



Hypothesis Testing and Chi-Square Tests

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

In this booster

34.1 The Logic of Significance Tests

34.2 Tests for Proportions and Means

34.3 Comparing Two Groups

34.4 Chi-Square Tests and Regression Inference

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.

34.1 The Logic of Significance Tests

Learn how a significance test works: writing hypotheses, interpreting p-values, making decisions with a significance level, and understanding Type I errors, Type II errors, and power.

Why This Matters

A significance test answers a yes or no style question about data: is what we saw in the sample just the kind of thing that chance produces, or is it strong evidence that a claim about the population is false? Drug approvals, quality control, A/B tests in technology, and every research paper with a "p-value" rely on this logic. Learning it well makes the formulas of the next sections easy.

The logic is the reason students get stuck. They write hypotheses with sample statistics instead of parameters, misinterpret the p-value as "the probability the null is true," say they "accept" the null hypothesis, and mix up Type I and Type II errors. The routine in this section prevents each of these mistakes.

Core Ideas

A **significance test** starts with a claim about a population **parameter** and asks whether the sample data give convincing evidence against it. The claim being tested is the **null hypothesis** H_0 , and what we suspect instead is the **alternative hypothesis** H_a .

Hypotheses

H_0 states that the parameter equals a specific value (for example, $H_0 : p = 0.50$ or $H_0 : \mu = 12$). H_a says the parameter is greater than ($>$), less than ($<$), or different from ($\neq$) that value.

Both hypotheses are about **parameters** (p , μ), never about statistics like $\hat{p}$ or $\bar{x}$. Define the parameter in words with context.

P-value

The **p-value** is the probability, **assuming** H_0 is true, of getting a sample result at least as extreme (in the direction of H_a) as the one observed.

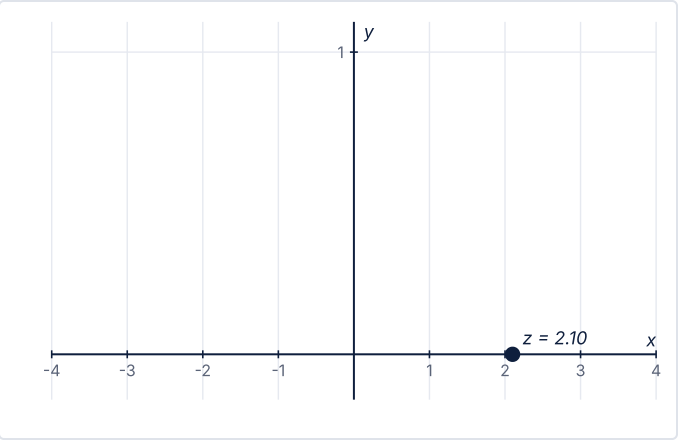
A small p-value means the result would be unlikely if H_0 were true, so it is evidence against H_0 . Compare the p-value with the **significance level** α : if $\text{p-value} \leq \alpha$, reject H_0 ; otherwise fail to reject H_0 .

The next figure shows a p-value for a right-tailed test. If the standardized test statistic is $z = 2.10$, the p-value is the area under the standard normal curve to the right of 2.10, which is about 0.0179.

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

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

BOOK NOW →



The shaded area to the right of $z = 2.10$ is the p-value, about 0.0179, for a right-tailed test.

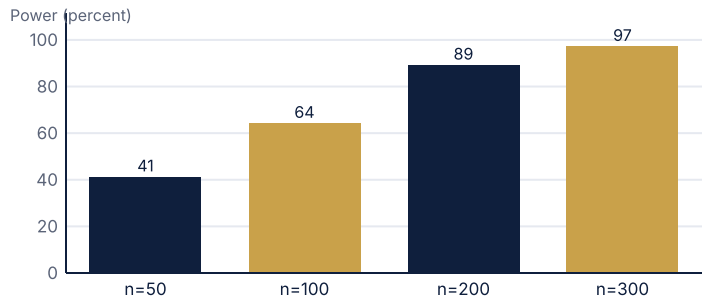
For a two-sided alternative ($\neq$) the p-value counts both tails, so it is double the one-tail area: $2(0.0179) \approx 0.0357$. Your conclusion has two parts. First the decision, linking the p-value to α . Then the statement in context: "Because the p-value of $___$ is less than $\alpha = ___$, we reject H_0 . There is convincing evidence that [H_a in context]." If the p-value is large: "we fail to reject H_0 . There is not convincing evidence that [H_a in context]."

Because a decision is based on a sample, it can be wrong in two ways. The table lists them. A **Type I error** is rejecting a true H_0 , and its probability is exactly α . A **Type II error** is failing to reject a false H_0 , with probability β .

