both groups, along with independent random samples and the 10 percent condition. Equal sample sizes are not required, and the populations need not be normal.
4. C. Group 1: 25 and 25, both fine. Group 2: 6 and 54, and 6 is less than 10 fails, so the condition fails for group 2 only.
5. D. $\sqrt{0.3(0.7)/100 + 0.2(0.8)/100} = \sqrt{0.0037} = 0.0608$. The value 0.0037 is the variance (no square root), and 0.022 comes from subtracting the variances.
6. A. $8^2/64 = 1$ and $6^2/36 = 1$, so the variances add to 2 and the SD is $\sqrt{2} \approx 1.41$. The choice 2 forgets the square root.
7. B. The histogram is centered near 0.10, and the center of the sampling distribution of the difference equals the true difference $p_1 - p_2$.
8. D. Doubling both sample sizes halves both variances, so the SD is multiplied by $1/\sqrt{2} \approx 0.71$, not by 1/2.
9. C. If the same students are measured twice, the two sets of results are related, so the independence needed for adding variances fails. This is a paired situation.
10. (a) Mean: $0.70 - 0.55 = 0.15$. SD: $\sqrt{0.70(0.30)/90 + 0.55(0.45)/110} = 0.0677$.
    (b) Independent random samples; each group is a small part of its population. Large Counts: $90(0.70) = 63$, $90(0.30) = 27$, $110(0.55) = 60.5$, $110(0.45) = 49.5$, all at least 10. So the difference is approximately normal.
11. The mean is 0.15 and the SD is 0.0677. $z = (0.20 - 0.15)/0.0677 = 0.74$, so $P(\hat{p}_F - \hat{p}_S > 0.20) \approx 0.2301$. About 0.2301 of pairs of samples would show freshmen more than 20 percentage points higher.
12. Mean: $450 - 440 = 10$. Variances: $12^2/36 = 4$ and $15^2/49 = 4.592$, total 8.592. SD $= 2.9312$. Both samples exceed 30 (or the populations are assumed approximately normal), so the CLT applies. $z = (0 - 10)/2.9312 = -3.41$, so $P(\bar{x}_A - \bar{x}_B < 0) \approx 0.0003$. About 0.0003 of sample pairs would show bakery A lighter than bakery B.
13. Each sample mean has its own random error. A difference combines both errors, and random errors do not cancel exactly: sometimes they add and sometimes they offset. Variances of independent quantities add, so for SDs 3 and 4 the difference has variance 9 + 16 = 25 and SD 5, which is larger than either SD. Subtracting (4 - 3 = 1) would claim the difference is less variable, which is wrong.
14. SD $= \sqrt{0.4(0.6)/50 + 0.4(0.6)/50} = 0.0980$. Large Counts: 20 and 30 in each group, fine. $z = 0.15/0.0980 = 1.53$, so $P(|\hat{p}_1 - \hat{p}_2| > 0.15) \approx 0.1258$. Even when the true proportions are identical, about 0.1258 of pairs of samples of 50 differ by more than 0.15, so small samples can show apparent differences that are just chance.
15. The statement is wrong. The CLT needs large samples. With n1 = 8 and n2 = 10 from strongly skewed populations, the sampling distribution of each mean is still skewed, so the distribution of the difference is not guaranteed to be approximately normal. A normal model should only be used if the populations are known to be nearly normal or both samples are large (about 30 or more).

T Unit 6 Test

UNIT TEST · UNIT 6 TEST

1. B. It comes from a sample of 400 voters, so it is a statistic ($\hat{p}$). A parameter would describe all registered voters.
2. D. The dots center near 0.50 (mean 0.499) and form a symmetric mound with SD about 0.067, so $\hat{p}$ looks unbiased and approximately normal. Adding dots does not change the true spread.
3. A. $\sqrt{0.20(0.80)/225} = \sqrt{0.000711} = 0.0267$. The value 0.0008 is the variance.
4. C. Large Counts needs np and n(1 - p) at least 10. Only n = 100, p = 0.30 passes (30 and 70). The others have np = 6, 8, and 5.
5. D. $\sigma = 0.0389$ and $z = (0.30 - 0.35)/0.0389 = -1.28$, so the probability is about 0.0996. The tempting 0.5 treats 0.30 as if it were the mean.
6. B. $24/\sqrt{64} = 24/8 = 3$. The answer 0.375 divides by 64 instead of its square root.
7. A. The CLT is about the sampling distribution of the sample mean being approximately normal for large n. It does not change the shape of the data themselves.
8. C. $\sigma_{\bar{x}} = 20/5 = 4$ and $z = (106 - 100)/4 = 1.5$, so $P(\bar{x} > 106) = 0.067$. Using SD 20 gives z = 0.3 and 0.382.
9. B. Quadrupling n halves the SD ($1/\sqrt{4}$). The mean stays μ. One fourth would occur only if SD were proportional to $1/n$.
10. D. $\sqrt{0.40(0.60)/100 + 0.30(0.70)/150} = \sqrt{0.0024 + 0.0014} = 0.0616$. The value 0.0038 is the variance (no square root), and 0.095 adds the two SDs.
11. A. $\sqrt{15^2/36 + 10^2/25} = \sqrt{6.25 + 4} = 3.20$. The value 4.5 adds the SDs, 0.5 subtracts them, and 10.25 is the variance.
12. C. A voluntary response call-in poll is biased: respondents are self-selected. Bias comes from the method, so a bigger n does not help, whereas random sampling avoids it.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. Scoring (6 points). (a) 2 points: random sample stated; 150 is less than 10 percent of 2400 (240); $np = 150(0.35) = 52.5 \geq 10$ and $n(1-p) = 97.5 \geq 10$ (1 point for 10 percent and random, 1 for both Large Counts with numbers). (b) 2 points: mean 0.35 (1); SD $\sqrt{0.35(0.65)/150} = 0.0389$ (1). (c) 2 points: $z = (0.30 - 0.35)/0.0389 = -1.28$ and probability about 0.0996 (1); interpretation in context (1).

Model response. (a) The sample is random. The sample of 150 is less than 10 percent of 2400 students (which is 240). Also $np = 52.5$ and $n(1-p) = 97.5$ are both at least 10. So $\hat{p}$ is approximately normal. (b) $\mu_{\hat{p}} = 0.35$, $\sigma_{\hat{p}} = 0.0389$. (c) $z = -1.28$, $P(\hat{p} < 0.30) \approx 0.0996$. About 0.0996 of random samples of 150 students would have fewer than 30 percent riding the bus.

14. Scoring (10 points). (a) 2: samples random and under 10 percent of the city; n = 40 is at least 30 so the CLT makes the sampling distribution approximately normal even though the population is skewed. (b) 2: mean 60, SD $12/\sqrt{40} = 1.8974$ (1); $z = (62 - 60)/1.8974 = 1.05$, probability about 0.1459 (1). (c) 3: mean $60 - 57 = 3$ (1); SD $\sqrt{12^2/40 + 9^2/50} = \sqrt{5.220} = 2.2847$ (1); $z = (5 - 3)/2.2847 = 0.88$, probability about 0.1907 (1). (d) 2: the student is wrong because the sample means are random: the probability that $\bar{x}_A > \bar{x}_B$ is $P(z > -3/2.2847) = 0.9054$, not 0.5, because the true difference is 3 and the SD of the difference is 2.2847 (1 for the correct reasoning, 1 for the value). (e) 1: with n = 160 the SD becomes $12/\sqrt{160} = 0.9487$, half as large, since quadrupling n halves the SD.

Model response. (a) Both samples are random and under 10 percent of their cities. Although commute times are skewed, n = 40 is at least 30, so by the CLT $\bar{x}_A$ is approximately normal. (b) $\mu = 60$, $\sigma = 12/\sqrt{40} = 1.8974$; $P(\bar{x}_A > 62) = P(z > 1.05) \approx 0.1459$. (c) The mean is 3 minutes and the SD is $\sqrt{5.220} = 2.2847$; both n are at least 30 so the difference is approximately normal; $P(\text{diff} > 5) = P(z > 0.88) \approx 0.1907$. (d) The student is incorrect. The true mean difference is 3 and the SD is 2.2847, so $P(\bar{x}_A > \bar{x}_B) = P(z > -1.31) \approx 0.9054$, higher than 0.5 but not certain. (e) Quadrupling the sample size to 160 halves the SD from 1.8974 to 0.9487.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 7: Inference for Proportions

Estimate and test proportions with confidence intervals and significance tests. 6 lessons, 104 practice and test items, all with full explanations.

In this unit · Exam Unit 3

7.1 The Logic of Confidence Intervals
7.2 One-Proportion z-Intervals, Margin of Error, and Sample Size
7.3 The Logic of Significance Tests and P-Values
7.4 One-Proportion z-Tests
7.5 Errors, Significance, and Power
7.6 Two-Proportion z-Intervals and Tests
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

7.1 The Logic of Confidence Intervals

How a single sample statistic becomes an honest range for a population parameter, and what the words 'confident' and '95 percent' really mean.

Why This Matters

Every poll you have ever seen makes the same quiet bet: that a small random sample can tell you something trustworthy about a huge population. A school newspaper surveys 400 students and reports that 45 percent bike to campus. Does that mean exactly 45 percent of the whole school bikes? Of course not. The sample could easily have landed a little high or a little low. A **confidence interval** is the tool that turns a single, shaky number into an honest statement about how far off we might be.

Students go wrong in two predictable ways. First, they treat the interval as describing individual students or as a range of sample results, when it is a statement about one fixed population parameter. Second, they confuse the *interval* (what we found) with the *confidence level* (how reliable the method is). This lesson sits at the start of Exam Unit 3 (Inference for Categorical Data: Proportions). On the AP exam, free-response questions routinely ask you to interpret an interval or a confidence level in context, and multiple-choice items test the common misinterpretations directly.

Core Ideas

A **parameter** is a fixed number describing a population, such as the true proportion p of all students who bike. A **statistic** is a number computed from a sample, such as the sample proportion $\hat{p}$. When we use a statistic to guess a parameter, the statistic is called a **point estimate**. The point estimate for a population proportion is $\hat{p} = \dfrac{\text{number of successes}}{n}$.

A point estimate is almost never exactly right, because a different random sample would give a different $\hat{p}$. This sample-to-sample change is **sampling variability**. We measure how much a typical $\hat{p}$ strays from p with the **standard error**, which estimates the standard deviation of the sampling distribution of $\hat{p}$. So instead of offering a lone number, we report a plausible range built around it.

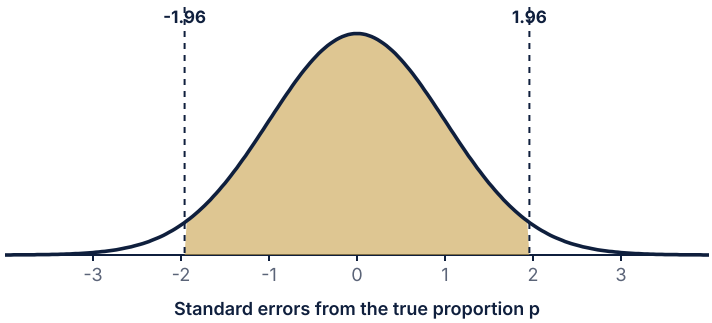
Confidence Interval Structure

Confidence interval = point estimate $\pm$ margin of error.
Margin of error = (critical value) $\times$ (standard error).
For a proportion: $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$, where z^* is the critical value for the chosen confidence level.
The interval is written as (lower, upper), and its midpoint is always $\hat{p}$.

The **critical value** z^* is the number of standard errors we must go on each side of $\hat{p}$ to capture the middle C percent of a normal curve. Common values are $z^* = 1.645$ for 90 percent, 1.960 for 95 percent, and 2.576 for 99 percent. The picture below shows the 95 percent case: the middle 95 percent of the sampling distribution lies within about 1.96 standard errors of the center.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



The shaded middle 95 percent of the sampling distribution of the sample proportion.

Here is the idea that makes the method work. About 95 percent of all random samples produce a $\hat{p}$ within 1.96 standard errors of the true p. If $\hat{p}$ is within 1.96 standard errors of p, then p is within 1.96 standard errors of $\hat{p}$. So an interval stretching 1.96 standard errors on both sides of $\hat{p}$ will capture p in about 95 percent of samples.

Two Different Interpretations

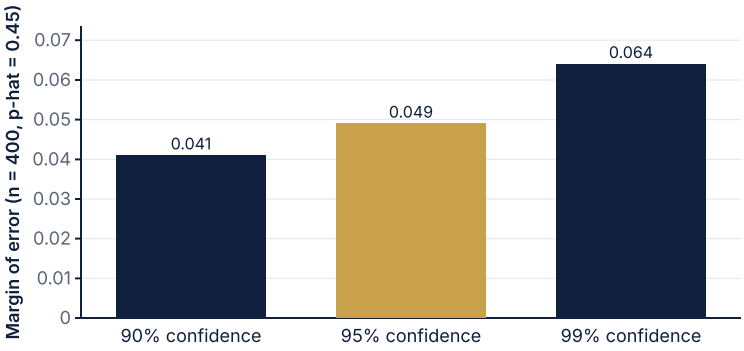
Interpreting the interval: We are C% confident that the interval from (lower) to (upper) captures the true value of the parameter, stated in context (for example, the true proportion of all students at this school who bike to campus).

Interpreting the confidence level: If we took many random samples of the same size and built an interval from each, about C% of those intervals would capture the true parameter.

The level describes the *method's long-run success rate*. It is not the probability that one finished interval is correct.

Why can we not say there is a 95 percent probability that p lies in a particular interval? Because p is a fixed number, not a random variable. Once the interval (0.401, 0.499) is computed, either p is in it or it is not. The randomness belongs to the sampling process, which is why we use the word *confident* rather than *probability* for a single finished interval.

What controls the width? The margin of error is $z^* \times$ SE, so it shrinks when the critical value shrinks (lower confidence) or when the standard error shrinks (larger n). Because n sits under a square root, you must multiply the sample size by 4 to cut the margin of error in half.



Higher confidence costs a wider interval from the same data.

Change you make	Effect on margin of error	Why
Increase confidence level	Gets larger	Larger z^*
Increase sample size n	Gets smaller	Smaller standard error
$\hat{p}$ closer to 0.5	Gets larger	$\hat{p}(1-\hat{p})$ is biggest at 0.5
Use a biased or nonrandom sample	Margin of error cannot fix it	Bias shifts the center, not the spread

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

FINDING A POINT ESTIMATE

A random sample of 150 students at a school is asked whether they own a reusable water bottle. Sixty-three say yes. Give a point estimate for the proportion of all students at the school who own one.

SOLUTION

The parameter is p, the proportion of all students at the school who own a reusable bottle.

The point estimate is $\hat{p} = 63/150 = 0.42$.

Our best single guess is that 42 percent of all students own one, though the true value will probably differ somewhat because of sampling variability.

INTERPRETING AN INTERVAL

A 95 percent confidence interval for the proportion of all students who bike to campus is (0.401, 0.499), based on a random sample of 400 students. Interpret the interval.

SOLUTION

We are 95 percent confident that the interval from 0.401 to 0.499 captures the true proportion of all students at this school who bike to campus.

Note the three required parts: the confidence level, the interval values, and the parameter described in context (a population proportion, not a sample result).

INTERPRETING THE CONFIDENCE LEVEL

An analyst builds 200 independent 90 percent confidence intervals for a proportion, each from a new random sample of the same size. About how many intervals should capture the true proportion, and what does 90 percent mean here?

SOLUTION

We expect about $0.90 \times 200 = 180$ of the intervals to capture the true proportion.

The 90 percent describes the *method*: in the long run, 90 percent of intervals produced this way contain the parameter. It does not say that any one interval has a 90 percent chance, and the count will vary a bit from 180 just like any random outcome.

READING THE MARGIN OF ERROR

A poll reports the interval (0.52, 0.60) for the proportion of voters in a town who favor a new park. Find the point estimate and the margin of error.

Lower	Upper	Midpoint	Margin of error
0.52	0.60	0.56	0.04

SOLUTION

The point estimate is the midpoint: $\hat{p} = (0.52 + 0.60)/2 = 0.56$.

The margin of error is half the width: $(0.60 - 0.52)/2 = 0.04$.

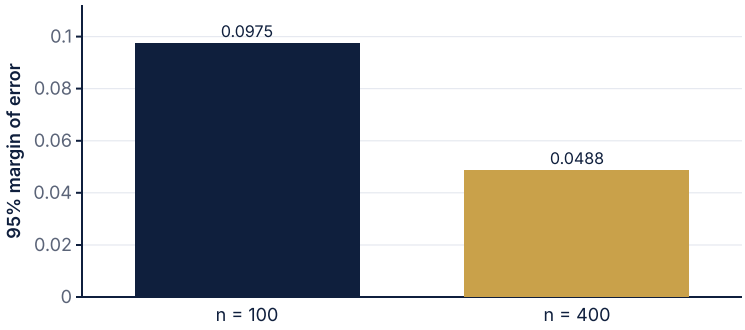
So the poll estimates 56 percent support, give or take 4 percentage points at the stated confidence level.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EFFECT OF SAMPLE SIZE

In the bike survey, $\hat{p} = 0.45$. Compare the 95 percent margin of error for $n = 100$ and $n = 400$.



Four times the sample, half the margin.

SOLUTION

For $n = 100$: $\text{SE} = \sqrt{0.45(0.55)/100} \approx 0.0498$, so $\text{ME} = 1.96(0.0498) \approx 0.0975$.

For $n = 400$: $\text{SE} = \sqrt{0.45(0.55)/400} \approx 0.0249$, so $\text{ME} = 1.96(0.0249) \approx 0.0488$.

Quadrupling the sample size cut the margin of error in half, from about 9.8 to about 4.9 percentage points.

CHANGING THE CONFIDENCE LEVEL

A 90 percent confidence interval for a proportion is (0.62, 0.70). Find the 95 percent interval from the same data.

SOLUTION

The center is $\hat{p} = 0.66$ and the 90 percent margin of error is 0.04.

Since $\text{ME} = z^* \cdot \text{SE}$, we get $\text{SE} = 0.04/1.645 \approx 0.02432$.

The 95 percent margin of error is $1.960(0.02432) \approx 0.0477$.

The 95 percent interval is 0.66 ± 0.0477, or about (0.612, 0.708). It is wider, as it must be, because more confidence requires a wider net.

Write It Like a 5

Model Response

Prompt. A random sample of 400 students at a school is asked how they get to campus. In the sample, 180 bike. A 95 percent confidence interval for the true proportion who bike is (0.401, 0.499). (a) Interpret the interval. (b) Interpret the confidence level. (c) A classmate says, 'So 95 percent of students bike between 40.1 and 49.9 percent of the time.' Explain the error.\n\n**(a)** We are 95 percent confident that the interval from 0.401 to 0.499 captures the true proportion of all students at this school who bike to campus.\n\n**(b)** If we repeatedly took random samples of 400 students from this school and built a 95 percent interval from each sample, about 95 percent of those intervals would capture the true proportion of students who bike.\n\n**(c)** The interval estimates a single population parameter, the proportion of students who bike. It says nothing about how often an individual student bikes, and it does not describe a percentage of students. The classmate has confused a statement about a parameter with a statement about individuals.

What the grader looks for

(1) States the confidence level and both endpoints in the interpretation. (2) Names the parameter in context as a population proportion for all students at this school. (3) Uses the language of capturing or being confident, not 'probability that p is in the interval'. (4) Describes the confidence level as a long-run percentage of intervals from repeated random samples. (5) Explains that the interval is about a parameter, not individuals or sample values.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Coach's Playbook

1. Name the parameter in words before doing anything else (the proportion of *what* in *which population*).
2. Compute $\hat{p}$ from counts, never from a rounded percentage you were given if counts exist.
3. Decide which question is being asked: interpret the interval, or interpret the level. They have different templates.
4. Check the answer's plausibility: the interval must be centered at $\hat{p}$, and its lower and upper values must stay between 0 and 1.
5. When comparing intervals, ask which has the larger z^* or the smaller n; that one is wider.

Common traps

1. Saying 'there is a 95 percent chance that p is in the interval.' The parameter is fixed. 2. Writing 'we are 95 percent confident that $\hat{p}$ is in the interval.' The statistic is already known and sits at the center. 3. Describing the interval as covering 95 percent of the data or of individual people. 4. Believing that a higher confidence level gives a more precise interval. It gives a more reliable one, but a wider one.

Mistake Clinic

Spot the error

Wrong solution. 'A 90 percent confidence interval for the proportion of teens who stream music daily is (0.71, 0.79). This means there is a 90 percent probability that the true proportion is between 0.71 and 0.79.'\n\n**Where it fails.** The true proportion is a fixed number, so probability statements about it are not appropriate once the interval is built. The 90 percent belongs to the method.\n\n**Correct.** We are 90 percent confident that the interval from 0.71 to 0.79 captures the true proportion of all teens in the population who stream music daily.

Spot the error

Wrong solution. 'To get a narrower interval and also be more certain, the pollster should switch from 90 percent to 99 percent confidence.'\n\n**Where it fails.** A larger confidence level uses a larger z^* (2.576 instead of 1.645), so the margin of error grows. Confidence and precision pull in opposite directions.\n\n**Correct.** To narrow the interval, keep the confidence level and increase the sample size, or accept a lower confidence level.

Spot the error

Wrong solution. 'The 95 percent interval for a proportion is (0.30, 0.40), so the point estimate is $\hat{p} = 0.30$ because that is the lower bound.'\n\n**Where it fails.** The interval is symmetric about the point estimate, so the estimate is the midpoint, not an endpoint.\n\n**Correct.** $\hat{p} = (0.30 + 0.40)/2 = 0.35$, and the margin of error is 0.05.

Real-World Connection

Election-night forecasts, public-health prevalence reports, and product satisfaction surveys all print a margin of error beside the headline number. When a report says 'support is 52 percent, plus or minus 3 points,' it is quoting a confidence interval, usually at 95 percent. A reader who knows the logic can tell when two candidates' intervals overlap so heavily that 'a lead' may only be sampling variability.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

A random sample of 250 customers gives $\hat{p} = 0.36$. Without computing a new interval, explain what must be true about the 95 percent interval from a second random sample of 1000 customers with the same $\hat{p}$, and compute both margins of error.\n\n**Solution.** The second interval has the same center, 0.36, but half the width because n is four times as large. Sample 1: $\text{SE} = \sqrt{0.36(0.64)/250} \approx 0.03036$, $\text{ME} \approx 1.96(0.03036) \approx 0.0595$. Sample 2: $\text{SE} \approx 0.01518$, $\text{ME} \approx 0.0297$. The ratio of margins is 2, matching the square-root law. The second sample does not guarantee that its interval lies inside the first one, since its $\hat{p}$ will not be exactly 0.36 in reality.

Practice

PRACTICE

The Logic of Confidence Intervals

Show your work. A calculator is allowed. Justify conclusions in context.

1. A random sample of 150 students includes 63 who own a reusable bottle. What is the point estimate of the population proportion who own one?
   (A) 0.63 **(B)** 0.42 **(C)** 0.58 **(D)** 63
2. A 95 percent confidence interval for the proportion of all households that recycle is (0.58, 0.66). Which is the correct interpretation?
   (A) There is a 95 percent probability that a sample proportion is between 0.58 and 0.66.
   (B) 95 percent of households recycle between 58 and 66 percent of their waste.
   (C) We are 95 percent confident that the true proportion of all households that recycle is between 0.58 and 0.66.
   (D) 95 percent of all samples have $\hat{p}$ between 0.58 and 0.66.
3. What does a 90 percent confidence level mean?
   (A) The parameter is within the interval 90 percent of the time.
   (B) 90 percent of the data fall inside the interval.
   (C) There is a 90 percent chance that the next sample proportion lies in the interval.
   (D) In the long run, about 90 percent of intervals built this way from repeated random samples would capture the parameter.
4. Which change would make a confidence interval for a proportion narrower?
   (A) Increasing the sample size **(B)** Increasing the confidence level
   (C) Using a z* from a higher confidence level **(D)** Making the sample proportion closer to 0.5
5. Two intervals use the same data. One is a 95 percent interval and the other is a 99 percent interval. Which statement is true?
   (A) The 99 percent interval is narrower because it is more accurate.
   (B) The two intervals have different centers.
   (C) The 99 percent interval is wider and has the same center.
   (D) They have the same width because the data are the same.
6. A researcher constructs 200 independent 90 percent confidence intervals, each from its own random sample. About how many should capture the true parameter?
   (A) 90 **(B)** 180 **(C)** 20 **(D)** 198
7. A confidence interval for a proportion is (0.38, 0.50). What are $\hat{p}$ and the margin of error?
   (A) $\hat{p} = 0.38$, margin of error 0.12 **(B)** $\hat{p} = 0.44$, margin of error 0.12
   (C) $\hat{p} = 0.50$, margin of error 0.06 **(D)** $\hat{p} = 0.44$, margin of error 0.06
8. A student says, 'The 95 percent interval is (0.41, 0.49), so there is a 95 percent probability that p is in this interval.' What is wrong with this statement?
   (A) Nothing, the statement is correct.
   (B) The parameter p is fixed, so the 95 percent describes the method, not this one interval.
   (C) The probability should be 5 percent.
   (D) The interval should be described using the sample mean.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. A poll has a margin of error of 6 percentage points. About how large must the sample be to cut the margin of error to 3 points (same confidence level and similar p-hat)?
   (A) 2 times as large **(B)** Half as large **(C)** 16 times as large
   (D) 4 times as large
10. A random sample of 240 teens includes 84 who say they sleep fewer than 7 hours on school nights. (a) Give the point estimate and name the parameter it estimates. (b) The 95 percent confidence interval is (0.290, 0.410). Interpret it in context.
11. A news site reports many 95 percent confidence intervals during an election season. Suppose it publishes 60 independent intervals, each from a new random sample. (a) About how many do you expect to miss the true parameter? (b) Explain what 'miss' means.
12. Critique each statement about a 90 percent confidence interval of (0.22, 0.30) for the proportion of a town's residents who use public transit weekly. (a) '90 percent of residents use transit between 22 and 30 percent of weeks.' (b) 'There is a 90 percent chance the sample proportion is in the interval.'
13. Two analysts use the same random sample of 500 voters. Analyst 1 reports a 90 percent interval and Analyst 2 reports a 99 percent interval. (a) Which interval is wider and why? (b) Analyst 2 says that the wider interval is 'worse'. Respond.
14. A 90 percent confidence interval for a proportion is (0.54, 0.62). (a) Find $\hat{p}$ and the margin of error. (b) Find the standard error. (c) Find the 95 percent interval from the same data.
15. A school's survey of a random sample of 300 students gives a 95 percent interval of (0.42, 0.54) for the proportion who support a later start time. The principal says, 'A majority support it.' (a) Is that conclusion justified by the interval? (b) What can be said honestly?

7.2 One-Proportion z-Intervals, Margin of Error, and Sample Size

Build and interpret one-proportion z-intervals with all conditions checked, find the margin of error, and plan the sample size you need.

Why This Matters

You now know what a confidence interval means. This lesson is about building one correctly and about planning a survey so the interval comes out as tight as you need. A pollster cannot call 5,000 people for a quick estimate and cannot afford to call only 20, so the real question is: how many people are enough? Students lose points here by skipping the conditions, by using p where only $\hat{p}$ is known, or by rounding a sample size down.

The one-proportion z-interval is a core procedure of Exam Unit 3 (Inference for Categorical Data: Proportions). Free-response questions expect the full four-part structure (parameter and procedure, conditions, calculation, conclusion in context), and multiple-choice questions often ask about the margin of error or the sample size needed.

Core Ideas

When we do not know p, we estimate the standard deviation of $\hat{p}$ by plugging in $\hat{p}$. The result is called the **standard error**: $\text{SE}_{\hat{p}} = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$. The interval then adds and subtracts z^* standard errors. Critical values come from the standard normal distribution: you want the z with area C between $-z^*$ and z^*.

Confidence level	Area in each tail	Critical value z^*
90%	0.05	1.645
95%	0.025	1.960
99%	0.005	2.576

One-Proportion z-Interval

Parameter: p = the true proportion of (population) that (has the characteristic).

Interval: $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

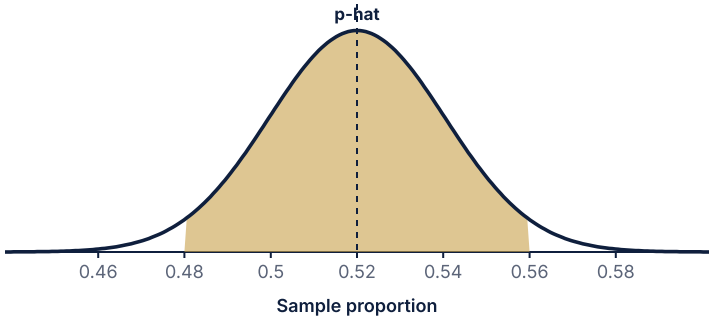
Conditions: (1) Random: the data come from a random sample or randomized experiment. (2) 10 percent: when sampling without replacement, n is less than 10 percent of the population. (3) Large counts: $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.

Conclusion: We are C% confident that the interval from ___ to ___ captures the true proportion of (population) that (has the characteristic).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Each condition protects one assumption. Randomness guards against bias; no formula can repair a biased sample. The 10 percent condition makes the observations nearly independent, which is what the standard error formula assumes. The large counts condition makes the sampling distribution of $\hat{p}$ close enough to normal that the z^* values are valid. Notice that for an interval we use $\hat{p}$ in the large counts check, because p is unknown.



Interval of 0.52 plus or minus 0.040, about 1.96 standard errors on each side.

Choosing the sample size. The margin of error is $\text{ME} = z^*\sqrt{\dfrac{\tilde{p}(1-\tilde{p})}{n}}$, where $\tilde{p}$ is a planning value. Solve for n by squaring. If you want ME no larger than a target, take $n \geq \left(\dfrac{z^*}{\text{ME}}\right)^2 \tilde{p}(1-\tilde{p})$ and **always round up**, since rounding down would give a margin slightly too big.

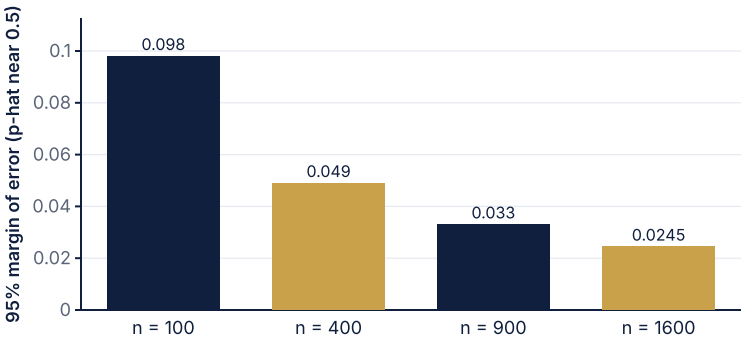
Sample Size for a Desired Margin of Error

$n \geq \left(\dfrac{z^*}{\text{ME}}\right)^2 \tilde{p}(1-\tilde{p})$, rounded up to the next whole number.

If a pilot study or earlier estimate gives a reasonable guess for p, use it as $\tilde{p}$.

If there is no guess, use $\tilde{p} = 0.5$. It maximizes $\tilde{p}(1-\tilde{p})$ and so guarantees the margin of error will be met (the **conservative** choice).

State ME as a decimal: 3 percentage points is 0.03.



Diminishing returns: each extra step costs far more respondents for a smaller gain.

Worked Examples

CRITICAL VALUES

Find the critical value z^* for a 98 percent confidence interval.

SOLUTION

A 98 percent interval leaves 2 percent in the two tails, so 1 percent in each tail.

We need the z with area 0.99 to its left: on the built-in calculator, invNorm(0.99) ≈ 2.326.

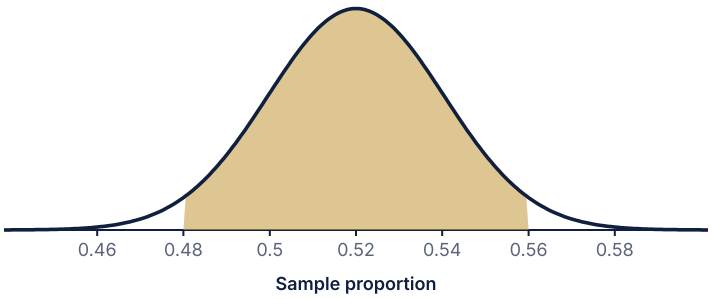
So $z^* \approx 2.326$. It lies between 1.960 and 2.576, as expected for a level between 95 and 99 percent.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A COMPLETE 95 PERCENT INTERVAL

A city surveys a random sample of 600 of its roughly 80,000 residents. Of these, 312 support building a new bike lane. Construct and interpret a 95 percent confidence interval for the proportion of all residents who support it.



Middle 95 percent of the sampling distribution.

SOLUTION

Parameter and procedure. Let p be the true proportion of all residents who support the bike lane. We use a one-proportion z-interval.

Conditions. Random: the sample is random. 10 percent: $600 < 8000$, which is 10 percent of 80,000. Large counts: $n\hat{p} = 312 \geq 10$ and $n(1-\hat{p}) = 288 \geq 10$.

Calculation. $\hat{p} = 312/600 = 0.52$ and $\text{SE} = \sqrt{0.52(0.48)/600} \approx 0.0204$. The margin is $1.960(0.0204) \approx 0.0400$, so the interval is 0.52 ± 0.040, or (0.480, 0.560).

Conclusion. We are 95 percent confident that the interval from 0.480 to 0.560 captures the true proportion of all residents who support the bike lane.

MARGIN OF ERROR FROM AN INTERVAL

A 95 percent interval for a proportion is (0.38, 0.46). Find $\hat{p}$, the margin of error, and roughly how many people were sampled.

SOLUTION

$\hat{p} = (0.38 + 0.46)/2 = 0.42$.

$\text{ME} = (0.46 - 0.38)/2 = 0.04$.

Since $\text{ME} = 1.960\sqrt{0.42(0.58)/n}$, solve: $n = (1.960/0.04)^2(0.42)(0.58) \approx 584.9$. The sample had about 585 people. This reverse-engineering is a quick way to check whether a reported interval and sample size agree.

CHECKING THE CONDITIONS

A technician tests a random sample of 50 phone batteries from one large shipment and finds 4 that are defective. She wants a 95 percent interval for the shipment's defect rate. Can she use the z-interval?

SOLUTION

Random: yes. 10 percent: 50 is less than 10 percent of a large shipment. Large counts: $n\hat{p} = 4 < 10$, so the condition **fails**.

With so few defects, the sampling distribution of $\hat{p}$ is skewed right, so the normal critical value cannot be trusted.

She should collect a larger sample (enough for at least 10 defectives and 10 non-defectives) before using this procedure. A z-interval here could report endpoints that are not credible.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

PLANNING A SAMPLE SIZE

A health club wants to estimate the proportion of members who use the pool, within 3 percentage points with 95 percent confidence. An earlier pilot found that about 30 percent use it. How many members should be sampled? How many if there were no prior information?

Planning value	Computation	Required n
0.30 (pilot)	4268.44 × 0.21	897
0.50 (no information)	4268.44 × 0.25	1068

SOLUTION

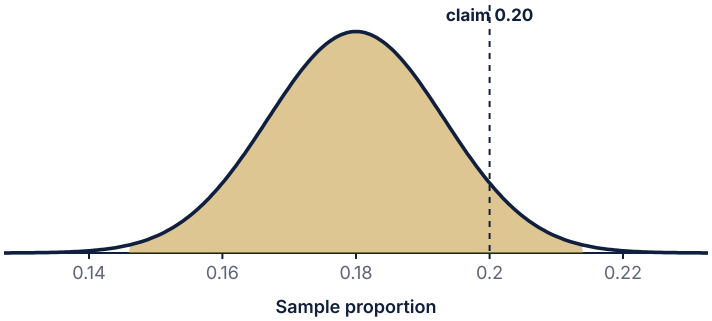
With the pilot: $n \geq (1.960/0.03)^2(0.30)(0.70) = 4268.44(0.21) \approx 896.3$, so $n = 897$.

With no information, use $\tilde{p} = 0.5$: $n \geq 4268.44(0.25) \approx 1067.1$, so $n = 1068$.

The conservative value costs about 170 more members. Both answers are rounded **up**.

INTERVAL AND A CLAIM

A random sample of 850 shoppers (far fewer than 10 percent of all shoppers at a chain) shows that 153 use a store's loyalty app. Build a 99 percent interval and decide whether the chain's claim that 20 percent of its shoppers use the app is plausible.



The claimed value 0.20 sits inside the middle 99 percent.

SOLUTION

$\hat{p} = 153/850 = 0.18$. Counts: $n\hat{p} = 153$ and $n(1-\hat{p}) = 697$, both at least 10, and sampling is random, so the procedure applies.

$\text{SE} = \sqrt{0.18(0.82)/850} \approx 0.01318$ and $\text{ME} = 2.576(0.01318) \approx 0.0339$. The interval is (0.146, 0.214).

Because 0.20 lies inside the interval, the claim is plausible. The data do not provide convincing evidence that the true proportion differs from 20 percent. This connects to the logic of significance tests, coming next.

Write It Like a 5

Model Response

Prompt. A random sample of 240 of the 3,000 students at a school shows that 84 regularly use the library. Construct and interpret a 95 percent confidence interval for the proportion of all students at the school who regularly use the library.\n\n**Parameter and procedure.** Let p be the true proportion of all students at the school who regularly use the library. I will construct a one-proportion z-interval.\n\n**Conditions.** The students were selected at random. The sample of 240 is less than 10 percent of 3,000 students, which is 300. The counts $240(0.35) = 84$ and $240(0.65) = 156$ are both at least 10, so the sampling distribution of $\hat{p}$ is approximately normal.\n\n**Calculation.** $\hat{p} = 84/240 = 0.35$, $\text{SE} = \sqrt{0.35(0.65)/240} \approx 0.03079$, and $z^* = 1.960$. The interval is $0.35 \pm 1.960(0.03079) = 0.35 \pm 0.0604$, or (0.290, 0.410).\n\n**Conclusion.** We are 95 percent confident that the interval from 0.290 to 0.410 captures the true proportion of all students at this school who regularly use the library.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

(1) Identifies the parameter in context and names the procedure (one-proportion z-interval). (2) Checks all three conditions with numbers, including the 10 percent condition against the population size. (3) Correct standard error, critical value, and interval. (4) A conclusion that states the level, the endpoints, and the parameter in context. (5) Uses the language of confidence rather than probability.

Coach's Playbook

1. Write the parameter first. If you cannot say it in a sentence, you cannot interpret it.
2. Run the three condition checks and put numbers on the large counts check.
3. Compute $\hat{p}$, SE, ME, and the interval; keep at least four decimals until the final rounding.
4. On the built-in calculator, use the one-proportion z-interval tool as a check, but still show the formula and values in writing.
5. Conclude in context, naming level, endpoints, and parameter. For sample size, round up and state the planning value you used.

Common traps

1. Using p in the large counts check for an interval; use $\hat{p}$ because p is unknown. 2. Rounding a sample size to the nearest whole number instead of up. 3. Entering the margin of error as 3 instead of 0.03 in the sample size formula. 4. Believing a larger sample cures bias; a nonrandom sample stays biased at any size.

Mistake Clinic

Spot the error

Wrong solution. 'We need a margin of error of 0.04 with 90 percent confidence and no prior estimate. $n = (1.645/0.04)^2(0.5) = 845.6$, so $n = 846$.'\n\n**Where it fails.** The formula needs $\tilde{p}(1-\tilde{p})$, which is 0.25 when $\tilde{p} = 0.5$. The student multiplied by 0.5 only.\n\n**Correct.** $n \geq (1.645/0.04)^2(0.25) \approx 422.7$, so $n = 423$.

Spot the error

Wrong solution. 'For 4 successes out of 50, $\hat{p} = 0.08$. The 95 percent interval is $0.08 \pm 1.96\sqrt{0.08(0.92)/50} = (0.005, 0.155)$. The conditions are fine because the sample is random.'\n\n**Where it fails.** Randomness is one condition of three. $n\hat{p} = 4 < 10$ violates large counts, so the interval is not trustworthy.\n\n**Correct.** State that the large counts condition fails and that the z-interval should not be used with this sample.

Spot the error

Wrong solution. 'A 95 percent interval based on 200 people is (0.30, 0.40). Because the sample is large, 95 percent of the 200 people have values in that range.'\n\n**Where it fails.** Each person either has or lacks the characteristic. The interval is about the population proportion, not individual values.\n\n**Correct.** We are 95 percent confident that the interval from 0.30 to 0.40 captures the true population proportion.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Before any national survey is fielded, statisticians pick a target margin of error and compute the sample size, exactly as above, and then add budget for people who will not respond. That is why so many polls land near 1,000 respondents: with the conservative $\tilde{p} = 0.5$ and 95 percent confidence, 1,068 people give a margin of about 3 points, and going further gets expensive fast.

Challenge Corner

A marketing team wants a 95 percent interval for the proportion of customers who prefer a new flavor, with width (not margin) at most 0.06. A previous survey suggests the proportion is between 0.20 and 0.40. What sample size guarantees the requirement?\n\n**Solution.** A width of 0.06 means ME ≤ 0.03. To be safe across the plausible range, choose the value in [0.20, 0.40] that makes $\tilde{p}(1-\tilde{p})$ largest, which is the value closest to 0.5, namely 0.40. Then $n \geq (1.960/0.03)^2(0.40)(0.60) = 4268.44(0.24) \approx 1024.4$, so $n = 1025$.

Practice

PRACTICE

One-Proportion z-Intervals, Margin of Error, and Sample Size

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which is the critical value z^* for a 90 percent confidence interval?
   (A) 1.282 **(B)** 1.960 **(C)** 1.645 **(D)** 2.326
2. A random sample of 60 students is used to estimate a proportion, and $\hat{p} = 0.10$. Why can a one-proportion z-interval NOT be used safely?
   (A) The sample is larger than 10 percent of the population.
   (B) $n\hat{p} = 6$ is less than 10, so the large counts condition fails.
   (C) A z-interval can only be used when n is at least 100.
   (D) The sample is not random because it is small.
3. For $\hat{p} = 0.40$ and $n = 200$, what is the standard error of $\hat{p}$?
   (A) 0.0012 **(B)** 0.0693 **(C)** 0.2400 **(D)** 0.0346
4. In a random sample of 120 members, 54 prefer morning classes. Which is the 95 percent confidence interval for the population proportion?
   (A) (0.405, 0.495) **(B)** (0.310, 0.590) **(C)** (0.361, 0.539)
   (D) (0.450, 0.539)
5. A club wants a 95 percent margin of error of at most 0.04 and has no prior estimate of the proportion. What minimum sample size is needed?
   (A) 600 **(B)** 601 **(C)** 385 **(D)** 1201
6. An earlier study says $p \approx 0.20$. The margin of error must be at most 0.05 at 95 percent confidence. What sample size is required?
   (A) 246 **(B)** 245 **(C)** 385 **(D)** 246.2
7. If a researcher quadruples the sample size and keeps everything else the same, the margin of error for a proportion will
   (A) be cut in half. **(B)** be cut to one fourth. **(C)** double.
   (D) stay the same.
8. Which of the following is a required condition for a one-proportion z-interval?
   (A) The population is approximately normal.
   (B) The sample size is at least 30.
   (C) The population standard deviation is known.
   (D) The counts $n\hat{p}$ and $n(1-\hat{p})$ are both at least 10.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. A 99 percent interval for a proportion from a random sample is (0.46, 0.54). Which statement is best supported?
   (A) It is 99 percent likely that p is less than 0.50.
   (B) The sample proves that exactly 50 percent have the characteristic.
   (C) Values of p from 0.46 to 0.54 are plausible, so p = 0.50 cannot be ruled out.
   (D) The interval contains 99 percent of the sample values.
10. A random sample of 500 of the 12,000 customers of a coffee shop shows that 210 buy a pastry with their drink. Construct and interpret a 95 percent confidence interval for the proportion of all customers who buy a pastry.
11. A random sample of 40 of the 2,000 players in a league is asked whether they have had a sports injury this season; 3 say yes. (a) Check the conditions for a one-proportion z-interval. (b) What should the researcher do?
12. A school board wants to estimate the proportion of families who support a new calendar with 99 percent confidence and a margin of error no larger than 0.03. There is no prior information. Find the required sample size and explain the choice of planning value.
13. In a random sample of 300 residents (far fewer than 10 percent of the population), 132 favor extending library hours. (a) Construct a 95 percent confidence interval. (b) A council member claims that a majority of residents favor extending hours. Does the interval support this claim?
14. A pilot study of 60 teens finds that 21 own a smart watch. The district then wants a 90 percent interval with margin of error at most 0.04. (a) Use the pilot as the planning value to find the sample size. (b) If the district instead used 0.5, what would happen to the required sample size, and why?
15. Explain, using the formula for margin of error, two different ways a pollster could reduce the margin of error from a survey, and state the cost of each.

7.3 The Logic of Significance Tests and P-Values

How a significance test weighs a claim against sample evidence: hypotheses, the P-value and what it does and does not mean, and how to reach a conclusion.

Why This Matters

A confidence interval estimates a parameter. A **significance test** answers a different question: is the evidence in this sample strong enough to doubt a specific claim? A drink maker says 30 percent of its customers choose the new flavor; a random sample shows 38 percent. Is that gap real, or could it easily happen by chance when the claim is true? The logic of a test is a little like a courtroom, and the language of a test is where most points are won or lost.

Students typically stumble in three places: writing hypotheses with statistics instead of parameters, misinterpreting the P-value as the probability that the null is true, and writing 'accept the null' at the end. This lesson, part of Exam Unit 3, builds the logic before we do any formal calculation. In the next lesson we put numbers into the machinery.

Core Ideas

A test begins with two competing claims about a population parameter. The **null hypothesis** H_0 is the claim we assume true while we weigh the evidence; it always states a single value, using equality, such as $H_0 : p = 0.30$. The **alternative hypothesis** H_a is the claim we hope to find evidence for. It uses $<$, $>$, or $\neq$. Both hypotheses are about the parameter p, never about $\hat{p}$, because we already know $\hat{p}$ exactly.

Writing Hypotheses

State the parameter in words: p = the true proportion of (population) that (characteristic).

$H_0 : p = p_0$, where p_0 is the claimed value.

$H_a : p > p_0$, $p < p_0$, or $p \neq p_0$. The direction comes from the research question, set before seeing the data.

A two-sided alternative ($\neq$) is used when the question is only whether the proportion differs from the claim.

We then ask how surprising the sample is if H_0 were true. A **P-value** is the probability, computed assuming H_0 is true, of getting a sample result at least as extreme as the one observed, in the direction of H_a. A small P-value means the observed result would be rare under the null, which is evidence against H_0. A large P-value means the result is quite ordinary under the null.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

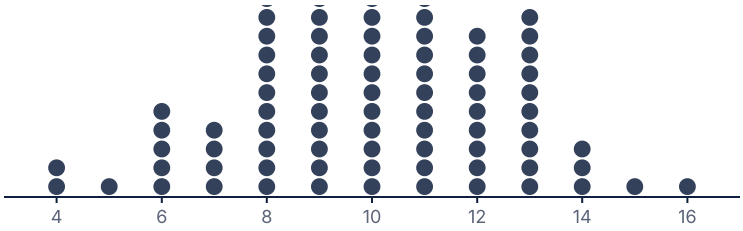
Interpreting a P-Value

Template: 'Assuming that H_0 is true (state it in context), the probability of getting a sample result (statistic in context) as extreme as, or more extreme than, the one observed is (P-value).'

It is NOT the probability that H_0 is true, and NOT the probability that the result happened by chance alone.

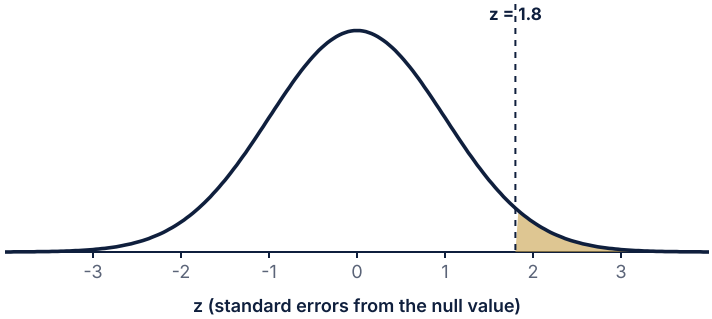
Smaller P-value means stronger evidence against H_0; it does not measure the size or importance of an effect.

Where does the P-value come from? One way is simulation. Suppose a music student claims she can tell the better of two audio clips. If she is just guessing, each of 20 trials is a 50-50 outcome. We simulate 100 guessers and count how many of 20 each gets right. She actually got 15 right. Only 2 of the 100 simulated guessers did that well, so the estimated P-value is $2/100 = 0.02$. Getting 15 or more right is rare for a guesser.

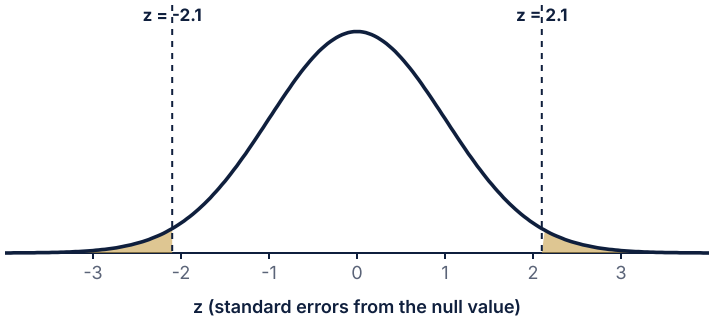


Correct answers out of 20 in a simulated guesser

With a normal model, the P-value is an area under the curve beyond the test statistic. The upper-tail case below shades the area to the right of $z = 1.8$, and the two-sided case shades both tails because a result equally extreme in the other direction also counts.



One-sided (upper tail): P-value = 0.0359.



Two-sided: both tails beyond |z| = 2.1 are shaded, P-value = 0.0357.

To make a decision, we compare the P-value with a **significance level** α, chosen before collecting data (commonly 0.05). If P-value $\leq \alpha$, we **reject** H_0; the result is *statistically significant*, and we have convincing evidence for H_a. If P-value $> \alpha$, we **fail to reject** H_0; we do not have convincing evidence for H_a. We never say 'accept H_0' or 'prove H_0', because a large P-value only means the data are consistent with the null, not that the null is true.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Conclusion Templates

If P-value $\leq \alpha$: 'Because the P-value of ___ is less than $\alpha =$ ___, we reject H_0. We have convincing evidence that (state H_a in context).'

If P-value $> \alpha$: 'Because the P-value of ___ is greater than $\alpha =$ ___, we fail to reject H_0. We do not have convincing evidence that (state H_a in context).'

Putting the pieces together. The logic of a test is a proof by contradiction with probability. We assume the null is true, see how surprising the data are under that assumption, and if the data would be very surprising we doubt the assumption. Notice what the argument cannot do. It cannot show the null is true, because many different null values would produce similar data. It also cannot tell us how large the departure from the null is. For size, we need a confidence interval.

Common misreading	Why it is wrong	Better wording
The P-value is the probability that H_0 is true.	H_0 is a fixed claim, not a random event; the P-value is computed assuming it.	If H_0 is true, the probability of results this extreme is the P-value.
A P-value of 0.04 means a 4% chance the result is due to chance.	Chance is the model used to compute it, not an event with its own probability.	Under random sampling variability alone, such an extreme result is rare.
Failing to reject means the null is true.	Large P-values reflect consistency, not proof.	We do not have convincing evidence for H_a.
A smaller P-value means a bigger effect.	Sample size also shrinks P-values.	Compare estimates and intervals to judge size.

One-sided or two-sided? Choose by the research question before seeing data. 'Did the rate go up?' is one-sided. 'Is the rate different?' is two-sided. A one-sided P-value is half of the two-sided value for a symmetric reference distribution when the sample lies in the direction of the alternative, which is why switching sides after the fact is a form of cheating. In practice, if you are unsure which to use, a two-sided test is the safer, more conservative choice.

Statistical versus practical significance. With a very large sample, even a tiny difference between $\hat{p}$ and p_0 can give a small P-value. A website with a million visitors might find that a new button color raises sign-ups from 12.00 to 12.05 percent, with a P-value of 0.01. The result is statistically significant but probably not worth the redesign. Always ask how big the difference is, not just whether it is statistically detectable.

- **Significance level α:** the cutoff for 'rare enough', set before the data. Common choices are 0.10, 0.05, and 0.01. A smaller α demands stronger evidence before rejecting H_0.
- **Statistically significant:** the P-value is at most α, so the sample result would be unusual if H_0 were true.
- **Test statistic:** a number that measures how many standard errors the sample result lies from the null value. A large $|z|$ goes with a small P-value.
- **Conclusion:** always about H_a in the context of the problem, using the phrase 'convincing evidence' and never 'prove'.

Which α should you use? If the problem gives one, use it. If the consequences of wrongly rejecting a true null are severe (for example, approving an unsafe product), a small α such as 0.01 is wise. If missing a real effect is the bigger worry, a larger α may be acceptable. The next lessons make this tradeoff precise using Type I and Type II errors.

Worked Examples

WRITING HYPOTHESES

A school says that 30 percent of its students ride the bus. A club suspects the percentage has increased since a new route was added. Write the hypotheses in words and symbols.

SOLUTION

Let p be the true proportion of students at the school who ride the bus.

$H_0 : p = 0.30$ and $H_a : p > 0.30$. The alternative is one-sided because the club suspects an *increase*.

If the club only wanted to know whether the proportion had *changed*, we would use $H_a : p \neq 0.30$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

FIXING FAULTY HYPOTHESES

A student writes $H_0 : \hat{p} = 0.30$ and $H_a : \hat{p} \neq 0.34$ for a poll where 34 percent of a sample chose the new flavor. What is wrong, and what are the correct hypotheses for testing the maker's claim that 30 percent of all customers prefer it?

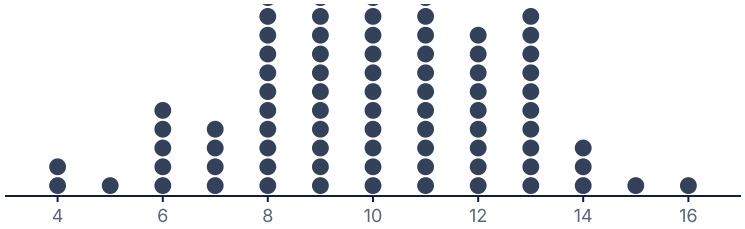
SOLUTION

There are two errors: the hypotheses use the statistic $\hat{p}$ instead of the parameter p, and the alternative uses the sample value 0.34, which is data, instead of the claim 0.30.

Correct: $H_0 : p = 0.30$ and $H_a : p \neq 0.30$, where p is the true proportion of all customers who prefer the new flavor.

A SIMULATED P-VALUE

In the guessing simulation above, the student got 15 of 20 right. Use the dot plot to estimate the P-value and state a conclusion at $\alpha = 0.05$.



Correct answers out of 20 in a simulated guesser

SOLUTION

$H_0 : p = 0.5$ (she is guessing); $H_a : p > 0.5$ (she can really tell). Of the 100 simulated guessers, 2 had 15 or more correct.

Estimated P-value $= 2/100 = 0.02$. Assuming she is guessing, the chance of 15 or more correct out of 20 is about 0.02.

Because $0.02 < 0.05$, we reject H_0. We have convincing evidence that she can tell the clips apart better than guessing.

INTERPRETING A P-VALUE

A test of $H_0 : p = 0.30$ against $H_a : p > 0.30$ for the proportion of customers who prefer a new flavor gives a P-value of 0.032. Interpret it in context.

SOLUTION

If the true proportion of customers who prefer the new flavor were 0.30, there is about a 3.2 percent chance of getting a sample proportion as large as the one observed, or larger, just from random sampling.

This is a statement about how unusual the data are when the null is true. It is not the probability that the null hypothesis is true.

DECISIONS AT DIFFERENT LEVELS

A test gives P-value $= 0.032$. State the decision and conclusion at (a) $\alpha = 0.05$ and (b) $\alpha = 0.01$ for H_a: the proportion who prefer the new flavor exceeds 0.30.

Level	Compare	Decision
0.05	0.032 < 0.05	Reject H0
0.01	0.032 > 0.01	Fail to reject H0

SOLUTION

(a) $0.032 < 0.05$, so we reject H_0. We have convincing evidence that more than 30 percent of customers prefer the new flavor.

(b) $0.032 > 0.01$, so we fail to reject H_0. We do not have convincing evidence that more than 30 percent prefer it.

The data did not change; the standard of evidence did. This is why α must be set before looking at the data.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

P-VALUES FROM Z

A two-sided test gives $z = -2.10$. A one-sided upper-tailed test gives $z = 1.80$. Find each P-value and decide which result is stronger evidence against its null.

SOLUTION

Two-sided: $P = 2 \cdot P(Z < -2.10) = 2(0.0179) \approx 0.0357$.

One-sided: $P = P(Z > 1.80) \approx 0.0359$.

The P-values are nearly equal, so the two results are about equally strong evidence, even though $|z|$ differs. The two-sided test needs a bigger $|z|$ to reach the same P-value because it counts both tails.

Write It Like a 5

Model Response

Prompt. A manufacturer claims that 90 percent of its replacement screens are free of defects. An inspector suspects the rate is lower. (a) State the hypotheses. (b) A test gives P-value $= 0.018$. Interpret it. (c) Give a conclusion at $\alpha = 0.05$.\n\n**(a)** Let p be the true proportion of replacement screens from this manufacturer that are free of defects. $H_0 : p = 0.90$ and $H_a : p < 0.90$.\n\n**(b)** If the true proportion of defect-free screens were 0.90, the probability of getting a sample proportion as small as the one observed, or smaller, is 0.018.\n\n**(c)** Because the P-value of 0.018 is less than $alpha = 0.05$, we reject H_0. We have convincing evidence that the true proportion of defect-free replacement screens is less than 0.90.

What the grader looks for

(1) Parameter defined in context with the correct population. (2) Correct null with equality and an alternative whose direction matches the suspicion. (3) A P-value interpretation that is conditional on H_0 being true and refers to a result as extreme or more extreme. (4) A comparison of the P-value with α and a decision about H_0. (5) A conclusion in context phrased as 'convincing evidence', not 'proved'.

Coach's Playbook

1. Read the question for the direction of the claim: is it 'greater', 'less', or 'different'?
2. Write the parameter in words, then H_0 with an equals sign and H_a with the matching direction.
3. Remember the P-value is computed assuming H_0 is true; say so when you interpret it.
4. Compare the P-value with α, then state the decision about H_0 (reject or fail to reject).
5. Close with a sentence in context that uses the alternative's wording.

Common traps

1. Writing hypotheses with $\hat{p}$ or with a sample value. 2. Saying the P-value is the probability that H_0 is true. 3. Writing 'accept H_0' after a large P-value. 4. Choosing a one-sided alternative after seeing which direction the data went.

Mistake Clinic

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Wrong solution. 'The P-value is 0.04, so there is a 4 percent chance that the null hypothesis is true.'\n\n**Where it fails.** H_0 is either true or false; it is not random. The P-value is computed assuming H_0 is true and measures how unusual the data would be.\n\n**Correct.** If H_0 is true, there is a 4 percent chance of obtaining a result at least this extreme.

Spot the error

Wrong solution. 'The P-value of 0.31 is bigger than $\alpha = 0.05$, so we accept H_0 and conclude that the proportion is exactly 0.30.'\n\n**Where it fails.** A large P-value only says the data are consistent with H_0; many other values would be consistent too.\n\n**Correct.** Fail to reject H_0. We do not have convincing evidence that the proportion differs from 0.30.

Spot the error

Wrong solution. 'The sample of 400 gave $\hat{p} = 0.34$, so $H_0 : p = 0.34$ and $H_a : p > 0.34$.'\n\n**Where it fails.** The null value is the claim being tested, not the sample result.\n\n**Correct.** If the claim is that the true proportion is 0.30, then $H_0 : p = 0.30$, and the alternative follows from the research question.

Real-World Connection

Drug trials, quality checks, and website experiments all ask whether a change is real or just sampling noise. When a site reports that a new layout improved sign-ups with 'p = 0.02', it is quoting exactly the kind of tail probability computed here. A small P-value alone does not tell you whether the improvement is large enough to matter; that takes the interval estimate too. Analysts in these fields usually report the estimate, an interval, and the P-value together so decision makers can weigh both the evidence and the size of the change.

Challenge Corner

The same sample is used for two tests of $H_0 : p = 0.50$. Test 1 uses $H_a : p > 0.50$ and gives P-value 0.03 because the sample proportion is above 0.50. What is the P-value for the two-sided test with the same data, and for the one-sided test $H_a : p < 0.50$?\n\n**Solution.** The two-sided P-value counts both tails, so it is $2(0.03) = 0.06$ when the distribution is symmetric. For $H_a : p < 0.50$, the P-value is the area on the other side of the observed statistic, so it is $1 - 0.03 = 0.97$. The same data can be significant at 0.05 one-sided but not two-sided, which is why the alternative must be decided before the data are seen.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE The Logic of Significance Tests and P-Values

Show your work. A calculator is allowed. Justify conclusions in context.

1. A company claims that 15 percent of its tablets are returned. A reviewer suspects the true return rate is higher. Which are the correct hypotheses?

   (A) $H_0 : p > 0.15$, $H_a : p = 0.15$ **(B)** $H_0 : \hat{p} = 0.15$, $H_a : \hat{p} > 0.15$ **(C)** $H_0 : p = 0.15$, $H_a : p > 0.15$ **(D)** $H_0 : p = 0.15$, $H_a : p \neq 0.15$

2. Which statement is a valid null hypothesis for a population proportion?

   (A) $p = 0.62$ **(B)** $p > 0.62$ **(C)** $\hat{p} = 0.62$ **(D)** $p \neq 0.62$

3. What does a small P-value indicate?

   (A) The null hypothesis is very likely true.
   (B) The observed result would be unlikely if the null hypothesis were true.
   (C) The effect is large and important.
   (D) The sample was probably biased.

4. A test of H0: p = 0.30 gives P-value 0.04. Which is the correct interpretation?

   (A) The probability that H0 is true is 0.04.
   (B) There is a 96 percent chance that the alternative is true.
   (C) If the true proportion is 0.30, the probability of a sample result at least as extreme as the one observed is 0.04.
   (D) The probability that the result occurred by chance alone is 0.04.

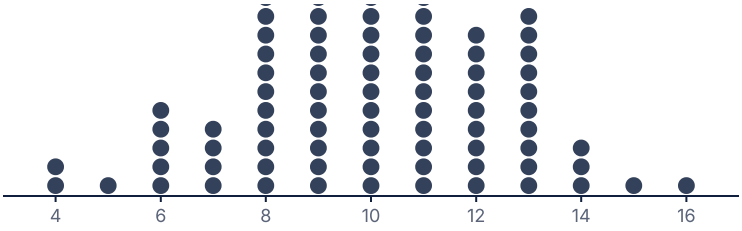
5. A test gives a P-value of 0.03. At the 0.05 significance level, what is the correct conclusion?

   (A) Fail to reject H0 because 0.03 is small.
   (B) Reject H0 and conclude H0 is proven false.
   (C) Accept H0 because 0.03 is close to zero.
   (D) Reject H0; there is convincing evidence for the alternative.

6. A test gives P-value 0.20 at alpha = 0.05. Which conclusion is best?

   (A) We do not have convincing evidence for the alternative; the data are consistent with H0.
   (B) We have shown that H0 is true.
   (C) We have convincing evidence for the alternative with 80 percent confidence.
   (D) We reject H0 because 0.20 is greater than 0.05.

7. The dot plot shows 100 simulated counts of correct answers from a guesser doing 20 trials. A student got 15 correct. What is the estimated P-value for the alternative that she does better than guessing?



Correct answers out of 20 in a simulated guesser

   (A) 0.15 **(B)** 0.98 **(C)** 0.02 **(D)** 0.20

8. Which situation calls for a two-sided alternative?

   (A) A researcher wants to know whether a new method lowers the defect rate.
   (B) A researcher wants to know whether the rate increased.
   (C) A researcher saw an increase in the sample and tests for an increase.
   (D) A researcher wants to know whether the proportion of defective parts differs from the stated 4 percent.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. For a two-sided test, z = -2.50. What is the P-value?
   (A) 0.0062 **(B)** 0.0124 **(C)** 0.9938 **(D)** 0.0250
10. Write the parameter, null hypothesis, and alternative hypothesis for each. (a) A coach claims 60 percent of the team's free throws go in; a player thinks it is lower. (b) A café says 25 percent of orders are iced drinks; a manager asks whether that is still accurate.
11. A test of $H_0 : p = 0.90$ versus $H_a : p < 0.90$ about the proportion of defect-free replacement screens gives P-value 0.018. (a) Interpret the P-value in context. (b) State the conclusion at $\alpha = 0.05$.
12. A student doing a simulation wants to see if a spinner is fair, with $H_0 : p = 0.25$ for landing on red and $H_a : p > 0.25$. In 120 simulated spinners where the true chance is 0.25, a sample of 40 spins gave a proportion of red as large as the observed one in 9 of the 120 runs. (a) Estimate the P-value. (b) State a conclusion at $\alpha = 0.05$.
13. A P-value is 0.032. (a) State the conclusion at $\alpha = 0.10$, 0.05, and 0.01. (b) Explain why α must be chosen before data are collected.
14. Critique each statement about a test with P-value 0.08 and $\alpha = 0.05$. (a) 'There is an 8 percent chance the null is true.' (b) 'We accept the null hypothesis.' (c) 'The result is not significant, so there is no effect.'
15. Two researchers examine the same data for $H_0 : p = 0.50$. One uses $H_a : p > 0.50$ and gets P-value 0.04 with the sample proportion above 0.50. The other uses $H_a : p \neq 0.50$. (a) What P-value does the second get? (b) At $\alpha = 0.05$, do they agree? (c) Why is it improper for the first researcher to choose the one-sided test after seeing the data?

7.4 One-Proportion z-Tests

Carry out the complete four-step one-proportion z-test: hypotheses, conditions based on the null value, test statistic, P-value, and a conclusion in context.

Why This Matters

Last lesson gave you the logic of a significance test. Now we turn the crank. A manufacturer says at most 8 percent of its parts are defective, but a quality inspector finds 11 percent in a sample. Is the gap evidence against the claim, or is it the kind of wobble you expect from sampling? The **one-proportion z-test** converts the gap into a number of standard errors and then into a P-value.

This is one of the most heavily tested procedures in Exam Unit 3. Free-response questions award separate points for hypotheses, for checking conditions, for the test statistic and P-value, and for a conclusion linked to the P-value in context. Students most often lose points by computing the standard error with $\hat{p}$ instead of p_0, by checking conditions with $\hat{p}$, or by writing a conclusion with no link to the P-value.

Core Ideas

A test begins by assuming the null hypothesis is true. If $H_0 : p = p_0$ is true, then the sampling distribution of $\hat{p}$ is centered at p_0 with standard deviation $\sqrt{p_0(1-p_0)/n}$. That is why the test statistic uses p_0, not $\hat{p}$, in the denominator: we are describing what a typical sample would look like *in the world where the null is true*.

One-Proportion z-Test: The Four Steps

State: Define the parameter p in context. Write $H_0 : p = p_0$ and H_a ($>$, $<$, or $\neq$). Give the significance level α (0.05 if none is given).

Plan: Name the procedure (one-proportion z-test). Check conditions: Random (random sample or randomized experiment); 10 percent ($n < 0.10N$ when sampling without replacement); Large counts ($np_0 \geq 10$ and $n(1-p_0) \geq 10$).

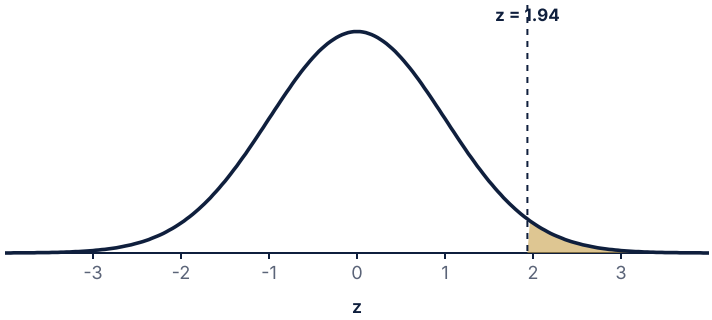
Do: Compute $\hat{p}$, then $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$, then the P-value from the normal curve (upper tail for $>$, lower tail for $<$, both tails for $\neq$).

Conclude: Compare the P-value with α, decide about H_0, and state a conclusion about H_a in context.

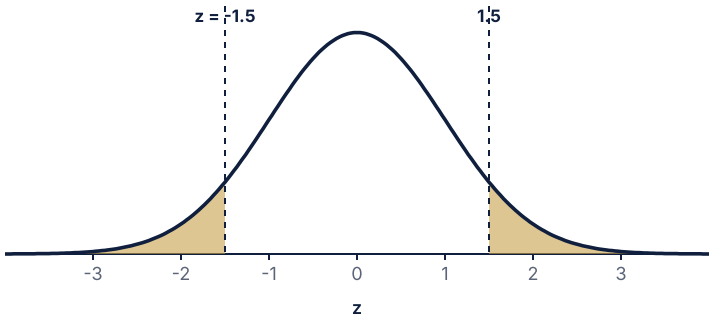
The z statistic is a standardized distance: it counts how many standard errors the observed $\hat{p}$ sits above or below the null value. A value near 0 means the sample agrees with the null; $|z|$ above 2 is the classic signal that the data are in the tails. The picture shows the P-value as a shaded tail area. For a **one-sided** alternative we shade one tail. For a **two-sided** alternative we shade both tails because a result equally far in the other direction would have been equally convincing.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Upper-tailed test: P-value = 0.0264.



Two-sided test: both tails beyond |z| = 1.5, P-value = 0.1336.

Alternative	P-value	Calculator tail
$H_a: p > p_0$	$P(Z \geq z)$	Upper tail: 1 - normalcdf(-big, z) or normalcdf(z, big)
$H_a: p < p_0$	$P(Z \leq z)$	Lower tail: normalcdf(-big, z)
$H_a: p \neq p_0$	$2P(Z \geq \lvert z \rvert)$	Double the smaller tail

Why the Conditions Use p0

In a test we assume H_0 is true, so the model for the sampling distribution uses p_0. The large counts check is $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.

In an interval, there is no assumed value, so large counts use $\hat{p}$ in place of p. Mixing these up is a classic error.

Intervals and two-sided tests. A C percent confidence interval and a two-sided test at $\alpha = 1 - C/100$ carry the same information. If the null value p_0 falls *outside* the C percent interval, the two-sided test at $\alpha = 1 - C$ rejects H_0; if p_0 is *inside*, it fails to reject. The match is close but not exact for proportions, because the interval uses $\hat{p}$ in the standard error while the test uses p_0. The interval has a bonus: it also shows how large the difference might be. A one-sided test at level α corresponds to a confidence interval with level $1 - 2\alpha$.

Reading the numbers. Suppose a test gives $z = 1.94$. Three questions should pop into your head. First, which direction? Positive means the sample proportion is above the claimed value. Second, how far? Nearly two standard errors, which is unusual but not outrageous. Third, which tail or tails does the alternative call for? If $H_a: p > p_0$ we want the upper tail only; the area is about 0.026. If $H_a: p \neq p_0$ we double it to 0.053. The same data, a different research question, a different P-value. This is why the hypotheses are written in the State step before any arithmetic is done.

What each condition protects. The random condition is the foundation: a biased sample invalidates everything that follows, and no calculation will repair it. The 10 percent condition guards independence: when we sample without replacement from a small population, each selection changes the makeup of what remains, and the true variability of $\hat{p}$ is smaller than the formula claims. The large counts condition guards the normal model: with fewer than 10 expected successes or failures, the sampling distribution of $\hat{p}$ is visibly skewed and tail areas from the normal curve are not accurate.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Step	What to write	Points a grader can award
State	Parameter in words, H_0, H_a, α	Hypotheses correct and in terms of p
Plan	Procedure name and three conditions with numbers	Procedure plus all conditions verified
Do	$\hat{p}$, z, P-value, a labeled sketch if helpful	Correct statistic and correct tail(s)
Conclude	Compare P-value with α, decision, context	Decision linked to the P-value and stated in context

Using the built-in calculator. The built-in calculator offers a 1-proportion z-test. You enter p_0, the number of successes x, the sample size n, and the direction of the alternative, and it returns z and the P-value. Treat this as a check, not a replacement. On the free-response section, writing only a calculator output earns little; you must show the formula or the named procedure, the values substituted, and the interpretation. On multiple-choice questions, the calculator is the fastest route, but you still must pick the correct tail.

A note on rounding. Carry at least four decimal places of the standard error through the calculation and round z to two decimals at the end, since normal tables give areas for two-decimal z values. A P-value of 0.0264 and one of 0.0268 will both lead to the same decision, so do not worry about the last digit; worry about using the correct tail and the correct standard error.

Reading Calculator Output

Exam questions sometimes print the output of a one-proportion test instead of asking you to compute it. Learn to read each line. The table below is output for the defect-rate test from the worked examples. Notice that it reports the null value, the sample proportion, the standard error computed with the null value, the z statistic, and the P-value for the stated alternative. Everything you need for the Do step is there; your job is to turn it into a conclusion.

Output line	Value	Meaning
Null hypothesis	p = 0.08	The claimed defect rate p_0
Alternative	p > 0.08	Upper-tailed test
Sample proportion	0.11	33 defective out of 300
Standard error	0.01566	Computed from p_0 and n
z statistic	1.92	About 1.9 standard errors above the claim
P-value	0.0277	Upper-tail area beyond z

When you read output like this, check three things before trusting it: the alternative direction matches the research question, the sample proportion matches the counts, and the conditions were verified (software does not check them for you). Then write the interpretation: if the true defect rate were 8 percent, a sample proportion of 11 percent or higher would happen in about 2.8 percent of random samples of 300 parts.

Sample size and what a result means. The same $\hat{p}$ gives very different evidence at different sample sizes. A sample proportion of 0.58 against a claim of 0.55 gives $z = 1.48$ when $n = 600$ but only $z = 0.65$ when $n = 120$. Larger samples have smaller standard errors, so the same difference is more surprising. When a test fails to reject, ask whether the sample was large enough to detect a difference that matters; that question leads directly to the idea of power in the next lesson.

Why a z and not a t? The standard deviation of $\hat{p}$ is determined by p_0 and n once the null is assumed, so no extra quantity has to be estimated from the data. That is why proportions use the normal curve directly. For means in the next unit, the population standard deviation is unknown and must be estimated, which brings in t-distributions. Keeping this contrast in mind will save you from reaching for the wrong tool.

A worked mini-narrative. Imagine the bakery's claim again. If it is really true that 70 percent of customers choose whole grain, then sample proportions from random samples of 120 customers will cluster around 0.70 with a standard error of about 0.042. Our sample came in at 0.617, which is about two standard errors low. Only about 4.6 percent of samples would be that far from 0.70 in either direction, so either the bakery was unlucky in this sample or its claim is overstated. That is the whole argument of a significance test in two sentences.

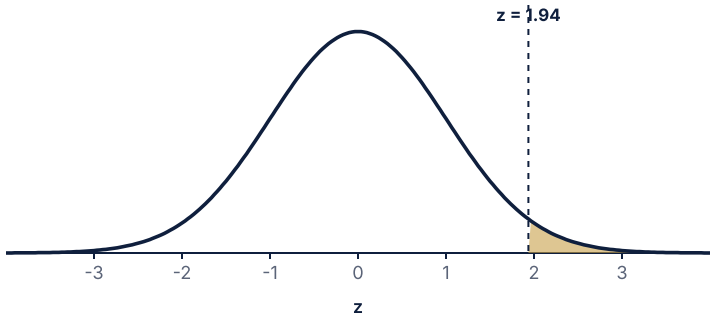
Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

COMPUTING THE TEST STATISTIC AND P-VALUE

A company claims that 60 percent of its customers renew their subscription. In a random sample of 250 customers (far fewer than 10 percent of all customers), 165 renewed. Find the test statistic and the P-value for $H_a : p > 0.60$.



P-value = 0.0264.

SOLUTION

$\hat{p} = 165/250 = 0.66$ and the standard error under H_0 is $\sqrt{0.60(0.40)/250} \approx 0.03098$.

$z = (0.66 - 0.60)/0.03098 \approx 1.94$.

Upper-tail area: $P(Z \geq 1.94) \approx 0.0264$. If the alternative had been two-sided, the P-value would be $2(0.0264) = 0.0528$.

CHECKING CONDITIONS WITH P0

A school nurse samples 60 of the 900 students at random and tests $H_0 : p = 0.70$ about the proportion who get at least 8 hours of sleep. Check the conditions. Then repeat the large counts check for $H_0 : p = 0.02$ with $n = 200$.

SOLUTION

Random: stated. 10 percent: $60 < 90$. Large counts: $np_0 = 60(0.70) = 42 \geq 10$ and $n(1 - p_0) = 60(0.30) = 18 \geq 10$. All conditions hold, so a z-test is appropriate.

For $p_0 = 0.02$, $n = 200$: $np_0 = 200(0.02) = 4 < 10$. The large counts condition fails, so the normal model is not reliable and the z-test should not be used. Expected successes under the null are too few.

A COMPLETE ONE-SIDED TEST

A plant claims at most 8 percent of its parts are defective. An inspector randomly selects 300 parts from a very large batch and finds 33 defective. Is there convincing evidence that the defect rate exceeds 8 percent? Use $\alpha = 0.05$.

Quantity	Value
p-hat	0.11
Standard error under H0	0.01566
z	1.92
P-value	0.0277

SOLUTION

State. Let p be the true proportion of defective parts. $H_0 : p = 0.08$, $H_a : p > 0.08$, $\alpha = 0.05$.

Plan. One-proportion z-test. Random: the parts were randomly selected. 10 percent: 300 is much less than 10 percent of a very large batch. Large counts: $300(0.08) = 24 \geq 10$ and $300(0.92) = 276 \geq 10$.

Do. $\hat{p} = 33/300 = 0.11$. $z = (0.11 - 0.08)/\sqrt{0.08(0.92)/300} = 0.03/0.01566 \approx 1.92$. P-value $= P(Z \geq 1.92) \approx 0.0277$.

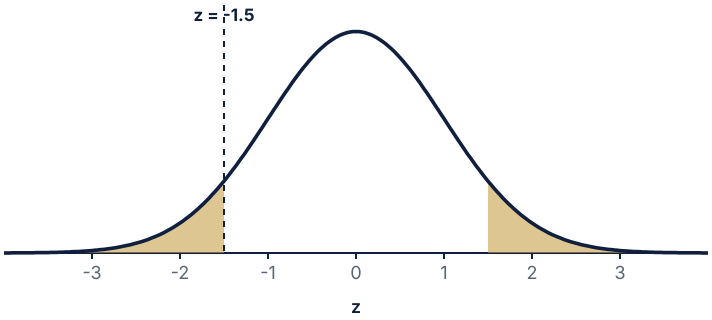
Conclude. Because $0.0277 < 0.05$, we reject H_0. We have convincing evidence that the true proportion of defective parts exceeds 0.08.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A COMPLETE TWO-SIDED TEST

A club says that 40 percent of students at a large school ride a bicycle at least once a week. A random sample of 150 students finds 51 who do. Test whether the true proportion differs from 40 percent at $\alpha = 0.05$.



Two-tailed P-value = 0.1336.

SOLUTION

State. p = true proportion of students at this school who ride a bicycle weekly. $H_0 : p = 0.40$, $H_a : p \neq 0.40$, $\alpha = 0.05$.

Plan. One-proportion z-test. Random sample; 150 is less than 10 percent of a large school; $150(0.40) = 60$ and $150(0.60) = 90$ are at least 10.

Do. $\hat{p} = 51/150 = 0.34$; $z = (0.34 - 0.40)/\sqrt{0.40(0.60)/150} = -0.06/0.04 = -1.50$. P-value $= 2P(Z \leq -1.50) = 2(0.0668) = 0.1336$.

Conclude. Since $0.1336 > 0.05$, we fail to reject H_0. We do not have convincing evidence that the true proportion differs from 0.40.

HOW MANY SUCCESSES ARE NEEDED?

In a random sample of 100, a test of $H_0 : p = 0.50$ against $H_a : p > 0.50$ is run at $\alpha = 0.05$. What is the smallest number of successes that leads to rejecting H_0?

SOLUTION

The standard error under H_0 is $\sqrt{0.5(0.5)/100} = 0.05$, so $z = (x/100 - 0.5)/0.05$.

For $x = 58$: $z = 1.60$ and the P-value is $0.0548 > 0.05$, so we do not reject.

For $x = 59$: $z = 1.80$ and the P-value is $0.0359 < 0.05$, so we reject.

The smallest count is 59. Equivalently, $\hat{p}$ must exceed $0.50 + 1.645(0.05) = 0.582$.

CONNECTING THE TEST TO AN INTERVAL

Using the bicycle data above ($\hat{p} = 0.34$, $n = 150$), build a 95 percent confidence interval and explain how it agrees with the two-sided test.

SOLUTION

$\text{SE} = \sqrt{0.34(0.66)/150} \approx 0.0387$ and $\text{ME} = 1.960(0.0387) \approx 0.0758$. The interval is (0.264, 0.416).

The interval contains 0.40, so 0.40 is a plausible value of p. That agrees with failing to reject $H_0 : p = 0.40$ at $\alpha = 0.05$.

The interval adds information the test cannot: the true proportion could plausibly be anywhere from about 26 percent to 42 percent.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt. A bakery claims that 70 percent of its customers choose a whole-grain loaf. A food club randomly samples 120 of the bakery's many customers and finds 74 chose whole grain. Do the data provide convincing evidence that the true proportion differs from 0.70? Use $\alpha = 0.05$.\n\n**State.** Let p be the true proportion of the bakery's customers who choose a whole-grain loaf. $H_0 : p = 0.70$ and $H_a : p \neq 0.70$, with $\alpha = 0.05$.\n\n**Plan.** One-proportion z-test. Random: customers were randomly selected. 10 percent: 120 is less than 10 percent of the bakery's many customers. Large counts: $120(0.70) = 84 \geq 10$ and $120(0.30) = 36 \geq 10$.\n\n**Do.** $\hat{p} = 74/120 \approx 0.617$. $z = (0.617 - 0.70)/\sqrt{0.70(0.30)/120} = -0.0833/0.04183 \approx -1.99$. P-value $= 2P(Z \leq -1.99) \approx 0.0464$.\n\n**Conclude.** Because $0.0464 < 0.05$, we reject H_0. We have convincing evidence that the true proportion of this bakery's customers who choose a whole-grain loaf differs from 0.70.

What the grader looks for

(1) Parameter and both hypotheses correct, with the right alternative. (2) Procedure named and all three conditions verified with numbers using p_0. (3) Correct test statistic with p_0 in the standard error. (4) Correct P-value for a two-sided alternative. (5) Conclusion that compares the P-value with α, rejects or fails to reject H_0, and states the result in context.

Coach's Playbook

1. Underline the claimed value in the problem: that is p_0. Underline the direction words: they set H_a.
2. Compute the standard error from p_0 and n. Never use $\hat{p}$ here.
3. Sketch the normal curve, mark z, and shade the tail(s). The picture prevents the wrong-tail error.
4. Use the built-in calculator's 1-proportion z-test to check the z and P-value after you have computed by hand.
5. Write the conclusion with a P-value link: 'Because the P-value of ___ is less (greater) than ___, we ...'

Common traps

1. Using $\hat{p}$ in the standard error of the test statistic. 2. Checking large counts with $\hat{p}$ instead of p_0. 3. Forgetting to double the tail area for a two-sided alternative. 4. A conclusion that says only 'reject H_0' with no context.

- **Sanity check the sign.** If $\hat{p} < p_0$ then z must be negative. A positive z with a smaller sample proportion signals an arithmetic slip.
- **Match the tail to the alternative.** Greater than means upper tail, less than means lower tail, not equal means double the smaller tail.
- **Large P-value, small effect?** Ask whether a larger sample might have shown the difference to be real. A fail-to-reject result is not evidence of no difference.
- **Link to the exam.** In Exam Unit 3 free-response questions, the conclusion point is lost most often because the P-value is not compared with α or the context is missing.

Mistake Clinic

Spot the error

Wrong solution. 'To test $H_0 : p = 0.25$ with $\hat{p} = 0.30$ and $n = 400$, $z = (0.30 - 0.25)/\sqrt{0.30(0.70)/400} = 2.18$.'\n\n**Where it fails.** The standard error must use $p_0 = 0.25$ because the test assumes H_0 is true.\n\n**Correct.** $\text{SE} = \sqrt{0.25(0.75)/400} = 0.02165$ and $z = 0.05/0.02165 \approx 2.31$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Wrong solution. 'For $H_a: p \neq 0.50$ the test statistic is $z = 1.85$, so the P-value is $P(Z \geq 1.85) = 0.0322$, and since $0.0322 < 0.05$ we reject.'\n\n**Where it fails.** A two-sided alternative needs both tails. The P-value is doubled.\n\n**Correct.** $P = 2(0.0322) = 0.0643 > 0.05$, so we fail to reject H_0.

Spot the error

Wrong solution. 'The P-value is 0.0643, which is greater than 0.05, so we accept H_0 and conclude the proportion equals 0.50.'\n\n**Where it fails.** A large P-value does not prove the null. The conclusion also fails to refer to the context.\n\n**Correct.** Fail to reject H_0. We do not have convincing evidence that the true proportion differs from 0.50.

Real-World Connection

Quality teams run one-proportion tests every day: Is the defect rate above its limit? Has the opt-in rate on a sign-up form changed after a redesign? In each case the claim is a fixed value p_0 from an earlier measurement, and the sample says how far today's result has strayed. The test statistic tells the team how many standard errors that stray is, and the P-value tells them how surprising it would be if nothing had changed. A team that finds a significant result then asks the follow-up question: is the change large enough to matter for cost, safety, or customers?

Challenge Corner

A coach claims that a player makes 75 percent of free throws. In a random sample of 80 throws the player makes 52. Run the test with $H_a: p < 0.75$ and then explain why the large counts condition passes but is worth checking carefully.\n\n**Solution.** $\hat{p} = 52/80 = 0.65$. $\text{SE} = \sqrt{0.75(0.25)/80} = 0.04841$, so $z = (0.65 - 0.75)/0.04841 \approx -2.07$, giving a lower-tail P-value of about 0.0194. Reject at $\alpha = 0.05$. Large counts: $80(0.75) = 60$ and $80(0.25) = 20$, both at least 10. The failures count, 20, is the one that comes closest to the cutoff, which is why we check both counts, not just the larger one. Now add a confidence-interval check: $\hat{p} = 0.65$ with $\text{SE} = \sqrt{0.65(0.35)/80} = 0.0533$ gives a 90 percent one-sided-equivalent interval $0.65 \pm 1.645(0.0533)$, or (0.562, 0.738). The value 0.75 lies above this interval, which agrees with rejecting a one-sided test at 0.05.

Practice

PRACTICE **One-Proportion z-Tests**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A test of $H_0: p = 0.25$ uses a random sample of $n = 400$. What is the standard error used in the test statistic?
   (A) 0.0012 **(B)** 0.0462 **(C)** 0.0004 **(D)** 0.0217
2. Test $H_0: p = 0.25$ with $\hat{p} = 0.30$ and $n = 400$. What is the test statistic?
   (A) 2.31 **(B)** 2.18 **(C)** 0.05 **(D)** 1.15
3. For $H_a: p > 0.25$ and $z = 2.31$, what is the P-value?
   (A) 0.0209 **(B)** 0.0105 **(C)** 0.9895 **(D)** 2.31
4. For $H_a: p \neq 0.25$ and $z = 2.31$, what is the P-value?
   (A) 0.0105 **(B)** 0.9791 **(C)** 0.0209 **(D)** 0.0420
5. Which condition is used to check the large counts requirement in a one-proportion z-test of $H_0: p = p_0$?
   (A) $np_0 \geq 10$ and $n(1 - p_0) \geq 10$ **(B)** $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$ **(C)** $n \geq 30$ **(D)** $np_0 \geq 5$ only

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. A 95 percent confidence interval for a proportion is (0.52, 0.60). For a two-sided test of H0: p = 0.50 at alpha = 0.05, what is the appropriate conclusion?
   (A) Fail to reject H0, because 0.50 is close to the interval.
   (B) Reject H0, because the interval does not contain p-hat.
   (C) Fail to reject H0, because the interval is narrow.
   (D) Reject H0, because 0.50 is outside the interval.
7. In a test, a student computes z using p-hat = 0.34 in the denominator and p0 = 0.40 in the numerator. Which describes the error?
   (A) The standard error should be computed with p0, not p-hat.
   (B) The numerator should use p-hat in both places.
   (C) The test statistic should be divided by n.
   (D) Nothing, this is the correct formula.
8. A test has P-value 0.0105 for H_a: p > 0.25 about the proportion of users who adopt a feature. Which conclusion at alpha = 0.05 is best?
   (A) The true proportion is exactly 0.30.
   (B) There is a 1.05 percent chance that H0 is true.
   (C) There is convincing evidence that the true proportion of users who adopt the feature exceeds 0.25.
   (D) We fail to reject H0 because 0.0105 is small.
9. A researcher plans to test H0: p = 0.05 with a random sample of n = 150 from a large population. Which is true?
   (A) All conditions are met because n is larger than 100.
   (B) The large counts condition fails because 150(0.05) = 7.5 is less than 10.
   (C) The 10 percent condition fails because 0.05 is small.
   (D) The random condition fails because n is not a multiple of 100.
10. A cafeteria manager claims that 40 percent of students buy lunch. A random sample of 400 students (from a population of 6,000) finds that 170 bought lunch. Is there convincing evidence that the proportion differs from 0.40? Use $\alpha = 0.05$.
11. A music app says that 55 percent of its users keep notifications on. In a random sample of 600 users (far fewer than 10 percent of all users), 348 keep them on. Test whether the proportion is greater than 0.55 at $\alpha = 0.05$.
12. A transit agency claims at least 80 percent of its trains arrive on time. An auditor randomly selects 250 trains from thousands and finds 184 on time. Do the data give convincing evidence that fewer than 80 percent arrive on time? Use $\alpha = 0.05$.
13. A 95 percent confidence interval for the proportion of voters in a town who support a bond is (0.52, 0.60). (a) Use the interval to decide a two-sided test of $H_0 : p = 0.50$ at $\alpha = 0.05$. (b) Do the same for $H_0 : p = 0.55$. (c) What does the interval tell us that the test result does not?
14. A random sample of 200 from a large population is used to test $H_0 : p = 0.30$ against $H_a : p > 0.30$ at $\alpha = 0.05$. Find the smallest number of successes that leads to rejecting H_0.
15. A student tests $H_0 : p = 0.50$ with $\hat{p} = 0.56$ and $n = 150$, writing 'z = (0.56-0.50)/sqrt(0.56(0.44)/150) = 1.48, P-value = 0.0694, so we accept H0 because 0.0694 > 0.05.' Identify and fix all errors. Then state the correct conclusion for a two-sided alternative.

7.5 Errors, Significance, and Power

Type I and Type II errors in context, their consequences, how the significance level relates to them, and what raises the power of a test.

Why This Matters

Tests are decisions, and decisions can be wrong. A fire alarm that never rings is dangerous; one that rings every day gets ignored. A significance test faces the same dilemma: reject a true null and you raise a false alarm, fail to reject a false null and you miss a real problem. Knowing which mistake is worse in a given context tells you how cautious your test should be.

The vocabulary here (Type I error, Type II error, significance level, power) appears in Exam Unit 3 mostly in conceptual form. Multiple-choice items ask you to match the definition or consequence to a scenario, and free-response items ask you to describe an error in context and explain its consequence. Students lose points by stating errors abstractly ('rejecting a true null') without tying them to the story.

Core Ideas

There are exactly two ways a test can go wrong. We never know whether we made one, but we can study the chances.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

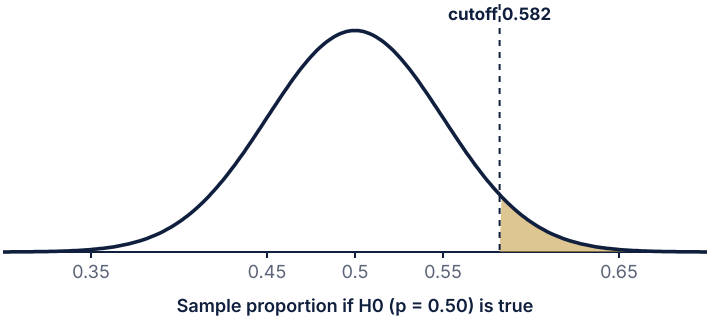
	Truth: H_0 is true	Truth: H_0 is false (H_a true)
Decision: reject H_0	**Type I error** (false positive), probability α	Correct decision, probability = power
Decision: fail to reject H_0	Correct decision, probability $1-\alpha$	**Type II error** (false negative), probability β

Errors, Alpha, Beta, and Power

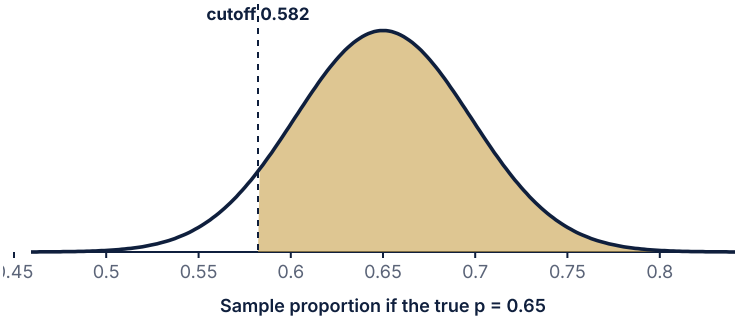
Type I error: rejecting H_0 when H_0 is true. $P(\text{Type I}) = \alpha$, the significance level.
Type II error: failing to reject H_0 when H_0 is false. Its probability β depends on which alternative value is true.
Power: the probability that the test correctly rejects H_0 when a specific alternative is true. Power $= 1 - \beta$.
In a written answer, describe each error in the context of the problem, not as a bare definition.

Notice the logic. If H_0 is really true, a test at $\alpha = 0.05$ will still reject it 5 percent of the time, by pure sampling variability. That is not a flaw; it is the price we agreed to pay when we set α. If we run 10 independent true-null tests at $\alpha = 0.05$, the probability of at least one false rejection is $1 - 0.95^{10} \approx 0.40$, which is why researchers who test many hypotheses must be wary of 'significant' findings.

The two curves below show where α and power come from in a one-sided test of $H_0 : p = 0.50$ with $n = 100$ and $\alpha = 0.05$. The test rejects whenever $\hat{p} > 0.582$, because that is the cutoff leaving 5 percent in the upper tail of the null curve, $0.50 + 1.645(0.05) = 0.582$. If the truth is actually $p = 0.65$, the sampling distribution of $\hat{p}$ moves to the right, and most of its area falls beyond the cutoff.



Shaded area beyond the cutoff is alpha = 0.05, the probability of a Type I error.



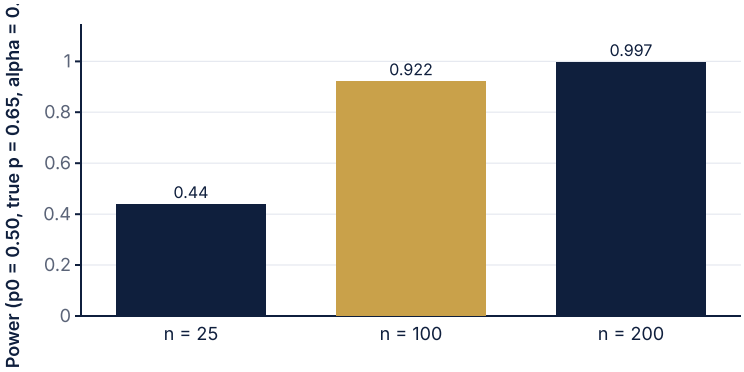
Shaded area beyond the cutoff is the power, about 0.922; the unshaded left area is beta, about 0.078.

