



Sampling, Study Design, and Confidence Intervals

Four focused lessons, worked examples, a mistake clinic, and 60 practice problems with full explanations.

In this booster

33.1 Sampling Methods and Bias

33.2 Experiments vs Observational Studies

33.3 Confidence Intervals for Proportions

33.4 Confidence Intervals for Means

Every section teaches the idea, walks through the most common problem types, flags the traps students fall into, and ends with fifteen practice problems. Full answers and explanations are at the back.

33.1 Sampling Methods and Bias

Learn how samples are chosen, how to tell stratified, cluster, and systematic samples apart, and how to name bias precisely. These ideas decide whether any later inference can be trusted.

Why This Matters

Every poll, clinical trial, and quality check begins with the same question: how did the data get chosen? A sample is only useful if it lets you say something trustworthy about a larger **population**. In AP Statistics, most of the later machinery (confidence intervals, significance tests) quietly assumes that the data came from a random process. If it did not, the formulas still produce numbers, but the numbers mean nothing.

Students most often get stuck in four places. They confuse **stratified** and **cluster** samples because both split the population into groups. They think a larger sample fixes a biased method (it does not). They describe bias in vague words like "unfair" instead of naming the type and the direction. And they forget that a sample can be chosen with some randomness and still not be a simple random sample. This section gives you precise definitions and a routine for writing complete AP-style answers.

Core Ideas

The **population** is the entire group you want to learn about. A **sample** is the part of the population you actually examine. A number that describes a population is a **parameter** (for example, the true proportion p of all students who walk to school). A number computed from a sample is a **statistic** (for example, the sample proportion $\hat{p}$). We use statistics to estimate parameters.

Simple Random Sample (SRS)

An SRS of size n is chosen so that **every group of n individuals in the population has an equal chance of being the sample**.

Consequences: every individual has an equal chance of selection, and every possible combination of n individuals is equally likely.

The number of possible samples is $\binom{N}{n}$.

To pick an SRS in practice you label every individual, then use a random number generator or a table of random digits. With a table, use labels with the same number of digits, read groups of digits, **skip labels that do not exist, 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 →

Method	How it works	Best used when	Main risk
Simple random	Every group of size n equally likely	A complete list exists	Can miss small subgroups by chance
Stratified	Split into similar groups (strata), take an SRS from each stratum	Groups differ on the thing you measure	Needs information to form strata
Cluster	Split into mixed groups (clusters), randomly choose whole clusters, survey everyone in them	The population is spread out and a list is hard to get	Clusters may be alike inside, which adds variability
Systematic	Choose a random start, then every kth individual	Individuals come in a natural order	A hidden pattern in the list can bias results

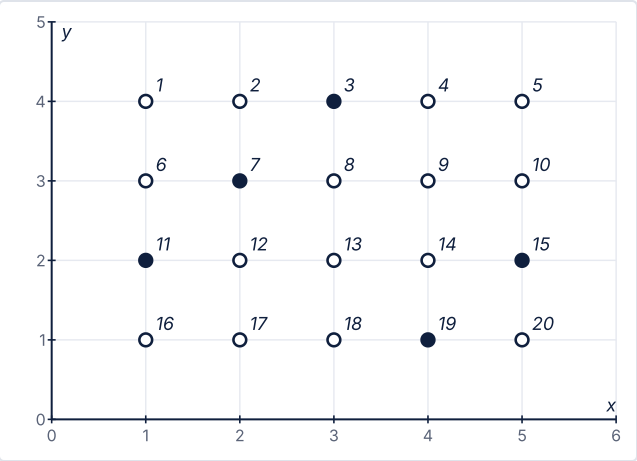
Stratified versus Cluster

Stratified: groups are **similar inside** and **different from each other**; you sample from **every** group.

Cluster: each group is a **miniature version of the whole population**; you sample **some** groups completely.

Stratifying reduces variability. Clustering saves time and money but usually increases variability.

A seating chart makes systematic sampling concrete. In the figure below, the 20 seats are numbered across the rows. Choosing a random start of 3 and then every 4th seat selects seats 3, 7, 11, 15, and 19 (the solid points).



A systematic sample from 20 numbered seats: start at 3, then every 4th seat. Solid points are selected.

Methods that are **not** random are the main sources of bias. A sampling method is **biased** if it systematically favors some outcomes, so the statistic tends to miss the parameter in the same direction again and again.

Common Sources of Bias

Convenience sample: you take whoever is easy to reach.

Voluntary response: people choose themselves to respond, so strong opinions are overrepresented.

Undercoverage: some part of the population has no chance of being chosen.

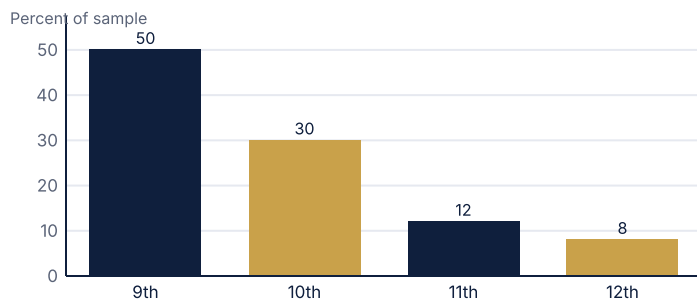
Nonresponse: chosen individuals cannot be reached or refuse.

Response bias: answers are untrue because of wording, the interviewer, or social pressure.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

[BOOK NOW →](#)



A survey taken only in the morning bus line. The students in the sample were 50, 30, 12, and 8 percent from grades 9 to 12, but the school has 35, 30, 20, and 15 percent. The method undercovers older students who drive.

Two ideas are easy to mix up. **Bias** is about the *method*: it does not shrink when you collect more data. **Variability** is about *chance*: it shrinks when n grows. A huge voluntary response survey is still biased. A small SRS is unbiased but varies more from sample to sample.

Worked Examples

EXAMPLE 1

A school has 1,200 students. The principal wants an opinion on a new schedule. For each plan, name the sampling method. (a) Number all students and use a random number generator to select 80. (b) Separate students by grade and randomly select 20 from each grade. (c) Randomly select 4 homerooms and survey every student in them. (d) Pick a random start between 1 and 15 on the alphabetical roster, then take every 15th student.

SOLUTION

- (a) Every group of 80 students is equally likely: a **simple random sample**.
 (b) Grades are the groups, and a random sample is taken from **each** group: a **stratified random sample**.
 (c) Whole groups are chosen at random and everyone inside is surveyed: a **cluster sample**.
 (d) Random start, then a fixed interval: a **systematic sample** ($1200/15 = 80$ students).

EXAMPLE 2

A club has 40 members labeled 01 to 40. Use this string of random digits to select an SRS of 4 members: 71829 40653 12974 38806 55291 37046 18.

SOLUTION

Labels have two digits, so read the digits two at a time: 71, 82, 94, 06, 53, 12, 97, 43, 88, 06, 55, 29, 13, 70, 46, 18. Skip any number above 40 (71, 82, 94, 53, 97, 43, 88, 55) and skip the repeated 06. The first four valid labels are **06, 12, 29, 13**.

EXAMPLE 3

An SRS of 5 students is chosen from a class of 25. (a) What is the chance that a particular student, Mia, is selected? (b) How many different samples of 5 are possible?

SOLUTION

(a) Each student has an equal chance, so $P(\text{Mia}) = \frac{5}{25} = 0.20$.

(b) The number of samples is $\binom{25}{5} = \frac{25 \cdot 24 \cdot 23 \cdot 22 \cdot 21}{5!} = 53,130$.

Each of these 53,130 samples is equally likely, with probability $\frac{1}{53130}$.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

EXAMPLE 4

A manager orders 600 shipping boxes numbered 1 to 600. She wants to inspect 30 using a systematic sample with a random start.

(a) Find the interval k . (b) If the random start is 7, list the first three boxes and the last box. (c) Is this an SRS?

SOLUTION

(a) $k = \frac{600}{30} = 20$. The start is chosen at random from 1 to 20.

(b) The boxes are 7, 27, 47, and so on. The last is $7 + 29 \cdot 20 = 587$.

(c) No. Only 20 different samples are possible (one for each start), so most groups of 30 boxes, such as boxes 1 to 30, have **no chance** of being chosen. Every individual has an equal chance, but not every group does.

EXAMPLE 5

Identify the main type of bias and the likely direction. (a) A radio station asks listeners to call in about whether the town should build a new stadium. (b) A phone survey about exercise is conducted only on landlines at 2 p.m. on weekdays. (c) A survey asks, "Do you agree that wasteful spending on the school's bloated sports budget should end?"

SOLUTION

(a) **Voluntary response**. People with strong opinions are most likely to call, so the results tend to overstate extreme views.

(b) **Undercoverage** (people with only cell phones) and **nonresponse** (people at work). The sample likely overrepresents retired people and at-home workers, so exercise rates may be biased in the direction of that group.

(c) **Response bias from wording**. Words like "wasteful" and "bloated" push respondents toward agreeing, so support for cutting the budget will be overstated.

EXAMPLE 6

A school of 1,200 students has 420 freshmen, 360 sophomores, 240 juniors, and 180 seniors. A counselor will take a stratified sample of 100 students using proportional allocation. How many should be chosen from each grade, and why might stratifying help?

SOLUTION

Each grade's share of the sample should equal its share of the school: $\frac{100}{1200} = \frac{1}{12}$.