	H_0 is true	H_0 is false
Reject H_0	Type I error (probability α)	Correct decision (probability = power)
Fail to reject H_0	Correct decision	Type II error (probability β)

Power

The **power** of a test against a specific alternative is the probability of rejecting H_0 when that alternative is true: $\text{power} = 1 - \beta$. Power **increases** when the sample size increases, when α increases, when the true parameter is farther from the null value, and when variability decreases.



Power of a right-tailed test of $H_0: p = 0.5$ at $\alpha = 0.05$ when the true p is 0.6. Larger samples make the test more likely to detect the difference.

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 →

EXAMPLE 1

Write the hypotheses and define the parameter. (a) A school claims that 30% of its students ride the bus; a student thinks it is less. (b) A machine should fill bottles to a mean of 12 ounces; an inspector checks whether the mean has changed. (c) A coach claims her players make at least 60% of free throws; a reporter doubts that this is true.

SOLUTION

(a) p = the true proportion of all students at the school who ride the bus. $H_0 : p = 0.30$, $H_a : p < 0.30$.

(b) μ = the true mean fill amount (in ounces) for all bottles from the machine. $H_0 : \mu = 12$, $H_a : \mu \neq 12$ ("has changed" means either direction).

(c) p = the true proportion of free throws her players make. $H_0 : p = 0.60$, $H_a : p < 0.60$ (the reporter doubts the claim of "at least 60%").

EXAMPLE 2

A test of $H_0 : p = 0.50$ versus $H_a : p > 0.50$ gives a p-value of 0.032. (a) Interpret the p-value. (b) State the conclusion at $\alpha = 0.05$ and at $\alpha = 0.01$.

SOLUTION

(a) If the true proportion is 0.50, there is a 0.032 probability of getting a sample proportion as large as the one observed (or larger) just by chance.

(b) At $\alpha = 0.05$: $0.032 < 0.05$, so reject H_0 . There is convincing evidence that the true proportion is greater than 0.50.

At $\alpha = 0.01$: $0.032 > 0.01$, so fail to reject H_0 . There is not convincing evidence that the true proportion is greater than 0.50. The same data can lead to different decisions at different significance levels.

EXAMPLE 3

A test statistic is $z = 2.10$. Find the p-value if (a) H_a is right-tailed, (b) H_a is two-sided, (c) the test statistic was $z = -1.80$ and H_a is two-sided.

SOLUTION

(a) $P(Z > 2.10) = 1 - 0.9821 = 0.0179$.

(b) Two tails: $2(0.0179) = 0.0357$.

(c) $P(Z < -1.80) = 0.0359$, doubled gives 0.0719. The sign of z only tells us the side; for two-sided tests we double the smaller tail.

EXAMPLE 4

A new treatment for a skin condition is compared with a standard one. The null hypothesis is that the new treatment is no better than the standard. (a) Describe a Type I error and a Type II error in context. (b) Which is more serious if the new treatment is expensive and has side effects?

SOLUTION

(a) **Type I:** concluding that the new treatment is better when it is actually no better. **Type II:** concluding that there is not enough evidence that the new treatment is better when it really is better.

(b) A Type I error would lead patients to pay for and suffer side effects of a treatment that does no extra good, so it may be more serious here; therefore a small α (such as 0.01) is reasonable. In other situations, a Type II error (missing a helpful treatment) may be worse.

EXAMPLE 5

A test of $H_0 : p = 0.50$ versus $H_a : p > 0.50$ at $\alpha = 0.05$ uses $n = 100$. Find the power when the true proportion is $p = 0.60$.

SOLUTION

Reject H_0 when $\hat{p}$ exceeds the cutoff: $0.50 + 1.645\sqrt{\frac{0.5(0.5)}{100}} = 0.50 + 1.645(0.05) = 0.5822$.

When $p = 0.60$, $\hat{p}$ has mean 0.60 and standard deviation $\sqrt{0.6(0.4)/100} = 0.0490$.

Power = $P(\hat{p} > 0.5822) = P\left(Z > \frac{0.5822 - 0.60}{0.0490}\right) = P(Z > -0.36) \approx 0.64$.

So the test detects this difference about 64% of the time. For $n = 200$, the same steps give a power of about 0.89.

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 6

A basketball player claims to make 70% of her free throws. In a sample of 50 shots she makes 38 (76%). A simulation of 1,000 samples of 50 shots with a true probability of 0.70 gave 224 samples with at least 38 makes. Estimate the p-value for $H_a : p > 0.70$ and state a conclusion at $\alpha = 0.05$.