What Increases Power

Larger sample size n: smaller standard error, curves separate. (The most practical lever.)
Larger significance level α: the cutoff moves toward the center, but Type I error probability rises.
True parameter farther from p_0: a bigger effect is easier to detect.
Less variability: for proportions, a true p nearer 0 or 1 gives smaller $p(1-p)$, though in practice n is what you control.
Lowering α lowers power unless n rises. For a fixed n and p, α and β trade off.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



More data, more power.

Choosing which error to guard against. Context decides. If H_0 says a bridge design is safe, a Type II error (failing to detect an unsafe design) is the dangerous one, and we might accept a larger α to gain power. If H_0 says a defendant is innocent, a Type I error (convicting an innocent person) is the graver one, and we choose a small α. There is no formula for this decision; it is a judgment about consequences, and the AP exam rewards a clear argument about the context.

Statistical significance and the number of tests. Because α is the long-run false-alarm rate when nothing is going on, a study that reports 1 significant result out of 20 true-null tests is entirely consistent with chance. Power has a similar caution: a failed test with low power means little, because even a real effect would likely have been missed. Always ask whether the sample was large enough to detect a difference that would matter.

Worked Examples

DESCRIBING ERRORS IN CONTEXT

A bottling plant tests $H_0 : p = 0.02$ against $H_a : p > 0.02$, where p is the proportion of bottles that are underfilled. Describe a Type I error and a Type II error and a consequence of each.

SOLUTION

Type I: We conclude that more than 2 percent of bottles are underfilled when the true rate is 2 percent. Consequence: the line is stopped and adjusted unnecessarily, wasting time and money.

Type II: We fail to find convincing evidence that the rate exceeds 2 percent when it truly does. Consequence: underfilled bottles keep shipping to customers, risking complaints and lost trust.

WHICH ERROR IS WORSE?

A city tests H_0: the water in a school's fountains meets the safety standard, against H_a: it does not. Describe each error and decide which has more serious consequences.

SOLUTION

Type I: concluding the water is unsafe when it is safe. Consequence: fountains are shut and repaired needlessly.

Type II: failing to conclude the water is unsafe when it is. Consequence: students drink unsafe water.

The Type II error is more serious for health, so the city should favor a test with high power, even if that means accepting a larger α such as 0.10.

PROBABILITY OF A TYPE I ERROR

A lab runs 10 independent tests at $\alpha = 0.05$, and in every case the null hypothesis is true. What is the probability of at least one Type I error? How many do we expect?

SOLUTION

Each test has a 0.05 chance of a false rejection. $P(\text{at least one}) = 1 - (0.95)^{10} \approx 0.401$.

The expected number is $10(0.05) = 0.5$.

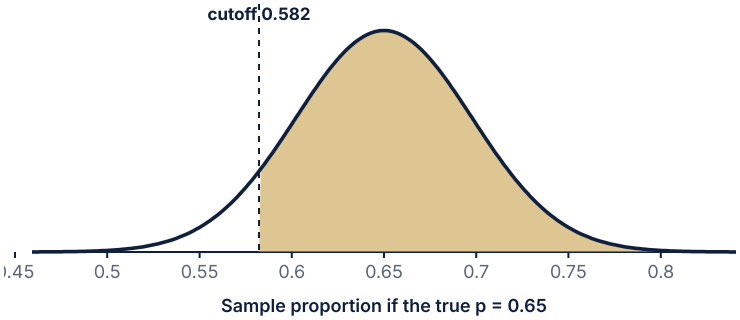
So there is about a 40 percent chance that the lab will record a 'significant' finding even though nothing is going on in any test. The lesson: α applies to each test, not to the whole collection.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

COMPUTING POWER

A one-sided test of $H_0 : p = 0.50$ against $H_a : p > 0.50$ uses $n = 100$ and $\alpha = 0.05$. Find the power when the true proportion is 0.65.



Shaded area beyond the cutoff is the power, about 0.922; the unshaded left area is beta, about 0.078.

SOLUTION

Step 1: cutoff. Reject when $\hat{p} > 0.50 + 1.645\sqrt{0.5(0.5)/100} = 0.582$.

Step 2: shift to the truth. If $p = 0.65$, $\hat{p}$ is approximately normal with mean 0.65 and SD $\sqrt{0.65(0.35)/100} \approx 0.0477$.

Step 3: area. Power $= P(\hat{p} > 0.582) = P\left(Z > \dfrac{0.582 - 0.65}{0.0477}\right) = P(Z > -1.42) \approx 0.922$.

So $\beta \approx 0.078$: about a 7.8 percent chance of missing the effect.

WHAT CHANGES POWER?

For the same test as above (true $p = 0.65$ unless stated), the table shows power under different settings. Explain each pattern.

Setting	n	alpha	True p	Power
Baseline	100	0.05	0.65	0.922
Larger sample	200	0.05	0.65	0.997
Smaller alpha	100	0.01	0.65	0.760
Smaller effect	100	0.05	0.60	0.641
Smaller sample	25	0.05	0.65	0.440

SOLUTION

Increasing n from 100 to 200 raises power from 0.922 to 0.997, because the standard errors shrink and the curves overlap less.

Lowering α from 0.05 to 0.01 lowers power from 0.922 to 0.760: a stricter cutoff makes it harder to reject, even when the null is false.

Moving the true value closer to the null ($p = 0.60$) lowers power to 0.641, because small effects are harder to detect than large ones.

TRADEOFF BETWEEN ALPHA AND BETA

A quality engineer tests $H_0 : p = 0.50$ with $n = 100$ and decides to lower α from 0.05 to 0.01 to avoid false alarms. Explain what happens to β when the true $p = 0.65$, and name two ways to restore power.

SOLUTION

At $\alpha = 0.01$ the cutoff is $0.50 + 2.326(0.05) = 0.616$, so power is $P(Z > (0.616 - 0.65)/0.0477) = P(Z > -0.71) \approx 0.760$. Thus β rises from about 0.078 to 0.240.

Lowering α reduces false alarms but raises the chance of missing a real shift. Two ways to restore power: increase the sample size n, or accept a larger α (a less strict cutoff). A third way, available only in design work, is to reduce variability in measurement.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt. A website team tests H_0: the proportion of visitors who click a new banner is 0.10, against H_a: the proportion is greater than 0.10. (a) Describe a Type I error and a Type II error in context. (b) Which would be more costly if adopting the banner requires a large redesign, and why? (c) State two ways to increase the power of the test.\n\n**(a)** A Type I error is concluding that the proportion of visitors who click the new banner is greater than 0.10 when it is really 0.10. A Type II error is failing to find convincing evidence that the proportion is greater than 0.10 when it really is greater.\n\n**(b)** The Type I error is more costly here: the team would spend a lot on a redesign that does not actually bring more clicks. The Type II error would only forgo a possible benefit, so a small significance level such as 0.01 is reasonable.\n\n**(c)** Power increases if the team collects data from more visitors (larger n), or if it uses a larger significance level
$alpha$ (accepting more risk of a Type I error).

What the grader looks for

(1) Type I error described as rejecting a true null, in the context of the clicks. (2) Type II error described as failing to reject a false null, in context. (3) A consequence tied to the story, with a reason for which is worse. (4) Two valid ways to raise power, with the tradeoff for increasing α mentioned or implied. (5) Language is specific to the banner, not generic.

Coach's Playbook

1. Write H_0 and H_a in words first; errors are then easy to describe.
2. Type I: the test says 'yes, something is going on' but nothing is. Type II: the test says 'not convinced' but something is going on.
3. Add a consequence sentence: what happens in the real world because of the error?
4. Remember α is chosen by you, β and power depend on the true value and the sample size.
5. For power questions, think of the two curves: the null curve fixes the cutoff, the alternative curve gives the power.

Common traps

1. Describing a Type I error as 'a wrong decision' without saying which decision. 2. Confusing power with the P-value or with α. 3. Believing that decreasing α is always good, ignoring the rise in β. 4. Interpreting a failure to reject as evidence the null is true, forgetting that low power could be the reason.

Mistake Clinic

Spot the error

Wrong solution. 'The P-value is 0.03, so the probability of a Type I error is 0.03.'\n\n**Where it fails.** The probability of a Type I error is the significance level α set in advance, not the P-value computed from data.\n\n**Correct.** If $\alpha = 0.05$ then $P(\text{Type I}) = 0.05$ whenever H_0 is true; the P-value only measures the strength of evidence in the sample.

Spot the error

Wrong solution. 'We failed to reject H_0, so we cannot have made an error.'\n\n**Where it fails.** A decision to fail to reject can still be wrong if H_0 is false; that is exactly a Type II error.\n\n**Correct.** If the null is actually false, failing to reject it is a Type II error. We cannot know which error occurred, only the chance of each.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Wrong solution. 'To get more power, change α from 0.05 to 0.01.'\n\n**Where it fails.** A smaller α moves the cutoff farther from the null and lowers power.\n\n**Correct.** Raise α (accepting more Type I risk) or, better, raise the sample size.

Real-World Connection

Medical screening illustrates both errors: a false positive (Type I) sends a healthy person for follow-up testing, while a false negative (Type II) lets disease go undetected. Test designers pick cutoffs by weighing those costs. Clinical trials are planned with a target power such as 80 or 90 percent for the smallest effect worth detecting, and the sample size is chosen to meet it.

Challenge Corner

For the test $H_0 : p = 0.30$ versus $H_a : p > 0.30$ with $\alpha = 0.05$, find the power when the true $p = 0.40$ for $n = 80$ and for $n = 320$.\n\n**Solution.** For $n = 80$: cutoff $= 0.30 + 1.645\sqrt{0.3(0.7)/80} = 0.3843$. The alternative SD is $\sqrt{0.4(0.6)/80} = 0.0548$, so power $= P(Z > (0.3843 - 0.40)/0.0548) = P(Z > -0.29) \approx 0.613$. For $n = 320$: cutoff $= 0.30 + 1.645(0.02562) = 0.3421$; alternative SD $= 0.02739$; power $= P(Z > -2.11) \approx 0.983$. Quadrupling n raised the power from about 61 percent to about 98 percent.

Practice

PRACTICE **Errors, Significance, and Power**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A Type I error occurs when we
   - **(A)** fail to reject a null hypothesis that is false.
   - **(B)** reject a null hypothesis that is true.
   - **(C)** reject a null hypothesis that is false.
   - **(D)** fail to reject a null hypothesis that is true.
2. A school tests H0: the new schedule does not change the on-time rate, against Ha: it increases the on-time rate. Which describes a Type II error?
   - **(A)** Concluding the schedule increases the on-time rate when it really does not.
   - **(B)** Concluding the on-time rate decreased.
   - **(C)** Concluding there is no convincing evidence of an increase when the new schedule really does increase the rate.
   - **(D)** Rejecting H0 with a P-value below alpha.
3. If a test is conducted at alpha = 0.05 and H0 is true, what is the probability of making a Type I error?
   (A) 0.05 **(B)** 0.95 **(C)** The P-value
   (D) 0.50
4. Which of the following would increase the power of a test, holding everything else constant?
   - **(A)** Decreasing the significance level
   - **(B)** Choosing a true parameter value closer to the null value
   - **(C)** Using a smaller sample to reduce variability
   - **(D)** Increasing the sample size
5. The power of a test against a specific alternative is 0.82. What is the probability of a Type II error for that alternative?
   (A) 0.82 **(B)** 0.05 **(C)** 0.18 **(D)** 0.90

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. Lowering the significance level from 0.05 to 0.01 (same n, same true value) will
   (A) decrease both error probabilities.
   (B) decrease the Type I error probability and increase the Type II error probability.
   (C) increase the power.
   (D) increase the Type I error probability.
7. A drug company tests H0: a new medicine is no better than the current one. A Type I error would mean
   (A) concluding the new medicine is no better when it is actually better.
   (B) concluding the current medicine is better.
   (C) failing to run the study.
   (D) concluding the new medicine is better when it is not.
8. For a test of H0: p = 0.5 with n = 100, the power is higher if the true p is 0.65 than if it is 0.55 because
   (A) a larger difference from the null value is easier to detect.
   (B) the sample size is larger for p = 0.65.
   (C) alpha is larger for p = 0.65.
   (D) the null hypothesis is different.
9. A researcher runs 40 independent tests, each at alpha = 0.05, and all null hypotheses are true. About how many Type I errors are expected?
   (A) 0 **(B)** 1 **(C)** 2 **(D)** 5
10. A site tests whether the proportion of visitors who click a new button exceeds the current 0.10. (a) Describe Type I and Type II errors in context. (b) Describe a consequence of each.
11. A one-sided test of $H_0 : p = 0.20$ versus $H_a : p > 0.20$ has $n = 150$ and $\alpha = 0.05$. Find the power when the true $p = 0.30$.
12. A test has power 0.61 against a particular alternative. (a) What is the probability of a Type II error? (b) Give two changes that would raise the power, and state the cost of each.
13. A county tests H_0: a river's bacteria level meets the safe standard, against H_a: it exceeds the standard. (a) Describe each error and its consequence. (b) Which error should be guarded against more strongly, and how can the county do that?
14. For $H_0 : p = 0.30$ versus $H_a : p > 0.30$ with $\alpha = 0.05$ and true $p = 0.40$, the power is about 0.61 when $n = 80$ and about 0.98 when $n = 320$. Explain in terms of sampling distributions why the larger sample has more power.
15. A researcher failed to reject H_0 in a pilot study with $n = 25$ and concluded, 'There is no effect.' (a) Explain what is wrong. (b) For $H_0 : p = 0.50$ and true $p = 0.65$, power with $n = 25$ is only 0.44. Use this to argue the pilot's conclusion is weak. (c) What should the researcher do?

7.6 Two-Proportion z-Intervals and Tests

Compare two proportions with a z-interval and a pooled z-test, check conditions for each sample, and know when a difference supports cause and effect.

Why This Matters

So far every inference compared one sample with a fixed claim. Real questions are usually comparisons: Do teens in one city stream more music than teens in another? Did a redesigned app raise the share of users who finish onboarding? A **two-proportion procedure** compares two groups by studying the difference $p_1 - p_2$ between two population (or treatment) proportions.

This lesson finishes the proportion toolkit of Exam Unit 3. The exam expects both procedures, the interval and the test, with conditions checked for *each* sample, and it expects you to know why the test uses a pooled proportion while the interval does not. Typical lost points: using paired reasoning for independent groups, checking large counts for the combined sample only, and concluding causation from an observational comparison.

Core Ideas

The parameter is $p_1 - p_2$, the difference between two proportions. Its natural point estimate is $\hat{p}_1 - \hat{p}_2$. When the two samples are independent, the variances add: $\text{Var}(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$. Variances add even though we subtract the proportions, because uncertainty from both samples contributes to uncertainty in the difference.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Conditions for Two-Proportion Procedures

Random and independent: the two groups come from independent random samples, or from a randomized experiment with subjects randomly assigned to two treatments.

10 percent: each sample is less than 10 percent of its population when sampling without replacement.

Large counts for each sample: in each group, the numbers of successes and failures are at least 10. For an interval check $n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$, $n_2(1-\hat{p}_2)$. For a test with pooled $\hat{p}_c$, check $n_1\hat{p}_c$, $n_1(1-\hat{p}_c)$, $n_2\hat{p}_c$, $n_2(1-\hat{p}_c)$ (many courses accept checking with the observed counts).

Two-Proportion z-Interval

$$(\hat{p}_1-\hat{p}_2) \pm z^*\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Interpretation: We are C% confident that the interval from ___ to ___ captures the true difference p_1-p_2 in the proportion of (context).

If the interval contains 0, a difference of zero is plausible. If it is entirely positive, group 1 is plausibly higher; entirely negative, group 2 is.

For a test, the null hypothesis is $H_0: p_1-p_2=0$, or equivalently $p_1=p_2$. If the null is true, both groups share one common proportion, so the best estimate of it combines the data from both samples. This is the **pooled (combined) proportion**: $\hat{p}_c=\frac{x_1+x_2}{n_1+n_2}$, the total number of successes over the total number of subjects. We use it in the standard error because the standard error should describe a world in which H_0 is true.

Two-Proportion z-Test

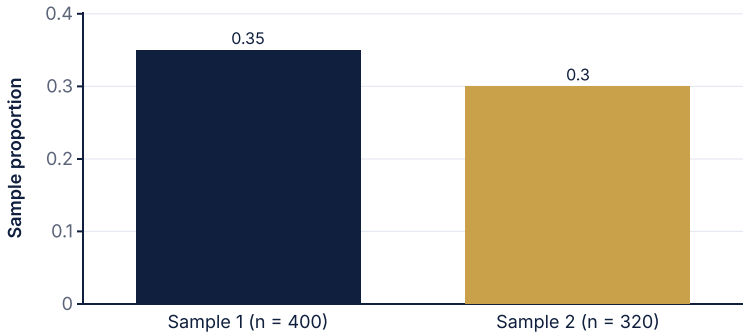
$H_0: p_1-p_2=0$; $H_a: p_1-p_2>0$, <0, or $\neq 0$.

Pooled proportion: $\hat{p}_c=\frac{x_1+x_2}{n_1+n_2}$.

$$z=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

P-value from the standard normal curve using the tail(s) matching H_a; compare with α; conclude in context.

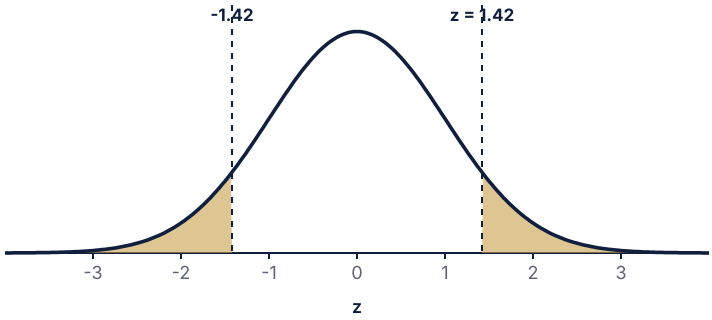
The interval does not pool because it does not assume $p_1=p_2$; it estimates the size of the difference, so each group keeps its own variability. In the test, the null asserts equality, so combining the samples is natural. A good memory aid: **test = pool, interval = do not pool.**



A 5-point gap in sample proportions: is it real?

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Two-sided P-value for z = 1.42 is 0.1555.

Experiments versus observational comparisons. If subjects were randomly assigned to treatments and a significant difference results, we may conclude that the treatment *caused* the difference. If two independent random samples were drawn from two existing populations, we can generalize to those populations but cannot claim cause. The statistical machinery is identical; the wording of the conclusion is not.

Why do variances add? Imagine two bathroom scales, each off by a little in an unpredictable way. If you weigh a box on one and a bag on the other and then subtract, the errors do not cancel; they stack, because each scale wobbles on its own. The same thing happens with sample proportions. Each of $\hat{p}_1$ and $\hat{p}_2$ has its own sampling variability, and since the samples are independent, the variability of the difference is larger than either alone. In the bus example, each proportion has a standard error of about 0.024 and 0.026, but the difference has a standard error of about 0.035, which is $\sqrt{0.024^2 + 0.026^2}$.

Naming the groups. Decide which group is Group 1 before you compute anything and write it down. The order sets the sign of your answer and the direction of the alternative. If Group 1 is the app and Group 2 is the notes-only group, then a positive interval means the app did better. If the order were reversed the interval would flip signs, and a conclusion that does not match the labeling will cost you the communication point.

Situation	Procedure	Standard error
One sample compared with a claimed value	One-proportion z-test	Uses p_0 and n
Estimate one proportion	One-proportion z-interval	Uses $\hat{p}$ and n
Two independent groups, test of equality	Two-proportion z-test	Pooled $\hat{p}_c$ with both sample sizes
Two independent groups, size of the difference	Two-proportion z-interval	Separate $\hat{p}_1$ and $\hat{p}_2$
Same individuals measured twice	Not a two-proportion procedure	Needs a paired approach

Are the groups independent? This is the hidden condition that students skip. Independent means that knowing the outcome of one subject tells you nothing about another subject in the other group. Two random samples from two different schools are independent. A before-and-after measurement on the same students is not. Siblings, roommates, or pairs matched on a characteristic are also dependent. In an experiment, random assignment of subjects to two treatment groups gives groups that are treated as independent for these procedures.

What the interval adds to the test. Suppose a test concludes that the proportions differ. A natural follow-up is, by how much? The confidence interval for the difference answers that. An interval of (0.03, 0.25) says the gap is positive but could be small or fairly large. An interval of (0.20, 0.24) says the gap is positive and precisely measured. Both give the same test decision at 0.05, yet they support very different business or policy choices. When the question asks you to estimate or to determine plausible values, use the interval. When it asks whether there is convincing evidence of a difference, use the test.

Pooled versus unpooled: do they ever disagree? Because the test pools and the interval does not, the two can disagree in borderline cases where a P-value is just above or below 0.05 and the interval just barely contains or excludes 0. When that happens, trust the procedure that matches the question being asked: use the P-value for a test conclusion and the interval for an estimate. On the exam, the question will state which procedure to carry out, so you will not usually face a conflict.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Wording Templates You Can Reuse

Interval: 'We are C% confident that the interval from ___ to ___ captures the true difference (Group 1 minus Group 2) in the proportion of (population) that (characteristic).'

Test, reject: 'Because the P-value of ___ is less than $alpha =$ ___, we reject H_0. We have convincing evidence that the proportion of (Group 1) that (characteristic) is (greater than / less than / different from) the proportion of (Group 2).'

Test, fail to reject: 'Because the P-value of ___ is greater than $alpha =$ ___, we fail to reject H_0. We do not have convincing evidence that the two proportions differ.'

Experiment add-on: 'Because subjects were randomly assigned, we can conclude that the treatment caused the difference.'

How sample size changes the story. Suppose two groups have sample proportions of 0.35 and 0.30, a five-point gap. With around 400 per group, that gap is only about 1.4 standard errors, which is unremarkable. With 4,000 per group, the standard error shrinks to about a third of its size and the same five-point gap is about 4.5 standard errors, which is overwhelming. The gap did not change; our ability to distinguish it from noise did. When a comparison fails to reach significance, check the sample sizes before concluding that the groups are alike, and when it succeeds, check the size of the gap before concluding that it matters.

An annotated walk-through of an experiment. Consider the flashcard study. The researcher recruits volunteers, flips a coin for each to decide whether they use the app or notes, and then compares quiz pass rates. The random assignment is what makes this an experiment, and it is why the groups are comparable at the start: motivated and unmotivated students end up in both groups in roughly equal numbers. The parameter is the difference between the pass rate every volunteer like these would have under the app and the pass rate they would have under notes. The test statistic is about 2.7 standard errors, so the observed 21-point gap would be rare if the app did nothing. Because the volunteers were not randomly sampled from all students, we can generalize only to similar volunteers; because the treatments were randomly assigned, we can state cause and effect within that group. Two different kinds of randomness answer two different questions.

Common wording errors to avoid. First, do not describe the interval as capturing the difference between the sample proportions; that difference is known and sits at the center. Second, do not say 'the proportions are the same' when you fail to reject; say that you lack convincing evidence of a difference. Third, do not leave the direction ambiguous. A statement like 'there is a difference' is incomplete when the alternative was one-sided; say which group is higher.

Worked Examples

DEFINING THE PARAMETER AND STANDARD ERROR

In a random sample of 400 students from School A, 140 take the bus. In an independent random sample of 320 students from School B, 96 take the bus. Define the parameter and find the estimate and unpooled standard error.

SOLUTION

Let p_1 be the true proportion of School A students who take the bus and p_2 the true proportion for School B. The parameter is $p_1 - p_2$.

$\hat{p}_1 = 140/400 = 0.35$, $\hat{p}_2 = 96/320 = 0.30$, so the estimate is $\hat{p}_1 - \hat{p}_2 = 0.05$.

$$\text{SE} = \sqrt{\frac{0.35(0.65)}{400} + \frac{0.30(0.70)}{320}} = \sqrt{0.000569 + 0.000656} \approx 0.0350.$$

A COMPLETE TWO-PROPORTION INTERVAL

Using the bus data above (each school has thousands of students), construct and interpret a 95 percent confidence interval for $p_1 - p_2$.

SOLUTION

Procedure. Two-proportion z-interval for $p_1 - p_2$.

Conditions. Independent random samples; both samples are less than 10 percent of their schools; counts are $140, 260, 96, 224$, all at least 10.

Calculation. $0.05 \pm 1.960(0.0350) = 0.05 \pm 0.0686$, so the interval is (-0.019, 0.119).

Conclusion. We are 95 percent confident that the interval from -0.019 to 0.119 captures the true difference (School A minus School B) in the proportion of students who take the bus. Because 0 is in the interval, we cannot conclude the proportions differ.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

CHECKING CONDITIONS IN A COMPARISON

A researcher compares the proportion of players with a recent injury in two leagues. Sample 1: 25 of 30 players in a league with 400 players say no injury. Sample 2: 70 of 150 players in a league with 1,200 players say no injury. Which conditions fail?

SOLUTION

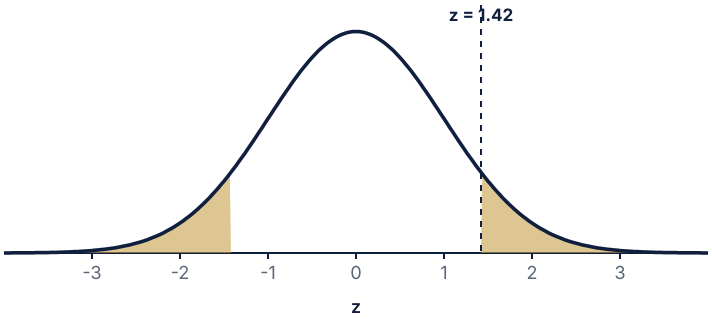
Sample 1: 30 is less than 10 percent of 400 (40), fine. Counts: successes 25, failures 5. The count of 5 is less than 10, so large counts **fails** for sample 1.

Sample 2: 150 is more than 10 percent of 1,200 (120), so the 10 percent condition **fails** for sample 2. The counts, 70 and 80, are fine.

Because a condition fails in each group, a two-proportion z-procedure is not appropriate. Consider larger samples or a different method.

THE POOLED PROPORTION AND TEST STATISTIC

Using the bus data (140 of 400 and 96 of 320), test $H_0 : p_1 = p_2$ versus $H_a : p_1 \neq p_2$ at $\alpha = 0.05$.



Both tails beyond |z| = 1.42.

SOLUTION

State. $H_0 : p_1 - p_2 = 0$, $H_a : p_1 - p_2 \neq 0$.

Plan. Two-proportion z-test. Conditions as before (random, independent, 10 percent, large counts with pooled $\hat{p}_c = 0.328$: $400(0.328) = 131$, $400(0.672) = 269$, $320(0.328) = 105$, $320(0.672) = 215$).

Do. $\hat{p}_c = (140 + 96)/(400 + 320) = 236/720 \approx 0.3278$. $\text{SE} = \sqrt{0.3278(0.6722)(1/400 + 1/320)} \approx 0.0352$. $z = 0.05/0.0352 \approx 1.42$. P-value $= 2P(Z \geq 1.42) \approx 0.1555$.

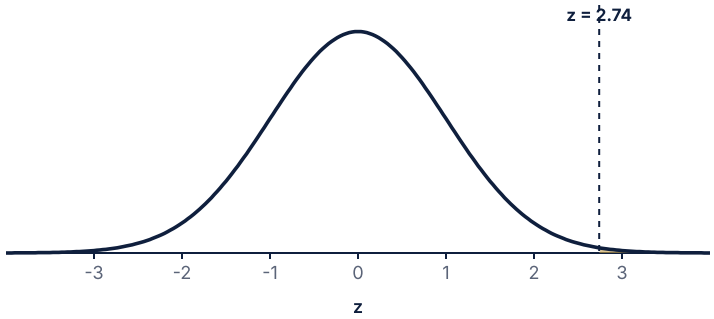
Conclude. Since $0.1555 > 0.05$, we fail to reject H_0. We do not have convincing evidence that the proportions of students who take the bus differ between the two schools. This agrees with the interval, which contained 0.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A RANDOMIZED EXPERIMENT

In a randomized experiment, 80 volunteers study with a new flashcard app and 85 volunteers study with notes only. Afterward, 58 of the 80 app users and 44 of the 85 notes users pass a quiz. Is there convincing evidence that the app increases the passing rate? Use $\alpha = 0.05$.



Upper-tail P-value = 0.0031.

SOLUTION

State. p_1 = true passing proportion for volunteers like these using the app, p_2 = same for notes. $H_0 : p_1 - p_2 = 0$, $H_a : p_1 - p_2 > 0$.

Plan. Two-proportion z-test. Random assignment; counts are 58, 22, 44, 41 (all at least 10). Volunteers are not a random sample, so we generalize only to similar volunteers.

Do. $\hat{p}_1 = 0.725$, $\hat{p}_2 = 0.518$, $\hat{p}_c = 102/165 \approx 0.618$. $\text{SE} = \sqrt{0.618(0.382)(1/80 + 1/85)} \approx 0.0757$. $z = 0.2074/0.0757 \approx 2.74$. P-value ≈ 0.0031.

Conclude. Since $0.0031 < 0.05$, we reject H_0. We have convincing evidence that the app increases the passing rate for volunteers like these, and because treatments were randomly assigned, the app can be credited as the cause.

INTERVAL AND TEST TOGETHER

For the experiment above, construct a 90 percent confidence interval for $p_1 - p_2$ and explain how it supports the test.

SOLUTION

Unpooled SE $= \sqrt{0.725(0.275)/80 + 0.518(0.482)/85} \approx 0.0737$.

Estimate $0.2074 \pm 1.645(0.0737) = 0.2074 \pm 0.1212$, or (0.086, 0.329).

We are 90 percent confident that the interval from 0.086 to 0.329 captures the true increase in passing proportion for the app over notes. The interval lies entirely above 0, matching the rejection in the one-sided test, and it shows the improvement could be modest (about 9 points) or large (about 33 points).

Write It Like a 5

Model Response

Prompt. A random sample of 150 students from a large school Cavern High finds 72 who use the library weekly; an independent random sample of 160 from a large school Ridge High finds 54. Is there convincing evidence that the proportions differ? Use $\alpha = 0.05$.\n\n**State.** Let p_1 and p_2 be the true proportions of students at Cavern High and Ridge High who use the library weekly. $H_0 : p_1 - p_2 = 0$ and $H_a : p_1 - p_2 \neq 0$, $\alpha = 0.05$.\n\n**Plan.** Two-proportion z-test. Samples are random and independent; each is less than 10 percent of its school. Pooled $\hat{p}_c = 126/310 \approx 0.4065$. Counts: $150(0.4065) = 61$, $150(0.5935) = 89$, $160(0.4065) = 65$, $160(0.5935) = 95$, all at least 10.\n\n**Do.** $\hat{p}_1 = 0.48$, $\hat{p}_2 = 0.3375$. $\text{SE} = \sqrt{0.4065(0.5935)(1/150 + 1/160)} \approx 0.0558$. $z = 0.1425/0.0558 \approx 2.55$. P-value $= 2P(Z \geq 2.55) \approx 0.0107$.\n\n**Conclude.** Because $0.0107 < 0.05$, we reject H_0. We have convincing evidence that the true proportions of students who use the library weekly differ between the two schools.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

(1) Both parameters defined in context and hypotheses about their difference. (2) Names the two-proportion z-test and verifies random independent samples, the 10 percent condition for each school, and large counts for each group. (3) Pooled proportion used correctly in the standard error. (4) Correct z and a two-sided P-value. (5) Decision linked to the P-value and a conclusion in context without claiming causation.

Coach's Playbook

1. Label the groups 1 and 2, say what each is, and define $p_1 - p_2$ in words.
2. Decide: sampling from two populations, or randomized experiment? That sets the wording of your conclusion.
3. Verify conditions for each group separately, with numbers.
4. Interval: separate standard errors. Test: pooled standard error. Say which you are using.
5. Conclude in context. For an interval, report the direction of the difference and whether 0 is plausible.

Common traps

1. Pooling for an interval or failing to pool for a test. 2. Checking large counts only for the combined data. 3. Treating two groups measured on the same people as independent samples. 4. Claiming cause and effect after comparing two existing groups.

- **Label first.** Write 'Group 1 = ..., Group 2 = ...' before computing and keep it consistent in the hypotheses, the estimate, and the conclusion.
- **Run the counts for each group.** Write four numbers: successes and failures in each group. If any is below 10, say so and stop.
- **Sketch the curve.** A quick normal curve with the tail(s) shaded keeps your P-value in the right direction for the alternative you stated.
- **Check the sign.** If $\hat{p}_1 > \hat{p}_2$, the test statistic must be positive and the interval should lean positive.
- **Match the wording to the design.** 'Caused' for randomized experiments, 'is associated with' or 'differs between the populations' for samples.

Mistake Clinic

Spot the error

Wrong solution. 'The 95 percent interval for $p_1 - p_2$ is (-0.019, 0.119). Since 0 is in the interval, the two proportions are equal.'\n\n**Where it fails.** An interval containing 0 shows that no difference is plausible; it does not prove equality. Other differences in the interval are plausible too.\n\n**Correct.** There is not convincing evidence of a difference.

Spot the error

Wrong solution. 'For the test, SE $= \sqrt{0.35(0.65)/400 + 0.30(0.70)/320}$.'\n\n**Where it fails.** That is the unpooled standard error, appropriate for intervals. A test of $H_0 : p_1 = p_2$ should use the pooled proportion.\n\n**Correct.** $\text{SE} = \sqrt{\hat{p}_c(1-\hat{p}_c)(1/n_1 + 1/n_2)}$ with $\hat{p}_c = 0.328$, giving 0.0352 rather than 0.0350.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Wrong solution. 'Volunteers who studied with the app passed more often, so the app caused the improvement.' (The volunteers chose which method to use.)\n\n**Where it fails.** Without random assignment, those who chose the app may differ in motivation. Association is not causation.\n\n**Correct.** Only a randomized experiment supports a causal conclusion; here we can only report an association.

Real-World Connection

Online experiments ('A/B tests') are two-proportion problems: visitors are randomly shown version A or B, and the outcome is whether they click or sign up. Teams compute the pooled z-test to decide whether a layout change truly lifts conversion, and report the interval for the difference to say how large the lift might be. The random assignment is what lets them say the new layout *caused* the lift. Public-health surveys do the same thing with existing groups, such as comparing vaccination rates between two regions, but there they must speak of differences between populations, not causes.

Challenge Corner

Group A has 90 subjects with 63 successes and Group B has 100 subjects with 48 successes, in a randomized experiment. Test $H_a: p_A > p_B$ and compute a 95 percent interval for $p_A - p_B$. Explain why the test and interval agree.\n\n**Solution.** $\hat{p}_A = 0.70$, $\hat{p}_B = 0.48$, $\hat{p}_c = 111/190 = 0.584$. $\text{SE}_{\text{pooled}} = \sqrt{0.584(0.416)(1/90 + 1/100)} = 0.0716$ and $z = 0.22/0.0716 \approx 3.07$, P-value ≈ 0.0011. Unpooled SE $= \sqrt{0.7(0.3)/90 + 0.48(0.52)/100} = 0.0695$; interval $0.22 \pm 1.96(0.0695) = (0.084, 0.356)$. The interval excludes 0, matching the strong rejection, and shows that the true difference could range from about 8 to 36 percentage points.

Practice

PRACTICE **Two-Proportion z-Intervals and Tests**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Two independent random samples are used to compare the proportion of boys and the proportion of girls in a school who own a bicycle. Which is the parameter of interest?
   (A) $\hat{p}_1 - \hat{p}_2$
   (B) $p_1 + p_2$
   (C) $p_1 - p_2$, the difference between the two population proportions
   (D) $\hat{p}_c$
2. Sample 1: $\hat{p}_1 = 0.50$, $n_1 = 100$. Sample 2: $\hat{p}_2 = 0.40$, $n_2 = 150$. What is the standard error (not pooled) of $\hat{p}_1 - \hat{p}_2$?
   (A) 0.0041 (B) 0.0640 (C) 0.0370 (D) 0.0910
3. In sample 1, 36 of 80 succeed. In sample 2, 54 of 120 succeed. What is the pooled proportion $\hat{p}_c$?
   (A) 0.90 (B) 0.225 (C) 0.40 (D) 0.45
4. Which is a required condition for a two-proportion z-test comparing two existing populations?
   (A) The two samples are independent random samples.
   (B) The two sample sizes are equal.
   (C) Both samples have more than 30 observations.
   (D) The same individuals appear in both samples.
5. A 95 percent confidence interval for p1 - p2 is (-0.02, 0.12). Which conclusion is best?
   (A) The proportions are equal.
   (B) Group 1 is definitely higher than Group 2.
   (C) There is a 95 percent probability that the sample proportions differ.
   (D) A difference of 0 is plausible, so there is not convincing evidence that the proportions differ.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. In a two-proportion z-test of H0: p1 = p2, the standard error uses the pooled proportion because
   (A) the null hypothesis assumes both groups share one common proportion.
   (B) the samples are small.
   (C) the groups must be the same size.
   (D) pooling makes the interval narrower.
7. In a randomized experiment comparing two treatments, a two-proportion test gives a small P-value. Which conclusion is justified?
   (A) The result applies to every person in the world.
   (B) The treatment difference caused the difference in success rates (for subjects like these).
   (C) The groups started out different.
   (D) The null hypothesis is certainly false.
8. In a test of H0: p1 = p2, 45 of 100 and 30 of 100 succeed. The pooled proportion is 0.375. What is z?
   (A) 1.06 **(B)** 3.10 **(C)** 2.19 **(D)** 0.15
9. Sample 1 has n = 40 with 7 successes; Sample 2 has n = 200 with 90 successes. For a two-proportion z-interval, which condition fails?
   (A) Large counts for Sample 2 **(B)** Large counts for Sample 1
   (C) The sample size of Sample 2 **(D)** Nothing fails
10. Random sample of 300 students at Elm Middle (a large school): 156 have a smartphone. Independent random sample of 280 students at Pine Middle (large): 129 have one. Construct and interpret a 95 percent confidence interval for $p_1 - p_2$.
11. A random sample of 150 students at Cavern High finds 72 who use the library weekly; an independent random sample of 160 at Ridge High finds 54. Both schools are large. Test whether the proportions differ at $\alpha = 0.05$.
12. In a randomized experiment, 90 subjects use Plan A and 100 use Plan B. Successes: 63 for A and 48 for B. Test $H_a : p_A > p_B$ at $\alpha = 0.05$ and state whether Plan A can be said to cause greater success.
13. A study compares injury rates: Group 1 has 30 players with 5 injuries and Group 2 has 150 players with 40 injuries, from two leagues of about 2,000 players each. Evaluate each condition for a two-proportion z-interval and state whether the procedure is appropriate.
14. For the library data (72 of 150 versus 54 of 160), construct a 95 percent confidence interval for $p_1 - p_2$ and explain how it agrees with the two-sided test at $\alpha = 0.05$.
15. A student tests whether a new routine improves free-throw rates by measuring the same 40 players before and after, then runs a two-proportion z-test with 40 shots 'before' and 40 'after'. Identify what is wrong. Separately, explain why a significant result from two independent self-selected groups would not show the routine caused the change.

T Unit 7 Test

Inference for proportions: confidence intervals, significance tests, errors and power, and two-proportion procedures.

Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

UNIT TEST

Unit 7 Test

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. A 95 percent confidence interval for a population proportion is computed from a random sample. Which is the best interpretation of the 95 percent?
   (A) There is a 95 percent chance that the true proportion is in the interval from this sample.
   (B) 95 percent of the individuals in the population fall inside the interval.
   (C) If this procedure were repeated with many random samples of the same size, about 95 percent of the intervals would capture the true proportion.
   (D) The sample proportion is within the interval 95 percent of the time.
2. A random sample of 80 students is used to build a z-interval. Only 6 of the 80 say they walk to school. Which is true?
   (A) The large counts condition fails because only 6 successes were observed.
   (B) The interval is valid because n is at least 30.
   (C) The 10 percent condition fails because 6 is small.
   (D) The interval is valid as long as the sample is random.
3. A pollster wants a 95 percent margin of error of at most 0.025 and has no prior information about the proportion. What is the smallest sample size that guarantees this?
   (A) 1536 **(B)** 385 **(C)** 3073 **(D)** 1537
4. In a random sample of 250 residents, 90 favor a new tax. Which is the 95 percent confidence interval for the proportion of all residents who favor it?
   (A) (0.330, 0.390) **(B)** (0.300, 0.420) **(C)** (0.280, 0.440)
   (D) (0.360, 0.420)
5. A factory claims that 5 percent or fewer of its items are defective. An inspector suspects the true rate is higher. Which are the correct hypotheses?
   (A) $H_0 : p = 0.05$, $H_a : p > 0.05$ **(B)** $H_0 : p > 0.05$, $H_a : p = 0.05$ **(C)** $H_0 : \hat{p} = 0.05$, $H_a : \hat{p} > 0.05$
   (D) $H_0 : p = 0.05$, $H_a : p /= 0.05$
6. A test of H0: p = 0.40 gives a P-value of 0.03. Which is the correct interpretation?
   (A) There is a 3 percent chance that H0 is true.
   (B) The sample proportion is 3 percent away from 0.40.
   (C) If the true proportion is 0.40, there is a 3 percent chance of getting a sample proportion at least as extreme as the one observed.
   (D) There is a 97 percent chance that Ha is true.
7. A random sample of 80 (from a large population) has 44 successes. For H0: p = 0.50 versus Ha: p > 0.50, what is the P-value?
   (A) 0.186 **(B)** 0.371 **(C)** 0.814 **(D)** 0.050
8. A 99 percent confidence interval for p is (0.41, 0.49). For a two-sided test of H0: p = 0.50 at alpha = 0.01, what is the conclusion?
   (A) Reject H0 because 0.50 is not in the interval.
   (B) Fail to reject H0 because 0.50 is close to the interval.
   (C) Reject H0 because the interval contains p-hat.
   (D) Fail to reject H0 because the interval is narrow.
9. A school tests H0: the new schedule does not change the proportion of students who are on time, against Ha: it increases the proportion. Which describes a Type II error?
   (A) Concluding the schedule increases the proportion when it does not.
   (B) Failing to find convincing evidence of an increase when the schedule really does increase the proportion.
   (C) Concluding the proportion decreased.
   (D) Rejecting H0 when the P-value is greater than alpha.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. Which change would increase the power of a one-sided test of a proportion against a specific alternative?
    (A) Decreasing the significance level
    (B) Moving the true value closer to the null value
    (C) Increasing the sample size
    (D) Using a two-sided alternative
11. Group 1: 45 successes of 90. Group 2: 30 successes of 110. What is the pooled proportion used in a test of H0: p1 = p2?
    (A) 0.375 (B) 0.386 (C) 0.500 (D) 0.273
12. In a large population, 40 percent of voters support a measure. For random samples of 200 voters, approximately what is the probability that the sample proportion exceeds 0.45?
    (A) 0.050 (B) 0.074 (C) 0.372 (D) 0.926
13. [6 points] A school claims that 70 percent of its graduates enroll in a college program. A random sample of 200 graduates, taken from the 3,000 graduates over the past decade, finds that 128 enrolled. A skeptic believes the true proportion is lower than 70 percent. Do the data provide convincing evidence for the skeptic? Use a significance level of 0.05. Show the four steps: (a) state hypotheses [1 point], (b) name the procedure and check conditions [2 points], (c) compute the test statistic and P-value [2 points], (d) conclude in context [1 point].
14. [10 points] A researcher randomly assigns 250 volunteers to receive reminders by text message (120 volunteers) or by email (130 volunteers). At the end of the month, 84 of the 120 text volunteers and 66 of the 130 email volunteers completed an online health survey. (a) Construct and interpret a 95 percent confidence interval for the difference in the true completion proportions (text minus email); check conditions. [4 points] (b) Does the interval give evidence that the methods differ? Explain. [1 point] (c) Test whether text reminders give a higher completion proportion than email at $\alpha = 0.05$. [3 points] (d) In the context of part (c), describe a Type I error and a Type II error, and state whether the researcher can conclude that text reminders cause higher completion. [2 points]

Answers Inference for Proportions: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

7.1 The Logic of Confidence Intervals

PRACTICE · THE LOGIC OF CONFIDENCE INTERVALS

1. B. The point estimate is the sample proportion 63/150 = 0.42. Dividing by 100 or reporting the raw count are the tempting slips.
2. C. The interval makes a statement about the population proportion. The other choices describe sample statistics or individuals, which is not what the interval estimates.
3. D. The level is the long-run capture rate of the method. Saying the parameter is inside the interval 90 percent of the time treats a fixed number as moving, and the data-range wording confuses an interval for a parameter with a range of data values.
4. A. A larger n shrinks the standard error. The other changes increase the critical value or the product p-hat(1 - p-hat), both of which widen the interval.
5. C. Only the critical value changes, so the center stays at p-hat while the width grows from 1.960 to 2.576 standard errors. Calling the wider interval 'more accurate' while narrower is a common confusion of reliability with precision.
6. B. 90 percent of 200 is 180. Answering 90 forgets to scale by 200, and 198 would correspond to a 99 percent level.
7. D. The estimate is the midpoint 0.44 and the margin is half the width, 0.06. Reporting 0.12 uses the full width instead of half of it.
8. B. Once computed, the interval either contains p or it does not. The 95 percent refers to the long-run success rate of the procedure.
9. D. The margin of error is proportional to 1/sqrt(n), so halving it requires four times the sample. Doubling the sample is the tempting linear guess.
10. (a) $\hat{p} = 84/240 = 0.35$. It estimates p, the proportion of all teens in the population who sleep fewer than 7 hours on school nights.
    (b) We are 95 percent confident that the interval from 0.290 to 0.410 captures the true proportion of all teens in the population who sleep fewer than 7 hours on school nights.
11. (a) The miss rate is 5 percent, so about $0.05 \times 60 = 3$ intervals should miss.
    (b) An interval misses when the sample proportion falls far enough from the true parameter that the interval built around it does not contain the true parameter. The count will vary from 3 by chance.
12. (a) Incorrect. The interval is about the proportion of residents who use transit, a single parameter, not about how many weeks individuals ride.
    (b) Incorrect. The sample proportion is the midpoint, 0.26, and is known exactly. The 90 percent describes the method's long-run capture rate of the population proportion.
13. (a) The 99 percent interval is wider because its critical value (2.576) is larger than the 90 percent critical value (1.645), and margin of error equals the critical value times the standard error.
    (b) A wider interval is less precise but more reliable: the method captures the true proportion in a larger share of repeated samples. The choice is a tradeoff, not a flaw.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. (a) $\hat{p} = (0.54 + 0.62)/2 = 0.58$ and ME $= 0.04$.
    (b) SE $= 0.04/1.645 \approx 0.02432$.
    (c) ME $= 1.960(0.02432) \approx 0.0477$, so the 95 percent interval is 0.58 ± 0.0477, or about (0.532, 0.628).
15. (a) No. The interval includes values below 0.50 (such as 0.42 to 0.49), so it is plausible that fewer than half of all students support the change.
    (b) We are 95 percent confident that between 42 percent and 54 percent of all students at this school support the later start time. The data do not clearly establish a majority.

7.2 One-Proportion z-Intervals, Margin of Error, and Sample Size

PRACTICE · ONE-PROPORTION Z-INTERVALS, MARGIN OF ERROR, AND SAMPLE SIZE

1. C. The 90 percent level leaves 5 percent in each tail, giving 1.645. The value 1.282 is the cutoff leaving 10 percent in one tail only, and 1.960 belongs to 95 percent.
2. B. The large counts condition requires at least 10 successes and 10 failures. Here only 6 successes are expected. Size alone (n = 60) says nothing about the 10 percent condition without the population size.
3. D. SE = sqrt(0.40 × 0.60 / 200) = sqrt(0.0012) = 0.0346. Forgetting the square root gives 0.0012, and forgetting to divide by n gives about 0.49.
4. C. p-hat = 0.45, SE = sqrt(0.45 × 0.55 / 120) = 0.0454, ME = 1.96 × 0.0454 = 0.0890. The interval is 0.45 +/- 0.089. The narrow interval uses a margin that is too small and the asymmetric one is not centered at p-hat.
5. B. n >= (1.96/0.04)^2 (0.5)(0.5) = 600.25, and sample sizes round up, so 601. Rounding to 600 would give a margin slightly above 0.04.
6. A. n >= (1.96/0.05)^2 (0.20)(0.80) = 245.86, round up to 246. Using 0.5 as the planning value would give 385, which is larger than necessary when a good estimate exists, and a sample size cannot be a decimal.
7. A. The margin of error is proportional to 1/sqrt(n). Multiplying n by 4 divides the margin by 2, not by 4.
8. D. The large counts condition is what justifies the normal approximation for p-hat. Normality of the population and n at least 30 are rules for means, not proportions, and a proportion has no separate sigma to know.
9. C. Because 0.50 is inside the interval, it is a plausible value. An interval never proves an exact value, and it makes no probability statement about p or about sample values.
10. Parameter: p = true proportion of all customers who buy a pastry. Procedure: one-proportion z-interval.
    Conditions: random sample; $500 < 1200$ (10 percent of 12,000); $500(0.42) = 210 \geq 10$ and $500(0.58) = 290 \geq 10$.
    Calculation: $\hat{p} = 0.42$, SE $= \sqrt{0.42(0.58)/500} \approx 0.02207$, ME $= 1.960(0.02207) \approx 0.0433$. Interval: (0.377, 0.463).
    Conclusion: We are 95 percent confident that the interval from 0.377 to 0.463 captures the true proportion of all customers of this coffee shop who buy a pastry.
11. (a) Random: yes. 10 percent: $40 < 200$, satisfied. Large counts: $n\hat{p} = 3 < 10$, which fails (and $n(1 - \hat{p}) = 37 \geq 10$ passes).
    (b) The researcher should not use the z-interval. A larger sample is needed so that the expected number of injured players is at least 10, which would take well over 100 players if the injury rate is near 7.5 percent.
12. Use the conservative value $\tilde{p} = 0.5$ because there is no prior estimate; it maximizes $\tilde{p}(1 - \tilde{p})$ and so guarantees the margin is met.
    $n \geq (2.576/0.03)^2(0.25) = 7373.2(0.25) \approx 1843.3$. Round up: $n = 1844$ families.
13. (a) $\hat{p} = 0.44$, SE $= \sqrt{0.44(0.56)/300} \approx 0.02866$, ME $= 1.960(0.02866) \approx 0.0562$. Conditions hold: 132 and 168 are at least 10. Interval: (0.384, 0.496).
    (b) No. Every value in the interval is below 0.50, so the data suggest that fewer than half of residents favor extending hours. We are 95 percent confident the true proportion is between 0.384 and 0.496.
14. (a) $\tilde{p} = 21/60 = 0.35$. $n \geq (1.645/0.04)^2(0.35)(0.65) = 1691.3(0.2275) \approx 384.8$, so $n = 385$.
    (b) With $\tilde{p} = 0.5$: $n \geq 1691.3(0.25) \approx 422.8$, so 423. The required size rises because $0.5(0.5) = 0.25$ is the largest possible value of $\tilde{p}(1 - \tilde{p})$; the conservative choice guarantees the margin but may over-sample.
15. ME $= z^* \sqrt{\hat{p}(1 - \hat{p})/n}$. (1) Increase n: this shrinks the standard error, but costs more time and money, and quadrupling n is needed to halve the margin. (2) Lower the confidence level: this reduces z^* (for example from 1.960 to 1.645), but the method then captures the parameter in a smaller percentage of samples, so the result is less reliable.

7.3 The Logic of Significance Tests and P-Values

PRACTICE · THE LOGIC OF SIGNIFICANCE TESTS AND P-VALUES

1. C. The null states the claimed value with equality and the alternative matches the suspicion of a higher rate. The statistic version confuses p-hat with p, and the two-sided alternative ignores the stated direction.
2. A. The null hypothesis is an equality about the parameter. Statements with inequality signs are alternatives, and the statistic p-hat is known from the sample, so it is never tested.
3. B. A small P-value means data this extreme are rare when H0 is true, which is evidence against H0. It does not measure the size of an effect or the quality of the sampling.
4. C. The P-value is a probability about data computed assuming H0 is true. The wording about H0 being true or about chance alone turns it into a probability for the hypothesis, which it is not.
5. D. Since 0.03 is less than 0.05 we reject H0 and have convincing evidence for Ha. Failing to reject would require a P-value above alpha, and a test never proves the null false with certainty.
6. A. A P-value above alpha leads to failing to reject H0, which means the evidence is not convincing. It does not show that H0 is true, and rejecting would need the P-value to be at most alpha.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. C. Only 2 of the 100 simulated results were 15 or more, so the P-value estimate is 2/100 = 0.02. Using the tail on the wrong side gives 0.98, and the other values count bars below 15.
8. D. A two-sided alternative is used when the question is only whether the parameter is different from the claim. The other situations name a direction, and choosing a direction after seeing the sample is not allowed.
9. B. Two-sided P-value = 2 x P(Z < -2.50) = 2(0.0062) = 0.0124. The value 0.0062 is only one tail, and 0.9938 is the area to the left of 2.50, not a tail.
10. (a) p = the true proportion of this team's free throws that go in. $H_0 : p = 0.60$, $H_a : p < 0.60$.

    (b) p = the true proportion of this café's orders that are iced drinks. $H_0 : p = 0.25$, $H_a : p \neq 0.25$ (two-sided because the manager asks only whether it is still accurate).
11. (a) If the true proportion of defect-free screens is 0.90, the probability of getting a sample proportion as small as the one observed or smaller is 0.018.

    (b) Since $0.018 < 0.05$, we reject H_0. We have convincing evidence that the true proportion of defect-free screens is less than 0.90.
12. (a) The estimated P-value is $9/120 = 0.075$.

    (b) Since $0.075 > 0.05$, we fail to reject H_0. We do not have convincing evidence that the spinner lands on red more than 25 percent of the time.
13. (a) At 0.10: $0.032 < 0.10$, reject H_0. At 0.05: $0.032 < 0.05$, reject H_0. At 0.01: $0.032 > 0.01$, fail to reject H_0.

    (b) If α were chosen after seeing the P-value, a researcher could pick whichever level gives the desired decision. Setting it in advance keeps the standard of evidence honest.
14. (a) Wrong: the P-value is computed assuming H_0 is true; it is the probability of data this extreme, not the probability of H_0.

    (b) Wrong: we fail to reject H_0; large P-values never prove it.

    (c) Wrong: lack of significance means the evidence is not convincing. A real but small effect, or a small sample, could still be hiding.
15. (a) $2(0.04) = 0.08$.

    (b) No. The first rejects H_0 ($0.04 < 0.05$) while the second fails to reject ($0.08 > 0.05$).

    (c) Choosing the direction after seeing which way the data lean halves the P-value artificially, which makes the test easier to pass and invalidates its stated error rate. The alternative must come from the research question before the data are seen.

7.4 One-Proportion z-Tests

PRACTICE · ONE-PROPORTION Z-TESTS

1. D. SE = sqrt(0.25 × 0.75 / 400) = 0.02165 under H0. The other values come from omitting the square root or the denominator, or from using a wrong p.
2. A. z = (0.30 - 0.25) / 0.02165 = 2.31. The value 2.18 comes from using p-hat in the standard error, which is a mistake in a test.
3. B. The alternative is one-sided upward, so the P-value is the upper tail area 0.0105. Doubling gives the two-sided value 0.0209, and 0.9895 is the area to the left.
4. C. A two-sided alternative doubles the tail area: 2 × 0.0105 = 0.0209. Using one tail ignores the other direction.
5. A. The test assumes H0 is true, so the expected counts come from p0. Using p-hat belongs to intervals, and n at least 30 is a rule for means.
6. D. A value outside the 95 percent interval is rejected by the two-sided test at 0.05. The interval is always centered at p-hat, so the statement about p-hat is false, and narrowness is irrelevant.
7. A. Under H0 the sampling distribution has standard deviation sqrt(p0(1 - p0)/n). The mixed formula belongs to an interval-style calculation.
8. C. Since the P-value is below alpha, we reject H0 and state a conclusion about Ha in context. The sample value is not the parameter, and the P-value is not the probability of H0.
9. B. Expected successes under H0 are 7.5, which is below 10, so the normal model is not reliable. Sample size alone does not guarantee that the counts are large enough.
10. State: p = true proportion of students who buy lunch; $H_0 : p = 0.40$, $H_a : p \neq 0.40$, $\alpha = 0.05$.

    Plan: one-proportion z-test. Random sample; $400 < 600$; $400(0.40) = 160$ and $400(0.60) = 240$ are at least 10.

    Do: $\hat{p} = 0.425$, SE $= \sqrt{0.4(0.6)/400} = 0.02449$, $z = 1.02$, P-value $= 2(0.1537) = 0.3074$.

    Conclude: since $0.3074 > 0.05$, we fail to reject H_0. We do not have convincing evidence that the proportion of students who buy lunch differs from 0.40.
11. State: p = true proportion of users who keep notifications on; $H_0 : p = 0.55$, $H_a : p > 0.55$.

    Plan: one-proportion z-test; random; 600 is less than 10 percent of users; $600(0.55) = 330$ and $600(0.45) = 270$ at least 10.

    Do: $\hat{p} = 0.58$; SE $= \sqrt{0.55(0.45)/600} = 0.02031$; $z = 1.48$; P-value $= 0.0698$.

    Conclude: $0.0698 > 0.05$, so fail to reject H_0. We do not have convincing evidence that more than 55 percent of users keep notifications on. (The result would be significant at 0.10, which shows how the conclusion depends on α.)
12. State: p = true proportion of trains arriving on time; $H_0 : p = 0.80$, $H_a : p < 0.80$.

    Plan: random; 250 is less than 10 percent of thousands of trains; $250(0.8) = 200$ and $250(0.2) = 50$ at least 10.

    Do: $\hat{p} = 0.736$; SE $= \sqrt{0.8(0.2)/250} = 0.0253$; $z = -2.53$; P-value $= 0.0057$.

    Conclude: since $0.0057 < 0.05$, reject H_0. We have convincing evidence that fewer than 80 percent of the agency's trains arrive on time.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. (a) 0.50 is outside the interval, so we reject $H_0 : p = 0.50$ at $\alpha = 0.05$.
    (b) 0.55 is inside the interval, so we fail to reject $H_0 : p = 0.55$.
    (c) The interval shows the range of plausible values (about 52 to 60 percent) and so how far from 0.50 the proportion might be; a test result gives only a reject or fail decision.
14. SE $= \sqrt{0.3(0.7)/200} = 0.03240$. For 70 successes, $z = 1.54$ and P-value $= 0.0614 > 0.05$. For 71 successes, $\hat{p} = 0.355$, $z = 1.70$, P-value $= 0.0448 < 0.05$. The smallest count is 71 (equivalent to $\hat{p} > 0.30 + 1.645(0.0324) = 0.353$).
15. Errors: (1) the standard error used $\hat{p}$ instead of p_0; (2) a two-sided alternative requires doubling the tail; (3) 'accept H_0' is not allowed; (4) no context.
    Correct: SE $= \sqrt{0.5(0.5)/150} = 0.04082$, $z = 0.06/0.04082 = 1.47$, P-value $= 2(0.0708) = 0.1416$. Since $0.1416 > 0.05$, we fail to reject H_0. We do not have convincing evidence that the true proportion differs from 0.50.

7.5 Errors, Significance, and Power

PRACTICE · ERRORS, SIGNIFICANCE, AND POWER

1. B. A Type I error is a false rejection of a true null. Failing to reject a false null is Type II, and the other two decisions are correct.
2. C. A Type II error misses a real effect: the test fails to reject a false null. The description of concluding an increase when none exists is the Type I error.
3. A. The significance level is exactly the Type I error probability when H0 is true. The value 0.95 is the probability of correctly not rejecting.
4. D. Larger samples reduce the standard error, so the two curves overlap less. Lowering alpha or looking for a smaller effect both reduce power, and a smaller sample increases variability.
5. C. Power = 1 - beta, so beta = 1 - 0.82 = 0.18. The value 0.82 is the power itself, not the error.
6. B. A stricter cutoff makes rejection harder, so false alarms fall but misses become more common. With a fixed sample size, the two errors trade off.
7. D. Type I rejects a true null: the company claims a benefit that does not exist. The reverse mistake is a Type II error.
8. A. The farther the true value is from the null, the less the two sampling distributions overlap. Sample size and alpha are unchanged between the two scenarios.
9. C. The expected number is 40 × 0.05 = 2. Expecting zero ignores that alpha applies to every test, and 5 confuses 0.05 with 5 tests.
10. (a) Type I: concluding that the proportion who click the new button exceeds 0.10 when it actually equals 0.10. Type II: failing to find convincing evidence that the proportion exceeds 0.10 when it really does.
    (b) Type I: the company may adopt a button that does not improve clicks and waste design effort. Type II: the company keeps the old button and misses a real improvement in clicks.
11. Cutoff: $0.20 + 1.645\sqrt{0.2(0.8)/150} = 0.20 + 1.645(0.03266) = 0.2537$. Under $p = 0.30$, SD $= \sqrt{0.3(0.7)/150} = 0.0374$. Power $= P(Z > (0.2537 - 0.30)/0.0374) = P(Z > -1.24) \approx 0.892$. The test detects a rise from 0.20 to 0.30 about 89 percent of the time.
12. (a) $\beta = 1 - 0.61 = 0.39$.
    (b) Increase the sample size (costs time and money to collect more data). Increase α (accepts a greater chance of a Type I error).
13. (a) Type I: concluding the river exceeds the standard when it does not; consequence: unneeded closure of swimming areas. Type II: failing to detect that the river exceeds the standard when it does; consequence: swimmers exposed to unsafe water.
    (b) The Type II error is more serious for public health, so the county should seek high power, by taking a large sample and perhaps using a larger significance level (such as 0.10) so that real contamination is rarely missed.
14. A larger sample reduces the standard error of $\hat{p}$, so both the null sampling distribution and the alternative sampling distribution become narrower. The rejection cutoff moves closer to 0.30 (about 0.342 instead of 0.384) while the center of the alternative distribution stays at 0.40. A much larger share of the narrow alternative curve then lies beyond the cutoff, which is the power.
15. (a) Failing to reject does not show the null is true; it only means the evidence was not convincing.
    (b) With power 0.44, a real shift to 0.65 would be missed 56 percent of the time, so the pilot's failure to reject is quite plausible even if an effect exists.
    (c) Collect a larger sample (for example $n = 100$ gives power about 0.92 for the same effect) before concluding anything about the absence of an effect.

7.6 Two-Proportion z-Intervals and Tests

PRACTICE · TWO-PROPORTION Z-INTERVALS AND TESTS

1. C. The parameter describes the populations; p-hat values are sample statistics. The pooled proportion is a tool used inside a test, not the target of inference.
2. B. SE = sqrt(0.25/100 + 0.24/150) = sqrt(0.0025 + 0.0016) = 0.0640. Forgetting the square root gives 0.0041, and subtracting the variances under the root gives a much smaller value.
3. D. Pooled = (36 + 54)/(80 + 120) = 90/200 = 0.45. Dividing by 100 or averaging unrelated numbers does not give the combined success rate.
4. A. Independent random samples are needed so the variances can be added. Equal sizes and a count of 30 are not required, and measuring the same individuals twice would create dependence.
5. D. Because 0 lies in the interval, no difference is a plausible value. The interval cannot prove equality, and it does not show Group 1 is higher.
6. A. Under H0 there is one common success rate, and combining the data estimates it best. Pooling has nothing to do with sample size or interval width.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. B. Random assignment balances other factors, so a significant difference supports cause and effect. Without random sampling, generalization is limited to subjects like those in the study.
8. C. SE = sqrt(0.375 × 0.625 x (1/100 + 1/100)) = 0.0685, so z = 0.15 / 0.0685 = 2.19. The value 0.15 is just the difference in proportions, not a z statistic.
9. B. Sample 1 has only 7 successes, fewer than 10. Sample 2 has 90 successes and 110 failures, so its counts are fine.
10. Conditions: independent random samples; each less than 10 percent of its school; counts 156, 144, 129, 151 all at least 10.

    $\hat{p}_1 = 0.52$, $\hat{p}_2 = 0.4607$, difference 0.0593. $\text{SE} = \sqrt{0.52(0.48)/300 + 0.4607(0.5393)/280} = 0.04146$. $\text{ME} = 1.960(0.04146) = 0.0813$. Interval: (-0.022, 0.141).

    We are 95 percent confident that the interval from -0.022 to 0.141 captures the true difference (Elm minus Pine) in the proportions of students with a smartphone. Since 0 is in the interval, we cannot conclude a difference exists.
11. $H_0 : p_1 - p_2 = 0$, $H_a : p_1 - p_2 \neq 0$. Conditions verified: random, independent, 10 percent, pooled counts at least 10.

    $\hat{p}_c = 126/310 = 0.4065$; $\text{SE} = \sqrt{0.4065(0.5935)(1/150 + 1/160)} = 0.0558$; $z = 0.1425/0.0558 = 2.55$; P-value $= 0.0107$.

    Since $0.0107 < 0.05$, reject H_0. We have convincing evidence that the proportions of students who use the library weekly differ between Cavern High and Ridge High.
12. $H_0 : p_A - p_B = 0$, $H_a : p_A - p_B > 0$. Random assignment; counts 63, 27, 48, 52 at least 10.

    $\hat{p}_c = 111/190 = 0.584$; $\text{SE} = 0.0716$; $z = 0.22/0.0716 = 3.07$; P-value $= 0.0011$.

    Reject H_0: we have convincing evidence that Plan A yields a higher success proportion than Plan B for subjects like these. Because of random assignment, Plan A can be credited as the cause.
13. Random and independent: assume both are random samples from separate leagues. 10 percent: $30 < 200$ and $150 < 200$ (10 percent of 2,000), both fine. Large counts: Group 1 has 5 injuries (less than 10), which fails; Group 2 has 40 and 110, fine.

    Because large counts fail for Group 1, the z-interval is not appropriate; collect a larger sample from the first league.
14. $\text{SE} = \sqrt{0.48(0.52)/150 + 0.3375(0.6625)/160} = 0.0553$. $\text{ME} = 1.960(0.0553) = 0.1084$. Interval: 0.1425 ± 0.1084, or (0.034, 0.251).

    The interval excludes 0, consistent with rejecting $H_0 : p_1 = p_2$ at $\alpha = 0.05$. It also shows the difference may be small (3 points) or large (25 points).
15. The before and after observations come from the same players, so the samples are dependent; the two-proportion procedure requires independent samples. A paired approach, comparing each player to himself, would be needed.

    With self-selected groups there is no random assignment, so other differences (motivation, skill) could explain the difference. Only a randomized experiment supports a causal conclusion.

T Unit 7 Test

UNIT TEST · UNIT 7 TEST

1. C. The level is the long-run capture rate of the method. The parameter is fixed, so it has no probability of being in a finished interval, and individuals or the sample proportion are not what the interval describes.
2. A. A z-interval needs at least 10 successes and 10 failures in the sample. Randomness is necessary but not sufficient, and size alone does not rescue small counts.
3. D. Use 0.5 as the planning value: n >= (1.96/0.025)^2 (0.25) = 1536.64, which rounds up to 1537. Rounding down fails to guarantee the margin, and 385 corresponds to a margin of 0.05.
4. B. p-hat = 0.36, SE = sqrt(0.36 × 0.64 / 250) = 0.0304, ME = 1.96 × 0.0304 = 0.0595, so the interval is 0.36 +/- 0.0595. The others are either not centered at 0.36 or use the wrong margin.
5. A. The null is an equality about the parameter and the alternative reflects the suspected increase. Using p-hat or a two-sided alternative does not match the problem.
6. C. A P-value is computed assuming H0 is true and describes how unusual the data are. It is not a probability for either hypothesis.
7. A. p-hat = 0.55, SE = sqrt(0.5 × 0.5 / 80) = 0.0559, z = 0.89, and the upper-tail area is 0.186. Doubling gives 0.371, which is for a two-sided alternative.
8. A. A two-sided test at level 0.01 corresponds to a 99 percent interval, and values outside the interval are rejected. Closeness and narrowness are irrelevant.
9. B. A Type II error is failing to reject a false null, which misses a real effect. The first alternative describes a Type I error.
10. C. A larger sample shrinks the standard error so the null and alternative distributions overlap less. The other changes lower the power.
11. A. Pooled = (45 + 30)/(90 + 110) = 75/200 = 0.375. Averaging the two group proportions (0.5 and 0.273) gives 0.386, which ignores the different group sizes.
12. B. The sampling distribution of p-hat is approximately normal with mean 0.40 and SD sqrt(0.4 × 0.6 / 200) = 0.0346. Then z = 1.44 and the upper-tail area is 0.074. The value 0.926 is the opposite tail.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. Scoring guide.

(a) 1 point: $H_0 : p = 0.70$, $H_a : p < 0.70$ with p defined as the true proportion of this school's graduates over the past decade who enroll in a college program.

(b) 2 points: one-proportion z-test named (1); conditions checked with numbers (1): random sample; $200 < 300$ (10 percent of 3,000); $200(0.70) = 140 \geq 10$ and $200(0.30) = 60 \geq 10$.

(c) 2 points: $\hat{p} = 128/200 = 0.64$; $z = (0.64 - 0.70)/\sqrt{0.70(0.30)/200} \approx -1.85$ (1); P-value $= P(Z \leq -1.85) \approx 0.032$ (1).

(d) 1 point: links the P-value to α and concludes in context.

Model response. Let p be the true proportion of the school's graduates over the past decade who enroll in a college program. $H_0 : p = 0.70$ and $H_a : p < 0.70$, $\alpha = 0.05$. I will use a one-proportion z-test. The sample is random; 200 is less than 10 percent of 3,000 graduates (300); $200(0.70) = 140$ and $200(0.30) = 60$ are both at least 10. $\hat{p} = 0.64$, SE $= \sqrt{0.70(0.30)/200} = 0.0324$, so $z = -0.06/0.0324 \approx -1.85$ and the P-value is about 0.032. Since $0.032 < 0.05$, we reject H_0. We have convincing evidence that the true proportion of this school's graduates who enroll in a college program is less than 0.70.

14. Scoring guide.

(a) 4 points: parameter defined and procedure named (1); conditions verified (1): randomized assignment, counts $84, 36, 66, 64$ all at least 10; calculation (1): $\hat{p}_1 = 0.70$, $\hat{p}_2 = 0.508$, difference 0.192, SE $= \sqrt{0.70(0.30)/120 + 0.508(0.492)/130} \approx 0.0606$, interval $0.192 \pm 1.96(0.0606) = (0.074, 0.311)$; interpretation in context (1).

(b) 1 point: the interval is entirely positive, so 0 is not plausible; there is evidence that text reminders have a higher completion proportion.

(c) 3 points: hypotheses $H_0 : p_1 - p_2 = 0$, $H_a : p_1 - p_2 > 0$ (1); pooled $\hat{p}_c = 150/250 = 0.60$, SE $= \sqrt{0.6(0.4)(1/120 + 1/130)} \approx 0.0620$, $z \approx 3.10$ (1); P-value $\approx 0.001 < 0.05$, reject, convincing evidence in context (1).

(d) 2 points: Type I: concluding text reminders give a higher completion proportion when they do not (1); Type II: failing to find convincing evidence of a higher proportion when text reminders really do give higher completion (0.5); because of random assignment, a causal conclusion for volunteers like these is justified (0.5).

Model response. (a) Let p_1 and p_2 be the true proportions of volunteers like these who would complete the survey with text and with email reminders. Two-proportion z-interval. Subjects were randomly assigned; successes and failures are 84, 36, 66, and 64, all at least 10. $\hat{p}_1 - \hat{p}_2 = 0.70 - 0.508 = 0.192$; SE $= 0.0606$; interval $(0.074, 0.311)$. We are 95 percent confident that the interval from 0.074 to 0.311 captures the true difference (text minus email) in completion proportions. (b) Yes: all values are positive, so 0 is not plausible. (c) $H_0 : p_1 = p_2$, $H_a : p_1 > p_2$. $\hat{p}_c = 0.60$, $z \approx 3.10$, P-value ≈ 0.001. Since $0.001 < 0.05$, reject H_0: convincing evidence that text reminders yield a higher completion proportion. (d) Type I: concluding text reminders are better when the two methods really have the same completion proportion. Type II: not finding convincing evidence that text is better when it truly is. Because volunteers were randomly assigned, the researcher can conclude that text reminders caused the higher completion for volunteers like these (but not necessarily for all people, since the volunteers were not a random sample).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 8: Inference for Means

Use t procedures for one mean, two means, and paired data. 6 lessons, 104 practice and test items, all with full explanations.

In this unit · Exam Unit 4

8.1 The t-Distributions and Conditions for Means
8.2 One-Sample t-Intervals
8.3 One-Sample t-Tests
8.4 Two-Sample t-Intervals
8.5 Two-Sample t-Tests
8.6 Paired Data and Matched Pairs
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

8.1 The t-Distributions and Conditions for Means

Estimating a mean with an unknown population spread requires t-distributions. Learn how they differ from the Normal curve, how to use the table and calculator, and how to check the conditions.

Why This Matters

Almost every real question about a mean starts with a problem: we do not know the population standard deviation σ. A coach wants to estimate the average sprint time of a team, a lab wants to know the mean battery life of a new phone, and in each case the only spread we can compute is the sample standard deviation s. Replacing σ by s adds extra uncertainty, and the **t-distributions** are the correction that pays for it.

Think of this lesson as the foundation for the whole unit. The next five lessons build one-sample intervals and tests, two-sample intervals and tests, and procedures for paired data. All of them use the same idea: a statistic $= \dfrac{\text{estimate} - \text{parameter}}{\text{standard error}}$ compared to a t-distribution, with the same three conditions. Mastering the t-distribution and the conditions here means you only have to learn new formulas, not new ideas, afterward.

Students lose points in two places: they use a z critical value where a t critical value belongs, and they run a t procedure without checking that the data could plausibly come from a roughly Normal population. On the AP exam this lesson supports everything in Exam Unit 4 (Inference for Quantitative Data: Means), and a free-response question will often give you only one point for a correct, in-context check of conditions. This lesson builds both skills so the rest of the unit goes smoothly.

Core Ideas

Suppose a random sample of size n comes from a Normal population with mean μ. If σ were known, the statistic $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ would follow the standard Normal distribution. In practice we compute the **standard error** $\dfrac{s}{\sqrt{n}}$ instead, and the statistic becomes

The t statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

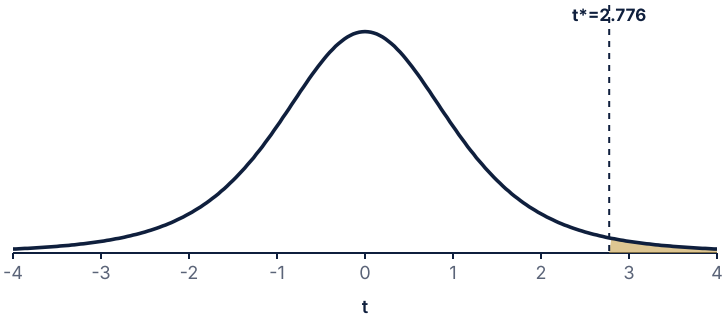
This follows a t-distribution with $df = n - 1$ degrees of freedom when the population is Normal.

The quantity $s/\sqrt{n}$ is the standard error of $\bar{x}$: the typical distance between a sample mean and the true mean.

Because s changes from sample to sample, the t statistic is more variable than a z statistic. That is why the t-distributions have heavier tails. The picture below shows the density curve for a t-distribution with only 4 degrees of freedom. Notice that it is centered at 0, symmetric and bell-shaped, but a noticeable amount of area sits far out in the tails.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Density of a t-distribution with df = 4. The shaded right tail holds about 2.5 percent of the area, whereas a Normal curve would put its 2.5 percent beyond z = 1.96.

Facts about t-distributions

Each t-distribution is symmetric, unimodal and centered at 0, with total area 1.

Its shape is set by one parameter, the degrees of freedom df. For one sample, $df = n - 1$.

Compared with the standard Normal, a t-curve is shorter at the center and has heavier tails. The smaller the df, the heavier the tails.

As df grows, the t-curve approaches the standard Normal curve. Around $df = 30$ they are already close, but they are never identical.

Because the tails are heavier, a t critical value t^* is always larger than the matching z^*, and a t p-value is larger than the matching z p-value.

Why $n - 1$? The sample standard deviation s is built from the n deviations $x_i - \bar{x}$, but those deviations always add to zero, so once $n - 1$ of them are known the last one is forced. Only $n - 1$ pieces of information are free to vary, and that count is the degrees of freedom. It also explains the pattern in the curves: a sample of 3 values gives a very shaky estimate of the spread, so the $df = 2$ curve is spread out and flat, while a sample of 100 values estimates the spread well, so the $df = 99$ curve is nearly the Normal curve.

A good habit is to picture the two curves layered on one another. At the center the Normal curve is a little taller. Out past about 2 units the t-curve is higher. The smaller the df, the earlier the crossover and the larger the difference, which is why a result of 'two standard errors above the mean' is genuinely less impressive when it comes from a tiny sample.

The effect on critical values is easy to see in numbers. Each row below is the critical value for 95 percent confidence, which leaves 2.5 percent in each tail.

Degrees of freedom	t^* (95 percent)	Compare to z^* = 1.960
2	4.303	larger by 2.343
4	2.776	larger by 0.816
10	2.228	larger by 0.268
24	2.064	larger by 0.104
50	2.009	larger by 0.049
100	1.984	larger by 0.024

Notice how fast the values fall: with only 2 degrees of freedom the multiplier is more than twice the Normal value, but by $df = 24$ the gap is under 0.1. In a printed table you would find the row for your df and the column for your tail probability. For example, the row $df = 24$ and upper-tail area 0.025 shows 2.064. The same table also answers the reverse question approximately: if $t = 2.30$ with $df = 24$, the table shows the tail area is between 0.01 and 0.025, so a one-sided p-value is somewhere in that range. The calculator gives the exact value, and either method earns credit as long as you show what you used.

Using the table or the calculator. A printed t table lists t^* by degrees of freedom and by tail area. If your exact df is not listed, use the next smaller df in the table, which is the conservative choice. The built-in calculator is more accurate: use its inverse t function with the area to the left and the degrees of freedom to find t^*, and use its t cumulative distribution function to find tail areas for a computed t. For a two-sided p-value, double the one-tail area.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Conditions for t procedures

Random: the data come from a random sample or a randomized experiment.

Independent (10 percent condition): when sampling without replacement, $n \leq 0.10N$, where N is the population size.

Normal/Large Sample: the population is approximately Normal, or the sample is large ($n \geq 30$). For smaller samples, graph the data and check that there is no strong skewness and no outliers.

The Normal/Large Sample condition is about the *sampling distribution* of $\bar{x}$. By the central limit theorem, $\bar{x}$ is close to Normal for large n no matter what the population looks like. With a small sample we cannot lean on that theorem, so we look at the sample itself. A dotplot, histogram, boxplot or Normal probability plot is a fair check, with the understanding that a small sample cannot prove Normality; we only look for obvious violations. When the graph shows a clear outlier or strong skew and $n < 30$, the t procedure is not trustworthy.

The t procedures are **robust**: they still work reasonably well if the population is only roughly Normal, especially if the sample is not tiny and the distribution is roughly symmetric. They are *not* resistant, though. Because $\bar{x}$ and s are both pulled around by outliers, a single extreme value can ruin a small-sample t analysis. Robustness does not excuse skipping the check.

Sentence frames for conditions

Random: 'The sample was a random sample of ___ from ___.'

10 percent: 'The sample size $n =$ ___ is less than 10 percent of the population of ___.'

Normal/Large Sample: '$n =$ ___ ≥ 30' or 'The ___ (graph) shows no strong skewness or outliers, so a Normal population is plausible.'

Always connect the check to the specific context. A generic 'conditions are met' earns no credit.

Each frame forces a fact from the problem. If a problem says the data come from a randomized experiment instead of a random sample, replace the first line with 'Treatments were randomly assigned' and drop the 10 percent condition, because there is no sampling from a finite population. If the problem gives only summary statistics for a small sample and no graph, you cannot verify Normality and you should say so: the best response is to state that the Normal condition cannot be checked without the data, or that it is assumed as stated in the problem.

Graphs for Checking Normality

Which graph should you draw? Any display of a single quantitative variable can reveal trouble, and each has a different strength:

- **Dotplot or stemplot:** best for very small samples because every value is visible; look for gaps and isolated points.
- **Histogram:** shows overall shape and skew; for n under 15 or so the bars can be misleading because a few values change the picture.
- **Boxplot:** shows skew through the median's position and shows outliers explicitly; it cannot show gaps or multiple peaks.
- **Normal probability plot:** points near a straight line suggest an approximately Normal population; systematic curves suggest skew.

Whatever graph you use, describe what you see in words. 'Roughly symmetric, single peak, no outliers' supports the Normal condition, while 'strongly skewed right with an outlier' does not. Graders look for that sentence, not for the picture alone.

Sample size	What you need to see in the data
$n < 15$	Close to symmetric with no outliers. Any clear skew or outlier means t procedures are not safe.
$15 \leq n < 30$	A single peak, little skew, and no outliers. Mild skew is usually tolerable.
$n \geq 30$	The central limit theorem applies. Still look for extreme outliers or very heavy skew, which can matter even here.

These sample-size ranges are guidelines used in practice, not rigid laws; the key words on the exam are *n at least 30*, or *graph shows no strong skewness or outliers*.

t or z? Choosing the Right Distribution

Students sometimes freeze at the first step: which distribution do I use? The decision depends on the parameter being studied and on what is known about the population spread.

Situation	Distribution for the statistic
Inference about a mean, σ unknown (almost always)	t with $df = n - 1$
Inference about a mean, σ truly known (rare, usually stated)	standard Normal
Inference about a proportion	standard Normal (covered in Unit 7)
Difference of two means, both σ unknown	t with df from the calculator (Lesson 8.4 and 8.5)

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Situation	Distribution for the statistic
Mean of paired differences	t with $df = n - 1$, where n is the number of pairs (Lesson 8.6)

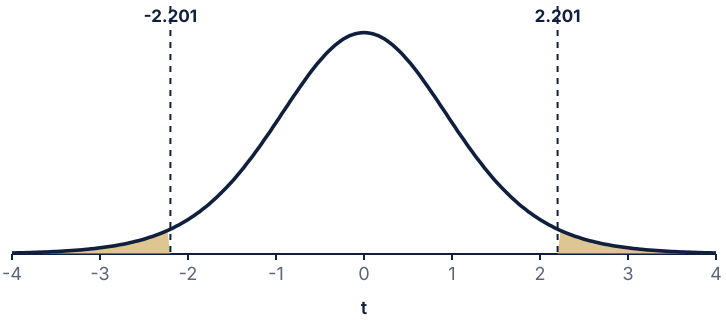
On the AP exam, a problem that gives a sample standard deviation s is a t problem. A problem that gives a population standard deviation σ as a known value is a z problem. If a problem gives neither, you are asked to find one from data, and that means s and a t procedure. Use the word 'unknown' in your explanation: 'Since σ is unknown, we use a t procedure.'

There is also a pattern worth remembering about what changes when you replace a z procedure with a t procedure. The center of the interval (the sample mean) does not move. The standard error is computed with s instead of σ. The multiplier gets bigger. So the t interval is wider than the z interval from the same data, and the t test statistic gives a larger p-value than the z statistic would. This is not a flaw; it is an honest accounting for the fact that s is itself an estimate.

Worked Examples

EXAMPLE 1: FINDING A CRITICAL VALUE

A student plans a 95 percent confidence interval for a mean from a random sample of $n = 12$ measurements. Find the critical value t^* and compare it to z^*.



Central 95 percent of a t-distribution with df = 11.

SOLUTION

The degrees of freedom are $df = n - 1 = 11$.

With 95 percent confidence, 2.5 percent of the area lies above t^*, so t^* has area 0.975 to its left. Using the inverse t function (or the table row $df = 11$, upper tail 0.025) gives $t^* = 2.201$.

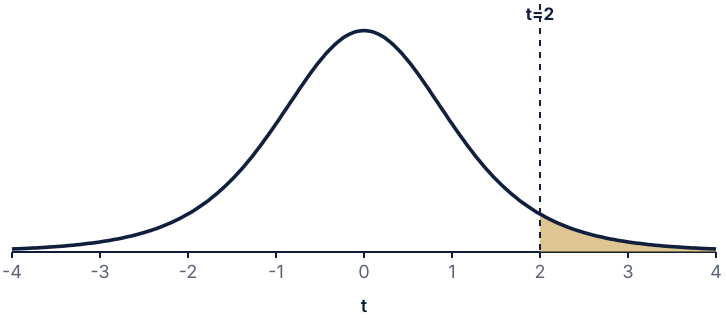
The z critical value is 1.960, so the t value is larger by 0.241. The extra width reflects the extra uncertainty from estimating σ with s.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: HEAVIER TAILS IN ACTION

For a t-distribution with $df = 5$, find $P(T > 2)$ and compare it to $P(Z > 2)$ for the standard Normal.



P(T greater than 2) with df = 5 is about 0.051.

SOLUTION

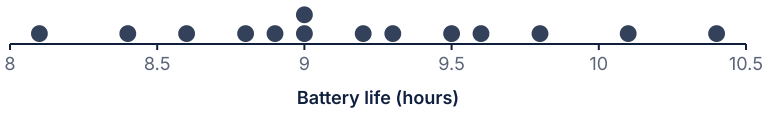
Use the t cumulative distribution function with lower bound 2, upper bound a large number such as 1000, and $df = 5$: $P(T > 2) = 0.0510$.

For the Normal, $P(Z > 2) = 0.0228$.

The t tail area is about 2.2 times as large. A result 2 standard errors above the mean is less surprising when we had to estimate the spread from a small sample.

EXAMPLE 3: CHECKING CONDITIONS WITH A DOTPLOT

A tech reviewer randomly selects 14 phones of one model from a very large shipment and records the battery life, in hours, of each. The dotplot is shown. Can the reviewer use a t procedure to estimate the mean battery life?



SOLUTION

Random: the phones were randomly selected, so the condition is met.

10 percent: the shipment is very large, so 14 phones are far less than 10 percent of it.

Normal/Large Sample: $n = 14 < 30$, so we look at the graph. The dotplot is roughly symmetric and mound-shaped with no outliers and no strong skew, so a Normal population is plausible.

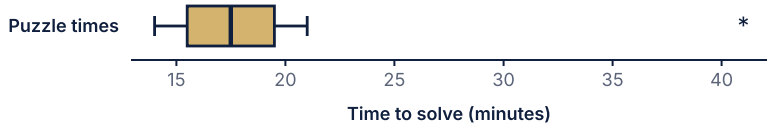
All three conditions are met, so a t procedure with $df = 13$ is reasonable. Notice that we never claimed the data "are Normal"; we only said that nothing in the graph contradicts it. The sample has $\bar{x} = 9.19$ hours and $s = 0.65$ hours.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 4: A SMALL SAMPLE WITH AN OUTLIER

A teacher randomly chooses 12 students from a large school and times how long each takes to solve a logic puzzle, in minutes. The boxplot summarizes the times. Is a t interval appropriate?



Q1 = 15.5, median = 17.5, Q3 = 19.5; the value 41 is an outlier.

SOLUTION

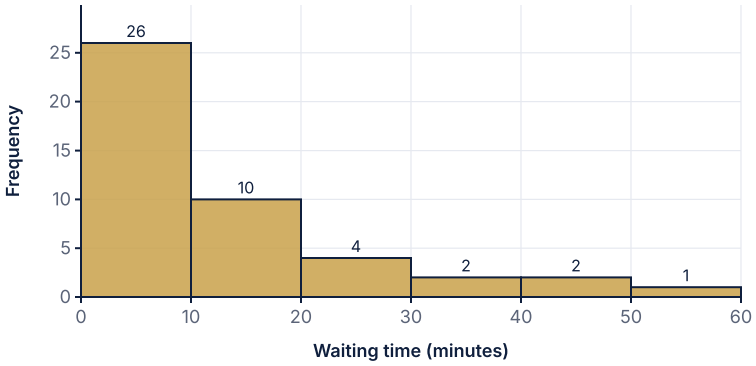
The sample is random and 12 is far below 10 percent of a large school, so those two conditions hold.

But $n = 12 < 30$, and the boxplot shows an extreme high outlier at 41 minutes (the other times run from 14 to 21). The distribution is skewed right.

With a small sample and a clear outlier we cannot treat $\bar{x}$ as approximately Normal, so the t interval is **not** justified. A reasonable next step is to find out whether 41 is a recording error or a real student, and to collect a larger sample.

EXAMPLE 5: SKEWED DATA BUT A LARGE SAMPLE

A city records the number of minutes 45 randomly chosen commuters spend waiting for a bus. The histogram is strongly right skewed. The commuters come from a city of about 80,000 regular bus riders. Are the conditions for a t procedure met?



Strong right skew, n = 45.

SOLUTION

Random: stated. **10 percent:** 45 is much less than 10% of 80,000.

Normal/Large Sample: the histogram is skewed, so the population is clearly not Normal. But $n = 45 \geq 30$, so the central limit theorem makes the sampling distribution of $\bar{x}$ approximately Normal.

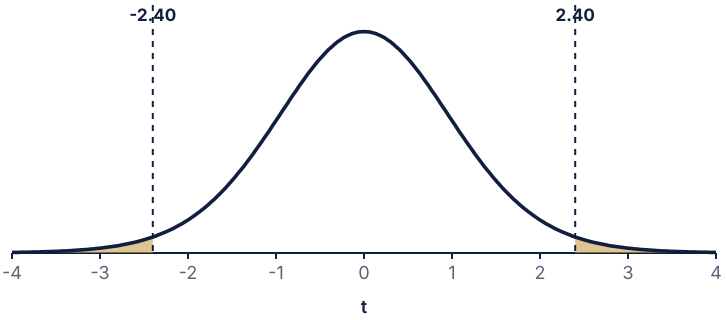
Conditions are met, and this is exactly the situation the central limit theorem was built for: the individual waiting times are skewed, but the average of 45 of them is not. We still note that the mean is a questionable summary for such skewed data (here $\bar{x} = 12.9$ while the median is 8), so we should say the interval is for the *mean* wait, not the typical wait.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 6: A TWO-SIDED P-VALUE

A test statistic computed from a sample of $n = 18$ is $t = -2.40$. Find the two-sided p-value and say how it differs from the Normal-based value.



Both tails beyond |t| = 2.40, df = 17.

SOLUTION

$df = 18 - 1 = 17$. The area below -2.40 is $P(T < -2.40) = 0.0141$.

A two-sided p-value doubles it: $2(0.0141) = 0.0281$.

With a standard Normal, $2P(Z < -2.40) = 0.0164$. The t-based p-value is larger because of the heavier tails, so using z would overstate the evidence.

Write It Like a 5

Prompt. A quality analyst takes a random sample of 10 bags of trail mix from a day's production of thousands of bags and weighs each. The dotplot of the weights shows one bag far below the others (an outlier), and the other nine are tightly clustered. The analyst wants to build a 95 percent confidence interval for the mean weight with a t procedure. Explain whether the conditions are satisfied and state what the analyst should do. [4 points]

Model Response

The bags were randomly selected, so the Random condition is met. Ten bags is far less than 10 percent of the thousands of bags produced in a day, so the 10 percent condition is met. The sample size is $n = 10$, which is less than 30, so the Normal/Large Sample condition depends on the shape of the data. The dotplot shows a clear outlier far below the other nine weights, so the sample is not approximately symmetric and we cannot assume the population is approximately Normal. Therefore the Normal/Large Sample condition is not met, and a t interval would not be trustworthy because one extreme value would heavily influence both $\bar{x}$ and s. The analyst should investigate the low bag (a filling error or a real part of the process), and either correct the data or collect a larger sample (at least 30 bags) before using a t procedure. If the low bag turns out to be a one-time filling error that is removed, the remaining data should be rechecked for shape before proceeding, and the analyst should report that the analysis excluded that bag and why.

What the grader looks for

1 point: confirms the Random condition using the stated random selection. 1 point: checks the 10 percent condition with a reason about population size. 1 point: identifies $n < 30$ and uses the graph, naming the outlier as evidence against Normality. 1 point: reaches a correct conclusion in context (t procedure not justified) and gives a sensible next step. Note: these are Scholar Coach practice rubrics, not official scoring.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Name the parameter and the population in words before anything else.
2. Write the three conditions as a short list: Random, 10 percent, Normal/Large Sample. Check each with a fact from the problem.
3. If $n < 30$, sketch or read the graph. Use the words *skewed*, *outlier*, *roughly symmetric* explicitly.
4. Compute $df = n - 1$ and find t^* or the tail area with the built-in calculator.
5. State in one sentence what the number means in context.

- Before writing, ask: is this about a mean? Is the spread known? Is the data one sample, two samples or pairs?
- Mental model: a t-curve is a Normal curve that is 'nervous' about s, so it keeps extra area in the tails.
- If df is not in the table, round down. The calculator avoids this problem.
- For a quick sense check, t^* for 95 percent confidence is always above 1.96 and falls toward it as n grows.
- A graph of a small sample cannot prove Normality. Say 'no strong skewness or outliers, so Normality is plausible'.

Calculator routine. To find t^* for confidence level C, compute the area to the left as $1 - \frac{1-C}{2}$ and use the inverse t function with that area and the degrees of freedom. To find a p-value for a computed t, use the t cumulative distribution function with a lower and upper bound set to the tail of interest (use ± 1000 as 'infinity'), then double for two-sided tests. Write one line showing what you entered so a grader can follow your work, for example 'tcdf(2, 1000, 5) = 0.051'.

Common traps

1) Using $z^* = 1.96$ with a small sample. 2) Writing $df = n$ instead of $n - 1$. 3) Saying 'the sample is Normal' instead of 'the population is approximately Normal' or 'the sampling distribution is approximately Normal'. 4) Treating a small-sample t procedure as safe without looking at a graph for skewness and outliers.

Mistake Clinic

Spot the error

A student finds the critical value for a 99 percent interval with $n = 8$ by writing '$z^* = 2.576$ because 99 percent confidence.' **Where it fails:** the standard deviation is estimated from a small sample, so the statistic follows a t-distribution with $df = 7$. **Correct solution:** 0.5 percent lies in each tail, so $t^* = 3.499$, noticeably larger than 2.576.

Spot the error

A student says: 'My sample has $n = 40$ so the Normal condition is met.' for a random sample of 40 students from a school of 300 students. **Where it fails:** 40 is more than 10 percent of 300 (the limit is 30), so the independence (10 percent) condition is violated even though the sample is large. **Correct solution:** state that $40 > 0.10(300) = 30$, so the standard error formula would overstate the variability and t procedures should not be used as stated.

Spot the error

A student with $t = 1.90$, $df = 9$ reports the p-value for a two-sided test as $P(T > 1.90) = 0.0449$. **Where it fails:** that is only one tail. **Correct solution:** the alternative is two-sided, so the p-value is $2(0.0449) = 0.0899$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Manufacturers, health clinics and school districts rarely know the true spread of what they measure. A small clinic that checks the resting heart rate of 15 volunteers, or a bakery that weighs 12 loaves, relies on the t-distribution every time it reports a margin of error. Choosing the right distribution is the difference between an honest margin of error and an overconfident one. The same thinking applies to a streaming service testing a small batch of servers, an environmental team sampling a dozen streams, or a school music program timing a few dozen rehearsals: when the sample is small, the honest report widens the margin and checks the data's shape first.

Challenge Corner

For which sample size n does a 95 percent t critical value first drop below 2.10? Find the smallest such n and explain why the t value can never fall to 1.96. **Solution:** we need $t^*(n-1) < 2.10$. Checking: $df = 18$ gives 2.101 and $df = 19$ gives 2.093, so the smallest n is 20. Since the t curve always keeps more area in its tails than the Normal curve for any finite df, the 97.5th percentile is strictly above 1.96, and it approaches 1.96 only as $df \to \infty$.

Practice

PRACTICE The t-Distributions and Conditions for Means

Show your work. A calculator is allowed. Justify conclusions in context.