Freshmen $420/12 = 35$, sophomores $360/12 = 30$, juniors $240/12 = 20$, seniors $180/12 = 15$. The total is 100.

Stratifying guarantees every grade is represented in the right proportion. If opinions differ by grade, this removes grade-to-grade chance variation, so the estimate is **more precise** than an SRS of the same size.

Coach's Playbook

- 1 **Name the population and the parameter first.** Many errors come from not knowing who the results should apply to.
- 2 **Ask who could never be chosen.** If a group has zero chance, that is undercoverage.
- 3 **Ask who chooses to answer.** If people select themselves, it is voluntary response.
- 4 **Check the groups.** Similar inside and different from each other means stratify. Each a mini-population and you pick some of them means cluster.
- 5 **Describe the direction of bias.** Say whether the statistic will be too high or too low and why.
- 6 **Never say a bigger sample fixes bias.** It only reduces variability.

A useful mental model: **stratify to be precise, cluster to be cheap**. Another: an SRS is a lottery on groups, not just on individuals. When a problem asks you to "describe a procedure," give four parts: the labeling, the random mechanism, how repeats are handled, and what you record. The scorers look for 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 →

Common traps

1. Calling a sample "random" just because the selection involved chance, when not every group of size n was possible.
2. Mixing up strata and clusters. Strata are all sampled; clusters are only some chosen.
3. Saying "the sample is too small" as the cause of bias. Small samples cause variability, not bias.
4. Blaming nonresponse bias on people "not caring" instead of explaining how the nonrespondents differ from respondents.

Mistake Clinic**Spot the error 1**

A student writes: "To get a random sample, I will stand at the school entrance and survey every 10th student who walks in. This is an SRS because I chose students by a fixed rule, not by my own preference." Where is the error? A fixed rule is systematic, not SRS: groups such as two friends who arrive together can never both be chosen, so not every group is equally likely. Also, only students who arrive at that entrance can be sampled (undercoverage). A correct plan: label all students on the roster and use a random number generator to select the sample.

Spot the error 2

A student writes: "A county has 30 high schools. I will randomly choose 5 high schools and then randomly choose 20 students at each. This is a stratified sample because I used groups." Where is the error? Only 5 of the 30 schools are used, but stratified sampling requires taking a sample from **every** group. This design is a **multistage cluster sample**. A stratified sample would take students from all 30 schools.

Spot the error 3

A student writes: "The online poll got 8,000 responses, and the margin of error is tiny, so the results are very reliable." Where is the error? Sample size does not remove bias. An online poll where people choose to respond is voluntary response, so those with strong opinions are overrepresented. The huge n only makes the biased estimate more consistent. A correct statement: the results cannot be generalized to the population because the method is biased.

Real-World Connection

Public health teams estimate how many people in a region have had a recent illness. They often use cluster sampling: choose random neighborhoods, then test everyone in the chosen areas. Travel costs drop, but because neighbors share conditions, the team needs a larger sample to match the precision of an SRS.

Challenge Corner

A town has 5,000 households: 3,000 in apartments and 2,000 in houses. A survey of 200 households uses proportional stratified sampling. Then 40 percent of the sampled apartment households and 70 percent of the sampled house households say yes to a proposal. Find the stratified estimate of the proportion of all households that say yes, and compare with a plan that surveys only 200 house households. Solution: Apartments: $200 \cdot 0.6 = 120$ sampled, and $0.40 \cdot 120 = 48$ yes. Houses: 80 sampled, and $0.70 \cdot 80 = 56$ yes. Estimate = $\frac{48 + 56}{200} = 0.52$. The house-only plan would give about 0.70, which is far too high because apartments have no chance of being included (undercoverage).

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

Practice

PRACTICE Sampling Methods and Bias

Show your work on paper. Use a calculator only when a problem says so, or if it is allowed at your level.

- Which of the following describes a simple random sample of 10 students from a school?
 - Choose 10 students from each grade at random
 - Choose a random start and then every 50th student on the roster
 - Every group of 10 students has the same chance of being the sample
 - Choose the first 10 students who volunteer
- A zoo wants to estimate the average age of its visitors. It randomly chooses 6 days of the year and surveys every visitor on those days. This is a
 - stratified sample
 - cluster sample
 - systematic sample
 - simple random sample
- A district surveys parents by calling only listed home phone numbers. Which problem is most likely?
 - Undercoverage
 - Voluntary response
 - Response bias from wording
 - Small sample size
- Which survey question is most likely to produce response bias?
 - Do you plan to attend the school play?
 - On average, how many hours do you sleep on school nights?
 - Do you agree that our excellent teachers deserve more support?
 - Which grade are you in?
- An SRS of 3 students is chosen from a club of 10. How many different samples are possible?
 - 30
 - 120
 - 720
 - 1,000
- A list of 450 invoices is checked using a systematic sample of 30 with a random start of 4. Which invoice is in the sample?
 - 124
 - 120
 - 130
 - 135
- When is stratified sampling most helpful?
 - When the population is spread over a large area and has no list
 - When the groups differ from each other on the variable being measured
 - When the sample size is very small
 - When every individual refuses to respond
- A pollster doubles the sample size of a survey that uses voluntary response. Which statement is true?
 - The bias is cut in half
 - The bias is eliminated
 - The bias remains but the variability decreases
 - The method becomes an SRS
- A town has 1,500 residents, and 450 live in the north district. A proportional stratified sample of 60 is taken. How many come from the north district?
 - 15
 - 18
 - 30
 - 45
- A coach numbers 60 swimmers from 1 to 60. She picks a random number from 1 to 6, then selects every 6th swimmer after it. Name the method. Is it a simple random sample? Explain.
- A principal posts an online poll on the school website asking if homework should be reduced, and 300 students respond. Identify the type of bias and state which way it likely pushes the result.
- A club has 48 members labeled 01 to 48. Use the random digits 63179 04117 08255 23644 to select an SRS of 6 members. List the labels.
- A county has 2,400 voters: 1,080 in District A, 840 in District B, and 480 in District C. Describe how many voters a proportional stratified sample of 120 should take from each district.
- A city has 40 apartment buildings. A researcher randomly chooses 3 buildings and surveys every unit in them. (a) What is the chance a particular building is chosen? (b) Why is this not an SRS of units?

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

15. A mail survey is sent to 500 people. Only 120 return it, and 90 of those 120 favor a proposal. (a) What is the response rate? (b) Suppose 40 percent of the 380 nonrespondents also favor it. Find the true proportion in the 500 and explain why the 75 percent from respondents is misleading.

33.2 Experiments vs Observational Studies

Learn to tell experiments from observational studies, design experiments with control, randomization, blinding, and blocking, and decide what conclusions a study really supports.

Why This Matters

News headlines announce that a food "causes" a disease or that a habit "leads to" better grades. Whether such a claim is justified depends entirely on how the data were produced. AP Statistics asks you to look at a study and decide, in one or two sentences, what the design allows you to conclude. This is the skill that separates evidence from a story.

Students struggle with the vocabulary (treatment, experimental unit, block, confounding variable) and with the final step: choosing the right conclusion. The two most common slips are saying that an observational study "proves" cause and effect, and forgetting that **random assignment** (not random selection) is what supports a cause-and-effect conclusion. You will also learn how to write a complete experimental design, which shows up on free-response questions again and again.

Core Ideas

In an **observational study** you observe individuals and measure variables, but you do not impose any treatment. In an **experiment** you deliberately impose a **treatment** on **experimental units** (called **subjects** when they are people) and then measure a **response variable**. The variable you manipulate is the **explanatory variable**; its values are the treatments.

Three Principles of Experimental Design

Control: compare treatments to each other or to a control group (a placebo or an existing standard), and keep other conditions the same.

Random assignment: use chance to assign units to treatments so the groups are alike before treatments start.

Replication: use enough units in each group so that real differences can be separated from chance variation.

A **placebo** is a dummy treatment that looks like the real one. It matters because simply receiving a treatment can change responses. In a **single-blind** study, either the subjects or the people measuring the response do not know which treatment was given. In a **double-blind** study, neither knows.

Design	Description	Use it when
Completely randomized	All units are randomly assigned to the treatments	Units are fairly similar
Randomized block	Group similar units into blocks , then randomize treatments within each block	A known variable (age, ability, field location) affects the response
Matched pairs	Each unit gets both treatments in random order, or similar units are paired and one gets each treatment at random	You can control for individual differences

Confounding

Two variables are **confounded** when their effects on the response cannot be separated. A **lurking variable** is a variable that is not among the explanatory variables but may influence the response and the explanatory variable.

In an observational study, a difference between groups can always be explained by a confounding variable. Random assignment balances confounders across groups in an experiment, so differences are then attributable to the treatments.

Blocking and random assignment do different jobs. **Blocking** removes the variation due to a known variable. **Random assignment** protects against the variables you did not think of. The diagram below shows a completely randomized design for comparing two study methods. Notice that random assignment comes first, and that the same response is measured in both groups.

Want a plan built around **your** student?