1. A random sample of 20 values will be used for a 95 percent confidence interval for a mean with σ unknown. What is the critical value t^*?
   (A) 1.960 **(B)** 2.093 **(C)** 2.086 **(D)** 1.729
2. Which statement correctly compares a t-distribution with 4 degrees of freedom to the standard Normal distribution?
   (A) The t-distribution has heavier tails and is lower at its center.
   (B) The t-distribution has lighter tails and is taller at its center.
   (C) The t-distribution is skewed right, while the Normal is symmetric.
   (D) The t-distribution has mean 4, while the Normal has mean 0.
3. For a t-distribution with 8 degrees of freedom, what is $P(T > 2.5)$?
   (A) 0.0062 **(B)** 0.0369 **(C)** 0.0185 **(D)** 0.0205
4. A researcher computes a t statistic from a random sample of 26 observations. How many degrees of freedom does the statistic have?
   (A) 26 **(B)** 24 **(C)** 52 **(D)** 25
5. A random sample of 12 delivery times from a very large service has the boxplot shown. Which statement is best?

   (A) A t procedure is not justified because the sample is small and has an extreme outlier.
   (B) A t procedure is justified because the sample is random.
   (C) A t procedure is justified because the median lies near the center of the box.
   (D) A t procedure is justified because 12 is under 10 percent of the population.
6. A school randomly surveys 60 students from a population of 400 students and wants to use a t procedure for the mean hours of sleep. Which condition fails?
   (A) The Random condition **(B)** The Normal/Large Sample condition
   (C) The 10 percent condition **(D)** None of the conditions fails

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. A 95 percent confidence interval is built from samples of several sizes. As the sample size grows, the critical value t^*
   (A) decreases toward 1.960 but never reaches it
   (B) increases without bound
   (C) stays at 2.576
   (D) becomes smaller than 1.960 for large enough n
8. A dotplot of 10 randomly selected times (population very large) is roughly symmetric with no outliers. Why is a t procedure reasonable?
   (A) The graph gives no evidence against a Normal population, so the Normal condition is plausible.
   (B) Because $n = 10$ guarantees that the sampling distribution is Normal.
   (C) Because the sample mean must equal the population mean.
   (D) Because t procedures require no conditions for small samples.
9. For a 90 percent confidence interval from a random sample of $n = 9$, the critical value is
   (A) 1.833 **(B)** 1.860 **(C)** 1.645 **(D)** 2.306
10. (a) Find the critical value t^* for a 98 percent confidence interval based on a sample of size 16.
    (b) How does it compare with z^* for 98 percent confidence, and why?
11. A t statistic is $t = 1.75$ with $df = 12$.
    (a) Find the p-value for a one-sided test with $H_a : \mu > \mu_0$.
    (b) Find the p-value for a two-sided test.
    (c) What p-value would the standard Normal give for the one-sided test, and why is it different?
12. A cooking class randomly selects 15 cakes from the many baked in one semester and records the baking time in minutes. The times are 3.1, 3.4, 3.6, 3.7, 3.9, 4.0, 4.1, 4.2, 4.4, 4.5, 4.7, 4.9, 5.2, 5.4 and 5.6 minutes; the dotplot of these values is roughly symmetric with no outliers. Check the conditions for a t procedure for the mean baking time.
13. A student says: 'Our sample of 25 values is very skewed, but the 10 percent condition holds, so a t-interval is fine.' (a) Identify the missing condition check. (b) Explain what the student should do or say.
14. A researcher randomly samples 35 homes from a town of 12,000 homes and records the monthly energy use. The histogram is skewed right with no extreme outliers. (a) Are the conditions for t procedures met? (b) Explain why the answer would change if only 8 homes were sampled.
15. Without a calculator, give a reason why the middle 95 percent of a t-distribution with $df = 3$ extends farther from 0 than ± 1.96. Then use your calculator to give the cutoff.

8.2 One-Sample t-Intervals

Build and interpret a confidence interval for one population mean when sigma is unknown. Check conditions, find the margin of error, and explain the meaning of the confidence level in context.

Why This Matters

A confidence interval turns a single sample mean into an honest statement about the population mean. Instead of claiming that the team's true average sprint time is exactly 12.5 seconds, we say we are 90 percent confident that it lies between two numbers. The one-sample t-interval is the tool for a mean when the population standard deviation is unknown, which is nearly every real situation.

Typical losses of points: forgetting to check conditions, using z^* instead of t^*, and interpreting the confidence level incorrectly (for example, saying 95 percent of the data lie in the interval). On the AP exam this skill is tested in Exam Unit 4 as a multiple-choice calculation, a computer-output reading task, and as a four-part free-response question: state, plan (conditions), do, and conclude in context.

Core Ideas

One-sample t-interval for a mean

$\bar{x} \pm t^* \dfrac{s}{\sqrt{n}}$

$\bar{x}$ is the sample mean, s the sample standard deviation, n the sample size.

t^* is the critical value for the confidence level C with $df = n - 1$ (area $\frac{1-C}{2}$ in each tail).

The part after the plus-or-minus sign is the **margin of error**: $ME = t^* \dfrac{s}{\sqrt{n}}$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Every confidence interval has the same skeleton: **statistic $\pm$ (critical value)(standard error of the statistic)**. For one mean the statistic is $\bar{x}$ and the standard error is $s/\sqrt{n}$. The interval is centered at $\bar{x}$, so $\bar{x}$ is the midpoint of the two endpoints, and the margin of error is half the width.

Four-step procedure (State, Plan, Do, Conclude)

State: name the parameter in context, for example 'μ = true mean sprint time of all players on the team', and the confidence level.

Plan: name the procedure (one-sample t-interval for μ) and check Random, 10 percent, and Normal/Large Sample with facts and a graph when $n < 30$.

Do: give $\bar{x}$, s, n, df, t^*, and the interval. Show the formula with numbers.

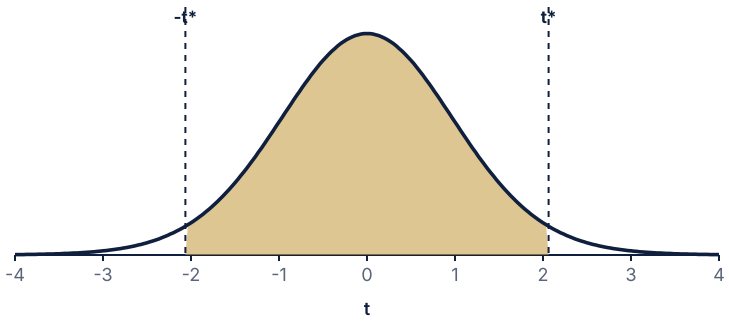
Conclude: 'We are C percent confident that the interval from ___ to ___ captures the true mean ___ (in context).'

What the confidence level means. The 90 percent describes the *method*. If we repeated the sampling process many times and built an interval each time with the same method, about 90 percent of those intervals would capture the true mean. It is not the probability that this one interval is correct (the true mean is a fixed number, so the interval either contains it or does not), and it is not a statement about individual values or about where 90 percent of the data lie.

What controls the width? The margin of error depends on three things, each of which you can reason about without a calculator:

Change	Effect on margin of error	Reason
Higher confidence level	Larger	t^* grows to cover more of the curve
Larger sample size n	Smaller	$s/\sqrt{n}$ shrinks (and t^* falls slightly)
More variable data (larger s)	Larger	The standard error grows

To cut the margin of error in half you must roughly quadruple the sample size, because n enters through $\sqrt{n}$. Increasing the confidence level always costs precision: you cannot make an interval both more confident and narrower without more data.



For a 95 percent interval with n = 25 (df = 24), the middle 95 percent of the t curve lies between -2.064 and 2.064.

Finding t^* fast. Use the inverse t function with area $1 - \frac{1-C}{2}$ and $df = n - 1$. For 90 percent confidence that area is 0.95, for 95 percent it is 0.975, and for 99 percent it is 0.995. Some calculators have a built-in t-interval feature that accepts either summary statistics or a list of data; use it to check your hand work, but still show the formula on the exam because a bare calculator answer earns only partial credit.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: FROM SUMMARY STATISTICS

A random sample of 25 students (from a large school) gives a mean of 48.2 minutes of weekly practice on an instrument, with a standard deviation of 6.5 minutes. Assume the data show no strong skewness or outliers. Construct a 95 percent confidence interval for the mean practice time.

SOLUTION

Parameter: μ = true mean weekly practice time (minutes) for all students at the school. Conditions: random sample; 25 is under 10 percent of a large school; the data show no strong skewness or outliers, so Normality is plausible.

$df = 24$, so $t^* = 2.064$. The standard error is $\frac{6.5}{\sqrt{25}} = 1.3$, so $ME = 2.064(1.3) = 2.68$.

Interval: $48.2 \pm 2.68 = (45.52,\ 50.88)$.

We are 95 percent confident that the true mean weekly practice time for all students at this school is between 45.5 and 50.9 minutes.

EXAMPLE 2: FROM RAW DATA

A coach randomly selects 16 players from a league with hundreds of players and records each player's 100-meter time in seconds. The times are shown below. Construct a 90 percent confidence interval for the league's mean time.

SOLUTION

Parameter: μ = true mean 100-meter time of all players in the league. Conditions: random sample; the league has hundreds of players, so 16 is under 10 percent of them; the dotplot is roughly symmetric with no outliers.

Enter the data in a list: $\bar{x} = 12.525$, $s = 0.465$, $n = 16$, $df = 15$. For 90 percent confidence $t^* = 1.753$.

$ME = 1.753 \cdot \frac{0.465}{\sqrt{16}} = 0.204$, so the interval is $(12.32,\ 12.73)$.

We are 90 percent confident that the true mean 100-meter time for all players in the league is between 12.32 and 12.73 seconds.

EXAMPLE 3: INTERPRETING AND WIDENING

In Example 2 the 90 percent interval was (12.32, 12.73). Explain what happens to the interval if the coach wants 99 percent confidence from the same data, and what happens if the coach quadruples the sample size at 90 percent confidence while the sample mean and standard deviation stay the same.

SOLUTION

For 99 percent confidence with $df = 15$, $t^* = 2.947$ instead of 1.753. The margin of error grows from 0.204 to 0.343, so the interval becomes (12.18, 12.87). It is centered at the same $\bar{x}$ but wider.

With $n = 64$ and 90 percent confidence, $df = 63$ and $t^* = 1.669$, so $ME = 1.669 \cdot \frac{0.465}{8} = 0.097$, roughly half of the original margin of error.

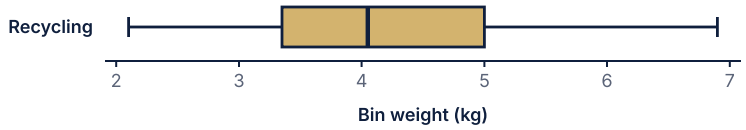
Greater confidence costs width; more data buys precision.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 4: CHECKING CONDITIONS WITH A BOXPLOT

A recycling center randomly selects 20 of the many bins it collects in one week and weighs the contents in kilograms. The boxplot is shown. Construct a 99 percent confidence interval for the mean bin weight.



Boxplot of 20 bin weights.

SOLUTION

Conditions: random sample; 20 bins are under 10 percent of the many bins; $n = 20 < 30$, but the boxplot is roughly symmetric (median near the center of the box) with no outliers, so a Normal population is plausible.

$\bar{x} = 4.215$, $s = 1.237$, $n = 20$, $df = 19$, $t^* = 2.861$ for 99 percent confidence.

$ME = 2.861 \cdot \dfrac{1.237}{\sqrt{20}} = 0.792$ and the interval is $(3.42,\ 5.01)$.

We are 99 percent confident that the true mean weight of all bins collected that week is between 3.42 and 5.01 kilograms.

EXAMPLE 5: USING AN INTERVAL TO EVALUATE A CLAIM

The recycling center's manager claims the mean bin weight is 3.5 kg. Does the interval from Example 4 provide evidence against the claim?

SOLUTION

The 99 percent interval is (3.42, 5.01). The claimed value 3.5 kg is inside this interval.

Because 3.5 is a plausible value for the mean, the data do not give convincing evidence against the claim at the 1 percent significance level (a two-sided test at $\alpha = 0.01$ corresponds to a 99 percent interval).

Note the wording: we did not prove the claim true, we only failed to find convincing evidence against it.

EXAMPLE 6: WORKING BACKWARD FROM AN INTERVAL

A random sample of 30 phone users gives a 95 percent t-interval for the mean daily screen time (minutes) of (68.11, 76.69). Find the sample mean, the margin of error, and the sample standard deviation.

SOLUTION

The mean is the midpoint: $\bar{x} = \dfrac{68.11 + 76.69}{2} = 72.4$.

The margin of error is half the width: $ME = \dfrac{76.69 - 68.11}{2} = 4.294$.

With $df = 29$, $t^* = 2.045$. Then $SE = \dfrac{ME}{t^*} = 2.100$ and $s = SE \cdot \sqrt{30} = 11.50$ minutes.

Write It Like a 5

Prompt. A music teacher randomly samples 20 students from a district of several thousand and records how many minutes each student practices on a school day. The sample mean is 41.5 minutes and the standard deviation is 9.8 minutes; a boxplot of the data is roughly symmetric with no outliers. Construct and interpret a 95 percent confidence interval for the mean daily practice time in the district. [4 points]

Model Response

State. Let μ be the true mean number of minutes per school day that students in the district practice. We construct a 95 percent one-sample t-interval for μ. **Plan.** Random: the students were randomly sampled. 10 percent: 20 students are less than 10 percent of several thousand. Normal/Large Sample: $n = 20 < 30$, but the boxplot is roughly symmetric with no outliers, so the population is plausibly approximately Normal. **Do.** $df = 19$ and $t^* = 2.093$, so $41.5 \pm 2.093\dfrac{9.8}{\sqrt{20}} = 41.5 \pm 4.59$, giving $(36.91,\ 46.09)$. **Conclude.** We are 95 percent confident that the true mean daily practice time for students in the district is between 36.91 and 46.09 minutes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

1 point: defines the parameter and names the procedure. 1 point: checks Random, 10 percent and Normal/Large Sample with context and uses the boxplot. 1 point: correct interval with correct df and t^* (the formula or calculator input is shown). 1 point: conclusion states the confidence level, the interval, and the parameter in context. These are Scholar Coach practice rubrics, not official scoring.

Coach's Playbook

1. Write the parameter first, with units and the word *true* or *population*.
2. List the three conditions and attach one fact to each.
3. Compute df, then t^* with the inverse t function. Write it down.
4. Compute ME and the endpoints; round only at the end (keep an extra decimal).
5. Sentence template: 'We are C percent confident that the interval from ___ to ___ captures the true mean ___.'

- Check reasonableness: the interval must be centered at $\bar{x}$, and a 95 percent interval for a mean is usually about $\bar{x} \pm 2$ standard errors.
- The interval estimates a **mean**, not an individual value. Most individual data points lie outside it.
- If a problem asks 'what would happen if ...', answer in terms of t^* and $\sqrt{n}$ before computing anything.
- If you use the calculator's interval feature, still write the formula and the numbers you entered.

Common traps

1) Using $df = n$ rather than $n - 1$. 2) Interpreting the level as 'there is a 95 percent chance the true mean is in my interval' or '95 percent of the sample falls in the interval'. 3) Dividing by n instead of $\sqrt{n}$ in the standard error. 4) Skipping the graph when $n < 30$ and writing only 'conditions are met'.

Mistake Clinic

Spot the error

A student computes $41.5 \pm 2.093(9.8/20)$ for the 20-student sample above. **Where it fails:** the denominator used $n = 20$ instead of $\sqrt{20}$. **Correct solution:** $SE = 9.8/\sqrt{20} = 2.191$, so $ME = 2.093(2.191) = 4.59$.

Spot the error

A student writes: 'There is a 95 percent probability that the true mean practice time is between 36.9 and 46.1.' **Where it fails:** the true mean is a fixed (unknown) number; the 95 percent describes the long-run success rate of the method. **Correct solution:** 'We are 95 percent confident that the interval from 36.9 to 46.1 minutes captures the true mean daily practice time of students in the district.'

Spot the error

A student says the interval (36.9, 46.1) means 95 percent of students practice between those times. **Where it fails:** the interval is about the mean, not individuals. With $s = 9.8$ the individual times vary much more (roughly $41.5 \pm 2(9.8)$), so far more than the interval's width. **Correct solution:** 'The interval estimates the mean practice time of all students.'

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Real-World Connection

Health clinics, environmental labs, streaming companies and factories all publish intervals rather than single averages, because a single average hides how uncertain it is. When a news report says that the average commute is 31 minutes 'plus or minus 2', the plus-or-minus is a margin of error, and the quality of the estimate depends on whether the sample was random and large enough.

Challenge Corner

A pilot sample of 25 gives $s = 6.5$ and a 95 percent margin of error of 2.68. Roughly what sample size would give a margin of error of 1.0 at 95 percent confidence, assuming s stays about 6.5? **Solution:** use $z^* \approx 1.96$ as a first guess: $n \geq (1.96 \cdot 6.5/1.0)^2 = 162.3$, so about 163 students. Recheck with t^*: for $n = 163$, $df = 162$ and $t^* = 1.975$, giving $ME = 1.975(6.5)/\sqrt{163} = 1.005$, which is slightly above 1.0, so go up to $n = 165$ where the margin is just under 1.0. The lesson: you cannot find n exactly because t^* depends on n, but z gives a close start.

Practice

PRACTICE **One-Sample t-Intervals**

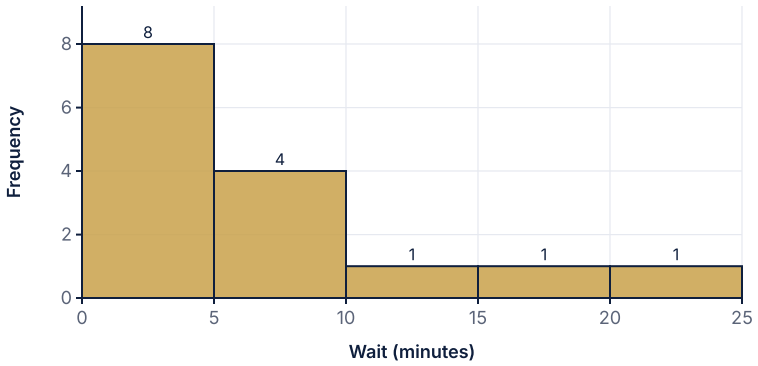
Show your work. A calculator is allowed. Justify conclusions in context.

1. A random sample of 36 batteries (from a large shipment) has mean life 21.4 hours and standard deviation 3.0 hours. What is the margin of error for a 95 percent confidence interval for the mean life?
   (A) 1.015 **(B)** 0.980 **(C)** 0.169 **(D)** 6.084
2. Which is the correct interpretation of a 90 percent confidence interval of (12.2, 12.8) seconds for a mean sprint time?
   (A) There is a 90 percent probability that a randomly chosen player runs between 12.2 and 12.8 seconds.
   (B) We are 90 percent confident the true mean sprint time is between 12.2 and 12.8 seconds.
   (C) 90 percent of the sample times were between 12.2 and 12.8 seconds.
   (D) The sample mean is between 12.2 and 12.8 seconds with probability 0.90.
3. All else equal, which change would make a confidence interval for a mean narrower?
   (A) Increasing the confidence level
   (B) Using a sample with a larger standard deviation
   (C) Increasing the sample size
   (D) Decreasing the sample size
4. A 95 percent t-interval for a mean is (40.6, 50.2) from a random sample of 20 values. What is the sample mean?
   (A) 4.8 **(B)** 50.2 **(C)** 9.6 **(D)** 45.4
5. A 95 percent t-interval from a sample of 25 has margin of error 2.06. About what is the standard error $s/\sqrt{n}$ of the sample mean?
   (A) 2.06 **(B)** 1.00 **(C)** 4.25 **(D)** 0.41

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. A random sample of 15 wait times from a very large call center is summarized by the histogram. Is a one-sample t-interval for the mean wait appropriate?



(A) Yes, because the sample is random.
(B) Yes, because 15 is less than 10 percent of the population.
(C) No, because $n < 30$ and the data are strongly skewed right.
(D) No, because the sample standard deviation is unknown.

7. What is the critical value t^* for a 99 percent confidence interval for a mean based on a random sample of 12 values?
(A) 3.106 **(B)** 3.055 **(C)** 2.576 **(D)** 2.718

8. A researcher wants to cut the margin of error of a t-interval in about half without changing the confidence level. What should be done to the sample size, assuming s stays the same?
(A) Multiply it by about 2 **(B)** Multiply it by about 4 **(C)** Increase it by 4
(D) Multiply it by about 8

9. Ten randomly chosen soil samples (from a very large field) have pH values 6.8, 7.1, 6.9, 7.4, 7.0, 7.3, 6.7, 7.2, 7.1, 6.9, with no outliers or strong skew. Which is the 95 percent confidence interval for the mean pH?
(A) (6.90, 7.18) **(B)** (6.91, 7.17) **(C)** (6.54, 7.54)
(D) (6.88, 7.20)

10. A random sample of 40 students (from a school of 1,500) spends a mean of 3.2 hours per week reading for pleasure, with a standard deviation of 1.1 hours. (a) Check the conditions. (b) Construct a 90 percent confidence interval. (c) Interpret the interval in context.

11. A bakery randomly selects 12 loaves from a day's production of several hundred and records baking times in minutes: 41, 44, 38, 46, 43, 40, 45, 42, 39, 44, 47, 43. A dotplot shows no outliers and no strong skew. Construct and interpret a 95 percent confidence interval for the mean baking time. (Show the conditions.)

12. A student states: 'Because my 95 percent interval for mean commute time is (24.1, 29.5) minutes, 95 percent of commuters have commutes between 24.1 and 29.5 minutes.' (a) Explain the error. (b) Write a correct interpretation. (c) Would a 99 percent interval from the same data be wider or narrower? Why?

13. A 95 percent t-interval for a mean, based on a random sample of 30 values, is (18.6, 23.4).
(a) Find the sample mean and the margin of error.
(b) Find the standard error and the sample standard deviation.

14. A cereal maker claims boxes contain a mean of 500 grams. A random sample of 16 boxes (from a very large production run) weighs 497, 503, 495, 499, 501, 494, 498, 496, 500, 497, 495, 499, 502, 496, 498, 497 grams; the data show no outliers or strong skew. Build a 99 percent confidence interval and say whether it gives convincing evidence against the claim.

15. Two random samples are taken from the same population with the same standard deviation $s = 8$: one of size 25 and one of size 100. Compute the 95 percent margin of error for each and explain the relationship in terms of sample size.

8.3 One-Sample t-Tests

Test a claim about one population mean with the four-step process: hypotheses, conditions, t statistic and p-value, and a conclusion in context, including errors and power.

Why This Matters

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A hypothesis test answers a yes-or-no style question with data: does a sample give convincing evidence that a mean differs from a claimed value? A snack company says its bags average 50 grams, a coach says a new training plan lowers sprint times, a school says students sleep 8 hours on average. In each case the data come from one random sample and the population spread is unknown, so the **one-sample t-test** is the tool.

The common errors are not arithmetic. Students write hypotheses with $\bar{x}$ instead of μ, forget to verify conditions, misread the p-value as the probability that the null hypothesis is true, or finish with 'reject' and no sentence about the context. On the AP exam, Exam Unit 4 tests this lesson through a four-step free-response question and several multiple-choice items that interpret p-values, hypotheses and errors.

Core Ideas

A significance test begins with two competing claims about a parameter. The **null hypothesis** H_0 is the claim being tested, always with an equal sign: $H_0 : \mu = \mu_0$. The **alternative hypothesis** H_a says what we suspect instead and is one of $\mu < \mu_0$, $\mu > \mu_0$ or $\mu \neq \mu_0$. The direction comes from the research question, written before looking at the data. Both hypotheses are about the population parameter μ and never about $\bar{x}$.

One-sample t-test for a mean

$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ with $df = n - 1$.

The numerator measures how far the sample mean is from the null value; the denominator is the standard error.

p-value: the probability, assuming H_0 is true, of getting a sample mean at least as far from μ_0 as the observed one, in the direction(s) of H_a. For $H_a : \mu \neq \mu_0$ double the one-tail area.

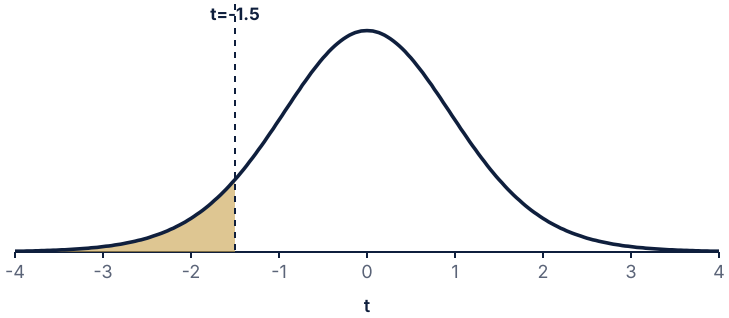
Four-step test (State, Plan, Do, Conclude)

State: H_0 and H_a in symbols, the parameter in context, and the significance level α (use 0.05 if none is given).

Plan: name the one-sample t-test for a mean and verify Random, 10 percent, and Normal/Large Sample (graph if $n < 30$).

Do: compute $\bar{x}$, s, n, t, df and the p-value; sketch or describe the tail area.

Conclude: compare the p-value to α and state the decision about H_0, then the conclusion about H_a in context.



Lower-tail p-value for a left-tailed test with df = 14: the area at or below the observed t.

Decisions. If the p-value is less than or equal to α, we **reject** H_0: the data provide convincing evidence for H_a. If the p-value is greater than α, we **fail to reject** H_0: the data do not provide convincing evidence for H_a. We never say 'accept H_0' or 'prove H_0', because a sample can fail to detect a real difference.

What a p-value means. A small p-value says the observed result would be surprising if H_0 were true. It is *not* the probability that H_0 is true, and it is not the probability that the result happened by chance. A p-value of 0.03 in a test of $H_0 : \mu = 50$ means: if the true mean were 50, there would be about a 3 percent chance of getting a sample mean this far or farther from 50 (in the direction of H_a) just from random sampling.

Errors and the significance level

Type I error: rejecting H_0 when it is actually true. Its probability is α.

Type II error: failing to reject H_0 when H_a is actually true.

Power = $1 - P(\text{Type II error})$. Power rises with larger n, larger α, and a true mean farther from μ_0.

Connection to intervals. For a two-sided test at level α, the decision agrees with a confidence interval of level $1 - \alpha$: if μ_0 is outside the interval, reject H_0; if μ_0 is inside, fail to reject. This shortcut works only for two-sided alternatives.

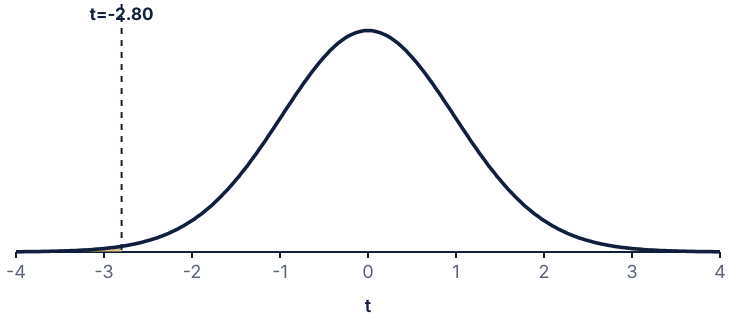
Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Worked Examples

EXAMPLE 1: A FULL TEST FROM SUMMARY STATISTICS

A tablet maker claims a charging time of 45 minutes. A random sample of 40 tablets from a large production run has mean charging time 42.3 minutes with standard deviation 6.1 minutes. The engineer suspects the true mean is lower than the claim. Test at the 5 percent level of significance.



The p-value is the area to the left of the observed t.

SOLUTION

State. $H_0 : \mu = 45$ and $H_a : \mu < 45$, where μ is the true mean charging time of all tablets from this run (minutes). $\alpha = 0.05$.

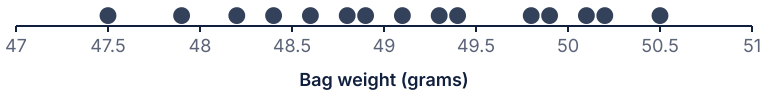
Plan. One-sample t-test for a mean. Random: stated. 10 percent: 40 is less than 10 percent of a large run. Normal/Large Sample: $n = 40 \geq 30$.

Do. $t = \dfrac{42.3 - 45}{6.1/\sqrt{40}} = -2.80$, $df = 39$, p-value $= P(T < -2.80) = 0.0040$.

Conclude. Since 0.0040 is less than 0.05, we reject H_0. There is convincing evidence that the true mean charging time of tablets from this run is less than 45 minutes.

EXAMPLE 2: RAW DATA, SMALL SAMPLE

A snack company claims bags weigh 50 grams on average. An inspector randomly selects 15 bags from a very large day's production, weighs each, and suspects the true mean is lower. The weights (grams) are 48.6, 49.8, 50.2, 47.9, 49.1, 48.8, 50.5, 49.4, 48.2, 49.9, 47.5, 50.1, 48.9, 49.3 and 48.4. The dotplot is shown. Test at $\alpha = 0.05$.



SOLUTION

State. $H_0 : \mu = 50$ vs. $H_a : \mu < 50$, where μ is the true mean weight (g) of bags from that day.

Plan. One-sample t-test. Random: stated. 10 percent: 15 bags is far less than 10 percent of a very large production. Normal: $n = 15 < 30$, but the dotplot is roughly symmetric with no outliers.

Do. $\bar{x} = 49.107$, $s = 0.891$, $n = 15$. $t = \dfrac{49.107 - 50}{0.891/\sqrt{15}} = -3.88$, $df = 14$, p-value $= 0.0008$.

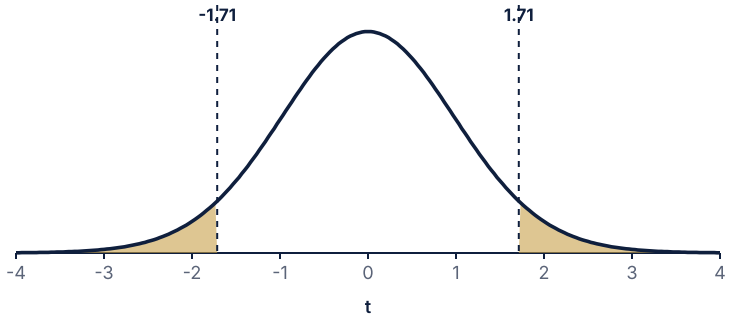
Conclude. Since 0.0008 is less than 0.05, we reject H_0. There is convincing evidence that the true mean weight of bags that day is less than 50 grams.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: TWO-SIDED TEST AND THE MATCHING INTERVAL

A school claims students average 3.1 hours of homework per night. A random sample of 36 students (from a large school) averages 2.9 hours with standard deviation 0.7 hours. Test whether the true mean differs from 3.1 at $\alpha = 0.05$ and compare with a 95 percent confidence interval.



Two-sided p-value: both tails beyond |t|.

SOLUTION

State. $H_0 : \mu = 3.1$ vs. $H_a : \mu \neq 3.1$.

Plan. Random and 10 percent given; $n = 36 \geq 30$ satisfies Normal/Large Sample.

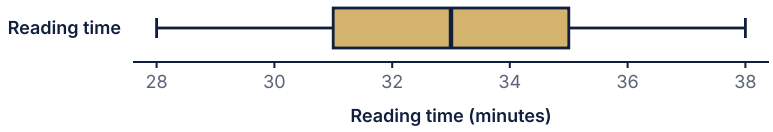
Do. $t = \dfrac{2.9 - 3.1}{0.7/\sqrt{36}} = -1.714$, $df = 35$. Two-sided p-value $= 2P(T < -1.714) = 0.0953$.

Conclude. Since 0.0953 is greater than 0.05, we fail to reject H_0. The data do not provide convincing evidence that the true mean homework time differs from 3.1 hours.

Check with an interval. The 95 percent interval is $2.9 \pm 2.030(0.7/6) = (2.66,\ 3.14)$, which contains 3.1, consistent with failing to reject.

EXAMPLE 4: A RIGHT-TAILED TEST WITH A BOXPLOT

A teacher believes a new routine increases the mean time students spend reading to more than the 30-minute goal. A random sample of 18 students from a large school reads for the times (minutes) 31, 34, 29, 36, 33, 28, 35, 32, 37, 30, 31, 34, 29, 38, 33, 35, 32, 36; the boxplot is shown. Test at $\alpha = 0.01$.



SOLUTION

State. $H_0 : \mu = 30$ vs. $H_a : \mu > 30$; μ is the true mean reading time of students at the school after the routine.

Plan. Random: stated. 10 percent: 18 is less than 10 percent of a large school. Normal: $n = 18 < 30$, but the boxplot is roughly symmetric with no outliers.

Do. $\bar{x} = 32.944$, $s = 2.920$, $t = \dfrac{32.944 - 30}{2.920/\sqrt{18}} = 4.28$, $df = 17$, p-value $= P(T > 4.28) = 0.0003$.

Conclude. Since 0.0003 is less than 0.01, we reject H_0. There is convincing evidence that the true mean reading time exceeds 30 minutes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: INTERPRETING P-VALUES AND ERRORS

In Example 2 the p-value is 0.0008. (a) Interpret it. (b) If the true mean weight were actually 50 grams, which error would the inspector have made? (c) Describe a Type II error in context.

SOLUTION

(a) If the true mean bag weight were 50 grams, the probability of getting a sample mean as low as 49.11 g or lower from a random sample of 15 bags is about 0.0008.

(b) Rejecting a true H_0 is a **Type I error**: concluding the mean is below 50 grams when it is really 50 grams.

(c) A **Type II error** would be failing to find convincing evidence that the mean is below 50 grams when the bags really do average less than 50 grams, so a real underfilling problem would go unnoticed.

EXAMPLE 6: READING COMPUTER OUTPUT

A random sample of 24 tutoring sessions (from many) is used to test whether the mean session length exceeds 15 minutes. Software reports the output below, and the histogram of lengths is roughly symmetric. What is the conclusion at $\alpha = 0.05$?

N	Mean	StDev	SE Mean	T	P
24	16.2	3.9	0.80	1.51	0.073

SOLUTION

State. $H_0 : \mu = 15$ vs. $H_a : \mu > 15$.

Plan. The conditions were checked (random, 24 is less than 10 percent of many sessions, roughly symmetric graph).

Do. The output gives $t = 1.51$ with $df = 23$ and one-sided p-value 0.0727. As a check, $\frac{16.2 - 15}{3.9/\sqrt{24}} = 1.51$.

Conclude. Since 0.0727 is greater than 0.05, we fail to reject H_0. The data do not provide convincing evidence that the true mean session length is greater than 15 minutes.

Write It Like a 5

Prompt. A company claims its LED bulbs last 800 hours on average. A consumer group tests 12 randomly selected bulbs from a very large batch and finds a mean of 771.5 hours with a standard deviation of 47.3 hours. A dotplot shows no outliers and no strong skew. Is there convincing evidence that the mean lifetime is less than 800 hours? Use $\alpha = 0.05$. [4 points]

Model Response

State. $H_0 : \mu = 800$ and $H_a : \mu < 800$ where μ is the true mean lifetime (hours) of bulbs in the batch; $\alpha = 0.05$. **Plan.** One-sample t-test for a mean. Random: bulbs were randomly selected. 10 percent: 12 is less than 10 percent of a very large batch. Normal: $n = 12 < 30$, but the dotplot shows no outliers or strong skew, so a Normal population is plausible. **Do.** $t = \frac{771.5 - 800}{47.3/\sqrt{12}} = -2.09$, $df = 11$, p-value $= 0.0305$. **Conclude.** Since the p-value 0.0305 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that the true mean lifetime of bulbs in the batch is less than 800 hours.

What the grader looks for

1 point: correct hypotheses in terms of μ with the parameter defined. 1 point: names the test and checks all three conditions with context. 1 point: correct t, df and p-value. 1 point: decision linked to the p-value and α, with a conclusion about the mean lifetime in context. Practice rubric, not official scoring.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Decide the direction of H_a from the question wording before touching the data: 'lower' means $<$, 'higher' means $>$, 'different' means $\neq$.
2. Write the conditions with context.
3. Compute t and df, then the p-value from the correct tail(s).
4. Compare p to α and write the decision in the language 'reject / fail to reject H_0'.
5. Finish with a sentence that names the parameter and the direction of H_a.

- The sign of t should match the direction of the data: $\bar{x} < \mu_0$ gives negative t. If the sign disagrees with H_a, the p-value is above 0.5 and you will fail to reject.
- On the calculator, use the built-in t-test for a mean with either summary statistics or a list, and choose the alternative. Copy t, df and p-value into your response.
- A large sample can make a tiny difference statistically significant. Always say what the difference means in practice.
- The words 'convincing evidence' belong in the conclusion, and 'at the 5 percent significance level' can be added.

Common traps

1) Writing hypotheses with $\bar{x}$. 2) Using the wrong tail or forgetting to double for a two-sided test. 3) 'Accepting H_0' when p is large. 4) Interpreting the p-value as $P(H_0 \text{ is true})$ or giving a conclusion without context.

Mistake Clinic

Spot the error

A student tests whether bags weigh less than 50 grams and writes '$H_0 : \bar{x} = 50$ vs. $H_a : \bar{x} < 50$.' **Where it fails:** hypotheses describe the population mean, not the observed sample mean, which is known. **Correct solution:** $H_0 : \mu = 50$ vs. $H_a : \mu < 50$, where μ is the true mean weight of the bags.

Spot the error

With p-value 0.0953 in the homework example, a student writes 'We accept H_0 and conclude the mean is exactly 3.1.' **Where it fails:** a large p-value means the sample gives no convincing evidence against H_0; it does not prove H_0. **Correct solution:** 'We fail to reject H_0; the data do not provide convincing evidence that the true mean homework time differs from 3.1 hours.'

Spot the error

A student reports for the two-sided test that p-value $= P(T < -1.714) = 0.0477$. **Where it fails:** the alternative is two-sided, so both tails count. **Correct solution:** $2(0.0477) = 0.0953$, which is greater than 0.05, so the decision is to fail to reject.

Real-World Connection

Quality control teams, clinical researchers and product testers use one-sample t-tests whenever a claimed average must be checked against a small sample. The p-value is the standard way to say how surprising the sample would be if the claim were true, and an honest conclusion always mentions what could have gone wrong (Type I or Type II error) and what the practical difference is.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

The snack test in Example 2 ($n = 15$, $s = 0.891$) rejects $H_0 : \mu = 50$ in favor of $\mu < 50$ at $\alpha = 0.05$ whenever $t \le t_{0.05} = -1.761$. **Question:** what is the largest sample mean for which the inspector would reject H_0? **Solution:** solve $\dfrac{\bar{x} - 50}{0.891/\sqrt{15}} \le -1.761$, giving $\bar{x} \le 50 + (-1.761)(0.230) = 49.59$ grams. Any sample mean at or below 49.59 g leads to rejection, so the observed 49.11 g is well inside the rejection region.

Practice

PRACTICE

One-Sample t-Tests

Show your work. A calculator is allowed. Justify conclusions in context.

1. A researcher wants to test whether the mean amount of sleep of students at a school is less than 8 hours. Which hypotheses are correct?

   (A) $H_0 : \mu = 8$, $H_a : \mu < 8$ **(B)** $H_0 : \bar{x} = 8$, $H_a : \bar{x} < 8$ **(C)** $H_0 : \mu < 8$, $H_a : \mu = 8$

   (D) $H_0 : \mu = 8$, $H_a : \mu \neq 8$

2. A one-sample t-test of $H_a : \mu < \mu_0$ gives $t = -2.10$ with a sample of 20. What is the p-value?

   (A) 0.0493 **(B)** 0.0247 **(C)** 0.0179 **(D)** 0.9753

3. The p-value of a test is 0.032. Which is the best interpretation?

   (A) There is a 3.2 percent chance that H_0 is true.

   (B) There is a 96.8 percent chance that H_a is true.

   (C) If H_0 were true, there is about a 3.2 percent chance of a sample result at least as extreme as the one observed.

   (D) 3.2 percent of the population differs from the null value.

4. At $\alpha = 0.05$ a test gives p-value 0.18. What is the correct conclusion?

   (A) Reject H_0; there is convincing evidence for H_a.

   (B) Accept H_0; the null hypothesis is true.

   (C) Reject H_0 because 0.18 is greater than 0.05.

   (D) Fail to reject H_0; the data do not provide convincing evidence for H_a.

5. In a test of $H_0 : \mu = 50$ vs. $H_a : \mu \neq 50$ at $\alpha = 0.05$, a 95 percent confidence interval for μ is (50.8, 54.2). What conclusion does the interval support?

   (A) Reject H_0, since 50 is not in the interval.

   (B) Fail to reject H_0, since the interval is narrow.

   (C) Reject H_0, since the interval contains the sample mean.

   (D) Fail to reject H_0, since 50.8 is above 50.

6. A random sample of 14 times from a very large population is used to test $H_0 : \mu = 20$ vs. $H_a : \mu > 20$. The boxplot of the data is shown. Which statement is correct?

   (A) All conditions are met because the sample is random.

   (B) The Normal condition is not met because of the strong outlier with $n < 30$.

   (C) The 10 percent condition fails because $n = 14$.

   (D) The test is valid because the median is above 20.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

7. A pharmacy tests $H_0 : \mu = 100$ (mean fill volume in mL) against $H_a : \mu < 100$. What is a Type I error in this context?
   (A) Concluding the mean fill is 100 mL when it is actually less than 100 mL.
   (B) Choosing the wrong sample size.
   (C) Concluding the mean fill is less than 100 mL when it is actually 100 mL.
   (D) Computing the standard deviation incorrectly.
8. A random sample of 25 values has $\bar{x} = 103$ and $s = 10$. For testing $H_0 : \mu = 100$, what is the t statistic?
   (A) 7.50 **(B)** 0.30 **(C)** 0.60 **(D)** 1.50
9. Which of the following would increase the power of a one-sample t-test, all else equal?
   (A) Decrease the significance level α
   (B) Increase the sample size
   (C) Move the true mean closer to μ_0
   (D) Use a sample with larger variability
10. Write the null and alternative hypotheses (in symbols, defining μ) for each claim.
    (a) A bakery says its muffins average 110 grams; an inspector thinks they are heavier.
    (b) A coach thinks the new warm-up changes the mean sprint time from 12.5 seconds.
    (c) A school claims the mean commute is 25 minutes; a parent thinks it is longer.
11. A school says students average 8 hours of sleep. A random sample of 10 students from a large school reports 8.2, 7.4, 8.8, 7.9, 8.5, 7.1, 8.0, 7.6, 8.4 and 7.3 hours; a dotplot is roughly symmetric with no outliers. Is there convincing evidence that the true mean differs from 8 hours? Use $\alpha = 0.05$ and the four-step process.
12. A streaming service claims the mean load time is 23.5 seconds. A random sample of 50 loads (from millions) has mean 24.8 and standard deviation 5.9 seconds. (a) Test whether the true mean is greater than 23.5 at $\alpha = 0.05$. (b) Describe a Type I error in this setting.
13. A test of $H_0 : \mu = 50$ vs. $H_a : \mu < 50$ for the mean weight of snack bags gives a p-value of 0.021. (a) Interpret the p-value in context. (b) State the conclusion at $\alpha = 0.05$ and at $\alpha = 0.01$. (c) Why does the conclusion not prove the mean is below 50?
14. A 95 percent confidence interval for the true mean battery life is (11.4, 13.2) hours. (a) What does it say about a two-sided test of $H_0 : \mu = 12$ at $\alpha = 0.05$? (b) What about $H_0 : \mu = 14$? (c) Can the interval be used for a one-sided test at $\alpha = 0.05$? Explain.
15. A manager tests $H_0 : \mu = 4$ vs. $H_a : \mu > 4$ using the waiting times (minutes) 3, 4, 3, 5, 4, 3, 4, 22, 3, 4, 5, 3 from a random sample of 12 customers. Explain whether a one-sample t-test is appropriate and what the manager should do.

8.4 Two-Sample t-Intervals

Estimate the difference between two population means using independent samples. Learn the unpooled standard error, degrees of freedom from the calculator, conditions for each group, and interpretation.

Why This Matters

Many questions compare two groups: do students recall more words in silence than with background music, does fertilizer A grow taller plants than fertilizer B, do bike commuters arrive faster than walkers? When the two groups are separate and unrelated, the right tool is the **two-sample t-interval for a difference in means**. It estimates $\mu_1 - \mu_2$ and tells us, with a stated confidence, how large the difference could be.

Students lose points by using one-sample formulas, by writing the interval without saying which group is first, by forgetting that each sample needs its own Normal/Large Sample check, and by concluding 'the means are equal' when the interval contains 0. This lesson belongs to Exam Unit 4 and shows up in free-response questions that ask you to construct, interpret and use an interval to compare groups.

Core Ideas

Take independent random samples (or two randomly assigned treatment groups) with sample sizes n_1 and n_2. The statistic $\bar{x}_1 - \bar{x}_2$ estimates $\mu_1 - \mu_2$. Its variability combines the variability of both sample means, so its standard error is

Two-sample t-interval (unpooled)

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The standard error $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ comes from adding variances of independent quantities, then taking the square root.
t^* is the critical value for the confidence level, with degrees of freedom found by the built-in calculator (a non-integer value is fine).
Do not pool the standard deviations and do not assume $\sigma_1 = \sigma_2$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Degrees of freedom. The exact degrees of freedom for this standard error come from a complicated formula; the built-in calculator evaluates it automatically when you use its two-sample t-interval feature. The result always lies between the smaller of $n_1 - 1$ and $n_2 - 1$ and the total $n_1 + n_2 - 2$. If you must work by hand, using $\min(n_1 - 1, n_2 - 1)$ is a safe, conservative choice that gives a slightly larger t^* and a slightly wider interval.

Conditions for the two-sample t-interval

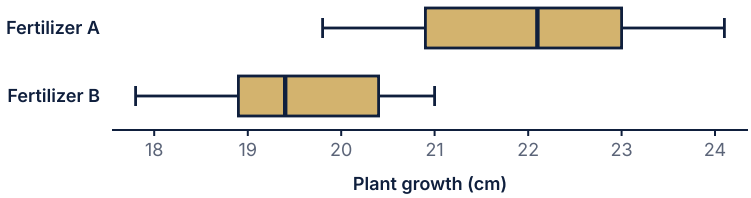
Random: both samples are independent random samples, or the subjects were randomly assigned to two treatments in an experiment.

10 percent: for sampling without replacement, each sample is less than 10 percent of its population. (Not needed for randomized experiments.)

Normal/Large Sample, for each group: each $n \geq 30$, or the graph of that group's data shows no strong skewness or outliers. Check the two groups separately.

Independent groups: knowing one group's values tells nothing about the other's. If the data are naturally matched, use paired data instead (Lesson 8.6).

A comparison of the two shapes is easiest with side-by-side boxplots drawn on the same axis. Look at the medians, the spreads and any outliers in each group before computing anything.



Side-by-side boxplots: both groups are roughly symmetric with no outliers (n = 10 each).

Interpreting the interval. Always say which group is first: 'We are 90 percent confident that the true mean growth with fertilizer A is between ___ and ___ cm greater than the true mean growth with fertilizer B.' If the interval is entirely positive, group 1 has the larger mean; if entirely negative, group 2 does. If the interval **contains 0**, a difference of 0 is plausible, so the data do not provide convincing evidence that the means differ. It does not prove they are equal.

Reading an interval for $\mu_1 - \mu_2$

All values positive: $\mu_1 > \mu_2$ is plausible; group 1 has the greater mean.

All values negative: $\mu_1 < \mu_2$ is plausible; group 2 has the greater mean.

Interval includes 0: no convincing evidence of a difference.

Switching the order of subtraction flips the signs of the endpoints but not the width or the meaning.

Worked Examples

EXAMPLE 1: FROM SUMMARY STATISTICS

Two random samples of students from a large school take the same practice exam. Section 1: $n_1 = 30$, mean 78.4, $s_1 = 9.2$. Section 2: $n_2 = 32$, mean 72.1, $s_2 = 10.5$ (the sections are separate, and each is less than 10 percent of its population of students). Give a 95 percent confidence interval for $\mu_1 - \mu_2$, assuming no strong skew or outliers.

SOLUTION

State. $\mu_1 - \mu_2$ is the difference in the true mean exam scores for students in section 1 and section 2. 95 percent confidence.

Plan. Two-sample t-interval for a difference in means. Random: independent random samples. 10 percent: stated. Normal/Large Sample: $n_1 = 30 \geq 30$ and $n_2 = 32 \geq 30$.

Do. $SE = \sqrt{\frac{9.2^2}{30} + \frac{10.5^2}{32}} = 2.503$. The calculator gives $df \approx 59.7$ and $t^* = 2.000$. $ME = 5.01$, so $6.3 \pm 5.01 = (1.29,\ 11.31)$.

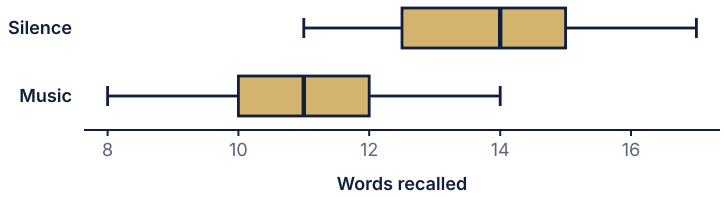
Conclude. We are 95 percent confident that the true mean score in section 1 is between 1.29 and 11.31 points higher than the true mean score in section 2. Because the interval does not contain 0, there is convincing evidence of a difference.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 2: RANDOMIZED EXPERIMENT WITH RAW DATA

Twenty-four volunteers were randomly assigned to memorize a word list either in silence (12 people) or with background music (12 people). The numbers of words recalled are: silence 14, 12, 16, 13, 15, 11, 14, 17, 13, 15, 12, 14; music 11, 10, 13, 9, 12, 10, 14, 11, 8, 12, 10, 11. Dotplots of both groups are roughly symmetric with no outliers. Construct a 95 percent confidence interval for the difference (silence minus music).



SOLUTION

State. $\mu_S - \mu_M$ = true mean words recalled in silence minus with music, for volunteers like these.

Plan. Two-sample t-interval. Random: random assignment to treatments. 10 percent: not required for an experiment. Normal: both $n = 12 < 30$, but both graphs show no strong skew or outliers.

Do. Silence: $\bar{x}_1 = 13.83$, $s_1 = 1.749$. Music: $\bar{x}_2 = 10.92$, $s_2 = 1.676$. $SE = 0.699$, $df \approx 22.0$, $t^* = 2.074$. Interval: $2.92 \pm 1.45 = (1.47,\ 4.37)$.

Conclude. We are 95 percent confident that the true mean number of words recalled in silence is between 1.47 and 4.37 words greater than with background music. The interval is entirely positive, so silence appears to give higher recall.

EXAMPLE 3: INTERVAL FOR TWO PLANT FERTILIZERS

Ten seedlings were randomly assigned to fertilizer A and ten to fertilizer B (growth in cm, shown in the boxplots above: A has mean 22.00 and $s = 1.374$; B has mean 19.55 and $s = 1.012$). Construct a 90 percent confidence interval for $\mu_A - \mu_B$ and interpret it.

SOLUTION

Conditions: random assignment; both boxplots are roughly symmetric with no outliers, so Normal is plausible for $n = 10$.

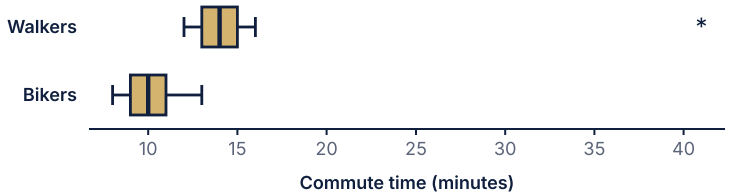
$SE = 0.540$, $df \approx 16.5$, $t^* = 1.742$, $ME = 0.94$. Interval: $2.45 \pm 0.94 = (1.51,\ 3.39)$ cm.

We are 90 percent confident that the true mean growth with fertilizer A is between 1.51 and 3.39 cm greater than with fertilizer B.

If the interval had been computed as B minus A, it would read (negative endpoints) and the sentence would say B is lower; the information is identical.

EXAMPLE 4: ONE SAMPLE FAILS THE CONDITIONS

A random sample of 12 walkers and a random sample of 14 bike commuters (from a large city) report their commute times in minutes: walkers 12, 14, 13, 15, 16, 12, 14, 13, 15, 14, 13, 41; bikers 9, 11, 10, 12, 8, 10, 11, 9, 12, 10, 11, 13, 9, 10. Can we build a two-sample t-interval?



The walkers' value 41 is an outlier.

SOLUTION

Random: both samples are random. 10 percent: both samples are far less than 10 percent of a large city. Independent: walkers and bikers are different people.

Normal/Large Sample is checked **for each sample**. The bikers' boxplot is roughly symmetric with no outliers, so that group is fine.

The walkers have $n = 12 < 30$ and an extreme outlier at 41 minutes. The Normal condition fails for this group, so the two-sample t-interval is **not** appropriate until the outlier is investigated or more data are collected.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: WHERE DO THE DEGREES OF FREEDOM COME FROM?

Two random samples: $n_1 = 15$, $\bar{x}_1 = 5.2$, $s_1 = 1.1$ and $n_2 = 40$, $\bar{x}_2 = 4.7$, $s_2 = 0.9$. Compare the 95 percent intervals for $\mu_1 - \mu_2$ using the calculator's degrees of freedom and the conservative choice $\min(n_1 - 1, n_2 - 1)$.

SOLUTION

$SE = \sqrt{\frac{1.1^2}{15} + \frac{0.9^2}{40}} = 0.318$ and $\bar{x}_1 - \bar{x}_2 = 0.5$.

Calculator: $df \approx 21.4$, $t^* = 2.077$, interval $0.5 \pm 0.660 = (-0.160, 1.160)$.

Conservative: $df = 14$, $t^* = 2.145$, interval $0.5 \pm 0.681 = (-0.181, 1.181)$.

The conservative interval is a little wider. Both lead to the same conclusion here, but on the exam the calculator's df is expected when you have the calculator available.

EXAMPLE 6: USING THE INTERVAL TO COMPARE

A music class reports a 95 percent confidence interval of (-1.4, 3.8) beats per minute for the difference in mean practice tempo between students taught with a metronome and without. What can be concluded?

SOLUTION

The interval contains 0, so a difference of 0 is plausible.

There is not convincing evidence of a difference in the true means. We cannot conclude that metronome use has no effect; the interval is also compatible with a difference as large as 3.8 beats per minute in favor of the metronome group (or 1.4 in favor of the other group).

The width (5.2) shows that the study is not very precise; a larger sample would shrink it.

Write It Like a 5

Prompt. A tech company randomly assigns 30 employees to use a new scheduling app and another 30 to continue the old method. After a month, each employee records the number of hours of meeting time per week. The new-app group has mean 9.4 and standard deviation 2.1; the old-method group has mean 11.2 and standard deviation 2.6. Graphs of both groups show no strong skewness or outliers. Construct and interpret a 90 percent confidence interval for the difference (new minus old) in true mean weekly meeting hours. [4 points]

Model Response

State. Let $\mu_N - \mu_O$ be the difference in the true mean weekly meeting hours (new app minus old method) for employees like these. **Plan.** Two-sample t-interval for $\mu_N - \mu_O$. Random: employees were randomly assigned to the two methods. Normal/Large Sample: both groups have $n = 30 \geq 30$ and the graphs show no strong skewness or outliers. The groups are independent. **Do.** $SE = \sqrt{\frac{2.1^2}{30} + \frac{2.6^2}{30}} = 0.610$, $df \approx 55.5$, $t^* = 1.673$. $(9.4 - 11.2) \pm 1.673(0.610) = -1.8 \pm 1.02$, or (-2.82, -0.78). **Conclude.** We are 90 percent confident that the true mean weekly meeting time with the new app is between 0.78 and 2.82 hours less than with the old method. Since the interval does not contain 0, there is convincing evidence the new app lowers mean meeting time.

What the grader looks for

1 point: defines the parameter as a difference and names the order (new minus old) and the procedure. 1 point: checks Random (random assignment), Normal/Large Sample for both groups, and independence. 1 point: correct standard error, df, t^* and interval. 1 point: interpretation says which group is higher or lower, in context, with the confidence level; uses 0 correctly. Practice rubric, not official scoring.

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Choose a direction for the subtraction and say it in the parameter definition ('group 1 minus group 2').
2. Verify the conditions for each sample, not just overall.
3. Use the built-in calculator's two-sample t-interval (unpooled) and record $\bar{x}_1$, s_1, n_1, $\bar{x}_2$, s_2, n_2, df, t^*.
4. Check the answer against SE and t^*: $ME = t^* \sqrt{s_1^2/n_1 + s_2^2/n_2}$.
5. Write the sentence that names both groups, the direction, and the units.

- The interval is centered at $\bar{x}_1 - \bar{x}_2$; make sure it lines up with the direction of subtraction.
- If both n are at least 30 you are rarely stopped by the Normal condition, but you still must say so.
- In an experiment, random assignment replaces random sampling, and the 10 percent condition is not needed.
- A bigger sample in either group narrows the interval, but the smaller sample affects the width more.

Common traps

1) Adding the margins of error or the standard deviations instead of adding variances. 2) Using a pooled standard deviation. 3) Saying 'the means are equal' when the interval contains 0. 4) Checking Normality only for the larger sample or for the combined data.

Mistake Clinic

Spot the error

A student computes the standard error as $\sqrt{\frac{(9.2+10.5)^2}{62}}$ for the sections example. **Where it fails:** variances add, not standard deviations. **Correct solution:** $SE = \sqrt{9.2^2/30 + 10.5^2/32} = 2.503$.

Spot the error

A student with the interval (-1.4, 3.8) writes 'The two teaching methods have the same effect.' **Where it fails:** containing 0 only means no convincing evidence of a difference, and the interval allows differences up to 3.8. **Correct solution:** 'There is not convincing evidence that the true mean tempos differ.'

Spot the error

A student checks Normality for walkers and bikers by combining all 26 times and says 'n = 26 is close enough to 30.' **Where it fails:** each group must be checked separately, and the walkers' sample has an outlier with $n = 12$. **Correct solution:** state that the walkers group violates the Normal condition, so the interval is not justified.

Real-World Connection

Two-sample intervals power comparisons in medical trials, product tests and A/B tests of web pages. A company testing a new checkout page does not only want to know whether the mean time is different; it wants to know how much, and the interval gives a plausible range for that difference.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

In the meeting-time prompt, how many employees per group (equal sizes) would be needed for a 90 percent margin of error of about 0.5 hours, if the standard deviations stay 2.1 and 2.6? **Solution:** $ME \approx 1.645\sqrt{(2.1^2 + 2.6^2)/n} \le 0.5$ gives $n \ge (1.645/0.5)^2(11.17) = 120.9$, so about 121 per group. Recheck with the t value: for $n = 121$, $t^* \approx 1.651$ and $ME = 0.502$, which meets the goal.

Practice

PRACTICE **Two-Sample t-Intervals**

Show your work. A calculator is allowed. Justify conclusions in context.

1. Two independent random samples have $n_1 = 40$, $s_1 = 8$ and $n_2 = 50$, $s_2 = 6$. What is the standard error of $\bar{x}_1 - \bar{x}_2$?
   (A) 1.523 **(B)** 1.054 **(C)** 2.113 **(D)** 2.320

2. Which expression gives the margin of error for a two-sample t-interval for $\mu_1 - \mu_2$?
   (A) $t^* \left(\frac{s_1}{\sqrt{n_1}} + \frac{s_2}{\sqrt{n_2}} \right)$ **(B)** $t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ **(C)** $t^* \sqrt{\frac{s_1^2 - s_2^2}{n_1 + n_2}}$
   (D) $z^* \sqrt{\frac{s_1^2}{n_1}} \cdot \sqrt{\frac{s_2^2}{n_2}}$

3. A 95 percent confidence interval for $\mu_A - \mu_B$, the difference in mean test scores for two teaching methods, is (2.1, 6.3). Which statement is correct?
   (A) 95 percent of students taught with method A scored 2.1 to 6.3 points higher than students with method B.
   (B) There is a 95 percent chance that the sample mean for A exceeds that for B by 2.1 to 6.3.
   (C) We are 95 percent confident that the true mean for method A is between 2.1 and 6.3 points higher than for method B.
   (D) The true means differ by exactly 4.2 points.

4. A 95 percent confidence interval for $\mu_1 - \mu_2$ is (-0.8, 3.6). Which conclusion is best?
   (A) The true means are equal.
   (B) Group 1 has a larger mean than group 2.
   (C) Group 2 has a larger mean than group 1.
   (D) There is not convincing evidence that the true means differ, since 0 is in the interval.

5. Twelve participants were randomly assigned to each of two treatments. Which condition must be checked to use a two-sample t-interval?
   (A) The data for each treatment group show no strong skewness or outliers, since each $n < 30$.
   (B) Each group's standard deviation is the same.
   (C) Each sample is less than 10 percent of a population.
   (D) The two groups have exactly the same sample size.

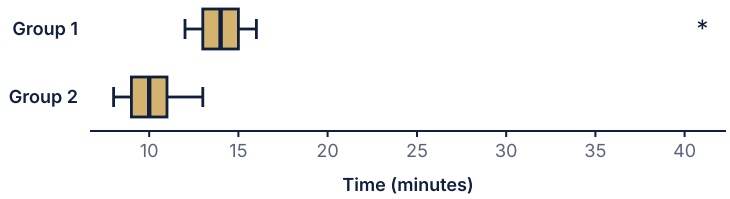
6. Independent random samples give $n_1 = 25$, $\bar{x}_1 = 20.1$, $s_1 = 3.2$ and $n_2 = 28$, $\bar{x}_2 = 18.4$, $s_2 = 4.0$. Which is the 99 percent confidence interval for $\mu_1 - \mu_2$ with the calculator's degrees of freedom?
   (A) (-1.07, 4.47) **(B)** (-0.85, 4.25) **(C)** (-4.35, 0.95)
   (D) (-0.95, 4.35)

7. When the built-in calculator computes a two-sample t-interval without pooling, the degrees of freedom are
   (A) always $n_1 + n_2 - 2$
   (B) usually not a whole number and lie between the smaller of $n_1 - 1$, $n_2 - 1$ and $n_1 + n_2 - 2$
   (C) always the smaller of n_1 and n_2
   (D) always $n_1 - 1$

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

8. Independent random samples of sizes 12 and 14 are summarized by the boxplots. Why is a two-sample t-interval not justified?



(A) The groups have different sample sizes.
(B) The medians are different.
(C) Group 1 has an extreme outlier and a small sample size.
(D) The boxes do not overlap.

9. A two-sample t-interval is computed with sample sizes 20 and 20. If both sample sizes are quadrupled to 80 and 80, with the same standard deviations and the same confidence level, the margin of error will be about

(A) one-fourth as large **(B)** half as large **(C)** the same
(D) twice as large

10. A random sample of 45 students who use flash cards has a mean quiz score of 64.2 with $s = 8.6$. An independent random sample of 48 students who do not use flash cards has mean 59.9 and $s = 9.4$. Both samples are less than 10 percent of their populations. Construct and interpret a 90 percent confidence interval for $\mu_1 - \mu_2$ (flash cards minus no flash cards).

11. Eight randomly selected plants received a nutrient mix (heights 52, 55, 49, 58, 54, 51, 57, 53 cm) and nine others, randomly selected, did not (heights 47, 50, 45, 52, 48, 44, 49, 46, 51 cm). Both dotplots are roughly symmetric with no outliers. Construct a 95 percent confidence interval for the difference (mix minus no mix) in true mean height and interpret it.

12. A researcher compares the mean reaction times (ms) of two groups. Group 1: $n = 10$, histogram strongly right skewed with an outlier. Group 2: $n = 40$, roughly symmetric. (a) Check the Normal/Large Sample condition for each group. (b) Is a two-sample t-interval appropriate? What could the researcher do?

13. A 95 percent confidence interval for $\mu_1 - \mu_2$ (mean hours of sleep, athletes minus non-athletes) is (-0.15, 0.62) hours. (a) Interpret the interval in context. (b) Does it give convincing evidence of a difference? (c) A classmate says the interval proves athletes and non-athletes sleep the same amount. Respond.

14. Using the memory experiment data (silence: 14, 12, 16, 13, 15, 11, 14, 17, 13, 15, 12, 14; music: 11, 10, 13, 9, 12, 10, 14, 11, 8, 12, 10, 11), compute a 95 percent confidence interval for the difference (music minus silence) and explain how it relates to the interval (silence minus music).

15. A 90 percent confidence interval for $\mu_1 - \mu_2$ is (1.2, 5.8). (a) Find the point estimate and the margin of error. (b) If the confidence level is changed to 99 percent with the same data, will the interval be wider or narrower? (c) Does the interval give convincing evidence that $\mu_1 \neq \mu_2$?

8.5 Two-Sample t-Tests

Test whether two population means differ using independent samples. Set up hypotheses with a consistent order, check conditions for each group, use the calculator, and conclude in context.

Why This Matters

Does a new teaching method raise scores, do energy drinks speed up reaction time, is route A really slower than route B? When two separate groups are compared, a **two-sample t-test for a difference in means** weighs the observed difference between the sample means against what random variation alone would produce. It is the hypothesis-test partner of the interval in Lesson 8.4.

The common mistakes are choosing the wrong tail because the order of subtraction was not tracked, mixing up the two sample standard deviations, using a paired method for independent groups (or the reverse), and ending the conclusion with a statistical verdict but no context. The topic is in Exam Unit 4 and is a frequent free-response task: state hypotheses, check conditions for both samples, find the p-value, and explain in context.

Core Ideas

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Hypotheses for two means

$H_0 : \mu_1 - \mu_2 = 0$ (equivalently $\mu_1 = \mu_2$).

H_a is one of $\mu_1 - \mu_2 > 0$, $\mu_1 - \mu_2 < 0$, or $\mu_1 - \mu_2 \neq 0$.

Define μ_1 and μ_2 in context, and let the order of subtraction match the order in the question.

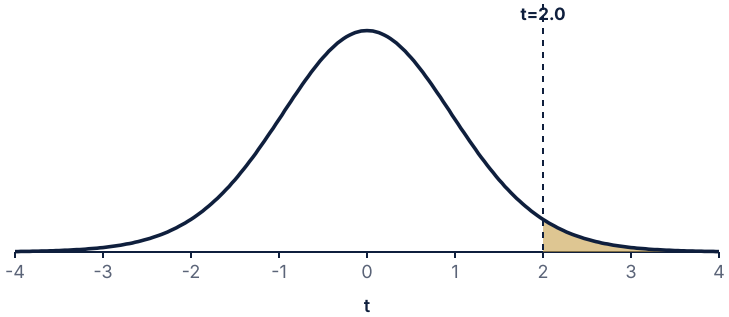
Two-sample t-test statistic

$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ with df from the built-in calculator.

The p-value is the area in the tail(s) of the t-distribution beyond the observed t, in the direction of H_a; double for $\neq$.

The calculator's two-sample t-test (unpooled) returns t, the p-value, and df directly.

The numerator measures how far apart the sample means are. The denominator is the standard error of the difference, which says how far apart they would typically be just from chance. A ratio far from 0 (large $|t|$) means the difference is big compared to chance variation.



Right-tailed test: the p-value is the area above the observed t (df is about 20 in this picture).

Conditions (check for each sample)

Random: independent random samples, or random assignment of subjects to two treatments.

10 percent: each sample is less than 10 percent of its population (sampling without replacement; skip for experiments).

Normal/Large Sample: each group has $n \geq 30$ or a graph with no strong skewness or outliers.

Independent groups: the two samples are not matched or related.

Four steps. State the hypotheses and parameters with α. Plan the procedure and check conditions for both groups. Do the computation: the means, standard deviations, t, df and p-value. Conclude by comparing the p-value to α and answering in context with 'convincing evidence' language. A **statistically significant** result means the difference is unlikely to be due to chance alone; it does not say the difference is large. Always ask how large the difference is and whether it matters.

Tests and intervals together. A two-sided test at significance level α agrees with a $(1 - \alpha)$ confidence interval for $\mu_1 - \mu_2$: reject H_0 when 0 is outside the interval. An interval adds information that the test lacks, namely the plausible size of the difference, so reporting both is good practice.

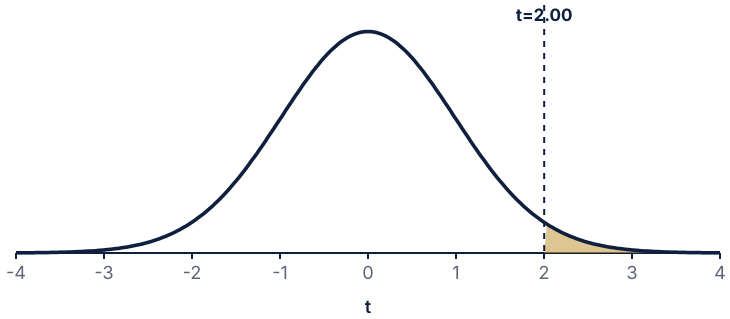
Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: ONE-SIDED TEST FROM SUMMARY STATISTICS

A delivery company compares two routes using independent random samples of trips from a very large number of trips. Route A: $n_1 = 35$, mean 28.4 minutes, $s_1 = 5.1$. Route B: $n_2 = 40$, mean 26.1 minutes, $s_2 = 4.8$. Dispatchers suspect route A takes longer on average. Test at $\alpha = 0.05$.



Upper-tail p-value for Example 1.

SOLUTION

State. $H_0 : \mu_A - \mu_B = 0$ vs. $H_a : \mu_A - \mu_B > 0$, where μ_A and μ_B are the true mean trip times (minutes) on routes A and B.

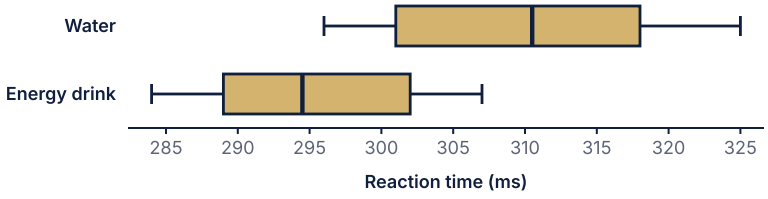
Plan. Two-sample t-test. Random: independent random samples. 10 percent: each is far less than 10 percent of a very large number of trips. Normal/Large Sample: $35 \geq 30$ and $40 \geq 30$.

Do. $SE = \sqrt{\frac{5.1^2}{35} + \frac{4.8^2}{40}} = 1.149$, $t = \frac{28.4 - 26.1}{1.149} = 2.00$, $df \approx 70.3$, p-value $= 0.0245$.

Conclude. Since 0.0245 is less than 0.05, we reject H_0. There is convincing evidence that the true mean trip time on route A is greater than on route B.

EXAMPLE 2: RANDOMIZED EXPERIMENT WITH RAW DATA

Twenty volunteers were randomly assigned to drink water (10) or an energy drink (10) and then took a reaction-time test. Reaction times (ms): water 312, 298, 325, 305, 318, 309, 301, 322, 296, 314; energy drink 291, 305, 288, 299, 302, 284, 296, 293, 307, 289. Boxplots of both groups are roughly symmetric with no outliers. Is there convincing evidence that the energy drink lowers the true mean reaction time? Use $\alpha = 0.05$.



SOLUTION

State. $H_0 : \mu_W - \mu_E = 0$ vs. $H_a : \mu_W - \mu_E > 0$, where μ_W and μ_E are the true mean reaction times (ms) after water and the energy drink for volunteers like these.

Plan. Two-sample t-test. Random: random assignment. Normal: $n = 10 < 30$ in each group, but both boxplots show no strong skew or outliers.

Do. Water: $\bar{x} = 310.0$, $s = 10.00$. Drink: $\bar{x} = 295.4$, $s = 7.71$. $t = 3.66$, $df \approx 16.9$, p-value $= 0.0010$.

Conclude. Since 0.0010 is less than 0.05, we reject H_0. There is convincing evidence that the energy drink lowers the true mean reaction time compared with water for volunteers like these.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: TWO-SIDED TEST AND AN INTERVAL

A researcher compares mean daily steps for two random samples from a large community: group 1 has $n = 22$, mean 14.2, $s = 3.1$ (thousands of steps) and group 2 has $n = 25$, mean 12.9, $s = 3.6$. Both graphs are roughly symmetric with no outliers. Test whether the true means differ at $\alpha = 0.05$ and check with a 95% interval.

SOLUTION

State. $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 \neq 0$.

Plan. Two-sample t-test; random samples; each sample is under 10 percent of a large community; both graphs show no strong skew or outliers.

Do. $t = \dfrac{14.2 - 12.9}{0.977} = 1.33$, $df \approx 45.0$, two-sided p-value $= 0.1902$.

Conclude. Since 0.1902 is greater than 0.05, we fail to reject H_0. The data do not provide convincing evidence that the true mean steps differ for the two groups.

Interval check. The 95% interval for $\mu_1 - \mu_2$ is (-0.67, 3.27), which includes 0 and agrees with the test.

EXAMPLE 4: READING SOFTWARE OUTPUT

Software compares the true mean scores of two groups (group 1: $n = 18$, group 2: $n = 20$, both random samples with no outliers) and reports: Difference = 4.5, SE of difference = 2.53, T = 1.78, DF = 36.0, P-value (two-sided) = 0.0838. A researcher had predicted that group 1 would score higher ($H_a : \mu_1 - \mu_2 > 0$). What is the conclusion at $\alpha = 0.05$?

Difference	SE	T	DF	P (two-sided)
4.5	2.53	1.78	36.0	0.0838

SOLUTION

The output is for a two-sided alternative. Because the observed difference is positive, matching the direction of H_a, the one-sided p-value is half: 0.0838/2 $= 0.0419$.

Since 0.0419 is less than 0.05, we reject H_0. There is convincing evidence that the true mean for group 1 is greater than for group 2.

Warning: halving works only when the sign of the difference matches H_a. If the difference had been negative, the one-sided p-value would be $1-$ (half the printed value), which is large.

EXAMPLE 5: CHOOSING THE TAIL AND THE ORDER

A coach tests whether a new drill reduces mean sprint time. Group N (new drill) and group O (old drill) are independent random samples. Write the hypotheses if the parameter is (a) $\mu_N - \mu_O$ and (b) $\mu_O - \mu_N$.

SOLUTION

(a) The new drill should reduce time, so $\mu_N < \mu_O$: $H_0 : \mu_N - \mu_O = 0$ vs. $H_a : \mu_N - \mu_O < 0$.

(b) Reversing the order reverses the inequality: $H_0 : \mu_O - \mu_N = 0$ vs. $H_a : \mu_O - \mu_N > 0$.

The two versions give the same p-value and conclusion. What matters is that H_a, the sign of t, and the tail area all use the same order.

EXAMPLE 6: SIGNIFICANT BUT SMALL

Two random samples of 1,500 students each (from very large populations) take a 100-point exam in two formats. Format 1 has mean 71.3 and format 2 has mean 70.5, and both have $s = 10$. Is the difference significant at $\alpha = 0.05$? Does it matter in practice?

SOLUTION

$SE = \sqrt{\dfrac{10^2}{1500} + \dfrac{10^2}{1500}} = 0.365$, $t = \dfrac{0.8}{0.365} = 2.19$, $df \approx 3000$, two-sided p-value $= 0.0285$.

Since 0.0285 is less than 0.05, we reject H_0: there is convincing evidence that the true mean scores differ.

A difference of 0.8 points on a 100-point exam is small compared with the standard deviation of 10 (about 0.08 standard deviations). With very large samples, tiny differences can be statistically significant, so always report the size of the difference too.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Prompt. A school randomly assigns 30 students to a tutoring program and 34 to a control group. After the term, the tutored students have a mean exam score of 82.5 with a standard deviation of 9.1, and the control students have a mean of 77.9 with a standard deviation of 10.2. Both distributions are roughly symmetric with no outliers. Is there convincing evidence at the 5 percent significance level that the tutoring program increases the true mean exam score? [4 points]

Model Response

State. Let μ_T and μ_C be the true mean exam scores for students like these with and without tutoring. $H_0 : \mu_T - \mu_C = 0$ vs. $H_a : \mu_T - \mu_C > 0$, $\alpha = 0.05$. **Plan.** Two-sample t-test for a difference in means. Random: students were randomly assigned to the two groups. Normal/Large Sample: $n_T = 30 \geq 30$ and $n_C = 34 \geq 30$, and the graphs show no strong skewness or outliers. The two groups are independent. **Do.** $t = \dfrac{82.5 - 77.9}{\sqrt{9.1^2/30 + 10.2^2/34}} = 1.91$, $df \approx 62.0$, p-value $= 0.0306$.

Conclude. Since 0.0306 is less than 0.05, we reject H_0. There is convincing evidence that the tutoring program increases the true mean exam score.

What the grader looks for

1 point: correct hypotheses with parameters defined and the order of subtraction clear. 1 point: names the test and checks Random and Normal/Large Sample for both groups. 1 point: correct test statistic, degrees of freedom and p-value. 1 point: compares the p-value to α and concludes in context with the direction of H_a. Practice rubric, not official scoring.

Coach's Playbook

1. Write the two parameters and the order of subtraction, then H_a in that order.
2. Check the conditions for each group separately and for the design (random sampling or random assignment).
3. Use the built-in two-sample t-test, unpooled, with summary statistics or lists. Record t, df and the p-value.
4. Decide: p-value versus α.
5. Conclude in context. Add a sentence on size or practical meaning when the question asks.

- Sanity check the sign: if $\bar{x}_1 < \bar{x}_2$ then t is negative; a right-tailed test with a negative t cannot be significant.
- Group 1 is whichever group the question names first. Do not reorder to make the numbers look nicer.
- A one-sided p-value can be found from the printed two-sided value only when the observed difference points in the direction of H_a.
- 'Convincing evidence that the means differ' is a conclusion. 'The means are different' stated as fact is not.

Common traps

1) Reversing the order in H_a but not in the computation. 2) Pooling the standard deviations or using $df = n_1 + n_2 - 2$ with the unpooled method. 3) Checking conditions for only one group. 4) Treating independent random assignment as a pair-by-pair match.

Mistake Clinic

Spot the error

A student writes $H_0 : \bar{x}_1 = \bar{x}_2$ and $H_a : \bar{x}_1 > \bar{x}_2$ for the route example. **Where it fails:** hypotheses are about the population means μ_1 and μ_2. **Correct solution:** $H_0 : \mu_A - \mu_B = 0$ vs. $H_a : \mu_A - \mu_B > 0$ with both parameters defined.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

For the energy-drink example a student uses the group sizes only: 'Both samples have $n = 10 < 30$, so the Normal condition fails.' **Where it fails:** a small sample does not fail the condition if the graph shows no strong skew or outliers. **Correct solution:** state that each boxplot is roughly symmetric with no outliers, so the condition is met.

Spot the error

A student reads the printed two-sided p-value 0.0838 in a test with $H_a : \mu_1 > \mu_2$ and concludes 'p = 0.0838 is greater than 0.05, so fail to reject.' The observed difference was positive. **Where it fails:** the alternative is one-sided, so the p-value is half the printed value: 0.0419. **Correct solution:** since 0.0419 is less than 0.05, reject H_0.

Real-World Connection

A/B tests in technology, clinical trials in health, and agricultural field trials all use the two-sample t-test (or close cousins of it) to decide whether a new version beats an old one. A fair test uses random assignment, so that other differences between groups average out and a significant difference can reasonably be attributed to the treatment.

Challenge Corner

In Example 2 the energy-drink test had $t = 3.66$. Suppose a researcher repeated the experiment with the same means and standard deviations but only 5 volunteers per group. **Question:** what is the new t statistic and p-value? **Solution:** the standard error becomes $\sqrt{10^2/5 + 7.71^2/5} = 5.646$ (it grows by $\sqrt{2}$), so $t = 14.6/5.646 = 2.59$, $df \approx 7.5$, and the one-sided p-value is 0.0170. Smaller samples make the same observed difference less convincing.

Practice

PRACTICE **Two-Sample t-Tests**

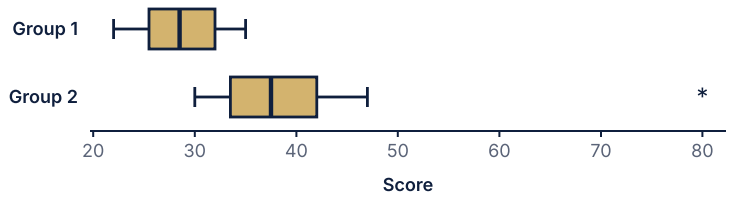
Show your work. A calculator is allowed. Justify conclusions in context.

1. A researcher suspects that students who use a study app score lower, on average, than those who do not. With μ_1 for app users and μ_2 for non-users, which hypotheses are appropriate?
   (A) $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 < 0$
   (B) $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 > 0$
   (C) $H_0 : \mu_1 - \mu_2 < 0$, $H_a : \mu_1 - \mu_2 = 0$
   (D) $H_0 : \bar{x}_1 - \bar{x}_2 = 0$, $H_a : \bar{x}_1 - \bar{x}_2 < 0$
2. Independent random samples: $n_1 = 30$, $\bar{x}_1 = 55.2$, $s_1 = 6$; $n_2 = 36$, $\bar{x}_2 = 52.0$, $s_2 = 7$. What is the two-sample t statistic?
   (A) 2.82 (B) 2.00 (C) 1.41 (D) 2.83
3. A two-sample t-test with $H_a : \mu_1 < \mu_2$ yields $t = -2.20$ with $df \approx 28$. What is the p-value?
   (A) 0.0362 (B) 0.9819 (C) 0.0181 (D) 0.0070
4. A test gives p-value 0.04 at $\alpha = 0.05$ for $H_a : \mu_1 > \mu_2$ comparing mean sleep for two groups of teenagers. Which conclusion is best?
   (A) Group 1 sleeps more than group 2 in every case.
   (B) There is a 4 percent chance that the means are equal.
   (C) We fail to reject H_0 because the p-value is small.
   (D) There is convincing evidence that the true mean sleep for group 1 is greater than for group 2.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. Two random samples of 12 each are compared with the boxplots shown. Which is correct?



(A) A two-sample t-test is not justified because group 2 has an outlier and $n < 30$.
(B) A two-sample t-test is justified since the sample sizes are equal.
(C) A two-sample t-test is justified since the medians differ.
(D) A paired t-test should be used because there are two groups.

6. A 95 percent confidence interval for $\mu_1 - \mu_2$ is (-0.7, 3.3). For $H_a : \mu_1 - \mu_2 \neq 0$ at $\alpha = 0.05$, the conclusion is
(A) fail to reject H_0; there is not convincing evidence the means differ.
(B) reject H_0; there is convincing evidence the means differ.
(C) reject H_0 since the interval is mostly positive.
(D) accept H_0; the means are equal.

7. In a two-sample t-test, a Type I error would be
(A) failing to find a difference that really exists.
(B) concluding the means differ when they are actually equal.
(C) using the wrong degrees of freedom.
(D) rejecting H_0 with a very small p-value.

8. Which situation calls for a two-sample t-test (not a paired test)?
(A) Comparing each student's score before and after a course.
(B) Comparing the two eyes of each patient on a vision test.
(C) Comparing the mean heights of two separate random samples of plants grown with different soils.
(D) Comparing a runner's time on the same track in two weather conditions on the same days.

9. Two random samples of 1,500 each show a statistically significant but tiny difference in means (p-value 0.03). Which statement is best?
(A) The difference must be large.
(B) The null hypothesis is proven false.
(C) The result is unreliable because the samples are large.
(D) The difference is unlikely to be due to chance, but it may be too small to matter in practice.

10. A random sample of 48 teenagers has a mean of 15.6 hours per week of screen time with $s = 3.4$. An independent random sample of 52 adults has mean 14.3 and $s = 3.9$ (both samples are far less than 10 percent of their populations). Test whether the true means differ at $\alpha = 0.05$.

11. A teacher randomly assigns 10 students to method X and 11 to method Y. Final scores: X: 41, 38, 45, 36, 43, 40, 47, 39, 42, 44; Y: 35, 32, 39, 30, 37, 33, 36, 31, 38, 34, 29. Dotplots are roughly symmetric with no outliers. Is there convincing evidence that method X produces a higher true mean score? Use $\alpha = 0.05$.

12. Software output for a test of $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 > 0$ (random samples, conditions met) shows: estimate of difference = 3.1, T = 1.84, DF = 33.7, P-value (one-sided, greater than) = 0.037. (a) State the conclusion at $\alpha = 0.05$. (b) State it at $\alpha = 0.01$. (c) Interpret the p-value.

13. In a test of two means, $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 \neq 0$ is rejected at $\alpha = 0.05$. (a) Describe the Type I error in the context of comparing the mean sleep of teens who use phones at night with those who do not. (b) Describe the Type II error. (c) Which error is possible in this case?

14. Two independent random samples give $n_1 = 40$, $\bar{x}_1 = 6.8$, $s_1 = 1.5$ and $n_2 = 45$, $\bar{x}_2 = 6.2$, $s_2 = 1.7$. (a) Perform a two-sided test of equal means at $\alpha = 0.05$. (b) Compute a 95 percent confidence interval for $\mu_1 - \mu_2$ and explain how it agrees with the test.

15. A student wants to compare mean prices of used bicycles in two cities. She takes a random sample of 9 bicycles from each city. City 1 prices are strongly right skewed with an outlier; city 2 prices are roughly symmetric. (a) Which conditions are met? (b) Is a two-sample t-test justified? (c) What could she do?

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

8.6 Paired Data and Matched Pairs

Recognize paired designs, form differences, and use one-sample t procedures on them. Compare paired and two-sample methods and interpret results in context.

Why This Matters

Sometimes the two sets of numbers you want to compare are not from separate groups at all. The same students are timed before and after a training program; two sneaker brands are tried by the same runners; twins are split between two treatments. These are **paired data**. Treating them as two independent samples throws away information and usually hides real effects, while treating independent samples as pairs makes no sense. Knowing which situation you are in is the key skill of this lesson.

The paired procedure itself is easy: compute the difference within each pair and run a one-sample t procedure on those differences. The AP exam in Exam Unit 4 tests this in two ways: identifying whether data are paired, and carrying out and interpreting a matched-pairs interval or test, with conditions stated for the differences. Choosing the wrong procedure is one of the most frequent lost points.

Core Ideas

Paired data (matched pairs)

Each observation in one group is naturally linked to exactly one observation in the other group.

Typical designs: **before and after** measurements on the same subject; **two treatments on the same subject** (in random order); **matched subjects** (twins, spouses, subjects matched on age or ability) with one assigned to each treatment.

Compute one difference per pair, $d = (\text{second}) - (\text{first})$ in a direction you state. The analysis then uses only the n differences.

Paired t-procedures

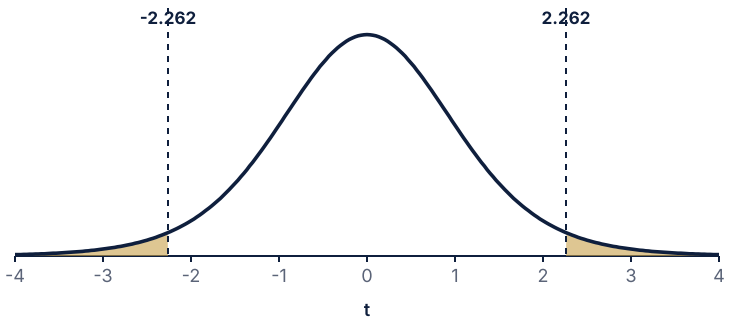
Let μ_d be the true mean difference. $\bar{x}_d$ and s_d are the mean and standard deviation of the sample differences; n is the number of **pairs**.

Interval: $\bar{x}_d \pm t^* \dfrac{s_d}{\sqrt{n}}$ with $df = n - 1$.

Test: $H_0 : \mu_d = 0$; $t = \dfrac{\bar{x}_d - 0}{s_d/\sqrt{n}}$ with $df = n - 1$.

Everything else (p-value, decision, conclusion) is exactly the one-sample t procedure.

Conditions are checked for the differences, not for the two original lists: (1) **Random**: the pairs were randomly sampled, or (for experiments) treatments were randomly assigned within each pair, for example by a coin flip deciding the order. (2) **10 percent**: when sampling, $n \le 0.10N$. (3) **Normal/Large Sample**: the number of pairs $n \ge 30$, or the graph of the *differences* shows no strong skewness or outliers. A dotplot or boxplot of the differences is the best evidence.



Two-sided 5 percent rejection region for a paired test with 10 pairs (df = 9).

Why pairing helps. People differ widely from one another: some type fast, some slowly. That subject-to-subject variability inflates the standard error of a two-sample comparison. When you take differences within each person, the person-to-person variation drops out and only the change remains. If the pairs are strongly related, s_d is much smaller than the spread of either original list, so the paired test is more powerful and the paired interval is narrower.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Question to ask	If yes	If no
Is each value in one list naturally linked to one value in the other (same subject, twin, matched pair)?	Paired t on differences	Two-sample t for independent groups
Are the two lists different sizes?	Cannot be paired	May be either; look at the design
Were subjects randomly assigned to groups with no matching?	Two-sample t	Check whether pairing is present

The design, not the data, decides. Two lists of the same length are not automatically paired, and a design in which the same person provides both numbers is automatically paired.

Worked Examples

EXAMPLE 1: PAIRED OR TWO-SAMPLE?

Decide whether each study calls for a paired or a two-sample procedure. (a) 25 randomly chosen students take a quiz before and after a review session. (b) 30 students are randomly assigned to one of two review methods and then take the same quiz. (c) 20 pairs of twins; one twin from each pair is randomly assigned to each of two diets. (d) 15 runners each run the same course with shoe A and with shoe B in random order.

SOLUTION

(a) **Paired**: the same student gives both scores; analyze each student's after minus before.

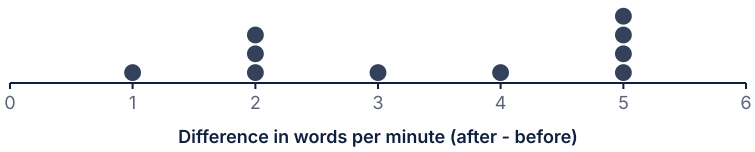
(b) **Two-sample**: two independent groups formed by random assignment, with no link between a student in one group and a student in the other.

(c) **Paired**: each twin pair is a matched pair, with one twin in each group.

(d) **Paired**: each runner is measured under both conditions, and randomizing the order of the shoes protects against order effects.

EXAMPLE 2: A PAIRED TEST FROM RAW DATA

A typing instructor randomly selects 10 students from a large course and records their words per minute before and after a practice program. Before: 42, 55, 38, 61, 47, 52, 44, 58, 40, 50. After: 46, 58, 40, 66, 49, 57, 45, 63, 45, 52. The dotplot of the differences is shown. Is there convincing evidence that the program increases the true mean typing speed? Use $\alpha = 0.05$.



SOLUTION

State. μ_d = true mean change (after minus before) in words per minute for students in the course. $H_0 : \mu_d = 0$ vs. $H_a : \mu_d > 0$.

Plan. Paired t-test (one-sample t-test on differences). Random: students were randomly selected. 10 percent: 10 students is under 10 percent of a large course. Normal: $n = 10 < 30$, but the dotplot of the differences shows no outliers and no strong skew.

Do. The differences are 4, 3, 2, 5, 2, 5, 1, 5, 5, 2. $\bar{x}_d = 3.4$, $s_d = 1.578$, $t = \dfrac{3.4}{1.578/\sqrt{10}} = 6.82$, $df = 9$, p-value $= lessthan 0.0001$.

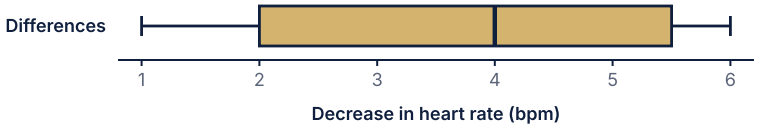
Conclude. Since less than 0.0001 is less than 0.05, we reject H_0. There is convincing evidence that the program increases the true mean typing speed of students in the course.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 3: A PAIRED INTERVAL

A fitness club randomly selects 12 members from hundreds and measures resting heart rate (beats per minute) before and after an eight-week program. Before: 72, 68, 80, 75, 70, 66, 78, 74, 69, 82, 71, 77. After: 68, 66, 75, 73, 69, 62, 72, 70, 67, 76, 70, 71. Construct a 95 percent confidence interval for the true mean drop (before minus after).



Boxplot of the 12 differences.

SOLUTION

State. μ_d = true mean decrease in resting heart rate (before minus after) for club members. 95 percent confidence.

Plan. Paired t-interval. Random: random selection. 10 percent: 12 is under 10 percent of hundreds. Normal: $n = 12 < 30$, but the boxplot of the differences is roughly symmetric with no outliers.

Do. Differences: 4, 2, 5, 2, 1, 4, 6, 4, 2, 6, 1, 6. $\bar{x}_d = 3.583$, $s_d = 1.929$, $df = 11$, $t^* = 2.201$. $ME = 1.23$; interval (2.36, 4.81).

Conclude. We are 95 percent confident that the true mean decrease in resting heart rate for club members is between 2.36 and 4.81 beats per minute.

EXAMPLE 4: TEST FROM SUMMARY STATISTICS OF THE DIFFERENCES

A random sample of 40 employees (from thousands) tries two keyboard layouts in random order and the difference in completion time (layout 1 minus layout 2, seconds) is computed for each. The differences have mean 1.8 and standard deviation 5.9. Test whether the true mean difference is nonzero at $\alpha = 0.05$.

SOLUTION

State. $H_0 : \mu_d = 0$ vs. $H_a : \mu_d \neq 0$, where μ_d is the true mean difference (layout 1 minus layout 2) in completion time.

Plan. Paired t-test. Random order of layouts and random sample; 40 is under 10 percent of thousands; $n = 40 \geq 30$.

Do. $t = \dfrac{1.8}{5.9/\sqrt{40}} = 1.93$, $df = 39$, two-sided p-value $= 0.0610$.

Conclude. Since 0.0610 is greater than 0.05, we fail to reject H_0. There is not convincing evidence that the true mean completion times for the two layouts differ.

EXAMPLE 5: PAIRED VERSUS TWO-SAMPLE ON THE SAME DATA

Run a two-sample t-test on the typing data of Example 2, as if the before and after lists were independent samples, and compare with the paired test.

SOLUTION

Two-sample (wrong design): after mean 52.1, $s = 8.60$; before mean 48.7, $s = 7.82$. $SE = 3.67$, $t = 0.93$, p-value $= 0.1836$.

Paired (correct design): $t = 6.82$ and p-value $= lessthan 0.0001$.

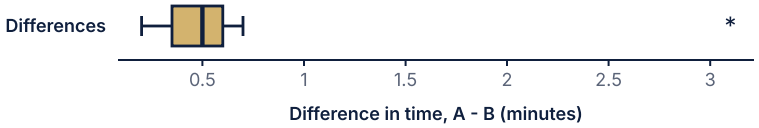
Typing speeds vary a lot from one student to the next (from the high 30s to the low 60s), and that variation swamps the 3.4 words per minute average improvement in the two-sample analysis. Taking differences within each student removes this variation, so the improvement stands out clearly. The two-sample method is not just less powerful here: it answers the wrong question for these data, because the two lists are not independent.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 6: CONDITIONS FAIL ON THE DIFFERENCES

Twelve randomly chosen runners (from a large club) run the same course in shoe A and shoe B in random order. The difference in times (A minus B, minutes) for each runner is 0.4, 0.6, 0.3, 0.5, 0.7, 0.2, 0.5, 0.4, 0.6, 0.3, 0.5, 3.1. Is a paired t-procedure appropriate?



SOLUTION

Random order of shoes and random selection of runners: satisfied. 12 runners are under 10 percent of a large club: satisfied.

Normal: the condition is checked on the **differences**. With $n = 12 < 30$, the boxplot shows a high outlier at 3.1 minutes, so the condition is not met.

Do not proceed with a t procedure. Investigate whether 3.1 is a timing error or a real unusual performance (for example an injury), then correct, remove with justification, or collect more pairs.

Write It Like a 5

Prompt. A coach wants to know whether a new warm-up routine changes vertical jump height. She randomly selects 15 players from a very large league and records each player's jump height (in cm) after a standard warm-up and after the new warm-up on two different days, in random order. The mean of the differences (new minus standard) is 1.9 cm with standard deviation 2.8 cm. A dotplot of the differences is roughly symmetric with no outliers. Is there convincing evidence that the new warm-up changes the true mean jump height? Use $\alpha = 0.05$. [4 points]

Model Response

State. Let μ_d be the true mean difference in jump height (new warm-up minus standard warm-up) for players in the league. $H_0 : \mu_d = 0$ vs. $H_a : \mu_d \neq 0$; $\alpha = 0.05$. **Plan.** Paired t-test (one-sample t-test on the differences). Random: players were randomly selected and the order of warm-ups was randomized. 10 percent: 15 players are less than 10 percent of a very large league. Normal: $n = 15 < 30$, but the dotplot of the differences is roughly symmetric with no outliers. **Do.** $t = \dfrac{1.9}{2.8/\sqrt{15}} = 2.63$, $df = 14$, p-value $= 0.0199$. **Conclude.** Since 0.0199 is less than 0.05, we reject H_0. There is convincing evidence that the new warm-up changes the true mean jump height of players in the league.

What the grader looks for

1 point: identifies the data as paired and defines μ_d with the direction of subtraction, with correct hypotheses. 1 point: checks Random, 10 percent and Normal/Large Sample using the differences. 1 point: correct t, $df = n - 1 = 14$ and p-value. 1 point: decision tied to p-value and α, with a conclusion about the mean change in context. Practice rubric, not official scoring.

Coach's Playbook

1. Ask: is each value linked to exactly one value in the other list? If yes, it is paired.
2. Choose and state the direction of the difference; define μ_d in words.
3. Compute or list the differences; graph them (dotplot or boxplot) and check conditions on the differences.
4. Run the one-sample t procedure on the differences with $df = n - 1$, where n counts pairs.
5. Conclude about the true **mean difference** in context, using the same direction as your definition.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

- If summary statistics are given for the differences, you can go straight to the one-sample formulas.
- If summary statistics are given only for the two original lists, you often cannot do a paired analysis, because s_d depends on how strongly the pairs are related.
- With a matched-pairs experiment, randomize which treatment each subject receives first (or which twin gets which treatment).
- A confidence interval for μ_d that contains 0 means there is not convincing evidence of a change.

Common traps

1) Using the two-sample formula on paired data. 2) Counting $2n$ instead of n in df. 3) Checking Normality of the original lists rather than the differences. 4) Reversing the direction of the difference between the hypotheses and the data.

Mistake Clinic

Spot the error

A student analyzes the before and after typing speeds of the same 10 students with a two-sample t-test and reports $df \approx 18$. **Where it fails:** the same students supply both lists, so the samples are not independent. **Correct solution:** paired t-test on the 10 differences with $df = 9$.

Spot the error

A student writes $H_0 : \mu_{before} = \mu_{after}$ and uses 'both lists are roughly symmetric' for the Normal condition. **Where it fails:** for paired data the parameter is μ_d and the condition concerns the distribution of the differences, which can look quite different from the two original lists. **Correct solution:** $H_0 : \mu_d = 0$, and graph the differences.

Spot the error

A student has $\bar{x}_d = -3.4$ (before minus after) and a conclusion 'convincing evidence that the program decreases speed.' **Where it fails:** with before minus after, a negative mean means the after values are larger, so speed increased. **Correct solution:** define the direction carefully; here the program increases mean typing speed by an estimated 3.4 words per minute.

Real-World Connection

Matched designs are everywhere in medicine and sport: patients are measured before and after treatment, and athletes try both the old and the new equipment. By letting each person act as their own control, researchers make a study far more sensitive to real effects, which means smaller, cheaper studies can still deliver convincing evidence.

Challenge Corner

In the typing example, how much does pairing narrow a 90 percent interval? **Solution:** the paired interval uses $\bar{x}_d = 3.4$, $s_d = 1.578$: $3.4 \pm 1.833(1.578/\sqrt{10}) = (2.49, 4.31)$, width 1.83. A two-sample interval for the same lists has $SE = 3.67$, $df \approx 17.8$, $t^* = 1.735$, giving width 12.75, more than 6 times wider even though it is centered on the same difference. Pairing was worth far more than collecting extra students.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE Paired Data and Matched Pairs

Show your work. A calculator is allowed. Justify conclusions in context.

1. Which study should be analyzed with a paired t-procedure?
   - **(A)** Each of 20 patients has blood pressure measured before and after taking a medication.
   - **(B)** Two separate random samples of 20 patients, one taking the medication and one taking a placebo.
   - **(C)** One random sample of 40 patients whose blood pressures are compared with a known national mean.
   - **(D)** Two surveys of different numbers of patients at two clinics.
2. A paired study has 18 patients, each measured twice. How many degrees of freedom does the paired t-test have?
   - **(A)** 35 **(B)** 17 **(C)** 18 **(D)** 34
3. In a paired design, which hypotheses test whether the mean of 'after minus before' differs from 0?
   - **(A)** $H_0 : \mu_{after} = \mu_{before}$ vs. $H_a : \mu_{after} > \mu_{before}$ using two samples
   - **(B)** $H_0 : \bar{x}_d = 0$ vs. $H_a : \bar{x}_d /= 0$
   - **(C)** $H_0 : \mu_d = 0$ vs. $H_a : \mu_d /= 0$
   - **(D)** $H_0 : \mu_d /= 0$ vs. $H_a : \mu_d = 0$
4. A random sample of 25 pairs gives differences with mean 2.4 and standard deviation 5.0. What is the t statistic for testing $H_0 : \mu_d = 0$?
   - **(A)** 0.48 **(B)** 12.00 **(C)** 4.80 **(D)** 2.40
5. A random sample of 14 pairs gives the boxplot of differences shown. What should be concluded about using a paired t-procedure?

   - **(A)** The Normal condition is in doubt because of the strong outlier in the differences with a small number of pairs.
   - **(B)** The condition is met because the original lists were roughly Normal.
   - **(C)** The condition is met because there are 14 pairs and 14 is less than 30.
   - **(D)** A two-sample test should be used instead because of the outlier.
6. A random sample of 16 pairs has mean difference 2.5 and standard deviation 3.2. What is the 95 percent confidence interval for μ_d?
   - **(A)** (0.93, 4.07) **(B)** (0.79, 4.21) **(C)** (1.35, 3.65)
   - **(D)** (-4.32, 9.32)
7. Why can a paired design be more powerful than a design with two independent samples of the same total size?
   - **(A)** Pairing makes the sample size larger.
   - **(B)** Pairing guarantees that the conditions are met.
   - **(C)** Differences within pairs remove variation between subjects.
   - **(D)** Paired data always have larger means.
8. A paired study computed 'before minus after' differences with $\bar{x}_d = -3.4$ minutes of study time. Which is the correct interpretation of the sign?
   - **(A)** The after values were greater on average by about 3.4 minutes.
   - **(B)** The before values were greater on average by about 3.4 minutes.
   - **(C)** The study contains an error because differences cannot be negative.
   - **(D)** The result is not statistically significant.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. A 95 percent confidence interval for μ_d (mean of after minus before differences) from 20 randomly selected pairs is (-0.4, 2.8). What is the best conclusion?
   (A) The true mean change is 2.4.
   (B) There is convincing evidence that scores increased.
   (C) There is convincing evidence that scores decreased.
   (D) There is not convincing evidence that the true mean change differs from 0.
10. For each study, state whether the data are paired or come from two independent samples, and explain.
    (a) A shoe company has 30 runners run a mile in each of two shoe types, in random order.
    (b) A teacher randomly assigns 15 students to flash cards and 15 to a video review and compares their test scores.
    (c) A researcher compares the heights of 25 randomly chosen fathers and the heights of their 25 adult sons.
11. A tutor randomly selects 9 students from a large school and records their scores before and after a workshop. Before: 68, 74, 59, 81, 70, 63, 77, 66, 72. After: 72, 79, 60, 85, 76, 64, 82, 70, 75. A dotplot of the differences is roughly symmetric with no outliers. Is there convincing evidence that the workshop increases the true mean score? Use $\alpha = 0.05$.
12. A researcher randomly selects 35 adults and measures the time (in seconds) to complete a task in a quiet room and in a noisy room, in random order. The differences (noisy minus quiet) have mean 0.82 seconds and standard deviation 1.9 seconds. Construct and interpret a 90 percent confidence interval for the true mean difference.
13. Twenty pairs of twins are used to compare two reading programs: one twin in each pair is randomly assigned to program 1 and the other to program 2. The differences in reading-test scores (program 1 minus program 2) are strongly right skewed with one extreme value. (a) List the conditions for a paired t-procedure. (b) Which is in doubt, and why? (c) What should the researcher do?
14. Eight randomly selected sprinters ran the same course in shoe A and shoe B in random order. Times (s): A: 12.8, 13.4, 12.1, 14.0, 13.1, 12.6, 13.7, 12.9; B: 12.6, 13.1, 12.0, 13.6, 12.9, 12.3, 13.5, 12.5. (a) Test for a difference in true mean time with the correct procedure at $\alpha = 0.05$ (assume the differences show no strong skew or outliers). (b) Explain why a two-sample t-test would be the wrong procedure and what p-value it would give.
15. Software output for a paired t-test (differences = after minus before, n = 12 pairs) reads: Mean difference = 2.1, SE of mean = 0.9, T = 2.33, DF = 11, P (two-sided) = 0.040. (a) State the hypotheses. (b) Give the conclusion at $\alpha = 0.05$. (c) Find a 95 percent confidence interval and explain how it agrees with the test.

T Unit 8 Test

Check your command of t-distributions, one-sample and two-sample t procedures, and paired data. Show all work and justify conclusions in context.

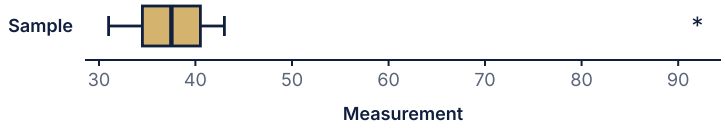
Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Part A: Multiple Choice

UNIT TEST **Unit 8 Test**

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. A random sample of 10 values will be used for a 99 percent confidence interval for a mean. What is the critical value?
   (A) 2.576 **(B)** 3.250 **(C)** 3.169 **(D)** 2.262
2. Compared with the standard Normal distribution, a t-distribution with 5 degrees of freedom has
   (A) more area in the tails and a lower peak **(B)** less area in the tails and a higher peak
   (C) a mean of 5 **(D)** the same shape, only shifted
3. A random sample of 12 values (from a very large population) has the boxplot shown. A student wants a t-interval for the mean. Which is correct?



   (A) All conditions are met because the sample is random.
   (B) The Normal condition is not met because of the outlier and $n < 30$.
   (C) The 10 percent condition fails.
   (D) The t-interval is fine because the median is in the box.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

4. A random sample of 22 phone batteries has mean life 15.8 hours and standard deviation 3.1 hours; no outliers or strong skew. What is the 95 percent confidence interval for the true mean life?
   (A) (14.50, 17.10) **(B)** (15.51, 16.09) **(C)** (14.43, 17.17)
   (D) (9.35, 22.25)

5. A 95 percent confidence interval for the mean time spent on homework is (1.8, 2.6) hours. Which interpretation is correct?
   (A) 95 percent of students do homework between 1.8 and 2.6 hours.
   (B) There is a 95 percent probability that the sample mean is between 1.8 and 2.6 hours.
   (C) The true mean is within 0.4 hours of 2.2 hours 95 percent of the time for this interval.
   (D) We are 95 percent confident that the true mean homework time is between 1.8 and 2.6 hours.

6. A random sample of 28 values (no outliers) has mean 52.1 and standard deviation 6.4. For $H_0 : \mu = 50$ vs. $H_a : \mu > 50$, what is the p-value?
   (A) 0.0470 **(B)** 0.0939 **(C)** 0.0413 **(D)** 0.9530

7. A test of $H_0 : \mu = 100$ vs. $H_a : \mu \neq 100$ gives a p-value of 0.07. Which conclusion is appropriate at $\alpha = 0.05$?
   (A) Reject H_0; the mean differs from 100.
   (B) Accept H_0; the mean equals 100.
   (C) Fail to reject H_0; there is not convincing evidence that the mean differs from 100.
   (D) Fail to reject H_0; the probability that H_0 is true is 0.07.

8. Independent random samples: $n_1 = 32$, $\bar{x}_1 = 21.4$, $s_1 = 5.2$; $n_2 = 35$, $\bar{x}_2 = 18.9$, $s_2 = 4.6$. What is the 95 percent confidence interval for $\mu_1 - \mu_2$ with the calculator's degrees of freedom?
   (A) (-0.83, 5.83) **(B)** (0.09, 4.91) **(C)** (0.80, 4.20)
   (D) (-2.31, 7.31)

9. Independent random samples ($n_1 = 40$, $\bar{x}_1 = 66$, $s_1 = 8$; $n_2 = 45$, $\bar{x}_2 = 62$, $s_2 = 9$) are used to test $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 > 0$. What is the p-value?
   (A) 0.0329 **(B)** 0.9835 **(C)** 0.0150 **(D)** 0.0165

10. Which study requires a paired t-procedure?
    (A) A random sample of 16 students is timed on a puzzle with and without background music, in random order.
    (B) Two random samples of 16 students are timed on the puzzle, one with music and one without.
    (C) One random sample of 16 students is compared with a claimed mean time of 5 minutes.
    (D) Students in two different schools are compared on their puzzle times.

11. Paired data from 15 randomly chosen subjects give differences with mean -1.6 and standard deviation 4.2 (no outliers). For $H_0 : \mu_d = 0$ vs. $H_a : \mu_d \neq 0$, what is the p-value?
    (A) 0.081 **(B)** 0.151 **(C)** 0.162 **(D)** 0.056

12. In an earlier unit you built a z-interval for a proportion. A random sample of 400 voters has $\hat{p} = 0.42$. What is the margin of error of a 95 percent confidence interval for p?
    (A) 0.967 **(B)** 0.048 **(C)** 0.0012 **(D)** 0.097

13. [6 points] A cafe owner wants to check that the mean volume of its large cups is 25 centiliters. She randomly selects 12 cups from a very large number filled in one week and measures the volume (centiliters): 24.1, 26.3, 22.8, 25.0, 27.2, 23.5, 25.8, 24.6, 26.0, 23.9, 25.4, 24.9. A dotplot of the data is roughly symmetric with no outliers.
    (a) [2 points] Verify the conditions for a t procedure.
    (b) [2 points] Construct a 95 percent confidence interval for the true mean volume.
    (c) [2 points] Interpret the interval and say what it indicates about the claim of 25 centiliters.

14. [10 points] A teacher wants to know whether practicing with a timed quiz app improves exam performance. She has 20 volunteers and plans a comparison of two groups.
    (a) [2 points] Describe how to randomly assign the 20 volunteers to the app group and the no-app group, and explain why random assignment is used.
    (b) [2 points] After the term, exam scores are: app group: 85, 92, 78, 88, 95, 81, 90, 86, 83, 91; no-app group: 76, 84, 80, 79, 88, 72, 82, 77, 75, 85. Side-by-side boxplots are roughly symmetric with no outliers. Verify the conditions for a two-sample t-test.
    (c) [4 points] Test whether the data provide convincing evidence that the app increases the true mean exam score. Use $\alpha = 0.05$.
    (d) [2 points] Construct a 90 percent confidence interval for the difference in true mean scores (app minus no app) and interpret it.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Answers

Inference for Means: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

8.1 The t-Distributions and Conditions for Means

PRACTICE · THE T-DISTRIBUTIONS AND CONDITIONS FOR MEANS

1. B. The sample gives $df = n - 1 = 19$, and 2.5 percent lies in each tail, so $t^* = 2.093$. Using 1.960 treats the problem as a z problem, and 2.086 comes from $df = 20$ (using n instead of $n - 1$).
2. A. Estimating σ with s adds variability, which puts more area in the tails and so lowers the center. Every t-distribution is symmetric about 0; the degrees of freedom control the spread, not the mean.
3. C. Use the t cumulative distribution function with $df = 8$ to get 0.0185. The value 0.0062 is the Normal tail, which is too small; 0.0369 doubles it, which is only right for a two-sided alternative.
4. D. For one sample, $df = n - 1 = 25$. Choosing 26 forgets to subtract 1, and 52 would come from adding the sample size to itself rather than a one-sample setting.
5. A. All three listed conditions must hold. Randomness and the 10 percent condition are fine, but with $n = 12 < 30$ the graph shows a strong outlier, so Normality of the population is not plausible. Each wrong choice speaks to only one condition.
6. C. The sample is 60, but 10 percent of 400 is 40, so $n > 0.10N$. The sample is random and $n \geq 30$ takes care of Normality, which is why the tempting choices are wrong.
7. A. More degrees of freedom make the t curve look more like the Normal curve, so t^* falls toward $z^* = 1.960$. It stays slightly above because the t curve always keeps more tail area than the Normal curve.
8. A. With $n < 30$ we rely on the shape of the data to judge Normality. Sample size 10 does not invoke the central limit theorem, and t procedures do have conditions at every sample size.
9. B. With $df = 8$ and 5 percent in each tail, $t^* = 1.860$. The value 1.645 is the z value, and 2.306 leaves 2.5 percent per tail (a 95 percent interval).
10. (a) $df = 15$, 1 percent in each tail: $t^* = 2.602$.
    (b) The z value is 2.326. The t value is larger because s replaces σ, which creates additional variability, and the t curve has heavier tails. The difference shrinks as n grows.
11. (a) $P(T > 1.75) = 0.0528$.
    (b) Double it: 0.1056.
    (c) $P(Z > 1.75) = 0.0401$, smaller, since the t curve has heavier tails. Using z would overstate the evidence against H_0.
12. Random: the cakes were randomly selected. 10 percent: 15 cakes are well under 10 percent of the many cakes baked. Normal/Large Sample: $n = 15 < 30$, but the dotplot is roughly symmetric and has no outliers, so a Normal population is plausible. All conditions are met, so a t procedure with $df = 14$ is reasonable.
13. (a) With $n = 25 < 30$ and strong skewness, the Normal/Large Sample condition is not satisfied; the 10 percent condition only addresses independence.
    (b) The student should not use a t interval as it stands: gather a larger sample ($n \geq 30$, and still check for extreme outliers) or use a different method. The student should also report the shape and any outliers.
14. (a) Random: stated. 10 percent: 35 is far less than 1,200. Normal/Large Sample: the data are skewed but $n = 35 \geq 30$, so the sampling distribution of $\bar{x}$ is approximately Normal by the central limit theorem. Conditions are met.
    (b) With $n = 8$ the large-sample argument no longer applies, and the right skew would be evidence against a Normal population, so the Normal condition would fail.
15. A t-curve with few degrees of freedom has heavier tails, so more of its area lies beyond 1.96, and to capture the middle 95 percent we must go farther out. For $df = 3$, the 97.5th percentile is $t^* = 3.182$, so the middle 95 percent lies between -3.182 and 3.182.

8.2 One-Sample t-Intervals

PRACTICE · ONE-SAMPLE T-INTERVALS

1. A. Use $t^* = 2.030$ with $df = 35$ and $SE = 3.0/\sqrt{36} = 0.5$, so $ME = 1.015$. Using 1.96 ignores that s was estimated, and dividing by n instead of $\sqrt{n}$ gives a margin far too small.
2. B. The interval is a statement about the true mean, and the confidence refers to the method. The other choices describe individual values or the sample mean, which is known exactly from the data.
3. C. A larger n lowers $s/\sqrt{n}$ (and slightly lowers t^*). A higher confidence level and a larger s both widen the interval, and a smaller n does too.
4. D. The interval is centered at $\bar{x}$, so the mean is the midpoint $(40.6 + 50.2)/2 = 45.4$. The value 4.8 is the margin of error, and 9.6 is the full width.
5. B. $SE = ME/t^* = 2.06/2.064 \approx 1.00$. Taking the margin of error itself as the SE skips the division by t^*, and multiplying by t^* goes the wrong way.
6. C. Random and 10 percent are fine, but with $n = 15$ the strong skew means the Normal condition fails. An unknown σ is exactly why we use t, so it is not a reason to avoid it.
7. A. With $n = 12$, $df = 11$ and 0.5 percent lies in each tail, so $t^* = 3.106$. The value 2.576 is a z critical value, and 3.055 comes from using $df = n$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

8. B. The margin of error is proportional to $1/\sqrt{n}$, so halving it requires $\sqrt{n}$ to double, which means n must quadruple. Doubling n only reduces the margin by a factor of about 1.41.
9. D. $\bar{x} = 7.040$, $s = 0.222$, $df = 9$, $t^* = 2.262$, so $ME = 0.159$. The z-based interval is too narrow, using the 90 percent multiplier is also too narrow, and omitting $\sqrt{n}$ makes the interval far too wide.
10. (a) Random: stated. 10 percent: 40 is less than 150, which is 10 percent of 1,500. Normal/Large Sample: $n = 40 \geq 30$. (b) $df = 39$, $t^* = 1.685$; $3.2 \pm 1.685(1.1/\sqrt{40}) = 3.2 \pm 0.293$, or (2.91, 3.49). (c) We are 90 percent confident that the true mean weekly reading time for all students at the school is between 2.91 and 3.49 hours.
11. Random: the loaves were randomly selected. 10 percent: this requires the production to be at least 120 loaves; several hundred is stated. Normal: $n = 12 < 30$ but the dotplot has no strong skew or outliers. $\bar{x} = 42.67$, $s = 2.774$, $df = 11$, $t^* = 2.201$, $ME = 1.76$, interval (40.90, 44.43). We are 95 percent confident that the true mean baking time of all loaves from that day is between 40.90 and 44.43 minutes.
12. (a) The interval estimates the population mean, not individual commutes; individual commutes vary far more than the mean does. (b) We are 95 percent confident that the interval from 24.1 to 29.5 minutes captures the true mean commute time of the population sampled. (c) Wider: a higher confidence level uses a larger t^*, so the margin of error increases.
13. (a) $\bar{x} = (18.6 + 23.4)/2 = 21.0$; $ME = (23.4 - 18.6)/2 = 2.4$. (b) With $df = 29$, $t^* = 2.045$, so $SE = 2.4/2.045 = 1.173$ and $s = SE\sqrt{30} = 6.43$.
14. Conditions: random, 16 is under 10 percent of a very large run, and the data show no strong skew or outliers. $\bar{x} = 497.938$, $s = 2.594$, $df = 15$, $t^* = 2.947$, so the interval is (496.03, 499.85). Because 500 is outside the interval, the data provide convincing evidence at the 1 percent level that the true mean differs from 500 grams.
15. $n = 25$: $ME = 2.064(8/5) = 3.302$. $n = 100$: $ME = 1.984(8/10) = 1.587$. Quadrupling the sample size roughly halves the margin of error (ratio about 2.08), because the standard error is proportional to $1/\sqrt{n}$.

8.3 One-Sample t-Tests

PRACTICE · ONE-SAMPLE T-TESTS

1. A. The null hypothesis contains the equal sign and the alternative reflects the research question, 'less than'. Hypotheses concern the parameter μ, not the statistic $\bar{x}$, and a two-sided alternative would not match the wording.
2. B. With $df = 19$, $P(T < -2.10) = 0.0247$. Doubling is for a two-sided test, the Normal value 0.0179 ignores the heavier tails, and 0.9753 is the upper tail, which goes in the wrong direction for this alternative.
3. C. A p-value is a probability computed assuming H_0 is true, about the data. The other statements treat it as the probability of a hypothesis or of a share of the population.
4. D. A p-value larger than α means we fail to reject H_0. That is not the same as showing H_0 true, which is what 'accept' wrongly suggests.
5. A. A two-sided test at α agrees with a $1 - \alpha$ interval: values outside the interval are rejected. The interval always contains $\bar{x}$, and its width does not decide the matter.
6. B. With $n = 14$ the shape matters, and the outlier at 58 is evidence against a Normal population. A random sample does not guarantee Normality, and 14 is not more than 10 percent of a very large population.
7. C. A Type I error means rejecting a true null hypothesis. The first alternative describes failing to detect a real shortfall, which is a Type II error.
8. D. $t = (103 - 100)/(10/\sqrt{25}) = 3/2 = 1.50$. Using 7.5 divides by s/n rather than $s/\sqrt{n}$, 0.30 divides by s alone, and 0.60 uses a standard error of 5.
9. B. A larger n reduces the standard error, making real differences easier to detect. Smaller α, a true mean nearer μ_0 and more variability all lower power.
10. (a) $H_0 : \mu = 110$, $H_a : \mu > 110$, where μ is the true mean muffin weight (g). (b) $H_0 : \mu = 12.5$, $H_a : \mu \neq 12.5$, where μ is the true mean sprint time (s) with the new warm-up. (c) $H_0 : \mu = 25$, $H_a : \mu > 25$, where μ is the true mean commute time (minutes) for the school.
11. State: $H_0 : \mu = 8$ vs. $H_a : \mu \neq 8$, μ = true mean nightly sleep of students at the school; $\alpha = 0.05$. Plan: one-sample t-test; random sample, 10 is under 10 percent of a large school, and the roughly symmetric dotplot with no outliers supports Normality for $n = 10$. Do: $\bar{x} = 7.92$, $s = 0.563$, $t = -0.45$, $df = 9$, two-sided p-value 0.6640. Conclude: since 0.6640 is greater than 0.05, we fail to reject H_0; there is not convincing evidence that the true mean nightly sleep differs from 8 hours.
12. (a) $H_0 : \mu = 23.5$ vs. $H_a : \mu > 23.5$. Random sample, 50 is under 10 percent of millions, $n = 50 \geq 30$. $t = \frac{24.8 - 23.5}{5.9/\sqrt{50}} = 1.56$, $df = 49$, p-value $= 0.0628$. Since 0.0628 is greater than 0.05, we fail to reject H_0; there is not convincing evidence that the true mean exceeds 23.5 seconds. (b) A Type I error would be concluding that the true mean load time is greater than 23.5 seconds when it is actually 23.5 seconds.
13. (a) If the true mean bag weight were 50 grams, there is about a 2.1 percent chance of getting a sample mean this far below 50 grams or farther just by random sampling. (b) At $\alpha = 0.05$: 0.021 is less than 0.05, so reject H_0 and conclude convincing evidence the true mean weight is below 50 grams. At $\alpha = 0.01$: 0.021 is greater than 0.01, so fail to reject H_0 and conclude the data do not give convincing evidence at that stricter level. (c) A significance test measures how surprising the data are under H_0; with a Type I error probability of α a true H_0 is still occasionally rejected.
14. (a) 12 is inside the interval, so fail to reject H_0: there is not convincing evidence that the true mean differs from 12 hours. (b) 14 is outside the interval, so reject H_0: there is convincing evidence that the true mean differs from 14 hours (and the interval suggests it is lower). (c) Not directly: a 95 percent interval matches a two-sided test at $\alpha = 0.05$; a one-sided test at $\alpha = 0.05$ corresponds to a 90 percent interval.
15. Random and 10 percent conditions are fine, but $n = 12 < 30$ and the data contain a strong outlier (22), so the Normal condition is not satisfied; the outlier would also inflate both $\bar{x}$ and s. The manager should investigate whether 22 is a data entry error or a genuine long wait, and collect a larger sample (at least 30 customers) before using a t-test.

8.4 Two-Sample t-Intervals

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

PRACTICE · TWO-SAMPLE T-INTERVALS

1. A. $SE = \sqrt{8^2/40 + 6^2/50} = \sqrt{1.6 + 0.72} = 1.523$. Adding the two standard errors skips the square root of the sum, and 2.320 forgets the square root entirely.
2. B. Variances of independent sample means add, and then the square root is taken. Adding the two standard errors overstates the variability, and the other options do not combine the variances correctly.
3. C. The interval is about the difference of true means, and the confidence level describes the method. Sample means are known exactly, and the interval does not describe individual students.
4. D. An interval containing 0 means a difference of 0 is plausible, so no convincing evidence of a difference. It does not show the means are equal, and negative and positive values are both plausible.
5. A. For small groups, each group must be plausibly Normal. The 10 percent condition is for sampling, not experiments, and unpooled methods do not require equal standard deviations or equal sample sizes.
6. D. $SE = 0.990$, $df \approx 50.4$, $t^* = 2.677$, so $ME = 2.65$ and the interval is 1.7 ± 2.65. The conservative $df = 24$ gives a slightly wider interval, the z value a narrower one, and reversing the order flips the signs.
7. B. The unpooled method estimates degrees of freedom from both sample sizes and both standard deviations, which usually gives a non-integer value in that range. The value $n_1 + n_2 - 2$ belongs only to the pooled method, which we do not use.
8. C. The Normal condition must hold for each group, and group 1's outlier with $n = 12$ violates it. Different sample sizes and different centers are normal in a comparison.
9. B. Each standard error is proportional to $1/\sqrt{n}$, and quadrupling n halves it. The multiplier t^* also shrinks slightly, so the margin of error is a little under half the original.
10. Conditions: random samples, each under 10 percent of its population, both $n \geq 30$, groups independent. $SE = \sqrt{8.6^2/45 + 9.4^2/48} = 1.867$, $df \approx 90.9$, $t^* = 1.662$. Interval: $4.3 \pm 3.10 = (1.20,\ 7.40)$. We are 90 percent confident that the true mean quiz score for students using flash cards is between 1.20 and 7.40 points higher than for students not using them.
11. Random samples (assumed less than 10 percent of the populations), independent groups; both small samples show no strong skew or outliers. Mix: $\bar{x} = 53.62$, $s = 3.021$. No mix: $\bar{x} = 48.00$, $s = 2.739$. $SE = 1.405$, $df \approx 14.3$, $t^* = 2.141$, $ME = 3.01$. Interval: (2.62, 8.63). We are 95 percent confident that the true mean height with the nutrient mix is between 2.62 and 8.63 cm greater than without it.
12. (a) Group 2: $n = 40 \geq 30$, so the condition is met. Group 1: $n = 10 < 30$ and the graph shows strong skew and an outlier, so the condition fails. (b) No, because the condition must hold for both groups. The researcher should investigate the outlier, collect a larger sample for group 1 (at least 30), or use a different method.
13. (a) We are 95 percent confident that the true mean sleep for athletes is between 0.15 hours less and 0.62 hours more than for non-athletes. (b) No: 0 is in the interval, so a difference of 0 is plausible and there is not convincing evidence of a difference. (c) Not correct: the interval also contains differences of up to about 0.6 hours; failing to find evidence of a difference does not prove the means are equal.
14. $\bar{x}_M - \bar{x}_S = -2.92$, $SE = 0.699$, $df \approx 22.0$, $t^* = 2.074$, $ME = 1.45$. Interval: (-4.37, -1.47) words. It is the mirror image of the silence-minus-music interval (1.47, 4.37): the signs flip and the width is unchanged. We are 95 percent confident that the true mean recall with music is between 1.47 and 4.37 words lower than in silence.
15. (a) The point estimate is the midpoint $(1.2 + 5.8)/2 = 3.5$ and the margin of error is $(5.8 - 1.2)/2 = 2.3$. (b) Wider: a higher confidence level uses a larger critical value. (c) Yes at the 10 percent level, since 0 is not in the 90 percent interval; the plausible differences are all positive, so μ_1 appears greater than μ_2.

8.5 Two-Sample t-Tests

PRACTICE · TWO-SAMPLE T-TESTS

1. A. Lower scores for app users mean $\mu_1 < \mu_2$, so the difference is negative. Choosing the right tail depends on the order, the null must carry the equal sign, and hypotheses use parameters, not statistics.
2. B. The denominator is $\sqrt{6^2/30 + 7^2/36} = 1.600$, so $t = 3.2/1.600 = 2.00$. Averaging the variances, adding the standard errors or multiplying by $\sqrt{2}$ gives different, incorrect values.
3. C. The alternative is one-sided and the direction matches, so the p-value is the lower tail: 0.0181. Doubling is for a two-sided test, the upper tail goes the wrong way, and the last value is neither tail properly.
4. D. Reject H_0 since 0.04 is less than 0.05, and state the conclusion about the population means in context. A p-value is not the probability that H_0 is true, and a small p-value leads to rejecting, not failing to reject.
5. A. The Normal condition must hold for each group, and the outlier in group 2 with only 12 values violates it. Equal sample sizes and different medians say nothing about Normality.
6. A. 0 is inside the interval, so the two-sided test at $\alpha = 0.05$ fails to reject. Mostly positive is not the same as entirely positive, and 'accept' wrongly claims proof.
7. B. A Type I error means rejecting a true H_0. Missing a real difference is a Type II error, and a small p-value is not an error in itself.
8. C. Plants in separate groups with no natural pairing form independent samples. The other three situations match each observation with a related one, which calls for a paired method.
9. D. Large samples have small standard errors, so even a small difference can produce a small p-value. Statistical significance does not measure how large the difference is.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10. State: $H_0 : \mu_T - \mu_A = 0$ vs. $H_a : \mu_T - \mu_A \neq 0$ (T = teens, A = adults). Plan: two-sample t-test; random samples, 10 percent satisfied, both $n \geq 30$. Do: $t = 1.78$, $df \approx 97.7$, p-value $= 0.0782$. Conclude: since 0.0782 is greater than 0.05, we fail to reject H_0; there is not convincing evidence that the true mean weekly screen time differs for teenagers and adults.
11. State: $H_0 : \mu_X - \mu_Y = 0$ vs. $H_a : \mu_X - \mu_Y > 0$. Plan: two-sample t-test; random assignment; both dotplots show no strong skew or outliers. Do: X: mean 41.5, $s = 3.37$; Y: mean 34.00, $s = 3.32$; $t = 5.13$, $df \approx 18.7$, p-value $= lessthan 0.0001$. Conclude: since less than 0.0001 is less than 0.05, we reject H_0; there is convincing evidence that the true mean score is higher with method X than with method Y.
12. (a) 0.037 is less than 0.05, so reject H_0: there is convincing evidence that the true mean for group 1 is greater than for group 2. (b) 0.037 is greater than 0.01, so fail to reject H_0: the data do not provide convincing evidence at that level. (c) If the true means were equal, the probability of seeing a difference in sample means of 3.1 or more (in favor of group 1) just from random sampling or random assignment is about 0.037.
13. (a) Type I error: concluding that the true mean sleep differs between the groups when it really does not. (b) Type II error: failing to find convincing evidence of a difference when the true means really differ. (c) Because H_0 was rejected, the only error that can have been made is a Type I error (probability $\alpha = 0.05$ under H_0).
14. (a) $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 \neq 0$; random samples, both $n \geq 30$. $t = 1.73$, $df \approx 83.0$, p-value $= 0.0876$. Since 0.0876 is greater than 0.05, fail to reject H_0; there is not convincing evidence of a difference in true means. (b) Interval: (-0.09, 1.29), which contains 0, consistent with failing to reject at the 5 percent level.
15. (a) Random sampling is met, and (assuming each city has far more than 90 bicycles for sale) the 10 percent condition holds; groups are independent. City 2 may satisfy the Normal condition. (b) No: city 1 has $n = 9 < 30$ with strong skew and an outlier, so the Normal condition fails for that group and the p-value would be unreliable. (c) She could take larger samples (at least 30 in each city), investigate the outlier, or use a method not based on Normality.

8.6 Paired Data and Matched Pairs

PRACTICE · PAIRED DATA AND MATCHED PAIRS

1. A. The same patients provide the two measurements, so the data are paired. The other designs have unrelated groups or only one sample.
2. B. The test uses the 18 differences, so $df = n - 1 = 17$. The value 34 or 35 would come from treating the data as two independent samples.
3. C. The parameter is the population mean of the differences, μ_d. Hypotheses about sample statistics are invalid, the null has the equal sign, and the two-mean version treats the groups as independent.
4. D. $t = 2.4/(5.0/\sqrt{25}) = 2.4/1.0 = 2.40$. Dividing by 5 gives 0.48, which forgets the square root of 25 in the standard error, and 4.80 uses a standard error of 0.5.
5. A. For paired data the condition applies to the differences. With $n = 14 < 30$, an extreme outlier is evidence against Normality; changing to a two-sample test would ignore the pairing.
6. B. With $n = 16$ pairs, $df = 15$, $t^* = 2.131$ and $ME = 2.131 \cdot 3.2/4 = 1.71$. Using 1.96 ignores the t correction, doubling the pairs for df treats the data as 32 independent values, and omitting the square root gives a very wide interval.
7. C. Subject-to-subject variability inflates the standard error of independent samples, while the differences in pairs remove it. Pairing does not add observations, and it does not guarantee any condition.
8. A. Before minus after is negative when after is larger. Negative differences are ordinary, and the sign alone says nothing about significance.
9. D. 0 is inside the interval, so no change is a plausible value. The midpoint is 1.2, not 2.4, and neither a decrease nor an increase is established.
10. (a) Paired: each runner provides a time in both shoe types; analyze the 30 differences. (b) Two independent samples: students in one group are not linked to students in the other. (c) Paired: each father is naturally matched to his son, so the 25 father-son differences are analyzed.
11. State: μ_d = true mean change (after minus before) for students at the school; $H_0 : \mu_d = 0$ vs. $H_a : \mu_d > 0$. Plan: paired t-test; random sample, 9 is under 10 percent of a large school, and the differences show no strong skew or outliers. Do: differences 4, 5, 1, 4, 6, 1, 5, 4, 3; $\bar{x}_d = 3.67$, $s_d = 1.732$, $t = 6.35$, $df = 8$, p-value $= 0.0001$. Conclude: since 0.0001 is less than 0.05, we reject H_0; there is convincing evidence that the workshop increases the true mean score.
12. Conditions: random sample, 35 adults under 10 percent of all adults, $n = 35 \geq 30$ pairs. $df = 34$, $t^* = 1.691$, $ME = 1.691(1.9/\sqrt{35}) = 0.543$. Interval: (0.28, 1.36) seconds. We are 90 percent confident that the true mean completion time in the noisy room is between 0.28 and 1.36 seconds longer than in the quiet room for adults.
13. (a) Random: random assignment within each pair of twins; 10 percent: not needed for the experiment; Normal/Large Sample: 20 pairs is under 30, so the differences must show no strong skewness or outliers. (b) The Normal condition is in doubt: the differences are strongly skewed with an extreme value and there are only 20 pairs. (c) Investigate the extreme value (error or real), consider collecting at least 30 pairs, and do not rely on the t-procedure's p-value until the condition is satisfied.
14. (a) μ_d = true mean (A minus B) time difference; $H_0 : \mu_d = 0$ vs. $H_a : \mu_d \neq 0$. Differences: 0.2, 0.3, 0.1, 0.4, 0.2, 0.3, 0.2, 0.4; $\bar{x}_d = 0.262$, $s_d = 0.106$, $t = 7.00$, $df = 7$, p-value $= 0.0002$. Since 0.0002 is less than 0.05, reject H_0: there is convincing evidence that the true mean times differ (shoe A is slower by about 0.262 seconds). (b) Each sprinter ran in both shoes, so the lists are linked, not independent. A two-sample test gives $t = 0.89$ and p-value 0.3892, which hides the effect because runner-to-runner differences are large compared with the shoe difference.
15. (a) $H_0 : \mu_d = 0$ vs. $H_a : \mu_d \neq 0$, where μ_d is the true mean change (after minus before). (b) 0.040 is less than 0.05, so reject H_0: there is convincing evidence that the true mean change is not 0 (it is positive, an increase). (c) $df = 11$, $t^* = 2.201$; $2.1 \pm 2.201(0.9) = (0.12, 4.08)$, which excludes 0, agreeing with rejecting H_0 at the 5 percent level.

T Unit 8 Test

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

UNIT TEST · UNIT 8 TEST

1. B. With $n = 10$, $df = 9$ and 0.5 percent in each tail, $t^* = 3.250$. The value 2.576 is the z critical value, and 3.169 uses $df = n$.
2. A. Estimating σ with s adds variability, so t-curves have heavier tails. All t-distributions are centered at 0, and shape depends on the degrees of freedom.
3. B. With a small sample the data must show no strong skew or outliers, and 92 is a clear outlier. Randomness does not address Normality, and 12 is far below 10 percent of a very large population.
4. C. $df = 21$, $t^* = 2.080$, $ME = 2.080 \cdot 3.1/\sqrt{22} = 1.37$. Using 1.96 is too narrow, dividing by 22 is far too narrow, and omitting the division by $\sqrt{n}$ gives an extremely wide interval.
5. D. The interval estimates the population mean, and 95 percent describes the method's long-run capture rate. The sample mean is known (2.2), and the interval says nothing about individual students.
6. A. $t = (52.1 - 50)/(6.4/\sqrt{28}) = 1.74$ and $df = 27$, so the upper-tail area is 0.0470. Doubling is for a two-sided test, the Normal tail is too small, and the lower tail is the wrong direction.
7. C. 0.07 is greater than 0.05, so we fail to reject. That does not prove the mean equals 100, and the p-value is not the probability that H_0 is true.
8. B. $SE = \sqrt{5.2^2/32 + 4.6^2/35} = 1.204$, $df \approx 62.2$, $t^* = 1.999$, so $ME = 2.41$. Adding the two standard errors, pooling the variances, or doubling the margin each gives a different, incorrect interval.
9. D. $SE = \sqrt{8^2/40 + 9^2/45} = 1.844$, $t = 4/1.844 = 2.17$ and $df \approx 83.0$, so the upper-tail p-value is 0.0165. The doubled value is for a two-sided test, the lower tail is the wrong direction, and the Normal tail is slightly too small.
10. A. Each student gives two measurements, so the data are paired. The other designs have separate groups or only one sample.
11. C. $t = -1.6/(4.2/\sqrt{15}) = -1.48$ with $df = n - 1 = 14$; the two-sided p-value is 0.162. The half value is the one-sided p-value, and 0.151 comes from wrongly using $df = 29$.
12. B. For a proportion use $z^*\sqrt{\hat{p}(1-\hat{p})/n} = 1.96\sqrt{0.42(0.58)/400} = 0.048$. A t-distribution is not used for proportions, and omitting the division by n or the square root gives the wrong size.
13. Scoring guide. (a) 2 points: 1 for Random and 10 percent with context (random selection; 12 cups is under 10 percent of a very large number), 1 for the Normal condition using the graph (n = 12 is under 30, but the dotplot is roughly symmetric with no outliers). (b) 2 points: 1 for correct $\bar{x}$, s, df and t^*, 1 for the interval. (c) 2 points: 1 for a correct interpretation (confidence level, parameter, context), 1 for linking 25 to the interval.

    Model response. (a) The cups were randomly selected, and 12 is under 10 percent of a very large number of cups. With $n = 12 < 30$ we check the dotplot: it is roughly symmetric with no outliers, so a Normal population is plausible. (b) $\bar{x} = 24.96$, $s = 1.265$, $df = 11$, $t^* = 2.201$, so $24.96 \pm 2.201(1.265/\sqrt{12}) = (24.15, 25.76)$. (c) We are 95 percent confident that the true mean volume of large cups filled that week is between 24.15 and 25.76 centiliters. Since 25 is inside this interval, the data do not provide convincing evidence against the claim of 25 centiliters.
14. Scoring guide. (a) 2 points: 1 for a valid random method (for example, label volunteers 1 to 20, use a random number generator to pick 10 for the app group, the rest are the control), 1 for the purpose (balances other variables between groups so a difference can be attributed to the app). (b) 2 points: 1 for random assignment (and 10 percent not needed for experiments), 1 for the Normal condition for each group (both n = 10 under 30, but boxplots show no strong skew or outliers). (c) 4 points: 1 for hypotheses with parameters, 1 for naming the test, 1 for correct t, df, p-value, 1 for decision and conclusion in context. (d) 2 points: 1 for the interval, 1 for interpretation with direction.

    Model response. (a) Number the volunteers 1 to 20, generate random integers from 1 to 20 without repeats, and put the first 10 distinct numbers in the app group; the others form the control group. Random assignment makes the groups similar in other respects (prior ability, motivation) on average. (b) Treatments were randomly assigned; the 10 percent condition does not apply to an experiment. Each group has n = 10 under 30, but each boxplot is roughly symmetric with no outliers, so Normality is plausible for each group. (c) $H_0 : \mu_A - \mu_N = 0$ vs. $H_a : \mu_A - \mu_N > 0$, where μ_A and μ_N are the true mean exam scores with and without the app for volunteers like these. Two-sample t-test: app mean 86.9, $s = 5.30$; no-app mean 79.8, $s = 4.98$; $t = 3.09$, $df \approx 17.9$, p-value $= 0.0032$. Since 0.0032 is less than 0.05, we reject H_0; there is convincing evidence that the app increases the true mean exam score. (d) With $t^* = 1.734$, the 90 percent interval is 7.1 ± 3.99, or (3.11, 11.09). We are 90 percent confident that the true mean exam score with the app is between 3.11 and 11.09 points higher than without it.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 9: Inference for Categorical Data: Chi-Square

Test distributions and relationships with chi-square procedures. 4 lessons, 74 practice and test items, all with full explanations.

In this unit · Beyond the 2027 core (see Part 1)

9.1 The Chi-Square Statistic and Goodness of Fit
9.2 Chi-Square Test for Homogeneity
9.3 Chi-Square Test for Independence
9.4 Choosing and Interpreting Chi-Square Tests
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

9.1 The Chi-Square Statistic and Goodness of Fit

Learn the chi-square statistic, its right-skewed distribution and the goodness-of-fit test, including expected counts, degrees of freedom, conditions and contributions.

Beyond the 2027 Exam Core

College Board's revision page lists the chi-square goodness-of-fit test as removed from the exam starting in 2027, so treat this lesson as extension material. It is still taught in college statistics courses and appears on earlier exams and in many classrooms. Check your teacher's current course description to see whether you will be tested on it.

Why This Matters

Suppose a snack company claims that 40% of its fans prefer pretzels, 30% prefer popcorn, 20% prefer chips and 10% prefer nuts. You survey 100 fans and see counts that do not match exactly. Is that mismatch just sampling variation, or is the claimed distribution wrong? The **chi-square goodness-of-fit test** compares a set of observed counts with the counts a claimed distribution would predict, and it gives one number that measures how far apart they are.

Students go wrong by testing counts as if they were proportions, by using the sample size instead of the number of categories to find the degrees of freedom, and by forgetting that the test only says whether the claim fits, not how it fails. This test belongs to categorical-data inference, alongside the proportion tests of the exam's Unit 3, but it handles variables with more than two categories.

Core Ideas

A categorical variable has k categories. The null hypothesis claims a specific proportion p_i for each category. The **observed counts** O_i are what you actually see. The **expected counts** $E_i = n \cdot p_i$ are what you would see on average if the claim were true. Large gaps between O_i and E_i are evidence against the claim.

The chi-square statistic

$\chi^2 = \sum \dfrac{(O-E)^2}{E}$, summed over all categories.

Each term $\dfrac{(O-E)^2}{E}$ is the **contribution** of one category. The statistic is never negative, and it is 0 only when every observed count equals its expected count.

Large χ^2 means the data are far from the claimed distribution; small χ^2 means they are close.

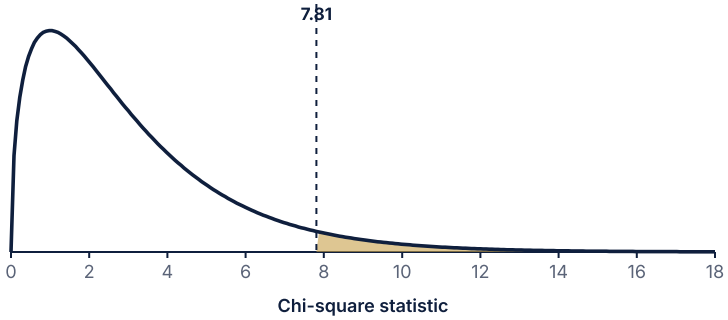
Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Squaring the differences keeps positive and negative gaps from cancelling, and dividing by E puts the size of a gap in perspective: a gap of 10 matters more when you expected 20 than when you expected 200.

The Chi-Square Distribution

When the null hypothesis is true and the conditions hold, χ^2 follows (approximately) a **chi-square distribution**. Every chi-square distribution is determined by its **degrees of freedom**, df. The curve starts at 0, is **skewed to the right**, and has mean equal to df. As df grows, the curve spreads out, shifts to the right and becomes more symmetric.



Chi-square distribution with 3 degrees of freedom. The shaded right tail beyond 7.81 has area 0.05.

The goodness-of-fit test: procedure

Hypotheses. H_0: the population proportions are $p_1, \ldots, p_k$ as claimed. H_a: at least one proportion is different.

Conditions. (1) Random: the data come from a random sample or randomized experiment. (2) Independent: if sampling without replacement, the sample is less than 10% of the population. (3) Large counts: every **expected** count is at least 5.

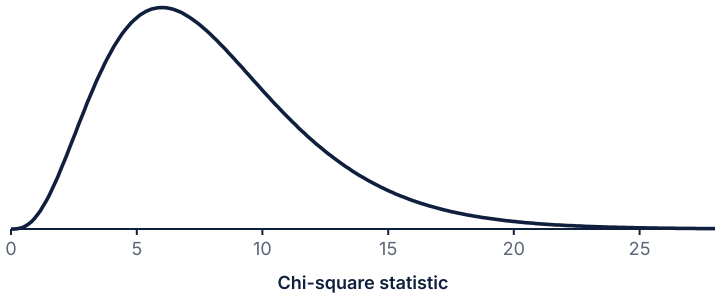
Statistic and df. $\chi^2 = \sum (O - E)^2 / E$ with $df = k - 1$, where k is the number of categories.

P-value. Always the **right tail**: $P(\chi^2 \geq \text{observed})$. Use a table or the built-in calculator's chi-square cdf from the statistic to 1E99.

Conclusion. Compare the p-value with α and state the decision in context.

The test is one-sided in a special way: because large values of χ^2 are the only evidence against H_0, the p-value is always the area to the right, even though H_a is written as "different." The conditions also have a subtle twist. The large counts condition uses **expected** counts, not observed counts.

df	1	2	3	4	5
Critical value for upper tail 0.05	3.84	5.99	7.81	9.49	11.07



With 8 degrees of freedom the curve is flatter and more symmetric, with its center near 8.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

COMPUTING THE STATISTIC

A spinner has four equal sectors. In 80 spins, the counts are 26, 15, 21 and 18. Test whether the sectors are equally likely. Find the expected counts, χ^2 and the p-value.

Sector	A	B	C	D
Observed	26	15	21	18
Expected	20	20	20	20

SOLUTION

Expected: $80(0.25) = 20$ per sector, all at least 5.

Contributions: $(26-20)^2/20 = 1.80$, $(15-20)^2/20 = 1.25$, $(21-20)^2/20 = 0.05$, $(18-20)^2/20 = 0.20$.

$\chi^2 = 3.30$ with $df = 4-1 = 3$. P-value $= P(\chi^2_3 \geq 3.30) = 0.348$.

Since 0.348 is large, there is no convincing evidence that the sectors are unequally likely.

READING A TABLE

A test statistic is $\chi^2 = 7.2$ with $df = 4$. Use the critical value table to bound the p-value, and give the calculator value.

SOLUTION

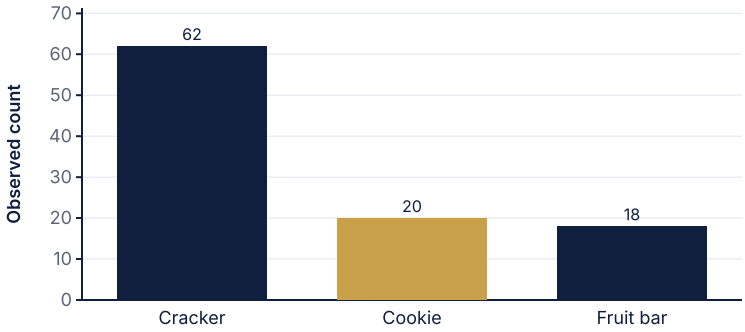
For $df = 4$, the critical value for upper tail 0.05 is 9.49. Since $7.2 < 9.49$, the p-value is greater than 0.05.

The calculator gives χ^2cdf(7.2, 1E99, 4) = 0.126.

Because 0.126 is larger than 0.05, the data do not provide enough evidence to reject H_0 at the 5% level.

A TEST THAT REJECTS

A vending machine's maker claims that 50% of purchases are crackers, 30% are cookies and 20% are fruit bars. In a random sample of 100 purchases there were 62 crackers, 20 cookies and 18 fruit bars. Test the claim at $\alpha = 0.05$.



Observed purchases (expected 50, 30, 20).

SOLUTION

H_0: the distribution is 50%, 30%, 20%. H_a: at least one proportion differs.

Conditions: random sample; 100 is under 10% of all purchases; expected counts are 50, 30, 20, all at least 5.

$\chi^2 = \frac{(62-50)^2}{50} + \frac{(20-30)^2}{30} + \frac{(18-20)^2}{20} = 2.88 + 3.33 + 0.20 = 6.41$, $df = 2$, p-value $= 0.041$.

Since $0.041 < 0.05$, reject H_0. There is convincing evidence that the distribution of purchases differs from the claimed 50/30/20. Cookies contributed most (3.33): fewer cookies than claimed.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

CHECKING THE LARGE COUNTS CONDITION

A random sample of 44 students is classified as 30 walkers, 10 bikers and 4 riders. A school claims the shares are 60%, 30%, 10%. Can the chi-square test be used? If not, what can be done?

SOLUTION

Expected counts: $44(0.6) = 26.4$, $44(0.3) = 13.2$, $44(0.1) = 4.4$.

The last expected count 4.4 is less than 5, so the large counts condition fails.

Fix: combine the last two categories (bikers and riders: 14 observed, expected 17.6) or collect more data. With the combined categories, $\chi^2 = 0.491 + 0.736 = 1.227$, $df = 1$, p-value 0.268.

Combining changes the hypotheses: now the claim is about 60% versus 40%.

WHICH CATEGORY CONTRIBUTES MOST?

Absences at a school by weekday in a random sample of 150 absences are Mon 40, Tue 25, Wed 22, Thu 28, Fri 35. Test for equal absence rates and say which day contributes most.

SOLUTION

$E = 150/5 = 30$ per day. Contributions: Mon 3.33, Tue 0.83, Wed 2.13, Thu 0.13, Fri 0.83.

$\chi^2 = 7.27$, $df = 4$, p-value $= 0.122$.

Do not reject H_0 at $\alpha = 0.05$: there is not convincing evidence that absences are unevenly distributed across weekdays.

Monday contributes most (3.33), because 40 absences were observed versus 30 expected.

FULL FOUR-STEP TEST

A city's earlier survey said that 40% of students ride the bus, 30% walk, 20% are driven and 10% bike. In a new random sample of 100 students, 55 ride the bus, 20 walk, 15 are driven and 10 bike. Is there convincing evidence that the distribution has changed?

SOLUTION

State: H_0: the proportions are 0.40, 0.30, 0.20, 0.10. H_a: at least one differs. Use $\alpha = 0.05$.

Plan: chi-square goodness-of-fit test. Random sample; 100 is under 10% of all students; expected counts 40, 30, 20, 10 are at least 5.

Do: $\chi^2 = 5.625 + 3.333 + 1.250 + 0 = 10.21$, $df = 3$, p-value $= 0.017$.

Conclude: since $0.017 < 0.05$, reject H_0. There is convincing evidence that the distribution of how students get to school has changed. The largest contribution is from the bus category (5.63): more students ride the bus than expected.

Write It Like a 5

Model Response

Prompt: A music app claims that its four listening platforms are used equally. In a random sample of 120 users, 18 use phones, 30 use tablets, 32 use speakers and 40 use computers. Do the data provide convincing evidence that platform use is not equal? H_0: the four platforms are used equally often ($p = 0.25$ each). H_a: at least one proportion is different. Conditions: random sample of users; 120 is under 10% of all users; expected counts are $120(0.25) = 30$ in each platform, which are all at least 5. Test: chi-square goodness-of-fit. $\chi^2 = \frac{(18-30)^2}{30} + \frac{(30-30)^2}{30} + \frac{(32-30)^2}{30} + \frac{(40-30)^2}{30} = 4.80 + 0 + 0.13 + 3.33 = 8.27$, $df = 3$, p-value $= 0.041$. Conclusion: because $0.041 < 0.05$, we reject H_0. There is convincing evidence that the four platforms are not used equally. Phones (contribution 4.8) are used less than expected and computers more.

What the grader looks for

1 point: correct hypotheses in context. 1 point: names the test and checks conditions (random, 10%, expected counts at least 5). 1 point: correct expected counts. 1 point: correct test statistic and degrees of freedom. 1 point: correct p-value. 1 point: decision linked to the p-value and a conclusion in context.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Coach's Playbook

1. Write the claimed proportions in a table next to the observed counts.
2. Multiply each proportion by n to get expected counts. Make sure they add to n.
3. Check the expected counts are at least 5, then compute contributions.
4. Add to get χ^2; use $df = k - 1$ to find the right-tail p-value.
5. State the decision and, if you reject, name the category that contributes most.

- On the built-in calculator, enter observed counts in one list and expected counts in another, then run the chi-square goodness-of-fit test. Still show the formula setup on paper.
- Sum of expected counts equals sum of observed counts. If not, a proportion is wrong.
- Doubling every count while keeping the proportions doubles χ^2: larger samples detect smaller departures.
- A rejected H_0 says the claim fails somewhere; it does not say where. Use contributions to point to the biggest gaps.

Common traps

1. Using observed counts in the denominator instead of expected. 2. Using $df = n - 1$ (sample size) rather than $k - 1$ (number of categories). 3. Checking the large counts condition with observed counts. 4. Concluding that "the distribution matches the claim" after a large p-value. You can only say there is not enough evidence against it.

Mistake Clinic

Spot the error

A student tests a four-category claim with 80 observations and uses $df = 79$. Error: the degrees of freedom depend on the number of categories, not the sample size. Correct: $df = 4 - 1 = 3$.

Spot the error

A student computes $\sum(O - E)/E$ and gets 0 for every data set. Error: the differences $O - E$ add to zero. Correct: square the differences first, $\sum(O - E)^2/E$, so every term is non-negative.

Spot the error

With $\chi^2 = 3.30$, $df = 3$, p-value 0.348, a student concludes: "The spinner is fair." Error: failing to reject does not prove the null. Correct: there is no convincing evidence that the sectors are unequally likely, but the data cannot prove they are exactly equal.

Real-World Connection

Geneticists compare observed offspring counts with the ratios predicted by a model, and election analysts compare a poll's category counts with known population shares. Anytime a model says "these categories should occur in these proportions," the goodness-of-fit test checks whether the data are consistent with the model.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

In the commute example, the new sample has 100 students. Suppose a second sample of 200 students has exactly the same proportions (55%, 20%, 15%, 10%). Find the new χ^2 and explain why the evidence is stronger. Solution: all counts and expected counts double, so each contribution doubles and $\chi^2 = 20.42$ with $df = 3$ and p-value 0.0001. The same proportions give stronger evidence with a larger sample, since random variation shrinks as n grows.

Practice

PRACTICE **The Chi-Square Statistic and Goodness of Fit**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A goodness-of-fit test has 6 categories. The degrees of freedom are:
   (A) 4 **(B)** 5 **(C)** 6
   (D) The sample size minus 1
2. A random sample of 150 is tested against a claim that 30% fall in category 2. The expected count for category 2 is:
   (A) 30 **(B)** 150 **(C)** 50 **(D)** 45
3. Which describes the shape of a chi-square distribution?
   (A) Skewed to the right, starting at 0 **(B)** Symmetric and bell-shaped
   (C) Skewed to the left **(D)** Uniform
4. A category has observed count 28 and expected count 20. Its contribution to the chi-square statistic is:
   (A) 0.4 **(B)** 8 **(C)** 3.2 **(D)** 64
5. A random sample of 40 is tested against a claim with proportions 0.5, 0.4, 0.07, 0.03. Which condition fails?
   (A) Random **(B)** Independence (10%)
   (C) Expected counts at least 5 **(D)** The null hypothesis is stated
6. A goodness-of-fit test gives $\chi^2 = 10.21$ with $df = 3$. The p-value is closest to:
   (A) 0.017 **(B)** 0.17 **(C)** 0.98 **(D)** 0.35
7. A goodness-of-fit test gives p-value 0.041 and $\alpha = 0.05$. The best conclusion is:
   (A) The sample is not random
   (B) The claimed distribution is exactly correct
   (C) There is not enough evidence to reject the claim
   (D) There is convincing evidence the claimed distribution is not correct
8. Which is the correct null hypothesis for a test of whether four platforms are used equally?
   (A) $\chi^2 = 0$ **(B)** $p_1 = p_2 = p_3 = p_4 = 0.25$
   (C) At least one platform is used more **(D)** $\mu_1 = \mu_2 = \mu_3 = \mu_4$
9. If every observed count in a goodness-of-fit test is doubled, with the same claimed proportions, χ^2 will:
   (A) Double **(B)** Stay the same **(C)** Quadruple
   (D) Be cut in half
10. A six-color spinner is spun 120 times with counts 25, 17, 22, 15, 24, 17. Test at $\alpha = 0.05$ whether all six colors are equally likely. Show hypotheses, conditions, statistic, p-value and conclusion.
11. A survey of 80 randomly chosen students on favorite season gave counts Winter 22, Spring 18, Summer 10, Fall 30. A club claims preferences are 20%, 20%, 20%, 40%. (a) Compute the expected counts. (b) Do the conditions hold? (c) Compute χ^2 and the degrees of freedom.
12. A random sample of 90 students shows the number who studied in each of five locations: library 20, home 27, cafe 13, school hall 10, other 20. Test whether students use the five locations equally often at $\alpha = 0.05$.
13. In the study-location data above (observed 20, 27, 13, 10, 20 against 18 each), compute each category's contribution and say which locations differ most from the expected count. Explain why this does or does not explain the result of the test.
14. Explain why the chi-square statistic can never be negative and why the p-value is always the area to the right of the statistic, even though the alternative hypothesis says "at least one proportion differs."

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

15. A student says: "My p-value was 0.30, so the claimed distribution is correct." Explain the flaw in this statement and give a correct conclusion. Also state the critical value for $df = 4$ at $\alpha = 0.05$ and what it means.

9.2 Chi-Square Test for Homogeneity

Compare the distribution of a categorical variable across several populations or treatments using two-way tables, expected counts, df = (r-1)(c-1) and contributions.

Confirm With Your Teacher

College Board's revision page names only the goodness-of-fit test as removed for the 2027 exam, but the 2027 unit titles do not include a chi-square unit. Treat the chi-square test for homogeneity as extension material and confirm with your teacher whether it is part of your course.

Why This Matters

Are the commuting habits of ninth graders, tenth graders and eleventh graders the same? Does a new study technique change the pass rate more than two older ones? Both questions compare a **categorical distribution across several groups**. With two groups you could use a two-proportion z-test, but once you have three groups or more than two outcomes you need a tool that handles a whole table at once. That tool is the **chi-square test for homogeneity**.

Typical errors: computing expected counts from the wrong totals, using the number of cells instead of $(r-1)(c-1)$ as degrees of freedom, and concluding that all groups differ when the test only says that at least one differs. The skill connects to Unit 3's two-proportion procedure, of which it is the generalization.

Core Ideas

In a test of homogeneity, you take **separate random samples from two or more populations** (or randomly assign subjects to two or more treatments) and classify every individual by one categorical variable. The question: is the distribution of that variable the same in every population or treatment? Data are summarized in a two-way table with one row per group and one column per category.

Hypotheses for homogeneity

H_0: the distribution of the categorical variable is the same for all populations (or treatments).

H_a: the distribution is not the same for all of them: at least one population differs.

Never write H_0 as "the groups are equal" without naming the variable. Always say what is being compared in context.

Expected counts, df and conditions

Expected count for a cell: $E = \dfrac{(\text{row total})(\text{column total})}{\text{table total}}$. It is what you would see if the distributions were identical.

$\chi^2 = \sum \dfrac{(O-E)^2}{E}$ over every cell, with $df = (r-1)(c-1)$ for r rows and c columns.

Conditions: (1) separate random samples or a randomized experiment; (2) independence within and between groups, with each sample under 10% of its population; (3) all expected counts at least 5.

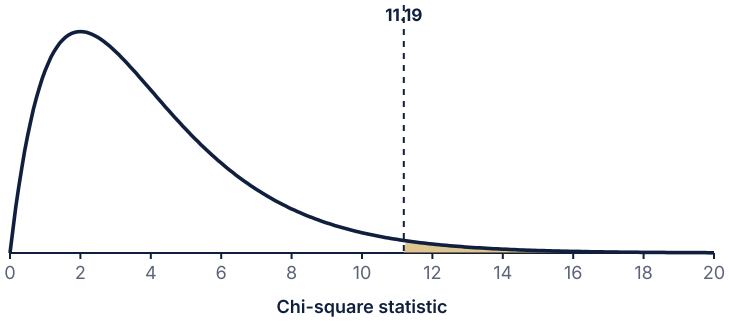
The p-value is the right-tail area $P(\chi^2 \geq \text{observed})$.

Where does the expected-count formula come from? If the distribution of the variable is the same in each group, then in every row the column proportions equal the overall column proportions. The expected count in a cell is that overall column proportion times the size of the group: $\dfrac{\text{column total}}{\text{table total}} \times \text{row total}$, which is the formula above.

Table size	Rows	Columns	df
2 × 2	2	2	1
2 × 3	2	3	2
3 × 3	3	3	4
4 × 3	4	3	6

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Chi-square with 4 degrees of freedom. The area beyond 11.19 is about 0.0245.

A rejected test of homogeneity points to the cells with the largest contributions $(O-E)^2/E$. Comparing the conditional percentages (for example the share of each group that chooses each category) tells you the direction. A bar chart of those percentages is often the clearest evidence in your explanation.

Worked Examples

EXPECTED COUNTS AND DF

Random samples of 100 students from each of grades 9, 10 and 11 are classified by how they travel to school: bus, walk or car. The counts are in the table. Find the expected count for each cell, state the hypotheses and give the degrees of freedom.

Grade	Bus	Walk	Car	Total
9	45	30	25	100
10	38	34	28	100
11	25	33	42	100
Total	108	97	95	300

SOLUTION

Column totals: bus 108, walk 97, car 95; each row total 100; table total 300.

Expected for bus in every row: $100(108)/300 = 36.0$. Walk: $100(97)/300 = 32.33$. Car: $100(95)/300 = 31.67$.

H_0: the distribution of travel mode is the same for grades 9, 10 and 11. H_a: the distribution of travel mode is not the same for all three grades.

$df = (3-1)(3-1) = 4$.

A 2 X 2 TABLE AND THE Z-TEST

In random samples of 50 students from each of two classes, 34 in Class X and 22 in Class Y improved their quiz score after using a study app. Test at $\alpha = 0.05$ whether the improvement rates differ.

SOLUTION

H_0: the distribution of outcome (improved or not) is the same in both classes. H_a: it is not.

Expected counts: improved $50(56)/100 = 28$ in each class; not improved 22 in each. All at least 5. Samples are random and under 10% of each class population.

$\chi^2 = \frac{(34-28)^2}{28} + \frac{(16-22)^2}{22} + \frac{(22-28)^2}{28} + \frac{(28-22)^2}{22} = 5.84$, $df = 1$, p-value $= 0.0156$.

Reject H_0: convincing evidence that the improvement rates differ. Class X improved more often (68% versus 44%).

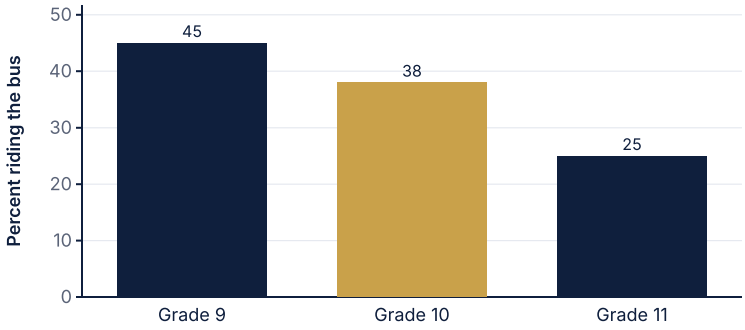
Check: the two-proportion z statistic is 2.417, and $2.417^2 = 5.84$. For a 2 × 2 table, the chi-square test and the two-sided two-proportion z-test agree.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

FULL TEST WITH A 3 X 3 TABLE

Use the grade/commute data from the first example to test at $\alpha = 0.05$ whether the distribution of travel mode is the same across grades.



Bus riders drop with grade level.

SOLUTION

Conditions: three separate random samples (each under 10% of its grade); all expected counts are at least 31.

Contributions: grade 9 (2.25, 0.17, 1.40), grade 10 (0.11, 0.09, 0.43), grade 11 (3.36, 0.01, 3.37).

$\chi^2 = 11.19$, $df = 4$, p-value $= 0.0245$.

Since $0.0245 < 0.05$, reject H_0. There is convincing evidence that the distribution of travel mode is not the same for all three grades. The largest contributions are bus and car in grade 11: eleventh graders ride the bus less (25%) and drive more (42%) than expected.

CHECKING CONDITIONS

Random samples of 23 and 38 students are classified as using 'bus', 'walk' or 'other' (counts: Group 1: 15, 5, 3; Group 2: 20, 12, 6). Can the test be used? What should be done?

SOLUTION

Column totals: 35, 17, 9; table total 61.

Expected counts: Group 1 is $23(9)/61 = 3.39$ for 'other', which is less than 5. The large counts condition fails.

Combine 'walk' and 'other' into one category: Group 1 has 15 and 8, Group 2 has 20 and 18. Now all expected counts are at least 9.8.

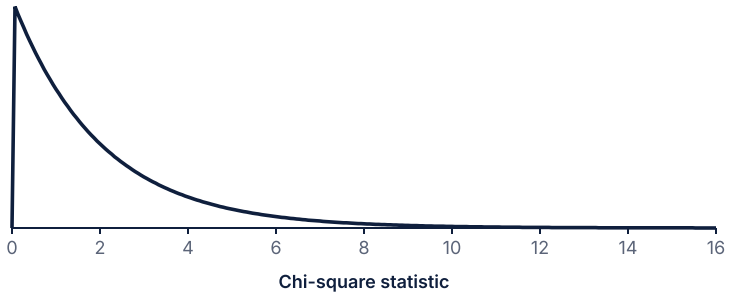
The new test gives $\chi^2 = 0.93$, $df = 1$, p-value $= 0.335$. There is no convincing evidence that the distributions differ.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

HOMOGENEITY IN AN EXPERIMENT

Of 150 volunteers, 50 were randomly assigned to each of three study techniques. Passing a unit quiz: technique A 18 passed and 32 did not; B 27 and 23; C 39 and 11. Is the pass rate the same for all three techniques?



The statistic 18.02 is far into the right tail for 2 degrees of freedom.

SOLUTION

H_0: the distribution of pass/fail is the same for all three techniques. H_a: it is not.

Conditions: random assignment; expected counts are $50(84)/150 = 28$ passing and 22 not passing in each row, all at least 5.

Contributions: A (3.57, 4.55), B (0.04, 0.05), C (4.32, 5.50).

$\chi^2 = 18.02$, $df = 2$, p-value $= 0.0001$. Reject H_0: very strong evidence that the pass rates are not all equal. Technique C had the highest pass rate (78%) and A the lowest (36%). Because treatments were randomly assigned, we can conclude the technique causes the difference.

INTERPRETING CALCULATOR OUTPUT

Two samples of 100 teens each chose their favorite music type: pop, rock or other. Output shows observed counts with expected counts in parentheses: City 1 has pop 30 (40), rock 45 (40), other 25 (20); City 2 has pop 50 (40), rock 35 (40), other 15 (20). $\chi^2 = 8.75$. Find df, the p-value, and describe the conclusion.

	Pop	Rock	Other
City 1	30 (40)	45 (40)	25 (20)
City 2	50 (40)	35 (40)	15 (20)

SOLUTION

Table is 2 × 3, so $df = (2-1)(3-1) = 2$.

P-value $= P(\chi^2_2 \geq 8.75) = 0.0126$.

All expected counts (40, 40, 20) are at least 5.

Since $0.0126 < 0.05$ we reject H_0: there is convincing evidence that the distribution of favorite music is not the same in the two cities. The cells contributing most are pop in both cities (2.5 each): City 2 prefers pop more than expected.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt: Random samples of 100 users of each of two apps were asked which feature they use most: messaging, photos or games. For App 1, the counts were 48, 32, 20; for App 2, 30, 40, 30. Is there convincing evidence that the distribution of most-used feature differs between the apps? Use $\alpha = 0.05$. H_0: the distribution of most-used feature is the same for the two apps. H_a: the distributions are not the same. Conditions: two separate random samples, each presumably under 10% of its app's users; expected counts are $100(78)/200 = 39$, 36, 25 in both rows, all at least 5. Test: chi-square test for homogeneity. $\chi^2 = \frac{(48-39)^2}{39} + \frac{(32-36)^2}{36} + \frac{(20-25)^2}{25} + \frac{(30-39)^2}{39} + \frac{(40-36)^2}{36} + \frac{(30-25)^2}{25} = 7.04$, $df = (2-1)(3-1) = 2$, p-value $= 0.030$. Conclusion: because $0.030 < 0.05$, we reject H_0. There is convincing evidence that the distribution of most-used feature is not the same for the two apps. Messaging is more common for App 1 (48% vs 30%).

What the grader looks for

1 point: states the hypotheses about the distribution of the variable across the groups. 1 point: names the test and checks both the sampling condition and expected counts. 1 point: correct expected counts. 1 point: correct statistic and degrees of freedom. 1 point: correct p-value. 1 point: decision and conclusion in context, naming a direction or the biggest contributions.

Coach's Playbook

1. Decide whether you have separate samples from each group (homogeneity). Name the grouping variable and the categorical response variable.
2. Build the observed table with row and column totals.
3. Compute expected counts with (row total)(column total)/(table total), and check they are all at least 5.
4. Compute χ^2, $df = (r-1)(c-1)$ and the right-tail p-value.
5. State the conclusion in context; point to the cells with the largest contributions and compare the conditional percentages.

- Enter the observed table as a matrix and run the calculator's chi-square test; the expected counts are stored in a second matrix.
- With two groups and two outcomes, the test matches a two-proportion z-test: $\chi^2 = z^2$.
- When groups have different sizes, compare percentages, not counts.
- A small p-value says the distributions differ somewhere; it doesn't say all groups differ.

Common traps

1. Using the sum of all counts instead of the row and column totals when finding expected counts. 2. Writing $df = r \cdot c$ or $r + c - 2$ instead of $(r-1)(c-1)$. 3. Concluding that "all three groups are different" from a significant result. 4. Treating an observational comparison as proof of cause. Only a randomized experiment supports causal wording.

Mistake Clinic

Spot the error

A student computes the expected count for the grade 9 bus cell in the commute table as 108/9 = 12. Error: dividing by the number of cells ignores the group size. Correct: $100(108)/300 = 36$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student writes H_0: $\mu_1 = \mu_2 = \mu_3$ for the travel-mode test. Error: the data are categorical, so there are no means. Correct: H_0: the distribution of travel mode is the same for grades 9, 10 and 11.

Spot the error

A student finds p-value 0.0245 and concludes "every grade has a different commuting pattern." Error: rejecting H_0 shows only that at least one distribution differs. Correct: "There is convincing evidence that the distribution is not the same for all three grades; grade 11 differs most from expected."

Real-World Connection

Health researchers compare outcome categories (improved, stayed the same, worsened) across several treatment groups; market analysts compare which product customers choose across regions. Whenever you compare how a categorical variable is distributed across separate groups, homogeneity is the test.

Challenge Corner

Two samples of sizes 50 and 70 have category counts (22, 18, 10) and (30, 28, 12). Compute the test and explain why the unequal sample sizes do not matter. Solution: column totals are 52, 46, 22; table total 120. Expected: row 1 $= 21.67, 19.17, 9.17$; row 2 $= 30.33, 26.83, 12.83$. $\chi^2 = 0.26$, $df = 2$, p-value $= 0.88$. The expected counts scale with the row totals, so unequal sample sizes are built into the calculation. No evidence the distributions differ: the percentages are nearly identical (44, 36, 20 versus 43, 40, 17).

Practice

PRACTICE **Chi-Square Test for Homogeneity**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A chi-square test uses a table with 4 rows and 3 columns. The degrees of freedom are:
   (A) 5 **(B)** 6 **(C)** 7 **(D)** 12
2. In a two-way table, the row total is 60, the column total is 45 and the table total is 150. The expected count for that cell is:
   (A) 15 **(B)** 45 **(C)** 30 **(D)** 18
3. Which scenario calls for a chi-square test of homogeneity?
   (A) Separate random samples from three schools, each classified by preferred lunch
   (B) One random sample of 200 adults classified by age group and favorite sport
   (C) One sample of 60 coin flips tested for fairness
   (D) Measuring the heights of two groups and comparing means
4. Which is the correct null hypothesis for a homogeneity test comparing the lunch preference across three schools?
   (A) Lunch preference and school are associated
   (B) The mean lunch preference is the same
   (C) The distribution of lunch preference is the same at all three schools
   (D) At least one school differs
5. A 2 × 3 table of counts has row totals 20 and 30, column totals 25, 15 and 10, and table total 50. Which statement is true?
   (A) All conditions are met
   (B) The degrees of freedom are 6
   (C) The test cannot be done because the table is not square
   (D) The large counts condition fails because an expected count is 4

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. A 2 × 2 homogeneity test gives $\chi^2 = 4.51$ with $df = 1$. The p-value is closest to:
   (A) 0.034 **(B)** 0.34 **(C)** 0.66 **(D)** 0.95
7. The two-proportion z-statistic for a 2 × 2 table is 2.12. The chi-square statistic for the same table is about:
   (A) 1.46 **(B)** 2.12 **(C)** 4.49 **(D)** 8.98
8. A homogeneity test gives p-value 0.0245 with $\alpha = 0.05$. The best conclusion is:
   (A) All populations have different distributions
   (B) The distribution of the variable is not the same in all populations
   (C) The populations have identical distributions
   (D) The sample sizes are too small
9. In a randomized experiment, 50 subjects are assigned to each of three treatments and the outcome is success or failure. The test of whether success rates are equal has degrees of freedom:
   (A) 1 **(B)** 49 **(C)** 3 **(D)** 2
10. Random samples of 100 students from each of two schools were asked whether they use public transit weekly: School 1 had 60 yes and 40 no; School 2 had 45 yes and 55 no. Test at $\alpha = 0.05$ whether the proportions differ using a chi-square test. Include hypotheses, conditions, expected counts, statistic, p-value and conclusion.
11. Random samples of 100 residents from each of two cities were classified by favorite music: pop, rock or other. City 1: 48, 32, 20; City 2: 30, 40, 30. (a) Find all expected counts. (b) Compute χ^2 and df. (c) Give the p-value and conclusion at $\alpha = 0.05$.
12. A random sample of 50 students from School X and 70 from School Y were classified by travel mode (bus, walk, other). School X: 22, 18, 10. School Y: 30, 28, 12. Test whether the distributions are the same at $\alpha = 0.05$.
13. Samples of 23 and 38 students are classified into three categories with counts (15, 5, 3) and (20, 12, 6). (a) Show that the large counts condition fails. (b) Combine categories and carry out the test.
14. In the three-technique experiment (A: 18 passed, 32 did not; B: 27, 23; C: 39, 11), compute the contributions to χ^2 and describe which cells drive the result. Can we conclude that the technique caused the difference? Explain.
15. Explain why the chi-square test for a 2 × 2 table gives the same conclusion as a two-sided two-proportion z-test, and give the relationship between the two statistics. What does the chi-square test add when there are more than two groups?

9.3 Chi-Square Test for Independence

Test whether two categorical variables in a single sample are associated using expected counts from independence, df = (r-1)(c-1) and contributions to chi-square.

Confirm With Your Teacher

College Board's revision page names only the goodness-of-fit test as removed for the 2027 exam, but the 2027 unit titles do not include a chi-square unit. Treat the chi-square test for independence as extension material and confirm with your teacher whether it is part of your course.

Why This Matters

Do students who get enough sleep also tend to have different screen-time habits? Is club membership related to grade point average? These are questions about whether two **categorical variables** in one population are related. Instead of comparing groups you chose in advance, you take a single sample, record two categorical variables for each individual, and ask whether knowing one variable tells you anything about the other. The **chi-square test for independence** (also called the test of association) answers that.

Students most often stumble over the wording: the null is that the variables are **independent**, not that they are "equal," and a significant result shows association, not cause. The calculation is identical to the homogeneity test, so the real skill is recognizing the study design and writing the right hypotheses. This links to the two-way table ideas of Unit 1 and to the independence rule $P(A \text{ and } B) = P(A)P(B)$ from the probability unit.

Core Ideas

In a two-way table of counts, rows give categories of one variable and columns give categories of the other. Two variables are **independent** if the distribution of one is the same no matter which category of the other you look at. Otherwise they are **associated**.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Hypotheses for independence

H_0: the two variables are independent in the population (there is no association between them).

H_a: the two variables are associated (not independent) in the population.

State both variables by name and context, such as "sleep category and screen-time group."

Procedure for the independence test

Design: one random sample from a single population, with two categorical variables recorded for each individual.

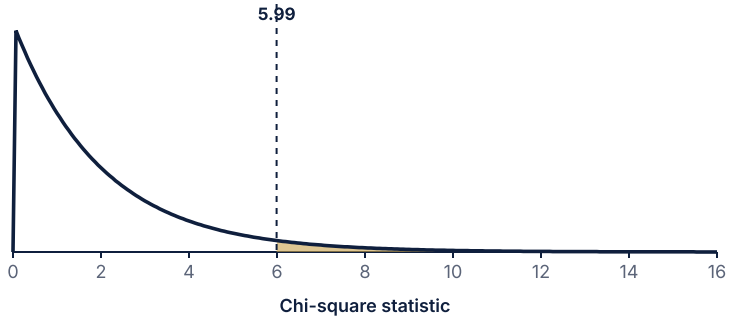
Expected count: $E = \dfrac{(\text{row total})(\text{column total})}{\text{table total}}$. It comes from independence: $P(\text{row and column}) = P(\text{row})P(\text{column})$, scaled by n.

Statistic: $\chi^2 = \sum \dfrac{(O-E)^2}{E}$, $df = (r-1)(c-1)$.

Conditions: random sample; sample under 10% of the population; all expected counts at least 5.

P-value: right tail of the chi-square distribution. Reject H_0 when the p-value is below α.

Why that expected-count formula? If sleep and screen time were independent, the probability that a student is in the "enough sleep" row **and** the "low screen" column would be the product of the two marginal probabilities. With 80 of 200 students in the first row and 70 of 200 in the first column, the expected count is $200 \times \dfrac{80}{200} \times \dfrac{70}{200} = 28$. The formula $\dfrac{80 \cdot 70}{200}$ does exactly this.



For 2 degrees of freedom, a statistic above 5.99 has a right-tail area below 0.05.

Independence Versus Homogeneity

The two tests use the same formulas but answer different questions because of how the data were collected.

	Homogeneity	Independence
How data are collected	Separate samples from each group (or treatments randomly assigned)	One sample, two variables measured on each individual
Null hypothesis	Same distribution of the variable in every group	The two variables are independent
Conclusion if rejected	The distributions are not all the same	The variables are associated
Calculation	(r-1)(c-1) df, same expected counts	(r-1)(c-1) df, same expected counts

A significant result tells you that the variables are associated, not that one causes the other. Unless the data come from a randomized experiment, lurking variables can create association. Describe the pattern with conditional percentages, such as the percent in each sleep category within each screen-time group.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXPECTED COUNTS FROM INDEPENDENCE

A random sample of 200 students is classified by sleep (enough or not enough) and daily screen time (low, medium, high). The table shows the counts. Find the expected counts and the degrees of freedom.

	Low screen	Medium	High	Total
Enough sleep	40	25	15	80
Not enough	30	45	45	120
Total	70	70	60	200

SOLUTION

Row totals: 80 and 120. Column totals: 70, 70, 60. Table total: 200.

Row 1: $80(70)/200 = 28$, 28, $80(60)/200 = 24$. Row 2: $120(70)/200 = 42$, 42, 36.

$df = (2-1)(3-1) = 2$. All expected counts are at least 5.

Interpretation: if sleep and screen time were independent, we would expect 28 students with enough sleep in the low-screen group; we observed 40.

A 2 X 2 TEST

In a random sample of 140 students, 70 play an instrument and 70 do not. Of those who play, 36 are on the honor roll; of those who do not play, 24 are on the honor roll. Test at $\alpha = 0.05$ whether instrument playing and honor roll status are independent.

SOLUTION

H_0: playing an instrument and honor roll status are independent for students at this school. H_a: they are associated.

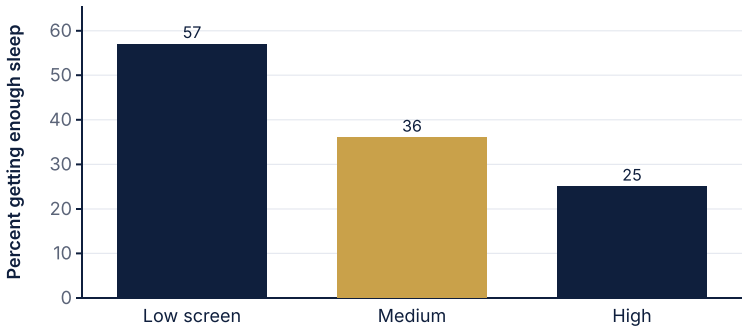
Conditions: random sample; 140 under 10% of students; expected counts are 30, 30, 40, 40, all at least 5.

$\chi^2 = \frac{(36-30)^2}{30} + \frac{(24-30)^2}{30} + \frac{(34-40)^2}{40} + \frac{(46-40)^2}{40} = 1.2 + 1.2 + 0.9 + 0.9 = 4.2$, $df = 1$, p-value $= 0.040$.

Since $0.040 < 0.05$, reject H_0. There is convincing evidence of an association: 51% of instrument players are on the honor roll versus 34% of non-players. The data are observational, so we cannot say playing an instrument causes the difference.

FULL TEST OF ASSOCIATION

Use the sleep and screen-time table to test at $\alpha = 0.05$ whether sleep and screen time are independent.



Percent with enough sleep in each screen-time group.

SOLUTION

H_0: sleep category and screen-time group are independent. H_a: they are associated.

Conditions: random sample of 200 students (under 10% of all students); all expected counts are at least 24.

Contributions: row 1 (5.14, 0.32, 3.38) and row 2 (3.43, 0.21, 2.25). $\chi^2 = 14.73$, $df = 2$, p-value $= 0.0006$.

Reject H_0: very strong evidence that sleep and screen time are associated. The largest contributions: low-screen students are more likely to get enough sleep (57%) than high-screen students (25%).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A 3 X 3 TABLE AND CONTRIBUTIONS

A random sample of 200 students is classified by age group (under 14, 14 to 15, 16 and over) and how they prefer to study (in a group, alone, online). Counts: under 14: 16, 24, 10; 14 to 15: 14, 36, 20; 16 and over: 10, 30, 40. Test for association and identify the biggest contributors.

SOLUTION

H_0: age group and study preference are independent. H_a: they are associated. $df = (3-1)(3-1) = 4$.

Expected counts (rows 50, 70, 80; columns 40, 90, 70): 10.0, 22.5, 17.5; 14.0, 31.5, 24.5; 16.0, 36.0, 28.0. All at least 5.

$\chi^2 = 3.60 + 0.10 + 3.21 + 0 + 0.64 + 0.83 + 2.25 + 1.00 + 5.14 = 16.78$, p-value $= 0.0021$.

Reject H_0 at $\alpha = 0.05$: convincing evidence that age group and study preference are associated. Biggest contributors: 16 and over studying online (5.14, observed 40 versus 28 expected) and under 14 studying in a group (3.60) and online (3.21, observed 10 versus 17.5).

WHEN THE CONDITIONS FAIL

A random sample of 50 students is classified by whether they ride the bus (yes or no) and whether they own a bike (yes or no). Counts: bus and bike 4, bus and no bike 14, no bus and bike 2, no bus and no bike 30. May the chi-square test be used?

SOLUTION

Row totals: 18 (bus), 32 (no bus). Column totals: 6 (bike), 44 (no bike).

Expected count for bus and bike: $18(6)/50 = 2.16$; for no bus and bike: $32(6)/50 = 3.84$. Both are less than 5.

The large counts condition is not met, so the p-value from the chi-square distribution would not be trustworthy (the calculated $\chi^2 = 2.78$ would give 0.095).

Options: collect a larger sample, or use a different method such as an exact test or a simulation. Do not combine 'bike' with 'no bike': with only two categories there is nothing to combine.

FILLING IN THE TABLE AND INTERPRETING

Out of 200 randomly chosen adults, 120 say they use a library and 80 do not; also, 100 are under 40 and 100 are 40 or older. Of the adults under 40, 75 use a library. Complete the table, test for association, and interpret.

SOLUTION

Under 40 and library: 75, so under 40 not library: $100 - 75 = 25$. For 40 and older: library $120 - 75 = 45$, not library $80 - 25 = 55$.

Expected counts: $120(100)/200 = 60$ and 40 in each column.

$\chi^2 = \frac{15^2}{60} + \frac{15^2}{60} + \frac{15^2}{40} + \frac{15^2}{40} = 18.75$, $df = 1$, p-value $= 0.00001$.

Very strong evidence that age group and library use are associated: 75% of those under 40 use a library versus 45% of those 40 and older. This is an observational study, so it does not show that age causes library use.

Write It Like a 5

Model Response

Prompt: A random sample of 140 students was asked about club membership and grade level (A, B or C) in their favorite class. Club members: 28 A, 12 B, 10 C. Non-members: 22 A, 38 B, 30 C. Is there convincing evidence of an association between club membership and grade in the class? Use $\alpha = 0.05$. H_0: club membership and grade in the class are independent for students at this school. H_a: club membership and grade are associated. Conditions: random sample; 140 is under 10% of all students at the school; expected counts: row totals 50 and 90, column totals 50, 50, 40; so $E = 17.86, 17.86, 14.29$ in row 1 and $32.14, 32.14, 25.71$ in row 2. All are at least 5. Test: chi-square test for independence. $\chi^2 = 5.76 + 1.92 + 1.29 + 3.20 + 1.07 + 0.71 = 13.95$, $df = (2-1)(3-1) = 2$, p-value $= 0.0009$. Conclusion: Since $0.0009 < 0.05$, reject H_0. There is convincing evidence of an association between club membership and class grade. Club members earn A's more often (56% versus 24% of non-members).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

What the grader looks for

1 point: states hypotheses about independence/association with both variables named. 1 point: names the test and verifies the random sample, 10% and expected counts conditions. 1 point: expected counts or the formula. 1 point: correct statistic and degrees of freedom. 1 point: correct p-value. 1 point: conclusion in context linking the p-value to the decision, noting association, and describing the pattern.

Coach's Playbook

1. Ask: was this one sample with two variables (independence) or separate samples (homogeneity)?
2. Name the two variables. Write H_0 as "independent" and H_a as "associated."
3. Total the rows and columns. Compute expected counts and check they are at least 5.
4. Compute χ^2, $df = (r-1)(c-1)$ and the p-value on the calculator's chi-square test (matrix input).
5. Conclude in context, describe the pattern with conditional percentages, and add the correlation-is-not-causation note when appropriate.

- The table of contributions $(O-E)^2/E$ shows which cells deviate most from independence.
- In a table with exactly independent variables, every row is a multiple of every other row, so $O = E$ and $\chi^2 = 0$.
- A big sample can make even a tiny association significant. Look at the percentages to judge whether the pattern matters in practice.
- Association does not mean cause. Only a randomized experiment supports cause-and-effect conclusions.

Common traps

1. Writing the null as "the proportions are equal" without saying that the variables are independent. 2. Using row percentages in one place and column percentages in another when describing the pattern. 3. Rejecting H_0 and claiming one variable causes the other. 4. Running the test on percentages or proportions instead of counts.

Mistake Clinic

Spot the error

A student enters the percentages (51%, 34%) rather than counts into the test and gets χ^2 very small. Error: chi-square needs counts, because the statistic depends on sample size. Correct: enter the counts 36, 24, 34, 46.

Spot the error

A student writes H_a: club membership causes better grades. Error: the test of independence shows association only, and the data were from an observational sample. Correct: H_a: club membership and grade are associated.

Spot the error

A student with a 3 × 4 table uses $df = 12$. Error: that counts the cells. Correct: $df = (3-1)(4-1) = 6$.

Real-World Connection

Public health researchers study whether smoking status and a disease category are associated using a sample of patients, and school districts check whether attendance category and test band are associated. The chi-square test of independence is the standard first look at two categorical variables in a single sample.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Challenge Corner

Explain why a 3 × 3 table with exactly independent variables has $\chi^2 = 0$ and construct an example. Solution: if the variables are independent, then $O = \frac{(\text{row total})(\text{col total})}{n}$ in every cell. For example, rows 20, 30 and 50 and columns 40, 40 and 20 with $n = 100$ give observed counts 8, 8, 4; 12, 12, 6; 20, 20, 10. Every row has the same column proportions (0.4, 0.4, 0.2), so $\chi^2 = 0$ and $df = 4$.

Practice

PRACTICE

Chi-Square Test for Independence

Show your work. A calculator is allowed. Justify conclusions in context.

1. A researcher records the favorite sport and the grade level of each student in one random sample of 300 students. Which test is appropriate to see whether the two variables are related?
   (A) Chi-square goodness-of-fit
   (B) Chi-square test for independence
   (C) Chi-square test for homogeneity
   (D) Two-sample t test
2. Which is the correct null hypothesis for a chi-square test of independence on sleep category and screen-time group?
   (A) The mean screen time is the same
   (B) The variables are associated
   (C) The variables are independent
   (D) The distributions of screen time are different
3. A two-way table has 3 rows and 4 columns. The degrees of freedom are:
   (A) 6 (B) 7 (C) 11 (D) 12
4. In a sample of 200, a cell is in a row with total 80 and a column with total 70. Its expected count under independence is:
   (A) 14 (B) 75 (C) 35 (D) 28
5. A test of independence gives $\chi^2 = 14.73$ with $df = 2$. The p-value is closest to:
   (A) 0.06 (B) 0.0006 (C) 0.25 (D) 0.94
6. A test of independence yields $\chi^2 = 4.2$, $df = 1$, p-value 0.040. At $\alpha = 0.05$, the best conclusion is:
   (A) Do not reject H_0; the variables are independent
   (B) Do not reject H_0; the sample is too small
   (C) Reject H_0; one variable causes the other
   (D) Reject H_0; there is convincing evidence of an association
7. Which table (rows are groups, columns are outcomes) shows exactly independent variables?
   (A) Row 1: 10, 20; Row 2: 20, 20
   (B) Row 1: 15, 25; Row 2: 35, 25
   (C) Row 1: 20, 30; Row 2: 40, 60
   (D) Row 1: 5, 10; Row 2: 15, 15
8. A significant chi-square test of independence from a random sample (not an experiment) allows which conclusion?
   (A) There is an association between the variables
   (B) The explanatory variable causes the response
   (C) The variables are perfectly related
   (D) The null hypothesis is true
9. Which of these is NOT a condition for the chi-square test of independence?
   (A) Random sample
   (B) The sample is less than 10% of the population
   (C) All expected counts are at least 5
   (D) The population is approximately normal
10. A random sample of 200 adults is classified by age (under 40, 40 or older) and smart speaker use (yes, no). Counts: under 40: 60 yes and 40 no; 40 or older: 52 yes and 48 no. Test at $\alpha = 0.05$ whether age group and smart speaker use are independent.
11. A random sample of 140 students is classified by club membership (member or not) and grade in a favorite class (A, B, C). Club members: 28, 12, 10. Non-members: 22, 38, 30. (a) Find the expected counts. (b) Compute χ^2 and df. (c) Find the p-value and conclude at $\alpha = 0.05$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

12. A random sample of 320 residents is classified by region (north, central, south) and favorite transport (bike, bus, car): North 30, 50, 20; Central 40, 30, 30; South 20, 40, 60. Find the degrees of freedom, expected counts and the test conclusion at $\alpha = 0.05$.
13. A random sample of 50 students: 4 ride the bus and own a bike, 14 ride the bus and own no bike, 2 do not ride and own a bike, 30 neither. (a) Find the expected counts. (b) Explain why the chi-square test is not appropriate. (c) Suggest what the researchers could do.
14. Out of 200 randomly chosen adults, 120 use a library and 80 do not; 100 are under 40 and 100 are 40 or older. Of those under 40, 75 use a library. Complete the table, find χ^2 and p-value, and interpret the result. Why can't we say that age causes library use?
15. Explain the difference between a chi-square test of homogeneity and a chi-square test of independence in terms of data collection, hypotheses and conclusion. Why are the calculations identical? Give an example of each.