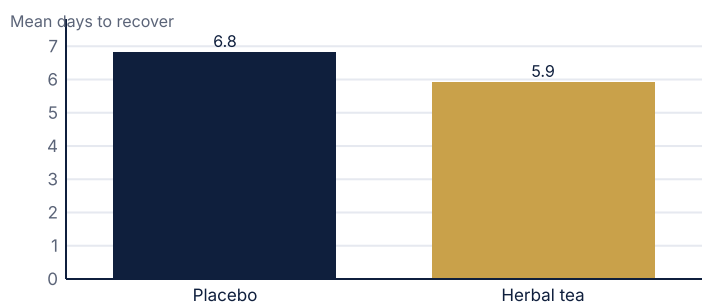
Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

Step	Group 1	Group 2
Random assignment	20 subjects	20 subjects
Treatment	Method A	Method B
Measure	Score on same quiz	Score on same quiz
Compare	Mean score of Group 1	Mean score of Group 2

Finally, you need to know what conclusion a design supports. This is called the **scope of inference**. Random *selection* of units from a population allows generalizing to that population. Random *assignment* to treatments allows a cause-and-effect conclusion.

	Random assignment	No random assignment
Random selection	Cause and effect, and the result can be generalized to the population	Can generalize to the population, but no cause and effect
No random selection	Cause and effect, but only for units like those in the study	Neither: only describes the sample



Results of an experiment with 30 volunteers per group: mean recovery time in days. The groups were created by random assignment.

Worked Examples

EXAMPLE 1

Decide whether each is an observational study or an experiment, and name the explanatory and response variables. (a) Researchers record how many hours 150 students sleep and their quiz scores. (b) Of 60 volunteers, a random half drink a sports drink before a run and the others drink water; run times are compared. (c) A hospital compares recovery times for patients who chose surgery with patients who chose physical therapy.

SOLUTION

- (a) **Observational.** No treatment is imposed. Explanatory: hours of sleep. Response: quiz score.
- (b) **Experiment.** The researchers impose the drink and use random assignment. Explanatory: type of drink. Response: run time.
- (c) **Observational.** Patients chose their own treatment, so the groups may differ in other ways (severity, age). Explanatory: type of treatment. Response: recovery time.

EXAMPLE 2

A gardener has 30 identical plants and wants to compare 3 fertilizers (A, B, C). Describe a completely randomized design.

SOLUTION

Label the plants 01 to 30. Use a random number generator to put the labels in a random order (a random permutation). The first 10 labels receive fertilizer A, the next 10 receive B, and the last 10 receive C. Equal group sizes make the comparison clean. Give every plant the same water, light, and soil, and after a fixed time measure the same response (for example, height in centimeters). Compare the three group means. Random assignment makes the three groups similar before treatment, so a large difference in height can be attributed to the fertilizer.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

EXAMPLE 3

A coach compares two shoe types on 24 runners: 12 experienced and 12 novices. Experience strongly affects times. Describe a randomized block design and explain why blocking helps.

SOLUTION

Blocks: experienced runners and novices.

Within the experienced block, randomly assign 6 runners to shoe type 1 and 6 to type 2. Do the same within the novice block.

Compare the times for the two shoe types within each block, then combine the comparisons.

Blocking removes the variability due to experience from the comparison, so a real difference between shoes is easier to detect.

EXAMPLE 4

A teacher wants to compare two practice apps. Each of 20 students will use both. Design a matched pairs experiment.

SOLUTION

Each student is their own pair, because each uses both apps.

For each student, flip a coin: heads means app 1 first, tails means app 2 first. This randomizes the order and protects against a practice effect.

Measure the quiz score after each app.

Analyze the **differences** (app 1 score minus app 2 score) for each student. Using the same student controls for ability.

EXAMPLE 5

A study of 2,000 adults finds that people who drink coffee daily have a lower rate of a certain illness. A news site says coffee prevents the illness. Is this justified? Identify a possible confounder.

SOLUTION

This is an observational study; people chose whether to drink coffee, so no cause-and-effect conclusion is justified.

A possible confounding variable: overall health habits. People who are active and eat well might also drink coffee more, and those habits could lower illness rates.

Because the effect of coffee cannot be separated from the effect of health habits, the data show an **association**, not a cause.

EXAMPLE 6

A researcher recruits 60 volunteers for a diet study. The volunteers are randomly assigned to two groups. One group follows Diet 1 and the other Diet 2. Diet 1 shows significantly more weight loss. (a) What conclusion about cause is justified? (b) To whom does it apply? (c) How many ways can the 60 volunteers be split into two groups of 30?

SOLUTION

(a) Random assignment was used, so differences are attributable to the diets: Diet 1 caused more weight loss in these volunteers.

(b) The volunteers were not randomly selected from a population, so the conclusion only extends to people similar to the volunteers, not to all adults.

(c) The number of splits is $\binom{60}{30}$, which is about 1.18×10^{17} , a very large number. Random assignment picks one of these equally likely splits.

Coach's Playbook

- 1 **Start with the verb.** Did the researchers impose a treatment? If yes, it is an experiment. If they only measured, it is observational.
- 2 **List the pieces.** Units, explanatory variable (treatments), response variable. Many answers lose points because one is missing.
- 3 **Write the random assignment procedure.** Say how you label, what random tool you use, and how groups are formed.
- 4 **Say the same response is measured in all groups under the same conditions.** This is control.
- 5 **Name the comparison.** For example, "compare the mean response for the groups."
- 6 **Finish with the justified conclusion.** Look at random selection and random assignment separately.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

When asked why a feature of a design is used, tie it to a purpose. Random assignment: to create groups similar in all respects before the treatment. Blinding: to avoid expectations changing the measurement or the behavior. Placebo: to account for the effect of receiving any treatment. Replication: to reduce the role of chance. Blocking: to reduce variability from a known variable.

Common traps

1. Writing "randomly select" when you mean "randomly assign." Selection chooses units; assignment places chosen units into treatments. 2. Calling a variable confounding without explaining how it relates to both the explanatory variable and the response. 3. Treating a block as a treatment. Blocks are groups of similar units, never randomly assigned. 4. Claiming that a bigger sample removes confounding in an observational study. It does not.

Mistake Clinic

Spot the error 1

A student writes: "To test a new fertilizer, I will put it on the plants in the sunny window and use no fertilizer on the plants in the dark corner. Then I will compare heights." Where is the error? The treatment is completely confounded with light. If the fertilized plants grow taller, we cannot tell whether the fertilizer or the sunlight caused it. A correct design randomly assigns plants to fertilizer or no fertilizer, with all plants in the same conditions (or blocks by location and randomizes inside each block).

Spot the error 2

A student writes: "In my experiment, the first 20 students who arrive get the new tutoring and the next 20 get the old tutoring. Since I used a rule, it is random." Where is the error? A fixed rule is not random assignment. Early arrivals may differ from late arrivals (more motivated, different schedules), which creates confounding. Correct: label the 40 students, randomly select 20 with a random number generator for the new tutoring, and use the old tutoring for the remaining 20.

Spot the error 3

A student writes: "We randomly selected 500 adults and asked about their exercise and stress levels. People who exercise have lower stress, so exercise reduces stress for all adults." Where is the error? Random selection supports generalizing to all adults, but there was no random assignment, so there is no cause-and-effect conclusion. The correct statement: there is an association between exercise and lower stress in adults, but this study cannot show that exercise causes the lower stress.

Real-World Connection

Drug trials use nearly every idea in this section: a randomized, double-blind, placebo-controlled design. Neither the patients nor the doctors measuring symptoms know who got the real drug, so hopes and expectations cannot change the results. Only then can a difference be linked to the drug itself.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

Challenge Corner

A farmer wants to test two irrigation methods on 40 plots. Twenty plots are on a hillside (drier soil) and 20 are in a valley (wetter soil). (a) If she uses a completely randomized design, what could go wrong? (b) Describe a better design. (c) In the better design, how many ways can the hillside plots be assigned? Solution: (a) By chance, many hillside plots could land in one method, confounding method with soil moisture, and the extra variability makes real differences harder to see. (b) Use a randomized block design with blocks hillside and valley; in each block, randomly assign 10 plots to method 1 and 10 to method 2. (c) $\binom{20}{10} = 184,756$ ways in each block.

Practice

PRACTICE Experiments vs Observational Studies

Show your work on paper. Use a calculator only when a problem says so, or if it is allowed at your level.

- Which of the following is an experiment?
 - Researchers survey 300 students about how long they study
 - Researchers compare test scores for students who chose to take a prep class and those who did not
 - Researchers randomly assign 60 students to study with music or in silence and compare quiz scores
 - Researchers record the heights of the plants in a garden
- What is the main purpose of a placebo in a drug trial?
 - To account for the effect of receiving a treatment, apart from the drug itself
 - To make sure every patient receives the same dose
 - To make the sample size larger
 - To allow the results to be generalized to all patients
- In a double-blind experiment,
 - the subjects are randomly assigned twice
 - there are two response variables
 - the experiment is repeated on two groups of different ages
 - neither the subjects nor the people measuring the response know who received which treatment
- A researcher compares two teaching methods on 40 students, 20 of whom are in grade 9 and 20 in grade 12. Grade level strongly affects scores. The best design is
 - to assign the grade 9 students to method 1 and the grade 12 students to method 2
 - to randomly assign methods within the grade 9 students and within the grade 12 students
 - to let students choose their method
 - to give only the grade 12 students a treatment
- A study finds that students who eat breakfast have higher test scores. A confounding variable would be
 - a variable such as family routine that is related to both breakfast and test scores
 - the number of students in the study
 - the difference between the two mean scores
 - the fact that the test was timed
- The most important reason to use random assignment in an experiment is to
 - make sure the sample represents the population
 - allow the subjects to know which group they are in
 - remove all variability from the response
 - make the groups similar before the treatments are applied

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

7. A random sample of 400 residents is surveyed on both their coffee habits and sleep. The data show an association. Which conclusion is justified?
 - (A) Coffee causes poor sleep in all residents
 - (B) Coffee causes poor sleep in the sample only
 - (C) The association applies to the residents, but cause and effect cannot be concluded
 - (D) Nothing can be concluded about the residents
8. Which is a matched pairs design?
 - (A) Each of 30 subjects tries both cereals, in random order, and rates each
 - (B) 30 subjects are randomly assigned to one of two cereals
 - (C) Subjects choose their favorite cereal
 - (D) Two different groups are compared on a survey
9. Twelve subjects are to be randomly assigned to two groups of 6. How many different assignments are possible?
 - (A) 12
 - (B) 66
 - (C) 924
 - (D) 1,728
10. Forty students volunteer for a study to compare two study methods, flashcards and rereading. Describe a completely randomized design with the response being a score on the same quiz.
11. A farmer has 30 tomato plants, 15 on the sunny side of a garden and 15 on the shady side. She wants to compare 3 fertilizers. Describe a randomized block design, and give the number of plants per treatment within each block.
12. A study finds that children who own more books have higher reading scores. Name a possible confounding variable, and explain how it could create this association without books being the cause.
13. A researcher uses a coin flip for each of 20 subjects (heads for the treatment group, tails for the control group). (a) What is the problem with unequal group sizes? (b) Find the probability of exactly 10 in each group, to three decimal places.
14. A random sample of 150 adults from a city was surveyed, and those who volunteered for a weekly walking program were compared with those who did not. The walkers report better moods. State what conclusion is and is not justified.
15. In an experiment, 24 subjects were randomly assigned to a new tutoring method or a standard method. The difference in mean scores is significant, which means it is unlikely to happen by chance if there is no difference. All 24 were volunteers from one school. What can be concluded about cause and effect, and about which population?

33.3 Confidence Intervals for Proportions

Build and interpret one-sample z intervals for a proportion, check the conditions, understand what controls the margin of error, and find the sample size needed for a target precision.

Why This Matters

A single sample proportion is almost never exactly the population proportion. If a poll of 200 students says 62 percent walk to school, the real percentage for the whole school is probably close to 62 percent, but how close? A **confidence interval** answers that question with a range of plausible values and an honest statement of how reliable the method is. Intervals appear in news polls, medical studies, quality control, and every later topic in inference.

Students get stuck in three places. They skip the conditions, which cost points on every free-response question. They interpret the confidence level wrongly (it is about the method, not the probability that a single interval is right). And they mix up the margin of error formula with the sample size formula. This section gives a four-step routine, a clean interpretation sentence, and a way to plan a sample size.

Core Ideas

A **point estimate** is a single number computed from the sample, such as $\hat{p} = x/n$. A **confidence interval** has the form

$$\text{point estimate} \pm \text{margin of error},$$

and the margin of error is a **critical value** times the **standard error** of the estimate. The standard error of $\hat{p}$ is $\sqrt{\hat{p}(1 - \hat{p})/n}$, our estimate of the spread of sample proportions.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

One-Sample z Interval for a Proportion

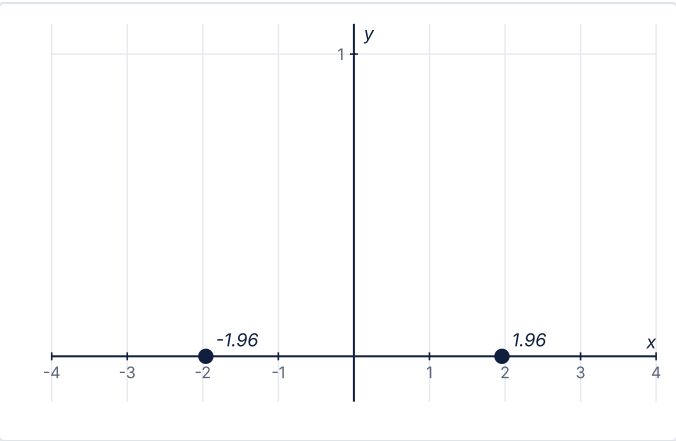
p-hat ± z* √(p-hat(1-p-hat)/n)

where z* is the critical value of the standard normal curve with area C between -z* and z*.

Conditions: (1) Random: the data come from a random sample or a randomized experiment. (2) 10%: if sampling without replacement, n ≤ 0.10N. (3) Large Counts: np ≥ 10 and n(1 - p-hat) ≥ 10.

Confidence level	90%	95%	99%
Critical value z*	1.645	1.960	2.576

The picture below shows where z* comes from for a 95% interval. The shaded region covers the middle 95% of the standard normal curve, and the boundaries are at ±1.96.



The middle 95% of a standard normal distribution lies between -1.96 and 1.96.

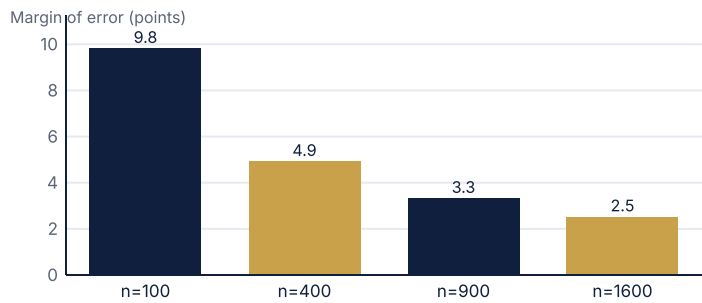
Interpreting a Confidence Interval

Interval: "We are C% confident that the interval from ___ to ___ captures the true proportion of [population and context]."

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 proportion."

The level describes the method. The true p is a fixed number; it is the interval that changes from sample to sample.

Three things control the width of an interval. A larger sample size makes it narrower. A higher confidence level makes it wider, because you need a larger z* to be more certain. A p-hat close to 0.5 makes it wider than a p-hat near 0 or 1. The margin of error shrinks with √n, so to cut it in half you must quadruple the sample size. The chart shows this for p-hat = 0.5 at 95% confidence.



Margin of error in percentage points for p-hat = 0.5 at 95% confidence. Each time n quadruples from 100 to 400 to 1600, the margin of error is cut in half.

Want a plan built around your student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

Sample Size for a Desired Margin of Error

To have margin of error at most m :

$$n \geq \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*).$$

Use a previous estimate for p^* if you have one. Otherwise use $p^* = 0.5$, which gives the largest (safest) n . Always **round up**.

Worked Examples

EXAMPLE 1

A random sample of 200 students from a school of 3,000 is surveyed, and 124 say they walk to school. Construct and interpret a 95% confidence interval for the proportion of all students at the school who walk.

SOLUTION

Parameter: p is the true proportion of all students at the school who walk.

Conditions: Random sample (given). 10%: $200 \leq 0.10(3000) = 300$. Large Counts: $200(0.62) = 124 \geq 10$ and $200(0.38) = 76 \geq 10$.

Calculate: $\hat{p} = 124/200 = 0.62$. $SE = \sqrt{0.62(0.38)/200} \approx 0.0343$. Margin of error = $1.96(0.0343) \approx 0.0673$. Interval: $0.62 \pm 0.0673 = (0.553, 0.687)$.

Conclude: We are 95% confident that the interval from 0.553 to 0.687 captures the true proportion of all students at the school who walk.

EXAMPLE 2

A news report says, "A 95% confidence interval for the proportion of voters who support a measure is 0.44 to 0.52." Which statements are correct? (a) There is a 95% chance that p is between 0.44 and 0.52. (b) About 95% of intervals built this way capture p . (c) 95% of voters are between 44% and 52%.

SOLUTION

(a) Incorrect as worded. The proportion p is fixed. After the interval is computed, it either contains p or it does not, so we cannot assign a 95% chance to this particular interval.

(b) **Correct.** The 95% describes the long-run success rate of the method.

(c) Incorrect. The interval is about the population proportion, not a statement about individual voters.

EXAMPLE 3

A confidence interval for a proportion is (0.44, 0.56). Find the sample proportion and the margin of error. If the confidence level is 95%, about how large was the sample?

SOLUTION

The center is the point estimate: $\hat{p} = \frac{0.44 + 0.56}{2} = 0.50$.

The margin of error is half the width: $m = \frac{0.56 - 0.44}{2} = 0.06$.

Solve $m = 1.96\sqrt{\frac{0.5(0.5)}{n}}$: $\sqrt{n} = \frac{1.96(0.5)}{0.06} \approx 16.33$, so $n \approx 266.8$. The sample size was about 267.

EXAMPLE 4

A city planner wants to estimate the proportion of residents who support a new bike lane to within 3 percentage points with 95% confidence. How many residents should be surveyed (a) with no prior information, (b) if an earlier study suggests $p \approx 0.30$?

SOLUTION

(a) Use $p^* = 0.5$: $n \geq \left(\frac{1.96}{0.03}\right)^2 (0.5)(0.5) \approx 1067.1$. Round up: **1,068** residents.

(b) Use $p^* = 0.30$: $n \geq \left(\frac{1.96}{0.03}\right)^2 (0.30)(0.70) \approx 896.4$. Round up: **897** residents.