9.4 Choosing and Interpreting Chi-Square Tests

Decide among goodness of fit, homogeneity and independence from the study design, then interpret results with contributions, residuals, percentages and cautions about cause.

Why This Matters

You now know three chi-square procedures, and on a test or in real work the hardest part is not the arithmetic. It is deciding **which test fits the study**. All three compute the same statistic from observed and expected counts, so a student who cannot tell them apart will write the wrong hypotheses and the wrong conclusion, even with perfect calculations. The decision is made from how the data were collected and what question is being asked.

Beyond choosing, you must interpret: a rejected null says the model fails or the variables are associated, but not why, and not by how much. This lesson draws on the categorical-data ideas from the exam's Unit 3 and on the study-design ideas from Unit 1 (random sampling versus random assignment, association versus cause). Remember to confirm with your teacher which chi-square tests are part of your course, since College Board's revision page lists goodness of fit as removed from the 2027 exam.

Core Ideas

Ask three questions in order. First, how many categorical variables are measured, and how many groups were sampled? Second, is there a claimed distribution to compare against? Third, are the groups chosen separately, or does one sample give two variables?

Choosing the correct chi-square test

Goodness of fit: ONE categorical variable, ONE sample, compared with a claimed set of proportions. $df = k - 1$. Expected counts come from the claim: $E_i = np_i$.

Homogeneity: ONE categorical variable measured in TWO OR MORE separate samples (or treatment groups). H_0: same distribution in every group. $df = (r-1)(c-1)$.

Independence: ONE sample, TWO categorical variables on each individual. H_0: the variables are independent. $df = (r-1)(c-1)$.

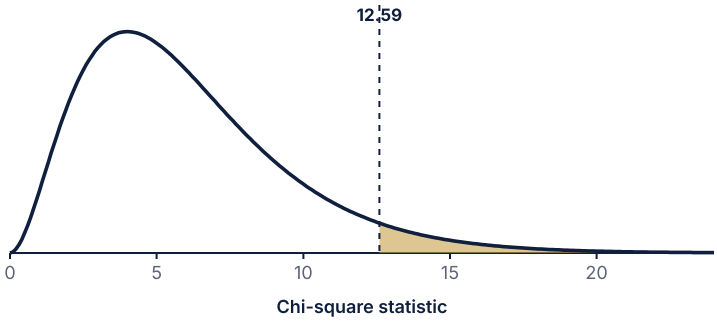
For a 2 × 2 table, the chi-square test is equivalent to a two-sided two-proportion z-test: $\chi^2 = z^2$.

Test	Data	Claim or question	df
Goodness of fit	1 sample, 1 variable	Does the distribution match the claimed proportions?	k - 1
Homogeneity	Several samples, 1 variable	Is the distribution the same in every group?	(r-1)(c-1)
Independence	1 sample, 2 variables	Are the variables associated?	(r-1)(c-1)

Notice that a homogeneity table and an independence table can look exactly the same. What differs is the **sampling plan**, and with it the wording of the hypotheses and the conclusion. If researchers fixed the group sizes (100 from each school), it is homogeneity; if they took 200 students and then recorded school and travel mode afterwards, it is independence.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



For 6 degrees of freedom the 5% critical value is 12.59. Reject H0 when the statistic exceeds it.

Interpreting a chi-square result

A small p-value is evidence against H_0, not a measure of the size of the effect. Describe the pattern with percentages.

Find the cells with the largest contributions $(O-E)^2/E$. A **Pearson residual** $(O-E)/\sqrt{E}$ also shows direction: positive means more than expected, negative means fewer. Residuals beyond about 2 in size are notable.

Association is not cause. Only a randomized experiment supports a causal conclusion.

A large p-value means the data are consistent with H_0; it never proves H_0 true.

Chi-square tests answer a yes/no question about whether something differs. They do not estimate a difference. When you need an estimate of how large a difference is, use a confidence interval, as in the two-proportion z-interval.

Worked Examples

CLASSIFY THE STUDY

Name the correct chi-square test for each scenario. (a) A bag of candy claims 30% red, 25% blue, 20% green, 15% yellow, 10% orange, and a random sample of 200 candies is counted. (b) Random samples of 100 students from each of four schools are asked about their favorite season. (c) One random sample of 250 adults is classified by education level and preferred news source. (d) 90 volunteers are randomly assigned to one of three diets, and weight-loss success (yes or no) is recorded.

SOLUTION

(a) Goodness of fit: one variable, one sample, a claimed distribution; $df = 5 - 1 = 4$.

(b) Homogeneity: one variable, four separate samples; $df = (4-1)(4-1) = 9$.

(c) Independence: one sample, two variables; if education has 4 levels and news source 3, $df = (4-1)(3-1) = 6$.

(d) Homogeneity: treatments are the groups, and the outcome is one categorical variable; $df = (3-1)(2-1) = 2$.

SAME TABLE, TWO DESIGNS

Counts of preferred event (concert, sports, dance) for random samples of 100 students from grade 9 and 100 from grade 10: grade 9: 35, 45, 20; grade 10: 25, 50, 25. Test for a difference, at $\alpha = 0.05$. How would the hypotheses read if the 200 students were a single random sample classified by grade and event?

SOLUTION

Design: separate samples of fixed size, so homogeneity. H_0: the distribution of preferred event is the same in grades 9 and 10. H_a: it is not.

Expected counts: $100(60)/200 = 30$, 47.5, 22.5 in each row, all at least 5.

$\chi^2 = 0.83 + 0.13 + 0.28 + 0.83 + 0.13 + 0.28 = 2.49$, $df = 2$, p-value $= 0.289$. Do not reject H_0: there is no convincing evidence that the preferred-event distribution differs between the grades.

If one sample of 200 had been classified by both variables, the test (same numbers) would be a test of independence: H_0: grade and preferred event are independent; H_a: they are associated. The conclusion would be: no convincing evidence of an association.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

GOODNESS OF FIT WITH A CONCLUSION

A pet store owner believes dog sizes brought in are 10% toy, 30% small, 40% medium and 20% large. A random sample of 100 dogs shows 12, 28, 44, 16. Test the belief at $\alpha = 0.05$.

SOLUTION

H_0: the sizes are distributed 10%, 30%, 40%, 20%. H_a: at least one differs.

Expected: 10, 30, 40, 20, all at least 5; random sample, under 10% of dogs.

$\chi^2 = \frac{4}{10} + \frac{4}{30} + \frac{16}{40} + \frac{16}{20} = 0.40 + 0.13 + 0.40 + 0.80 = 1.73$, $df = 3$, p-value $= 0.63$.

Do not reject H_0: the data are consistent with the owner's belief. This does not prove the belief; it says the sample gives no convincing evidence against it.

PEARSON RESIDUALS

In the study-preference table (age group by study mode), the expected count for 'age 16 and over, online' is 28 and the observed count is 40. Others: under 14, group: observed 16, expected 10; under 14, online: 10 observed, 17.5 expected. Find the Pearson residuals and interpret them.

SOLUTION

16 and over, online: $(40 - 28)/\sqrt{28} = 2.27$.

Under 14, group: $(16 - 10)/\sqrt{10} = 1.90$. Under 14, online: $(10 - 17.5)/\sqrt{17.5} = -1.79$.

The 16-and-over online cell is above 2: noticeably more of the oldest students prefer online study than independence predicts. The younger students are more group-oriented and less online than expected.

Check: the square of the residual is the contribution: $2.27^2 = 5.14$, which matches the largest contribution in the table.

CHI-SQUARE OR Z-TEST?

Two random samples of 100 each: Group A, 48 successes; Group B, 63 successes. (a) Test whether the success rates differ. (b) A researcher wants to show Group B has the higher rate. Which test is appropriate and what is the p-value?

SOLUTION

(a) 2 × 2 homogeneity test: expected 55.5 and 44.5 in each row; $\chi^2 = 4.56$, $df = 1$, p-value $= 0.033$. Reject H_0 at 0.05: convincing evidence that the success rates differ.

(b) A chi-square test only supports the two-sided alternative "the proportions differ." For the one-sided alternative $p_B > p_A$, use a two-proportion z-test: $z = 2.13$, one-sided p-value $= 0.016$ (half of 0.033).

Conclusion: convincing evidence that Group B has the higher success rate. Note that $z^2 = 4.56$ equals the chi-square value.

HANDLING A SMALL EXPECTED COUNT

A survey of 60 randomly selected teens has five categories of preferred snack, and one expected count is 3.2. The researcher says to run the test anyway and report p = 0.04. What should be said?

SOLUTION

The large counts condition requires all expected counts at least 5. An expected count of 3.2 means the chi-square distribution may not approximate the sampling distribution of the statistic well, so the p-value is unreliable.

Remedies: combine a sparse category with a logically similar one (changing the question slightly) or collect a larger sample.

If neither is possible, the researcher should report the analysis as unreliable or use an exact or simulation-based method rather than claim p = 0.04 as reliable evidence.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt: A school wants to know whether lunch choice (pizza, salad, sandwich) differs among grade levels. Random samples of 80 students from each of grades 9, 10 and 11 gave counts: grade 9: 30, 20, 30; grade 10: 25, 25, 30; grade 11: 15, 30, 35. Choose the appropriate test, carry it out at $\alpha = 0.05$ and state a conclusion. Test: because there are three separate random samples and one categorical variable, I use a chi-square test for homogeneity. H_0: the distribution of lunch choice is the same for grades 9, 10 and 11. H_a: the distribution of lunch choice is not the same for all three grades. Conditions: three random samples, each under 10% of its grade; expected counts are $80(70)/240 = 23.3$, $80(75)/240 = 25$, $80(95)/240 = 31.7$ in each row, all at least 5. Calculation: $\chi^2 = 7.53$, $df = (3-1)(3-1) = 4$, p-value $= 0.111$. Conclusion: since $0.111 > 0.05$, we do not reject H_0. There is not convincing evidence that the distribution of lunch choice differs among grades 9, 10 and 11.

What the grader looks for

1 point: identifies the correct test (homogeneity) with a reason based on the sampling design. 1 point: correct hypotheses in context. 1 point: verifies random samples, 10% and expected counts at least 5. 1 point: correct statistic and df. 1 point: correct p-value. 1 point: conclusion linked to the p-value, in context, without claiming that the null is true.

Coach's Playbook

1. Count the categorical variables and the number of samples.
2. Decide: claimed distribution (goodness of fit), several separate samples (homogeneity) or one sample with two variables (independence).
3. Write hypotheses in the language of that test and in context.
4. Check the three conditions with expected, not observed, counts.
5. Compute, then interpret with percentages and the largest contributions or residuals.

- Say the design in a short phrase: "three random samples, one variable" or "one sample, two variables." It locks in the correct test and the right wording.
- Use $df = k - 1$ for goodness of fit and $(r-1)(c-1)$ for tables.
- If the answer choices say "means," "slopes" or "differences of means," the data are quantitative and chi-square does not apply.
- A table of percentages is often easier to interpret than a table of counts; compute the conditional percentages.

Common traps

1. Calling a two-sample comparison of proportions with a one-sided alternative a chi-square test. Use the z-test. 2. Using homogeneity wording ("same distribution") when the data are one sample with two variables, or the reverse. 3. Checking the large counts condition with observed counts. 4. Saying "accept H_0" or "the variables are definitely independent" after a large p-value.

Mistake Clinic

Spot the error

Random samples of 100 from each of two schools are compared on favorite season. A student writes: H_0: season and school are independent. Error: separate samples from each group signal homogeneity. Correct: H_0: the distribution of favorite season is the same at both schools.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student says, "The p-value was 0.289, so grade and event are independent." Error: failing to reject does not prove independence. Correct: there is not convincing evidence of an association.

Spot the error

A student calls a one-sample table of 300 students' favorite color a test of homogeneity. Error: one variable and one sample compared with an expected distribution is a goodness-of-fit situation, if a claimed distribution exists. Correct: goodness of fit with $df = k - 1$.

Real-World Connection

Researchers, pollsters and quality analysts often start with counts in categories. Which test they use depends on the design: a claimed distribution (goodness of fit), a comparison of groups (homogeneity), or a search for relationships in a single sample (independence). Naming the design correctly is half the analysis.

Challenge Corner

A table of 300 students (rows: three class years; columns: three study spots) has $\chi^2 = 10.55$ with 100 per row. Determine df, the p-value, and explain how the conclusion wording changes between a homogeneity and an independence design. Solution: $df = (3-1)(3-1) = 4$ and the p-value is 0.032. At $\alpha = 0.05$ reject. Homogeneity: "the distribution of study spot is not the same for the three class years." Independence: "class year and study spot are associated." The numbers are identical; only the framing changes.

Practice

PRACTICE **Choosing and Interpreting Chi-Square Tests**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A candy maker claims a color distribution. A random sample of 300 candies is counted by color. Which test is appropriate?
   (A) Chi-square test for homogeneity **(B)** Chi-square goodness-of-fit test
   (C) Chi-square test for independence **(D)** Two-proportion z-test
2. Random samples of 100 students from each of four cities are classified by favorite season. Which test is appropriate?
   (A) Goodness of fit **(B)** Independence **(C)** Homogeneity
   (D) One-sample t test
3. One random sample of 250 adults is classified by major and preferred study location. Which statement is correct?
   (A) The test is for independence, and H_0 says major and location are independent
   (B) The test is for homogeneity, and H_0 says the means are equal
   (C) The test is goodness of fit with $df = 248$
   (D) No chi-square test applies
4. Which pair gives the degrees of freedom for (i) a goodness-of-fit test with 5 categories and (ii) an independence test with a 3 × 4 table?
   (A) 5 and 6 **(B)** 5 and 12 **(C)** 4 and 12 **(D)** 4 and 6
5. A researcher wants to show that the success rate in Group B is higher than in Group A using two random samples. The best choice is:
   (A) Two-proportion z-test with a one-sided alternative
   (B) Chi-square homogeneity test, because it is one-sided
   (C) Chi-square goodness-of-fit test
   (D) Paired t test

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. The Pearson residual for a cell with observed count 40 and expected count 28 is closest to:
   (A) 0.43 **(B)** 12 **(C)** 2.27 **(D)** 5.14
7. A chi-square test of independence gives p-value 0.289. The best statement is:
   (A) The variables are independent
   (B) The null hypothesis is false
   (C) The variables are weakly associated
   (D) There is not convincing evidence of an association
8. A chi-square test rejects H_0 with a small p-value on a large random sample. What can be concluded?
   (A) The effect is large
   (B) There is evidence against the null, but the practical size must be judged from percentages
   (C) The variables are causally related
   (D) The null hypothesis is certainly false
9. A table has an expected count of 2.8. The best action is:
   (A) Run the test anyway; chi-square is robust **(B)** Use observed counts in the denominators
   (C) Combine categories or collect more data **(D)** Use a z-test for means
10. Classify each as goodness of fit, homogeneity or independence and give the degrees of freedom: (a) 200 candies counted by color (6 colors) against a claimed distribution; (b) random samples of 80 students from each of 3 schools, classified by preferred sport (4 choices); (c) one random sample of 150 adults classified by age group (3 levels) and TV preference (4 types).
11. Random samples of 80 students from each of two schools were classified by preferred event: concert, sports or dance. School 1: 20, 35, 25. School 2: 30, 20, 30. Choose the correct test, check conditions, compute χ^2 and the p-value, and conclude at $\alpha = 0.05$.
12. In a random sample of 80 students, 40 were tutored and 40 were not. Of the tutored students, 26 passed and 14 did not; of the others, 14 passed and 26 did not. (a) Which test is appropriate and why? (b) Compute χ^2 and the p-value. (c) Show how the z-test relates and give a conclusion.
13. A spinner has four equal sectors. In 100 random spins the counts were 30, 22, 28, 20. Choose the test, state hypotheses, and carry it out at $\alpha = 0.05$.
14. The 3 × 3 table of age group by study mode has observed counts 16, 24, 10; 14, 36, 20; 10, 30, 40 and expected counts 10, 22.5, 17.5; 14, 31.5, 24.5; 16, 36, 28. Compute the Pearson residuals for the first row and the last row, and explain which cells show the clearest departures from independence.
15. A researcher compares success in two groups of 100: Group A 48 and Group B 63. She runs a chi-square test (p = 0.033), then claims that Group B has a higher rate and says the p-value for that claim is 0.033. Explain the mistake and give the correct p-value and conclusion.

T Unit 9 Test

Test your skills with the chi-square distribution, goodness of fit, homogeneity, independence, df, expected counts, contributions and choosing the correct test.

Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Part A: Multiple Choice

UNIT TEST **Unit 9 Test**

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. A goodness-of-fit test compares a sample with a claimed distribution that has 7 categories. The degrees of freedom are:
   (A) 5 **(B)** 6 **(C)** 7
   (D) The sample size minus 1
2. A random sample of 250 is tested against a claim that 18% fall in a category. The expected count for that category is:
   (A) 18 **(B)** 232 **(C)** 55 **(D)** 45
3. Which statement about the chi-square distribution is true?
   (A) It is skewed to the right with mean equal to its degrees of freedom
   (B) It is symmetric about 0
   (C) It is skewed to the left
   (D) It can take negative values

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

4. A category has observed count 14 and expected count 20. Its contribution to the chi-square statistic is:
   (A) 0.3 **(B)** 6 **(C)** 1.8 **(D)** 36
5. Random samples of 90 patients from each of three hospitals are classified by satisfaction rating (low, medium, high). Which test is appropriate?
   (A) Goodness of fit **(B)** Independence **(C)** Homogeneity
   (D) Two-sample t test
6. A chi-square test is run on a two-way table with 4 rows and 5 columns. The degrees of freedom are:
   (A) 12 **(B)** 9 **(C)** 16 **(D)** 20
7. In a sample of 300, a cell is in a row with total 90 and a column with total 120. Its expected count is:
   (A) 30 **(B)** 110 **(C)** 75 **(D)** 36
8. Expected counts in a 2 × 3 table are 12.4, 10.1, 4.5, 17.6, 14.9, 6.5. Which statement is correct?
   (A) All conditions on expected counts are met **(B)** The test must be homogeneity
   (C) The large counts condition fails **(D)** The degrees of freedom are 6
9. A chi-square test gives $\chi^2 = 9.2$ with $df = 4$ (p-value 0.056). At $\alpha = 0.05$, the best conclusion is:
   (A) Reject H_0 because 9.2 is large
   (B) Do not reject H_0; there is not convincing evidence against it
   (C) The null hypothesis is true
   (D) The sample is not random
10. A random sample of students is classified by sleep and screen-time group. A test of independence is significant. The best conclusion is:
    (A) More screen time causes less sleep
    (B) Every screen-time group has a different sleep distribution
    (C) Sleep and screen time are independent
    (D) Sleep and screen time are associated
11. A 2 × 2 chi-square test statistic is 6.25. The equivalent two-proportion z statistic has absolute value:
    (A) 2.5 **(B)** 3.13 **(C)** 6.25 **(D)** 39.1
12. The critical value of the chi-square distribution with $df = 1$ at the 5% level is 3.84. Which normal critical value does this correspond to?
    (A) 1.28 **(B)** 1.645 **(C)** 1.96 **(D)** 2.576
13. [6 points] A shop claims that customers' favorite flavors are 25% vanilla, 35% chocolate, 25% strawberry and 15% mint. In a random sample of 120 customers, 42 chose vanilla, 34 chocolate, 22 strawberry and 22 mint. (a) [2 points] State the hypotheses and the name of the test. (b) [1 point] Find the expected counts and verify the conditions. (c) [2 points] Compute the statistic, degrees of freedom and p-value. (d) [1 point] State the conclusion at $\alpha = 0.05$.
14. [10 points] A researcher takes independent random samples of 100 shoppers at each of three stores and records the payment method (cash, card, mobile). Store A: 45, 35, 20. Store B: 30, 40, 30. Store C: 25, 35, 40. (a) [2 points] Name the test and state the hypotheses. (b) [2 points] Show the expected count for the Store A cash cell, check the conditions, and state the degrees of freedom. (c) [3 points] The calculator gives $\chi^2 = 13.62$. Find the p-value and state a conclusion in context. (d) [2 points] Identify the two cells with the largest contributions (given: Store A cash 4.08, Store A mobile 3.33, Store C mobile 3.33, Store C cash 2.08) and describe the pattern. (e) [1 point] Explain why the researcher cannot conclude that the store causes the payment method.

Answers — Inference for Categorical Data: Chi-Square: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

9.1 The Chi-Square Statistic and Goodness of Fit

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

PRACTICE · THE CHI-SQUARE STATISTIC AND GOODNESS OF FIT

1. B. $df = k - 1 = 6 - 1 = 5$. Choice A subtracts one extra; the sample size (D) is not used for goodness of fit.
2. D. Expected count $= n \cdot p = 150(0.30) = 45$. Choice A treats 30% as a count; C divides by 3 (a common slip).
3. A. Chi-square curves start at 0, have a long right tail and get less skewed as df increases. A symmetric bell (B) describes a normal curve.
4. C. $(28 - 20)^2/20 = 64/20 = 3.2$. Choice A divides the raw difference by 20 without squaring; D (64) forgets to divide by the expected count.
5. C. Expected counts are $20, 16, 2.8, 1.2$. The last two are less than 5. The data being random and independent are not the issue here.
6. A. The calculator gives χ^2cdf(10.21, 1E99, 3) = 0.017. The p-value is a right tail; choosing 0.98 (C) mistakes it for the left tail.
7. D. The p-value is below α, so we reject H_0 and have convincing evidence against the claim. Choice B claims too much; C reverses the decision.
8. B. The null states each of the four proportions is 0.25. A (a statistic equals zero) and C (an alternative) are not hypotheses about proportions, and D is about means.
9. A. Each $(O - E)^2/E$ doubles when counts double: $(2d)^2/(2E) = 2d^2/E$. So χ^2 doubles. Quadrupling (C) forgets that E doubles too.
10. H_0: all six colors have $p = 1/6$. H_a: at least one differs. Conditions: random spins; $E = 120/6 = 20$ each, all at least 5. $\chi^2 = (25 - 20)^2/20 + (17 - 20)^2/20 + (22 - 20)^2/20 + (15 - 20)^2/20 + (24 - 20)^2/20 + (17 - 20)^2/20 = 1.25 + 0.45 + 0.20 + 1.25 + 0.80 + 0.45 = 4.40$, $df = 5$, p-value $= 0.493$. Since $0.493 > 0.05$, we do not reject H_0. There is not convincing evidence that the six colors are unequally likely.
11. (a) $E = 80(0.2) = 16$, 16, 16, $80(0.4) = 32$ for Winter, Spring, Summer, Fall.

    (b) The sample is random, 80 is under 10% of students, and all expected counts are at least 5. The conditions hold.

    (c) $\chi^2 = \frac{(22 - 16)^2}{16} + \frac{(18 - 16)^2}{16} + \frac{(10 - 16)^2}{16} + \frac{(30 - 32)^2}{32} = 2.25 + 0.25 + 2.25 + 0.125 = 4.875$, $df = 3$. (p-value ≈ 0.181.)
12. H_0: the five locations are used equally ($p = 0.2$). H_a: at least one differs. Random sample, 90 under 10% of students, $E = 18$ each (all at least 5). $\chi^2 = \frac{4 + 81 + 25 + 64 + 4}{18}$ with differences 2, 9, -5, -8, 2: $= 0.222 + 4.5 + 1.389 + 3.556 + 0.222 = 9.89$, $df = 4$, p-value $= 0.042$. Since $0.042 < 0.05$, reject H_0: convincing evidence that the five locations are not equally popular.
13. Contributions: library 0.22, home 4.50, cafe 1.39, school hall 3.56, other 0.22. Home (27 versus 18 expected) and school hall (10 versus 18) contribute most, together 8.06 of the 9.89. So the rejection is driven mainly by too many students at home and too few in the school hall. The test itself only says that at least one proportion differs; the contributions tell us where to look.
14. χ^2 sums squared differences divided by positive expected counts, so each term is at least 0, and the total is non-negative. Only large positive values of χ^2 indicate disagreement between data and claim, whichever category is off and in whichever direction. Because bigger χ^2 means stronger evidence, the p-value is the probability of getting a value at least as large, the right tail. The alternative "at least one differs" is not a one-sided or two-sided comparison of a single parameter.
15. A large p-value means the data are consistent with the claim; it does not prove the claim. Correct conclusion: with p = 0.30, there is not convincing evidence against the claimed distribution, so we cannot reject it. The critical value for $df = 4$ at $\alpha = 0.05$ is 9.49: a statistic greater than 9.49 would give a p-value less than 0.05 and lead us to reject H_0.

9.2 Chi-Square Test for Homogeneity

PRACTICE · CHI-SQUARE TEST FOR HOMOGENEITY

1. B. $df = (r - 1)(c - 1) = (4 - 1)(3 - 1) = 6$. Choice D multiplies $r \cdot c$, which counts the cells; choice A uses $r + c - 2$, the wrong formula.
2. D. $E = (60)(45)/150 = 18$. Choice A divides 150 by 10, and C is half the row total; neither uses all three totals.
3. A. Homogeneity needs separate samples from each population classified by one variable. B is a single sample classified by two variables (independence); C is goodness of fit; D involves means.
4. C. Homogeneity states that the distribution is the same across the populations. A is the independence null; B is about means; D is the alternative.
5. D. The expected count in the first row, third column is $20(10)/50 = 4$, which is less than 5. Whether the table is square is irrelevant, and $df = (2 - 1)(3 - 1) = 2$, not 6.
6. A. Using a calculator, $P(\chi^2_1 \geq 4.51) = 0.034$. The other choices are left-tail or unrelated values; the p-value is always the right tail.
7. C. For a 2 × 2 table, $\chi^2 = z^2$, so $2.12^2 = 4.49$. Choice B forgets to square.
8. B. Rejecting H_0 says only that at least one distribution differs. Choice A overstates this; C reverses the decision.
9. D. The table is 3 × 2, so $df = (3 - 1)(2 - 1) = 2$. Choice A (1) treats it as a 2 × 2 table; choice B (49) uses a sample size.
10. H_0: the proportion who use transit weekly is the same at both schools (the distribution of response is the same). H_a: it is not.

    Conditions: two separate random samples; each under 10% of its school; expected counts: yes $100(105)/200 = 52.5$, no 47.5 in each row, all at least 5.

    $\chi^2 = \frac{(60 - 52.5)^2}{52.5} + \frac{(40 - 47.5)^2}{47.5} + \frac{(45 - 52.5)^2}{52.5} + \frac{(55 - 47.5)^2}{47.5} = 1.071 + 1.184 + 1.071 + 1.184 = 4.51$, $df = 1$, p-value $= 0.034$.

    Since $0.034 < 0.05$, reject H_0: convincing evidence that the proportion who use transit weekly differs between the schools (60% versus 45%).
11. (a) Column totals: pop 78, rock 72, other 50. Expected (each row): $100(78)/200 = 39$, 36, 25.

    (b) $\chi^2 = \frac{81}{39} + \frac{16}{36} + \frac{25}{25} + \frac{81}{39} + \frac{16}{36} + \frac{25}{25} = 2.077 + 0.444 + 1 + 2.077 + 0.444 + 1 = 7.04$, $df = (2 - 1)(3 - 1) = 2$.

    (c) P-value $= 0.030$. Since $0.030 < 0.05$, reject H_0: convincing evidence that the distribution of favorite music differs between the cities. City 1 prefers pop more (48% versus 30%).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

12. H_0: the distribution of travel mode is the same at both schools. H_a: it is not. Conditions: two random samples under 10% of their schools; column totals are 52, 46, 22 with table total 120, so expected counts are 21.67, 19.17, 9.17 and 30.33, 26.83, 12.83, all at least 5. $\chi^2 = 0.26$, $df = 2$, p-value $= 0.878$. Since $0.878 > 0.05$, we do not reject H_0: there is no convincing evidence that the travel distribution differs between the schools. The percentages are nearly the same (44, 36, 20 versus 43, 40, 17).
13. (a) Column totals: 35, 17, 9 and table total 61. For Group 1 the 'other' expected count is $23(9)/61 = 3.39$, which is less than 5, so the condition fails.
(b) Combine 'walk' and 'other': Group 1 = 15 and 8; Group 2 = 20 and 18. Expected: 13.2, 9.8 and 21.8, 16.2. $\chi^2 = 0.925$, $df = 1$, p-value $= 0.335$. No convincing evidence that the distribution differs between the groups. (The combined categories change what we are comparing: transit use versus non-bus.)
14. Contributions: A pass 3.57, A not 4.55, B pass 0.04, B not 0.05, C pass 4.32, C not 5.50. The largest come from A and C: A passed less often than expected (36% vs 56% overall) and C passed more often (78%). Because the technique was randomly assigned, the experiment supports a causal conclusion: the differences in pass rate are attributable to the technique (not to who chose it), provided the experiment was well designed.
15. In a 2 × 2 table, there is just one parameter of interest, the difference in two proportions. The chi-square statistic equals the square of the z statistic ($\chi^2 = z^2$), and both give the same p-value as the two-sided test because $P(\chi^2_1 \geq z^2) = P(|Z| \geq |z|)$. The chi-square test generalizes: it can compare three or more groups or categories at once with $df = (r-1)(c-1)$, which the two-proportion z-test cannot do.

9.3 Chi-Square Test for Independence

PRACTICE · CHI-SQUARE TEST FOR INDEPENDENCE

1. B. One sample with two categorical variables recorded means a test of independence (association). Homogeneity (C) would need separate samples from each group; goodness-of-fit (A) uses one variable.
2. C. The null is that the variables are independent. A is about a mean; B is the alternative; D describes the alternative for a homogeneity test.
3. A. $df = (3-1)(4-1) = 2 \cdot 3 = 6$. Choice D counts cells; choice B uses $r + c$.
4. D. $E = (80)(70)/200 = 28$. Choice C (35) is half of 70, and A (14) multiplies the two proportions and scales by 100 instead of n.
5. B. $P(\chi^2_2 \geq 14.73) = 0.0006$. The p-value is the right tail; 0.94 would be the left tail.
6. D. Since $0.040 < 0.05$, reject H_0: convincing evidence of association. Choice C overreaches because the data are observational; choice A confuses failing to reject with proving independence.
7. C. In option C, both rows have the same column proportions: $20/50 = 40/100 = 0.4$. Options A, B and D have different ratios in each row. Exact independence means $O = E$ in every cell.
8. A. Rejecting H_0 shows association only; a causal claim requires a randomized experiment. Choice D is never concluded from a test.
9. D. The test needs random sampling, independence (10%) and large expected counts. It makes no normality assumption about a population because the variables are categorical.
10. H_0: age group and smart speaker use are independent among adults. H_a: they are associated.
Conditions: random sample; 200 under 10% of adults; expected counts: row totals 100 and 100, column totals 112 and 88, so 56, 44, 56, 44, all at least 5.
$\chi^2 = \frac{16}{56} + \frac{16}{44} + \frac{16}{56} + \frac{16}{44} = 1.30$, $df = 1$, p-value $= 0.255$.
Since $0.255 > 0.05$, we do not reject H_0: there is not convincing evidence of an association between age group and smart speaker use (60% versus 52%).
11. (a) Row totals 50 and 90; column totals 50, 50, 40. Expected: row 1 17.86, 17.86, 14.29; row 2 32.14, 32.14, 25.71.
(b) $\chi^2 = 5.76 + 1.92 + 1.29 + 3.20 + 1.07 + 0.71 = 13.95$, $df = 2$.
(c) P-value $= 0.0009$. Reject H_0: convincing evidence that club membership and grade are associated. Members get A's more often (56% versus 24%).
12. $df = (3-1)(3-1) = 4$. Row totals 100, 100, 120; column totals 90, 120, 110; $n = 320$. Expected: 28.1, 37.5, 34.4; 28.1, 37.5, 34.4; 33.75, 45, 41.25 (all at least 5). $\chi^2 = 32.05$, p-value ≈ 0.000002. Reject H_0: very strong evidence of an association between region and favorite transport. South prefers cars (50% versus 20% in the north).
13. (a) Row totals 18 and 32; column totals 6 and 44. Expected: 2.16 and 15.84 in row 1; 3.84 and 28.16 in row 2.
(b) Two expected counts (2.16 and 3.84) are less than 5, so the large counts condition fails and the chi-square p-value is not trustworthy.
(c) Collect a larger sample, or use an exact method such as a randomization test.
14. Table: under 40: 75 use, 25 do not; 40 or older: 45 use, 55 do not. Expected: 60 and 40 in each column. $\chi^2 = \frac{225}{60} + \frac{225}{60} + \frac{225}{40} + \frac{225}{40} = 18.75$, $df = 1$, p-value $= 0.00001$. Very strong evidence of an association (75% vs 45% use a library). We cannot conclude that age causes library use because this is an observational sample; lurking variables, such as work schedule or tech habits, could explain the pattern.
15. Homogeneity: separate random samples (or random assignment) from each of several populations, with one categorical variable; H_0 says the distribution is the same in each population; rejecting shows the distributions differ. Independence: one random sample with two categorical variables; H_0 says the variables are independent; rejecting shows an association. Both compare observed counts in an $r \times c$ table with counts expected if there were no difference or association, so both use $E = (\text{row})(\text{col})/n$, $\chi^2 = \sum(O-E)^2/E$ and $df = (r-1)(c-1)$. Example of homogeneity: compare travel modes across three random samples of grades. Independence: one random sample of students classified by sleep and screen time.

9.4 Choosing and Interpreting Chi-Square Tests

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

PRACTICE · CHOOSING AND INTERPRETING CHI-SQUARE TESTS

1. B. One variable, one sample, and a claimed distribution is goodness of fit. Homogeneity (A) needs several samples and independence (C) needs two variables.
2. C. Several separate samples with one categorical variable mean homogeneity. Independence (B) needs one sample with two variables.
3. A. One sample classified by two variables is a test of independence, with H_0 stating independence. Choices B and C describe different tests and use the wrong df or wrong parameter.
4. D. Goodness of fit has $df = k - 1 = 4$. Independence with a 3 × 4 table has $df = (3-1)(4-1) = 6$. Choice B uses 12 cells and 5 categories.
5. A. A chi-square test can only give a two-sided result, so a one-sided claim needs a z-test. Choice B is wrong because the test is not one-sided, and the data are categorical so t is not used.
6. C. $(40 - 28)/\sqrt{28} = 2.27$. Choice B is just $O - E$; choice D is the square (the contribution), not the residual.
7. D. A large p-value means no convincing evidence of an association. Saying the variables are independent (A) claims too much.
8. B. A small p-value shows evidence against H_0, and with a large sample can come from a tiny effect; use percentages to judge size. Certainty (D) and causation (C) are never established by a p-value alone.
9. C. Expected counts under 5 make the chi-square approximation unreliable; combine similar categories or collect more data. Running the test anyway (A) gives untrustworthy p-values.
10. (a) Goodness of fit, $df = 6 - 1 = 5$.

    (b) Homogeneity (three separate samples, one variable), $df = (3-1)(4-1) = 6$.

    (c) Independence (one sample, two variables), $df = (3-1)(4-1) = 6$.
11. Design: two separate random samples with one categorical variable: chi-square test for homogeneity. H_0: the distribution of preferred event is the same at both schools; H_a: it is not. Conditions: random samples, under 10% of each school; column totals 50, 55, 55, so expected counts are 25, 27.5, 27.5 in each row, all at least 5.
    $\chi^2 = \frac{25}{25} + \frac{56.25}{27.5} + \frac{6.25}{27.5} + \frac{25}{25} + \frac{56.25}{27.5} + \frac{6.25}{27.5} = 6.55$, $df = 2$, p-value $= 0.038$. Since $0.038 < 0.05$, reject H_0: convincing evidence that the distribution of preferred event differs between the schools (School 1 prefers sports more: 44% versus 25%).
12. (a) One random sample of 80 students classified by two variables (tutored, passed): test of independence. (If 40 tutored and 40 non-tutored were selected separately, it would be homogeneity; the calculation is the same.)

    (b) Expected counts: 20 in every cell. $\chi^2 = 4\left(\frac{36}{20}\right) = 7.20$, $df = 1$, p-value $= 0.0073$.

    (c) $\chi^2 = z^2$, so $z = \sqrt{7.2} = 2.68$, the two-proportion z statistic for 65% vs 35%. Since $0.0073 < 0.05$, reject H_0: convincing evidence that tutoring and passing are associated; students who were tutored passed more often. Because this is observational, we cannot say tutoring causes the difference.
13. Goodness of fit: one variable, one sample, claimed equal distribution. H_0: each sector has probability 0.25. H_a: at least one differs. Conditions: random spins; expected 25 in each sector, all at least 5. $\chi^2 = \frac{25 + 9 + 9 + 25}{25}$ with differences 5, -3, 3, -5: $= 1 + 0.36 + 0.36 + 1 = 2.72$, $df = 3$, p-value $= 0.437$. Do not reject H_0: no convincing evidence the sectors are unequally likely.
14. Row 1 (under 14): $(16 - 10)/\sqrt{10} = 1.90$; $(24 - 22.5)/\sqrt{22.5} = 0.32$; $(10 - 17.5)/\sqrt{17.5} = -1.79$. Row 3 (16 and over): $(10 - 16)/4 = -1.50$; $(30 - 36)/6 = -1.00$; $(40 - 28)/\sqrt{28} = 2.27$. The clearest departure is 16-and-over studying online (2.27, more than expected); the next are under 14 in groups (1.90 more) and under 14 online (-1.79 fewer). Younger students lean toward group study; the oldest toward online study.
15. The chi-square test only tests the two-sided claim "the success rates differ." To claim Group B is higher, she needs a one-sided alternative, so she should use a two-proportion z-test. Here $z = \sqrt{4.555} = 2.13$ and the one-sided p-value is $P(Z > 2.13) = 0.016$ (half of 0.033). Conclusion: convincing evidence that Group B has the higher success rate.

T Unit 9 Test

UNIT TEST · UNIT 9 TEST

1. B. $df = k - 1 = 7 - 1 = 6$. Choice A subtracts twice; the sample size (D) is not used.
2. D. $250(0.18) = 45$. Choice A treats the percent as a count, and 55 does not come from any formula.
3. A. Chi-square curves start at 0, are skewed to the right and have mean equal to df. Choice D is impossible: squared differences are non-negative.
4. C. $(14 - 20)^2/20 = 36/20 = 1.8$. Choice A divides $|O - E|$ by 20 without squaring; D forgets to divide by E.
5. C. Several separate samples of one categorical variable: homogeneity. Independence (B) would need a single sample with two variables.
6. A. $df = (4-1)(5-1) = 12$. Choice D (20) counts cells and choice B (9) uses $r + c$.
7. D. $E = (90)(120)/300 = 36$. Choice A (30) is the row total divided by 3 and does not use the column total.
8. C. The expected counts 4.5 are below 5, so the large counts condition fails. The table shape does not determine df: it is $(2-1)(3-1) = 2$.
9. B. Since 0.056 is greater than 0.05, do not reject H_0. Choice A ignores the p-value; C claims to prove the null.
10. D. Rejection indicates association, not cause (A) or independence (C). Choice B says every group differs, which overstates the result.
11. A. For a 2 × 2 table, $\chi^2 = z^2$, so $|z| = \sqrt{6.25} = 2.5$. Choice C is the square.
12. C. $1.96^2 = 3.84$. A chi-square with 1 df is the square of a standard normal, so the 5% critical value of the chi-square corresponds to the two-tailed 5% value 1.96.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. Scoring guide (6 points). (a) 2 points: H_0: the flavor distribution is 25%, 35%, 25%, 15% (1); H_a: at least one differs, and names the chi-square goodness-of-fit test (1). (b) 1 point: expected counts 30, 42, 30, 18, all at least 5, and random sample. (c) 2 points: $\chi^2 = \frac{144}{30} + \frac{64}{42} + \frac{64}{30} + \frac{16}{18} = 4.8 + 1.52 + 2.13 + 0.89 = 9.35$ (1); $df = 3$, p-value $= 0.025$ (1). (d) 1 point: since $0.025 < 0.05$ reject H_0; convincing evidence that the flavor distribution is different from the claim.
Model response: H_0: the distribution of favorite flavor is 25/35/25/15; H_a: at least one proportion differs. Chi-square goodness of fit. Expected 30, 42, 30, 18 (all at least 5; random sample). $\chi^2 = 9.35$, $df = 3$, p-value $= 0.025$. Since $0.025 < 0.05$, there is convincing evidence that the distribution of favorite flavors differs from the shop's claim; vanilla was chosen more than expected.

14. Scoring guide (10 points). (a) 2 points: chi-square test for homogeneity (1); H_0: the distribution of payment method is the same at all three stores, H_a: it is not (1). (b) 2 points: $E = 100(100)/300 = 33.3$ (1); random independent samples, 100 under 10% of each store's shoppers, all expected counts at least 30; $df = (3-1)(3-1) = 4$ (1). (c) 3 points: p-value $= 0.0086$ (1); reject H_0 since $0.0086 < 0.05$ (1); conclusion in context: convincing evidence that the distribution of payment method is not the same at all three stores (1). (d) 2 points: Store A has more cash (45%) and fewer mobile (20%) than expected, while Store C has more mobile (40%) and less cash (25%) (2). (e) 1 point: shoppers were not randomly assigned to stores; this is an observational comparison, so lurking variables (such as neighborhood or age) could explain the differences.
Model response: (a) Homogeneity test, three separate random samples. H_0: same payment-method distribution at all stores; H_a: not the same. (b) Expected for Store A cash $= 100(100)/300 = 33.3$; random samples, each under 10% of its store's shoppers, all expected counts at least 30; $df = 4$. (c) p-value $= 0.0086 < 0.05$: reject H_0; convincing evidence the distribution differs for at least one store. (d) Store A shoppers use cash more (45% vs 33% expected) and mobile less (20% vs 30%); Store C shoppers use mobile more (40%) and cash less (25%). (e) Observational study, so we cannot conclude the store causes the payment method.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Unit 10: Inference for Regression Slopes

Estimate and test the slope of a population regression line. 3 lessons, 59 practice and test items, all with full explanations.

In this unit · Exam Unit 5 (extension)

10.1 The Regression Model, Conditions, and Computer Output
10.2 Confidence Intervals for the Slope
10.3 Hypothesis Tests for the Slope
Unit Test mixed multiple choice and a free-response question, with a full scoring guide

10.1 The Regression Model, Conditions, and Computer Output

Extend regression to the population: state the model, check the LINER conditions with residual plots, and read slope, standard error, and degrees of freedom from computer output.

Beyond the 2027 Exam Core

College Board lists inference for regression slopes as removed from the 2027 AP Statistics exam. Treat this unit as an extension: it is standard in college statistics courses and on earlier exams. Check your teacher's current course description to see whether you will be assessed on it.

Why This Matters

In Unit 3 you fitted a least-squares line to a sample. But a sample line is only an estimate. The real question is about the **population**: is there truly a linear relationship, and how steep is it? To answer that, we treat the sample slope b as one outcome of a sampling distribution and build a model for what the whole population looks like.

Students typically stumble in three places: confusing the sample statistics a, b, s with the population parameters α, β, σ; checking conditions only by looking at the scatterplot instead of the residual plot; and misreading a computer output table. This lesson sets up the model, the conditions, and how to read output. It extends the regression ideas of Exam Unit 5.

Core Ideas

The **population regression model** says that for each value x, the response y varies around a line:

Population regression model

$y = \alpha + \beta x + \varepsilon$, where the deviations ε are independent and normal with mean 0 and the same standard deviation σ for every x.
α = intercept, β = slope, σ = standard deviation of the responses around the line (all parameters).

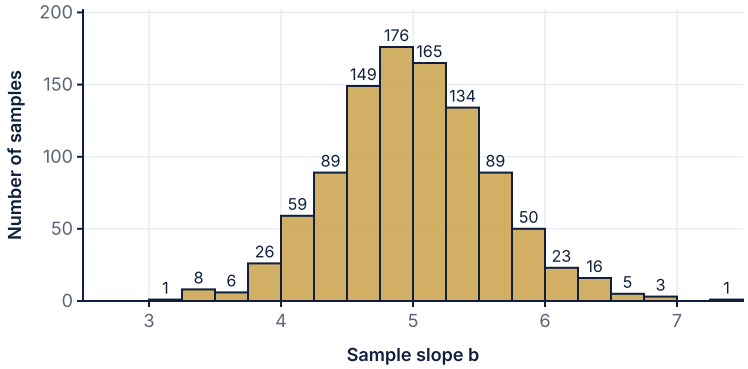
Estimates from one sample: a estimates α, b estimates β, and $s = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}$ estimates σ.

The mean response at x is $\mu_y = \alpha + \beta x$. The slope β is the change in the MEAN response per one-unit increase in x.

Different samples give different lines. The sample slope b is a statistic with a sampling distribution centered at β. The histogram shows 1000 simulated slopes from a population with $\beta = 5$, $n = 15$: the center is 4.99 and the SD is 0.592.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

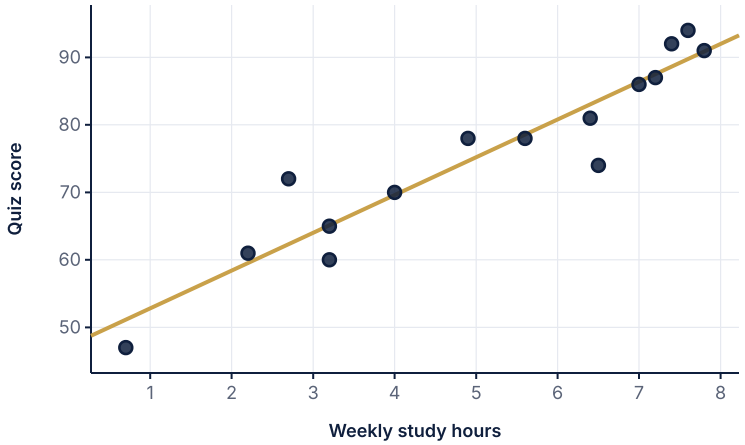


1000 simulated slopes: true beta = 5, n = 15

Conditions for regression inference: LINER

Linear: the true relationship is linear. Check the scatterplot and especially the residual plot for no curve.
Independent: observations are independent (for sampling without replacement, the sample is under 10 percent of the population).
Normal: for each x, the responses vary normally around the line. Check that residuals are roughly symmetric with no strong skew or outliers (histogram, dotplot, or normal probability plot).
Equal variance: the spread of residuals is the same for all x. Check that the residual plot has constant vertical spread (no fan).
Random: data come from a random sample or randomized experiment.

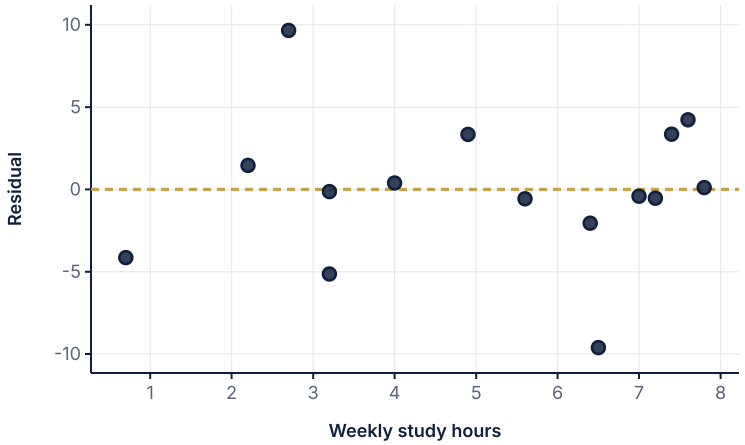
Here are the scatterplot and residual plot for 15 students who reported weekly study hours and their quiz scores. The residual plot has no curve and uniform spread, so Linear and Equal variance look reasonable.



Scatterplot with least-squares line, n = 15

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Residual plot: no pattern, even spread

Reading Computer Output

Statistical software gives a table like this one. The row labeled with the explanatory variable (here **Hours**) gives the slope b in the Coef column, and its standard error SE_b in the SE Coef column. The T-value is b/SE_b and the P-value is for a two-sided test of $H_0 : \beta = 0$.

Predictor	Coef	SE Coef	T-value	P-value
Constant	47.219	3.003	15.73	0.000
Hours	5.598	0.541	10.35	0.000

S	R-sq	R-sq(adj)
4.615	89.2%	88.3%

Standard error of the slope

$SE_b = \dfrac{s}{s_x\sqrt{n-1}}$, where s is the standard deviation of residuals and s_x is the standard deviation of the x values.

SE_b estimates how far a sample slope typically falls from the true β.

It is SMALLER when s is small (points close to the line), when the x values are spread out (large s_x), and when n is large.

Inference for slope uses a t distribution with $n-2$ degrees of freedom, because two parameters (α and β) were estimated.

Worked Examples

PARAMETERS VERSUS STATISTICS

A researcher fits a line to a random sample of 15 students: predicted score = 47.219 + 5.598(hours). Name each symbol: what is β, what is b, and what does s estimate?

SOLUTION

β is the true slope of the population line: the change in MEAN quiz score per extra weekly study hour for ALL students. It is unknown.

$b = 5.598$ is the slope of the sample line, a statistic that estimates β.

$s = 4.615$ is the standard deviation of the residuals. It estimates σ, the SD of quiz scores around the population line at any given hours value.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

READING THE OUTPUT

Use the regression output for the study-hours data (shown in Core Ideas). Give the regression equation, SE_b, s, r^2, and degrees of freedom for inference on the slope.

Predictor	Coef	SE Coef	T-value	P-value
Constant	47.219	3.003	15.73	0.000
Hours	5.598	0.541	10.35	0.000

SOLUTION

Equation: predicted score = 47.219 + 5.598(hours).

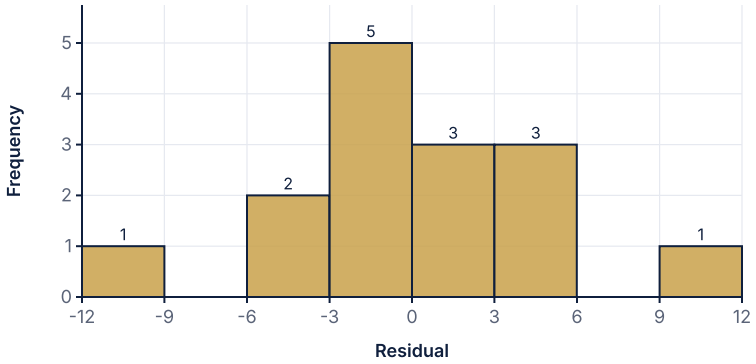
$SE_b = 0.541$ (row Hours, SE Coef).

$s = 4.615$ points; $r^2 = 89.2$ percent of the variation in quiz scores is explained by the linear relationship with hours.

Degrees of freedom: $n - 2 = 15 - 2 = 13$.

CHECKING LINER

For the study-hours data, the scatterplot is linear, the residual plot shows no pattern and constant spread, and a histogram of the residuals is symmetric. The 15 students were randomly selected from a school of 1800. Are the conditions for inference met?



Residuals from the study-hours regression

SOLUTION

Linear: the scatterplot is roughly linear and the residual plot has no curve.

Independent: random selection, and 15 is under 10 percent of 1800.

Normal: the residual histogram is roughly symmetric with no outliers.

Equal variance: the residual plot has constant spread.

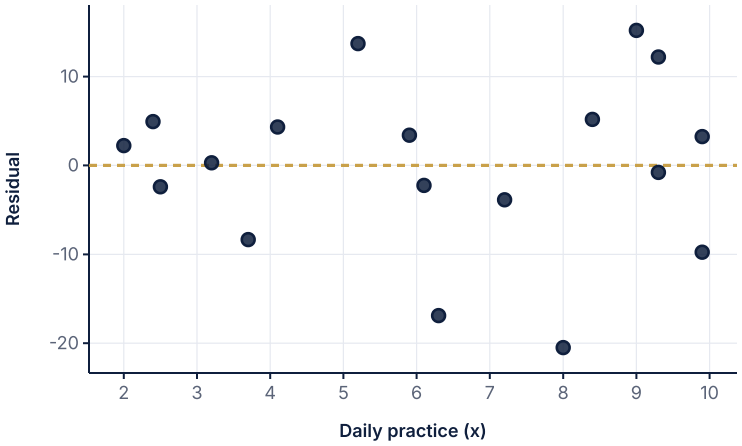
Random: stated. All conditions are met, so a t procedure for the slope is justified.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

SPOTTING A VIOLATED CONDITION

A sample of 18 shows the residual plot below: residuals are tightly clustered for small x and fan out for large x. Which condition fails, and what does that mean for inference?



Residual plot: fan shape

SOLUTION

The equal-variance condition fails. The spread of y around the line grows with x.

A single s cannot describe the scatter at all values of x, so SE_b and the t procedures would be unreliable.

Possible remedies: transform the data, or report that the model is not appropriate. The line itself may still describe the trend, but inference should not be trusted.

COMPUTING SE_B FROM SUMMARY STATISTICS

For a sample of $n = 20$ points, $s = 6.2$ and $s_x = 3.1$. Find SE_b, and explain what happens if the x values are twice as spread out.

SOLUTION

$$SE_b = \frac{6.2}{3.1\sqrt{19}} = \frac{6.2}{3.1(4.359)} = 0.459.$$

If s_x doubles (and s stays the same) SE_b is cut in half to about 0.229, because s_x is in the denominator.

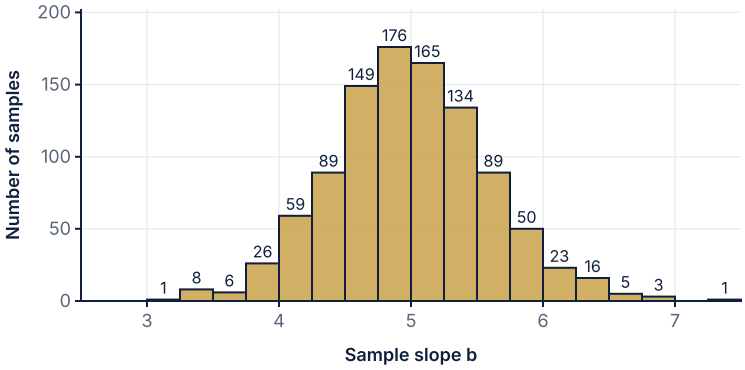
More spread in x gives a more stable slope estimate, like a seesaw with a longer lever.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

SAMPLING DISTRIBUTION OF THE SLOPE

In the simulated histogram of 1000 slopes (true $\beta = 5$, $n = 15$, $\sigma = 4.6$), describe the distribution and compare its SD to the SE in the study-hours output.



1000 simulated slopes: true beta = 5, n = 15

SOLUTION

Center: 4.99, essentially the true $\beta = 5$, so b is unbiased for β.

Shape: roughly symmetric, mound-shaped (approximately normal; with s estimated we use t).

Spread: SD about 0.592. The study-hours output has $SE_b = 0.541$, similar in size. The standard error from one sample estimates the SD of this sampling distribution.

Write It Like a 5

Model Response

Prompt. A random sample of 12 teens recorded daily practice minutes and a skill score. The software output gives the values listed below, and the residual plot shows no pattern. (a) Define the parameter β. (b) State and check the conditions for inference. (c) Identify the standard error of the slope. \nUse this output: Constant coefficient 39.683, Minutes coefficient 2.347 with SE Coef 0.409, S = 2.984, R-sq = 76.7 percent, n = 12. **(a)** β is the true slope of the population regression line: the change in the mean skill score for each additional daily practice minute for all teens in the population. **(b)** Linear: the scatterplot is roughly linear and the residual plot shows no curve. Independent: random sample, and 12 teens is under 10 percent of all teens. Normal: residuals are roughly symmetric with no outliers. Equal variance: residual spread is constant across x. Random: the teens were randomly selected. **(c)** $SE_b = 0.409$. This is the estimated typical amount by which the sample slope differs from the true slope β.

What the grader looks for

1 point: defines β as a POPULATION slope in context (change in mean response). 1 point: names the L, I, N, E, R conditions. 1 point: verifies at least the residual-plot conditions with evidence (not just 'the scatterplot looks linear'). 1 point: reads SE_b from the SE Coef row of the explanatory variable and explains it.

Coach's Playbook

1. Write what β means in words: change in the MEAN of y for each one-unit increase in x.
2. Draw or read the residual plot first. Look for curves (Linear) and fans (Equal variance).
3. Check residuals for skew or outliers (Normal), and describe how the sample was collected (Independent, Random).
4. Read output by row: the explanatory variable's row has b, SE_b, t, and P.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5 Remember df = n - 2, not n - 1.

Common traps

1. Confusing the intercept row with the slope row when reading output. 2. Checking linearity only on the scatterplot. Curves are far easier to see in the residual plot. 3. Using n - 1 degrees of freedom for regression. 4. Describing β as the change in an individual's y rather than the change in the MEAN of y.

Mistake Clinic

Spot the error

Student: 'The slope is 4.5 points per hour for every student.' **Where it fails:** b is a sample estimate and β describes the MEAN response. Individual students scatter around the line. **Correct:** 'For each additional weekly study hour, the mean predicted quiz score increases by about 4.5 points.'

Spot the error

Student: 'The scatterplot looks linear and r is large, so all conditions for inference are satisfied.' **Where it fails:** A large r does not check Equal variance, Normality of residuals, or Independence. **Correct:** Examine the residual plot and a residual graph, and verify how the data were collected.

Spot the error

Student: 'n = 15, so df = 14.' **Where it fails:** Regression estimates two parameters. **Correct:** df = n - 2 = 13.

Real-World Connection

Public-health and education researchers ask whether the slope relating a behavior to an outcome is really different from zero in a population. Regression output tables are how they report that evidence, along with checks of the residual plot.

Challenge Corner

Problem. Two studies use the same $s = 5$ and $n = 20$. Study 1 uses x values with $s_x = 2$; Study 2 uses x values with $s_x = 4$. Compare SE_b. **Solution.** Study 1: $5/(2\sqrt{19}) = 0.574$. Study 2: $5/(4\sqrt{19}) = 0.287$. Study 2 has half the standard error because its x values are twice as spread out, so its slope estimate is more precise. Design lesson: when you can choose x values, spread them widely.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE **The Regression Model, Conditions, and Computer Output**

Show your work. A calculator is allowed. Justify conclusions in context.

1. In the model $y = \alpha + \beta x + \varepsilon$, the symbol β represents

   (A) the slope of the sample line **(B)** the slope of the population regression line
   (C) the standard deviation of the residuals **(D)** the correlation

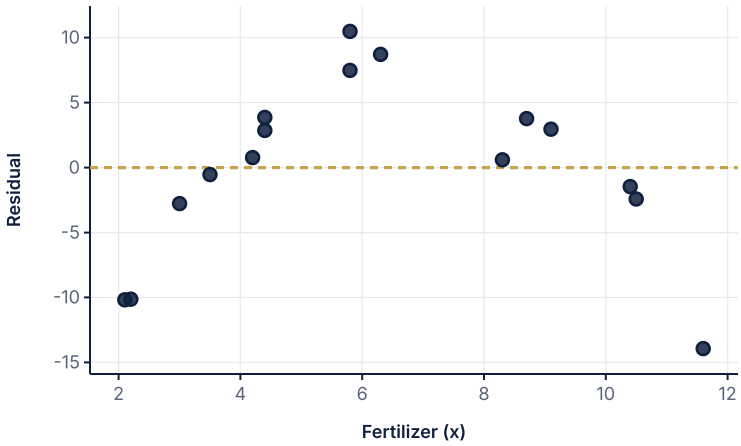
2. Which statistic estimates σ, the standard deviation of the responses about the population line?

   (A) r **(B)** b **(C)** $s = \sqrt{\sum (y_i - \hat{y}_i)^2 / (n-2)}$ **(D)** s_x

3. A researcher has 20 pairs of data. The degrees of freedom for t inference on the slope are

   (A) 20 **(B)** 19 **(C)** 17 **(D)** 18

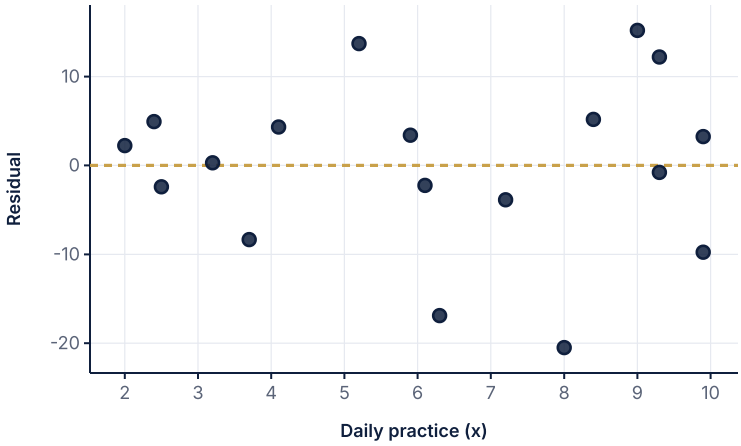
4. The residual plot below shows residuals that form a clear U shape. The condition that is violated is



Residual plot: clear curve

(A) Equal variance **(B)** Normality **(C)** Linear
(D) Independence

5. The residual plot shows residuals tightly grouped for small x and widely spread for large x. The violated condition is



Residual plot: fan shape

(A) Linear **(B)** Equal variance **(C)** Random
(D) Independent only

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

6. Which change would DECREASE the standard error of the slope, with everything else the same?
   (A) Using x values that are more spread out **(B)** Using x values that are closer together
   (C) Reducing the sample size **(D)** Increasing s
7. In the output table, the P-value in the row for the explanatory variable tests

Predictor	Coef	SE Coef	T-value	P-value
Constant	39.683	2.273	17.46	0.000
Minutes	2.347	0.409	5.74	0.000

   (A) H0: alpha = 0 **(B)** H0: rho = 1
   (C) H0: beta = 0 against a two-sided alternative **(D)** H0: s = 0
8. The output shows Constant 38.4 and Minutes 2.3 with SE Coef 0.44. The estimated standard deviation of the sampling distribution of b is
   (A) 0.44 **(B)** 2.3 **(C)** 38.4 **(D)** 5.2
9. A 'normal' condition for regression inference means that
   (A) the x values are normally distributed
   (B) the residuals are approximately normally distributed with constant spread
   (C) the scatterplot is symmetric
   (D) r is near 0
10. For the model $y = \alpha + \beta x + \varepsilon$, describe in words what α, β, and σ mean for a study of daily practice minutes (x) and skill score (y).
11. For the practice-minutes sample (n = 12), the regression output is: Constant 39.683, Minutes 2.347 with SE Coef 0.409, S = 2.984, R-sq = 76.7 percent.
    (a) Write the least-squares equation.
    (b) Interpret the slope in context.
    (c) Give the degrees of freedom and the standard error of the slope.
12. State the four conditions other than Random for inference about a slope, and for each tell which graph you would examine.
13. A sample of 25 points has $s = 4.5$ and $s_x = 3$. (a) Compute SE_b. (b) If the sample size were 100 with the same s and s_x, what is SE_b and what does this tell you?
14. A residual plot shows a clear curved pattern, but the correlation is r = 0.93. A student says inference is fine because r is high. Respond.
15. Explain why the sampling distribution of the slope b is centered at β, and what would make its spread smaller. Use a sentence about the simulation idea.

10.2 Confidence Intervals for the Slope

Build and interpret a t confidence interval for the population slope using b, its standard error, and n - 2 degrees of freedom, with conditions checked from residual plots.

Beyond the 2027 Exam Core

College Board lists inference for regression slopes as removed from the 2027 AP Statistics exam. This lesson is an extension for college-bound students and older exam practice. Confirm with your teacher whether it is part of your course.

Why This Matters

A slope estimate by itself is not enough. If a used-car regression gives $b = -1.8$ thousand dollars per year of age, how close is that to the TRUE rate of price decline for all such cars? A **confidence interval for the slope** answers: it gives a range of plausible values for the population slope β and a level of confidence in the method.

Students often use z^* instead of t^*, use $n - 1$ degrees of freedom, forget to read SE_b off the right row, or interpret the interval as a statement about individual cars instead of the mean change. Slope inference connects to the regression ideas of Exam Unit 5.

Core Ideas

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

The sampling distribution of the slope b is centered at β with spread estimated by SE_b. Because we estimate σ with s, the standardized statistic $\frac{b-\beta}{SE_b}$ follows a **t distribution with $n-2$ degrees of freedom** when the LINER conditions hold.

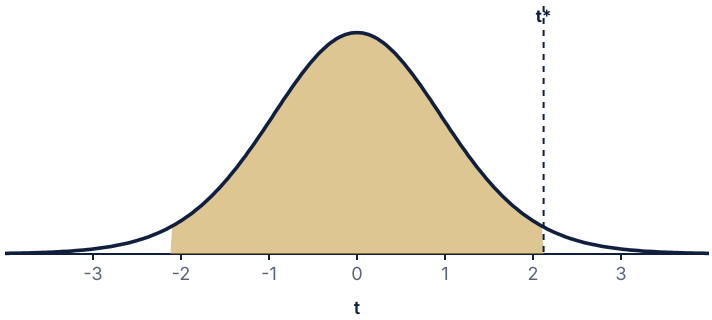
Confidence interval for a slope

Statistic $\pm$ (critical value)(standard error of statistic): $b \pm t^* \cdot SE_b$.

t^* is the critical value for the chosen confidence level from the t distribution with $df = n-2$ (use the built-in calculator's inverse t function, or a t table).

SE_b comes from computer output (SE Coef column of the explanatory-variable row) or from $s/(s_x\sqrt{n-1})$.

Margin of error $= t^* \cdot SE_b$.



t distribution with 16 df: middle 95 percent lies between -2.120 and 2.120

The critical value grows with confidence and shrinks with degrees of freedom. A few values appear below.

Confidence level	df = 10	df = 16	df = 25	df = 40
90%	1.812	1.746	1.708	1.684
95%	2.228	2.120	2.060	2.021
99%	3.169	2.921	2.787	2.704

Four-step procedure (State, Plan, Do, Conclude)

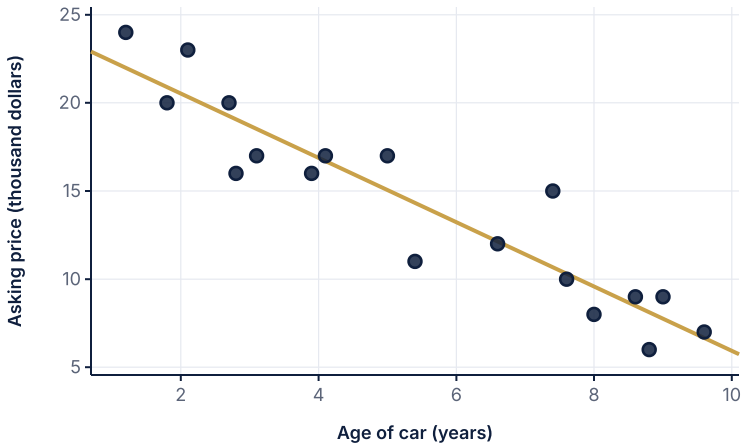
State: name the parameter β in context and the confidence level.

Plan: name the procedure (t interval for a slope) and check LINER with residual plots and the study design.

Do: give b, SE_b, df, t^*, and the interval.

Conclude: 'We are C percent confident that the true slope of the population regression line of y on x is between __ and __ (units of y per unit of x).'

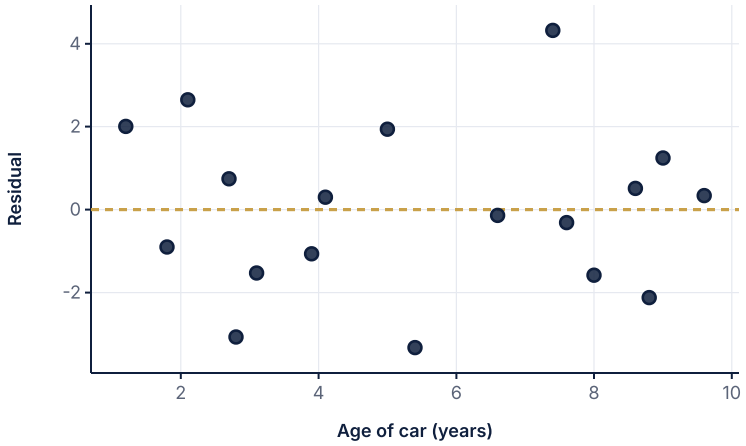
Here is the used-car dataset. The scatterplot shows a negative linear trend, the residual plot has no pattern, and the residual histogram is roughly symmetric, so LINER looks reasonable.



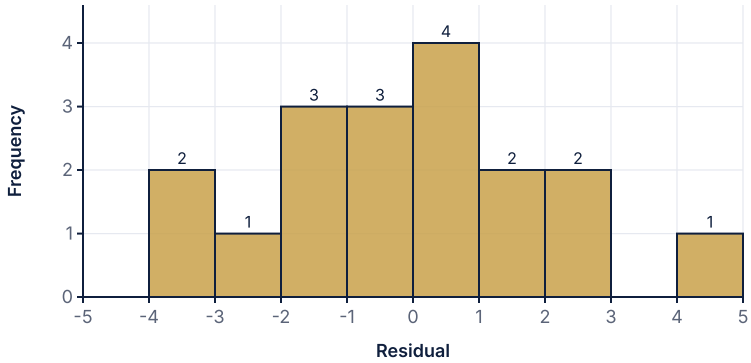
n = 18 used cars

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



Residual plot: no pattern



Residuals of price on age

The regression output for these 18 cars:

Predictor	Coef	SE Coef	T-value	P-value
Constant	24.185	1.081	22.38	0.000
Age	-1.825	0.178	-10.26	0.000

S	R-sq	R-sq(adj)
2.061	86.8%	86.0%

Worked Examples

FINDING T^*

Find t^* for (a) a 95 percent interval with n = 18, (b) a 90 percent interval with n = 25, (c) a 99 percent interval with n = 12.

SOLUTION

(a) df = 16 and $t^* = 2.120$.

(b) df = 23 and $t^* = 1.714$.

(c) df = 10 and $t^* = 3.169$.

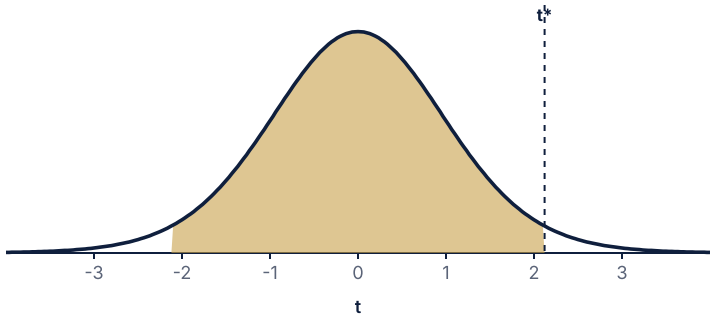
Always subtract 2 for df. Compare with z^* = 1.960, 1.645, 2.576: the t values are larger, especially for small samples.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

INTERVAL FROM OUTPUT

Use the used-car output to construct and interpret a 95 percent confidence interval for the slope of the true regression line of price (thousand dollars) on age (years). Conditions were checked above.



t distribution with 16 df: middle 95 percent lies between -2.120 and 2.120

SOLUTION

State: β = the true slope of the regression of asking price on age for all used cars of this type.

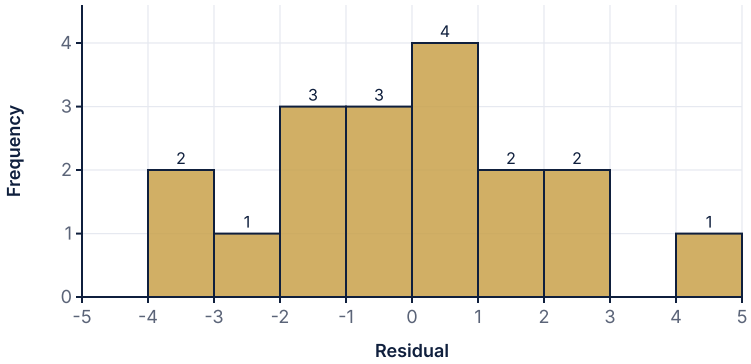
Plan: t interval for a slope; conditions (LINER) satisfied by the plots; cars were a random sample from a large listing.

Do: $b = -1.825$, $SE_b = 0.178$, df = 16, $t^* = 2.120$. Margin of error $= 2.120(0.178) = 0.377$. Interval: $-1.825 \\ pm 0.377 = (-2.202, -1.448)$.

Conclude: We are 95 percent confident that the true slope is between -2.202 and -1.448 thousand dollars per year. That is, each additional year of age is associated with a decrease in MEAN asking price of about 1.45 to 2.20 thousand dollars.

CHECKING CONDITIONS WITH PLOTS

For the used-car data, explain how the plots support the use of a t interval.



Residuals of price on age

SOLUTION

Linear: the scatterplot follows a straight trend and the residual plot shows no curve.

Equal variance: the residual plot has no fan; the vertical spread looks constant.

Normal: the residual histogram is roughly symmetric and mound-shaped, with no outliers.

Independent and Random: the cars are a random sample from a large listing, fewer than 10 percent of cars. All conditions are met.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

INTERVAL FROM SUMMARY STATISTICS

A random sample of 22 students gives $b = 3.1$ points per extra hour, $s = 8.4$, and $s_x = 5.2$. Conditions are satisfied. Find a 90 percent confidence interval for β.

SOLUTION

$SE_b = \dfrac{8.4}{5.2\sqrt{21}} = 0.353.$

df = 20, $t^* = 1.725$ for 90 percent confidence.

Margin of error $= 1.725(0.353) = 0.608$. Interval: $3.1 pm 0.608 = (2.492, 3.708)$.

We are 90 percent confident that the true slope is between 2.492 and 3.708 points per hour.

AN INTERVAL THAT CONTAINS 0

A random sample of 12 practice runs recorded outside temperature (degrees F) and laps completed. The output is shown below. Find and interpret a 95 percent confidence interval for the slope; does the data suggest a linear relationship in the population?

Predictor	Coef	SE Coef	T-value	P-value
Constant	20.737	5.043	4.11	0.002
Temp	0.037	0.069	0.54	0.603

SOLUTION

$b = 0.037$, $SE_b = 0.069$, df = 10, $t^* = 2.228$. Margin of error $= 0.155$.

Interval: $(-0.117, 0.192)$ laps per degree.

Interpretation: we are 95 percent confident that the true slope is between -0.117 and 0.192 laps per degree.

Because 0 is in the interval, zero slope (no linear relationship) is plausible. The data do not give convincing evidence of a linear relationship between temperature and laps. This agrees with the large P-value 0.603 in the output.

CONFIDENCE LEVEL AND PRECISION

Using the used-car data, compare 90, 95, and 99 percent intervals, and explain what would happen if the sample size were four times larger.

Confidence	t*	Interval	Width
90%	1.746	-2.136 to -1.515	0.621
95%	2.120	-2.202 to -1.448	0.754
99%	2.921	-2.345 to -1.306	1.039

SOLUTION

Higher confidence needs a larger t^*, so the intervals get wider (see the table).

A wider interval gives more confidence but less precision.

With four times as many cars (72) and similar s and s_x, SE_b would be about 0.087 instead of 0.178, roughly half, and t^* would also fall slightly, so the interval would be about half as wide.

Write It Like a 5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Prompt. A random sample of 18 used cars of one model gave the regression of asking price (thousand dollars) on age (years): slope -1.825, SE_b = 0.178. Residual plots show no pattern and the residuals are roughly normal. Construct and interpret a 95 percent confidence interval for the slope. **State.** Let β be the true slope of the population regression line relating asking price to age for used cars of this model. We want a 95 percent confidence interval for β. **Plan.** One-sample t interval for a slope. Conditions: the cars are a random sample (Random) and fewer than 10 percent of all such cars (Independent). The residual plot shows no curve (Linear) and constant spread (Equal variance), and the residuals are roughly symmetric (Normal). **Do.** df = 18 - 2 = 16, $t^* = 2.120$. $-1.825 pm 2.120(0.178) = (-2.202, -1.448)$. **Conclude.** We are 95 percent confident that the true slope of the regression line of price on age is between -2.202 and -1.448 thousand dollars per year. In other words, each extra year of age is associated with a drop in the mean asking price of about 1.45 to 2.20 thousand dollars.

What the grader looks for

1 point: defines β in context. 1 point: names the t interval for a slope and checks LINER with evidence. 1 point: correct df and t^*. 1 point: correct interval. 1 point: conclusion states the confidence level, the slope units, and refers to the population (mean change), not individual cars.

Coach's Playbook

1. Get b and SE_b from the explanatory-variable row of the output.
2. Find df = n - 2, then t* for the confidence level with the inverse t function on the built-in calculator.
3. Compute b plus or minus t* times SE_b. Do not mix up the SE Coef and Coef columns.
4. Check whether 0 is in the interval: if it is, a zero slope is plausible.
5. Write the conclusion with units, the population, and the words 'mean' or 'average'.

Common traps

1. Using z^* or $n - 1$ degrees of freedom. 2. Using the standard error of the intercept (the Constant row) instead of SE_b. 3. Writing 'we are 95 percent confident that each car loses...' The interval is about the MEAN change for the population, not each car. 4. Interpreting the confidence level as the probability that β is in this specific interval; it describes the method's long-run success rate.

Mistake Clinic

Spot the error

Student: 'For n = 18 and 95 percent, t* = 2.110, from df = 17.' **Where it fails:** df for a slope is n - 2. **Correct:** df = 16 and $t^* = 2.120$.

Spot the error

Student: 'The interval (-2.2, -1.4) means 95 percent of cars lose between 1.4 and 2.2 thousand dollars per year.' **Where it fails:** The interval estimates the population slope β, not the middle 95 percent of individual cars. **Correct:** We are 95 percent confident the true mean annual price drop is between 1.4 and 2.2 thousand dollars.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

Student: 'The interval for the slope includes 0, so we are sure that there is no relationship.' **Where it fails:** Including 0 only means no linear relationship is plausible; the sample may be too small to detect one. **Correct:** The data do not provide convincing evidence of a linear relationship; zero slope is plausible.

Real-World Connection

Economists, biologists, and engineers report slopes with intervals, such as 'each extra degree of warming is associated with a change of x to y days in the growing season.' The interval tells readers how precise the estimate is.

Challenge Corner

Problem. In the used-car data, a competing analyst wants a 95 percent interval for β with margin of error at most 0.15 thousand dollars per year. Using $s = 2.061$ and $s_x = 2.81$, roughly how many cars are needed? Use $t^* \approx 2.0$. **Solution.** Need $2.0 \cdot s/(s_x\sqrt{n-1}) \le 0.15$, so $\sqrt{n-1} \ge 2.0(2.061)/(2.81 \cdot 0.15) = 9.78$, so $n - 1 \ge 95.6$ and $n \ge 97$. About 97 cars would be needed, much more than the 18 collected.

Practice

PRACTICE

Confidence Intervals for the Slope

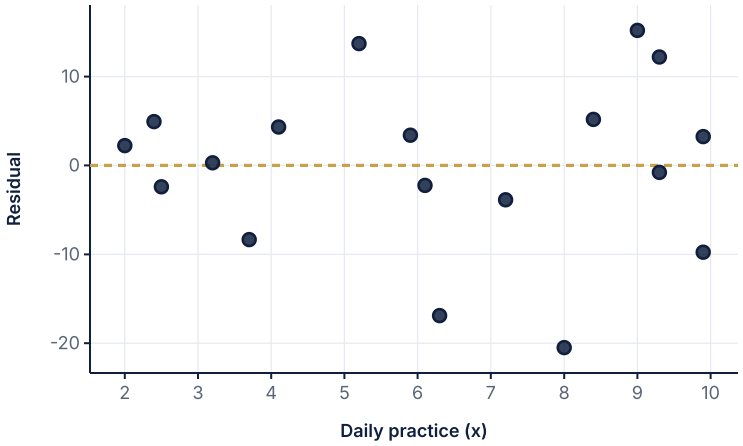
Show your work. A calculator is allowed. Justify conclusions in context.

1. For a 95 percent confidence interval for a slope based on 25 points, the degrees of freedom are
   (A) 25 **(B)** 24 **(C)** 23 **(D)** 22
2. The general form of a confidence interval for the slope is
   (A) b +/- z* times s **(B)** b +/- t* times SE_b **(C)** b +/- t* times s/n
   (D) r +/- t* times SE_b
3. The critical value t^* for a 95 percent interval from a sample of n = 12 pairs is closest to
   (A) 2.201 **(B)** 1.812 **(C)** 2.228 **(D)** 2.179
4. A 95 percent confidence interval for the slope of height (cm) on age (years) is (4.8, 6.2). The correct interpretation is
   (A) 95 percent of children grow between 4.8 and 6.2 cm per year
   (B) The interval contains 95 percent of the sample slopes
   (C) There is a 95 percent chance the sample slope is between 4.8 and 6.2
   (D) We are 95 percent confident that the mean height increases between 4.8 and 6.2 cm per year of age in the population
5. A 95 percent confidence interval for a slope is (-0.12, 0.31). Based on this interval, which conclusion is best?
   (A) There is convincing evidence of a positive linear relationship
   (B) There is convincing evidence of a negative linear relationship
   (C) Zero slope is plausible, so there is not convincing evidence of a linear relationship
   (D) The correlation must be exactly 0
6. Switching from 95 percent to 99 percent confidence with the same data will
   (A) narrow the interval **(B)** widen the interval
   (C) not change the interval **(D)** change the slope estimate
7. Which gives the narrowest interval for the slope, other things equal?
   (A) n = 10, widely spread x **(B)** n = 50, narrowly spread x
   (C) n = 50, widely spread x **(D)** n = 10, narrowly spread x
8. The output gives Coef 1.10 and SE Coef 0.21 for the explanatory variable with n = 14. The margin of error for a 99 percent interval is closest to
   (A) 0.64 **(B)** 0.45 **(C)** 0.35 **(D)** 0.21

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. A student wants a confidence interval for the slope but the residual plot shows a clear fan shape. The best advice is



Fan-shaped residual plot

(A) Proceed, since intervals are robust to this
(B) Do not trust the t interval because the Equal variance condition fails
(C) Use z* instead of t*
(D) Increase the confidence level

10. Find t^* for: (a) 90 percent confidence with n = 10; (b) 95 percent confidence with n = 30; (c) 99 percent confidence with n = 18. Give the degrees of freedom for each.
11. Using the output for the dose-response data (n = 14): Constant 38.586, Dose coefficient 1.459 with SE Coef 0.401, S = 3.794.
(a) Compute a 99 percent confidence interval for the slope.
(b) Interpret it in context, assuming conditions are met.
12. A random sample of 20 buildings gave $b = -2.4$ units of energy per extra insulation layer, with $s = 5.0$ and $s_x = 1.8$. Find a 95 percent confidence interval for β and interpret it.
13. A researcher reports the 95 percent interval for the slope of weight on height as (0.4, 1.9) pounds per inch. A reader says, 'So 95 percent of people gain between 0.4 and 1.9 pounds for each extra inch.' Explain what is wrong and rewrite the interpretation.
14. Explain how each of the following affects the width of a confidence interval for a slope: (a) a larger sample size; (b) a higher confidence level; (c) more spread in the x values; (d) a larger standard deviation of residuals.
15. For the temperature and laps data, the 95 percent interval for the slope is (-0.117, 0.192). Does it include 0? State what that suggests about the population slope, and explain why a small sample might fail to detect a real relationship.

10.3 Hypothesis Tests for the Slope

Test whether a population slope is zero (or another value) using t = (b - hypothesized value) / SE_b with n - 2 degrees of freedom, and write conclusions about association versus causation.

Beyond the 2027 Exam Core

College Board lists inference for regression slopes as removed from the 2027 AP Statistics exam. This test is an extension used in college courses and on earlier exams. Confirm with your teacher whether it applies to your course.

Why This Matters

Is there really a linear relationship in the population, or could the slope we see in a sample be just chance? A **significance test for the slope** answers this by measuring how far the sample slope b is from a hypothesized value in standard-error units.

Students often forget that the P-value in computer output is two-sided, mix up the null and alternative hypotheses, use $n - 1$ degrees of freedom, or conclude 'no relationship' after a large P-value. This lesson ties together the model and conditions from earlier in the unit with the logic of tests from Exam Unit 3. It extends Exam Unit 5.

Core Ideas

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

The most common null hypothesis is $H_0 : \beta = 0$: there is NO linear relationship between x and the mean of y in the population. The alternative can be two-sided ($\beta \neq 0$) or one-sided ($\beta > 0$ or $\beta < 0$) based on the research question stated BEFORE seeing the data.

t test for the slope

Hypotheses: $H_0 : \beta = \beta_0$ (usually 0) versus $H_a : \beta \neq \beta_0$, $\beta > \beta_0$, or $\beta < \beta_0$.

Test statistic: $t = \dfrac{b - \beta_0}{SE_b}$, which follows a t distribution with $df = n - 2$ when H_0 is true and LINER holds.

P-value: the probability, if H_0 is true, of getting a t at least as extreme as the one observed, in the direction of H_a.

Decision: if P-value $\leq \alpha$ reject H_0 and say the data provide convincing evidence of H_a in context; otherwise say the data do not provide convincing evidence.

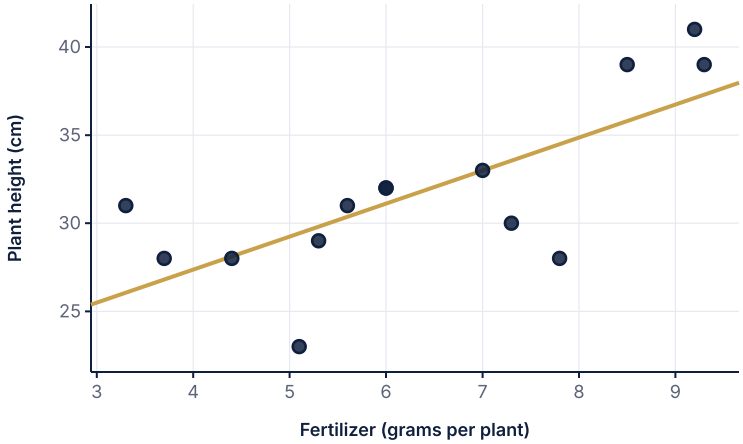
Reading output for a test

The row for the explanatory variable gives b (Coef), SE_b (SE Coef), $t = b/SE_b$ (T-value), and a TWO-SIDED P-value.

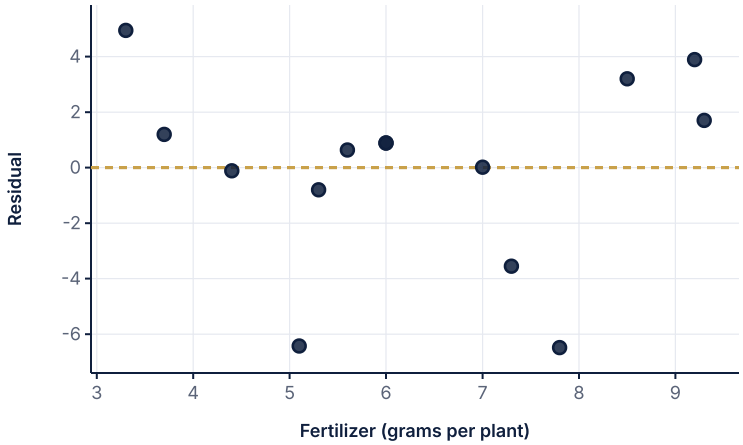
For a one-sided H_a: if b has the sign predicted by H_a, the one-sided P-value is HALF of the printed value. If b has the opposite sign, the P-value is greater than 0.5 and there is no evidence for H_a.

The test of $H_0 : \beta = 0$ has the same t and P-value as the test of $H_0 : \rho = 0$ for the correlation.

Here is a dataset: 14 plants each given a different amount of fertilizer, with height measured after six weeks. The scatterplot is linear, the residual plot has no pattern, and the residuals are roughly symmetric, so LINER is reasonable.



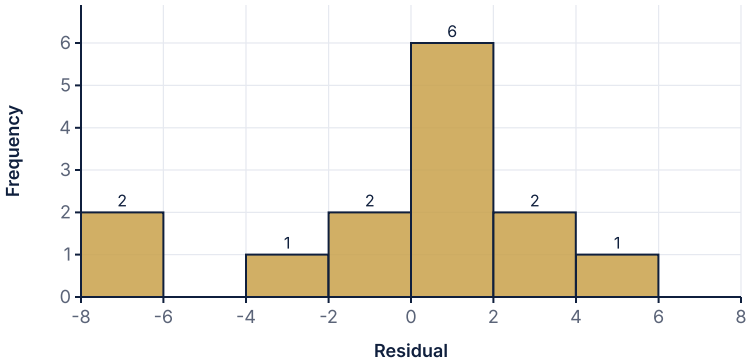
n = 14 plants



Residual plot: no pattern

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

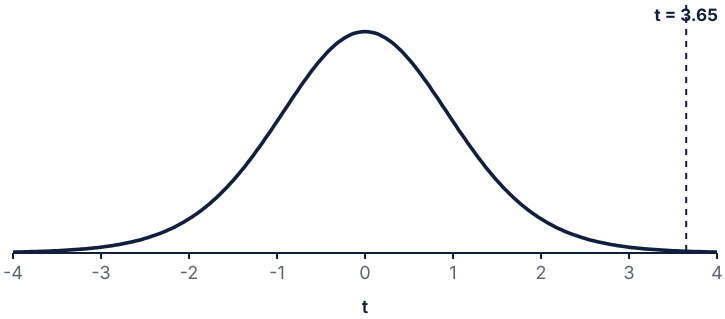


Residuals of height on fertilizer

Predictor	Coef	SE Coef	T-value	P-value
Constant	19.872	3.378	5.88	0.000
Fertilizer	1.873	0.513	3.65	0.003

S	R-sq	R-sq(adj)
3.568	52.7%	48.7%

From the table, $t = 1.873/0.513 = 3.65$ with $df = 12$. The picture shows the upper tail beyond this t.



t distribution, df = 12: upper-tail area beyond the observed t

The logic of the test. Suppose the true slope really is $\beta_0 = 0$. Then different random samples would still produce different sample slopes, scattered around 0 with a typical distance of about SE_b. The test statistic t counts how many standard errors the observed slope is from 0. If $|t|$ is small, the sample slope is the kind of thing that happens by chance when no relationship exists. If $|t|$ is large, a slope that far from 0 would be rare under H_0, so we doubt H_0. The P-value turns that statement into a probability: it is the area under the t curve beyond the observed t, counted in the direction(s) of H_a.

We use a t distribution, not a normal distribution, because we replaced the unknown σ with its estimate s inside SE_b. The t curve has heavier tails than the normal curve, and the extra weight shrinks as degrees of freedom grow. For regression, $df = n - 2$ because the line uses up two degrees of freedom: one for the intercept and one for the slope.

Printed P-value (two-sided)	Sign of b	Alternative	P-value to use
0.040	Positive	$\beta > 0$	0.020
0.040	Negative	$\beta > 0$	0.980
0.040	Negative	$\beta < 0$	0.020
0.040	Either	$\beta \neq 0$	0.040

Connecting the Test and the Interval

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

A two-sided test of $H_0 : \beta = 0$ at significance level α agrees with a confidence interval for β at confidence level $1 - \alpha$. If the 95 percent interval for β excludes 0, the two-sided test at $\alpha = 0.05$ rejects H_0; if the interval contains 0, the test fails to reject. For the fertilizer data the 95 percent interval is (0.756, 2.990), which excludes 0, matching the small P-value. The interval is more informative because it also shows how large the slope might be, not just that it differs from zero.

Statistical Versus Practical Significance

With a very large sample, even a tiny slope can give a very small P-value. A significant slope tells us a relationship is unlikely to be zero, but it does not say the relationship is important. Always report the slope and its interval in context, and ask whether the size of the effect matters for the real situation. A small P-value from a sample of thousands might correspond to an increase of only a fraction of a point.

- **Small P-value:** the sample slope is unlikely to be this far from the hypothesized value if H_0 were true.
- **Large P-value:** the sample slope is consistent with H_0. This is not proof that H_0 is true.
- **Type I error:** concluding there is a linear relationship when the true slope is 0. Its probability is α.
- **Type II error:** failing to detect a real nonzero slope. It is more likely with small samples, large scatter, or narrowly spread x values.

Worked Examples

STATE THE HYPOTHESES

A gardener asks: does more fertilizer increase plant height? Write the parameter and hypotheses for a test of the fertilizer-height slope, and say how the hypotheses would differ if she only asked whether there is ANY linear relationship.

SOLUTION

Let β = the true slope of the population regression line of plant height (cm) on fertilizer amount (grams).

$H_0 : \beta = 0$ (no linear relationship; fertilizer amount does not predict mean height) versus $H_a : \beta > 0$ (more fertilizer predicts greater mean height).

For 'any linear relationship' the alternative would be two-sided: $H_a : \beta \neq 0$.

COMPLETE TEST FROM OUTPUT

Use the fertilizer output (n = 14) to test $H_0 : \beta = 0$ against $H_a : \beta \neq 0$ at $\alpha = 0.05$. LINER has been checked.

Predictor	Coef	SE Coef	T-value	P-value
Constant	19.872	3.378	5.88	0.000
Fertilizer	1.873	0.513	3.65	0.003

SOLUTION

State: β = true slope of height on fertilizer. $H_0 : \beta = 0$, $H_a : \beta \neq 0$.

Plan: t test for a slope; random assignment of amounts (Random), residual plot shows no curve or fan, residuals roughly normal.

Do: $b = 1.873$, $SE_b = 0.513$, $t = 3.65$, df = 12. The two-sided P-value from the output is 0.003.

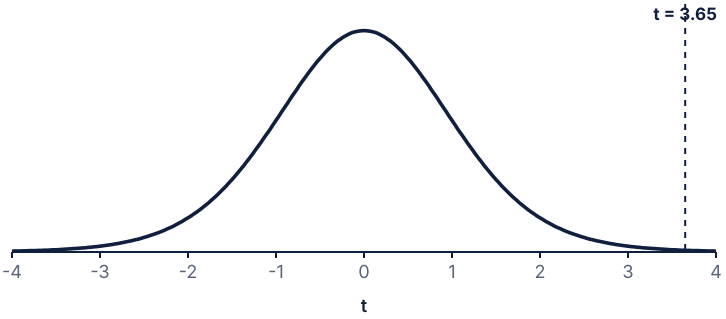
Conclude: Since 0.003 is less than 0.05, we reject H_0. There is convincing evidence of a linear relationship between fertilizer amount and mean plant height in the population.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

ONE-SIDED P-VALUE FROM OUTPUT

For the same data, test $H_a : \beta > 0$. What is the P-value, and what is the conclusion?



t distribution, df = 12: upper-tail area beyond the observed t

SOLUTION

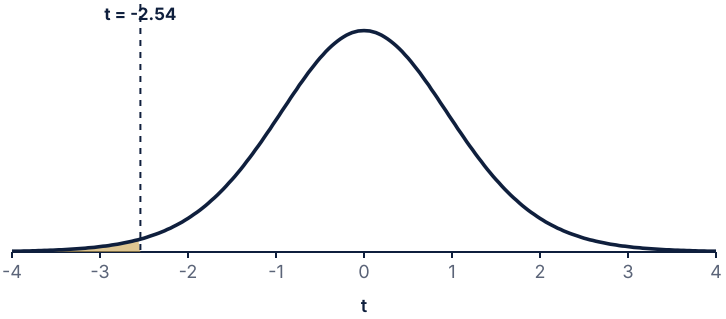
The output P-value 0.003 is two-sided. The sample slope 1.873 is positive, in the direction of H_a, so the one-sided P-value is half: about 0.0015.