Using the earlier information saves about 170 people because $0.3(0.7) = 0.21$ is smaller than 0.25.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

EXAMPLE 5

Using the data in Example 1 ($\hat{p} = 0.62$, $n = 200$), find the 90% and 99% intervals and compare their widths with the 95% interval.

SOLUTION

$SE \approx 0.0343$ for all three.

90%: $z^* = 1.645$, $m = 1.645(0.0343) \approx 0.0565$, interval (0.564, 0.676).

99%: $z^* = 2.576$, $m = 2.576(0.0343) \approx 0.0884$, interval (0.532, 0.708).

Widths: about 0.113 (90%), 0.135 (95%), 0.177 (99%). More confidence requires a wider interval: the price of greater certainty is less precision.

EXAMPLE 6

A random sample of 40 students shows 4 who own a particular kind of device. Can a z interval for the proportion be built? In a different case, a random sample of 200 from Example 1 gives the interval (0.553, 0.687). A club claims that exactly half of the students walk. What does the interval say?

SOLUTION

First case: $n\hat{p} = 40(0.1) = 4 < 10$, so the Large Counts condition fails. The z interval is not appropriate because the sampling distribution of $\hat{p}$ may be skewed.

Second case: The value 0.50 is **not** in (0.553, 0.687), so 0.50 is not a plausible value of p at the 95% confidence level.

The data provide convincing evidence that the true proportion is not 0.50. Because the whole interval is above 0.50, the evidence points toward more than half of the students walking.

Coach's Playbook

- 1 Write the parameter in words.** Include the population and the context, such as "the proportion of all students at the school who walk."
- 2 Check the three conditions with numbers.** Say "random", compute $10n$ versus N , and compute both $n\hat{p}$ and $n(1 - \hat{p})$.
- 3 Write the formula before you calculate.** This earns credit even if the arithmetic slips.
- 4 Compute $\hat{p}$, SE , then the margin of error.** Keep at least four decimals until the last step.
- 5 Interpret in context.** Use the sentence "We are C% confident that the interval from ___ to ___ captures the true proportion of ___."
- 6 Sanity check.** The center must be $\hat{p}$ and the interval must stay between 0 and 1.

A mental model: the interval is a net, and the confidence level is the fraction of nets that catch the fish over many throws. A wider net (higher confidence) catches more often but tells you less about where the fish is. A bigger sample gives a better, tighter net.

Common traps

1. Saying "there is a 95% probability that p is in the interval." The 95% describes the method. 2. Using $\hat{p}$ instead of $p^* = 0.5$ in the sample size formula when no earlier estimate exists, or forgetting to round up. 3. Checking Large Counts with np using a guessed p instead of the counts of successes and failures in the sample. 4. Forgetting the square root in the standard error, or writing the interval with the margin of error but not stating the endpoints.

Mistake Clinic

Spot the error 1

A student computes a 95% interval for $\hat{p} = 0.40$, $n = 500$ as $0.40 \pm 1.96\sqrt{0.40/500}$ and gets (0.345, 0.455). Where is the error? The student used $\hat{p}$ in place of $\hat{p}(1 - \hat{p})$ under the radical. The correct standard error is $\sqrt{0.40(0.60)/500} = 0.0219$. The margin of error is $1.96(0.0219) = 0.0429$, so the interval is (0.357, 0.443).

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

Spot the error 2

A student writes: "We are 95% confident that 95% of all students walk to school." Where is the error? The confidence level and the proportion are different things. The interval is about the proportion p , so the correct statement is: "We are 95% confident that the interval from 0.553 to 0.687 captures the true proportion of all students at the school who walk."

Spot the error 3

A student needs margin of error 0.04 at 95% confidence with no prior information and writes $n = \left(\frac{1.96}{0.04}\right)^2 = 2401$.

Where is the error? The factor $p^*(1 - p^*) = 0.25$ was left out. The correct size is $2401(0.25) = 600.25$, which rounds **up** to 601.

Real-World Connection

Election polls report "plus or minus 3 points." That is a margin of error at 95% confidence, and with $p^* = 0.5$ it requires about 1,068 respondents. This is why most national polls survey around a thousand people: more respondents cost a lot and shrink the margin only slowly.

Challenge Corner

A random sample of 80 people from a very large town gives $\hat{p} = 0.30$. Another researcher wants a margin of error half as large as the one from this sample at the same confidence level. How many people should be sampled, assuming $\hat{p}$ stays about 0.30? Solution: The margin of error is proportional to $1/\sqrt{n}$. Halving it requires $\sqrt{n}$ to double, so n must be quadrupled: $4(80) = 320$. Check: original $m = 1.96\sqrt{0.3(0.7)/80} \approx 0.1004$; new $m = 1.96\sqrt{0.3(0.7)/320} \approx 0.0502$, which is half.

Practice**PRACTICE Confidence Intervals for Proportions**

Show your work on paper. Use a calculator only when a problem says so, or if it is allowed at your level.

- A random sample of 45 gives $\hat{p} = 0.30$. Which statement about the Large Counts condition is correct?
 - It is not met because 45 is less than 100
 - It is met because $45(0.30) = 13.5$ and $45(0.70) = 31.5$ are both at least 10
 - It is not met because 13.5 is not a whole number
 - It cannot be checked without knowing the population proportion
- A 95% confidence interval for the proportion of students who walk is (0.55, 0.69). Which interpretation is correct?
 - There is a 95% probability that p is between 0.55 and 0.69
 - 95% of the students in the sample walk
 - 95% of all sample proportions lie between 0.55 and 0.69
 - We are 95% confident that the interval from 0.55 to 0.69 captures the true proportion of all students who walk
- What is the critical value z^* for a 90% confidence interval for a proportion?
 - 1.645
 - 1.960
 - 1.282
 - 2.576
- A 95% confidence interval for a proportion is computed with a sample of 250. If the researcher instead uses a sample of 1,000 with the same $\hat{p}$, the margin of error
 - doubles
 - stays the same
 - is cut in half
 - is cut to one fourth

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

5. A random sample of 500 voters from a large city shows 200 who support a measure. What is a 95% confidence interval for the proportion that supports it?
 (A) (0.378, 0.422) (B) (0.357, 0.443) (C) (0.314, 0.486)
 (D) (0.345, 0.455)
6. Which change alone would make a confidence interval for a proportion wider?
 (A) Increasing the sample size
 (B) Decreasing the confidence level from 95% to 90%
 (C) Using a sample with $\hat{p}$ much closer to 0
 (D) Increasing the confidence level from 95% to 99%
7. A researcher wants a 95% margin of error of no more than 0.04 for a proportion, with no prior information. What is the smallest sample size?
 (A) 600 (B) 2,401 (C) 601 (D) 1,201
8. A 90% confidence interval for a proportion is (0.32, 0.48). What are $\hat{p}$ and the margin of error?
 (A) $\hat{p} = 0.40$, margin of error 0.08 (B) $\hat{p} = 0.40$, margin of error 0.16
 (C) $\hat{p} = 0.48$, margin of error 0.08 (D) $\hat{p} = 0.32$, margin of error 0.16
9. A random sample of 100 members is taken from a club with 700 members, and $\hat{p} = 0.50$. Which condition for the z interval fails?
 (A) The random condition
 (B) The 10% condition, since 700 is less than 10(100)
 (C) The Large Counts condition
 (D) None of the conditions fail
10. A random sample of 150 households from a town of 5,000 shows that 36 recycle every week. Construct and interpret a 90% confidence interval for the proportion of households in the town that recycle every week. Include the conditions.
11. A 95% interval for a proportion of voters supporting a measure is (0.44, 0.52). Someone claims that exactly 40% of voters support it. Does the interval support or refute the claim? Explain.
12. Find the smallest sample size needed to estimate a proportion within 2 percentage points with 99% confidence when nothing is known about the proportion.
13. A random sample of 30 students shows that 27 prefer online homework. Explain why a z interval for the population proportion should not be used.
14. A poll of 385 randomly selected adults has $\hat{p} = 0.50$. Find the 95% margin of error. If the poll had used 1,540 adults, what would the margin of error be, and what is the relationship between the two?
15. A random sample of 1,000 voters from a large state shows 540 who support a candidate. (a) Construct a 95% confidence interval. (b) Does the interval give convincing evidence that more than half of the voters support the candidate? Explain, and describe what you cannot conclude.

33.4 Confidence Intervals for Means

Construct and interpret one-sample t intervals for a mean, check the Normal and Large Sample condition, handle paired data, and avoid the wording mistakes that cost points on free-response questions.

Why This Matters

Most real questions are about averages: the mean battery life, the mean time to finish a task, the mean change in a score. When we estimate a population mean μ from a sample, we do not know the population standard deviation σ , so we use the sample standard deviation s instead. That small change forces a new family of distributions, the **t distributions**, and it is the reason the procedure in this section is used far more often than a z interval for a mean.

Students struggle with four things: choosing t^* with the right degrees of freedom, checking the Normal/Large Sample condition (it depends on n and on the graph), distinguishing a statement about the *mean* from a statement about *individuals*, and wording the conclusion. A very common mistake is writing that 95% of the data lie in the interval. This section fixes all of these.

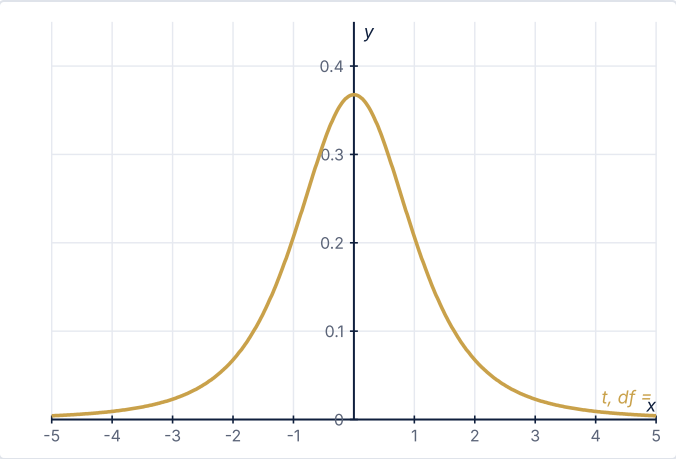
Core Ideas

When we replace σ by s , the standardized statistic $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ no longer follows a normal distribution. It follows a **t distribution with $n - 1$ degrees of freedom**. A t distribution is bell shaped and centered at 0, like the standard normal, but it has **heavier tails** (more probability far from the center) because s varies from sample to sample. As the degrees of freedom increase, the t distribution gets closer to the standard normal curve.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →



A standard normal curve (navy) and a t curve with 3 degrees of freedom (gold). The t curve is lower in the middle and has heavier tails.

One-Sample t Interval for a Mean

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where t^* is the critical value of the t distribution with $df = n - 1$ that has area C between $-t^*$ and t^* . The quantity $s/\sqrt{n}$ is the **standard error** of $\bar{x}$.

Conditions: (1) **Random:** random sample or randomized experiment. (2) **10%:** $n \leq 0.10N$ when sampling without replacement. (3) **Normal/Large Sample:** population approximately normal, or $n \geq 30$.

df	90% t^*	95% t^*	99% t^*
4	2.132	2.776	4.604
9	1.833	2.262	3.250
14	1.761	2.145	2.977
24	1.711	2.064	2.797
29	1.699	2.045	2.756
Normal (z^*)	1.645	1.960	2.576

Checking the Normal/Large Sample Condition

$n \geq 30$: the Central Limit Theorem makes the sampling distribution of $\bar{x}$ approximately normal, even for skewed data (very extreme outliers still deserve a comment).

$15 \leq n < 30$: the sample data should show no strong skewness and no outliers.

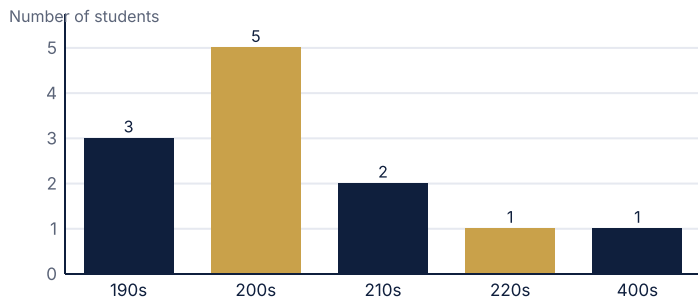
$n < 15$: the sample data should be roughly symmetric with no outliers. Always **graph the data** (dotplot, histogram, boxplot) and say what you see.

Interpretation uses the same structure as before, but about a **mean**: "We are C% confident that the interval from ___ to ___ captures the true mean [variable] for [population]." The confidence interval is **not** a statement about individual values. Individual data vary far more than the mean does. A bar chart of a small sample makes the Normal condition concrete. In the sample below, one reaction time is far from the rest.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →



Reaction times in milliseconds for 12 students. Eleven values lie between 195 and 220, but one value is 410.

Three factors change the width of a t interval: a larger sample size makes it narrower (both through $\sqrt{n}$ and through a smaller t^*), a larger s makes it wider, and a higher confidence level makes it wider. For **paired data** (two measurements on the same person or matched pairs), compute the difference for each pair and then apply the one-sample t interval to the differences.

Worked Examples

EXAMPLE 1

A random sample of 25 batteries from a large shipment has a mean life of 41.2 hours with standard deviation 5.5 hours, and a dotplot is roughly symmetric with no outliers. Construct and interpret a 95% confidence interval for the mean life of all batteries in the shipment.

SOLUTION

Parameter: μ is the true mean life of all batteries in the shipment.

Conditions: Random sample (given). 10%: the shipment is large, so 25 is far less than 10% of it. Normal/Large Sample: $n = 25 < 30$, but the dotplot is roughly symmetric with no outliers.

Calculate: $df = 24$, $t^* = 2.064$. $SE = 5.5/\sqrt{25} = 1.1$. Margin of error $= 2.064(1.1) = 2.270$. Interval: $41.2 \pm 2.270 = (38.93, 43.47)$.

Conclude: We are 95% confident that the interval from 38.93 to 43.47 hours captures the true mean life of all batteries in the shipment.

EXAMPLE 2

Find the critical value t^* for each situation and compare with z^* . (a) 90% confidence, $n = 12$. (b) 95% confidence, $n = 16$. (c) 99% confidence, $n = 30$.

SOLUTION

Use $df = n - 1$.

(a) $df = 11$: $t^* = 1.796$, compared with $z^* = 1.645$.

(b) $df = 15$: $t^* = 2.131$, compared with $z^* = 1.960$.

(c) $df = 29$: $t^* = 2.756$, compared with $z^* = 2.576$.

The t critical values are always larger than the z critical values, and the gap shrinks as n grows. This makes t intervals wider to account for the extra uncertainty from estimating σ with s .

EXAMPLE 3

Twelve reaction times (milliseconds) are 212, 198, 205, 220, 201, 195, 208, 215, 203, 199, 207, 410. A researcher wants a t interval for the mean reaction time of the class. Is that appropriate?

SOLUTION

The sample is small ($n = 12 < 15$), so we must check that the data are roughly symmetric with no outliers.

The value 410 is far above the others (the next largest is 220), and the graph shows it standing alone. The mean is 222.75, which is pulled up by the outlier, and $s \approx 59.4$ is inflated by it too.

Conclusion: the Normal/Large Sample condition is **not met**, so a t interval is not appropriate unless the outlier is investigated (for example, a recording error) and the situation is resolved.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

EXAMPLE 4

A 95% confidence interval for a mean, based on a random sample of 16 observations, is (38.4, 44.0). Find $\bar{x}$, the margin of error, and the sample standard deviation s .

SOLUTION

$$\bar{x} = \frac{38.4 + 44.0}{2} = 41.2.$$

$$\text{Margin of error} = \frac{44.0 - 38.4}{2} = 2.8.$$

$$df = 15, \text{ so } t^* = 2.131. \text{ From } 2.8 = 2.131 \cdot \frac{s}{\sqrt{16}} \text{ we get } s = \frac{2.8(4)}{2.131} \approx 5.26.$$

EXAMPLE 5

Eight students took a vocabulary quiz before and after a review session. The mean of the differences (after minus before) was 3.4 points with standard deviation 2.1 points, and a graph of the differences showed no strong skew or outliers. Construct a 95% confidence interval for the mean improvement of all students like these, assuming these students were randomly chosen.

SOLUTION

These are paired data, so we work with the **differences**. Parameter: μ_d is the true mean improvement.

Conditions: random (given); the differences show no strong skew or outliers, which suffices for $n = 8$.

$$df = 7, t^* = 2.365. SE = 2.1/\sqrt{8} \approx 0.7425. m = 2.365(0.7425) \approx 1.756.$$

Interval: $3.4 \pm 1.756 = (1.64, 5.16)$. We are 95% confident that the interval from 1.64 to 5.16 points captures the true mean improvement in quiz score for all such students. Since the whole interval is above 0, the review appears helpful.

EXAMPLE 6

Using the battery data of Example 1, a manufacturer claims that the mean life is 45 hours. (a) What does the interval say about the claim? (b) If a sample of 100 had given the same $\bar{x}$ and s , what would the 95% interval be?

SOLUTION

(a) The interval (38.93, 43.47) does not contain 45, so 45 is not a plausible value for μ at the 95% level. The data suggest the true mean life is less than 45 hours.

(b) $df = 99, t^* = 1.984. SE = 5.5/\sqrt{100} = 0.55. m = 1.984(0.55) \approx 1.091$. Interval: (40.11, 42.29).

The larger sample makes the interval about half as wide (from width 4.54 to 2.18), a more precise estimate of the same quantity.

Coch's Playbook

- 1 **Name the parameter.** " μ is the true mean ___ of all ___." Include units and the population.
- 2 **Write the conditions with evidence.** For Normal/Large Sample, state n and describe the graph if $n < 30$.
- 3 **Find $df = n - 1$ and t^* before computing.** Writing df shows you used a t procedure.
- 4 **Calculate $SE = s/\sqrt{n}$, then $m = t^* \cdot SE$.** Keep three or more decimals in SE .
- 5 **State the interval and interpret it.** Use "mean", the units, and the population.
- 6 **Check the width.** The center is $\bar{x}$. For a sample of 5 or more, the margin of error should be smaller than s ; if it is not, recheck your arithmetic.