Since 0.0015 is less than 0.05, we reject H_0. There is convincing evidence that the true slope is positive: more fertilizer is associated with a greater mean plant height.

If b had been negative, the P-value would exceed 0.5 and the data would provide no support for H_a.

TEST FROM SUMMARY STATISTICS

A random sample of 16 runners recorded weekly cross-training hours (x) and 5 km time in minutes (y). The slope is $b = -0.84$ minutes per hour, $s = 3.2$, $s_x = 2.5$. LINER is met. Test whether more cross-training is associated with a lower mean time at $\alpha = 0.05$.



Lower-tail area for the one-sided test, df = 14

SOLUTION

$H_0 : \beta = 0$, $H_a : \beta < 0$.

$SE_b = \dfrac{3.2}{2.5\sqrt{15}} = 0.330.$

$t = -0.84/0.330 = -2.54$, df = 14.

P-value $= P(t_{14} < -2.54) \approx 0.0117$ (built-in calculator: t cdf).

Since 0.0117 is less than 0.05, reject H_0. There is convincing evidence that the true slope is negative: more weekly cross-training is associated with a lower mean 5 km time.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

TESTING A VALUE OTHER THAN 0

A coach claims the fertilizer slope is 1.5 cm per gram. Test $H_0 : \beta = 1.5$ against $H_a : \beta \neq 1.5$ using the fertilizer data (b = 1.873, SE_b = 0.513, df = 12).

SOLUTION

$t = (b - 1.5)/SE_b = (1.873 - 1.5)/0.513 = 0.73$.

P-value (two-sided) $= 2P(t_{12} > |0.73|) \approx 0.480$.

This P-value is large. The data do not provide convincing evidence that the true slope differs from 1.5 cm per gram; the coach's claim is plausible. Note: the output P-value tests only $\beta = 0$, so for other values compute t yourself.

NO EVIDENCE OF A RELATIONSHIP

A random sample of 20 teens reported commute distance (km) and nightly sleep (hours). The output is shown below; LINER is met. Test for a linear relationship and give a 95 percent interval for the slope.

Predictor	Coef	SE Coef	T-value	P-value
Constant	7.2970	0.3543	20.60	0.000
Distance	-0.0235	0.0167	-1.41	0.177

SOLUTION

$H_0 : \beta = 0$, $H_a : \beta \neq 0$.

$t = -0.0235/0.0167 = -1.41$, df = 18, P-value = 0.177.

Since 0.177 is greater than 0.05, we fail to reject H_0. The data do not provide convincing evidence of a linear relationship between commute distance and sleep.

The 95 percent interval is (-0.0586, 0.0116) hours per km, which contains 0, consistent with the test. We cannot conclude that there is no relationship; the sample may be too small or the effect small.

Write It Like a 5

Model Response

Prompt. In the fertilizer experiment, 14 plants were randomly given different amounts of fertilizer. Regression of height on fertilizer gave b = 1.873, SE_b = 0.513, with residual plots showing no pattern and roughly normal residuals. Is there convincing evidence at the 5 percent level that a linear relationship exists? **State.** Let β be the true slope of the regression of plant height on fertilizer amount. $H_0 : \beta = 0$ and $H_a : \beta \neq 0$, $\alpha = 0.05$. **Plan.** t test for a slope. Random: amounts were randomly assigned. Linear and Equal variance: the residual plot shows no curve or fan. Normal: residuals are roughly symmetric with no outliers. Independent: plants are separate and the sample is a small part of all such plants. **Do.** $t = 1.873/0.513 = 3.65$, df = 12, P-value = 0.003. **Conclude.** Since the P-value 0.003 is less than 0.05, we reject H_0. There is convincing evidence of a linear relationship between the amount of fertilizer and mean plant height in the population.

What the grader looks for

1 point: parameter and both hypotheses in context. 1 point: names the t test for a slope and checks LINER with evidence. 1 point: correct t and df. 1 point: correct P-value. 1 point: comparison of P-value to alpha, decision about H_0, and a conclusion in context about a linear relationship in the population (not causation unless randomized).

Coach's Playbook

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. Write $H_0 : \beta = 0$ and the alternative in context before looking at the output.
2. Check LINER with residual plots and the design.
3. Read b, SE_b, t, and P from the explanatory-variable row. Halve the P-value for a one-sided test only if b has the right sign.
4. Compare P to alpha and write the conclusion about the population relationship in context.
5. If asked for causation, check whether the data come from a randomized experiment.

Common traps

1. Using the printed P-value for a one-sided alternative without halving it, or halving it when the sign is wrong. 2. Concluding 'there is no relationship' from a large P-value. 3. Using the Constant row instead of the explanatory-variable row. 4. Claiming causation from an observational study, even when the P-value is small.

Mistake Clinic

Spot the error

Student: 'The output P-value is 0.003, so for the one-sided test the P-value is also 0.003.' **Where it fails:** Output P-values are two-sided. **Correct:** For $H_a : \beta > 0$ with a positive slope, the P-value is 0.003/2 = 0.0015.

Spot the error

Student: 'P = 0.18 is greater than 0.05, so there is no linear relationship between commute distance and sleep.' **Where it fails:** A large P-value means insufficient evidence, not proof of no relationship. **Correct:** The data do not provide convincing evidence of a linear relationship.

Spot the error

Student: 'The slope is significant, so more fertilizer causes plants to grow taller.' **Where it fails:** Causation requires a well-designed randomized experiment. The statement should match the study design. **Correct:** Because amounts were randomly assigned in this experiment, the significant slope supports a causal link; for an observational study, only an association could be claimed.

Real-World Connection

Environmental scientists test whether the slope of temperature against year differs from zero; teachers test whether practice time predicts scores. The same t test applies, and the careful wording about association versus cause protects readers from over-claiming.

Challenge Corner

Problem. For the fertilizer data, show that the test of $H_0 : \rho = 0$ gives the same t as the test of $H_0 : \beta = 0$. **Solution.** With $r = 0.726$ and $n = 14$, $t = r\sqrt{n-2}/\sqrt{1-r^2} = 0.726\sqrt{12}/\sqrt{1-0.5268} = 3.65$, which equals the slope test statistic 3.65. The two tests are equivalent because the slope is zero exactly when the correlation is zero, and both use df = n - 2.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

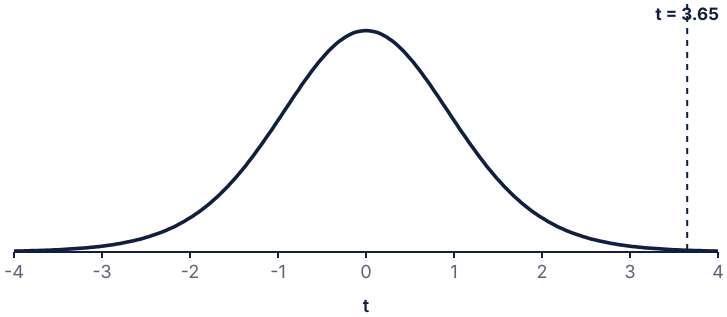
© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE Hypothesis Tests for the Slope

Show your work. A calculator is allowed. Justify conclusions in context.

1. The usual null hypothesis for a test about the slope of a population regression line is
   (A) $H_0 : b = 0$ **(B)** $H_0 : \beta = 0$ **(C)** $H_0 : r = 0$
   (D) $H_0 : \sigma = 0$
2. For a regression with n = 22 pairs, the test statistic $t = b/SE_b$ has how many degrees of freedom?
   (A) 22 **(B)** 21 **(C)** 19 **(D)** 20
3. The output shows Coef 2.5 and SE Coef 0.8 for the explanatory variable. The t statistic is
   (A) 3.13 **(B)** 1.7 **(C)** 0.32 **(D)** 2.0
4. The output P-value for a slope is 0.08 and the sample slope is positive. For $H_a : \beta > 0$ the P-value is
   (A) 0.16 **(B)** 0.08 **(C)** 0.04 **(D)** 0.92
5. The output P-value for a slope is 0.08 but the sample slope is negative. For $H_a : \beta > 0$ the P-value is closest to
   (A) 0.04 **(B)** 0.96 **(C)** 0.08 **(D)** 0.16
6. A test of $H_0 : \beta = 0$ gives P = 0.62. The best conclusion is
   (A) There is no linear relationship in the population
   (B) The relationship is strong but not linear
   (C) The slope is exactly 0
   (D) The data do not provide convincing evidence of a linear relationship
7. The shaded area below is the P-value for a test with n = 14. Which alternative does the shading represent?



t distribution, df = 12: upper-tail area beyond the observed t

   (A) $\beta < 0$ **(B)** $\beta > 0$ **(C)** $\beta \neq 0$ **(D)** $\beta = 0$
8. A researcher finds a very small P-value for the slope in an observational study. Which conclusion is supported?
   (A) The explanatory variable causes the response
   (B) The sample slope equals the population slope
   (C) There is convincing evidence of a linear association in the population
   (D) All individuals follow the line
9. The test of $H_0 : \beta = 0$ and the test of $H_0 : \rho = 0$ have
   (A) the same t statistic and the same P-value **(B)** different t statistics but the same P-value
   (C) the same t statistic but different df **(D)** different conclusions
10. For the commute-distance and sleep output (n = 20, b = -0.0235, SE = 0.0167):
    (a) State the hypotheses for a two-sided test.
    (b) Compute t and the degrees of freedom.
    (c) State a conclusion at $\alpha = 0.05$.
11. A random sample of 25 laptops records age in years (x) and repair cost in dollars (y). The regression gives $b = 18.4$ dollars per year, $s = 40$, and $s_x = 1.9$. LINER is met. Test $H_0 : \beta = 0$ versus $H_a : \beta > 0$ at $\alpha = 0.01$.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

12. A website claims the slope of time-on-site (minutes) versus page count is 2.0. A random sample of 30 sessions gives $b = 1.4$ and $SE_b = 0.35$. Test $H_0 : \beta = 2$ against $H_a : \beta \neq 2$. State a conclusion at the 5 percent level.
13. Explain why the P-value in a typical computer output must be adjusted for a one-sided test, and show how to adjust for $H_a : \beta < 0$ when the sample slope is negative and the printed P-value is 0.024.
14. A student states: 'P = 0.30, so we have proven that the slope is zero.' Identify two errors and write a correct conclusion in context of a study of weekly exercise hours and resting heart rate.
15. A study finds a significant positive slope between ice cream sales and number of sunburns in a sample of beach towns. Explain what the test does and does not allow us to conclude, and name a possible lurking variable.

T Unit 10 Test

Check your command of the regression model, LINER conditions, computer output, and t procedures for the slope.

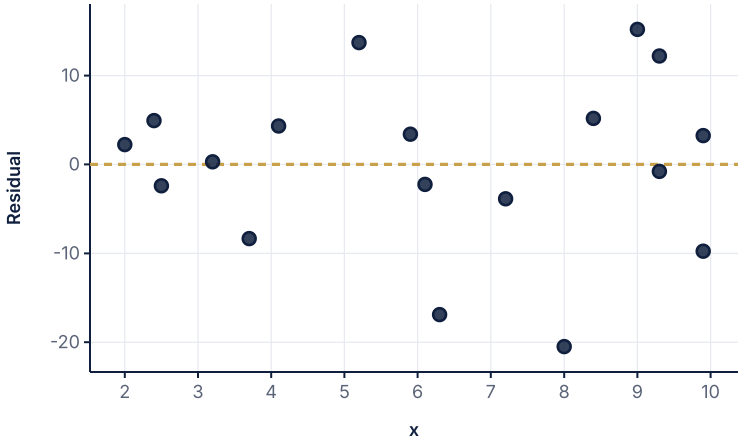
Time: about 55 minutes. Part A: 12 multiple choice (4 choices each). Part B: 2 free-response questions. Use a calculator. Score with the guide at the back of the unit.

Part A: Multiple Choice

UNIT TEST **Unit 10 Test**

Part A: choose the best answer for items 1 to 12. Part B: items 13 and 14 are free response; write full, in-context answers.

1. In the model $y = \alpha + \beta x + \varepsilon$, which statement about β is correct?
   (A) It is the slope of the sample regression line
   (B) It is the change in the mean response per one-unit increase in x in the population
   (C) It is the standard deviation of the residuals
   (D) It is always positive
2. A regression uses 16 data pairs. The t procedures for the slope use how many degrees of freedom?
   (A) 16 (B) 15 (C) 14 (D) 13
3. The residual plot shows a fan shape, with residuals spreading out as x increases. The violated condition is



Residual plot

   (A) Linear (B) Independent (C) Normal
   (D) Equal variance
4. From the output, the standard error of the slope is

Predictor	Coef	SE Coef	T-value	P-value
Constant	34.710	3.170	10.95	0.000
Reading	3.215	0.824	3.90	0.002

   (A) 3.215 (B) 0.824 (C) 6.200 (D) 34.71

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. From the same output, the test statistic for $H_0 : \beta = 0$ is closest to

Predictor	Coef	SE Coef	T-value	P-value
Constant	34.710	3.170	10.95	0.000
Reading	3.215	0.824	3.90	0.002

(A) 3.90 **(B)** 0.26 **(C)** 2.145 **(D)** 42.1

6. The critical value t^* for a 95 percent confidence interval for the slope with n = 16 is
(A) 1.960 **(B)** 2.131 **(C)** 2.145 **(D)** 2.120

7. The margin of error for a 95 percent interval for β using SE = 0.824 and $t^* = 2.145$ is closest to
(A) 1.77 **(B)** 0.82 **(C)** 3.22 **(D)** 1.62

8. A 95 percent confidence interval for the slope of vocabulary score on reading hours is (1.4, 5.0). The correct interpretation is
(A) 95 percent of students gain 1.4 to 5.0 points per hour
(B) We are 95 percent confident the mean score increases by 1.4 to 5.0 points per extra reading hour in the population
(C) The sample slope is between 1.4 and 5.0 with probability 0.95
(D) There is a 95 percent chance a student's score rises by 1.4 to 5.0

9. The printed two-sided P-value is 0.04 and the sample slope is positive. For $H_a : \beta > 0$ the P-value is
(A) 0.08 **(B)** 0.04 **(C)** 0.02 **(D)** 0.98

10. A test of $H_0 : \beta = 0$ gives P-value 0.41. The best conclusion is
(A) The slope equals 0
(B) The relationship is nonlinear
(C) There is no relationship of any kind
(D) The data do not provide convincing evidence of a linear relationship in the population

11. If the x values in a study are spread twice as widely (same s and n), the standard error of the slope will be about
(A) twice as large **(B)** half as large **(C)** the same
(D) four times as large

12. In the reading output, R-sq = 52.1 percent. The best interpretation is

S	R-sq	R-sq(adj)
6.201	52.1%	48.7%

(A) 52.1 percent of the variation in vocabulary scores is explained by the linear relationship with reading hours
(B) 52.1 percent of students read regularly
(C) The correlation is 0.521
(D) The slope is 52.1 percent

13. [6 points] Use the reading output (n = 16; Reading coefficient 3.215, SE Coef 0.824, T-value 3.90, P-value 0.002). The reading hours were measured for a random sample of students, and the residual plot shows no pattern with roughly normal residuals.
(a) [2 points] State the parameter and hypotheses to test whether there is a linear relationship between weekly reading hours and vocabulary score.
(b) [2 points] Give the test statistic, degrees of freedom, and P-value.
(c) [2 points] State a conclusion at the 5 percent level in context.

14. [10 points] A random sample of 25 tablets recorded battery age in months (x) and minutes of charge lost per day (y). Summary: $\bar{x} = 10$, $\bar{y} = 52$, $b = 2.8$, $s = 9.0$, $s_x = 3.4$. The residual plot shows random scatter with constant spread, a histogram of residuals is roughly symmetric, and the 25 tablets are under 10 percent of all tablets of this model.
(a) [2 points] Explain why the conditions for inference are met.
(b) [3 points] Find the standard error of the slope and a 95 percent confidence interval for β.
(c) [2 points] Interpret the interval in context.
(d) [1 point] Does the interval provide convincing evidence of a linear relationship? Explain.
(e) [2 points] Find the least-squares equation, the predicted loss for a 12-month-old tablet, and the residual for a tablet whose actual loss is 65 minutes.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Answers Inference for Regression Slopes: Answers and Explanations

Solutions for every practice problem and the unit test, numbered to match. Read each explanation, including the ones for problems you got right.

10.1 The Regression Model, Conditions, and Computer Output

PRACTICE · THE REGRESSION MODEL, CONDITIONS, AND COMPUTER OUTPUT

1. B. β is the parameter: the slope of the population regression line. The sample slope is b, a statistic.
2. C. $s = \sqrt{\sum(y-\hat{y})^2/(n-2)}$ estimates σ. The value b estimates β, and s_x describes the spread of the x values.
3. D. Regression estimates two parameters, so df = n - 2 = 18. The choice 19 incorrectly uses n - 1.
4. C. A U-shaped residual plot signals curvature: the relationship is not linear, so the Linear condition fails.
5. B. A fan shape means the spread of residuals changes with x, a failure of the Equal variance condition.
6. A. $SE_b = s/(s_x\sqrt{n-1})$ shrinks when s_x is large, i.e., when x values are spread out. Larger s or smaller n increases it.
7. C. The P-value on the explanatory variable's row is for the two-sided test of $H_0 : \beta = 0$. The intercept row tests alpha.
8. A. The standard error of the slope is in the SE Coef column of the Minutes row: 0.44. The value 2.3 is the slope itself.
9. B. The normal condition is about the residuals, which should be roughly normal around the line with equal spread. The x values need not be normal.
10. α: the mean skill score for teens who practice 0 minutes a day (the line's intercept for the population). β: the change in the MEAN skill score for each extra daily practice minute, across the whole population. σ: the standard deviation of individual skill scores around the population line at a given number of practice minutes.
11. (a) predicted score = 39.683 + 2.347(minutes).
    (b) For each additional daily practice minute, the mean predicted skill score increases by about 2.347 points.
    (c) df = n - 2 = 10 and $SE_b = 0.409$, the typical amount the sample slope differs from the true slope.
12. Linear: residual plot should show no curve. Independent: study design (random sample, under 10 percent of the population). Normal: histogram, dotplot, or normal probability plot of the residuals should be roughly symmetric with no outliers. Equal variance: the residual plot should have a constant vertical spread (no fan). Random: random sampling or random assignment.
13. (a) $SE_b = 4.5/(3\sqrt{24}) = 0.306$.
    (b) With n = 100: $4.5/(3\sqrt{99}) = 0.151$. A larger sample gives a smaller standard error and therefore a more precise estimate of the slope. The standard error shrinks about as $1/\sqrt{n}$.
14. The student is wrong. Curvature violates the Linear condition even when r is high. A high r only tells us the points are close to some straight line pattern; the residual plot reveals that a line is the wrong model. Inference for a slope would be unreliable, and a transformation or a curved model should be considered.
15. The slope b is unbiased for β: across many random samples, the sample slopes average to the population slope. In a simulation of many samples, a histogram of b is centered at β. The spread (SE) gets smaller when the sample size is larger, when the x values are more spread out, or when the points lie closer to the line (smaller s).

10.2 Confidence Intervals for the Slope

PRACTICE · CONFIDENCE INTERVALS FOR THE SLOPE

1. C. Slope inference uses df = n - 2 = 23. The choice 24 uses n - 1.
2. B. The interval uses the t distribution and the standard error of the slope SE_b. The standard deviation s alone is not the standard error of b.
3. A. With n = 12, df = 10 and $t^* = 2.228$ for 95 percent. The value 2.201 corresponds to df = 11, which would result from using n - 1.
4. D. The interval is about the MEAN change in the population (the true slope), with confidence in the method. It is not a range for individuals or for sample slopes.
5. C. Because 0 is inside the interval, a zero slope is plausible, so we lack convincing evidence of a linear relationship. Failing to reject is not proof of no relationship.
6. B. A larger confidence level needs a larger t^*, which widens the interval (and gives less precision).
7. C. The standard error is smallest when n is large and the x values are widely spread, since $SE_b = s/(s_x\sqrt{n-1})$.
8. A. df = 12 and $t^* = 3.055$ for 99 percent, so the margin of error is $3.055(0.21) = 0.64$. The value 0.45 comes from the 95 percent critical value 2.179.
9. B. A fan shape violates Equal variance, so a single s cannot describe the spread at every x and the t procedure is unreliable. Switching to z* or changing the confidence level does not fix this.
10. (a) df = 8, $t^* = 1.860$. (b) df = 28, $t^* = 2.048$. (c) df = 16, $t^* = 2.921$.
11. (a) df = 12, $t^* = 3.055$, margin of error $= 3.055(0.401) = 1.223$. Interval: $1.459 \pm 1.223 = (0.236, 2.683)$.
    (b) We are 99 percent confident that the true slope is between 0.236 and 2.683: each extra unit of dose is associated with an increase in the MEAN response of that amount in the population.
12. $SE_b = 5.0/(1.8\sqrt{19}) = 0.637$. df = 18, $t^* = 2.101$. Margin of error $= 1.339$. Interval: $-2.4 \pm 1.339 = (-3.739, -1.061)$. We are 95 percent confident that each extra insulation layer is associated with a decrease of between 1.061 and 3.739 energy units in the mean energy use of buildings of this type.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. The interval is a statement about the true slope β, the change in MEAN weight per inch for the population, not about the weights of 95 percent of individual people. Better: 'We are 95 percent confident that the mean weight in the population increases by 0.4 to 1.9 pounds for each extra inch of height.'
14. (a) Larger n decreases SE_b (and t^*), so the interval is narrower. (b) A higher confidence level increases t^*, so the interval is wider. (c) More spread in x increases s_x and decreases SE_b, so narrower. (d) A larger residual SD increases SE_b and widens the interval.
15. Yes, 0 is in the interval. A zero slope is plausible, so there is no convincing evidence of a linear relationship between temperature and laps. A small sample has a large SE_b and a wide interval, so even a real nonzero slope can produce an interval containing 0. Failing to find evidence is not the same as showing no relationship.

10.3 Hypothesis Tests for the Slope

PRACTICE · HYPOTHESIS TESTS FOR THE SLOPE

1. B. The null hypothesis refers to the population parameter β, not the sample statistic b. $H_0 : \beta = 0$ says there is no linear relationship.
2. D. For slope inference df = n - 2 = 20. Using n - 1 = 21 is a typical error.
3. A. $t = 2.5/0.8 = 3.125$. The value 0.32 is SE over b.
4. C. The printed P-value is two-sided. The slope has the sign predicted by H_a, so the one-sided P-value is half: 0.04.
5. B. The slope is on the wrong side for $H_a : \beta > 0$, so the one-sided P-value is 1 - 0.04 = 0.96, which is no evidence for H_a. Halving alone (0.04) would be wrong.
6. D. A large P-value means insufficient evidence; it never proves the null. The slope could be small or the sample too small to detect it.
7. B. The shaded area is the upper tail beyond a positive t, which corresponds to $H_a : \beta > 0$.
8. C. A small P-value gives evidence of a linear association. Causation requires a randomized experiment, and a sample slope does not equal the population slope.
9. A. They are equivalent: the t statistic and df (n - 2) are identical, so the P-values and conclusions are the same.
10. (a) $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$, where β is the true slope of sleep on commute distance.
    (b) $t = -0.0235/0.0167 = -1.41$, df = 20 - 2 = 18.
    (c) The P-value 0.177 is above 0.05, so we fail to reject H_0. The data do not provide convincing evidence of a linear relationship between commute distance and nightly sleep.
11. $SE_b = 40/(1.9\sqrt{24}) = 4.30$. $t = 18.4/4.30 = 4.28$, df = 23. P-value $= P(t_{23} > 4.28) \approx 0.0001$. Since 0.0001 is less than 0.01, reject H_0. There is convincing evidence that the true slope is positive: older laptops tend to have higher mean repair costs.
12. $t = (1.4 - 2.0)/0.35 = -1.71$, df = 28. Two-sided P-value ≈ 0.098. Since 0.098 is greater than 0.05, we fail to reject H_0. The data do not provide convincing evidence that the true slope differs from 2.0 minutes per page; the claim is plausible.
13. Software prints the two-sided P-value, which includes both tails. A one-sided alternative counts only one tail, so if the sample slope is in the direction of H_a, divide by 2. Here the slope is negative, matching $H_a : \beta < 0$, so the P-value is 0.024/2 = 0.012.
14. Error 1: A large P-value never proves the null; it only means the data do not provide convincing evidence against it. Error 2: a conclusion must be about the population in context. Correct: Since P = 0.30 is above the usual alpha, we fail to reject H_0. The data do not provide convincing evidence of a linear relationship between weekly exercise hours and resting heart rate in the population.
15. The test shows evidence of a linear association between sales and sunburns in the population of beach towns. It does not show that ice cream causes sunburns, because this is observational. A lurking variable such as hot, sunny weather (or number of visitors) can increase both.

T Unit 10 Test

UNIT TEST · UNIT 10 TEST

1. B. β is the population slope: the change in the MEAN response per unit of x. The sample slope is b, and the SD of residuals is σ.
2. C. Regression estimates two parameters, so df = n - 2 = 14. The choice 15 uses n - 1.
3. D. A fan shape means the residual spread changes with x, violating Equal variance. A curve would signal a Linear violation.
4. B. The standard error of the slope is in the SE Coef column of the Reading row: 0.824. The value 3.215 is the slope itself.
5. A. $t = b/SE_b = 3.215/0.824 = 3.90$. The value 0.26 is the reciprocal, and 2.145 is a critical value.
6. C. df = 14 and $t^* = 2.145$ for 95 percent. The value 1.960 is z*, and 2.131 belongs to df = 15 (using n - 1).
7. A. $2.145 \times 0.824 = 1.77$. Using only SE (0.82) forgets to multiply by t^*.
8. B. The interval estimates the MEAN change in the population (the true slope) with 95 percent confidence in the method. It does not describe individuals or the sample slope.
9. C. The printed P-value is two-sided; the slope is in the direction of H_a, so halve it: 0.02.
10. D. A large P-value means we lack convincing evidence. It does not prove the slope equals 0 or that no relationship exists.
11. B. $SE_b = s/(s_x\sqrt{n-1})$, so doubling s_x halves the standard error.
12. A. R-sq is the percent of the variation in the response explained by the linear model. It is not the correlation, which is r (the square root of R-sq, with the sign of the slope).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

13. Scoring (6 points). (a) 2 points: β defined as the true slope of vocabulary score on weekly reading hours (1); $H_0 : \beta = 0$, $H_a : \beta \neq 0$ (1). (b) 2 points: $t = 3.215/0.824 = 3.90$ (1); df = 14 and P-value 0.002 (1). (c) 2 points: compares P-value to 0.05 and rejects H_0 (1); conclusion in context about a linear relationship in the population (1).

Model response. (a) Let β be the true slope of the regression line of vocabulary score on weekly reading hours for all students in the population. $H_0 : \beta = 0$; $H_a : \beta \neq 0$. (b) $t = 3.90$, df = 16 - 2 = 14, P-value = 0.002. (c) Since 0.002 is less than 0.05, we reject H_0. There is convincing evidence of a linear relationship between weekly reading hours and vocabulary score for students in the population.

14. Scoring (10 points). (a) 2 points: Linear and Equal variance from the residual plot (1); Normal from the residual histogram, Independent from under 10 percent, Random from random sample (1). (b) 3 points: $SE_b = 9.0/(3.4\sqrt{24}) = 0.5403$ (1); df = 23, $t^* = 2.069$ (1); interval $2.8 \pm 2.069(0.5403) = (1.682, 3.918)$ (1). (c) 2 points: states 95 percent confidence (1); the true slope of the population line is between 1.682 and 3.918 minutes of charge lost per month of battery age, referring to the MEAN (1). (d) 1 point: yes, the interval is entirely above 0, so zero slope is not plausible. (e) 2 points: $a = 52 - 2.8(10) = 24$, so $\hat{y} = 24 + 2.8x$; at 12 months $\hat{y} = 57.6$ (1); residual $= 65 - 57.6 = 7.4$ minutes (1).

Model response. (a) The residual plot shows no pattern (Linear) and constant spread (Equal variance); the residual histogram is roughly symmetric (Normal); the tablets are a random sample and under 10 percent of all tablets (Random, Independent). (b) $SE_b = 0.5403$; df = 23; $t^* = 2.069$; the interval is (1.682, 3.918). (c) We are 95 percent confident that the true slope is between 1.682 and 3.918: each additional month of battery age is associated with an increase of that many minutes of charge lost per day in the mean for all tablets of this model. (d) Yes. The whole interval is above 0, so there is convincing evidence of a positive linear relationship. (e) $\hat{y} = 24 + 2.8x$. For x = 12, $\hat{y} = 57.6$ minutes. The residual is 65 - 57.6 = 7.4 minutes; this tablet lost more than the model predicts.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Capstone: Choosing Procedures and Investigative Tasks

Decide which procedure fits a question, then combine skills from every unit.

C.1 Which Procedure Do I Use? The Decision Guide

Use a four-question path to choose among one-proportion, two-proportion, one-sample, two-sample and paired procedures, chi-square and slope extensions, and probability models, then justify your choice in context.

Why This Matters

The hardest part of an inference question is often not the calculation but the first decision: which procedure fits this situation? A student who chooses the wrong procedure can compute everything correctly and still earn almost no credit. The exam expects you to choose among proportion and mean procedures, one-sample, two-sample and paired designs, and to explain the choice in words.

This guide pulls together Exam Units 1 to 5 into one decision path. Use it after finishing Units 7 and 8, and again when you work practice exams. The path always starts with the question being asked, then the type of variable, then how the data were collected.

Core Ideas

The four-question path

Four questions to choose a procedure

1. What is the goal? Describe data (graphs and summaries), find a probability from a model, estimate a parameter (interval), or test a claim (significance test).

2. What type of variable? Categorical (counts, proportions) leads to z procedures for proportions or chi-square. Quantitative (measurements) leads to t procedures for means, or regression for two quantitative variables.

3. How many groups and how are they related? One sample, two independent samples (or groups from randomly assigned treatments), or paired/matched data (the same individuals measured twice, or natural pairs).

4. Are the conditions met? Random sample or random assignment, independence (10% rule), and normal/large enough sample (large counts for proportions; no strong skewness or outliers for small samples).

Situation	Procedure	Parameter	Key conditions	Calculator function	Book lesson
One categorical variable, one sample (estimate)	One-proportion z-interval	p	Random; 10% rule; $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$	1-PropZInt	7.2
One categorical variable, one sample (test a claim)	One-proportion z-test	p	Random; 10% rule; $np_0 \geq 10$ and $n(1-p_0) \geq 10$	1-PropZTest	7.4
Compare two proportions	Two-proportion z-interval or test	$p_1 - p_2$	Two independent random samples or random assignment; large counts in each group	2-PropZInt, 2-PropZTest	7.6
One quantitative variable, one sample	One-sample t-interval or test	μ	Random; 10% rule; normal or $n \geq 30$, no strong skewness or outliers	TInterval, T-Test	8.2, 8.3
Compare two means, independent groups	Two-sample t-interval or test	$\mu_1 - \mu_2$	Independent random samples or random assignment; each group normal or large	2-SampTInt, 2-SampTTest	8.4, 8.5

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Situation	Procedure	Parameter	Key conditions	Calculator function	Book lesson
Same individuals measured twice, or natural pairs	Paired (matched pairs) t procedure on the differences	μ_d	Random; differences normal or $n \geq 30$	TInterval, T-Test on the list of differences	8.6

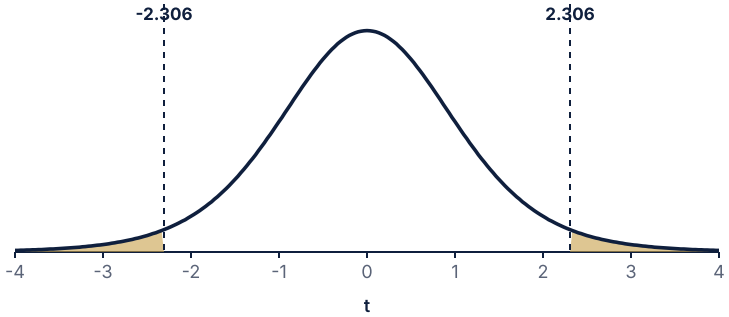
Writing the choice

Always write the procedure by name and give the reason in one phrase: "two-sample t-test, because the groups are independent and the response is quantitative."

Then define the parameter, write the hypotheses (or the interval target), and check the conditions before any calculation.

Extension procedures. The following appear in college courses and some older exams. The College Board revision page names the chi-square goodness-of-fit test and inference for regression slopes as removed for 2027. For chi-square homogeneity and independence, the revision page names only goodness of fit as removed, but the 2027 unit titles have no chi-square unit, so confirm with your teacher.

Situation	Procedure	Book lesson
One categorical variable compared with a claimed distribution	Chi-square goodness of fit (beyond the 2027 core)	9.1
One categorical variable, several populations or treatments	Chi-square test for homogeneity	9.2
Two categorical variables, one sample	Chi-square test for independence	9.3
Two quantitative variables, estimate or test a slope	t-interval or t-test for slope (beyond the 2027 core)	Unit 10



t procedures use a distribution with heavier tails than the normal. With df = 8, the 95% critical value is t* = 2.306, compared with z* = 1.96.

Cue words and traps in the wording

If the problem says	It probably means
"proportion," "percent," "fraction who"	Categorical; z procedure for proportions
"average," "mean," "typical measurement"	Quantitative; t procedure for means
"before and after," "twins," "each person tries both"	Paired data; analyze the differences
"two randomly selected groups," "randomly assigned to treatments"	Two independent groups; two-sample procedure
"are these two variables related?" (both categorical)	Two-way table; chi-square independence (confirm with your teacher)
"how likely is it that in a group of n exactly k..."	Probability with a binomial model, not inference
"is there convincing evidence"	A significance test
"estimate," "how much," "between what values"	A confidence interval

Probability versus inference. If the population proportion or mean is known and you ask about a sample, you are doing probability (a binomial, a normal, or a sampling distribution). If the population value is unknown and you use a sample to learn about it, you are doing inference.

Worked Examples

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 1: MATCHING SCENARIOS TO PROCEDURES

For each scenario, name the parameter and procedure: (a) A school wants to estimate the percent of students who walk to school. (b) A coach tests whether a new drill reduces mean sprint time, with each athlete timed before and after. (c) A teacher compares the average quiz score of two randomly chosen classes. (d) A manager asks whether more than 10% of invoices contain errors.

SOLUTION

(a) Parameter p, the true proportion who walk. One-proportion z-interval (estimate, categorical, one sample).

(b) Parameter μ_d, the true mean difference (after minus before). Paired t-test, because the same athletes are measured twice.

(c) Parameter $\mu_1 - \mu_2$. Two-sample t-interval or test for independent groups.

(d) Parameter p, with $H_0 : p = 0.10$ and $H_a : p > 0.10$. One-proportion z-test.

EXAMPLE 2: PAIRED T-TEST

Eight students take a timed practice quiz, then take a similar quiz after a review session. Before: 72, 65, 80, 58, 69, 75, 62, 70. After: 75, 66, 86, 60, 74, 79, 63, 76. Is there convincing evidence that the review improves scores? Assume the students are a random sample and the differences show no strong skewness or outliers.

SOLUTION

The same students are measured twice, so use a paired t-test on differences (after minus before): 3, 1, 6, 2, 5, 4, 1, 6.

Parameter: μ_d, the true mean improvement. $H_0 : \mu_d = 0$, $H_a : \mu_d > 0$.

$\bar{d} = 3.5$, $s_d = 2.07$, $n = 8$, $t = \frac{3.5}{2.07/\sqrt{8}} = 4.78$, $df = 7$, p-value ≈ 0.001.

Conclusion: Since 0.001 < 0.05, we reject H_0. There is convincing evidence that the mean score improves after the review session. A 95% interval for μ_d is (1.77, 5.23) points.

EXAMPLE 3: TWO-PROPORTION Z-TEST

In a random sample of 120 students taking Course A, 45 are enrolled in a club. In a random sample of 110 students taking Course B, 30 are enrolled in a club. Test whether the proportion in Course A is greater than the proportion in Course B at alpha = 0.05.

SOLUTION

Two independent random samples, categorical response: two-proportion z-test. $H_0 : p_A = p_B$, $H_a : p_A > p_B$.

Pooled proportion: $\hat{p}_c = \frac{45+30}{230} = 0.326$. Large counts: all of $120(0.326) = 39.1$, $120(0.674) = 80.9$, $110(0.326) = 35.9$, $110(0.674) = 74.1$ are at least 10.

$\hat{p}_A = 0.375$, $\hat{p}_B = 0.273$, $z = \frac{0.375 - 0.273}{\sqrt{0.326(0.674)(1/120 + 1/110)}} = 1.65$, p-value $= 0.049$.

Conclusion: Since 0.049 < 0.05, we reject H_0. There is convincing evidence that the proportion enrolled in a club is greater for Course A than for Course B (the result is borderline, so a larger study would help).

EXAMPLE 4: TWO-SAMPLE T-INTERVAL

Random samples of students from two schools took a reading speed test. School 1: n = 20, mean 52.1 words per minute (wpm), s = 6.8. School 2: n = 22, mean 47.3, s = 7.9. Assume both samples show no strong skewness or outliers. Construct a 95% interval for the difference in true means and interpret it.

SOLUTION

Two independent quantitative samples: two-sample t-interval for $\mu_1 - \mu_2$.

$SE = \sqrt{\frac{6.8^2}{20} + \frac{7.9^2}{22}} = 2.269$, and the calculator gives $df \approx 39.9$, so $t^* \approx 2.02$.

Interval: $(52.1 - 47.3) \pm 2.02(2.269) = (0.21, 9.39)$.

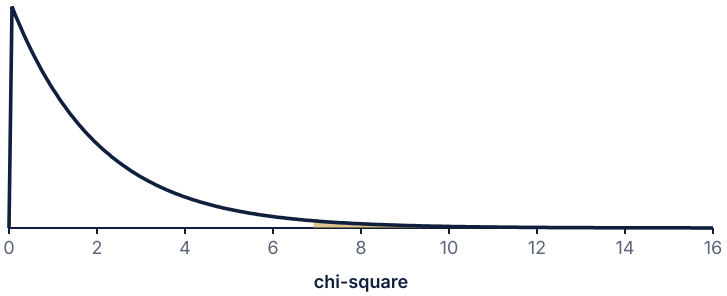
Interpretation: We are 95% confident that the true mean reading speed at School 1 is between 0.21 and 9.39 wpm higher than at School 2. Since 0 is not in the interval, this suggests the means differ (two-sided p-value ≈ 0.041).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

EXAMPLE 5: CHI-SQUARE TEST FOR INDEPENDENCE (EXTENSION)

A survey of 175 randomly chosen students records preferred study place and grade band. Counts: Library, Home, Cafe for lower grades 38, 22, 20 and for upper grades 27, 33, 35. Is there convincing evidence of an association? (Confirm with your teacher whether chi-square tests are in your course.)



p-value is the area to the right of 6.92

SOLUTION

One sample, two categorical variables: chi-square test for independence. H_0: study place and grade band are independent. H_a: they are associated.

Expected counts: row 1 total 80, row 2 total 95, column totals 65, 55, 55. Expected for row 1: 29.7, 25.1, 25.1; row 2: 35.3, 29.9, 29.9. All are at least 5.

$\chi^2 = \sum \frac{(O-E)^2}{E} = 6.92$, $df = (2-1)(3-1) = 2$, p-value ≈ 0.031.

Conclusion: Since 0.031 < 0.05, we reject H_0. There is convincing evidence of an association between grade band and preferred study place.

EXAMPLE 6: PROBABILITY OR INFERENCE?

A machine's flaw rate is known to be 5% on long runs. In a box of 40 independently produced parts, find the probability of at least 4 flawed parts. Then, separately, suppose a new machine's rate is unknown, and in a random sample of 80 parts, 52 pass inspection; give a 95% interval for the pass rate.

SOLUTION

First part: the flaw rate is known, so this is probability. $X \sim \text{Binomial}(40, 0.05)$. $P(X \geq 4) = 1 - \text{binomcdf}(40, 0.05, 3) = 0.138$.

Second part: the pass rate is unknown, so this is inference. One-proportion z-interval: $\hat{p} = 52/80 = 0.65$, large counts $52 \geq 10$ and $28 \geq 10$, $SE = 0.0533$, interval $0.65 \pm 1.96(0.0533) = (0.545,\ 0.755)$.

Interpretation: We are 95% confident that the true pass rate for the new machine is between 0.545 and 0.755.

The same scenario type, different goal: the first uses a **known** parameter to predict a sample, the second uses a sample to learn about an **unknown** parameter.

Write It Like a 5

Prompt: A trainer recruits 40 volunteer runners and randomly assigns 20 to Stretch Routine A and 20 to Stretch Routine B. After four weeks each runner's time in a 5K run is recorded. The trainer wants to know whether the mean time differs between the routines. Identify the appropriate inference procedure, state the hypotheses, and list the conditions you would check. What kind of conclusion is justified if the test is significant?

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model Response

Procedure: two-sample t-test for a difference in means, because there are two independent groups created by random assignment and the response (5K time) is quantitative. **Hypotheses:** let μ_A and μ_B be the true mean 5K times that would result from Routine A and B for runners like these. $H_0 : \mu_A - \mu_B = 0$ and $H_a : \mu_A - \mu_B \neq 0$. **Conditions:** (1) Random: the treatments were randomly assigned. (2) Independence: runners are assigned independently, and the group sizes (20 and 20) are not compared with a larger finite population. (3) Normal or large: each group has $n = 20 < 30$, so a graph of each group's times should show no strong skewness or outliers. **Conclusion:** If the p-value is less than alpha, there is convincing evidence that the routines give different mean times, and because treatments were randomly assigned we may conclude cause and effect. However, the volunteers were not randomly selected, so the conclusion applies only to runners like these volunteers.

What the grader looks for

(1) Correct procedure (two-sample t, not paired and not a z procedure). (2) Parameters defined in context and correct hypotheses. (3) Random assignment, independence and normality conditions, with the graph check for small groups. (4) A two-sided alternative. (5) Scope of inference: cause and effect from random assignment, limited generalization because of volunteers.

Coach's Playbook

1. **Underline the goal.** Estimate or test? If the problem says "convincing evidence," it is a test. If it says "estimate," "how large," or "between what values," it is an interval.
2. **Classify the variable.** Say aloud: categorical or quantitative? This single answer eliminates half of the procedures.
3. **Count the groups.** One group, two independent groups, or the same individuals measured twice. If you can pair each value in one list with a specific value in the other, it is paired.
4. **Name the parameter.** Write the symbol and a phrase: p for a proportion, μ for a mean, $p_1 - p_2$ or $\mu_1 - \mu_2$ for differences, μ_d for paired.
5. **Check conditions with numbers.** Write each condition and the check (for example, counts at least 10, or a graph of the data).
6. **Choose the function and write it.** Name the calculator function and the inputs, then compute and conclude in context.

Common traps

1. Using a two-sample procedure for paired data, which ignores the pairing and usually shrinks the t-statistic. 2. Using a z procedure for a mean (or t for a proportion). 3. Using inference when the parameter is known (use probability instead). 4. Claiming cause and effect from an observational study, or generalizing a volunteer sample to a whole population.

Mistake Clinic

Spot the error

A student treats the before and after quiz scores of the same eight students as two independent samples and runs a two-sample t-test, getting t = 0.88 and p = 0.39. Where does this fail? The scores are paired by student. The students' ability varies far more than the improvement does, so ignoring the pairing hides the effect. Correct: paired t on differences gives t = 4.78 and p = 0.001.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Spot the error

A student uses a one-proportion z-test to answer: "Is the mean number of hours of sleep different from 8?" Where does this fail? Hours of sleep is a quantitative variable and the parameter is a mean. Correct: a one-sample t-test for μ with $H_0 : \mu = 8$.

Spot the error

A student reads "A random sample of 50 voters found 24 in favor, and a second random sample of 60 voters from another city found 33 in favor," and uses a paired t-test. Where does this fail? The samples are independent and the response is categorical. Correct: two-proportion z procedure for $p_1 - p_2$.

Real-World Connection

In clinical research, marketing and engineering, the analysis plan is chosen before the data are collected, and the first questions are exactly the four above. An analyst who picks the wrong test costs money or risks harm, which is why teams document the choice.

Challenge Corner

Using the eight before-and-after scores from Example 2, run a two-sample t-test (not paired) and compare the result with the paired result. Explain why they differ. **Solution:** The before scores have mean 68.9 and the after scores have mean 72.4. The two-sample t-statistic is $t = 0.88$ with $df \approx 13.4$ and a two-sided p-value of 0.39, so no evidence of a difference. The paired result has $t = 4.78$. The difference arises because the two score lists are strongly correlated (r about 0.99): students who scored high before also scored high after, so differences between students swamp the improvement unless you analyze each student's difference. The pairing removes between-student variation.

Practice

PRACTICE **Which Procedure Do I Use? The Decision Guide**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A random sample of 80 voters finds that 52 support a ballot measure. A poll director wants to estimate the proportion of all voters who support it. Which procedure is appropriate?
   (A) Two-sample t-interval **(B)** One-sample t-interval
   (C) One-proportion z-interval **(D)** Chi-square test for independence
2. Thirty randomly chosen batteries of Brand X and 30 randomly chosen batteries of Brand Y are tested for lifetime in hours. To compare mean lifetimes, use
   (A) a paired t-test **(B)** a two-sample t-test **(C)** a two-proportion z-test
   (D) a one-proportion z-test
3. Ten volunteers have their resting pulse measured before and after a week of a breathing exercise. To test whether the exercise lowers mean pulse, use
   (A) a two-sample t-test **(B)** a two-proportion z-test
   (C) a chi-square test **(D)** a paired t-test on the differences
4. In an experiment, 90 patients are randomly assigned to a new plan and 85 to the standard plan, and 66 and 51 of them improve, respectively. The appropriate procedure to compare success rates is
   (A) a two-proportion z-test **(B)** a two-sample t-test **(C)** a paired t-test
   (D) a one-sample t-test

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

5. A quality inspector measures the diameter of 25 randomly selected parts and wants to test whether the mean diameter equals 5.0 cm. Which procedure fits?

   (A) One-proportion z-test
   (B) One-sample t-test for a mean
   (C) Two-sample t-test
   (D) Chi-square test for homogeneity

6. A random sample of students is classified by grade band (3 levels) and preferred lunch type (3 types). To test for an association, which is the usual procedure (confirm with your teacher whether chi-square is in your course)?

   (A) Chi-square test for independence
   (B) Two-sample t-test
   (C) Paired t-test
   (D) One-proportion z-test

7. Which statement correctly distinguishes a paired design from a two-sample design?

   (A) In a paired design, each value in one list is naturally matched with a value in the other, and you analyze the differences
   (B) In a paired design, the two groups must be the same size
   (C) In a two-sample design, the same subjects are always measured twice
   (D) The difference is only that paired uses z

8. Which study supports a cause-and-effect conclusion?

   (A) Students who choose to eat breakfast are compared with those who do not
   (B) A comparison of last year's and this year's averages
   (C) A survey of randomly selected adults about diet and weight
   (D) An experiment in which volunteers are randomly assigned to different diets

9. A machine produces flawed parts at a known rate of 10%. In a sample of 15 parts, what is the best way to find the probability of exactly 2 flawed parts?

   (A) A one-proportion z-test
   (B) Binomial probability, binompdf(15, 0.1, 2)
   (C) A one-sample t-interval
   (D) A normal probability using the sample mean

10. For each situation, name the parameter and the appropriate procedure.

    (a) A city wants to estimate the average commute time of its workers from a random sample of 150 workers.

    (b) A teacher tests whether two teaching methods lead to different proportions of students passing, using two randomly assigned classes.

    (c) A study measures the weight of 12 people before and after a program and tests for a decrease.

11. Ten students take a similar pre-test and post-test. Pre: 68, 72, 75, 80, 66, 71, 74, 78, 69, 73. Post: 71, 75, 74, 86, 69, 76, 75, 83, 72, 78. Assume the students are a random sample and the differences show no strong skewness.

    (a) Explain why a paired procedure is appropriate.

    (b) Test whether the mean score improved (alpha = 0.05).

    (c) Give a 95% confidence interval for the mean improvement.

12. In a randomized experiment, 90 volunteers receive Plan X and 85 receive Plan Y. Of those on Plan X, 66 improve; of those on Plan Y, 51 improve.

    (a) Check the conditions for a two-proportion z-test.

    (b) Test $H_0 : p_X = p_Y$ against $H_a : p_X > p_Y$ at alpha = 0.05.

    (c) Give a 90% confidence interval for $p_X - p_Y$.

13. Random samples of fitness-class members: Group 1 (n = 29, mean 10.2 minutes, s = 2.1) and Group 2 (n = 31, mean 12.0 minutes, s = 2.9) record time to complete a task. Assume both samples show no strong skewness or outliers.

    (a) Which procedure should be used? (b) Test whether the true means differ at alpha = 0.05. (c) Construct and interpret a 95% interval for $\mu_1 - \mu_2$.

14. A random sample of 100 adults is classified by sleep (adequate or short) and exercise (regular or rare). Counts: regular and adequate 30, regular and short 20, rare and adequate 18, rare and short 32. (Confirm with your teacher whether chi-square is in your course.)

    (a) Name the test and state the hypotheses.

    (b) Compute the expected counts and verify the condition.

    (c) Find the test statistic, degrees of freedom and p-value, and conclude.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

15. Researchers randomly assign 60 volunteers to Diet A or Diet B and measure each person's weight change in kilograms after 12 weeks; 58 finish. In the final data, Diet A (n = 29) has mean change -4.8 kg with s = 3.1 and Diet B (n = 29) has mean -2.9 kg with s = 3.4.

(a) Which procedure is appropriate and why?

(b) Carry out a test of $H_0 : \mu_A - \mu_B = 0$ versus $H_a : \neq 0$.

(c) What conclusion about cause and effect and about generalization is justified?

C.2 Investigative Tasks: Combining Skills Across Units

Work through six complete investigations that combine study design, graphs, probability models and inference, then practice writing coherent conclusions that respect what the design can and cannot show.

Why This Matters

Real statistical work does not arrive labeled "Unit 7." A single question, such as whether a fertilizer works or whether a policy is popular, requires you to design data collection, graph the data, build a probability model, and run an inference, and then communicate a single coherent conclusion. The free-response section does this by design: a question may ask you to critique a study design in part (a), analyze a graph in part (b), and test a claim in part (c).

This capstone lesson practices that integration. It reaches across all five exam units: data and design (Unit 1), probability and sampling distributions (Unit 2), proportions (Unit 3), means (Unit 4) and regression (Unit 5). Expect to use earlier lessons constantly and to make the decisions that Lesson C.1 describes.

Core Ideas

The investigation cycle

A five-stage investigation

1. Question and design: What is the parameter or relationship? Is the study an experiment or observational? How were units chosen or assigned? Name sources of bias and confounding.

2. Display and describe: Graph the data (boxplot, histogram, scatterplot, bar chart) and describe it with SOCS or form, direction, strength and outliers.

3. Model: Choose a probability model (normal, binomial, sampling distribution) only if its conditions hold.

4. Infer: Choose the procedure (Lesson C.1), check conditions, compute, and interpret a p-value or interval.

5. Conclude and critique: State the conclusion in context, limited by the design: causation only from random assignment, generalization only to the population sampled.

Scope of inference rules

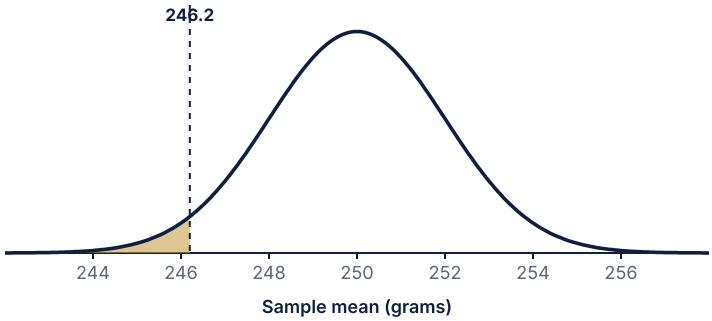
Random sampling from a population allows generalizing to that population. **Random assignment** to treatments allows a cause-and-effect conclusion.

Both present: generalize and conclude causation. Only random assignment: cause and effect for the units in the study. Only random sampling: generalize, but report association only. Neither: describe only these units.

Stage	Questions to ask	Typical exam wording
Design	How were individuals selected or assigned? Could bias or confounding exist?	"Describe a design," "identify a source of bias"
Graph	What does the shape, center, spread or pattern show? Any outliers?	"Compare the distributions," "describe the association"
Model	Is a binomial or normal model reasonable? What are the parameters?	"Find the probability that..."
Infer	Which procedure? Do the conditions hold?	"Is there convincing evidence...," "construct and interpret..."
Conclude	What can and cannot be said?	"Explain whether the result can be generalized or is evidence of cause"

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.



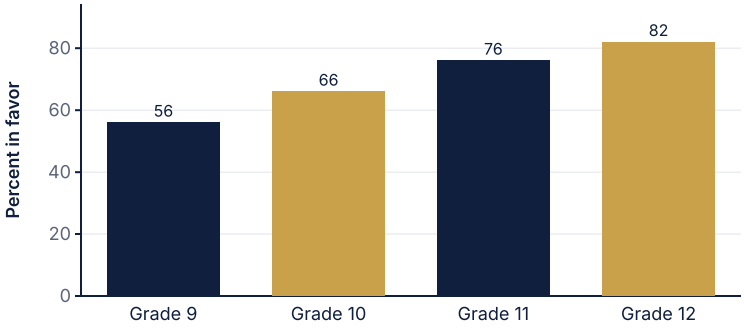
The sampling distribution of a mean for samples of 36 bags: mean 250, standard deviation 12/6 = 2. The shaded area is P(sample mean at most 246.2), about 0.029.

Sensitivity. A good investigator asks how much a conclusion depends on one unusual value or one assumption. Example 2 below shows how a single suspicious value can change a p-value from 0.07 to 0.002, which is a reason to check the value, not a license to delete it.

Worked Examples

INVESTIGATION 1: A SURVEY OF RECYCLING SUPPORT

A school of 2,400 students wants to estimate the proportion who support a new recycling program. The student council uses a list of all students, selects a simple random sample of 50 students from each of four grades (200 in all), and records the responses. Yes counts by grade: 28, 33, 38, 41. (a) Describe a potential source of bias and how the design avoids it. (b) Graph the results. (c) Construct and interpret a 95% interval for the overall proportion.



Support rises with grade level in this sample of 50 per grade

SOLUTION

(a) Asking only members of the environmental club would give a biased sample (selection bias). Random selection from a complete list prevents it, and selecting 50 from each grade ensures every grade is represented. Nonresponse is still a concern, so follow up with those who did not respond.

(b) The percentages by grade are 56%, 66%, 76% and 82%, rising with grade level.

(c) Overall $\hat{p} = 140/200 = 0.70$. Conditions: random sample; 200 is less than 10% of 2,400 (which is 240). Large counts: $140 \geq 10$ and $60 \geq 10$. $SE = \sqrt{0.7(0.3)/200} = 0.0324$, interval $0.70 \pm 1.96(0.0324) = (0.636, 0.764)$.

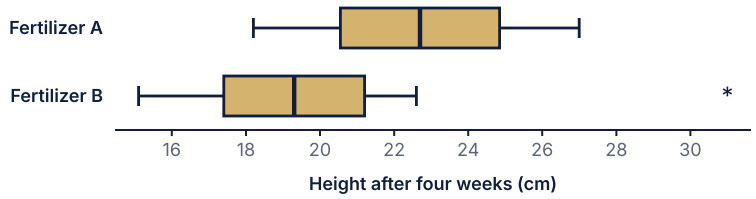
Interpretation: We are 95% confident that between 63.6% and 76.4% of all students support the program. Because equal numbers were sampled per grade, the overall figure treats grades equally, which is acceptable here only if grade sizes are similar.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

INVESTIGATION 2: PLANT GROWTH AND AN UNUSUAL VALUE

A greenhouse randomly assigns 24 identical seedlings, 12 to a new fertilizer (A) and 12 to the standard fertilizer (B). After four weeks the heights (cm) are recorded. A: 18.2, 19.5, 20.1, 21.0, 21.8, 22.4, 23.0, 23.7, 24.5, 25.2, 26.1, 27.0. B: 15.1, 16.4, 17.0, 17.8, 18.5, 19.0, 19.6, 20.2, 20.9, 21.5, 22.6, 31.0. (a) Describe the design. (b) Compare the distributions. (c) Test whether the fertilizers give different mean heights.



Heights of seedlings by fertilizer

SOLUTION

(a) This is a completely randomized experiment with one explanatory variable (fertilizer) and 12 replicates per group; random assignment of identical seedlings supports a cause-and-effect conclusion.

(b) A: Q1 20.55, median 22.7, Q3 24.85, mean 22.7, s = 2.73, roughly symmetric, no outliers. B: Q1 17.4, median 19.3, Q3 21.2, mean 19.97, s = 4.10, with a high outlier at 31.0 (upper fence $21.2 + 1.5(3.8) = 26.9$). The center of A is higher by 3.4 cm in median, and B is more variable because of the outlier.

(c) Two-sample t-test: $H_0 : \mu_A = \mu_B$, $H_a : \mu_A \neq \mu_B$. Random assignment holds; B has an outlier, so the normality condition is questionable. Calculation: $t = 1.93$, $df \approx 19.1$, p-value ≈ 0.069. Since 0.069 > 0.05, fail to reject H_0: there is not convincing evidence that the mean heights differ. The 95% interval is $(-0.23, 5.72)$ cm.

Sensitivity: Without the 31.0, $t = 3.58$ and the p-value is 0.002. A good investigator checks whether 31.0 is a recording error or a real plant before deciding; deleting it only because it hurts the result is not justified.

INVESTIGATION 3: QUALITY CONTROL, WHEN THE NORMAL APPROXIMATION FAILS

A factory claims that at most 8% of its parts are flawed. (a) In an inspection of 50 independent parts when the true rate is 8%, find the mean and standard deviation of the number flawed and the probability of 2 or fewer. (b) An auditor inspects a random sample of 60 parts and finds 9 flawed. Test $H_0 : p = 0.08$ versus $H_a : p > 0.08$.

SOLUTION

(a) $X \sim \text{Binomial}(50, 0.08)$: mean $= 4$, standard deviation $= \sqrt{50(0.08)(0.92)} = 1.918$. $P(X \leq 2) = \text{binomcdf}(50, 0.08, 2) = 0.226$.

(b) A z-test requires $np_0 \geq 10$: $60(0.08) = 4.8 < 10$, so the condition fails and the z-based p-value (0.023) is unreliable.

Use the exact binomial probability: if $p = 0.08$, then $P(X \geq 9) = 1 - \text{binomcdf}(60, 0.08, 8) \approx 0.048$.

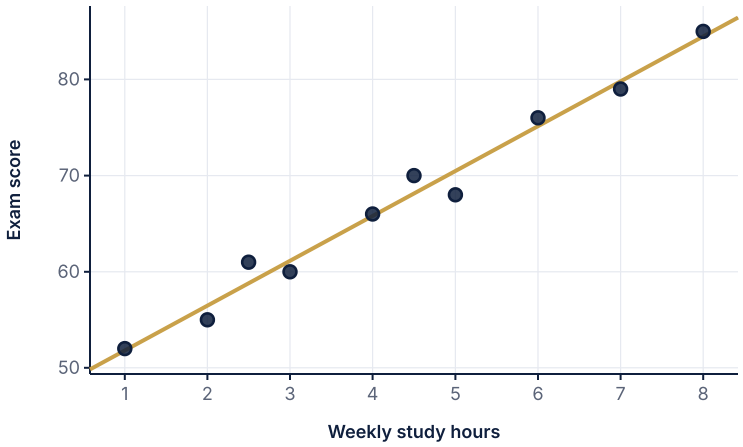
Conclusion: Since 0.048 < 0.05, we reject H_0 narrowly. There is convincing evidence that the true flaw rate exceeds 8%, though the evidence is borderline and the exact method should be named in the answer.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

INVESTIGATION 4: STUDY TIME AND SCORES

A teacher records weekly study hours (x) and exam score (y) for 10 randomly chosen students: x = 1, 2, 2.5, 3, 4, 4.5, 5, 6, 7, 8 and y = 52, 55, 61, 60, 66, 70, 68, 76, 79, 85. Describe the association, find the least-squares line, check it with a residual plot, and predict the score for 5.5 hours.



Scatterplot with least-squares line

SOLUTION

Description: the scatterplot is positive, linear and very strong, with no outliers.

Line: $\hat{y} = 47.12 + 4.669x$, $r = 0.990$, $r^2 = 0.981$, $s = 1.57$ points.

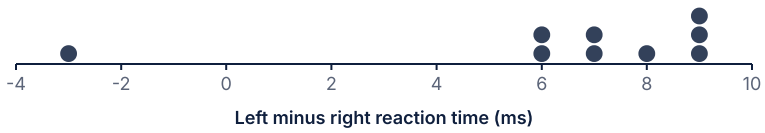
Residuals: 0.21, -1.46, 2.20, -1.13, 0.20, 1.87, -2.47, 0.86, -0.81, 0.53. The residual plot shows no pattern, so a linear model is appropriate.

Prediction at 5.5 hours: $47.12 + 4.669(5.5) = 72.8$ points, an interpolation.

Interpretation: For each additional hour of weekly study, the predicted exam score increases by about 4.7 points; about 98% of the variation in scores is accounted for by the linear relationship with study time. Because this is observational, we cannot conclude studying causes the higher scores (motivation could be a confounder).

INVESTIGATION 5: MATCHED PAIRS AND A CONDITION CHECK

Nine volunteers have their reaction time (milliseconds) measured with their left hand and with their right hand, with the order randomized. Left minus right differences: 7, 7, 9, -3, 8, 9, 6, 9, 6. Construct a 95% confidence interval for the mean difference and interpret it.



SOLUTION

Each volunteer provides both measurements, so use a paired t-interval on the differences (parameter μ_d).

Conditions: the order was randomized and volunteers are independent; with $n = 9$ the dotplot of differences should show no strong skewness or outliers. The dotplot has one low value (-3), but the remaining data are fairly compact, so the condition is acceptable.

$\bar{d} = 6.44$, $s_d = 3.75$, $df = 8$, $t^* = 2.306$: $6.44 \pm 2.306(3.75/\sqrt{9}) = (3.57, 9.32)$ ms.

Interpretation: We are 95% confident that the true mean reaction time with the left hand is between 3.57 and 9.32 ms slower than with the right hand. Since 0 is not in the interval, the data suggest a real difference (two-sided p-value about 0.0009).

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

INVESTIGATION 6: FROM A POPULATION TO A SAMPLE MEAN

The weight of a bag of rice is approximately normal with mean 250 g and standard deviation 12 g. (a) Find the probability that one bag weighs less than 245 g. (b) Find the probability that the mean of 36 bags is below 245 g. (c) A sample of 36 bags has mean 246.2 g. What does this suggest about the claim that $\mu = 250$?

SOLUTION

(a) $P(X < 245) = \text{normalcdf}(-1E99, 245, 250, 12) = 0.338$.

(b) The sampling distribution of $\bar{x}$ is normal with mean 250 and standard deviation $12/\sqrt{36} = 2$. $P(\bar{x} < 245) = \text{normalcdf}(-1E99, 245, 250, 2) = 0.006$. Averages vary far less than single bags.

(c) Suppose $\mu = 250$. Then $P(\bar{x} \le 246.2) = 0.029$ (from $z = -1.9$). This is unusual under the claim, so the data suggest the true mean is below 250 g. A formal test with a known σ would give a one-sided p-value of about 0.029.

Conclusion: At $\alpha = 0.05$ we would reject the claim, but this relies on $\sigma = 12$ being known and on the bags being a random sample.

Write It Like a 5

Prompt: A teacher randomly assigns 50 students to two note-taking methods, 25 per method, and compares end-of-unit scores. Method 1: mean 78.4, s = 9.6. Method 2: mean 72.1, s = 10.8. Assume scores in each group show no strong skewness or outliers. (a) Describe why this design allows a cause-and-effect conclusion. (b) Is there convincing evidence that the mean scores differ? (c) Estimate the difference in true means with 95% confidence.

Model Response

(a) Students were randomly assigned to the methods, which balances other variables such as prior ability across groups on average, so a difference in outcomes can be attributed to the method (if the students are volunteers from one class, the conclusion applies to students like these). (b) Let μ_1, μ_2 be the true mean scores with Methods 1 and 2. $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 \neq 0$. Two-sample t-test: random assignment, independent groups, and no strong skewness or outliers in either group of 25. $t = \frac{78.4 - 72.1}{\sqrt{9.6^2/25 + 10.8^2/25}} = 2.18$, $df \approx 47.3$, p-value ≈ 0.034. Since 0.034 < 0.05, reject H_0: there is convincing evidence that the mean scores differ. (c) The interval is $6.3 \pm 2.012(2.89) = (0.49, 12.11)$. We are 95% confident that the true mean score with Method 1 is between 0.49 and 12.11 points higher than with Method 2.

What the grader looks for

(1) Random assignment named as the reason for the causal conclusion, with the scope limit. (2) Correct procedure (two-sample t) and parameters in context. (3) Conditions: random assignment, independence, no strong skewness or outliers. (4) Correct t, df and p-value, with a conclusion linked to alpha. (5) Interval computed and interpreted in context as a statement about the difference in means.

Coach's Playbook

1. **Read the whole prompt first.** Mark each part (a), (b), (c) with the stage it asks for: design, graph, model or inference.
2. **Design before data.** Decide whether the data came from an experiment or an observational study; this limits the final sentence.
3. **Graph before test.** Sketch or describe the data to check shape and outliers before you trust a t-procedure.
4. **Model with care.** Check that binomial and normal conditions hold, and use the exact binomial method when counts are small.
5. **Tie parts together.** A later part often depends on an earlier one: if you find an outlier in (b), mention it when checking conditions in (c).
6. **End with scope.** Say what population the conclusion applies to and whether causation is supported.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Common traps

1. Running a test without checking conditions, especially normality in small groups. 2. Deleting an outlier only because it changes the result. 3. Claiming cause and effect from observational data, such as study hours and scores. 4. Forgetting that probability about a sample mean uses the standard deviation $\sigma/\sqrt{n}$, not σ.

Mistake Clinic

Spot the error

A student uses a z-test for $H_0 : p = 0.08$ with $n = 60$ and 9 flawed parts, reporting p = 0.023. Where does this fail? The expected count $np_0 = 4.8$ is less than 10, so the normal approximation is not appropriate. Correct: use the exact binomial probability $P(X \geq 9) = 0.048$.

Spot the error

A student finds P(sample mean of 36 bags < 245) with normalcdf(-1E99, 245, 250, 12) and gets 0.338. Where does this fail? That is the probability for a single bag. A mean of 36 bags has standard deviation $12/\sqrt{36} = 2$; the answer is normalcdf(-1E99, 245, 250, 2) = 0.006.

Spot the error

A student writes, "Study hours and exam scores have r = 0.99, so studying more causes higher scores." Where does this fail? The data come from an observational study, so confounding (such as motivation) is possible. Correct: "Study time and score are strongly positively associated; the data do not show that studying causes higher scores."

Real-World Connection

Public-health teams, product designers and sports analysts move through exactly this cycle: collect with a design, plot the data, model the chance variation, test, and then write what the evidence does and does not support. Practice with whole investigations is practice for working with data in any career.

Challenge Corner

For the plant data in Investigation 2, a student removes the outlier 31.0 and reports a p-value of 0.002 without comment. Write a justified recommendation. **Solution:** The outlier is 31.0 cm, far above the upper fence of 26.9. A responsible report states both analyses: with all data, $t = 1.93$, p = 0.069 (not significant); without it, $t = 3.58$, p = 0.002 (significant). The conclusion depends on whether 31.0 is a mistake or a true seedling, so the investigators should check the original record. If it is an error, correct or remove it and say so; if it is real, report the first analysis and note that the result is sensitive to one observation, perhaps recommending a larger experiment.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice

PRACTICE **Investigative Tasks: Combining Skills Across Units**

Show your work. A calculator is allowed. Justify conclusions in context.

1. A principal surveys only the students who stay for after-school clubs about the school's schedule. The most likely problem is
   (A) nonresponse bias because some students refuse
   (B) selection bias because the sample is not representative of all students
   (C) confounding between schedule and clubs
   (D) too small a margin of error
2. A study finds that people who drink more tea have lower blood pressure, but tea drinkers also tend to exercise more. Exercise is best described as
   (A) a confounding variable **(B)** a placebo **(C)** a response variable
   (D) a blocking variable
3. In an experiment, the purpose of blocking is to
   (A) increase the sample size
   (B) make the experiment single-blind
   (C) reduce variability due to a known variable by comparing units within similar groups
   (D) guarantee a cause-and-effect conclusion
4. A skewed right distribution of waiting times has an outlier. Which pair is the better summary of center and spread?
   (A) Mean and standard deviation **(B)** Median and standard deviation
   (C) Mean and IQR **(D)** Median and IQR
5. Scores are approximately N(70, 8). The proportion of scores above 82 is closest to
   (A) 0.067 **(B)** 0.933 **(C)** 0.015 **(D)** 0.250
6. A random variable X is Binomial(10, 0.2). P(X = 3) is closest to
   (A) 0.121 **(B)** 0.678 **(C)** 0.322 **(D)** 0.201
7. The standard deviation of the sampling distribution of $\hat{p}$ for n = 100 and p = 0.4 is closest to
   (A) 0.24 **(B)** 0.049 **(C)** 0.0024 **(D)** 0.40
8. A 95% confidence interval for a proportion has a margin of error of 0.06 with a sample of 200. If the sample size is increased to 800 (and everything else stays the same), the margin of error is about
   (A) 0.015 **(B)** 0.06 **(C)** 0.03 **(D)** 0.12
9. A regression line is $\hat{y} = 12 + 3x$. For x = 4 the observed y is 27. The residual is
   (A) -3 **(B)** 3 **(C)** 24 **(D)** 39
10. A coach wants to compare two training methods for sprint time using 30 athletes. Some athletes are experienced and some are novices.
    (a) Describe a randomized block design that accounts for experience.
    (b) Explain how random assignment should be carried out.
    (c) State the type of conclusion that is supported.
11. Exam scores are approximately normal with mean 70 and standard deviation 8.
    (a) Find the percent of students who score above 82.
    (b) Find the minimum score needed to be in the top 10%.
    (c) Find the probability that a random sample of 16 students has a mean above 74.
12. A random sample of 400 voters finds that 232 support a ballot measure. Test whether the true proportion differs from 0.50 at alpha = 0.05, and construct a 95% interval. Check conditions and conclude in context.
13. A sample of six students recorded hours of practice per week (x) and test scores (y): x = 10, 15, 20, 25, 30, 35; y = 22, 30, 31, 41, 44, 52.
    (a) Find the least-squares line, r and r-squared.
    (b) Predict y when x = 28.
    (c) Find the residual for x = 20.
    (d) Interpret the slope and r-squared.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. Seven typists take a speed test before and after a training course; the number of errors is: Before: 14, 18, 11, 16, 13, 20, 15. After: 12, 15, 11, 13, 10, 16, 14. Assume the typists are a random sample and the differences show no strong skewness.
    (a) Explain why a paired procedure fits.
    (b) Test whether the mean number of errors decreased (alpha = 0.05).
    (c) Give a 95% confidence interval for the mean decrease.
15. The waiting time at a clinic is skewed right with mean 8 minutes and standard deviation 3 minutes. A random sample of 40 patients is selected.
    (a) Describe the shape, mean and standard deviation of the sampling distribution of the sample mean.
    (b) Find the probability that the sample mean exceeds 9 minutes.
    (c) Explain why the probability that one randomly chosen patient waits more than 9 minutes cannot be found with a normal model.

Answers Answers and Explanations

C.1 Which Procedure Do I Use? The Decision Guide

PRACTICE · WHICH PROCEDURE DO I USE? THE DECISION GUIDE

1. C. The goal is to estimate one population proportion from a categorical response, so use a one-proportion z-interval (here $\hat{p} = 0.65$, interval about (0.545, 0.755)). A one-sample t-interval is for quantitative data and is the tempting mistake.
2. B. The two brands form two independent random samples of a quantitative measurement. A paired test would need natural pairs, which are not present.
3. D. The same volunteers are measured twice, so analyze the differences with a paired t-test. A two-sample test would ignore the pairing and lose power.
4. A. The response (improve or not) is categorical and the groups come from random assignment, so use a two-proportion z-test. A t-test is for means.
5. B. One sample of a quantitative measurement and a claimed mean call for a one-sample t-test. A proportion test is for categorical outcomes.
6. A. Two categorical variables from one sample are tested for association with the chi-square test for independence. A t-test needs a quantitative response.
7. A. In paired data each observation has a natural partner, and the analysis uses the differences. Choice C reverses the idea, because a design that measures the same subjects twice is paired.
8. D. Random assignment of treatments in an experiment is what supports cause and effect. Observational comparisons and surveys can show association only.
9. B. The rate is known and the question asks for the probability of a count, so use the binomial model with $n = 15$, $p = 0.1$, $k = 2$: $P = 0.267$. Inference procedures are for estimating an unknown parameter.
10. (a) Parameter μ, the true mean commute time of the city's workers: one-sample t-interval.
    (b) Parameter $p_1 - p_2$, the difference in true pass proportions for the two methods: two-proportion z-test.
    (c) Parameter μ_d, the true mean change in weight (after minus before): paired t-test.
11. (a) The same students take both tests, so each pre-test score is naturally matched with a post-test score.
    (b) Differences (post minus pre): 3, 3, -1, 6, 3, 5, 1, 5, 3, 5. $H_0 : \mu_d = 0$, $H_a : \mu_d > 0$. $\bar{d} = 3.3$, $s_d = 2.11$, $t = 4.94$, $df = 9$, p-value ≈ 0.0004. Since this is less than 0.05, reject H_0: there is convincing evidence that the mean score improves.
    (c) $3.3 \pm 2.262(2.11/\sqrt{10}) = (1.79, 4.81)$. We are 95% confident that the true mean improvement is between 1.79 and 4.81 points.
12. (a) Random assignment; independent groups; pooled proportion $\hat{p}_c = 117/175 = 0.669$; expected successes and failures $90(0.669) = 60.2$, $90(0.331) = 29.8$, $85(0.669) = 56.8$, $85(0.331) = 28.2$ are all at least 10.
    (b) $\hat{p}_X = 0.733$, $\hat{p}_Y = 0.600$, $z = 1.87$, p-value $= 0.031$. Since 0.031 < 0.05, reject H_0: there is convincing evidence that the true proportion who improve is greater with Plan X than with Plan Y.
    (c) $0.133 \pm 1.645(0.0707) = (0.017, 0.250)$. We are 90% confident that the true proportion who improve with Plan X is between 0.017 and 0.250 higher than with Plan Y.
13. (a) Two independent samples of a quantitative variable: two-sample t-test for $\mu_1 - \mu_2$.
    (b) $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 \neq 0$. $SE = 0.6506$, $t = \frac{-1.8}{0.6506} = -2.77$, $df \approx 54.7$, p-value ≈ 0.008. Since this is less than 0.05, reject H_0: there is convincing evidence that the true mean times differ.
    (c) $-1.8 \pm 2.004(0.6506) = (-3.10, -0.50)$. We are 95% confident that the true mean time for Group 1 is between 0.50 and 3.10 minutes shorter than for Group 2.
14. (a) Chi-square test for independence (one sample, two categorical variables). H_0: sleep adequacy and exercise are independent. H_a: they are associated.
    (b) Totals: regular 50, rare 50; adequate 48, short 52. Expected: regular and adequate 24, regular and short 26, rare and adequate 24, rare and short 26. All are at least 5.
    (c) $\chi^2 = 5.77$, $df = 1$, p-value ≈ 0.016. Since 0.016 < 0.05, reject H_0: there is convincing evidence of an association between exercise level and sleep adequacy in the population of adults.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

15. (a) Two independent groups created by random assignment and a quantitative response: two-sample t-test for a difference in means.
(b) $SE =$\sqrt{3.1^2/29+3.4^2/29}=0.8544;t=\frac{-4.8-(-2.9)}{0.8544}=-2.22,df\approx 55.5,$p-value$\approx 0.030.$Since$0.030 < 0.05,$reject H_0$: there is convincing evidence that the mean weight changes differ.
(c) Because treatments were randomly assigned, we can conclude that the difference was caused by the diets. Because the participants were volunteers and not a random sample, we can generalize only to people like these volunteers.

C.2 Investigative Tasks: Combining Skills Across Units

PRACTICE · INVESTIGATIVE TASKS: COMBINING SKILLS ACROSS UNITS

1. B. Club members are not representative of all students, so the sample is biased by selection. Nonresponse and margin of error are real concerns in surveys but they are not the main problem here.
2. A. Exercise is related to both the explanatory variable and the response, so it is a confounding variable. A blocking variable is part of an experiment, while this is an observational study.
3. C. Blocking groups similar units so treatments are compared within blocks, which reduces variability. It does not increase the sample size or by itself guarantee causation.
4. D. With skewness and an outlier, the median and IQR are resistant. The mean and standard deviation would be pulled by the outlier, which is the tempting but wrong pair.
5. A. $z = (82 - 70)/8 = 1.5$ and the upper tail is $1 - 0.9332 = 0.0668$. The value 0.933 is the area to the left of 82.
6. D. $P(X = 3) = \binom{10}{3}(0.2)^3(0.8)^7 = 0.201$. The value 0.678 is $P(X \le 3)$, which is the cumulative error.
7. B. $\sqrt{0.4(0.6)/100} = \sqrt{0.0024} = 0.049$. The value 0.24 forgets to divide by n and the root.
8. C. The margin of error is proportional to $1/\sqrt{n}$. Quadrupling n halves the margin: $0.06/2 = 0.03$. It does not shrink by a factor of 4.
9. B. Residual = actual - predicted. The predicted value is $12 + 3(4) = 24$, so the residual is $27 - 24 = 3$. The value -3 reverses the subtraction.
10. (a) Block the athletes by experience: put the experienced athletes in one block and the novices in another.
(b) Within each block, randomly assign half of the athletes to Method 1 and half to Method 2 (for example, by drawing names or using random digits).
(c) Because treatments are randomly assigned, cause and effect can be concluded for athletes like those in the study; generalization to others depends on how the 30 were chosen.
11. (a) $P(X > 82) = \text{normalcdf}(82, 1E99, 70, 8) = 0.067$, about 6.7% of students.
(b) invNorm(0.90, 70, 8) = 80.25, so a score of about 80.3.
(c) The sampling distribution of $\bar{x}$ is $N(70, 8/\sqrt{16} = 2)$, so $P(\bar{x} > 74) = \text{normalcdf}(74, 1E99, 70, 2) = 0.023$.
12. Parameter p: the true proportion of all voters in support. $H_0 : p = 0.50$, $H_a : p \ne 0.50$. Conditions: random sample; 400 is less than 10% of all voters; $400(0.5) = 200 \ge 10$ in both categories. $\hat{p} = 0.58$, $z = \frac{0.58 - 0.5}{\sqrt{0.25/400}} = 3.20$, p-value $= 0.0014$. Since this is less than 0.05, reject H_0: there is convincing evidence that the true proportion differs from 0.50 (it appears to be greater). The 95% interval is $0.58 \pm 1.96\sqrt{0.58(0.42)/400} = (0.532, 0.628)$; we are 95% confident that between 53.2% and 62.8% of all voters support the measure.
13. (a) $\hat{y} = 10.70 + 1.154x$, $r = 0.986$, $r^2 = 0.973$.
(b) At x = 28: $10.70 + 1.154(28) = 43.0$.
(c) Predicted at x = 20 is 33.78, so the residual is $31 - 33.78 = -2.78$ (the observed score is below the line).
(d) Slope: for each additional hour of weekly practice, the predicted test score increases by about 1.15 points. r^2: about 97% of the variation in test scores is accounted for by the linear relationship with practice hours.
14. (a) The same typists are measured twice, so each before value is matched with an after value; analyze the differences (before minus after): 2, 3, 0, 3, 3, 4, 1.
(b) $H_0 : \mu_d = 0$, $H_a : \mu_d > 0$. $\bar{d} = 2.29$, $s_d = 1.38$, $t = 4.38$, $df = 6$, p-value $= 0.0023$. Since 0.0023 < 0.05, reject H_0: there is convincing evidence that the mean number of errors decreased after the training.
(c) $2.29 \pm 2.447(1.38/\sqrt{7}) = (1.01, 3.56)$. We are 95% confident that the true mean decrease is between 1.01 and 3.56 errors.
15. (a) By the Central Limit Theorem, with $n = 40$ the sampling distribution of $\bar{x}$ is approximately normal, with mean 8 minutes and standard deviation $3/\sqrt{40} = 0.474$ minutes.
(b) $P(\bar{x} > 9) = \text{normalcdf}(9, 1E99, 8, 0.474) = 0.018$.
(c) A single waiting time is skewed right, so a normal model for individual patients is inappropriate; the Central Limit Theorem applies to the sample mean, not to individual values.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice Exam 1

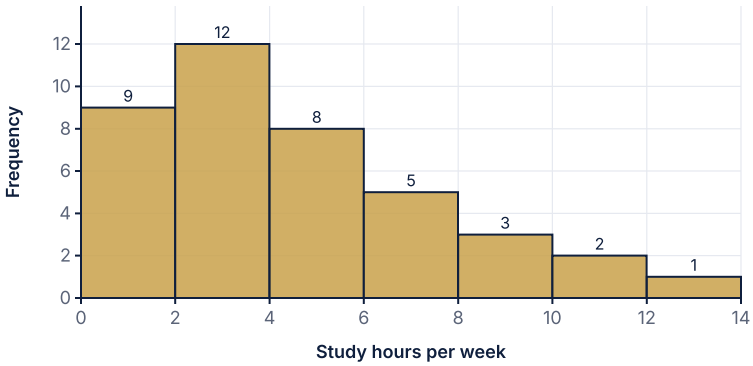
42 multiple-choice questions and 4 free-response questions in the digital exam format, followed by every answer, explanation, rubric, and a scoring guide.

Exam 1 Section I: Multiple Choice

Section I · 42 questions · 90 minutes · 50% of score

Choose the best answer for each question. A graphing calculator is allowed, and the formula sheet and tables in the appendix are available. Questions in a set share one stimulus.

1

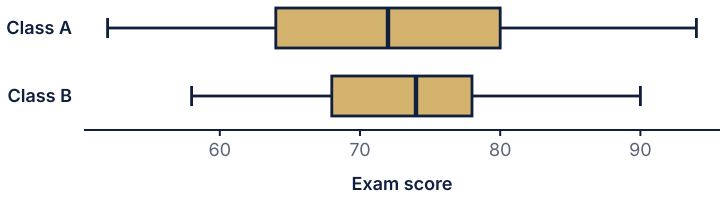


Weekly study hours, 40 students

The histogram shows the number of hours per week that 40 randomly chosen students spend studying outside of class. Which statement best describes the distribution?

(A) The distribution is skewed right, so the mean is less than the median.
(B) The distribution is skewed left, so the mean is less than the median.
(C) The distribution is roughly symmetric, so the mean is approximately equal to the median.
(D) The distribution is skewed right, so the mean is greater than the median.

2



Final exam scores by section

The boxplots compare final exam scores for two sections of the same course. Which conclusion is supported by the plots?

(A) Class B has the higher median score, and its scores are less variable in the middle 50% than those of Class A.
(B) Class A has the higher median score because its maximum score is larger.
(C) Class B has the larger range of scores because its box is shorter.
(D) About 75% of the students in Class A scored above 72.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

3 On a chemistry test, Maya scored 84 in a class with mean 75 and standard deviation 6. On a biology test, Jon scored 71 in a class with mean 62 and standard deviation 5. Both students scored 9 points above their class mean. Relative to their own classes, who performed better?

(A) Maya, because her raw score of 84 is higher than Jon's raw score of 71.
(B) Jon, because his z-score of 1.8 is greater than Maya's z-score of 1.5.
(C) They performed equally well, because each scored 9 points above the mean.
(D) Maya, because her class has the larger standard deviation.

4 The daily water temperature of a lake in degrees Celsius has mean 20 and standard deviation 3. Temperatures are converted to degrees Fahrenheit using $F = 1.8C + 32$. What are the mean and standard deviation of the temperatures in degrees Fahrenheit?

(A) Mean 68, standard deviation 37.4
(B) Mean 36, standard deviation 5.4
(C) Mean 68, standard deviation 5.4
(D) Mean 68, standard deviation 3

5 The daily air quality index in a city is approximately normally distributed with mean 52 and standard deviation 9. About what proportion of days have an index above 65?

(A) 0.9257
(B) 0.1486
(C) 0.4362
(D) 0.0743

6 The number of minutes students ride the bus to school has first quartile 14 and third quartile 26. Which of the following times would be identified as an outlier by the $1.5 \times IQR$ rule?

(A) 43 minutes
(B) 40 minutes
(C) -2 minutes
(D) 45 minutes

7 To estimate the proportion of students who favor a later school start time, the principal posts a link to an online survey in the weekly school newsletter and uses the responses of the 220 students who choose to reply. Which is the main concern with this method?

(A) It is a convenience sample, which is always unbiased as long as the sample is larger than 30.
(B) It is a voluntary response sample, so students with strong opinions are likely to be overrepresented.
(C) The sample size of 220 is too small for any survey to be reliable.
(D) The results are biased because the survey was posted online, and all online surveys are biased in the same direction.

8 A researcher compares two fertilizers on 24 garden plots. Plots on the upper part of the hillside tend to be drier than plots on the lower part, so the researcher groups the plots into 12 upper and 12 lower plots and randomly assigns each fertilizer to 6 plots within each group. What is the main purpose of grouping the plots this way?

(A) To control for variability in yield caused by moisture, so the fertilizers can be compared more precisely
(B) To make random assignment unnecessary, since the groups are already similar
(C) To make the experiment double-blind
(D) To make sure every plot receives both fertilizers

9 A random sample of 300 teenagers is surveyed, and teens who report more than 8 hours of sleep per night have a higher average grade point average than teens who report less. Which conclusion is justified?

(A) More sleep causes higher grades, because the sample was randomly selected.
(B) Higher grades cause teens to sleep more, because the grades were measured after the sleep was reported.
(C) There is an association between sleep and grades in the population the sample represents, but because this is an observational study, we cannot conclude that more sleep causes higher grades.
(D) No conclusion of any kind can be drawn because the study was not an experiment.

10 A county has 12 high schools. A researcher randomly selects 4 of the schools and surveys every senior at each of the 4 chosen schools. Which type of sampling is this?

(A) Stratified random sample
(B) Simple random sample of seniors in the county
(C) Systematic sample
(D) Cluster sample

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

11 Monthly rainfall in a region is approximately normally distributed with mean 40 mm and standard deviation 6 mm. A month is called 'very wet' if its rainfall is in the top 10% of all months. What is the minimum rainfall for a very wet month?

(A) 32.3 mm
(B) 49.9 mm
(C) 47.7 mm
(D) 40.6 mm

12 At a school fair, a player pays 3 dollars to spin a wheel. Let X be the prize, in dollars, that the player wins on one spin. The probability distribution of X is shown in the table, where k is the probability of the 20 dollar prize.

Prize X (dollars)	0	2	5	20
Probability	0.50	0.30	0.15	k

What is the value of k?

(A) 0.10
(B) 0.20
(C) 0.05
(D) 0.95

13

Prize X (dollars)	0	2	5	20
Probability	0.50	0.30	0.15	k

What is the expected value of the prize X on one spin?

(A) 6.75 dollars
(B) 2.35 dollars
(C) 5.35 dollars
(D) 0.65 dollars

14

Prize X (dollars)	0	2	5	20
Probability	0.50	0.30	0.15	k

What is the standard deviation of the prize X?

(A) 19.43 dollars
(B) 4.99 dollars
(C) 2.35 dollars
(D) 4.41 dollars

15

	Music club	No music club	Total
Plays a sport	18	42	60
Does not play a sport	30	30	60
Total	48	72	120

A survey of 120 students recorded whether each plays a sport and whether each is in the music club. If a student who is in the music club is chosen at random, what is the probability that the student plays a sport?

(A) 18/60 = 0.300
(B) 18/120 = 0.150
(C) 18/48 = 0.375
(D) 48/120 = 0.400

16 For events A and B, P(A) = 0.40, P(B) = 0.25, and P(A or B) = 0.55. Which statement is correct?

(A) A and B are mutually exclusive but not independent.
(B) A and B are independent but not mutually exclusive.
(C) A and B are both independent and mutually exclusive.
(D) A and B are neither independent nor mutually exclusive.

17 A basketball player makes 70% of her free throws, and the shots are independent. She attempts 8 free throws. Which is the probability that she makes at least 6 of them?

(A) 0.2965
(B) 0.2553
(C) 0.7447
(D) 0.5518

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

18 Which of the following random variables is most reasonably modeled as binomial?

(A) The heights, in centimeters, of 10 randomly selected students

(B) The number of defective bulbs in a random sample of 20 bulbs from a production line that is 3% defective and makes thousands of bulbs a day

(C) The number of left-handed students among 5 students chosen without replacement from a class of 12 that contains 4 left-handed students

(D) The total score from rolling a six-sided die 10 times

19 For samples of size 100 from a population in which 30% of members have a certain trait, the standard deviation of the sampling distribution of $\hat{p}$ is about 0.046. If the sample size is increased to 400, what is the standard deviation of the sampling distribution of $\hat{p}$?

(A) 0.023 **(B)** 0.011

(C) 0.046 **(D)** 0.092

20 A population of commute times is strongly skewed to the right with mean 25 minutes and standard deviation 10 minutes. A random sample of 64 commuters is selected. What is the approximate probability that the sample mean is less than 23 minutes?

(A) 0.0548 **(B)** 0.9452

(C) 0.4207 **(D)** less than 0.0001

21 A random sample of 40 students from a very large school is asked whether they ride the bus, and 5 say yes. A student wants to construct a one-sample z interval for the population proportion who ride the bus. Which statement about the Large Counts condition is correct?

(A) The condition is met because the sample size of 40 is greater than 30.

(B) The condition is not met because the number of successes, 5, is less than 10.

(C) The condition is met because 5 is less than 10 percent of the population.

(D) The condition is not met because the sample is random.

22 In a random sample of 150 students at a large school, 62 say they walk to school. Which is a 95% confidence interval for the proportion of all students at the school who walk?

(A) (0.347, 0.479) **(B)** (0.310, 0.517)

(C) (0.373, 0.454) **(D)** (0.335, 0.492)

23 A 95% confidence interval for the proportion of households in a city that recycle is (0.58, 0.66). Which is the best interpretation of the confidence level?

(A) There is a 95% probability that the true proportion is between 0.58 and 0.66 for this particular interval.

(B) 95% of households in the city recycle between 58% and 66% of the time.

(C) If we repeatedly took random samples of the same size and built an interval from each, about 95% of those intervals would capture the true proportion.

(D) 95% of all sample proportions from samples of this size fall between 0.58 and 0.66.

24 A school nurse believes that more than 25% of the school's students need vision correction. She plans to take a random sample of students to test her belief. Which are the appropriate hypotheses, where p is the true proportion of all students who need vision correction?

(A) $H_0 : p = 0.25$ and $H_a : p > 0.25$ **(B)** $H_0 : \hat{p} = 0.25$ and $H_a : \hat{p} > 0.25$

(C) $H_0 : p > 0.25$ and $H_a : p = 0.25$ **(D)** $H_0 : p = 0.25$ and $H_a : p \neq 0.25$

25 A significance test of $H_0 : p = 0.60$ versus $H_a : p > 0.60$ gives a P-value of 0.034. Which is the correct interpretation of the P-value?

(A) If the true proportion is 0.60, the probability of getting a sample proportion at least as large as the one observed is 0.034.

(B) The probability that the null hypothesis is true is 0.034.

(C) There is a 96.6% chance that the true proportion is greater than 0.60.

(D) The sample proportion is 0.034 greater than 0.60.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

26

Group	Attended	Total	Sample proportion
Text reminder	78	120	0.650
No reminder	62	110	0.564

A clinic randomly assigns appointments to receive a text reminder or no reminder, with the results shown. A test of $H_0 : p_1 = p_2$ versus $H_a : p_1 > p_2$ is conducted, where p_1 and p_2 are the true attendance proportions with and without a reminder. Which is the test statistic of the two-proportion z test using the pooled proportion?

(A) z = 2.68
(B) z = 1.98
(C) z = 1.34
(D) z = 0.09

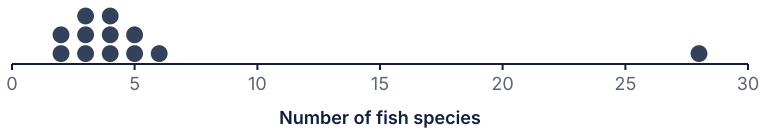
27 A polling group wants to estimate the proportion of voters in a district who support a ballot measure, with a margin of error of no more than 0.03 at 95% confidence. No prior estimate of the proportion is available. What is the minimum sample size needed?

(A) 1068
(B) 1067
(C) 1844
(D) 267

28 A random sample of 16 trout from a river has a mean weight of 12.4 ounces and a standard deviation of 2.8 ounces. Assume the weights are approximately normally distributed. Which is a 95% confidence interval for the mean weight of trout in the river?

(A) (11.03, 13.77)
(B) (10.91, 13.89)
(C) (12.03, 12.77)
(D) (10.34, 14.46)

29



A random sample of 12 local streams is selected, and the number of fish species found in each is shown in the dotplot. A researcher wants to construct a one-sample t interval for the mean number of species in all local streams. Is the procedure appropriate?

(A) Yes, because the t procedures can be used with any sample, since the t distribution adjusts for sample size.
(B) Yes, because the sample is random and the sample size is at least 10.
(C) No, because a t interval can only be used when the population standard deviation is known.
(D) No, because the sample is small and the data are strongly skewed with an extreme outlier.

30 A manufacturer claims that its batteries last an average of 50 hours. A consumer group tests the alternative that the mean is less than 50 hours. A random sample of 25 batteries has a mean life of 47.2 hours and a standard deviation of 6.5 hours, and the lifetimes are approximately normally distributed. What is the P-value of the test?

(A) 0.0415
(B) 0.0208
(C) 0.0156
(D) 0.9792

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

31

Group	n	Mean	Std. dev.
Online course	30	78.4	9.1
In-person course	32	74.1	10.3
t = 1.745	df = 59.8	P-value = 0.0862	(two-sided)

Students were randomly assigned to an online course or an in-person course, and final scores were compared with a two-sample t test of $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 \neq 0$. Assume the conditions for inference are met. The results are shown. What is the appropriate conclusion at the 5% significance level?

(A) Because 78.4 is greater than 74.1, there is convincing evidence that the online course produces higher mean scores.

(B) Because the P-value is greater than 0.05, we have shown that the two population means are equal.

(C) Because the P-value is greater than 0.05, there is not convincing evidence of a difference in mean final scores between the two types of course.

(D) Because the P-value is 0.086, there is a 8.6% chance that the null hypothesis is true.

32 A coach records the 5-kilometer running time of each of 15 runners before and after an eight-week training program. To determine whether the program reduces the mean time, which procedure is most appropriate?

(A) A two-sample t test comparing the mean before time and the mean after time

(B) A two-proportion z test comparing the proportion of runners who improved before and after

(C) A matched-pairs t test using the 15 before-minus-after differences

(D) A one-sample t test using only the 15 after times

33 A 95% confidence interval for $\mu_L - \mu_E$, the difference in mean nightly sleep (hours) between students at schools with late and early start times, is (0.12, 1.04). Assuming the conditions are met, which conclusion is justified?

(A) Because the interval is close to 0, there is convincing evidence that there is no difference in the means.