Mental model: the standard error is the typical distance from $\bar{x}$ to μ . The margin of error stretches this by t^* , a slightly larger number than z^* when the sample is small, to pay for estimating the spread. When you must pick an interpretation sentence, prefer one that includes the word **mean** and a population 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 →

Common traps

1. Saying "95% of the batteries last between 38.93 and 43.47 hours." The interval is for the mean, not for individual batteries. 2. Using $df = n$ instead of $n - 1$, or using z^* because the sample seems big. 3. Skipping the graph for small samples, or writing "normal" without evidence. 4. Treating paired data as two independent samples instead of working with the differences.

Mistake Clinic**Spot the error 1**

A student writes: "A 90% confidence interval for the mean time to finish a puzzle is (32.1, 36.9) minutes, so 90% of students finish between 32.1 and 36.9 minutes." Where is the error? The interval estimates the **mean** time for the population, not the spread of individual times. Individual times vary much more. The correct statement: "We are 90% confident that the interval from 32.1 to 36.9 minutes captures the true mean time for all students to finish the puzzle."

Spot the error 2

A student computes a 95% interval from a sample of 16 with $\bar{x} = 100$ and $s = 10$ as $100 \pm 1.96 \cdot \frac{10}{\sqrt{16}} = (95.1, 104.9)$.

Where is the error? The population standard deviation is unknown (we only have s), so a t critical value is required.

With $df = 15$, $t^* = 2.131$, and the margin of error is $2.131(2.5) = 5.33$. The correct interval is (94.67, 105.33), which is wider than the student's.

Spot the error 3

A student has 10 students' before and after scores and writes: "Because we have two groups, we use a two-sample interval, and the conditions are met because each group has 10 students." Where is the error? The scores come from the **same** students, so they are paired. The correct approach is to find each student's difference and construct a one-sample t interval for the mean difference, checking that the differences show no strong skew or outliers.

Real-World Connection

Engineers testing a new type of phone battery cannot drain every battery produced. They test a random sample of 20 and construct a t interval for the mean life. If the lower end of the interval is above the advertised minimum, the company can claim the advertised life with evidence behind it.

Challenge Corner

A random sample has $s = 12$. How large must the sample be so that the 95% t interval has a margin of error of at most 4.0? Start with $z^* = 1.96$ to get a first guess, then refine with t^* . Solution: With z^* : $1.96(12)/\sqrt{n} \leq 4$ gives $\sqrt{n} \geq 5.88$, so $n \geq 34.6$, about 35. Now check with t^* . For $n = 35$ ($df = 34$, $t^* = 2.032$): $m = 2.032(12)/\sqrt{35} = 4.12 > 4$, too big. For $n = 37$ ($df = 36$, $t^* = 2.028$): $m = 4.001$, still just above 4. For $n = 38$ ($df = 37$, $t^* = 2.026$): $m = 3.944 \leq 4$. So $n = 38$ is the smallest size. The t value makes the required n larger than the z estimate.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

Practice

PRACTICE Confidence Intervals for Means

Show your work on paper. Use a calculator only when a problem says so, or if it is allowed at your level.

- What is the critical value t^* for a 95% confidence interval for a mean based on a sample of size 20?
 (A) 2.086 (B) 1.960 (C) 2.093 (D) 1.729
- A researcher has a random sample of 10 measurements and wants a t interval for the mean. Which check is needed for the Normal/Large Sample condition?
 (A) A graph of the data showing roughly symmetric shape with no outliers
 (B) Nothing, because n is large enough by the Central Limit Theorem
 (C) That the population standard deviation is known
 (D) That the sample is at least 10% of the population
- Compared with a standard normal curve, a t distribution with 5 degrees of freedom
 (A) has a smaller center and lighter tails
 (B) is skewed to the right
 (C) has the same shape for every sample size
 (D) has heavier tails and approaches the normal curve as df increases
- A random sample of 36 values has $\bar{x} = 72$ and $s = 9$. Which is a 95% confidence interval for the population mean?
 (A) (69.06, 74.94) (B) (68.96, 75.05) (C) (63.0, 81.0)
 (D) (70.5, 73.5)
- A 95% interval for the mean number of hours of sleep of teenagers is (7.2, 7.8). Which interpretation is correct?
 (A) 95% of teenagers sleep between 7.2 and 7.8 hours
 (B) There is a 95% chance that the sample mean is between 7.2 and 7.8
 (C) The mean is 7.5 with 95% probability
 (D) We are 95% confident that the interval from 7.2 to 7.8 captures the true mean hours of sleep of teenagers
- Which change would make a t interval for a mean narrower, if everything else stays the same?
 (A) Using a higher confidence level (B) A sample with a larger standard deviation
 (C) Using a larger sample size (D) Using $df = n$
- Ten students are timed on a task before and after practice. Which procedure is appropriate for estimating the mean improvement?
 (A) A one-sample t interval on the 10 differences
 (B) A two-sample t interval comparing before and after
 (C) A one-proportion z interval
 (D) A z interval with σ from the before times
- A random sample of 18 values is used to build a t interval. How many degrees of freedom are used?
 (A) 18 (B) 17 (C) 16 (D) 19
- A 95% t interval for a mean from a random sample of 16 values is (12.4, 15.6). What is the sample standard deviation s ?
 (A) 6.0 (B) 1.6 (C) 3.0 (D) 0.75
- A random sample of 20 students has a mean time to finish a task of 34.5 seconds with standard deviation 6.2 seconds, and a boxplot shows no outliers roughly symmetric. Construct and interpret a 90% confidence interval for the mean time of all students (population is large).
- A 95% t interval for the mean number of hours that teenagers spend on homework per week is (3.1, 3.9). Write a correct interpretation, then explain why the statement "95% of teenagers do between 3.1 and 3.9 hours" is wrong.
- Eight values are 14, 15, 15, 16, 17, 17, 18, and 55. Is a t interval for the mean appropriate? Explain.
- Six matched pairs have differences (after minus before) of 4, -1, 3, 5, 2, 3. Assuming the conditions are met, construct a 95% confidence interval for the mean difference.
- Two random samples both have $\bar{x} = 100$ and $s = 10$, with $n = 16$ and $n = 64$. Find the 95% interval for each, and explain what happens to the width and why.
- A bottler claims that bottles contain 500 mL on average. A random sample of 30 bottles has mean 497.8 mL and standard deviation 4.5 mL, with a graph showing no strong skew. Construct a 95% confidence interval and say what it tells you about the claim.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

Answers**Answers and Explanations**

Complete solutions for every practice problem, numbered to match. Read the explanation even when your answer is right.

33.1 Sampling Methods and Bias**PRACTICE · SAMPLING METHODS AND BIAS**

1. C. An SRS means every group of 10 is equally likely. A is stratified, B is systematic, and D is voluntary response. B looks random but only allows a few possible samples.
2. B. Whole days (groups) are chosen at random and every visitor in them is surveyed, which is a cluster sample. Stratified would sample from every day.
3. A. People without a listed home phone have no chance of being chosen, which is undercoverage. The other issues are not described.
4. C. The word "excellent" pushes respondents toward agreeing. The other questions are neutral and factual.
5. B. The number of samples is $\binom{10}{3} = \frac{10 \cdot 9 \cdot 8}{6} = 120$. The choice 720 counts ordered selections, which is the tempting error.
6. A. $k = 450/30 = 15$, so the sample is $4 + 15m$. For $m = 8$ this gives 124. The others are not 4 plus a multiple of 15 (120 is not, since $120 - 4 = 116$).
7. B. Stratifying works when groups are different from each other but alike inside, so each group is sampled separately. A describes cluster sampling.
8. C. A bigger sample reduces random variation but cannot fix a flawed method. Voluntary response stays biased and does not become an SRS.
9. B. The share is $450/1500 = 0.30$, and $0.30 \cdot 60 = 18$. Choice 15 uses $450/30$, a wrong divisor.
10. Systematic sample. Only 6 different samples are possible (one for each start from 1 to 6), so most groups of 10 swimmers, such as swimmers 1 to 10, cannot occur. It is not an SRS, although each swimmer has a 1 in 6 chance.
11. Voluntary response bias. Students with strong feelings are more likely to answer, so the poll likely overstates the proportion who feel strongly (most likely students who want less homework).
12. Pairs: 63, 17, 90, 41, 17, 08, 25, 52, 36, 44. Skip 63, 90, and 52 (too large) and the repeated 17. Answer: 17, 41, 08, 25, 36, 44.
13. District A: 54, District B: 42, District C: 24. Reason: each share times 120 gives $1080/20 = 54$, $840/20 = 42$, $480/20 = 24$ (since $120/2400 = 1/20$). Total 120.
14. (a) $3/40 = 0.075$. (b) Groups of 3 buildings are the only possible samples, so units in different buildings cannot be sampled in arbitrary combinations. Groups of units that span many buildings have no chance of being chosen.
15. (a) $120/500 = 0.24$, or 24 percent. (b) Nonrespondents who favor it: $0.40 \cdot 380 = 152$. True count = $90 + 152 = 242$, so the proportion is $242/500 = 0.484$. The respondents' 75 percent is too high because nonrespondents differ from respondents (nonresponse bias).