(B) 95% of students at late-start schools sleep between 0.12 and 1.04 hours more than students at early-start schools.

(C) Because 0 is in the interval, we cannot conclude that the means differ.

(D) The interval contains only positive values, so there is convincing evidence at the 5% level that the mean sleep is greater at late-start schools.

34 A researcher is constructing a one-sample t interval for a population mean. Which change, if everything else stayed the same, would produce a narrower interval?

(A) Increasing the confidence level from 95% to 99%

(B) Increasing the sample size from 25 to 100

(C) Using a sample that has a larger standard deviation

(D) Using a sample of size 25 but with two outliers added

35

A teacher records the number of hours studied (x) and the score on a 100-point exam (y) for 12 students. A least-squares regression produced the computer output shown.

Predictor	Coefficient
Constant	50.83
Hours	4.58
S	1.82
R-sq	97.8%

Which is the correct interpretation of the slope of the least-squares regression line?

(A) For each additional hour studied, the exam score increases by exactly 4.58 points.

(B) When a student studies 0 hours, the predicted exam score is 4.58 points.

(C) For each additional hour studied, the predicted exam score increases by about 4.58 points.

(D) For each additional point on the exam, the predicted study time increases by 4.58 hours.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

36

Predictor	Coefficient
Constant	50.83
Hours	4.58
S	1.82
R-sq	97.8%

A student studied for 5 hours and scored 75 on the exam. What is the residual for this student?

(A) -1.3 points
(B) 1.3 points
(C) 24.2 points
(D) 73.7 points

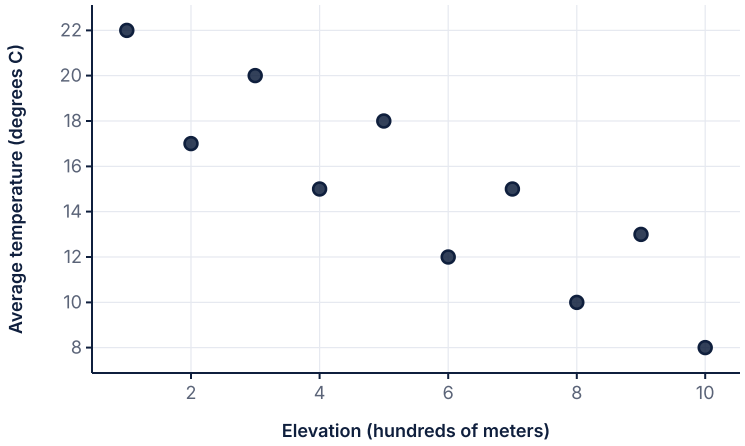
37

Predictor	Coefficient
Constant	50.83
Hours	4.58
S	1.82
R-sq	97.8%

Which is the correct interpretation of R-sq in the output?

(A) About 97.8% of the students had exam scores that were predicted exactly.
(B) About 97.8% of the variation in exam scores is accounted for by the linear relationship with hours studied.
(C) The correlation between hours studied and exam score is 0.978.
(D) The exam score increases by 97.8% for each additional hour studied.

38



10 weather stations

The scatterplot shows the elevation of 10 weather stations (in hundreds of meters) and the average temperature (degrees Celsius) at each station. Which is the best estimate of the correlation coefficient r?

(A) -0.9
(B) -0.5
(C) 0.9
(D) -0.1

39

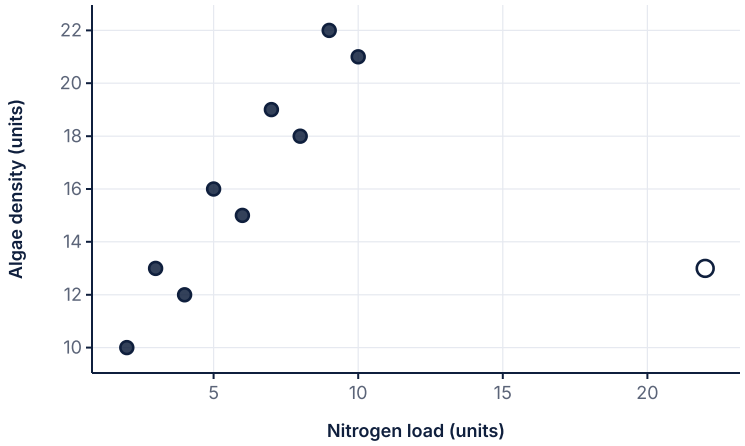
A least-squares regression of the amount of air pollution on distance from a highway has a negative slope and an R-sq of 64%. What is the correlation coefficient r?

(A) -0.80
(B) 0.80
(C) -0.64
(D) -0.41

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

40



10 lakes

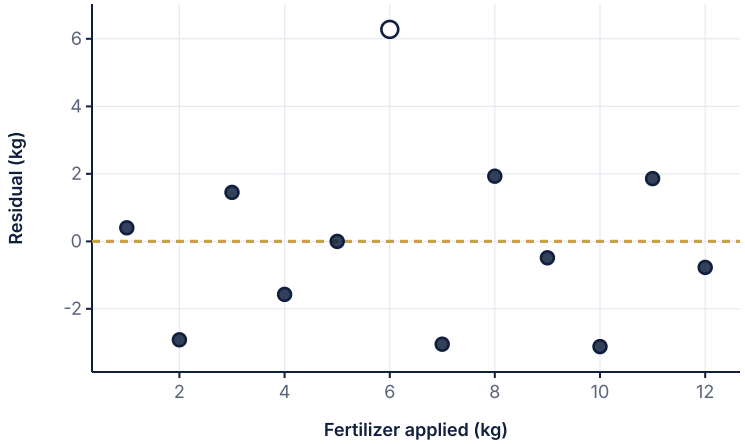
The scatterplot shows data for 10 lakes, with one lake highlighted far to the right. What effect would removing the highlighted point have on the correlation and on the slope of the least-squares line?

(A) The correlation would move closer to 0, and the slope would increase.
(B) The correlation would move closer to 1, and the slope would decrease.
(C) Neither would change much, because a single point out of 10 cannot affect the line.
(D) The correlation would move closer to 1, and the slope would increase.

41 A botanist measured the height of a seedling on days 5 through 30 after planting and found the least-squares line $\hat{y} = 2.1 + 1.8x$, where x is the day and $\hat{y}$ is the predicted height in centimeters, with a correlation of 0.99. She uses it to predict the height on day 90. Which statement is correct?

(A) The prediction of 164.1 cm is an extrapolation and may be unreliable because it is far outside the range of the data.
(B) The prediction of 164.1 cm is reliable because the correlation is so strong.
(C) The prediction of 164.1 cm is an interpolation, since it is a prediction from the line of best fit.
(D) The prediction of 164.1 cm is the residual for day 90.

42



Residual plot

The residual plot is from a least-squares regression of crop yield on the amount of fertilizer applied for 12 plots. The plot at fertilizer amount 6 is highlighted and has a residual of about 6 kg. Which is the correct interpretation?

(A) The actual yield of that plot was about 6 kg more than the yield predicted by the regression line.
(B) The actual yield of that plot was about 6 kg less than the yield predicted by the regression line.
(C) The regression line predicts a yield of 6 kg for that plot.
(D) The plot is an outlier in the x direction, so it is influential.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Exam 1

Section II: Free Response

Section II · 4 questions · 90 minutes · 50% of score

Each question is worth 10 points. Show your work and explain your reasoning in context, in complete sentences. Spend about 20 minutes on each of Questions 1 to 3 and the remaining time on Question 4 if you prefer, but plan your own pacing.

QUESTION 1 Homework and Sleep Survey · 10 points

A student council wants to know whether students at Lakeview High School, which has 1200 students (300 in each of grades 9 through 12), spend more time on homework than the council's goal of 60 minutes per night. A council member proposes posting a question on the school's social media page: 'Isn't it true that too much homework is ruining your sleep? How many minutes did you spend on homework last night?' The council then asks a statistics class to plan a better study. Later, a pilot group of 15 randomly selected students reports the following numbers of minutes spent on homework last night: 20, 25, 30, 30, 35, 40, 40, 45, 50, 55, 60, 60, 65, 75, 140.

(a) State a statistical question that the study could answer, and identify the population and the parameter of interest. [2 points]

(b) Describe a sampling method that would produce a sample of 80 students that is representative of all students at the school. Be specific about how randomness is used. [3 points]

(c) Identify one source of bias in the social media question, explain the direction in which it is likely to affect the results, and describe one way to eliminate it. [2 points]

(d) For the pilot data, find the five-number summary and use the 1.5 times IQR rule to decide whether 140 minutes is an outlier. [3 points]

QUESTION 2 Exercise and Resting Heart Rate · 10 points

A health clinic collected data on 10 adult patients: the minutes of exercise per week and the resting heart rate in beats per minute (bpm). A least-squares regression of resting heart rate on exercise time gives the following computer output: Constant = 79.74, Exercise slope = -0.0870, S = 1.25, R-sq = 96.6%.

Person	1	2	3	4	5	6	7	8	9	10
Exercise (min/week)	30	45	60	75	90	120	150	180	210	240
Resting heart rate (bpm)	78	76	75	72	73	68	66	64	60	61

(a) Interpret the slope of the regression line in context. [2 points]

(b) Interpret R-sq in context. [2 points]

(c) For the patient who exercises 90 minutes per week, find the predicted resting heart rate and the residual, and interpret the residual in context. [3 points]

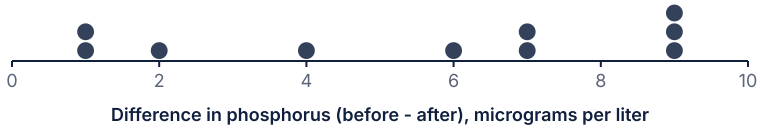
(d) A doctor concludes, 'The regression shows that exercising more causes a lower resting heart rate.' Explain why this conclusion is not justified by these data, and explain why using the line to predict the heart rate of a patient who exercises 600 minutes per week would be unwise. [3 points]

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

QUESTION 3 Lake Cleanup and Phosphorus · 10 points

Environmental scientists selected a random sample of 10 lakes from the lakes in a region that took part in a cleanup program. The phosphorus level (micrograms per liter) of each lake was measured before the program and again one year after the program. The differences, before minus after, are displayed in the dotplot. Before: 48, 62, 55, 70, 41, 66, 59, 52, 73, 64. After: 42, 55, 53, 61, 40, 57, 58, 45, 64, 60. The mean of the differences is 5.5 and the standard deviation of the differences is 3.274.



(a) Identify the procedure you would use to estimate the mean change in phosphorus level and define the parameter of interest. [2 points]

(b) Verify that the conditions for the procedure are met. [2 points]

(c) Construct and report a 95% confidence interval for the parameter. [3 points]

(d) Interpret the interval in context, and state what it suggests about whether the program lowered phosphorus levels in the region's lakes. [3 points]

QUESTION 4 Missed Appointments · 10 points

A community clinic has found that 8% of scheduled appointments are missed, and each appointment can be treated as independent. The clinic schedules 12 appointments for a morning, and then starts an outreach program of phone calls to patients. After the program starts, the clinic takes a random sample of 150 scheduled appointments and finds that 7 of them were missed.

(a) Before the program, let X be the number of missed appointments among the 12 scheduled in a morning. Find P(X >= 3). [2 points]

(b) Find the mean and standard deviation of X, and interpret the mean in context. [2 points]

(c) Assuming the true proportion of missed appointments is still 0.08, describe the sampling distribution of the sample proportion for random samples of 150 appointments (shape, center and spread), with justification. [2 points]

(d) Do the sample results provide convincing evidence at the 5% significance level that the program has reduced the proportion of missed appointments? Complete a significance test, including hypotheses, conditions, test statistic, P-value, and a conclusion in context. [4 points]

Key Practice Exam 1: Answers, Explanations, and Scoring

Score yourself honestly. For free response, grade each part against the rubric, awarding points only for work that is shown and explained in context.

SECTION I · ANSWER KEY

1 D 2 A 3 B 4 C 5 D 6 D 7 B 8 A 9 C 10 D 11 C 12 C 13 B 14 D 15 C 16 B 17 D 18 B 19 A 20 A 21 B 22 D 23 C 24 A 25 A 26 C 27 A 28 B 29 D 30 B 31 C 32 C 33 D 34 B 35 C 36 B 37 B 38 A 39 A 40 D 41 A 42 A

TOPIC MAP: WHICH QUESTIONS TEST WHICH UNIT

Topic	Questions
Exam Unit 1: One-Variable Data	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
Exam Unit 2: Probability	12, 13, 14, 15, 16, 17, 18, 19, 20
Exam Unit 3: Proportions	21, 22, 23, 24, 25, 26, 27
Exam Unit 4: Means	28, 29, 30, 31, 32, 33, 34
Exam Unit 5: Regression	35, 36, 37, 38, 39, 40, 41, 42

Sec I Multiple Choice Explanations

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA
The bulk of the data sits at low values (the tallest bars are at 2 to 4 hours) and a long tail stretches toward 14 hours, so the shape is skewed right. The 20th and 21st ordered values both fall in the 2 to 4 hour bar, so the median is at most 4, while the mean from bar midpoints is about 4.6. Long right tails pull the mean up toward the tail, so the mean exceeds the median. Calling it skewed left confuses the direction of the tail with the location of the peak, and the symmetric answer ignores the tail.

2. **Answer: A** · EXAM UNIT 1: ONE-VARIABLE DATA
The median line is at 74 for Class B and 72 for Class A, and the IQR is 78 - 68 = 10 for B versus 80 - 64 = 16 for A, so B is higher in the center and tighter in the middle half. The maximum does not determine the median, so the Class A claim fails. The range for B is 90 - 58 = 32, smaller than A's 94 - 52 = 42, so a short box does not mean a large range. Only half, not 75%, of Class A scored above its median of 72.

3. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA
A z-score measures how many standard deviations a value is from its mean: Maya's is (84 - 75)/6 = 1.5 and Jon's is (71 - 62)/5 = 1.8. Jon's score is farther above his class average in standard deviation units, so he did better relative to his class. Comparing raw scores ignores the different centers and spreads. Equal distances from the mean are not equal performances when the standard deviations differ. A larger standard deviation actually makes the same 9 points less impressive.

4. **Answer: C** · EXAM UNIT 1: ONE-VARIABLE DATA
Multiplying every value by 1.8 multiplies both the mean and the standard deviation by 1.8, and adding 32 shifts only the center. The mean is 1.8(20) + 32 = 68 and the standard deviation is 1.8(3) = 5.4. Adding 32 to the spread (giving 37.4) wrongly treats a shift as changing variability. Leaving out the +32 gives a mean of 36. Keeping the standard deviation at 3 forgets that rescaling stretches the distribution.

5. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA
Standardize: z = (65 - 52)/9 = 1.44, and the area to the right of z = 1.44 is about 0.0743. The value 0.9257 is the area to the left, the wrong tail for 'above'. The value 0.1486 doubles the tail as if the question asked about both extremes. The value 0.4363 comes from dividing by the variance 81 instead of the standard deviation 9, which gives a tiny z-score of 0.16.

6. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA
IQR = 26 - 14 = 12, so 1.5(IQR) = 18. The fences are 14 - 18 = -4 and 26 + 18 = 44. A value is an outlier only if it lies beyond a fence, so 45 is an outlier because it exceeds 44. The values 43 and 40 lie inside the upper fence. A time of -2 minutes is impossible in context, but by the rule it is still inside the lower fence of -4, so the rule does not flag it.

7. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA
When individuals decide for themselves whether to respond, those with strong feelings tend to answer and the rest do not, so the sample is likely to be biased and not representative of all students. A large sample size does not remove bias from a poor sampling method, which is why the 'larger than 30' claim fails. The size of 220 is not the problem; a well-chosen random sample of that size can be quite reliable. Online delivery alone does not push results in one fixed direction; the self-selection is what causes the bias.

8. **Answer: A** · EXAM UNIT 1: ONE-VARIABLE DATA
Grouping similar experimental units into blocks and randomizing within each block removes the variation due to the blocking variable (here, moisture) from the comparison of treatments. Random assignment is still required within each block, so it is not made unnecessary. Blinding concerns whether people know the treatments, which has nothing to do with the grouping. Each plot still receives only one fertilizer; applying both to each unit is a matched-pairs design, which is not what is described.

9. **Answer: C** · EXAM UNIT 1: ONE-VARIABLE DATA
Random sampling allows the association to be generalized to the population of teenagers sampled, but only random assignment of treatments in an experiment can support a cause-and-effect conclusion. Lurking variables such as a student's stress level or home environment could affect both sleep and grades. Random selection does not establish causation, so the first alternative fails. Reversing the direction of causation is also unsupported. An observational study can still reveal an association, so saying no conclusion can be drawn is too extreme.

10. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA
The schools are natural groups (clusters). The researcher randomly selects whole clusters and then surveys every member of the selected clusters, which is the definition of a cluster sample. A stratified sample would take a random sample from every group, which would require sampling from all 12 schools. It is not a simple random sample, because not every group of seniors in the county has an equal chance of being chosen; seniors at unselected schools have no chance. A systematic sample selects every kth individual from an ordered list, which was not done.

11. **Answer: C** · EXAM UNIT 1: ONE-VARIABLE DATA
The top 10% starts at the 90th percentile, where z = 1.28, so the cutoff is 40 + 1.28(6) = 47.7 mm. The value 32.3 mm uses z = -1.28, the 10th percentile, which is the cutoff for the driest months. The value 49.9 mm uses z = 1.645, the 95th percentile, so it is too high. The value 40.6 mm comes from multiplying 0.10 by the standard deviation instead of using the z-score for the 90th percentile.

12. **Answer: C** · EXAM UNIT 2: PROBABILITY
The probabilities of a discrete probability distribution must add to 1. The known probabilities sum to 0.50 + 0.30 + 0.15 = 0.95, so k = 1 - 0.95 = 0.05. The value 0.95 is the sum of the other probabilities, not the missing one. The values 0.10 and 0.20 do not make the total equal to 1 (the totals would be 1.05 and 1.15), so they would not be a valid distribution.

13. **Answer: B** · EXAM UNIT 2: PROBABILITY
The mean of a discrete random variable is the sum of each value times its probability: 0(0.50) + 2(0.30) + 5(0.15) + 20(0.05) = 0.60 + 0.75 + 1.00 = 2.35 dollars. The value 6.75 is the unweighted average of the four prizes, which ignores how likely each prize is. The value 5.35 adds the 3 dollar fee instead of subtracting it. The value 0.65 is the long-run average loss per play after paying 3 dollars, not the expected prize.

14. **Answer: D** · EXAM UNIT 2: PROBABILITY
The variance is the weighted average of squared deviations from the mean 2.35: (0 - 2.35)^2(0.50) + (2 - 2.35)^2(0.30) + (5 - 2.35)^2(0.15) + (20 - 2.35)^2(0.05) = 19.43. The standard deviation is the square root, about 4.41 dollars. The value 19.43 is the variance, which is in squared dollars. The value 4.99 is the square root of the mean of the squared prizes, forgetting to subtract the square of the mean. The value 2.35 is the mean.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

15. **Answer: C** · EXAM UNIT 2: PROBABILITY
This is a conditional probability, P(sport | music club), so the sample space is restricted to the 48 music club members; 18 of them play a sport, giving 18/48 = 0.375. The value 0.300 divides by the 60 athletes, which is P(music club | sport), the conditional probability in the reverse direction. The value 0.150 divides by all 120 students, giving the joint probability of both. The value 0.400 is the marginal probability of being in the music club.

16. **Answer: B** · EXAM UNIT 2: PROBABILITY
By the addition rule, P(A and B) = P(A) + P(B) - P(A or B) = 0.40 + 0.25 - 0.55 = 0.10. Since P(A)P(B) = 0.40(0.25) = 0.10 equals P(A and B), the events are independent. Because P(A and B) is not 0, they can occur together, so they are not mutually exclusive. Mutually exclusive events with positive probabilities can never be independent, so the second and third statements are impossible combinations, and the fourth fails because the product rule holds.

17. **Answer: D** · EXAM UNIT 2: PROBABILITY
Let X be the number made, a binomial random variable with n = 8 and p = 0.7. 'At least 6' means P(X \ge 6) = P(6) + P(7) + P(8) = 0.2965 + 0.1977 + 0.0576 = 0.5518. The value 0.2965 is only P(X = 6), leaving out 7 and 8. The value 0.2553 is P(X > 6), which leaves out 6. The value 0.7447 is P(X \le 6), which counts the wrong direction and includes outcomes with fewer than 6 made.

18. **Answer: B** · EXAM UNIT 2: PROBABILITY
A binomial setting has a fixed number of trials, each classified as success or failure, independent trials, and the same probability of success each time. The bulb inspection meets all four, and sampling 20 from thousands of bulbs makes the trials nearly independent. Heights are measured on a numerical scale, not classified as success or failure. Choosing 5 of 12 without replacement changes the probability of success from draw to draw. The total of die rolls adds values, rather than counting successes.

19. **Answer: A** · EXAM UNIT 2: PROBABILITY
The standard deviation of the sampling distribution of $\hat{p}$ is $\sqrt{p(1-p)/n}$, which depends on the square root of n. Multiplying n by 4 divides the standard deviation by 2, so it goes from 0.046 to 0.023. Dividing by 4 (0.011) treats the standard deviation as proportional to 1/n. Leaving it at 0.046 ignores sample size. Doubling it (0.092) goes in the wrong direction, because larger samples give less variability, not more.

20. **Answer: A** · EXAM UNIT 2: PROBABILITY
With n = 64, the Central Limit Theorem says the sampling distribution of $\bar{x}$ is approximately normal with mean 25 and standard deviation $10/\sqrt{64}$ = 1.25, even though the population is skewed. Then z = (23 - 25)/1.25 = -1.6 and the area below it is 0.0548. The value 0.9452 is the wrong tail. The value 0.4207 uses z = -0.2, which treats the standard deviation of one commute (10) as the standard deviation of the mean. Dividing by 64 instead of its square root gives a tiny probability.

21. **Answer: B** · EXAM UNIT 3: PROPORTIONS
The Large Counts condition for a z interval for a proportion requires at least 10 successes and at least 10 failures in the sample. Here there are 5 successes and 35 failures, so the number of successes is too small for the normal approximation. The rule of 30 applies to means, not to proportions. The 10% condition is a separate check about independence, not about large counts. Random selection is a different condition and does not make the Large Counts condition fail.

22. **Answer: D** · EXAM UNIT 3: PROPORTIONS
The point estimate is $\hat{p}$ = 62/150 = 0.413 and the standard error is $\sqrt{0.413(0.587)/150}$ = 0.040. The interval is 0.413 plus or minus 1.96(0.040), which gives (0.335, 0.492). The interval (0.347, 0.479) uses z* = 1.645, which is for 90% confidence. The interval (0.310, 0.517) uses z* = 2.576, which is for 99% confidence. The interval (0.373, 0.454) uses one standard error without a critical value, so it does not correspond to a 95% level.

23. **Answer: C** · EXAM UNIT 3: PROPORTIONS
The confidence level describes the long-run success rate of the method: in repeated sampling, about 95% of the intervals constructed this way capture the population proportion. Once a particular interval is calculated, the true proportion is a fixed number that is either inside or outside it, so no probability statement applies to this interval. The proportion refers to households, not to a rate of recycling over time. The interval is also not a range for sample proportions; it is an estimate of the population parameter.

24. **Answer: A** · EXAM UNIT 3: PROPORTIONS
Hypotheses are statements about a population parameter, and the null hypothesis is the claim of no effect, which has an equals sign. The nurse suspects the proportion is greater than 0.25, so the alternative is one-sided: $H_a : p > 0.25$. Writing the hypotheses in terms of $\hat{p}$ is wrong because a sample statistic is never the subject of a hypothesis. Putting the inequality in the null hypothesis reverses the roles. A two-sided alternative does not match her belief that the proportion is larger.

25. **Answer: A** · EXAM UNIT 3: PROPORTIONS
A P-value is computed assuming the null hypothesis is true, and it gives the probability of observing a result as extreme as, or more extreme than, the one obtained. It does not give the probability that the null hypothesis is true, since the null is a fixed claim and not a random event. It also does not give the probability that the alternative is true (96.6%). The size of the difference between the sample proportion and 0.60 is a different quantity entirely and cannot be read from the P-value.

26. **Answer: C** · EXAM UNIT 3: PROPORTIONS
The pooled proportion is (78 + 62)/(120 + 110) = 0.609, and the standard error is $\sqrt{0.609(0.391)(1/120 + 1/110)}$ = 0.064. The test statistic is z = (0.650 - 0.564)/0.064 = 1.34. Using 1/(n1 + n2) in the standard error gives 2.68, a formula that does not combine the two sample sizes correctly. Using only the first group's standard error gives 1.98. Reporting the raw difference of 0.09 is not a test statistic because it was never divided by a standard error.

27. **Answer: A** · EXAM UNIT 3: PROPORTIONS
With no prior estimate, use the conservative value 0.5 for the proportion: $n \ge (z^*/m)^2 p(1-p) = (1.96/0.03)^2(0.25)$, which is 1067.1. Because the sample must be at least that large to keep the margin of error at 0.03 or less, round up to 1068, not down to 1067. The value 1844 uses z* = 2.576 for 99% confidence. The value 267 uses a margin of error of 0.06, which is double the requirement.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

28. **Answer: B** · EXAM UNIT 4: MEANS
Because the population standard deviation is unknown, use a t interval with df = 15 and t* = 2.131. The standard error is $2.8/\sqrt{16} = 0.7$, so the margin of error is 2.131(0.7) = 1.49 and the interval is (10.91, 13.89). Using z* = 1.96 ignores the extra uncertainty from estimating the standard deviation and gives an interval that is too narrow. Dividing by 16 instead of $\sqrt{16}$ makes the interval far too narrow. Using t* = 2.947 corresponds to 99% confidence.

29. **Answer: D** · EXAM UNIT 4: MEANS
The t procedures require a random sample and a population that is approximately normal, or a large enough sample (roughly 30 or more) for the Central Limit Theorem to help. With only 12 values, the dotplot shows a strong right skew and an extreme value at 28, so there is no evidence of normality and n is too small to compensate. The t distribution corrects for estimating the standard deviation, not for non-normal data. A size of 10 is not enough when the data are this skewed. Finally, t procedures are used precisely when the population standard deviation is unknown.

30. **Answer: B** · EXAM UNIT 4: MEANS
The test statistic is $t = (47.2 - 50)/(6.5/\sqrt{25}) = -2.15$ with df = 24. Because the alternative is one-sided (less than), the P-value is the area below t = -2.15, which is 0.0208. The value 0.0415 is the two-sided P-value, which does not match the alternative hypothesis. The value 0.0156 uses a normal distribution instead of a t distribution with 24 degrees of freedom. The value 0.9792 is the area in the wrong tail.

31. **Answer: C** · EXAM UNIT 4: MEANS
The P-value of 0.086 is larger than 0.05, so we fail to reject the null hypothesis: the observed difference of 4.3 points is the kind of difference that could happen by chance about 9% of the time if the course type made no difference. The raw difference between sample means cannot be used on its own to conclude a real difference. Failing to reject never proves the null hypothesis true; it only says the evidence is insufficient. A P-value is not the probability that the null is true.

32. **Answer: C** · EXAM UNIT 4: MEANS
Each runner is measured twice, so the two sets of times are dependent and not independent samples. The correct approach is to compute the difference for each runner and perform a one-sample t procedure on those 15 differences. A two-sample t test assumes independent groups and would ignore the strong runner-to-runner variability the pairing removes. The data are quantitative times, so a proportion test is not appropriate. Using only the after times discards the before times, so it can say nothing about change.

33. **Answer: D** · EXAM UNIT 4: MEANS
A confidence interval for a difference in means that does not contain 0 provides convincing evidence of a difference, and its sign shows the direction: here the mean is greater at late-start schools by an estimated 0.12 to 1.04 hours. Being near zero does not mean that there is no difference, and an interval cannot prove equality. The interval estimates a difference in population means, not the range of individual students' differences. The final statement is false because 0 is not in the interval (the lower endpoint is 0.12).

34. **Answer: B** · EXAM UNIT 4: MEANS
The margin of error is $t^*(s/\sqrt{n})$. Quadrupling n from 25 to 100 divides $s/\sqrt{n}$ by 2 and also slightly lowers t*, so the interval is narrower. A higher confidence level requires a larger critical value, which widens the interval. A larger sample standard deviation increases $s/\sqrt{n}$ and widens it. Adding outliers would inflate s, again widening the interval, and it would also threaten the validity of the t procedure.

35. **Answer: C** · EXAM UNIT 5: REGRESSION
The slope is the predicted change in the response variable (exam score) for each one-unit increase in the explanatory variable (hours studied), so it is 4.58 points per hour. Regression gives predicted averages, not exact changes for every student, so 'increases by exactly' overstates what the model says. The statement about 0 hours describes the intercept, not the slope. Reversing the roles of the variables is wrong because the line predicts score from hours, not hours from score.

36. **Answer: B** · EXAM UNIT 5: REGRESSION
The predicted score for 5 hours is 50.83 + 4.58(5) = 73.7. A residual is the actual value minus the predicted value, so it is 75 - 73.7 = 1.3 points. The student scored about 1.3 points above the prediction, so the residual is positive. Reversing the subtraction (predicted minus actual) gives the wrong sign, -1.3. The value 24.2 subtracts only the intercept, and 73.7 is the predicted score itself, not a residual.

37. **Answer: B** · EXAM UNIT 5: REGRESSION
The coefficient of determination R-sq is the fraction of the variability in the response (exam score) that is explained by the linear regression on the explanatory variable, here about 97.8%. It is not a percentage of students whose scores are predicted correctly, and no individual prediction is exact. The correlation r is the square root of R-sq (r = 0.989, with the sign of the slope), not R-sq itself. R-sq says nothing about a percentage increase in scores per hour; that would be a slope interpretation.

38. **Answer: A** · EXAM UNIT 5: REGRESSION
The points fall close to a line that slopes downward, so the association is strong, linear, and negative, and r should be close to -1 but not exactly -1 because the points are scattered about the line. A value of -0.9 matches this. A value of -0.5 would show a much weaker association with a wider cloud. A positive value such as 0.9 has the wrong direction, because temperature decreases as elevation increases. A value of -0.1 would show almost no linear pattern.

39. **Answer: A** · EXAM UNIT 5: REGRESSION
The coefficient of determination is r squared, so $|r| = \sqrt{0.64} = 0.80$. Because the slope is negative, r has the same sign as the slope, so r = -0.80. Choosing 0.80 gives the wrong sign. The value -0.64 treats R-sq itself as the correlation. The value -0.41 squares 0.64 instead of taking the square root.

40. **Answer: D** · EXAM UNIT 5: REGRESSION
The highlighted point is far from the others in the x direction and well below the linear pattern of the rest, which makes it influential. It pulls the line down on the right and weakens the correlation, so with it removed the remaining points are almost perfectly linear. The slope (0.14 with the point, 1.43 without) and r (0.19 with, 0.95 without) both increase. A point that has an extreme x value can change the line a great deal, so the claim that one point cannot matter is false.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

41. **Answer: A** · EXAM UNIT 5: REGRESSION
Day 90 is far beyond the observed days 5 to 30, so using the line there is extrapolation. The linear pattern may not continue, since seedlings stop growing at some height. A strong correlation within the observed range does not guarantee the same pattern outside it. Interpolation means predicting within the range of the data, which is not the case. A residual is the difference between an observed and a predicted value, so a prediction itself is never a residual. The arithmetic 2.1 + 1.8(90) = 164.1 is correct but cannot be trusted.

42. **Answer: A** · EXAM UNIT 5: REGRESSION
A residual is the observed value minus the predicted value, so a positive residual of about 6 kg means the plot produced about 6 kg more than the line predicts for that amount of fertilizer. A negative residual would indicate a yield below the prediction, so the second statement has the wrong direction. The residual measures a vertical distance from the line, not the predicted value itself. Its position along the horizontal axis (6) is in the middle of the data, so it is not extreme in x and is not high leverage.

Sec II Free Response Scoring Guide and Model Responses

QUESTION 1 · HOMEWORK AND SLEEP SURVEY

Part / skill	Points	What earns the points
(a) Statistical question, population and parameter	2	1 point for a question with variability that can be answered with data (for example, what is the mean homework time of Lakeview students per night). 1 point for naming the population (all 1200 Lakeview students) and the parameter (the population mean time, or the proportion above 60 minutes).
(b) Valid random sampling plan	3	1 point for a probabilistic method such as a stratified random sample by grade. 1 point for explaining how randomness is used (list each grade, number the students, use a random number generator to pick 20 from each grade). 1 point for a sample that is representative or sized correctly (20 students from each of four equal grades gives 80).
(c) Identify bias, direction and fix	2	1 point for correctly identifying the leading wording (or voluntary response from posting on social media) and its likely direction (students reporting more time and agreeing homework is excessive). 1 point for a fix: use neutral wording, and use random selection of respondents instead of a voluntary post.
(d) Five-number summary and outlier	3	1 point for the five-number summary: min 20, Q1 30, median 45, Q3 60, max 140. 1 point for the IQR (30) and fences (-15 and 105). 1 point for concluding that 140 is greater than 105, so it is an outlier.

Model response

(a) A statistical question is: What is the mean number of minutes per night that Lakeview High School students spend on homework? The population is all 1200 Lakeview students. The parameter of interest is the population mean time spent on homework per night (mu), which can be compared with the goal of 60 minutes.

(b) I would use a stratified random sample, with the four grade levels as strata. Using a roster for each grade, I would number the 300 students in the grade 1 to 300 and use a random number generator to select 20 different students without repeats from each grade, giving 80 students in all. Stratifying guarantees that every grade is represented in proportion to its size, and the random selection within each stratum gives every student an equal chance of selection.

(c) The question is leading ('Isn't it true that too much homework is ruining your sleep?') and it also uses a voluntary post. Both would bias the results toward overestimating the time spent, since students who feel overloaded are most likely to answer and the wording suggests an answer. To fix this, I would ask a neutral question ('How many minutes did you spend on homework last night?') and survey only the randomly selected students.

(d) The data are already ordered. The minimum is 20, the median is the 8th value, 45, the first quartile is the median of the lower seven values, 30, the third quartile is the median of the upper seven values, 60, and the maximum is 140. The IQR is 60 - 30 = 30, so the fences are 30 - 1.5(30) = -15 and 60 + 1.5(30) = 105. Since 140 > 105, 140 minutes is an outlier by the 1.5 IQR rule.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Common point-losing errors

Describing a 'random sample' without saying how randomness is used (no list, no random number generator). Calling the sample stratified but then picking volunteers or friends from each grade. Describing the bias but not its direction, or naming a fix that is not tied to the bias (such as 'use a bigger sample'). In (d), using the mean to find fences, forgetting that Q1 and Q3 are medians of the halves, or comparing 140 to Q3 and not to the upper fence.

QUESTION 2 · EXERCISE AND RESTING HEART RATE

Part / skill	Points	What earns the points
(a) Slope interpretation	2	1 point for stating that the predicted resting heart rate changes by -0.087 bpm (decreases about 0.09 bpm) for each additional minute of weekly exercise. 1 point for context-specific wording about predicted or average value.
(b) R-sq interpretation	2	1 point for stating that about 97% of the variation in resting heart rate is explained by the linear relationship with exercise time. 1 point for context (resting heart rate, exercise minutes) and the word variation.
(c) Prediction, residual and interpretation	3	1 point for the predicted value of 71.9 bpm. 1 point for residual = 73 - 71.9 = 1.1 bpm (actual minus predicted). 1 point for interpreting it: this patient's actual rate is about 1.1 bpm above the rate predicted for 90 minutes of exercise.
(d) Causation and extrapolation	3	1 point for explaining that the data are observational (no random assignment), so association does not prove causation. 1 point for a lurking variable (for example, overall health, age, or diet). 1 point for identifying 600 minutes as far beyond the observed 30 to 240 minute range (extrapolation), so the linear pattern may not hold.

Model response

(a) For each additional minute of weekly exercise, the predicted resting heart rate decreases by about -0.087 bpm. (Equivalently, each additional hour is associated with a predicted decrease of about 5.2 bpm.)

(b) About 97% of the variation in the patients' resting heart rates is accounted for by the linear relationship with weekly exercise time. The remaining 3% is variation due to other factors and individual differences.

(c) The predicted resting heart rate for 90 minutes is 79.74 + (-0.0870)(90) = 71.9 bpm. The residual is actual minus predicted: 73 - 71.9 = 1.1 bpm. This patient's resting heart rate is about 1.1 bpm higher than the heart rate the regression line predicts for someone who exercises 90 minutes per week.

(d) The data come from an observational study in which the clinic recorded the patients' habits; no one was randomly assigned to an exercise level. Thus, the relationship is an association and does not demonstrate that exercise causes a lower heart rate. A lurking variable such as general fitness, age, or diet may influence both exercise time and heart rate. Also, the line was built from patients who exercise between 30 and 240 minutes per week. A prediction for 600 minutes is extrapolation, far outside the data, and the linear trend is unlikely to continue (the line would predict an unrealistically low heart rate), so the prediction would be unreliable.

Common point-losing errors

Interpreting the slope as an exact change for every person or omitting 'predicted' and the units (bpm per minute). Describing R-sq as the percentage of points on the line, or saying '97% of heart rate' is explained without the word variation. Computing the residual as predicted minus actual and getting the wrong sign. Claiming causation from an observational study, or mentioning 'correlation is not causation' without a specific lurking variable or the reason (no random assignment).

QUESTION 3 · LAKE CLEANUP AND PHOSPHORUS

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Part / skill	Points	What earns the points
(a) Procedure and parameter	2	1 point for a matched-pairs (one-sample) t interval for a mean difference. 1 point for defining mu_d as the true mean difference (before minus after) in phosphorus level for all lakes in the program in the region.
(b) Conditions	2	1 point for random sample of lakes and the 10% condition (or lakes fewer than 10% of lakes in the region). 1 point for normality: n = 10 is small, but the dotplot of differences shows no strong skewness or outliers, so approximate normality is reasonable.
(c) Interval calculation	3	1 point for correct t* with df = 9 (2.262). 1 point for correct standard error or margin of error (2.342). 1 point for the interval (3.16, 7.84).
(d) Interpretation and conclusion	3	1 point for interpreting the confidence level (95% confident that the interval captures the true mean difference). 1 point for stating the interval in context with units. 1 point for concluding that, since all values in the interval are positive (0 is not included), there is convincing evidence the program reduced the mean phosphorus level.

Model response

(a) Because each lake was measured twice, the data are matched pairs. I will use a one-sample t interval for a mean difference. Let mu_d be the true mean difference (before minus after) in phosphorus level, in micrograms per liter, for all lakes in this region that took part in the cleanup program.

(b) Conditions: Random: the 10 lakes are a random sample of the lakes in the region. 10% condition: the 10 lakes are less than 10% of all the lakes in the program. Normal/Large sample: the sample size is small (n = 10), so I need to check the dotplot of differences. The dotplot has no strong skewness and no outliers, so the differences are roughly symmetric and it is reasonable to treat the population of differences as approximately normal.

(c) The sample mean difference is 5.5 and the standard deviation is 3.274, with df = 10 - 1 = 9, so t* = 2.262. The interval is 5.5 +/- 2.342(3.274/sqrt(10)) = 5.5 +/- 2.342, which gives (3.16, 7.84) micrograms per liter.

(d) We are 95% confident that the true mean decrease in phosphorus level (before minus after) for lakes in this region's program is between 3.16 and 7.84 micrograms per liter. If we repeated this sampling many times, about 95% of the intervals constructed this way would capture the true mean difference. Because the entire interval is above 0, there is convincing evidence that the program lowers the mean phosphorus level of the lakes.

Common point-losing errors

Using a two-sample t interval for before and after values from the same lakes. Checking normality of the original before and after levels instead of the differences, or saying 'n < 30 so it is fine' without looking at the dotplot. Using z* = 1.96 or df = 10 instead of t with df = 9. Interpreting the confidence level as a probability that the sample mean is in the interval, or failing to connect the interval to the claim that 0 is not included, or omitting units and context.

QUESTION 4 · MISSED APPOINTMENTS

Part / skill	Points	What earns the points
(a) Binomial probability	2	1 point for identifying X as binomial with n = 12 and p = 0.08 (or equivalent setup). 1 point for P(X >= 3) = 1 - P(X <= 2) = 0.0652.
(b) Mean, SD and interpretation	2	1 point for mean 12(0.08) = 0.96 and SD sqrt(12(0.08)(0.92)) = 0.940. 1 point for interpreting the mean: over many mornings, the long-run average number of missed appointments out of 12 is about 0.96.
(c) Sampling distribution of p-hat	2	1 point for center 0.08 and standard deviation sqrt(0.08(0.92)/150) = 0.0222. 1 point for approximately normal shape justified by np = 12 >= 10 and n(1-p) = 138 >= 10.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Part / skill	Points	What earns the points
(d) Significance test	4	1 point for hypotheses H0: p = 0.08 and Ha: p < 0.08 and naming the one-sample z test for a proportion. 1 point for conditions (random sample, 10% condition, np0 >= 10 and n(1-p0) >= 10). 1 point for test statistic z = -1.50 and P-value = 0.0662. 1 point for a conclusion linked to the P-value in context: P-value > 0.05, so fail to reject H0; the evidence does not convincingly show the program reduced the missed rate.

Model response

(a) X is binomial with n = 12 and p = 0.08. P(X >= 3) = 1 - P(X <= 2) = 1 - binomcdf(12, 0.08, 2) = 0.0652.

(b) The mean is np = 12(0.08) = 0.96 and the standard deviation is sqrt(np(1-p)) = sqrt(12(0.08)(0.92)) = 0.940. Over many mornings with 12 scheduled appointments, the average number of missed appointments is about 0.96 per morning.

(c) For samples of n = 150, the sampling distribution of p-hat has a mean of 0.08 and a standard deviation of sqrt(0.08(0.92)/150) = 0.0222. Its shape is approximately normal because np = 150(0.08) = 12 >= 10 and n(1-p) = 150(0.92) = 138 >= 10.

(d) I will use a one-sample z test for a proportion. H0: p = 0.08 versus Ha: p < 0.08, where p is the true proportion of missed appointments after the outreach program. Conditions: the appointments are a random sample; 150 is less than 10% of all appointments the clinic will schedule; and np0 = 12 >= 10 and n(1-p0) = 138 >= 10, so the sampling distribution is approximately normal. The sample proportion is 7/150 = 0.0467, so z = (0.0467 - 0.08)/0.0222 = -1.50 and the P-value is P(Z < -1.50) = 0.0662. Since the P-value of 0.066 is greater than alpha = 0.05, we fail to reject H0. There is not convincing evidence that the outreach program reduced the proportion of missed appointments below 8%; the sample result could plausibly happen by chance if the rate were still 8%.

Common point-losing errors

In (a), computing P(X = 3) or P(X > 3) instead of P(X >= 3), or using the wrong complement (1 - P(X <= 3)). Writing the interpretation of the mean as 'exactly 0.96 appointments will be missed' instead of a long-run average. In (c), checking the large counts condition with p-hat in place of p, or omitting the justification for normality. In (d), using the sample proportion in the standard error of a test, writing the hypotheses with p-hat, using a two-sided P-value, or concluding 'accept H0' or 'the program does not work' instead of 'not convincing evidence'.

Score

Scoring Your Practice Exam

Section I counts 50% and Section II counts 50%. Convert each section to a 0 to 50 scale, then add.

Step	Calculation	Your result
Section I	(number correct / 42) x 50	
Section II	(total points / 40) x 50	
Composite	Section I + Section II (0 to 100)	

About the score estimate

The College Board sets its cut scores for each exam year and does not publish them in advance. Scholar Coach uses this rough guide for practice only: composite 80 or above is in the range of a 5, 65 to 79 a 4, 50 to 64 a 3, 35 to 49 a 2, and below 35 a 1. Treat these as estimates, not predictions, and look at the topic map to decide what to study next.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice Exam 2

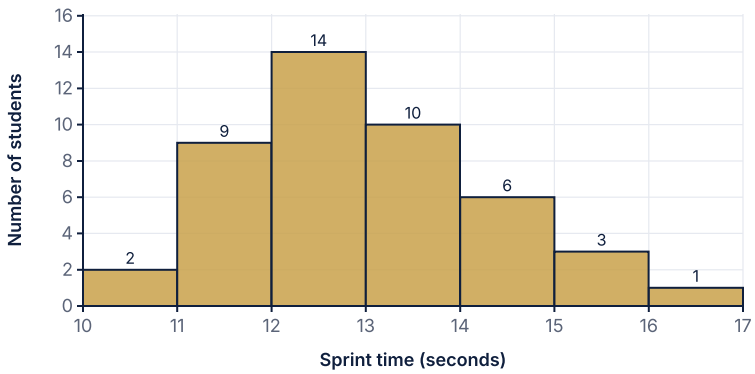
42 multiple-choice questions and 4 free-response questions in the digital exam format, followed by every answer, explanation, rubric, and a scoring guide.

Exam 2 Section I: Multiple Choice

Section I · 42 questions · 90 minutes · 50% of score

Choose the best answer for each question. A graphing calculator is allowed, and the formula sheet and tables in the appendix are available. Questions in a set share one stimulus.

1

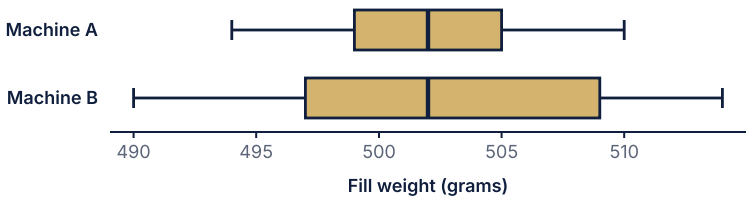


45 sprint times

The histogram shows the 100-meter sprint times, in seconds, of 45 sophomores in a school track program. Which statement about the center of this distribution is most accurate?

(A) The distribution is skewed right, so the mean is less than the median.
(B) The distribution is symmetric, so the mean equals the median.
(C) The distribution is skewed right, so the mean is greater than the median.
(D) The distribution is skewed left, so the mean is greater than the median.

2



Fill weights

A cereal plant checks two filling machines. The boxplots show the fill weights, in grams, of random samples of boxes from Machine A and Machine B. Which conclusion is best supported by the graph?

(A) Machine B has a larger median fill weight than Machine A.
(B) The two machines have the same median fill weight, but Machine B is more variable.
(C) Machine A fills boxes with more variability because its box is shorter.
(D) Machine A has more outliers than Machine B.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

3 The battery life of a certain tablet is approximately normally distributed with mean 14.0 hours and standard deviation 2.5 hours. About what proportion of these tablets have a battery life under 10 hours?

(A) 0.945
(B) 0.110
(C) 0.055
(D) 0.016

4 The heights of players in a basketball league have mean 75 inches and standard deviation 3.2 inches. A statistician converts every height to centimeters by multiplying by 2.54. What are the mean and standard deviation of the heights in centimeters?

(A) Mean 190.5 cm, standard deviation 3.2 cm
(B) Mean 190.5 cm, standard deviation 8.13 cm
(C) Mean 77.54 cm, standard deviation 5.74 cm
(D) Mean 190.5 cm, standard deviation 20.65 cm

5 The first and third quartiles of the boot-up times, in seconds, for a sample of game consoles are 34 and 46. Using the 1.5 x IQR rule, which of the following boot-up times would be identified as an outlier?

(A) 70 seconds
(B) 64 seconds
(C) 58 seconds
(D) 18 seconds

6 A music-streaming company adds a pop-up asking users to rate a new playlist feature. Of the 3,200 users who chose to respond, 84% gave a positive rating. Which is the most serious weakness in using the 84% to describe all users?

(A) Voluntary response bias: users with strong opinions are more likely to respond.
(B) The sample size of 3,200 is too small to generalize to all users.
(C) The question was necessarily worded to lead users toward a positive answer.
(D) The users were not randomly assigned to treatments.

7 A manufacturer has three production shifts. To estimate the mean number of defects per worker, the quality manager randomly selects 20 workers from each shift's roster and records each worker's defects. This sampling method is best described as

(A) a cluster sample, because the workers are grouped by shift.
(B) a simple random sample, because workers within a shift were randomly chosen.
(C) a stratified random sample, which guarantees every shift is represented.
(D) a systematic sample, because the same number is taken from each group.

8 Researchers want to compare two practice methods for 40 volunteer music students: 20 pianists and 20 violinists. Within each instrument group, students are randomly assigned to Method 1 or Method 2. What is the main purpose of grouping students by instrument before randomizing?

(A) To ensure the results can be generalized to all music students
(B) To make the experiment double-blind
(C) To eliminate the effect of every possible confounding variable
(D) To reduce variability in the response due to differences between instruments

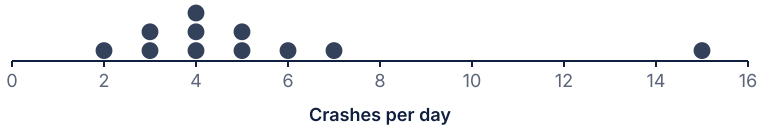
9 A survey of 500 high school athletes found that those who reported sleeping at least 8 hours per night had a higher average free-throw percentage than those who slept less. Which conclusion is justified?

(A) More sleep causes athletes to shoot better free throws.
(B) Athletes should sleep more because the survey shows it raises their free-throw percentage.
(C) There is no relationship, because a survey cannot measure free-throw percentage.
(D) There is an association between sleep and free-throw percentage, but cause and effect cannot be concluded.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10



The dotplot shows the number of crashes reported by a mobile app on each of 11 days. The mean is about 5.3 and the median is 4. Which statement is the best choice for summarizing the typical number of daily crashes?

(A) The mean of 5.3, because it uses every value in the data set.
(B) The mean of 5.3, because the distribution is symmetric.
(C) The median of 5, because it is the middle of the horizontal axis.
(D) The median of 4, because the one unusually high day pulls the mean upward.

11 Scores on a processor benchmark are approximately normally distributed with mean 500 and standard deviation 80. What minimum score is needed to be in the top 10% of all scores?

(A) Approximately 632
(B) Approximately 603
(C) Approximately 398
(D) Approximately 580

12 In a music app, let A be the event that a randomly selected user streams pop music and B the event that the user streams jazz. Suppose P(A) = 0.50, P(B) = 0.30, and P(A or B) = 0.65. Which statement is true?

(A) A and B are independent but not mutually exclusive.
(B) A and B are mutually exclusive but not independent.
(C) A and B are both independent and mutually exclusive.
(D) A and B are neither independent nor mutually exclusive.

13

	Defective	Not defective	Total
Plant X	18	482	500
Plant Y	24	276	300
Plant Z	8	192	200
Total	50	950	1000

A company tested 1,000 phones from three plants for a battery defect, with the results shown. A phone is chosen at random from the 50 that were defective. What is the probability it came from Plant Y?

(A) 0.48
(B) 0.08
(C) 0.024
(D) 0.30

14 A basketball player makes 70% of free throws, and attempts are independent. What is the probability that she makes exactly 6 of her next 8 attempts?

(A) 0.296
(B) 0.552
(C) 0.118
(D) 0.745

15 Each of 40 independent emails sent by a tech company is opened with probability 0.35. Let X be the number opened. What are the mean and standard deviation of X?

(A) Mean 14, standard deviation 9.1
(B) Mean 14, standard deviation 3.74
(C) Mean 26, standard deviation 3.02
(D) Mean 14, standard deviation 3.02

16 Suppose 60% of all voters in a large city favor a new stadium. Random samples of size 150 are repeatedly taken, and the sample proportion $\hat{p}$ who favor it is recorded each time. What is the standard deviation of the sampling distribution of $\hat{p}$?

(A) 0.040
(B) 0.0016
(C) 0.0033
(D) 0.490

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

17 The waiting time at a concert gate is strongly skewed right, with mean 12 minutes and standard deviation 9 minutes. For random samples of 50 fans, which describes the sampling distribution of the sample mean?

(A) Approximately normal with mean 12 and standard deviation about 1.27
(B) Skewed right with mean 12 and standard deviation 9
(C) Approximately normal with mean 12 and standard deviation 9
(D) Skewed right with mean 12 and standard deviation about 1.27

18 An audio app plays X advertisements during a random listening session. The probability distribution of X is shown in the table, where k is a constant.

x (ads)	0	1	2	3	4
P(X = x)	0.35	0.30	0.20	k	0.05

What is the value of k?

(A) 0.15
(B) 0.10
(C) 0.20
(D) 0.90

19

x (ads)	0	1	2	3	4
P(X = x)	0.35	0.30	0.20	k	0.05

What is the expected number of advertisements in a session?

(A) 2.0
(B) 1.36
(C) 0.30
(D) 1.2

20

x (ads)	0	1	2	3	4
P(X = x)	0.35	0.30	0.20	k	0.05

Given that a session has at least one advertisement, what is the probability that it has at least two?

(A) 0.350
(B) 0.538
(C) 0.308
(D) 0.650

21 A coffee shop wants to estimate the proportion of its customers who would buy a new seasonal drink. In a random sample of 40 customers out of thousands, 6 say they would buy it. Before constructing a one-proportion z-interval, which condition has NOT been satisfied?

(A) The sample must be randomly selected.
(B) The sample size must be less than 10% of the population.
(C) The sample must include at least 10 successes and at least 10 failures.
(D) The population of customers must be normally distributed.

22 A 95% confidence interval for the proportion of all fans who prefer an earlier game time is (0.42, 0.54). Which is the best interpretation of the 95% confidence level?

(A) 95% of all fans have a preference between 0.42 and 0.54.
(B) There is a 95% probability that the sample proportion lies between 0.42 and 0.54.
(C) If many random samples were taken and an interval computed from each, about 95% of the intervals would capture the true proportion.
(D) 95% of the sample proportions from future samples will fall between 0.42 and 0.54.

23 In a random sample of 300 listeners, 132 said they use a streaming service daily. Which is a 90% confidence interval for the proportion of all listeners who use it daily?

(A) (0.384, 0.496)
(B) (0.403, 0.477)
(C) (0.346, 0.534)
(D) (0.393, 0.487)

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

24 A player claims to make 50% of her shots. A coach tests H0: p = 0.50 versus Ha: p > 0.50 using a random sample of her shots and obtains a P-value of 0.032. Which is the correct interpretation of the P-value?

(A) The probability that the player's true proportion is 0.50 is 0.032.
(B) If the true proportion were 0.50, the probability of getting a sample proportion at least this large by chance alone is 0.032.
(C) The probability that the player's true proportion is greater than 0.50 is 0.968.
(D) The player makes 3.2% of her shots.

25 A factory claims at most 10% of its sensors are defective. In a random sample of 400 sensors, 52 are defective. For H0: p = 0.10 versus Ha: p > 0.10, what is the P-value (using a z-test)?

(A) 0.046
(B) 0.038
(C) 0.023
(D) 0.977

26

Statistic	Value
Version A: successes / n	96 / 200 (0.48)
Version B: successes / n	72 / 200 (0.36)
Difference (A - B)	0.12
z (pooled)	2.43
P-value (Ha: pA > pB)	0.0075

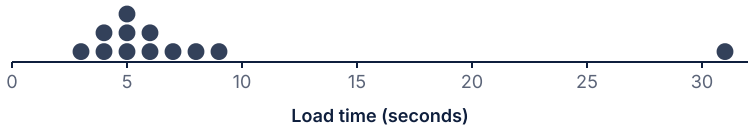
In a randomized experiment, 400 new users were assigned at random to Version A or Version B of a tutorial (200 each). The output for testing H0: pA - pB = 0 versus Ha: pA - pB > 0, where p is the proportion who completed the tutorial, is shown. Which conclusion is appropriate at the 5% significance level?

(A) Because the P-value is less than 0.05, we fail to reject H0.
(B) There is a 0.75% chance that the null hypothesis is true.
(C) The difference of 12 percentage points is not significant because it is not close to 0.75.
(D) There is convincing evidence that Version A leads to a higher completion rate than Version B.

27 A pollster wants a 95% confidence interval for a population proportion with margin of error no more than 0.03, and has no prior estimate of the proportion. Which is the smallest sample size that guarantees this?

(A) 1067
(B) 1068
(C) 385
(D) 2135

28



A website engineer records the load times (in seconds) of 12 randomly chosen page requests and wants to construct a t-interval for the mean load time. The dotplot of the sample is shown. Why is the t-interval not appropriate here?

(A) The sample is small and strongly skewed with an outlier, so the normality condition is not met.
(B) The population standard deviation is unknown, so a t-interval cannot be used.
(C) A sample of 12 is too small for any t-interval, regardless of shape.
(D) The requests are not independent because they were randomly chosen.

29 A random sample of 25 batteries has a mean life of 48.2 hours and a standard deviation of 6.5 hours. Assuming the data are roughly symmetric, which is a 95% confidence interval for the true mean battery life?

(A) (45.7, 50.8)
(B) (34.8, 61.6)
(C) (47.7, 48.7)
(D) (45.5, 50.9)

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

30 A machine is supposed to fill bottles with a mean of 500 mL. A random sample of 16 bottles has mean 496.2 mL and standard deviation 5.6 mL, and the volumes appear approximately normal. For H0: mu = 500 versus Ha: mu not equal to 500, which result and decision are correct at the 5% level?

(A) t = -2.71, P-value is about 0.008, reject H0
(B) t = -0.68, P-value is about 0.51, fail to reject H0
(C) t = -2.71, P-value is about 0.016, reject H0
(D) t = -2.71, P-value is about 0.016, fail to reject H0

31 Twelve runners each run a 5-kilometer course twice, once wearing Shoe A and once wearing Shoe B, with the order randomized. To test whether the mean time differs between shoe types, the appropriate procedure is

(A) a two-sample t-test comparing the mean times for Shoe A and Shoe B.
(B) a paired t-test on the 12 differences in time.
(C) a one-sample t-test on the Shoe A times only.
(D) a two-proportion z-test comparing the proportions of fast runs.

32

Statistic	Value
Platform A: n, mean, s	18, 62.4, 8.1
Platform B: n, mean, s	20, 57.9, 9.3
Difference in means	4.5
Standard error	2.82
t	1.59
df	35.97
P-value (two-sided)	0.120
95% confidence interval	(-1.23, 10.23)

Two independent random samples of students took a coding quiz after using different tutorial platforms. The two-sample t-test output for H0: muA - muB = 0 versus Ha: muA - muB not equal to 0 is shown. Which conclusion is appropriate at the 5% significance level?

(A) There is convincing evidence that the mean scores differ for the two platforms.
(B) There is not convincing evidence that the mean scores differ for the two platforms.
(C) There is convincing evidence that the two platforms have equal mean scores.
(D) There is a 12% chance that the null hypothesis is true.

33 A researcher constructs a 95% t-interval for a mean. Assuming the sample standard deviation stays about the same, how must the sample size change to cut the margin of error in half?

(A) Multiply the sample size by about 4.
(B) Multiply the sample size by about 2.
(C) Divide the sample size by 2.
(D) Increase the sample size by about 50%.

34 A 95% confidence interval for the mean battery life, in hours, of a new phone model is (12.3, 15.9). A manager uses the same data to test H0: mu = 16 versus Ha: mu not equal to 16 at the 5% significance level. What can be concluded?

(A) Fail to reject H0, because 16 is close to the upper end of the interval.
(B) Fail to reject H0, because the interval contains the sample mean.
(C) Reject H0, because 16 is not in the 95% confidence interval.
(D) Nothing can be concluded without the sample standard deviation.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

35 A coach records the height (in inches) and weight (in pounds) of players and finds the correlation is r = 0.62. If the heights are converted to centimeters and the weights to kilograms, what is the correlation between the new variables?

(A) 0.62 multiplied by 2.54 / 2.2
(B) Between 0 and 0.62, because the units are smaller
(C) Cannot be determined without the new data
(D) 0.62

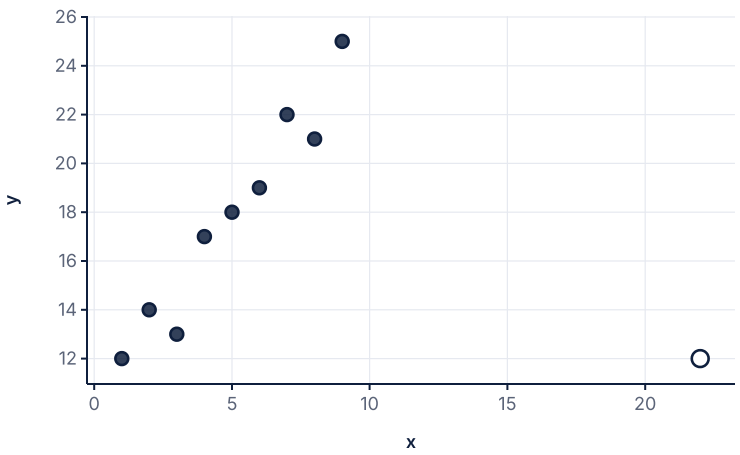
36 A trainer models an athlete's vertical jump y (in cm) from weekly training hours x, using the least-squares regression line y-hat = 12.4 + 3.1x. An athlete who trains 8 hours per week jumps 41 cm. What is the residual for this athlete?

(A) -3.8 cm
(B) 37.2 cm
(C) 3.8 cm
(D) 41 cm

37 For a sample of teenagers, a regression of hours of sleep on daily minutes of social media use gives r = -0.80. Which is the correct interpretation of r-squared?

(A) About 80% of the variation in sleep hours is explained by the linear relationship with social media minutes.
(B) Sleep hours decrease by 64% for each additional minute of social media.
(C) About 64% of the variation in sleep hours is explained by the linear relationship with social media minutes.
(D) 64% of the teenagers fall exactly on the regression line.

38



Highlighted point is (22, 12)

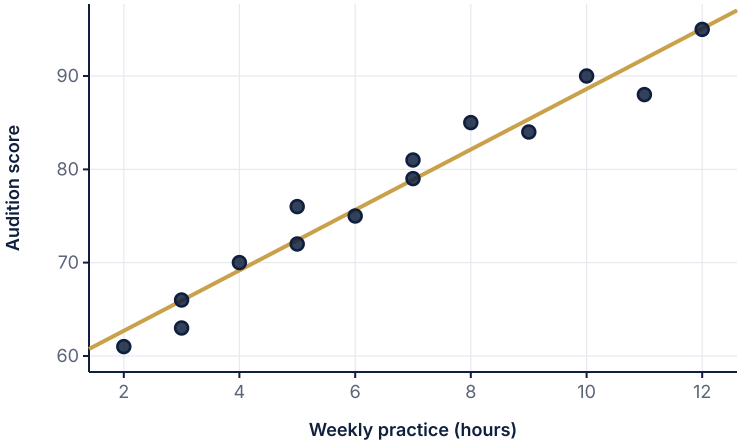
The scatterplot shows nine points that follow a clear increasing linear trend, plus one additional point labeled (22, 12) far to the right. Compared with the least-squares line fit to only the nine points, how will the line fit to all ten points differ?

(A) It will have a much smaller slope, and the correlation will be closer to 0.
(B) It will have a larger slope, and the correlation will be closer to 1.
(C) It will have a similar slope, because the point's y-value is within the range of the other y-values.
(D) It will have a smaller slope, and the correlation will be closer to 1.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

39 A music teacher recorded the weekly practice time x (hours) and the audition score y (points) for 14 students. The scatterplot and computer output from a least-squares regression are shown.



14 students

Term	Coefficient
Constant	56.20
Hours	3.241
S (typical residual size)	2.179
R-sq	95.9%

Which is the correct interpretation of the slope of the least-squares regression line?

(A) For each additional hour of weekly practice, the audition score increases by exactly 3.24 points.

(B) For each additional point of audition score, the practice time increases by 3.24 hours.

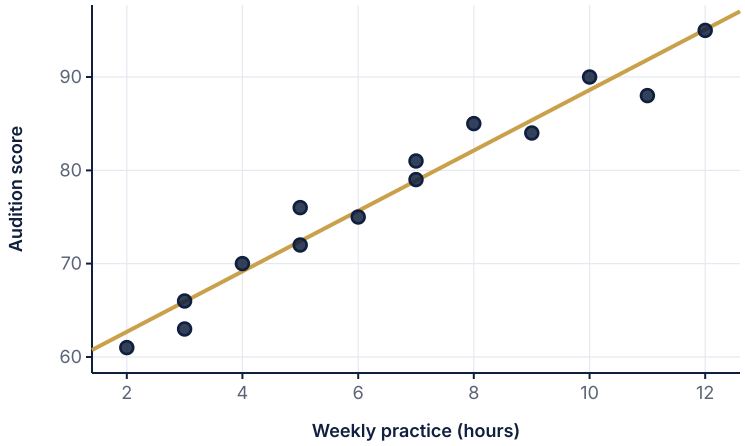
(C) 3.24% of the variation in audition scores is explained by practice time.

(D) For each additional hour of weekly practice, the predicted audition score increases by about 3.24 points.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

40



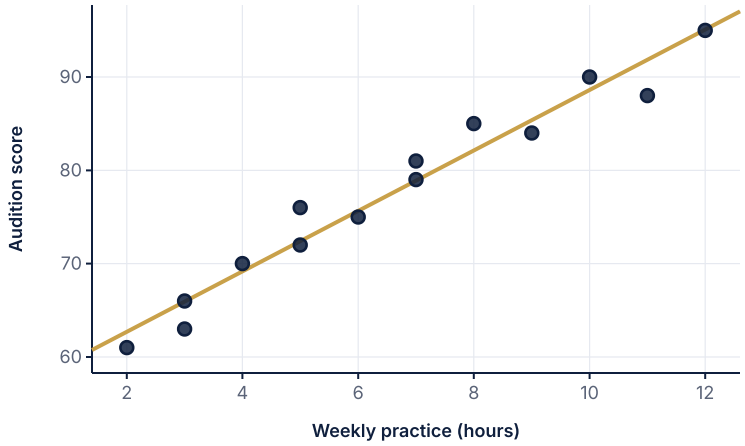
14 students

Term	Coefficient
Constant	56.20
Hours	3.241
S (typical residual size)	2.179
R-sq	95.9%

One student practices 9 hours per week and has an audition score of 84. What is the residual for this student?

(A) 1.37 points
(B) -1.37 points
(C) 85.37 points
(D) 84 points

41



14 students

Term	Coefficient
Constant	56.20
Hours	3.241
S (typical residual size)	2.179
R-sq	95.9%

What is the correlation coefficient r between weekly practice time and audition score?

(A) 0.959
(B) 0.979
(C) -0.979
(D) 0.920

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

42 A coach fits the line y-hat = 5.9 - 0.08x for sprint time y (seconds) and weeks of training x, using athletes who trained between 4 and 12 weeks. For which use of the model are the predictions most trustworthy?

(A) Predicting the sprint time of an athlete who trains for 8 weeks
(B) Predicting the sprint time of an athlete who trains for 30 weeks
(C) Predicting the sprint time of an athlete who has not trained (0 weeks)
(D) Predicting the sprint time of an athlete who trains for 20 weeks

Exam 2 Section II: Free Response

Section II · 4 questions · 90 minutes · 50% of score

Each question is worth 10 points. Show your work and explain your reasoning in context, in complete sentences. Spend about 20 minutes on each of Questions 1 to 3 and the remaining time on Question 4 if you prefer, but plan your own pacing.

QUESTION 1 Practice Time at a Music School · 10 points

A community music school has 1,800 students: 720 study piano, 600 study strings, and 480 study brass or percussion. The director wants to estimate the mean number of minutes students practice on a typical day. The director first proposes surveying the students who happen to be in the lobby on Saturday morning. A better plan is to select a sample of 30 students. A sample of 15 students gave the daily practice times (minutes) shown in the table.

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Minutes	20	25	30	30	35	40	40	45	45	50	55	60	65	75	130

(a) Describe how to select a stratified random sample of 30 students, using the three instrument groups as strata. State how many students come from each group and how the students would be chosen. [3 points]

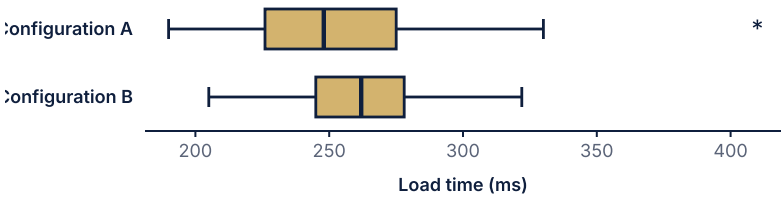
(b) Explain why the Saturday-lobby plan is likely to produce a biased estimate, and state the direction of the bias. [2 points]

(c) For the 15 practice times, find the five-number summary and use the 1.5 x IQR rule to determine whether there are any outliers. [3 points]

(d) Would the mean or the median be the better summary of a typical practice time for these students? Justify your choice using the data. [2 points]

QUESTION 2 Page Load Times for Two Server Configurations · 10 points

A web company tests two server configurations, A and B. For each configuration, the load time (in milliseconds) of 50 randomly selected page requests was recorded. The side-by-side boxplots summarize the data. For Configuration B, the load times are approximately normally distributed with mean 262 ms and standard deviation 24 ms.



50 requests per configuration

(a) Compare the distributions of load times for the two configurations, addressing center, spread, shape, and unusual values. [4 points]

(b) For Configuration B, what proportion of page requests take longer than 300 ms? [2 points]

(c) For Configuration B, what load time marks the 90th percentile? Interpret this value in context. [2 points]

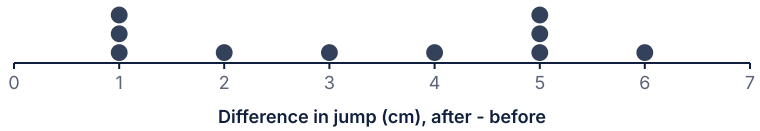
(d) The company wants customers to have consistent load times. Which configuration would you recommend? Justify your answer. [2 points]

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

QUESTION 3 A New Jump-Training Drill · 10 points

A coach claims that a new plyometric drill increases the vertical jump of players. A random sample of 10 players from a league of more than 200 players had their vertical jumps (in cm) measured before and after a 6-week program that included the drill. The table shows the data and the dotplot shows the 10 differences (after - before), which have mean 3.3 cm and standard deviation 1.95 cm.



Difference in jump (cm), after - before

Player	1	2	3	4	5	6	7	8	9	10
Before	52	58	61	49	66	55	60	57	63	54
After	55	60	66	50	71	56	64	62	64	60

(a) State the null and alternative hypotheses for a test of the coach's claim. Define any parameter you use. [2 points]

(b) Verify that the conditions for a paired t-test are met. [2 points]

(c) Calculate the test statistic and the P-value. [2 points]

(d) What conclusion do you draw at the 5% significance level? Answer in context. [2 points]

(e) Interpret the P-value in the context of this study. [2 points]

QUESTION 4 Drumstick Quality Control · 10 points

A factory produces drumsticks, and historically 8% of them have a defect. Sticks are packaged in packs of 12, and defects occur independently. The quality inspector also periodically takes random samples of 200 sticks from the production line.

(a) What is the probability that a randomly selected pack of 12 contains at least one defective stick? [2 points]

(b) Describe the sampling distribution of the sample proportion of defective sticks in random samples of size 200 when p = 0.08. Include shape (with a check), center, and spread. [3 points]

(c) Use your answer to (b) to find the probability that a sample proportion is 0.12 or greater when p = 0.08. [2 points]

(d) In today's sample of 200 sticks, 24 were defective (a sample proportion of 0.12). Does this give convincing evidence that the defect rate has increased above 8%? Use your result from (c) to justify your answer. [3 points]

Key Practice Exam 2: Answers, Explanations, and Scoring

Score yourself honestly. For free response, grade each part against the rubric, awarding points only for work that is shown and explained in context.

SECTION I · ANSWER KEY

1 C 2 B 3 C 4 B 5 A 6 A 7 C 8 D 9 D 10 D 11 B 12 A 13 A 14 A 15 D 16 A 17 A 18 B 19 D 20 B 21 C 22 C 23 D 24 B 25 C 26 D 27 B 28 A 29 D 30 C 31 B 32 B 33 A 34 C 35 D 36 C 37 C 38 A 39 D 40 B 41 B 42 A

TOPIC MAP: WHICH QUESTIONS TEST WHICH UNIT

Topic	Questions
Exam Unit 1: One-Variable Data and Data Collection	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
Exam Unit 2: Probability and Distributions	12, 13, 14, 15, 16, 17, 18, 19, 20
Exam Unit 3: Inference for Proportions	21, 22, 23, 24, 25, 26, 27
Exam Unit 4: Inference for Means	28, 29, 30, 31, 32, 33, 34
Exam Unit 5: Regression	35, 36, 37, 38, 39, 40, 41, 42

Sec I Multiple Choice Explanations

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. **Answer: C** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
The bars have a long tail toward the slow (large) times, so the distribution is skewed right. In a right-skewed distribution the few large values pull the mean upward, so the mean is greater than the median; for these counts the mean is about 13.0 seconds and the median about 12.5 seconds. The statement with mean less than median reverses the effect of skew. The symmetric claim ignores the clear tail to the right. Calling the shape skewed left confuses the direction of the tail with the location of the peak, which is near the left side of the graph.

2. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
Both boxplots have their median line at 502 grams, so the centers match. The interquartile range for Machine B is 509 - 497 = 12 grams, double the 6-gram IQR for Machine A, and B also has the wider range, so B is more variable. Saying B has the larger median misreads the graph. A short box means less spread, not more, so claiming A is more variable reverses the meaning. Neither plot shows any asterisks, so there are no outliers to compare.

3. **Answer: C** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
The z-score for 10 hours is (10 - 14.0)/2.5 = -1.6, and the area to the left of z = -1.6 in a standard normal curve is about 0.055. The value 0.945 is the area above 10 hours, so it answers the opposite question. The value 0.110 doubles the tail, which would be correct only for a two-sided question such as being more than 1.6 standard deviations from the mean in either direction. The value 0.016 is the area beyond z = -2.14 or so and results from dividing by a wrong standard deviation.

4. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
Multiplying every value by 2.54 multiplies both the mean and the standard deviation by 2.54, giving 75(2.54) = 190.5 and 3.2(2.54) = 8.128. Leaving the standard deviation at 3.2 treats it like a quantity that is unaffected by scaling, which is true only for adding a constant. Adding 2.54 to each value would shift the mean but not change the spread, so that choice mixes up the two transformations. Multiplying the standard deviation by 2.54 squared confuses it with the variance, which does scale by the square of the factor.

5. **Answer: A** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
The IQR is 46 - 34 = 12, so the upper fence is 46 + 1.5(12) = 64 and the lower fence is 34 - 18 = 16. A value is an outlier only if it lies beyond a fence, so 70 qualifies. A time of 64 sits exactly on the upper fence, not beyond it, so it is not flagged. A time of 58 is inside the fences. A time of 18 is above the lower fence of 16, so it is also not flagged; students who forget to subtract from Q1 may misplace the lower fence.

6. **Answer: A** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
Because users chose whether to respond, people with strong feelings are overrepresented, and the 84% may not reflect the typical user. A large sample does not repair this kind of bias, and 3,200 is actually not small. Nothing in the description suggests leading wording, so that choice invents a problem. Random assignment is a feature of experiments; this is a survey of opinions, so lack of random assignment is not the issue. The sampling method is the flaw.

7. **Answer: C** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
The population is divided into groups (shifts) and a separate random sample is taken from every group, which defines a stratified random sample. This design guarantees each shift is represented and reduces variability when shifts differ. A cluster sample would randomly choose whole shifts and then include everyone in them. A simple random sample gives every set of 60 workers the same chance, which is not true here because exactly 20 must come from each shift. A systematic sample selects every kth person from a list, which is not described.

8. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
Grouping by instrument is blocking: comparisons are made within similar students, so differences between pianists and violinists do not hide the effect of the practice method. Blocking does not provide generalization, because the subjects are volunteers. Blinding is a different design feature related to who knows the treatments. Random assignment, not blocking, balances out other lurking variables on average, and nothing can eliminate every possible confounder with certainty.

9. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
This is an observational study, so it can reveal an association but not cause and effect, since lurking variables such as overall fitness or training habits could affect both sleep and shooting. The two causal statements ignore this and treat the survey as if athletes had been randomly assigned to sleep amounts. Saying there is no relationship ignores the observed difference; the issue is the interpretation of the difference, not whether it exists. Only a randomized experiment could support a causal claim.

10. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
The value 15 is far from the other days, creating right skew, and the mean is not resistant to such extremes. The median, which is the sixth ordered value (4), is resistant and describes a typical day. Using every value is a property of the mean but is a disadvantage when an outlier is present. The distribution is not symmetric, so the second choice rests on a false claim. The median is determined by the ordered data, not by the midpoint of the axis.

11. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
The top 10% begins at the 90th percentile, where z is about 1.28, so the cutoff is 500 + 1.28(80), roughly 603. The value 632 uses z = 1.645, which corresponds to the top 5%. The value 398 uses z = -1.28 and marks the bottom 10%. The value 580 is only one standard deviation above the mean (z = 1), which puts about 16% of scores above it, not 10%.

12. **Answer: A** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
By the addition rule, P(A and B) = 0.50 + 0.30 - 0.65 = 0.15. Because 0.15 is greater than 0, the events can occur together, so they are not mutually exclusive. Because P(A)P(B) = 0.50(0.30) = 0.15 equals P(A and B), the events are independent. Mutually exclusive events with positive probabilities can never be independent, so the second and third statements fail. The last statement is wrong because the product check confirms independence.

13. **Answer: A** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
We restrict attention to the 50 defective phones, and 24 of them came from Plant Y, so the probability is 24/50 = 0.48. The value 0.08 is 24/300, the probability that a Plant Y phone is defective, which reverses the conditioning. The value 0.024 is 24/1000, the joint probability of being from Plant Y and defective. The value 0.30 is the share of all phones made at Plant Y, which ignores the information that the phone is defective.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

14. **Answer: A** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
This is a binomial setting with n = 8 and p = 0.7, so P(X = 6) = C(8,6)(0.7)^6(0.3)^2 = 28(0.1176)(0.09), about 0.296. The value 0.552 is P(X at least 6), which answers a different question. The value 0.118 is just 0.7 raised to the sixth power, omitting both the number of arrangements and the two misses. The value 0.745 is P(X at most 6), the cumulative probability up to and including 6.

15. **Answer: D** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
X is binomial with n = 40 and p = 0.35, so the mean is np = 14 and the standard deviation is the square root of np(1 - p) = sqrt(40(0.35)(0.65)) = sqrt(9.1), about 3.02. The value 9.1 is the variance and not the standard deviation. The value 3.74 is sqrt(np), which forgets the factor (1 - p). A mean of 26 would be the expected number NOT opened, so it uses the wrong probability.

16. **Answer: A** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The standard deviation of $\hat{p}$ is sqrt(p(1-p)/n) = sqrt(0.6(0.4)/150) = sqrt(0.0016) = 0.04. The value 0.0016 is the variance and has not been square-rooted. The value 0.0033 comes from placing n outside the square root, dividing sqrt(0.24) by 150 instead of sqrt(150). The value 0.490 is sqrt(p(1-p)), the standard deviation of a single yes/no response, and ignores division by the sample size.

17. **Answer: A** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The mean of the sampling distribution equals the population mean, 12. Its standard deviation is 9/sqrt(50), about 1.27. By the central limit theorem, a sample of 50 is large enough to make the distribution of the sample mean approximately normal even though the population is skewed. The other choices either keep the population shape despite the large sample, keep the population standard deviation, or both, and none of those is what happens when averaging 50 observations.

18. **Answer: B** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The probabilities in a discrete distribution must add to 1. The known values sum to 0.35 + 0.30 + 0.20 + 0.05 = 0.90, so k = 1 - 0.90 = 0.10. A value of 0.15 or 0.20 would make the total exceed 1 (1.05 or 1.10). The value 0.90 is the sum of the known probabilities, not the missing one; it mistakes the total of the other cells for the amount needed to complete the distribution.

19. **Answer: D** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The mean of a discrete random variable is the sum of each value times its probability: 0(0.35) + 1(0.30) + 2(0.20) + 3(0.10) + 4(0.05) = 0.30 + 0.40 + 0.30 + 0.20 = 1.2. The value 2.0 is the unweighted average of the five x-values, which ignores the probabilities. The value 1.36 is the variance of X, a different quantity. The value 0.30 is P(X = 1), a probability and not a mean.

20. **Answer: B** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The condition leaves X = 1, 2, 3, 4 with total probability 0.65. Within that group, P(X at least 2) = 0.35, so the conditional probability is 0.35/0.65, about 0.538. The value 0.350 is the unconditional probability of at least two ads, which ignores the given information. The value 0.308 is 0.20/0.65, counting only exactly two ads. The value 0.650 is the probability of at least one ad, the denominator, not the conditional probability.

21. **Answer: C** · EXAM UNIT 3: INFERENCE FOR PROPORTIONS
With $\hat{p}$ = 6/40 = 0.15, the sample has only 6 successes (and 34 failures), and the large-counts condition requires at least 10 of each. The sample is stated to be random, so that condition is met. Forty customers out of thousands is far below 10% of the population, so the independence condition is also met. A normally distributed population is not required for proportions; the shape of the sampling distribution is checked through the counts instead.

22. **Answer: C** · EXAM UNIT 3: INFERENCE FOR PROPORTIONS
The confidence level describes the long-run success rate of the interval-building method: about 95% of intervals made this way capture the population proportion. The first alternative applies a percentage to individual fans, but the interval is about one parameter. The sample proportion is already known (it is the midpoint, 0.48), so a probability statement about it is meaningless. Future sample proportions would vary around the true proportion, not around this interval, so the last choice also misreads the interval.

23. **Answer: D** · EXAM UNIT 3: INFERENCE FOR PROPORTIONS
The sample proportion is 132/300 = 0.44 and the standard error is sqrt(0.44(0.56)/300), about 0.0287. For 90% confidence, z* = 1.645, so the margin of error is about 0.047 and the interval is (0.393, 0.487). The interval (0.384, 0.496) uses z* = 1.96, which is 95% confidence. The interval (0.403, 0.477) uses z* = 1.28, which is 80% confidence. The interval (0.346, 0.534) has a margin about twice as large as it should be, likely from doubling the margin of error.

24. **Answer: B** · EXAM UNIT 3: INFERENCE FOR PROPORTIONS
A P-value is computed assuming the null hypothesis is true: it is the chance of a result at least as extreme as the one observed in the direction of Ha. The probability that the null is true is not what a P-value measures, so the second statement is wrong. Flipping the number to 0.968 and calling it the probability that the proportion exceeds 0.50 makes the same error, since the parameter is a fixed value and not random in this framework. The last choice confuses the P-value with a shooting percentage.

25. **Answer: C** · EXAM UNIT 3: INFERENCE FOR PROPORTIONS
The test statistic uses the hypothesized value in the standard error: z = (0.13 - 0.10)/sqrt(0.10(0.90)/400) = 0.03/0.015 = 2.00. The alternative is one-sided (greater), so the P-value is the upper-tail area, about 0.023. The value 0.046 doubles the tail, which belongs to a two-sided alternative. The value 0.038 results from using the sample proportion in the standard error, giving z of about 1.78. The value 0.977 is the area below z = 2, the wrong tail.

26. **Answer: D** · EXAM UNIT 3: INFERENCE FOR PROPORTIONS
The P-value of 0.0075 is less than 0.05, so we reject H0 and conclude that Version A has the higher completion rate. Because users were randomly assigned, the conclusion can be causal. Saying we fail to reject when the P-value is below the level reverses the decision rule. A P-value is not the probability that H0 is true; it assumes H0. Comparing the 12-point difference to the P-value of 0.0075 makes no sense because they are different quantities.

27. **Answer: B** · EXAM UNIT 3: INFERENCE FOR PROPORTIONS
With no prior estimate, use p = 0.5 to maximize the variance: n = (1.96/0.03)^2(0.5)(0.5), about 1067.1. Sample sizes must be rounded up to guarantee the margin of error, so 1068, not 1067. The value 385 is the sample size for a margin of error of 0.05. The value 2135 is about double the required size, which would be needed only for a much smaller margin of error, so it overshoots the requirement.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

28. **Answer: A** · EXAM UNIT 4: INFERENCE FOR MEANS
With n = 12, fewer than about 30, the t procedure relies on the sample showing no strong skewness or outliers, and the value near 31 seconds violates this. An unknown population standard deviation is the reason to use t in the first place, so that is not an obstacle. Small samples are acceptable when the data look roughly symmetric, so the claim about any t-interval is false. Random selection supports, not harms, independence; 12 requests are far less than 10% of all requests.

29. **Answer: D** · EXAM UNIT 4: INFERENCE FOR MEANS
The standard error is 6.5/sqrt(25) = 1.3, and with df = 24 the critical value is t* = 2.064, so the margin of error is 2.064(1.3), about 2.68, and the interval is (45.5, 50.9). The interval (45.7, 50.8) uses z* = 1.96, ignoring the extra uncertainty from estimating the standard deviation. The interval (34.8, 61.6) multiplies t* by s instead of s/sqrt(n). The interval (47.7, 48.7) divides s by n rather than by its square root, making it far too narrow.

30. **Answer: C** · EXAM UNIT 4: INFERENCE FOR MEANS
The test statistic is t = (496.2 - 500)/(5.6/sqrt(16)) = -3.8/1.4 = -2.71 with 15 degrees of freedom, and the two-sided P-value is about 0.016. Since 0.016 is less than 0.05, we reject H0. The one-sided value of 0.008 forgets to double the tail for the two-sided alternative. The statistic of -0.68 divides by s instead of the standard error s/sqrt(n). The last choice has the correct numbers but reverses the decision rule.

31. **Answer: B** · EXAM UNIT 4: INFERENCE FOR MEANS
Each runner supplies a measurement under both conditions, so the two sets of times are dependent. The right analysis finds each runner's difference and performs a one-sample t-test on those 12 differences. A two-sample test would ignore the pairing and treat the groups as independent, losing the benefit of controlling for runner-to-runner differences. Using only one shoe's times wastes the comparison. The response is quantitative (time), not a count of successes, so a proportion test does not apply.

32. **Answer: B** · EXAM UNIT 4: INFERENCE FOR MEANS
The P-value of 0.120 is greater than 0.05, so we fail to reject H0 and do not have convincing evidence of a difference; the 95% interval (-1.2, 10.2) also contains 0. Rejecting H0 would require a P-value below 0.05. Failing to reject never proves that the means are equal; it only means the data are consistent with equality. A P-value is computed assuming H0 is true and is not the probability that H0 is true.

33. **Answer: A** · EXAM UNIT 4: INFERENCE FOR MEANS
The margin of error is proportional to s/sqrt(n), so to halve it the square root of n must double, which requires n to quadruple. Doubling the sample size reduces the margin only by a factor of about 1/sqrt(2), roughly 0.71, which is a common mistake from assuming a linear relationship. Dividing the sample size by 2 would increase the margin of error. A 50% increase in n shrinks the margin by only about 18%.

34. **Answer: C** · EXAM UNIT 4: INFERENCE FOR MEANS
A two-sided test at the 5% level and a 95% confidence interval give matching decisions: values outside the interval are rejected. Since 16 lies outside (12.3, 15.9), the test rejects H0. Being close does not matter; any value not in the interval is not a plausible value of mu at this level. The interval always contains the sample mean (its midpoint, 14.1), so that fact says nothing about H0. The standard deviation is already built into the interval, so no further information is needed.

35. **Answer: D** · EXAM UNIT 5: REGRESSION
The correlation is calculated from standardized values, so linear changes of units, such as multiplying by a positive constant, do not change it. The converted data therefore still have r = 0.62. Multiplying r by the conversion factors applies a transformation that the correlation is designed to ignore. Smaller units would change the scale of the axes, not the strength of the linear association. No new data are needed, because r is unchanged by this kind of rescaling.

36. **Answer: C** · EXAM UNIT 5: REGRESSION
The predicted jump at x = 8 is 12.4 + 3.1(8) = 37.2 cm. The residual is observed minus predicted, 41 - 37.2 = 3.8 cm, positive because the line underpredicts. The value -3.8 reverses the subtraction. The value 37.2 is the prediction, not the residual. The value 41 is the observed value and has not been compared with the prediction at all.

37. **Answer: C** · EXAM UNIT 5: REGRESSION
The coefficient of determination is r squared = (-0.80)^2 = 0.64, so 64% of the variability in sleep hours is accounted for by the least-squares line using social media minutes. Using 80% mistakes r for r squared. The slope, not r squared, describes how the predicted response changes per unit of x, and r squared is not a rate. R squared also is not a count of points on the line; typically very few points lie exactly on a regression line.

38. **Answer: A** · EXAM UNIT 5: REGRESSION
The point (22, 12) has an x-value far from the rest, giving it high leverage, and it lies well below the trend, so it pulls the line downward on the right. This reduces the slope (from about 1.55 to nearly 0) and weakens the correlation toward 0, which makes the point influential. A larger slope and stronger correlation would require the point to lie above or on the trend. A y-value inside the y-range does not protect against influence when the x-value is extreme. Reducing the slope while strengthening r is inconsistent because the added point increases scatter.

39. **Answer: D** · EXAM UNIT 5: REGRESSION
The slope is the predicted change in the response for a one-unit increase in the explanatory variable, so 3.24 is the predicted increase in score per extra weekly hour. Describing the change as exact turns a prediction into a guarantee, but individual students vary around the line. Reversing the variables treats practice time as the response. The percent of variation explained is r squared (95.9%), not the slope.

40. **Answer: B** · EXAM UNIT 5: REGRESSION
The predicted score is 56.20 + 3.241(9), about 85.37. The residual is observed minus predicted: 84 - 85.37 = -1.37, indicating the model overpredicts for this student. The positive 1.37 reverses the subtraction. The value 85.37 is the predicted score and not the residual. The value 84 is the observed score itself, which has not been compared with the prediction.

41. **Answer: B** · EXAM UNIT 5: REGRESSION
The coefficient of determination is R-sq = 0.959, and r is the square root of R-sq with the same sign as the slope. The slope is positive, so r = +sqrt(0.959), about 0.979. The value 0.959 is r squared, not r. The value -0.979 has the wrong sign, as the scatterplot shows an increasing trend. The value 0.920 is r squared squared, the result of squaring when you should take a square root.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

42. Answer: A · EXAM UNIT 5: REGRESSION

A regression line is reliable only for x-values within the range of the data, and 8 weeks is inside the interval from 4 to 12, so this is interpolation. Predicting at 30, 20, or 0 weeks is extrapolation, beyond the observed range, where the linear pattern may not continue. For example, 30 weeks would give 3.5 seconds, which is unrealistically fast for most athletes, and the intercept of 5.9 seconds at 0 weeks describes a situation the data never examined. A strong fit inside the range gives no guarantee outside it.

Sec II Free Response Scoring Guide and Model Responses

QUESTION 1 · PRACTICE TIME AT A MUSIC SCHOOL

Part / skill	Points	What earns the points
(a) Proportional allocation	1	Sample sizes 12 piano, 10 strings, 8 brass/percussion (proportional to 720, 600, 480).
(a) Random selection within strata	2	Uses a random method (random numbers on each roster list) within each group separately, with no student chosen twice.
(b) Source of bias	1	Names undercoverage or convenience sampling: only students who are at school on Saturday can be chosen.
(b) Direction	1	Says estimates are likely too high because students at the lobby on Saturday are probably more committed and practice more.
(c) Five-number summary	2	Min 20, Q1 30, median 45, Q3 60, max 130.
(c) Outlier test	1	IQR = 30; fences -15 and 105; 130 is above 105, so it is an outlier.
(d) Choice	1	Chooses the median.
(d) Justification	1	Mean is pulled up by the outlier 130 (mean about 49.7 versus median 45; mean without 130 is about 43.9).

Model response

(a) The strata are piano (720), strings (600), and brass/percussion (480). Using proportional allocation, the sample of 30 should have 720/1800(30) = 12 piano students, 600/1800(30) = 10 string students, and 480/1800(30) = 8 brass or percussion students. I would list the students in each group, number them, and use a random number generator to select 12, 10, and 8 distinct students from the respective lists.

(b) Surveying only the students who are in the lobby on Saturday morning is a convenience sample and suffers from undercoverage, because students who do not attend on Saturday have no chance of being selected. Students who come in on Saturday are probably more dedicated, so they probably practice more than a typical student, and the bias will overestimate the mean practice time.

(c) The ordered data give a minimum of 20, a median of 45, a maximum of 130, Q1 = 30 (median of the lower seven values) and Q3 = 60 (median of the upper seven values). The IQR is 60 - 30 = 30, so the fences are 30 - 45 = -15 and 60 + 45 = 105. Since 130 > 105, it is an outlier; no value is below -15.

(d) The median is better. The outlier of 130 minutes pulls the mean up to about 49.7 minutes, while the median is 45; without the outlier the mean is about 43.9. The median is resistant to the outlier and better represents a typical student.

Common point-losing errors

Allocating equal numbers (10 each) to the strata without noting that this is not proportional; saying 'random' without describing how selection is done within each group; calling the lobby bias 'response bias' or 'sampling error' and not stating undercoverage and the direction; computing quartiles but forgetting to multiply the IQR by 1.5 or to compare 130 to the fence; choosing the mean 'because it uses all the data' while ignoring the outlier.

QUESTION 2 · PAGE LOAD TIMES FOR TWO SERVER CONFIGURATIONS

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Part / skill	Points	What earns the points
(a) Center	1	Compares medians in context (A: 248 ms, B: 262 ms; B is typically slower).
(a) Spread	1	Compares IQR or range (A: IQR 49; B: IQR 33), so A is more variable.
(a) Shape	1	Describes shape: A is slightly right-skewed; B is roughly symmetric.
(a) Unusual values	1	Notes the outlier at 410 ms for A and none for B.
(b) z-score and proportion	2	z = (300-262)/24 = 1.58; proportion about 0.057 (normal model).
(c) 90th percentile	1	z = 1.28 gives 262 + 1.28(24), about 292.8 ms.
(c) Interpretation	1	About 90% of requests in Configuration B load in under about 293 ms.
(d) Recommendation with evidence	2	Recommends B because of smaller spread (IQR 33 vs 49) and no outliers, acknowledging its slightly higher median.

Model response

(a) Center: the median load time for Configuration A is 248 ms, while B's median is 262 ms, so a typical request is about 14 ms slower with B. Spread: A is more variable, with an IQR of 275 - 226 = 49 ms compared with 278 - 245 = 33 ms for B. Shape: A is skewed to the right (the upper whisker and the outlier extend farther than the lower whisker), while B is roughly symmetric. Unusual values: A has an outlier at 410 ms, which is a very slow request, and B has no outliers.

(b) With the normal model N(262, 24), z = (300 - 262)/24 = 1.58. The proportion above 300 ms is P(Z > 1.58) = about 0.057, so approximately 5.7% of requests in Configuration B take longer than 300 ms.

(c) For the 90th percentile, z = 1.28, so the load time is 262 + 1.28(24) = about 292.8 ms. About 90% of page requests in Configuration B load in less than approximately 293 ms.

(d) Because the company values consistency, I recommend Configuration B. Its IQR (33 ms) is much smaller than A's (49 ms), its distribution is symmetric, and it has no extremely slow outliers like the 410 ms request for A. The trade-off is a slightly higher median.

Common point-losing errors

Comparing only centers and ignoring spread, shape, or outliers; giving a comparison without context (not naming load times or configurations) or using the word 'it' for both groups; using the upper-tail area incorrectly (reporting 0.943 instead of 0.057); interpreting the percentile as 'the 90% of the data are 293' rather than 'about 90% of requests are below 293 ms'; recommending a configuration without citing spread.