33.2 Experiments vs Observational Studies**PRACTICE · EXPERIMENTS VS OBSERVATIONAL STUDIES**

1. C. Only choice C imposes a treatment (music or silence) using random assignment. A and D only measure, and B lets students choose their group, so it is observational.
2. A. A placebo lets researchers separate the effect of the drug from the effect of simply receiving something. It does not set the dose (B) or allow generalizing (D).
3. D. Double-blind means both the subjects and the people evaluating responses do not know the assignments. This avoids expectations influencing results.
4. B. Grade level is a known variable, so make blocks and randomize within each block. A confounds the method with grade level.
5. A. A confounder is related to both the explanatory and response variables, such as a family routine that supports both breakfast and good study habits. The other choices do not relate the variables.
6. D. Random assignment balances all variables (known or unknown) across groups before treatments. Representing a population (A) comes from random selection.
7. C. Random selection lets the association be generalized to the residents, but without random assignment cause and effect cannot be concluded.
8. A. Each subject receives both treatments in random order, which is the definition of a matched pairs design. B is completely randomized.
9. C. The number of assignments is $\binom{12}{6} = 924$. The tempting choice 66 is $\binom{12}{2}$, and $1,728$ is 12^3 .
10. Label the 40 students 01 to 40. Use a random number generator to select 20 distinct labels for flashcards; the other 20 do rereading. Give both groups the same amount of time and the same quiz. Compare the mean quiz scores for the two groups. Random assignment makes the groups comparable, so a large difference can be attributed to the method.
11. Blocks: sunny side (15 plants) and shady side (15 plants). Within each block, randomly assign 5 plants to each of the three fertilizers, giving 5 plants per treatment per block (10 plants per fertilizer in total). Measure growth after the same time period and compare fertilizers within blocks.
12. A possible confounder: household income or parents' education. Families with more resources may buy more books and also provide more reading support, which raises reading scores. So the association could exist without books causing higher scores.
13. (a) Unequal groups (such as 7 and 13) reduce the information in the comparison, and the study loses efficiency. (b) $P = \binom{20}{10}/2^{20} = 184756/1048576 \approx 0.176$. A better method assigns exactly 10 to each group.
14. The study is observational because the volunteers chose their group, and it is not an experiment. It can be concluded that walkers in the sample report better moods (an association), and, because of random selection, this may apply to the city's adults. It cannot be concluded that walking causes better moods; a confounder such as general health could explain 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 →

15. Cause and effect: because subjects were randomly assigned, the new method caused the higher mean scores for these subjects. Population: the subjects were volunteers from one school, not randomly selected, so the conclusion is limited to students similar to them and not to all students.

33.3 Confidence Intervals for Proportions

PRACTICE · CONFIDENCE INTERVALS FOR PROPORTIONS

1. B. Large Counts needs expected successes and failures of at least 10: $45(0.30) = 13.5$ and $45(0.70) = 31.5$. The condition does not need whole numbers or a sample of 100.
2. D. The confidence level describes the method, and the interval estimates the population proportion in context. The first option treats p as random, which is the tempting error.
3. A. For 90% confidence there is 5% in each tail, so $z^* = 1.645$. The value 1.960 is for 95% and 2.576 is for 99%.
4. C. Margin of error is proportional to $1/\sqrt{n}$. Quadrupling n from 250 to 1,000 makes it $1/\sqrt{4} = 1/2$ as large, not one fourth.
5. B. $\hat{p} = 0.40$, $SE = \sqrt{0.4(0.6)/500} = 0.0219$, $m = 1.96(0.0219) = 0.0429$, so $(0.357, 0.443)$. The interval $(0.378, 0.422)$ forgets z^* , and $(0.345, 0.455)$ drops the factor $(1 - \hat{p})$.
6. D. A higher confidence level needs a larger z^* (2.576 versus 1.960), so the interval is wider. A larger sample or a lower confidence level makes it narrower.
7. C. $n \geq (1.96/0.04)^2(0.25) = 600.25$, which must be rounded up to 601. The value 2,401 leaves out the factor 0.25, and 600 rounds down.
8. A. The center is $(0.32 + 0.48)/2 = 0.40$ and the margin of error is half the width: $(0.48 - 0.32)/2 = 0.08$. The width 0.16 is not the margin of error.
9. B. For the 10% condition the population must be at least $10n = 1000$, but the club has only 700. Large Counts is satisfied since $100(0.5) = 50$.
10. Parameter: p , the proportion of all households in the town that recycle weekly. Conditions: random sample (given); $150 \leq 0.10(5000) = 500$; $150(0.24) = 36 \geq 10$ and $150(0.76) = 114 \geq 10$. $\hat{p} = 0.24$, $SE = \sqrt{0.24(0.76)/150} = 0.0349$, $m = 1.645(0.0349) = 0.0574$. Interval $(0.183, 0.297)$. We are 90% confident that the interval from 0.183 to 0.297 captures the true proportion of all households in the town that recycle weekly.
11. The claim is refuted: 0.40 is not in $(0.44, 0.52)$, so 0.40 is not a plausible value for the proportion of voters who support the measure at the 95% confidence level. The data suggest that the true proportion is above 0.40.
12. $n \geq (2.576/0.02)^2(0.25) = 4147.36$, so the smallest size is 4,148.
13. Large Counts fails: $\hat{p} = 27/30 = 0.90$, so $n(1 - \hat{p}) = 30(0.10) = 3 < 10$. The sampling distribution of $\hat{p}$ is not approximately normal, so the z interval's formula is not reliable.
14. For 385: $m = 1.96\sqrt{0.5(0.5)/385} \approx 0.0499$. For 1,540: $m = 1.96\sqrt{0.25/1540} \approx 0.0250$. Since $1540 = 4(385)$, the margin of error is half as large.
15. (a) $\hat{p} = 0.54$, $SE = \sqrt{0.54(0.46)/1000} = 0.01576$, $m = 1.96(0.01576) = 0.0309$. Interval $(0.509, 0.571)$. (b) Yes: every plausible value is above 0.50, so there is convincing evidence that more than half support the candidate. It does not give the exact support level, and the 95% level does not mean there is a 95% probability for this single interval.

33.4 Confidence Intervals for Means

PRACTICE · CONFIDENCE INTERVALS FOR MEANS

1. C. $df = 19$, so $t^* = 2.093$. The value 2.086 is for $df = 20$ (using n instead of $n - 1$), and 1.960 is the z value.
2. A. For small samples the data must be roughly symmetric with no outliers; this is checked with a graph. The 10% condition is about the population size, and the CLT needs $n \geq 30$.
3. D. t distributions have heavier tails than the normal curve and move closer to it as df increases. They are symmetric and their shape depends on df .
4. B. $df = 35$, $t^* = 2.030$, $SE = 9/6 = 1.5$, $m = 3.045$, giving $(68.96, 75.05)$. The interval $(69.06, 74.94)$ uses z^* , and $(70.5, 73.5)$ uses only the standard error.
5. D. The interval is about the population mean, and the confidence level describes the method. The other choices make claims about individuals or assign a probability to a fixed number.
6. C. A larger sample reduces the standard error $s/\sqrt{n}$ and t^* , so the interval narrows. A higher confidence level or a larger s widens it.
7. A. The measurements come from the same students, so they are paired: find the 10 differences and use a one-sample t interval on them.
8. B. $df = n - 1 = 17$. The value 18 is the tempting mistake of using n .
9. C. $m = (15.6 - 12.4)/2 = 1.6$ and $t^* = 2.131$ (for $df = 15$). Then $s = m\sqrt{n}/t^* = 1.6(4)/2.131 \approx 3.0$. The value 6.0 comes from using the full width 3.2 instead of the margin of error 1.6.
10. Parameter: μ , the true mean time for all students. Conditions: random (given), population large so 10% holds, boxplot shows no outliers or strong skew so Normal/Large Sample is met for $n = 20$. $df = 19$, $t^* = 1.729$, $SE = 6.2/\sqrt{20} = 1.386$, $m = 2.397$. Interval $(32.10, 36.90)$ seconds. We are 90% confident that the interval from 32.10 to 36.90 seconds captures the true mean time for all students.
11. Interpretation: We are 95% confident that the interval from 3.1 to 3.9 hours captures the true mean number of homework hours per week for all teenagers. The other statement is wrong because the interval is about the mean, and individual values vary much more than the mean does; many teenagers do less than 3.1 or more than 3.9 hours.
12. No. The sample is small ($n = 8 < 15$), and 55 is a strong outlier (the other values are 14 to 18). The mean of 20.9 and the standard deviation of 13.9 are both distorted, so the Normal/Large Sample condition is not met.
13. $\bar{x}_d = 16/6 \approx 2.667$, $s_d \approx 2.066$, $df = 5$, $t^* = 2.571$. $SE = 2.066/\sqrt{6} = 0.843$, $m = 2.571(0.843) = 2.168$. Interval $(0.50, 4.83)$. The mean difference is estimated to be positive because the whole interval is above 0.
14. $n = 16$: $df = 15$, $t^* = 2.131$, $m = 2.131(10/4) = 5.33$, interval $(94.67, 105.33)$, width 10.66. $n = 64$: $df = 63$, $t^* = 1.998$, $m = 1.998(10/8) = 2.50$, interval $(97.50, 102.50)$, width 5.00. The width is less than half because $s/\sqrt{n}$ is halved (quadrupling n) and t^* also decreases.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →

15. $df = 29$, $t^* = 2.045$, $SE = 4.5/\sqrt{30} = 0.8216$, $m = 1.680$. Interval (496.12, 499.48) mL. The value 500 is not in the interval, so 500 is not plausible at the 95% level. The data suggest that the true mean is less than 500 mL.

Want a plan built around **your** student?

Book a free consultation. The short intake takes about a minute and opens right away.

BOOK NOW →