QUESTION 3 · A NEW JUMP-TRAINING DRILL

Part / skill	Points	What earns the points
(a) Parameter and hypotheses	2	Defines mu_d as the true mean difference (after - before) in vertical jump for league players; H0: mu_d = 0, Ha: mu_d > 0.
(b) Random and independent	1	Random sample of players; 10 is less than 10% of more than 200 players.
(b) Normality	1	Differences show no strong skewness or outliers in the dotplot, which is acceptable for n = 10.
(c) Test statistic	1	t = (3.3 - 0)/(1.95/sqrt(10)) = 5.36 with df = 9 (allow small rounding).
(c) P-value	1	P-value is about 0.0002 (one-sided).
(d) Decision linked to P-value	1	P-value is less than 0.05, so reject H0.
(d) Conclusion in context	1	There is convincing evidence that the mean vertical jump after the program exceeds the mean before for players in the league.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Part / skill	Points	What earns the points
(e) Interpretation of P-value	2	If the mean improvement were really 0, the probability of observing a sample mean difference of 3.3 cm or more by chance is about 0.0002.

Model response

(a) Let mu_d be the true mean difference (after - before) in vertical jump, in cm, for all players in the league. H0: mu_d = 0 and Ha: mu_d > 0.

(b) The procedure is a one-sample (paired) t-test for a mean difference. Random: the 10 players were randomly selected. Independence: 10 players is less than 10% of a league of more than 200 players. Normality: the sample size is small (10), so I check the dotplot of the differences; it has no strong skewness and no outliers, so the population of differences can reasonably be considered approximately normal.

(c) The test statistic is t = (3.3 - 0)/(1.95/sqrt(10)) = 3.3/0.617 = 5.36 with df = 10 - 1 = 9. The P-value is P(t > 5.36), which is about 0.0002.

(d) Because the P-value of about 0.0002 is less than alpha = 0.05, we reject H0. There is convincing evidence that the mean vertical jump after the 6-week program is greater than the mean before it for players in this league.

(e) If the true mean change in vertical jump were 0 cm, there would be about a 0.02% chance of obtaining a sample mean difference of 3.3 cm or greater just by random sampling variability.

Common point-losing errors

Treating the data as two independent samples instead of paired differences; writing hypotheses in words only without defining mu_d or using sample statistics (x-bar) in the hypotheses; not checking normality by looking at the differences (checking the original columns instead); stating 'we accept H0' or concluding without context; interpreting the P-value as the probability that H0 is true or as the probability that the drill works.

QUESTION 4 · DRUMSTICK QUALITY CONTROL

Part / skill	Points	What earns the points
(a) Model	1	Identifies a binomial setting (n = 12, p = 0.08) or uses the complement rule 1 - 0.92^12.
(a) Probability	1	1 - 0.92^12 = 0.632.
(b) Shape with check	1	Approximately normal because np = 16 and n(1-p) = 184 are both at least 10.
(b) Center	1	Mean equals 0.08.
(b) Spread	1	SD = sqrt(0.08(0.92)/200) = 0.0192.
(c) Standardize	1	z = (0.12 - 0.08)/0.0192 = 2.09.
(c) Probability	1	P(p-hat >= 0.12) is about 0.0185.
(d) Link to probability	1	A sample proportion this high would occur only about 1.9% of the time if p were still 0.08.
(d) Compare and decide	1	This is below a typical 5% threshold, so it is unlikely to be due to chance alone.
(d) Conclusion in context	1	There is convincing evidence that the defect rate for the sticks is now greater than 8%.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model response

(a) The number of defective sticks in a pack is binomial with n = 12 and p = 0.08, since each stick is defective with the same probability independently of the others. P(at least one defective) = 1 - P(none) = 1 - (0.92)^12 = 1 - 0.368 = 0.632.

(b) The sampling distribution of p-hat is approximately normal because np = 200(0.08) = 16 and n(1 - p) = 200(0.92) = 184 are both at least 10. Its mean is mu = p = 0.08 and its standard deviation is sqrt(0.08(0.92)/200) = 0.0192.

(c) The z-score is z = (0.12 - 0.08)/0.0192 = 2.09, so P(p-hat >= 0.12) = P(Z >= 2.09) = 0.0185, or about 1.9%.

(d) If the defect rate were still 8%, the probability of seeing a sample proportion as high as 0.12 or higher in a random sample of 200 sticks is only about 0.019. This is a small probability, below the commonly used 5% level, so the result is unlikely to occur by chance alone. Therefore we have convincing evidence that the true proportion of defective drumsticks is now greater than 8%.

Common point-losing errors

Computing P(exactly one defective) or 12(0.08) in (a) rather than at least one; checking the normality condition with p instead of np and n(1-p), or omitting it; using p-hat in place of p in the standard deviation of the sampling distribution; in (d), saying that the probability of 0.0185 is the probability that the defect rate is 8%, or concluding without referring to the context (drumsticks and the 8% rate).

Score

Scoring Your Practice Exam

Section I counts 50% and Section II counts 50%. Convert each section to a 0 to 50 scale, then add.

Step	Calculation	Your result
Section I	(number correct / 42) x 50	
Section II	(total points / 40) x 50	
Composite	Section I + Section II (0 to 100)	

About the score estimate

The College Board sets its cut scores for each exam year and does not publish them in advance. Scholar Coach uses this rough guide for practice only: composite 80 or above is in the range of a 5, 65 to 79 a 4, 50 to 64 a 3, 35 to 49 a 2, and below 35 a 1. Treat these as estimates, not predictions, and look at the topic map to decide what to study next.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Practice Exam 3

42 multiple-choice questions and 4 free-response questions in the digital exam format, followed by every answer, explanation, rubric, and a scoring guide.

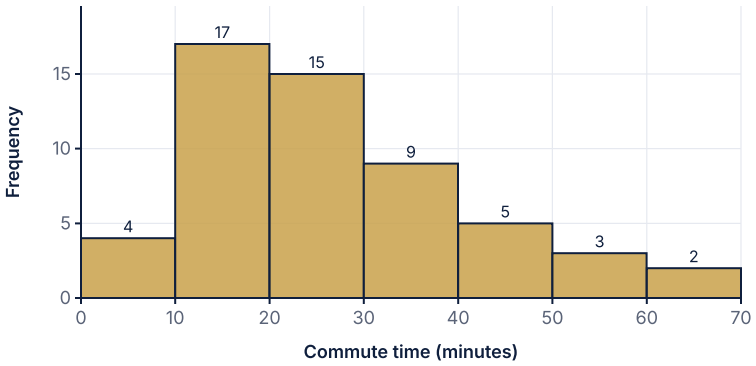
Exam 3

Section I: Multiple Choice

Section I · 42 questions · 90 minutes · 50% of score

Choose the best answer for each question. A graphing calculator is allowed, and the formula sheet and tables in the appendix are available. Questions in a set share one stimulus.

1

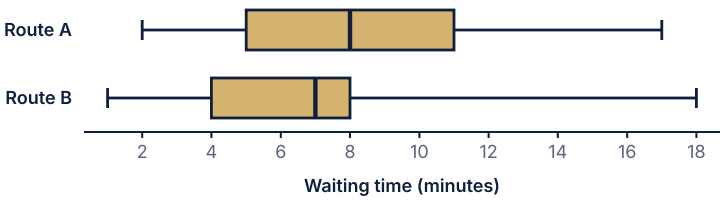


Commute times for 55 employees

The histogram shows the one-way commute times, in minutes, of 55 randomly chosen employees at a downtown firm. Which statement best describes the distribution?

(A) The distribution is skewed left, so the mean is likely less than the median.
(B) The distribution is skewed right, so the mean is likely greater than the median.
(C) The distribution is approximately symmetric, so the mean and median are about equal.
(D) The distribution is skewed right, so the mean is likely less than the median.

2



Waiting times on two routes

The boxplots compare the waiting times, in minutes, at stops on two bus routes. Which conclusion is supported by the plots?

(A) Route A has the larger median and the larger IQR.
(B) Route B has the larger median wait but less variability in its middle 50%.
(C) Route B is more variable because its range is larger, so its IQR must also be larger.
(D) The two routes have the same IQR, so their waiting times are equally consistent.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

3 A granola bar has 5.8 grams of fiber. Fiber content across bars has mean 4.2 g and standard deviation 0.8 g. A different bar has 11 grams of protein; protein content has mean 8 g and standard deviation 2.5 g. Which bar is relatively more unusual for its own nutrient?

(A) The protein bar, because 11 grams is a larger raw amount than 5.8 grams.
(B) The protein bar, because its standard deviation is larger.
(C) The fiber bar, because its z-score of 2.0 is larger than the protein bar's z-score of 1.2.
(D) Neither, because values measured in different nutrients cannot be compared.

4 The weights of bags of produce packed at a plant are approximately normal with mean 500 grams and standard deviation 12 grams. What proportion of bags weigh less than 480 grams?

(A) 0.952
(B) 0.023
(C) 0.096
(D) 0.048

5 A transit fare was originally recorded for each rider with mean 22 dollars and standard deviation 5 dollars. The agency raises every fare by 10 percent and then adds a flat 2 dollar service fee. What are the mean and standard deviation of the new fares?

(A) Mean 26.2 dollars, standard deviation 7.5 dollars
(B) Mean 24.2 dollars, standard deviation 5.5 dollars
(C) Mean 26.2 dollars, standard deviation 5 dollars
(D) Mean 26.2 dollars, standard deviation 5.5 dollars

6 A cereal company wants opinions about a new flavor. It divides its customer list into four regions, then takes a simple random sample of 150 customers from each region. This sampling method is best described as

(A) a cluster sample
(B) a stratified random sample
(C) a simple random sample of all customers
(D) a systematic sample

7 A bus agency posts a link on its website asking riders to rate service on a 1 to 5 scale, and 3 percent of visitors respond. Which is the most important concern about using these responses to estimate overall rider satisfaction?

(A) Voluntary response bias: riders with strong opinions are more likely to respond, so the results may not represent all riders.
(B) The sample is too small to calculate any statistics.
(C) Nonresponse is not a problem because the survey was available to every visitor.
(D) Random assignment was not used, so the results show no cause and effect.

8 Researchers compare two diets for lowering cholesterol in 80 volunteers. Because age strongly affects cholesterol, they first sort volunteers into three age groups and then randomly assign diets within each age group. What is the main purpose of grouping by age before assigning diets?

(A) To make sure that the two diets are assigned in a way that the subjects cannot know which diet they receive
(B) To reduce variability in the response that is due to age, making differences between diets easier to detect
(C) To allow the researchers to conclude that age causes changes in cholesterol
(D) To remove the need for random assignment

9 A data set of daily sugar intake (grams) has first quartile 24 and third quartile 36. Using the 1.5 times IQR rule, which of these values would be identified as an outlier?

(A) 7
(B) 55
(C) 48
(D) 54

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

10

```
  0 | 4 6 7 9
  1 | 1 2 2 4 5 7 8
  2 | 1 2 5
  3 | 1
Key: 1 | 4 means 14 grams
```

The stemplot shows the sugar content, in grams, of 15 yogurt cups. What is the median sugar content?

(A) 12 grams
(B) 15 grams
(C) 14.9 grams
(D) 14 grams

11

Age group	Car	Bus	Bike
Under 30	40	35	25
30 and over	120	50	30

Commuters in a survey were classified by age group and main way of getting to work. Among the commuters who mainly ride the bus, what percent are under 30?

(A) 35 percent
(B) About 12 percent
(C) 70 percent
(D) About 41 percent

12

In a household survey, 22 percent of households own an electric vehicle, 35 percent own a bicycle, and 8 percent own both. What is the probability that a randomly selected household owns an electric vehicle or a bicycle (or both)?

(A) 0.57
(B) 0.077
(C) 0.49
(D) 0.41

13

Plan	Satisfied	Not satisfied
Standard	168	72
Premium	135	25

A company surveyed customers about satisfaction with its standard and premium plans. If one satisfied customer is chosen at random, what is the probability that this customer is on the premium plan?

(A) About 0.446
(B) About 0.844
(C) About 0.338
(D) About 0.555

14

For a randomly selected shopper, let A be the event that the shopper buys groceries online each week and B the event that the shopper uses a loyalty app. Suppose P(A) = 0.30, P(B) = 0.40, and P(A or B) = 0.58. Which statement is correct?

(A) A and B are independent but not mutually exclusive.
(B) A and B are mutually exclusive and independent.
(C) A and B are mutually exclusive but not independent.
(D) A and B are neither independent nor mutually exclusive.

15

A rail station tracks how many of its three morning trains are delayed on a randomly chosen day. Let X be the number of delayed trains. The probability distribution is: P(X = 0) = 0.35, P(X = 1) = 0.30, P(X = 2) = k, and P(X = 3) = 0.10.

What is the value of k?

(A) 0.05
(B) 0.25
(C) 0.75
(D) 0.30

16

What is the expected number of delayed trains on a randomly chosen morning?

(A) 1.5
(B) 0.99
(C) 1.0
(D) 1.1

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

17 What is the standard deviation of X?

(A) About 0.99
(B) About 0.995
(C) About 1.1
(D) About 1.49

18 A phone survey calls 10 randomly chosen households. Each household independently answers with probability 0.3. What is the probability that exactly 3 households answer?

(A) About 0.650
(B) About 0.617
(C) About 0.267
(D) About 0.027

19 Suppose 40 percent of all commuters in a large region favor a proposed transit levy. A random sample of 150 commuters is taken. What is the approximate probability that the sample proportion favoring the levy exceeds 0.46?

(A) About 0.933
(B) About 0.134
(C) About 0.226
(D) About 0.067

20 The grocery bills of customers at a store have a right-skewed distribution with mean 85 dollars and standard deviation 40 dollars. For random samples of 64 bills, what is the best description of the sampling distribution of the sample mean?

(A) Right skewed with mean 85 dollars and standard deviation 40 dollars
(B) Approximately normal with mean 85 dollars and standard deviation 40 dollars
(C) Approximately normal with mean 85 dollars and standard deviation 0.625 dollars
(D) Approximately normal with mean 85 dollars and standard deviation 5 dollars

21 A random sample of 60 customers, chosen from a list of 5000, was asked whether they would buy a new product. Only 6 said yes. A student wants to use a one-sample z interval for the true proportion. Which condition is not met?

(A) The sample was not selected at random.
(B) The counts of successes and failures are not both at least 10.
(C) The sample is more than 10 percent of the population.
(D) The population is not normally distributed.

22 A 95 percent confidence interval for the proportion of residents who commute by transit is (0.41, 0.49). Which is the correct interpretation of the 95 percent confidence level?

(A) There is a 95 percent probability that the true proportion is between 0.41 and 0.49.
(B) 95 percent of residents commute by transit at a rate between 0.41 and 0.49.
(C) If many random samples were taken and an interval was built from each, about 95 percent of those intervals would contain the true proportion.
(D) 95 percent of all sample proportions will fall between 0.41 and 0.49.

23 In a random sample of 400 shoppers, 232 said they prefer store-brand pasta. Which is a 95 percent confidence interval for the proportion of all shoppers who prefer store-brand pasta?

(A) (0.539, 0.621)
(B) (0.516, 0.644)
(C) (0.532, 0.628)
(D) (0.578, 0.582)

24 A consumer group wants a 95 percent confidence interval for a proportion with a margin of error no larger than 0.03, and has no prior estimate. Using the conservative value p = 0.5, what minimum sample size is needed?

(A) 1067
(B) 1068
(C) 534
(D) 385

25 A company claims that at least 60 percent of its customers use its mobile app. A researcher tests H0: p = 0.60 against Ha: p < 0.60. Which describes a Type I error?

(A) Concluding that 60 percent of customers use the app when the true proportion is actually less than 60 percent
(B) Concluding that fewer than 60 percent of customers use the app when the true proportion is actually less than 60 percent
(C) Concluding that fewer than 60 percent of customers use the app when the true proportion is actually 60 percent
(D) Failing to find convincing evidence that the proportion is below 60 percent when it is in fact 60 percent

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

26 A random sample of 250 drivers includes 140 who use a transit app. To test H0: p = 0.5 against Ha: p not equal to 0.5, which gives the correct test statistic and P-value?

(A) z = 1.90, P-value = 0.058
(B) z = 1.90, P-value = 0.029
(C) z = 1.91, P-value = 0.056
(D) z = 0.06, P-value = 0.952

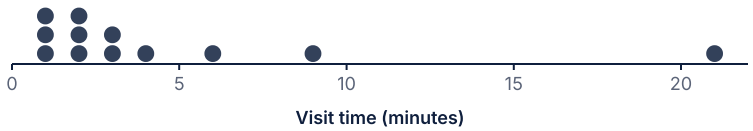
27

Region	n	Satisfied	Sample proportion
East	200	118	0.590
West	220	104	0.473
Difference (East - West)			0.117
z = 2.40			P-value = 0.016

A retailer compares customer satisfaction with its checkout app in two regions using independent random samples. The software output tests H0: p1 = p2 versus Ha: p1 not equal to p2. At the 0.05 significance level, which conclusion is appropriate?

(A) Fail to reject H0; there is convincing evidence that the proportions are equal.
(B) Reject H0; there is a 1.6 percent chance that the null hypothesis is true.
(C) Reject H0; there is convincing evidence that the true satisfaction proportions differ between the regions.
(D) Fail to reject H0, since the sample proportions of 0.59 and 0.47 are both above 0.05.

28



A biologist records the minutes that each of 12 randomly selected pollinators stays at a flower. A student wants to construct a 95 percent t interval for the mean visit time. Based on the dotplot, which is the most important concern?

(A) The sample size is not at least 30 times the size of the population.
(B) A t interval cannot be used because the population standard deviation is unknown.
(C) The sample is small and strongly skewed with an extreme value, so the normality condition is not satisfied.
(D) The data are discrete, and t procedures require continuous data.

29 A random sample of 25 trout from a lake has mean length 48.2 cm and standard deviation 6.5 cm. Assume the lengths are approximately normal. Which is a 95 percent confidence interval for the mean length of trout in the lake?

(A) (45.6, 50.8)
(B) (46.0, 50.4)
(C) (35.5, 60.9)
(D) (45.5, 50.9)

30 A random sample of 16 energy bars has mean 31.5 grams of protein and standard deviation 4.8 grams. To test H0: mu = 30 versus Ha: mu > 30, the test statistic is t = 1.25. Which is the correct P-value?

(A) About 0.115
(B) About 0.230
(C) About 0.885
(D) About 0.106

31

N	Mean	StDev	SE Mean	T	P-value
30	191.3	21.4	3.91	-2.23	0.017

A manufacturer claims that its snack packs average 200 calories. A nutrition class tests H0: mu = 200 versus Ha: mu < 200 using a random sample of 30 packs, and the software output is shown. At the 0.05 significance level, which conclusion is correct?

(A) Reject H0; there is convincing evidence that the mean calorie content is less than 200.
(B) Reject H0; there is a 1.7 percent probability that the null hypothesis is true.
(C) Fail to reject H0, because the sample mean of 191.3 is close to 200.
(D) Fail to reject H0, because the P-value is less than 0.05 but the sample is small.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

32 A commuter records travel time to work on the same 12 days using Route A and Route B, driving each route on the same day in random order. Which procedure is most appropriate to test whether the mean travel times differ?

(A) A two-sample t test comparing the Route A and Route B means
(B) A paired t test on the 12 daily differences
(C) A one-sample t test using only the Route A times
(D) A two-proportion z test

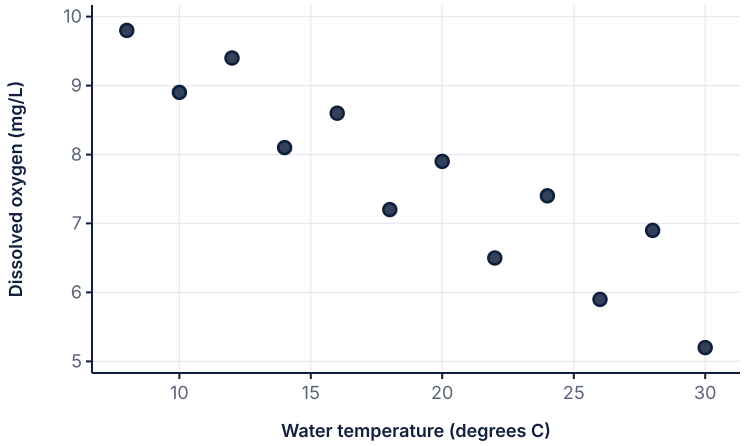
33 Independent random samples of two types of fertilizer plots give 95 percent confidence interval estimate (0.07, 5.33) for mu_A - mu_B, the difference in mean plant height (cm) for fertilizers A and B. Which conclusion is justified?

(A) Since the interval is wide, there is no evidence of any difference.
(B) Since the interval contains positive values, every plant with fertilizer A is taller than every plant with fertilizer B.
(C) Since 5.33 is the largest value, the true difference is exactly 5.33 cm.
(D) Since 0 is not in the interval, there is convincing evidence at the 5 percent level that the mean heights differ, with A larger.

34 Which change would increase the power of a one-sample t test of H0: mu = mu0, assuming everything else stays the same?

(A) Decreasing the significance level from 0.05 to 0.01
(B) Using a population with greater variability
(C) Choosing a true mean closer to mu0
(D) Increasing the sample size

35



Twelve river sites

Ecologists measured the water temperature and the dissolved oxygen concentration at 12 sites along a river. Which is the best estimate of the correlation between temperature and dissolved oxygen?

(A) -0.30
(B) 0.92
(C) -0.92
(D) -0.18

36 A regression of the number of bird species on the area of a forest patch gives r = 0.80. Which is the correct interpretation of r squared?

(A) About 64 percent of the variation in the number of bird species is accounted for by the linear relationship with forest area.
(B) About 64 percent of the forest patches have the predicted number of bird species.
(C) About 80 percent of the variation in the number of bird species is explained by forest area.
(D) The correlation between forest area and bird species is 0.64.

37 The least squares line relating weekly fuel cost (dollars) to miles driven is: predicted cost = 8.40 + 0.15(miles). Which is the correct interpretation of the slope?

(A) For each additional dollar of fuel cost, the predicted miles driven increases by 0.15.
(B) A driver who drives zero miles will have a predicted weekly fuel cost of 0.15 dollars.
(C) Each additional mile driven increases the weekly fuel cost by exactly 0.15 dollars.
(D) For each additional mile driven, the predicted weekly fuel cost increases by 0.15 dollars.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

38 For a set of snack items, the regression line is: predicted calories = 52 + 9.8(grams of fat). A snack with 14 grams of fat has 205 calories. What is the residual, and what does it mean?

(A) Residual = +15.8; the actual calories are 15.8 more than the line predicts.
(B) Residual = -15.8; the actual calories are 15.8 less than the line predicts.
(C) Residual = +15.8; the line overestimates the calories by 15.8.
(D) Residual = 189.2; this is the number of calories the line predicts.

39 Biologists studied 14 streams. For each, they recorded the width of the stream (in meters, ranging from 4 to 35 meters) and the number of fish species found. A least squares regression of number of fish species on stream width produced the computer output shown.

Predictor	Coef	SE Coef	T	P
Constant	4.436	1.489	2.98	0.012
Width	0.3756	0.0697	5.39	0.000
S = 2.55	R-sq = 70.8%			

Which is the correct interpretation of the slope of the regression line?

(A) For each additional meter of stream width, the predicted number of fish species increases by about 0.376.
(B) For each additional fish species, the predicted stream width increases by about 0.376 meters.
(C) For each additional meter of stream width, the number of fish species increases by exactly 0.376.
(D) A stream with 0 width is predicted to have 0.376 fish species.

40

Predictor	Coef	SE Coef	T	P
Constant	4.436	1.489	2.98	0.012
Width	0.3756	0.0697	5.39	0.000
S = 2.55	R-sq = 70.8%			

What is the correlation coefficient r between stream width and number of fish species?

(A) -0.84
(B) 0.71
(C) 0.84
(D) 0.51

41

Predictor	Coef	SE Coef	T	P
Constant	4.436	1.489	2.98	0.012
Width	0.3756	0.0697	5.39	0.000
S = 2.55	R-sq = 70.8%			

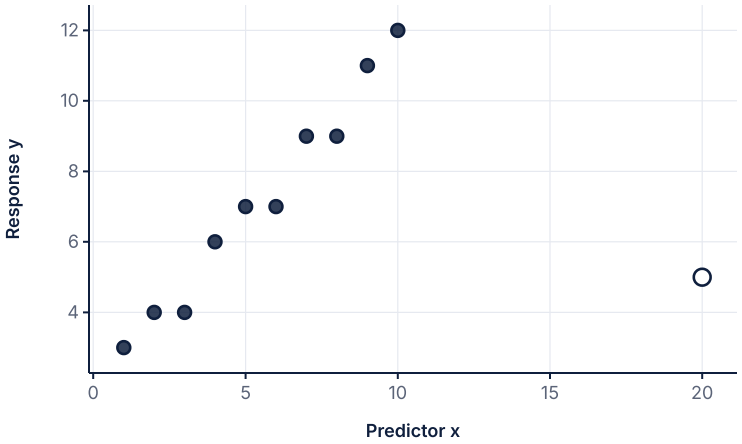
A researcher uses the line to predict the number of fish species in a river that is 60 meters wide. Which statement is best?

(A) The predicted value is about 27 species, but this prediction is unreliable because 60 meters is outside the range of the data.
(B) The predicted value is about 27 species, and it is reliable because the correlation is strong.
(C) The predicted value is about 22 species, because R-sq of 70.8% means the line is accurate for 70.8% of streams.
(D) No prediction can be made because the P-value for width is 0.000.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

42



Eleven observations; the highlighted point is (20, 5)

The scatterplot shows data for ten plots, plus one additional point (20, 5) shown as a highlighted point. What would happen to the correlation r and the slope of the least squares line if the highlighted point were removed?

(A) r would move closer to 0 and the slope would increase.

(B) r would move closer to 1 and the slope would increase.

(C) r would move closer to 1 and the slope would decrease.

(D) r and the slope would stay essentially the same, because the point is only one of eleven.

Exam 3 Section II: Free Response

Section II · 4 questions · 90 minutes · 50% of score

Each question is worth 10 points. Show your work and explain your reasoning in context, in complete sentences. Spend about 20 minutes on each of Questions 1 to 3 and the remaining time on Question 4 if you prefer, but plan your own pacing.

QUESTION 1 Rider Satisfaction with a New Express Bus · 10 points

A regional transit agency has launched a new express bus line and wants to know what proportion of the 8,000 people who rode it during its first month are satisfied with the service. The agency has an email list of all 8,000 riders who used the electronic fare card. As a pilot, it also recorded the waiting times, in minutes, of 15 randomly selected riders at one stop: 2, 3, 3, 4, 5, 5, 6, 7, 8, 8, 9, 11, 12, 14, 25.

(a) State the population and the parameter of interest for the agency's question. [2 points]

(b) Describe a method the agency could use to select a random sample of 400 riders from the email list. [3 points]

(c) Give one source of bias that could remain even if the sample is selected randomly, and explain its likely direction. [2 points]

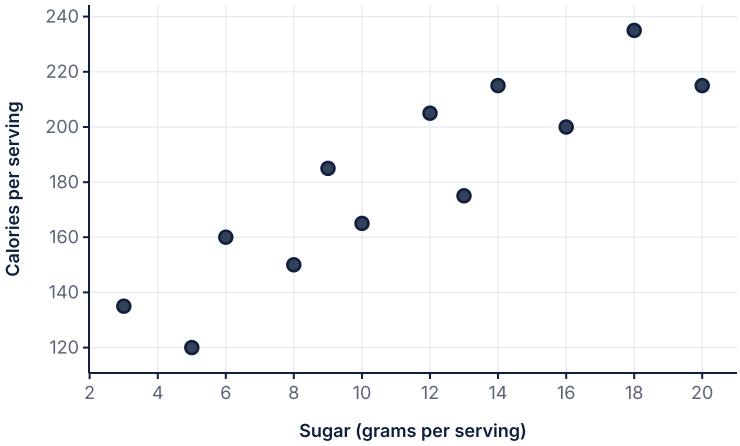
(d) For the 15 waiting times, compute the five-number summary and use the 1.5 times IQR rule to determine whether the longest waiting time is an outlier. [3 points]

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

QUESTION 2 Sugar and Calories in Snack Bars · 10 points

A nutrition class collected data on 12 snack bars. For each bar, the class recorded the grams of sugar per serving (x) and the calories per serving (y). A least squares regression produced the output below. The scatterplot shows the data. For the bar with 12 grams of sugar, the actual calories were 205.



Twelve snack bars

Predictor	Coef	SE Coef
Constant	113.40	11.60
Sugar	5.964	0.946
S = 16.6	R-sq = 79.9%	

(a) Write the equation of the least squares line and interpret the slope in context. [3 points]

(b) Calculate the predicted calories and the residual for the bar with 12 grams of sugar. Interpret the residual. [3 points]

(c) Interpret the value of R-sq in context. [2 points]

(d) A new bar with 2 grams of sugar and 300 calories is added to the data set. Describe, without calculating, how this single bar would affect the slope and the correlation, and explain why. [2 points]

QUESTION 3 Reusable Bags at a Grocery Chain · 10 points

A grocery chain claims that at least 60 percent of its customers bring reusable bags. A consumer group suspects the true proportion is lower. It selects a random sample of 250 customers at checkout across the chain's stores and finds that 138 of them have reusable bags.

(a) State the appropriate null and alternative hypotheses, defining any parameter. [2 points]

(b) Check the conditions for the procedure you will use. [2 points]

(c) Calculate the test statistic and the P-value. [2 points]

(d) At the significance level alpha = 0.05, state a conclusion in context. [2 points]

(e) Interpret the P-value in the context of this problem. [2 points]

QUESTION 4 Songbird Nests and Seedling Mulch · 10 points

Ecologists are studying a protected wetland. Part 1: they have 40 tree seedlings and want to compare the growth of seedlings that receive mulch type A with those that receive mulch type B, using 20 seedlings for each. Part 2: historically, 80 percent of songbird nests in the wetland survive to hatching. In a restored area, the team watched 15 nests that were independent of one another, and 9 survived.

(a) Describe a method to randomly assign the 40 seedlings to the two mulch types. [2 points]

(b) Suppose the seedlings in mulch A grew significantly more than those in mulch B. Explain why the ecologists can conclude that mulch A caused the greater growth. [2 points]

(c) Assume each nest survives independently with probability 0.80. Let X be the number of survivors among 15 nests. Name the distribution of X and calculate its mean and standard deviation. [3 points]

(d) Assuming p = 0.80, the probability of 9 or fewer survivors in 15 nests is about 0.061. Does the restored-area result provide convincing evidence at the 0.05 level that survival is lower there? Explain, and state the probability's meaning. [3 points]

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Key Practice Exam 3: Answers, Explanations, and Scoring

Score yourself honestly. For free response, grade each part against the rubric, awarding points only for work that is shown and explained in context.

SECTION I · ANSWER KEY

1 B 2 A 3 C 4 D 5 D 6 B 7 A 8 B 9 B 10 D 11 D 12 C 13 A 14 A 15 B 16 D 17 B 18 C 19 D 20 D 21 B 22 C 23 C 24 B 25 C 26 A 27 C 28 C 29 D 30 A 31 A 32 B 33 D 34 D 35 C 36 A 37 D 38 A 39 A 40 C 41 A 42 B

TOPIC MAP: WHICH QUESTIONS TEST WHICH UNIT

Topic	Questions
Exam Unit 1: One-Variable Data and Data Collection	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
Exam Unit 2: Probability and Distributions	12, 13, 14, 15, 16, 17, 18, 19, 20
Exam Unit 3: Proportions	21, 22, 23, 24, 25, 26, 27
Exam Unit 4: Means	28, 29, 30, 31, 32, 33, 34
Exam Unit 5: Regression	35, 36, 37, 38, 39, 40, 41, 42

Sec I Multiple Choice Explanations

1. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
   The bulk of commutes fall between 10 and 40 minutes, with a long tail stretching toward 70 minutes, so the shape is skewed right. A long right tail pulls the mean toward larger values while the median resists it, so the mean is likely greater than the median. The skewed-left choice reverses the direction of the tail. The symmetric choice ignores the clear tail. The last choice names the right shape but reverses the mean-median relationship, a common error.
2. **Answer: A** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
   Reading the plots, Route A has median 8 and quartiles 5 and 11 (IQR 6), while Route B has median 7 and quartiles 4 and 8 (IQR 4). So Route A has both the larger median and the larger IQR. The second choice reverses the medians. The third choice confuses range with IQR: Route B's range is larger only because of its long upper whisker, while its middle half is tighter. The fourth choice is false because the boxes have different lengths.
3. **Answer: C** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
   Standardizing puts different variables on a common scale. The fiber bar has $z = (5.8 - 4.2)/0.8 = 2.0$, while the protein bar has $z = (11 - 8)/2.5 = 1.2$. The fiber bar lies farther above its own mean in standard deviation units, so it is more unusual. Comparing raw amounts ignores the different means and spreads. A larger standard deviation actually makes a given distance less unusual. Z-scores exist precisely so that unlike variables can be compared.
4. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
   First standardize: $z = (480 - 500)/12 = -1.67$. The area to the left of $z = -1.67$ under the standard normal curve is about 0.048. The choice 0.952 is the area to the right, which answers the wrong tail. The choice 0.023 corresponds to $z = -2$, which would result from rounding 20/12 to 2 or from using a standard deviation of 10. The choice 0.096 doubles the tail area, which is appropriate only for a two-sided question about bags more than 20 grams from the mean in either direction.
5. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
   The new fare is $1.1x + 2$. The mean transforms with both steps: $1.1(22) + 2 = 26.2$. The standard deviation is changed only by the multiplication, because adding a constant shifts every value without changing spread: $1.1(5) = 5.5$. The choice with 7.5 wrongly adds the fee to the spread. The choice with mean 24.2 forgets to add the fee to the mean. The choice with standard deviation 5 ignores the 10 percent increase, which does stretch the spread.
6. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
   The customers are split into groups (regions) that are expected to differ, and a random sample is drawn from every group; that is a stratified random sample. In a cluster sample, whole groups would be randomly selected and every member of a chosen group surveyed, not a sample from each group. A simple random sample would pick 600 customers from the whole list with no grouping step, so region counts could vary by chance. A systematic sample takes every k-th name from an ordered list, which is not described here.
7. **Answer: A** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
   A voluntary response sample lets people choose whether to participate, and those with very positive or very negative views tend to respond. That makes it biased no matter how many people reply, and a large number of volunteers does not repair the bias. The claim about too small a sample misidentifies the problem, since bias, not sample size, is the flaw. Availability to everyone does not make participants representative. Random assignment applies to experiments and is irrelevant to a survey estimating satisfaction.
8. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
   Sorting subjects into age groups before randomizing is blocking. Blocks hold the influence of a known source of variation (age) apart from the comparison, so the diet effect is estimated more precisely. Hiding the diet from subjects is blinding, which is a different design feature. Blocking does not permit a conclusion about age as a cause because age is not randomly assigned. Random assignment is still carried out within each block, so blocking never replaces it.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

9. **Answer: B** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
The IQR is 36 - 24 = 12, so 1.5(IQR) = 18. The fences are 24 - 18 = 6 and 36 + 18 = 54. A value is an outlier only if it is below 6 or above 54. The value 55 exceeds 54, so it is flagged. The value 7 is above the lower fence of 6, so it is not an outlier. The value 48 is inside the upper fence. The value 54 sits exactly on the fence, which is not beyond it, so it is not flagged either.

10. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
With 15 values the median is the 8th value in order. Counting the leaves: the stem 0 row holds 4 values, and the stem 1 row holds the next 7 values (11, 12, 12, 14, 15, 17, 18), so the 8th value is the fourth leaf in the 1 row, which is 14. The choice 12 is a miscount by two positions. The choice 15 comes from counting past the middle. The value 14.9 is the mean (224/15), which the right tail pulls above the median; it is not the median.

11. **Answer: D** · EXAM UNIT 1: ONE-VARIABLE DATA AND DATA COLLECTION
The question asks for a conditional percent given the bus column. The bus column total is 35 + 50 = 85, and 35 of them are under 30, so 35/85 is about 41 percent. The choice 35 percent divides by the 100 people under 30 instead of the 85 bus riders, answering a different question. The choice 12 percent divides by the grand total of 300, which gives a joint percent. The choice 70 percent divides 35 by the 50 bus riders who are 30 or older, using the wrong group.

12. **Answer: C** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
Use the addition rule: P(A or B) = P(A) + P(B) - P(A and B) = 0.22 + 0.35 - 0.08 = 0.49. The value 0.57 adds the two probabilities without subtracting the overlap, double counting households that own both. The value 0.077 multiplies 0.22 by 0.35, which would be the probability of both only if the events were independent, and it is not the union anyway. The value 0.41 subtracts the overlap twice (0.57 - 0.08 - 0.08).

13. **Answer: A** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
There are 168 + 135 = 303 satisfied customers, and 135 of them are premium, so P(premium | satisfied) = 135/303, about 0.446. The value 0.844 is 135/160, which is P(satisfied | premium), the conditional probability in the reversed order. The value 0.338 is 135/400, the joint probability of being premium and satisfied. The value 0.555 is 168/303, the proportion of satisfied customers on the standard plan.

14. **Answer: A** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
From the addition rule, P(A and B) = 0.30 + 0.40 - 0.58 = 0.12. Because P(A and B) is not 0, the events can occur together, so they are not mutually exclusive. Because P(A)P(B) = 0.30(0.40) = 0.12 equals P(A and B), the events are independent. The second choice cannot be right because mutually exclusive events with positive probabilities are never independent, and these have a positive overlap. The third and fourth choices each wrongly deny independence or assert exclusivity.

15. **Answer: B** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The probabilities in a distribution must sum to 1, so 0.35 + 0.30 + k + 0.10 = 1, which gives k = 1 - 0.75 = 0.25. The value 0.05 would come from subtracting 0.95 instead of 0.75, perhaps leaving out one given probability. The value 0.75 is the sum of the other probabilities, not the missing one. The value 0.30 is a repeat of P(X = 1) with no justification from the sum rule.

16. **Answer: D** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The mean is the sum of each value times its probability: 0(0.35) + 1(0.30) + 2(0.25) + 3(0.10) = 0.30 + 0.50 + 0.30 = 1.1 trains. The value 1.5 is the simple average of the possible values 0, 1, 2, and 3, which ignores the unequal probabilities. The value 0.99 is the variance, not the mean. The value 1.0 reflects an arithmetic slip and does not match the weighted sum.

17. **Answer: B** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
First compute E(X squared) = 0(0.35) + 1(0.30) + 4(0.25) + 9(0.10) = 2.2. The variance is 2.2 - 1.1 squared = 2.2 - 1.21 = 0.99, so the standard deviation is the square root of 0.99, about 0.995. The value 0.99 is the variance and was not square rooted. The value 1.1 is the mean. The value 1.49 is the square root of 2.2, which skips subtracting the squared mean.

18. **Answer: C** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The count of answering households is binomial with n = 10 and p = 0.3, so P(X = 3) = C(10,3)(0.3)^3(0.7)^7, about 0.267. The value 0.650 is P(X at most 3), which includes 0, 1, and 2 answers. The value 0.617 is P(X at least 3), which includes 4 or more answers. The value 0.027 is (0.3)^3 alone, which leaves out both the number of arrangements C(10,3) and the probability 0.7^7 that the other seven do not answer.

19. **Answer: D** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The sampling distribution of the sample proportion has mean 0.40 and standard deviation sqrt(0.40(0.60)/150) = 0.04, and it is approximately normal because 150(0.4) = 60 and 150(0.6) = 90 are both at least 10. Then z = (0.46 - 0.40)/0.04 = 1.5, and the upper tail area is about 0.067. The value 0.933 is the lower tail. The value 0.134 doubles the area. The value 0.226 results from using a standard deviation of sqrt(0.4(0.6)) = 0.49 in the z calculation, forgetting to divide by sqrt of n.

20. **Answer: D** · EXAM UNIT 2: PROBABILITY AND DISTRIBUTIONS
The mean of the sampling distribution equals the population mean, 85, and its standard deviation is sigma/sqrt(n) = 40/8 = 5. Since the sample size 64 is large, the Central Limit Theorem says the distribution of the sample mean is approximately normal even though individual bills are skewed. The first choice describes the population, not the sampling distribution of the mean. The second choice gets the shape but uses the population standard deviation. The last choice divides by n instead of sqrt(n).

21. **Answer: B** · EXAM UNIT 3: PROPORTIONS
The large counts condition requires at least 10 expected successes and 10 failures; here there are only 6 successes (and 54 failures), so the sampling distribution of the proportion may not be approximately normal. The sample was explicitly random. Sixty is far less than 10 percent of 5000, so the 10 percent condition holds. Normality of the population is not a requirement for inference about a proportion, since the data are categorical.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

22. **Answer: C** · EXAM UNIT 3: PROPORTIONS
The confidence level describes the long-run success rate of the method: in repeated sampling, about 95 percent of intervals constructed this way capture the fixed, unknown population proportion. The first alternative treats the parameter as random after the interval is computed; the true proportion is fixed, so this particular interval either contains it or does not. The second alternative misreads the interval as a statement about individuals. The last confuses the interval with the sampling distribution of sample proportions, whose center is the true proportion and not this interval.

23. **Answer: C** · EXAM UNIT 3: PROPORTIONS
The sample proportion is 232/400 = 0.58, the standard error is sqrt(0.58(0.42)/400) = 0.0247, and the critical value is 1.96, so the margin is about 0.048 and the interval is (0.532, 0.628). The choice (0.539, 0.621) uses z* = 1.645, which is a 90 percent interval. The choice (0.516, 0.644) uses z* = 2.576, a 99 percent interval. The choice (0.578, 0.582) divides by n instead of sqrt(n) in the standard error, making the interval absurdly narrow.

24. **Answer: B** · EXAM UNIT 3: PROPORTIONS
Solve 1.96 sqrt(0.5(0.5)/n) at most 0.03, so n is at least (1.96/0.03)^2(0.25) = 1067.1. Sample sizes must be rounded up to guarantee the margin of error requirement, giving 1068. The value 1067 rounds down and would give a margin slightly above 0.03. The value 534 comes from using a margin of 0.06 or forgetting that halving the margin multiplies n by four. The value 385 is the sample size for a margin of error of 0.05.

25. **Answer: C** · EXAM UNIT 3: PROPORTIONS
A Type I error is rejecting a true null hypothesis. Here that means finding convincing evidence that p is below 0.60 when in fact p equals 0.60. The first alternative below the correct one describes a Type II error: failing to reject a false null. The next choice describes a correct decision because the null is false and rejected. The final choice also is a correct decision, since failing to reject a true null hypothesis is the right outcome.

26. **Answer: A** · EXAM UNIT 3: PROPORTIONS
Here the sample proportion is 140/250 = 0.56. For a test, the standard error uses the hypothesized value: sqrt(0.5(0.5)/250) = 0.0316, so z = 0.06/0.0316 = 1.90. Because the alternative is two-sided, the P-value doubles the upper tail: 2(0.0289) = 0.058. The choice with 0.029 forgets to double. The choice with z = 1.91 substitutes the sample proportion in the standard error, which belongs to interval estimation. The last choice uses the raw difference as z and is not a valid statistic.

27. **Answer: C** · EXAM UNIT 3: PROPORTIONS
The P-value of 0.016 is less than the significance level 0.05, so we reject H0 and conclude there is convincing evidence that the two population proportions differ. The first alternative is flawed because failing to reject never establishes that the proportions are equal, and here we reject anyway. The second misinterprets the P-value; it is a probability computed assuming H0 is true, not the probability that H0 is true. The last compares sample proportions to alpha, which are unrelated quantities.

28. **Answer: C** · EXAM UNIT 4: MEANS
With only 12 observations, the t procedures need the sample data to be roughly symmetric with no strong skew or outliers. The dotplot shows heavy right skew and a value far from the rest, so the interval would be unreliable. The first alternative states the 10 percent condition backwards. The second is reversed: t is used precisely because sigma is unknown. The third is not a requirement; times recorded to the minute are treated as measurements without difficulty.

29. **Answer: D** · EXAM UNIT 4: MEANS
With df = 24, the critical value is t* = 2.064, the standard error is 6.5/sqrt(25) = 1.3, and the margin of error is about 2.68, so the interval is (45.5, 50.9). The choice (45.6, 50.8) uses z* = 1.96, which is too small for a sample this size. The choice (46.0, 50.4) uses the wrong 90 percent critical value. The choice (35.5, 60.9) multiplies by the standard deviation rather than the standard error, so it describes a range of individual fish rather than the mean.

30. **Answer: A** · EXAM UNIT 4: MEANS
The test is upper tailed with df = 15, so the P-value is P(t > 1.25) with 15 degrees of freedom, about 0.115. The value 0.230 doubles the tail area, which is correct only for a two-sided alternative. The value 0.885 is the lower area to the left of 1.25. The value 0.106 uses the standard normal distribution instead of the t distribution; with a small sample, the t distribution has heavier tails and gives a larger P-value.

31. **Answer: A** · EXAM UNIT 4: MEANS
The P-value of 0.017 is below 0.05, so we reject H0 and have convincing evidence the true mean is below 200 calories. The second choice misinterprets the P-value as the probability that H0 is true. The third choice judges by the raw difference rather than comparing the P-value with alpha; whether the difference is meaningful depends on variability and sample size. The fourth contradicts itself, since a P-value below alpha leads to rejection, and a sample of 30 is adequate.

32. **Answer: B** · EXAM UNIT 4: MEANS
Each day produces a pair of times, one for each route, so the observations are not independent between routes. The right analysis is a paired t test on the differences (A - B). A two-sample t test assumes two independent samples and would ignore the pairing that removes day-to-day variability. A one-sample test of only Route A throws away half the data and cannot compare routes. A two-proportion test applies to categorical data, but travel times are quantitative.

33. **Answer: D** · EXAM UNIT 4: MEANS
A 95 percent interval that does not contain 0 corresponds to rejecting the hypothesis of no difference at the 5 percent significance level in a two-sided test. Every value in the interval is positive, so the mean for A appears larger. The first choice is wrong because the interval excludes 0; width alone does not determine whether a difference exists. The second wrongly applies a statement about means to individual plants. The third treats an endpoint as the parameter value, but the interval only gives plausible values.

34. **Answer: D** · EXAM UNIT 4: MEANS
Power is the probability of rejecting H0 when it is false. A larger sample reduces the standard error, so the test is better able to detect a given difference, increasing power. Lowering alpha makes rejection harder and reduces power while lowering the Type I error rate. Greater variability inflates the standard error, making true differences harder to detect. A true mean closer to mu0 is a smaller effect, which is harder to detect and reduces power.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

35. **Answer: C** · EXAM UNIT 5: REGRESSION
The pattern is a strong linear trend decreasing from left to right, so the correlation is negative and close to -1. The value -0.92 matches that strength. The value -0.30 would be a weak association with much more scatter than shown. The value 0.92 has the right strength but the wrong sign, because oxygen falls as temperature rises. The value -0.18 is the slope of the best-fit line, not the correlation, and is a mix-up of two different quantities.

36. **Answer: A** · EXAM UNIT 5: REGRESSION
The coefficient of determination is r squared = 0.80 squared = 0.64, and it gives the proportion of variability in the response that is accounted for by the linear model with the explanatory variable. The first alternative misreads it as a percent of data points. The second uses r rather than r squared. The last confuses r squared with the correlation itself. Remember that r squared is always a proportion of variation, not a count of cases.

37. **Answer: D** · EXAM UNIT 5: REGRESSION
The slope is the change in the predicted response per one-unit increase in the explanatory variable, so each extra mile is associated with 0.15 dollars more in predicted fuel cost. The first choice swaps the roles of x and y. The second choice describes the intercept's role (although the intercept would be 8.40 here, not 0.15). The last choice says exactly, but a regression line gives only predictions, so actual changes vary from driver to driver.

38. **Answer: A** · EXAM UNIT 5: REGRESSION
The predicted value is 52 + 9.8(14) = 189.2 calories. The residual is actual minus predicted: 205 - 189.2 = +15.8. A positive residual means the point lies above the line, so the actual calories exceed the prediction, and the line underestimates. The first choice reverses the subtraction. The second has the correct number but says overestimates, which is the wrong direction for a positive residual. The last choice reports the predicted value, not the residual.

39. **Answer: A** · EXAM UNIT 5: REGRESSION
The slope 0.3756 is the predicted change in the response (number of fish species) for each one meter increase in the explanatory variable (width). The first choice swaps explanatory and response variables. The second says exactly, claiming a deterministic relationship, whereas the line describes an average trend and individual streams vary. The third confuses the slope with the intercept, which is 4.436 here and would in any case not be meaningful because no stream has zero width.

40. **Answer: C** · EXAM UNIT 5: REGRESSION
The output gives R-sq = 70.8 percent, so $|r|$ = sqrt(0.708) = 0.84. The sign of r matches the sign of the slope; the slope is positive (0.3756), so r = +0.84. The choice -0.84 has the right size but the wrong sign. The choice 0.71 is R-sq as a decimal, not its square root. The choice 0.51 is roughly the square of 0.71 and results from squaring instead of taking a square root.

41. **Answer: A** · EXAM UNIT 5: REGRESSION
Substituting gives 4.436 + 0.3756(60), which is about 27.0 species. Because 60 meters is far beyond the observed widths of 4 to 35, this is extrapolation, and we cannot assume the linear pattern continues. A strong correlation does not justify extrapolating. The second choice misuses R-sq and gives an incorrect prediction. The last choice is wrong because a small P-value indicates a significant relationship; it never prohibits prediction, although extrapolating is risky.

42. **Answer: B** · EXAM UNIT 5: REGRESSION
The ten other points follow a tight, steep positive line (slope about 0.99, r about 0.99). The highlighted point is far to the right of the others in x and well below the trend, so it is an influential point that drags the line down, giving a slope of about 0.19 and r of about 0.34. Removing it restores a strong trend: r moves toward 1 and the slope rises. Any answer saying r moves toward 0 or the slope decreases contradicts that. A point extreme in x can have a large effect even in a small data set.

Sec II Free Response Scoring Guide and Model Responses

QUESTION 1 · RIDER SATISFACTION WITH A NEW EXPRESS BUS

Part / skill	Points	What earns the points
(a) Population and parameter	2	1 point: all riders of the express line during the first month (or the 8,000 fare-card riders). 1 point: parameter is the proportion p of those riders who are satisfied.
(b) Random selection method	3	1 point: number the 8,000 riders (or use the list order). 1 point: use a random number generator or equivalent to choose 400 distinct numbers with no repeats. 1 point: select those riders so every rider/group of 400 has an equal chance.
(c) Bias	2	1 point: names a plausible source (nonresponse, riders without a fare card or email excluded, wording). 1 point: explains the likely direction in context (for example, dissatisfied riders more likely to respond, overestimating or underestimating satisfaction).
(d) Five-number summary and outlier rule	3	1 point: correct five-number summary (2, 4.5, 7, 10, 25). 1 point: IQR = 5.5 and upper fence = 10 + 1.5(5.5) = 18.25. 1 point: concludes that 25 exceeds the fence so it is an outlier.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Model response

(a) The population is all riders who used the new express line during its first month (the 8,000 riders on the fare-card list). The parameter is p, the proportion of these riders who are satisfied with the service.

(b) Assign each rider on the email list a number from 1 to 8000. Use a random number generator to select 400 different integers between 1 and 8000, ignoring any repeats, and survey the riders with those numbers. This is a simple random sample, so each group of 400 riders is equally likely.

(c) Nonresponse bias could remain: riders who are unhappy are often more motivated to answer a satisfaction survey than satisfied riders. That would make the sample proportion of satisfied riders lower than the true proportion, an underestimate. Excluding riders who pay cash and have no fare card would be another possibility.

(d) The ordered data have minimum 2, Q1 = 4.5, median = 7, Q3 = 10, and maximum 25, so the five-number summary is 2, 4.5, 7, 10, 25. The IQR is 10 - 4.5 = 5.5. The upper fence is 10 + 1.5(5.5) = 18.25 and the lower fence is 4.5 - 8.25 = -3.75. Since 25 is greater than 18.25, a waiting time of 25 minutes is an outlier.

Common point-losing errors

Describing a convenience sample (such as emailing everyone or surveying riders at the stop) as random; forgetting to say that repeated numbers are ignored in a random selection method; naming bias without stating a direction in context; computing the fences using the median or the range instead of the quartiles and IQR.

QUESTION 2 · SUGAR AND CALORIES IN SNACK BARS

Part / skill	Points	What earns the points
(a) Equation and slope	3	1 point: predicted calories = 113.40 + 5.964(sugar). 1 point: slope interpreted as change in predicted calories per additional gram of sugar. 1 point: in context with predicted/on average language, about 5.96 calories.
(b) Prediction and residual	3	1 point: predicted = 113.40 + 5.964(12) = about 185.0. 1 point: residual = 205 - 185.0 = +20.0. 1 point: interpretation that the actual calories are about 20 higher than the line predicts (line underestimates).
(c) R-sq	2	1 point: about 79.9 percent of the variation in calories. 1 point: accounted for by the linear relationship with grams of sugar.
(d) New bar	2	1 point: slope decreases and correlation weakens (moves toward 0). 1 point: explains that the point has an extreme low x value and is far above the pattern, so it is influential and pulls the line up at the left end.

Model response

(a) The least squares line is predicted calories = 113.40 + 5.964(sugar). The slope means that for each additional gram of sugar per serving, the predicted number of calories increases by about 5.96 calories. This is an average trend, not an exact change for each bar.

(b) For 12 grams of sugar, the predicted calories are 113.40 + 5.964(12) = 184.97, or about 185.0 calories. The residual is the actual minus the predicted value: 205 - 185.0 = +20.0 calories. Because the residual is positive, this bar has about 20 more calories than the line predicts for a bar with 12 grams of sugar; it lies above the line and the line underestimates it.

(c) R-sq = 79.9 percent means that about 79.9 percent of the variation in calories per serving among these snack bars is accounted for by the linear relationship with grams of sugar. The remaining variation is due to other factors.

(d) The new bar has a very low x value (2 grams), at the left edge of the data, and its calories are far above the line. A point like this is influential: it would pull the left end of the line upward and make the line flatter, so the slope would decrease. The correlation would also weaken, moving toward 0, because the point does not follow the strong increasing pattern.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Common point-losing errors

Interpreting the slope without 'predicted' or without the units (calories per gram); computing the residual as predicted minus actual and getting the sign wrong; describing R-sq as the percent of bars that fall on the line; saying the new point has no effect because it is only one of 13 values.

QUESTION 3 · REUSABLE BAGS AT A GROCERY CHAIN

Part / skill	Points	What earns the points
(a) Hypotheses	2	1 point: parameter p = true proportion of all the chain's customers who bring reusable bags. 1 point: H0: p = 0.60 and Ha: p < 0.60.
(b) Conditions	2	1 point: random sample and the sample is less than 10 percent of all customers. 1 point: large counts: 250(0.6) = 150 and 250(0.4) = 100 are both at least 10.
(c) Test statistic and P-value	2	1 point: z = (0.552 - 0.60)/sqrt(0.6(0.4)/250) = -1.55. 1 point: P-value = P(Z < -1.55), about 0.061.
(d) Conclusion	2	1 point: compares the P-value 0.061 to alpha = 0.05 and fails to reject H0. 1 point: states in context that there is not convincing evidence that fewer than 60 percent of customers bring reusable bags.
(e) Interpretation of the P-value	2	1 point: describes a probability computed assuming the true proportion is 0.60. 1 point: probability of getting a sample proportion as low as 0.552 or lower (about 6.1 percent) by random sampling variation.

Model response

(a) Let p be the true proportion of all the chain's customers who bring reusable bags. H0: p = 0.60 and Ha: p < 0.60.

(b) I will use a one-sample z test for a proportion. Random: the sample of 250 customers was randomly selected. Independence: the sample is less than 10 percent of all of the chain's customers, which is surely true for a large chain. Large counts: 250(0.60) = 150 and 250(0.40) = 100 are both at least 10, so the sampling distribution of the sample proportion is approximately normal.

(c) The sample proportion is 138/250 = 0.552. The test statistic is z = (0.552 - 0.60)/sqrt(0.60(0.40)/250) = -0.048/0.0310 = -1.55. The P-value is P(Z < -1.55), which is about 0.061.

(d) Because the P-value of 0.061 is greater than alpha = 0.05, we fail to reject H0. There is not convincing evidence that the true proportion of customers who bring reusable bags is less than 0.60.

(e) If the true proportion of customers with reusable bags were 0.60, there would be about a 6.1 percent chance of getting a random sample of 250 customers in which the sample proportion is 0.552 or less, just by random sampling variability.

Common point-losing errors

Using the sample proportion in the standard error for the test; writing hypotheses in terms of the sample proportion; stating the conclusion as 'accept H0' or 'the claim is true'; interpreting the P-value as the probability that H0 is true; omitting the check that the sample is random or less than 10 percent of the population.

QUESTION 4 · SONGBIRD NESTS AND SEEDLING MULCH

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Part / skill	Points	What earns the points
(a) Random assignment	2	1 point: a valid chance mechanism (for example, labeling seedlings 1 to 40 and using random numbers or drawing 20 labels). 1 point: each seedling is assigned to exactly one treatment with 20 per group, no repeats.
(b) Causation	2	1 point: random assignment makes the two groups similar in all other characteristics (balances lurking variables). 1 point: so a difference larger than chance would produce can be attributed to the mulch type.
(c) Binomial model	3	1 point: X is binomial with n = 15 and p = 0.80. 1 point: mean = 15(0.8) = 12. 1 point: SD = sqrt(15(0.8)(0.2)) = 1.55.
(d) Decision	3	1 point: states that P(X <= 9) = 0.061 is computed assuming p = 0.80. 1 point: compares 0.061 to 0.05. 1 point: concludes in context that there is not convincing evidence that survival is lower in the restored area.

Model response

(a) Number the seedlings from 1 to 40. Use a random number generator to pick 20 different numbers between 1 and 40; those seedlings receive mulch A and the remaining 20 receive mulch B. Every seedling is assigned to exactly one treatment, and each group of 20 is equally likely.

(b) Random assignment tends to make the two groups of seedlings alike in every other way, such as sunlight, soil, and initial size, so these lurking variables are balanced between the groups. If the difference in growth is larger than chance variation alone would likely produce (statistically significant), the only systematic difference between the groups is the mulch type. So the ecologists can conclude that mulch A caused the greater growth for seedlings like these.

(c) Each nest has two outcomes, survival or not, the nests are independent, the number of nests is fixed at 15, and the probability of survival is 0.80 for each. So X is binomial with n = 15 and p = 0.80. The mean is np = 15(0.80) = 12 nests. The standard deviation is sqrt(15(0.80)(0.20)) = sqrt(2.4), about 1.55 nests.

(d) If the survival probability in the restored area were still 0.80, the probability of seeing 9 or fewer survivors among 15 nests is about 0.061, so such a result would occur about 6.1 percent of the time by chance. Since 0.061 is greater than 0.05, the result is not statistically significant. We do not have convincing evidence that the nest survival rate in the restored area is lower than 80 percent, although 9 of 15 is somewhat low.

Common point-losing errors

Describing random assignment as choosing 'every other seedling' or by letting the ecologists decide; stating causation without linking it to random assignment or statistical significance; using 15(0.8)(0.2) as the standard deviation (forgetting the square root); reporting P(X = 9) or P(X >= 9) instead of P(X <= 9); concluding the survival rate is the same when failing to reject.

Score Scoring Your Practice Exam

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Section I counts 50% and Section II counts 50%. Convert each section to a 0 to 50 scale, then add.

Step	Calculation	Your result
Section I	(number correct / 42) x 50	
Section II	(total points / 40) x 50	
Composite	Section I + Section II (0 to 100)	

About the score estimate

The College Board sets its cut scores for each exam year and does not publish them in advance. Scholar Coach uses this rough guide for practice only: composite 80 or above is in the range of a 5, 65 to 79 a 4, 50 to 64 a 3, 35 to 49 a 2, and below 35 a 1. Treat these as estimates, not predictions, and look at the topic map to decide what to study next.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Reference: Formulas, Conditions, Glossary, and Tables

Formulas, the conditions checklist for every procedure, a glossary, and the z, t, and chi-square tables.

R.1 Formula Reference

Every formula in the course, grouped by topic, with a note on when to use it. Extension items are flagged for topics College Board lists as removed from the 2027 exam core.

How to Use This Reference

Each table lists the name of a formula, the formula itself, and the situation that calls for it. Formulas marked **(extension)** belong to topics that College Board's revision page lists as removed from the 2027 exam core; they remain useful in college courses and on older exams, so check your teacher's current course description. The exam provides formulas and tables, but knowing which formula fits which situation is your job.

One-Variable Statistics

Name	Formula	When to use
Sample mean	$\bar{x} = \frac{\sum x_i}{n}$	Center of a quantitative sample; sensitive to outliers and skew.
Sample standard deviation	$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$	Typical distance of values from the mean in a sample; use with the mean.
Range	$\max - \min$	Crude measure of spread; uses only two values.
Interquartile range	$IQR = Q_3 - Q_1$	Spread of the middle half of the data; resistant to outliers.
Five-number summary	$\min,\ Q_1,\ \text{median},\ Q_3,\ \max$	Building a boxplot or describing a skewed distribution.
z-score	$z = \frac{x - \bar{x}}{s_x}$ or $z = \frac{x - \mu}{\sigma}$	Number of standard deviations a value lies from the mean; compare values from different distributions.
1.5 IQR outlier fences	$Q_1 - 1.5\,IQR$ and $Q_3 + 1.5\,IQR$	Flag a value as an outlier if it lies below the lower fence or above the upper fence.
Percentile rank	$\frac{\text{number of values below } x}{n} \times 100$	Describing the relative standing of one value in a data set.
Relative frequency	$\frac{\text{count in category}}{\text{total count}}$	Summarizing categorical data or comparing groups of different sizes.
Add a constant a to every value	$\bar{x}_{new} = \bar{x} + a,\ \ s_{new} = s,\ \ IQR_{new} = IQR$	Shifting a scale (for example, adding a bonus to every score): center moves, spread does not.
Multiply every value by b	$\bar{x}_{new} = b\bar{x},\ \ s_{new} = \lvert b \rvert\, s,\ \ IQR_{new} = \lvert b \rvert\, IQR$	Changing units (inches to centimeters): center and spread both scale.
Combined transformation $y = a + bx$	$\bar{y} = a + b\bar{x},\ \ s_y = \lvert b \rvert\, s_x$	Any linear change of units, such as Celsius to Fahrenheit.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Shape, Center, Spread

Skewed or outliers present: report the median and IQR. Roughly symmetric with no outliers: report the mean and standard deviation. A linear transformation never changes the shape of a distribution or the z-score of an individual value.

The Normal Model

Name	Formula	When to use
Normal density notation	$X \sim N(\mu, \sigma)$	Naming a normal model by its mean and standard deviation.
Standardizing	$z = \dfrac{x - \mu}{\sigma}$	Convert to the standard normal model with mean 0 and SD 1 before using a table.
Area from a z-score	$P(Z < z) = \text{normalcdf}(-\infty, z)$	Find a proportion below a value using the built-in calculator.
Value from an area	$x = \mu + z^* \sigma$	Find the cutoff for a given percentile: get z^* from the area, then unstandardize.
Empirical rule	$68\%,\ 95\%,\ 99.7\%$ within $1,\ 2,\ 3\ \sigma$	Quick estimates for approximately normal data.

Regression and Correlation

Name	Formula	When to use
Correlation coefficient	$r = \dfrac{1}{n-1} \sum \left(\dfrac{x_i - \bar{x}}{s_x} \right) \left(\dfrac{y_i - \bar{y}}{s_y} \right)$	Strength and direction of a linear relationship; $-1 \le r \le 1$.
Least-squares slope	$b = r \dfrac{s_y}{s_x}$	Predicted change in y for each one-unit increase in x.
Least-squares intercept	$a = \bar{y} - b\bar{x}$	Predicted y when $x = 0$; the line always passes through $(\bar{x}, \bar{y})$.
Regression line	$\hat{y} = a + bx$	Predicting a response from an explanatory value inside the observed range.
Residual	$e = y - \hat{y}$	How far an observed value is above (positive) or below (negative) the line.
Coefficient of determination	r^2	Proportion of variation in y accounted for by the linear relationship with x.
Residual standard deviation	$s = \sqrt{\dfrac{\sum (y_i - \hat{y}_i)^2}{n-2}}$	Typical size of a prediction error, in units of y.
Outlier or influential point check	Compare the line with and without the point	A point far from the others in x can change the slope and r a great deal.
Residual plot (extension)	e_i plotted against x_i or $\hat{y}_i$	Analyzing departures from linearity: a curved pattern means a line is a poor model.
Transformed model (extension)	$\ln \hat{y} = a + bx$ or $\hat{y} = a + b \ln x$	Straightening a curved relationship before fitting a line.

Probability

Name	Formula	When to use
Complement rule	$P(A^C) = 1 - P(A)$	Probability that an event does not happen; handy for 'at least one'.
General addition rule	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	Probability of A or B (or both).
Addition rule, mutually exclusive	$P(A \cup B) = P(A) + P(B)$	When A and B cannot both happen.
General multiplication rule	$P(A \cap B) = P(A)\,P(B \mid A)$	Probability that both events happen, in sequence.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Name	Formula	When to use
Conditional probability	$P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$	Probability of B given that A is known to have happened.
Independence test	$P(B \mid A) = P(B) \Leftrightarrow P(A \cap B) = P(A)P(B)$	Decide whether knowing A changes the probability of B.
Counting arrangements	${}_nP_r = \dfrac{n!}{(n-r)!}$	Ordered selections without repetition.
Counting combinations	$\binom{n}{r} = \dfrac{n!}{r!\,(n-r)!}$	Unordered selections; appears in the binomial formula.

Discrete Random Variables

Name	Formula	When to use
Expected value	$\mu_X = E(X) = \sum x_i\, p_i$	Long-run average value of a discrete random variable.
Variance and SD	$\sigma_X^2 = \sum (x_i - \mu_X)^2\, p_i, \quad \sigma_X = \sqrt{\sigma_X^2}$	Typical distance of X from its mean over many trials.
Linear change of a variable	$\mu_{a+bX} = a + b\mu_X, \;\; \sigma_{a+bX} = \lvert b \rvert \sigma_X$	Fees, bonuses, or unit changes applied to a random variable.
Sum or difference: mean (extension)	$\mu_{X \pm Y} = \mu_X \pm \mu_Y$	Combining random variables: means always add or subtract.
Sum or difference: SD (extension)	$\sigma_{X \pm Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$	Only when X and Y are independent; variances add, SDs do not.
Binomial probability	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$	Count of successes in n independent trials, fixed p each time.
Binomial mean and SD	$\mu_X = np, \quad \sigma_X = \sqrt{np(1-p)}$	Center and spread of a binomial count.
Geometric probability (extension)	$P(X = k) = (1-p)^{k-1} p$	Probability the first success occurs on trial k.
Geometric mean and SD (extension)	$\mu_X = \dfrac{1}{p}, \quad \sigma_X = \dfrac{\sqrt{1-p}}{p}$	Typical wait for a first success.
Geometric tail (extension)	$P(X > k) = (1-p)^k$	Probability of more than k failures before the first success.

Binomial or Geometric?

Binomial: fixed number of trials n, count the successes. Geometric: no fixed n, count the trials until the first success.

Both need independent trials with the same p each time. Without replacement from a small population, check the 10 percent condition.

Sampling Distributions

Name	Formula	When to use
Mean of $\hat{p}$	$\mu_{\hat{p}} = p$	The sample proportion is an unbiased estimator of p.
SD of $\hat{p}$	$\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$	Typical sampling variability; needs $n \le 10\%$ of the population.
Normal condition for $\hat{p}$	$np \ge 10$ and $n(1-p) \ge 10$	Check before using a normal model for $\hat{p}$.
Mean of $\bar{x}$	$\mu_{\bar{x}} = \mu$	The sample mean is an unbiased estimator of μ.
SD of $\bar{x}$	$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$	Larger samples give less variable means; needs $n \le 10\%$ of the population.
Central limit theorem	$n \ge 30$ gives an approximately normal $\bar{x}$ for any shape	Use when the population is not normal; strong skew or outliers need larger n.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Name	Formula	When to use
Difference of proportions	$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2,\ \sigma = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	Comparing two independent groups' sample proportions.
Difference of means	$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2,\ \sigma = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	Comparing two independent groups' sample means.

Confidence Intervals and Test Statistics for Proportions

Name	Formula	When to use
One-proportion z interval	$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	Estimate a population proportion p.
One-proportion z statistic	$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	Test $H_0 : p = p_0$; use p_0 in the denominator.
Two-proportion z interval	$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	Estimate $p_1 - p_2$.
Pooled proportion	$\hat{p}_c = \frac{x_1 + x_2}{n_1 + n_2}$	Test $H_0 : p_1 = p_2$; combine the two groups.
Two-proportion z statistic	$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	Test for a difference between two population proportions.
Standard error (general)	$SE = \text{estimated SD of a statistic}$	Replace unknown parameters with sample values.
General interval form	$\text{statistic} \pm (\text{critical value})(SE)$	Every confidence interval in the course has this shape.
General test statistic	$\text{test statistic} = \frac{\text{statistic} - \text{null value}}{SE}$	Every one-sample or two-sample test in the course has this shape.

Confidence Intervals and Test Statistics for Means

Name	Formula	When to use
One-sample t interval	$\bar{x} \pm t^* \frac{s}{\sqrt{n}},\ df = n - 1$	Estimate a population mean μ when σ is unknown.
One-sample t statistic	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},\ df = n - 1$	Test $H_0 : \mu = \mu_0$.
Two-sample t interval	$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Estimate $\mu_1 - \mu_2$ from two independent samples or groups.
Two-sample t statistic	$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	Test $H_0 : \mu_1 = \mu_2$; degrees of freedom from the built-in calculator.
Paired t interval	$\bar{x}_d \pm t^* \frac{s_d}{\sqrt{n}},\ df = n - 1$	Estimate a mean difference from matched pairs or before/after data.
Paired t statistic	$t = \frac{\bar{x}_d - 0}{s_d/\sqrt{n}}$	Test whether the mean difference is zero.

Chi-Square Procedures

Name	Formula	When to use
Chi-square statistic	$\chi^2 = \sum \frac{(O-E)^2}{E}$	Compare observed counts with the counts expected under H_0.
Expected count, goodness of fit (extension)	$E_i = n\,p_i$	One categorical variable against a stated distribution.
Degrees of freedom, goodness of fit (extension)	$df = \text{categories} - 1$	Chi-square goodness-of-fit test.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Name	Formula	When to use
Expected count, two-way table	$E = \frac{(\text{row total})(\text{column total})}{\text{table total}}$	Homogeneity and independence tests.
Degrees of freedom, two-way table	$df = (r-1)(c-1)$	Table with r rows and c columns.
Largest contribution	$\frac{(O-E)^2}{E}$ for each cell	Find which cells drive a significant result.

The College Board revision page names only goodness of fit as removed, but the 2027 unit titles have no chi-square unit, so confirm with your teacher whether homogeneity and independence are tested.

Inference for Regression Slopes (extension)

Name	Formula	When to use
Slope t statistic (extension)	$t = \frac{b-\beta_0}{SE_b},\ df = n-2$	Test $H_0 : \beta = 0$, which says there is no linear relationship.
Slope interval (extension)	$b \pm t^* SE_b$	Estimate the true slope β.
Standard error of slope (extension)	$SE_b = \frac{s}{s_x\sqrt{n-1}}$	Appears in computer output as 'SE Coef'.

Margin of Error and Sample Size

Name	Formula	When to use
Margin of error (proportion)	$ME = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	Half-width of an interval for p.
Margin of error (mean)	$ME = t^*\frac{s}{\sqrt{n}}$	Half-width of an interval for μ.
Sample size for a proportion	$n \geq \left(\frac{z^*}{ME}\right)^2 \hat{p}(1-\hat{p})$	Use $\hat{p} = 0.5$ if no estimate exists; round up.
Sample size for a mean	$n \geq \left(\frac{z^*\sigma}{ME}\right)^2$	Planning a study with a rough guess for σ; round up.
Common critical values	$z^* = 1.645,\ 1.960,\ 2.576$	90, 95, and 99 percent intervals for proportions.
Type I and II errors	$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true}),\ \text{power} = 1-\beta$	Interpreting error risks and the strength of a test.

Decision Rule for Every Test

If the p-value is less than or equal to α, reject H_0 and state convincing evidence for H_a in context.

If the p-value exceeds α, fail to reject H_0 and state that the evidence is not convincing; never say H_0 is proven.

A confidence interval that excludes the null value agrees with rejecting H_0 in a two-sided test at the matching level.

R.2 Conditions Checklist for Every Procedure

A one-stop checklist of the conditions behind every inference procedure, how to verify each in context, how to choose the procedure, and the four-step templates.

Why Conditions Come First

Every inference procedure is a mathematical model, and the model gives correct probabilities only when its assumptions are reasonable. Graders reward you for naming each condition and then checking it with facts from the problem, not for reciting a list. A condition is checked in context: write the condition, then the evidence, for example 'Random: students were chosen by a random number generator from the roster.' Procedures marked **(extension)** cover topics College Board lists as removed from the 2027 exam core; check your teacher's current course description.

Conditions for Each Procedure

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

One-Proportion z Procedures (interval and test)

Random: the data come from a random sample or a randomized experiment. Verify by quoting how individuals were selected.

10 percent: if sampling without replacement, $n \leq 0.10N$. Verify by estimating the population size N and comparing.

Large Counts: for a test, $np_0 \geq 10$ and $n(1-p_0) \geq 10$ using the null value. For an interval, $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$ using the sample proportion. Verify by computing both counts.

Two-Proportion z Procedures

Random: two independent random samples, or one randomized experiment with random assignment to two treatments.

10 percent: each sample is at most 10 percent of its own population, $n_1 \leq 0.10N_1$ and $n_2 \leq 0.10N_2$ (not needed for experiments).

Large Counts: at least 10 successes and 10 failures in each group. For a test, many courses compute the counts with the pooled $\hat{p}_c$; for an interval use each group's own counts. Verify with all four numbers $n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$, $n_2(1-\hat{p}_2)$.

Independence between groups: the two groups share no individuals and one group's outcomes do not affect the other's.

One-Sample t Procedures

Random: random sample from the population, or random assignment in an experiment with a single treatment.

10 percent: $n \leq 0.10N$ when sampling without replacement.

Normal or large sample: the population is approximately normal, or $n \geq 30$ so the central limit theorem applies. For smaller n, graph the sample (dotplot, boxplot, or histogram) and look for no strong skew and no outliers.

Two-Sample t Procedures

Random: independent random samples from two populations, or random assignment of subjects to two treatments.

10 percent: each sample is at most 10 percent of its population when sampling.

Normal or large sample: each group has $n \geq 30$, or each group's graph shows no strong skew or outliers. Check each group separately and state both results.

Paired t Procedures

Random: pairs are randomly selected, or the order of treatments within each pair is randomized.

10 percent: the number of pairs is at most 10 percent of all possible pairs.

Normal or large sample: graph the differences, not the original columns. The differences are approximately normal with no strong skew or outliers, or the number of pairs is at least 30.

Chi-Square Goodness of Fit (extension)

Random: one random sample, with each individual placed in exactly one category.

10 percent: $n \leq 0.10N$ when sampling without replacement.

Large Counts: every expected count $E_i = np_i$ is at least 5. Verify by listing all the expected counts; observed counts need not be large. Degrees of freedom are categories minus 1.

Chi-Square Homogeneity and Independence

Random: homogeneity uses independent random samples from several populations or random assignment to treatments; independence uses one random sample from one population classified two ways.

10 percent: each sample (or the single sample) is at most 10 percent of its population.

Large Counts: every expected count, $E = \frac{(\text{row total})(\text{column total})}{\text{table total}}$, is at least 5. Verify by computing or displaying the expected-count matrix, not the observed table.

The College Board revision page names only goodness of fit as removed, but the 2027 unit titles have no chi-square unit, so confirm with your teacher whether these two tests are covered.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Slope t Procedures, Linear Regression (extension)

Linear: the scatterplot looks roughly linear and the residual plot shows no curved pattern.

Independent: observations are independent; for sampling without replacement, $n \leq 0.10N$.

Normal: for each x, the residuals vary approximately normally around the line; check a dotplot or histogram of residuals for no strong skew or outliers.

Equal SD: the spread of the residuals is roughly the same at every x; check that the residual plot has no fan shape.

Random: data come from a random sample or randomized experiment. Together these are often remembered as LINER.

Matching Scenario Clues to Procedures

Clue in the problem	Procedure	Why
One categorical variable, one group, a population proportion or percent	1-proportion z	Parameter is p for a single group.
Compare success rates of two independent groups or two treatments	2-proportion z	Parameter is $p_1 - p_2$.
One quantitative variable, one sample, a mean is claimed or estimated	1-sample t	Parameter is μ, with σ unknown.
Quantitative response measured in two separate groups	2-sample t	Parameter is $\mu_1 - \mu_2$; the groups are unrelated.
Same people measured twice, twins, matched subjects, left versus right	Paired t	Analyze the differences as one sample.
Counts in categories compared with a claimed distribution (extension)	Chi-square goodness of fit	One categorical variable against a model.
Counts for several groups, same categorical variable, groups sampled separately	Chi-square homogeneity	Is the distribution the same across groups?
Two categorical variables measured on one sample	Chi-square independence	Are the variables associated?
Two quantitative variables and a question about the true slope (extension)	Slope t	Parameter is β.
Scores are before and after for the same students	Paired t, not 2-sample	Individuals are linked, so differences are used.
Percent of a sample who agree, compared with a stated target	1-proportion z test	The target is p_0.

Four-Step Templates

Four-Step Template for a Significance Test

The same four steps work for every test; only the details change.

1. **State.** Name the parameter in context. Write H_0 and H_a with the parameter symbol. State the significance level α if given, and decide one-sided or two-sided from the wording.
2. **Plan.** Name the procedure and verify every condition with evidence from the problem (Random, 10 percent, Large Counts or Normal or large sample).
3. **Do.** Give the test statistic, degrees of freedom if any, and the p-value. Show the substitution into the formula, or name the built-in calculator function used.
4. **Conclude.** Compare the p-value with α. Say reject or fail to reject H_0, and state in context whether there is convincing evidence for H_a.

Four-Step Template for a Confidence Interval

Intervals use the same structure with different content.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

1. **State.** Name the parameter in context and the confidence level.
2. **Plan.** Name the interval procedure and verify the conditions.
3. **Do.** Write the interval as statistic $\pm$ critical value $\times$ standard error, with the critical value and the final interval endpoints.
4. **Conclude.** Interpret: 'We are __ percent confident that the interval from __ to __ captures the true __ (parameter in context).' Link to a claim only if asked.

Model Phrases for Verifying Conditions

Condition	Weak statement	Full-credit statement
Random	The sample was random.	The 60 students were selected by a random number generator from the school roster, so the sample is random.
10 percent	Population is big.	There are about 1,800 students in the school and 60 is less than 10 percent of 1,800, so independence is reasonable.
Large Counts (proportion)	Sample is large.	$np_0 = 60(0.4) = 24 \geq 10$ and $n(1-p_0) = 36 \geq 10$, so the sampling distribution of $\hat{p}$ is approximately normal.
Normal or large sample	Data look normal.	The sample size is 18 and the dotplot is roughly symmetric with no outliers, so a t model is reasonable.
Expected counts	Counts are big.	The smallest expected count is $\frac{(40)(35)}{200} = 7 \geq 5$, so the chi-square model is reasonable.
Experiment	Groups were fair.	Subjects were randomly assigned to treatments, which supports a conclusion about cause and effect.

Common Mistakes with Conditions

Mistake 1: Checking Large Counts with the wrong proportion

In a test of $H_0 : p = p_0$, use p_0, not $\hat{p}$, for np_0 and $n(1-p_0)$. In an interval there is no null value, so use $\hat{p}$. Mixing these up is the most common lost point on proportion problems.

Mistake 2: Using n at least 30 without a sample

The rule $n \geq 30$ rescues the shape of the sampling distribution only for means. It does not apply to proportions, which need successes and failures, and it does not excuse a lack of randomization.

Mistake 3: Graphing the wrong thing in a paired design

In a paired t procedure the Normal condition applies to the differences. Plotting the two original columns separately earns no credit because the model is built on the one-sample differences.

Mistake 4: Verifying observed counts for chi-square

The 'at least 5' rule applies to expected counts. Write the expected count for the smallest cell, or show the expected matrix. Saying 'all observed counts are over 5' does not verify the condition.

Mistake 5: Skipping the 10 percent check or claiming it for experiments

For samples taken without replacement, the 10 percent condition keeps trials nearly independent. For randomized experiments there is no population to sample from, so do not write it; instead state that random assignment was used.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

R.3 Glossary

Over 160 key terms from the whole course, each with a precise one-sentence definition, in alphabetical order.

A to D

Term	Definition
Addition rule	The rule that the probability of A or B equals P(A) plus P(B) minus P(A and B).
Alternative hypothesis	The claim a test seeks evidence for, stating that the parameter differs from the null value in a specified direction or in either direction.
Association	A relationship in which values of one variable tend to go with particular values of another, so knowing one helps predict the other.
Bar chart	A display that shows the count or relative frequency of each category of a categorical variable using separated bars of equal width.
Bias	A systematic tendency of a sampling method, measurement, or estimator to miss the true value in the same direction every time.
Bimodal	Describing a distribution with two distinct peaks, often a sign that two different groups are mixed together.
Binomial distribution	The probability model for the number of successes in a fixed number of independent trials that each have the same success probability.
Binomial setting	A situation with a fixed number of trials, each with two outcomes, independent of each other, and with a constant probability of success.
Blinding	Keeping subjects, evaluators, or both unaware of which treatment each subject received so expectations cannot influence results.
Block	A group of experimental units that are similar in a way that may affect the response, within which treatments are randomly assigned.
Bounded distribution	A distribution that cannot take values beyond a fixed limit, which often creates skew near the limit.
Boxplot	A display of the five-number summary as a box from the first to the third quartile with a line at the median and whiskers to the extremes.
Categorical variable	A variable that places each individual into one of several groups or categories.
Census	A study that collects data from every individual in the population.
Central limit theorem	The result that the sampling distribution of a sample mean is approximately normal when the sample size is large, whatever the population shape.
Chi-square distribution	A right-skewed family of distributions, indexed by degrees of freedom, used to judge how far observed counts stray from expected counts.
Chi-square statistic	The sum over all cells of the squared difference between observed and expected count divided by the expected count.
Cluster sample	A sample formed by randomly choosing whole groups of individuals and then surveying everyone in each chosen group.
Coefficient of determination	The square of the correlation, giving the proportion of variation in the response that is accounted for by the linear model.
Combination	A selection of items in which the order does not matter, counted using n choose r.
Common response	A relationship in which a lurking variable causes changes in both of two variables, producing an association between them.
Comparative experiment	An experiment that applies different treatments to groups so their responses can be compared with each other.
Complement	The event consisting of all outcomes in the sample space that are not in a given event.
Conditional probability	The probability that one event occurs given that another event is known to have occurred.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Term	Definition
Conditions for inference	The checks on randomness, independence, and distribution shape that justify using a particular model for a procedure.
Confidence interval	A range of plausible values for a parameter computed from sample data using a method that captures the true value a stated percent of the time.
Confidence level	The long-run percent of intervals produced by a method that capture the parameter being estimated.
Confounding	A situation in which the effects of two variables on a response cannot be separated because they are mixed together.
Continuous random variable	A random variable that can take any value in an interval, with probabilities given by areas under a density curve.
Control group	The group of experimental units that receive no treatment or a standard treatment, serving as a baseline for comparison.
Convenience sample	A sample made up of individuals who are easiest to reach, which tends to be biased.
Correlation	A number between -1 and 1 that measures the strength and direction of the linear relationship between two quantitative variables.
Correlation versus causation	The reminder that a strong association alone, without a well-designed experiment, does not show that one variable causes the other.
Critical value	The cutoff on a standard model, such as z* or t*, that marks the middle area equal to the confidence level.
Cumulative relative frequency	The proportion of observations at or below a given value or interval.
Degrees of freedom	The number of independent pieces of information that determine the shape of a t or chi-square distribution.
Density curve	A smooth curve above the horizontal axis whose total area is 1 and whose area over an interval gives a proportion of values.
Dependent events	Events for which knowing that one occurred changes the probability of the other.
Descriptive statistics	Numbers and graphs that summarize the data in hand without drawing conclusions about a larger population.
Difference of means	The statistic found by subtracting the mean of one group from the mean of another to compare their centers.
Difference of proportions	The statistic found by subtracting one group's sample proportion from another's.
Discrete random variable	A random variable whose possible values can be listed, usually counts, each with its own probability.

D to N

Term	Definition
Distribution	A description of the values a variable takes and how often it takes each value.
Dotplot	A display that stacks a dot above a number line for each observation, suited to small data sets.
Double-blind	An experiment in which neither the subjects nor the people measuring the response know which treatment each subject received.
Empirical rule	The guideline that about 68, 95, and 99.7 percent of values in a normal distribution lie within 1, 2, and 3 standard deviations of the mean.
Event	A set of outcomes from a chance process to which a probability can be assigned.
Expected count	The number of observations predicted in a table cell if the null hypothesis were true.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Term	Definition
Expected value	The long-run average of a random variable, found by weighting each possible value by its probability.
Experiment	A study in which researchers impose treatments on units and measure a response in order to investigate cause and effect.
Experimental unit	The individual object, person, or group to which a treatment is assigned.
Explanatory variable	The variable used to explain or predict changes in the response variable.
Extrapolation	Using a regression model to predict at values of the explanatory variable outside the range of the data used to build it.
Factor	An explanatory variable in an experiment whose levels are assigned to units as treatments.
Failure to reject	The decision when the p-value is larger than the significance level, meaning the data do not provide convincing evidence for the alternative.
Five-number summary	The minimum, first quartile, median, third quartile, and maximum of a data set.
Frequency table	A table listing each category or interval along with the number of observations that fall in it.
Geometric distribution	The probability model for the number of trials needed to get the first success in independent trials with constant success probability.
Goodness of fit	A comparison of the observed distribution of a categorical variable with a distribution claimed by a model.
Histogram	A display of a quantitative distribution in which adjacent bars show the count or percent of values in equal-width intervals.
Homogeneity	The claim that several populations have the same distribution of a categorical variable.
Hypothesis test	A formal procedure that uses sample data to decide whether there is convincing evidence against a null hypothesis.
Independence of variables	The claim that two categorical variables have no association in a population.
Independent events	Events such that knowing one occurred does not change the probability of the other.
Inference	The use of sample data to draw conclusions about a population or process, with a statement of how reliable those conclusions are.
Influential point	A point whose removal would noticeably change the slope, intercept, or correlation of a regression line.
Interquartile range	The distance between the third and first quartiles, measuring the spread of the middle half of the data.
Joint probability	The probability that two events both occur.
Law of large numbers	The principle that the relative frequency of an outcome approaches its true probability as the number of repetitions grows.
Least-squares regression line	The line that makes the sum of the squared residuals as small as possible.
Level of a factor	One of the specific values or settings of a factor used in an experiment.
Linear transformation	A change of the form a plus bx applied to every value, which changes center and spread in predictable ways.
Lower fence	The value Q1 minus 1.5 times the IQR, below which an observation is flagged as a possible outlier.
Lurking variable	A variable that is not among those studied but influences both the explanatory and response variables, creating a misleading association.
Margin of error	The half-width of a confidence interval, showing how far the estimate is likely to fall from the parameter because of sampling variability.
Matched pairs design	An experimental or sampling design in which subjects are paired by similarity or measured twice, and the analysis uses the differences.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Term	Definition
Mean	The arithmetic average of a set of numbers, found by adding all the values and dividing by how many there are.
Median	The middle value of an ordered data set, with half of the values at or below it.
Mode	The value or category that occurs most often in a data set.
Multistage sample	A sample chosen by combining two or more sampling methods in successive steps.
Mutually exclusive events	Events that cannot occur together, so their intersection has probability zero.
Negative association	A pattern in which larger values of one variable tend to go with smaller values of the other.
Nonresponse bias	Bias that occurs when selected individuals do not respond and differ in relevant ways from those who do.
Normal distribution	A symmetric, bell-shaped model completely described by its mean and standard deviation.

N to S

Term	Definition
Null hypothesis	The default claim, usually of no difference or no effect, that a test assumes true when computing the p-value.
Observational study	A study that observes individuals and measures variables without imposing any treatment.
One-sided alternative	An alternative hypothesis that states the parameter is either greater than or less than the null value, but not both.
Outcome	One possible result of a chance process.
Outlier	A value that falls far from the rest of the distribution, often flagged using the 1.5 times IQR rule.
P-value	The probability, assuming the null hypothesis is true, of getting a result at least as extreme as the one observed.
Paired data	Two measurements that are linked to the same individual or to a matched pair.
Parameter	A number that describes a characteristic of a population.
Percentile	The value below which a stated percent of the observations in a distribution fall.
Placebo	An inactive treatment that looks like the real treatment, used to separate treatment effects from effects of expectation.
Point estimate	A single value computed from a sample that serves as a best guess for a parameter.
Pooled proportion	The overall proportion of successes found by combining two samples, used for a test of equal population proportions.
Population	The entire group of individuals about which we want information.
Positive association	A pattern in which larger values of one variable tend to go with larger values of the other.
Power	The probability that a test correctly rejects a false null hypothesis.
Probability	A number between 0 and 1 giving the long-run relative frequency of an outcome of a chance process.
Probability distribution	A list or model of every possible value of a random variable together with its probability.
Quantitative variable	A variable that takes numerical values for which arithmetic such as averaging makes sense.
Quartile	One of three values that divide an ordered data set into four parts holding about equal numbers of observations.
Random assignment	The use of chance to decide which treatment each experimental unit receives.
Random sample	A sample selected by a chance process so that every individual has a known chance of being chosen.
Random variable	A numerical quantity whose value is determined by the outcome of a chance process.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Term	Definition
Randomization	The use of chance in selecting samples or assigning treatments to prevent systematic differences.
Range	The difference between the largest and smallest values in a data set.
Relative frequency	The count in a category divided by the total number of observations, often written as a proportion or percent.
Replication	Using enough experimental units in each treatment group so that real differences can be separated from chance variation.
Residual	The observed response minus the response predicted by the regression line.
Residual plot	A scatterplot of residuals against the explanatory variable or the predicted values, used to judge whether a line fits well.
Resistant measure	A statistic that is not much affected by extreme values, such as the median and the IQR.
Response bias	Bias caused by a poorly worded question, a dishonest answer, or an interviewer who influences replies.
Response variable	The variable that measures the outcome of a study and is expected to depend on the explanatory variable.
Robust procedure	An inference method that still gives approximately correct results when its conditions are only slightly violated.
Sample	The subset of the population from which data are actually collected.
Sample space	The set of all possible outcomes of a chance process.
Sample survey	A study that gathers information from a sample of individuals to learn about a population.
Sampling distribution	The distribution of values a statistic takes over all possible samples of the same size from the same population.
Sampling frame	The list of individuals from which a sample is drawn.
Sampling variability	The natural variation in the value of a statistic from one random sample to the next.
Sampling without replacement	A selection process in which an individual cannot be chosen more than once, which slightly links successive choices.
Scatterplot	A graph that shows each individual as a point positioned by its values on two quantitative variables.
Shape	The overall pattern of a distribution, described by symmetry, skewness, peaks, and gaps.
Significance level	The cutoff probability, usually written alpha, below which a p-value is judged small enough to reject the null hypothesis.

S to Z

Term	Definition
Simple random sample	A sample in which every group of the chosen size has the same chance of being the sample.
Simulation	A procedure that uses random digits or software to imitate a chance process many times in order to estimate probabilities.
Skewed distribution	A distribution with a longer tail on one side, named by the direction of the longer tail.
Slope	The predicted change in the response for each one-unit increase in the explanatory variable.
Standard deviation	A measure of spread giving the typical distance of observations from the mean.
Standard error	The estimated standard deviation of a statistic, computed from sample data.
Standard error of the slope	The estimated standard deviation of the sample slope across repeated samples, used in slope intervals and tests.
Standard normal distribution	The normal distribution with mean 0 and standard deviation 1.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Term	Definition
Standardizing	Converting a value to a z-score by subtracting the mean and dividing by the standard deviation.
Statistic	A number that describes a characteristic of a sample and is used to estimate a parameter.
Statistically significant	Describing a result whose p-value is at or below the chosen significance level, so chance alone is an unlikely explanation.
Stemplot	A display that splits each value into a stem and a leaf, keeping the individual data values visible.
Stratified random sample	A sample formed by dividing the population into similar groups and taking a random sample from each group.
Subjects	The people who are the experimental units in an experiment.
Symmetric distribution	A distribution whose left and right sides are approximately mirror images.
Systematic sample	A sample formed by choosing a random starting point and then every kth individual from an ordered list.
t distribution	A symmetric, bell-shaped family with heavier tails than the normal, indexed by degrees of freedom, used when the population standard deviation is estimated.
Test statistic	A standardized value that measures how many standard errors a sample statistic lies from the value stated in the null hypothesis.
Transformation of data	Applying a function such as the logarithm to each value to change the shape or straighten a relationship.
Treatment	A specific condition or combination of factor levels applied to the experimental units.
Tree diagram	A diagram that shows the branches of a multistep chance process along with the probability of each branch.
Two-sided alternative	An alternative hypothesis that states the parameter differs from the null value in either direction.
Two-way table	A table of counts that classifies individuals by two categorical variables.
Type I error	Rejecting a null hypothesis that is actually true, with probability equal to the significance level.
Type II error	Failing to reject a null hypothesis that is actually false.
Unbiased estimator	A statistic whose sampling distribution is centered at the parameter it estimates.
Undercoverage	A bias that occurs when some members of the population cannot be chosen for the sample.
Unimodal	Describing a distribution with a single clear peak.
Unit of measurement	The scale in which a variable is recorded, such as minutes or kilograms, which must be named when interpreting results.
Upper fence	The value Q3 plus 1.5 times the IQR, above which an observation is flagged as a possible outlier.
Variability	The extent to which values in a distribution or a statistic differ from one another.
Variable	A characteristic that takes different values for different individuals.
Variance	The square of the standard deviation, found as an average of squared deviations from the mean.
Venn diagram	A picture using overlapping regions to show how events share and do not share outcomes.
Voluntary response sample	A sample consisting of individuals who choose to respond to an open invitation, which tends to overrepresent strong opinions.
Well-designed experiment	An experiment with comparison, random assignment, and replication, and with control of outside variables.
Whisker	The line on a boxplot that extends from the box to the smallest or largest value that is not an outlier.
Y-intercept	The predicted response when the explanatory variable equals zero, meaningful only if zero is a sensible value.
z-score	The number of standard deviations a value lies above or below the mean of its distribution.

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

Tables

Statistical Tables

Standard normal probabilities, t critical values, and chi-square critical values. Original computed tables, provided for practice; use the reference materials supplied on exam day for the real exam.

Table A. Standard Normal: area to the left of z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table B. t critical values t* by confidence level (two-sided)

df	80%	90%	95%	98%	99%
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

df	80%	90%	95%	98%	99%
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
80	1.292	1.664	1.990	2.374	2.639
100	1.290	1.660	1.984	2.364	2.626
1000	1.282	1.646	1.962	2.330	2.581

Table C. Chi-square critical values: area to the right

df	0.1	0.05	0.025	0.01	0.005
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300
13	19.812	22.362	24.736	27.688	29.819
14	21.064	23.685	26.119	29.141	31.319
15	22.307	24.996	27.488	30.578	32.801
16	23.542	26.296	28.845	32.000	34.267
17	24.769	27.587	30.191	33.409	35.718
18	25.989	28.869	31.526	34.805	37.156

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.

df	0.1	0.05	0.025	0.01	0.005
19	27.204	30.144	32.852	36.191	38.582
20	28.412	31.410	34.170	37.566	39.997
25	34.382	37.652	40.646	44.314	46.928
30	40.256	43.773	46.979	50.892	53.672
40	51.805	55.758	59.342	63.691	66.766
50	63.167	67.505	71.420	76.154	79.490
60	74.397	79.082	83.298	88.379	91.952
80	96.578	101.879	106.629	112.329	116.321
100	118.498	124.342	129.561	135.807	140.169

Want a plan built around your student?
Book a free consultation. The short intake takes about a minute and opens right away.
BOOK NOW →

© 2026 Scholar Coach, Inc. All rights reserved. Copying or reproducing strictly prohibited.