SOLUTION

$H_0 : p = 0.70$, $H_a : p > 0.70$, where p is the true probability she makes a free throw.

The estimated p-value is the fraction of simulated results at least as extreme as the observed one: $224/1000 = 0.224$.

Since $0.224 > 0.05$, we fail to reject H_0 . There is not convincing evidence that her true free throw percentage is greater than 70%. A result of 76% is quite plausible by chance alone when the true rate is 70%.

Coach's Playbook

- 1 Define the parameter and write H_0 and H_a .** Match the direction of H_a to the wording of the question ("better" is $>$, "different" is $\neq$).
- 2 Choose α before looking at data.** If none is given, use 0.05.
- 3 Get the p-value.** Use the standardized statistic and the area in the correct tail(s).
- 4 Compare and decide.** $p\text{-value} \leq \alpha$ means reject H_0 .
- 5 Conclude in context.** Use the phrase "convincing evidence" and mention H_a in the context of the problem.
- 6 If asked about errors,** write the error as a sentence about the context, not just "rejecting a true null."

Mental model: a trial. H_0 is "innocent until proven guilty," the p-value measures how surprising the evidence is if the defendant is innocent, and we only convict (reject H_0) when the evidence would be very unlikely under innocence. Acquittal (failing to reject) does not prove innocence.

Common traps

1. Writing the p-value as "the probability that H_0 is true." It is computed assuming H_0 is true. 2. Saying "accept H_0 " or "prove H_0 ." The correct phrase is "fail to reject H_0 " or "not convincing evidence for H_a ." 3. Using $\hat{p}$ or $\bar{x}$ in a hypothesis. Hypotheses are always about parameters. 4. Confusing the probability of a Type I error (α) with the p-value. α is chosen in advance; the p-value comes from the data.

Mistake Clinic**Spot the error 1**

A student writes: " $H_0 : \hat{p} = 0.30$, $H_a : \hat{p} < 0.30$ " for a survey of 80 students in which 20 percent ride the bus. Where is the error? The hypotheses are claims about the **population** proportion, not about the sample proportion. The sample proportion 0.20 is the data. Correct: $H_0 : p = 0.30$, $H_a : p < 0.30$, where p is the true proportion of all students at the school who ride the bus.

Spot the error 2

A student writes: "The p-value is 0.03, so there is a 3% chance that the null hypothesis is true." Where is the error? The p-value assumes H_0 is true; it is the probability of a result at least this extreme if H_0 is true. Correct: "If the null hypothesis is true, there is a 3% chance of getting a result at least this extreme by chance alone."

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 3

A student concludes: "The p-value of 0.18 is greater than 0.05, so we accept the null hypothesis and conclude that the true proportion is exactly 0.70." Where is the error? A large p-value only means the data are consistent with H_0 ; many other values of p would also be consistent. Correct: "Because 0.18 is greater than 0.05, we fail to reject H_0 . There is not convincing evidence that the true proportion differs from 0.70."

Real-World Connection

Websites run experiments on layouts and test whether a new design gets more clicks. The null hypothesis is that the click rate has not changed. A low p-value is evidence that the design change really matters, and the company checks the power of the test so that a real improvement is not missed because the sample was too small.

Challenge Corner

A test of $H_0 : p = 0.50$ versus $H_a : p > 0.50$ uses $n = 100$ and $\alpha = 0.05$. The true proportion is 0.60. Without recomputing everything, explain what happens to power if (a) α is changed to 0.01, (b) the true proportion is 0.70. Then verify with calculations. Solution: (a) A smaller α makes it harder to reject, so power decreases. The cutoff becomes $0.50 + 2.326(0.05) = 0.6163$ and power $= P(Z > (0.6163 - 0.60)/0.0490) = P(Z > 0.33) \approx 0.37$. (b) A larger gap between the true value and the null makes the difference easier to detect, so power increases. With the cutoff 0.5822: $P(Z > (0.5822 - 0.70)/0.0458) = P(Z > -2.57) \approx 0.995$.

Practice

PRACTICE The Logic of Significance Tests

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

- A researcher claims that the mean wait time at a clinic is greater than 50 minutes. Which are the correct hypotheses?
 (A) $H_0 : \bar{x} = 50$, $H_a : \bar{x} > 50$ (B) $H_0 : \mu > 50$, $H_a : \mu = 50$ (C) $H_0 : \mu = 50$, $H_a : \mu > 50$
 (D) $H_0 : \mu = 50$, $H_a : \mu \neq 50$
- Which best describes a p-value?
 (A) The probability that the null hypothesis is true
 (B) The probability that the alternative hypothesis is true
 (C) The probability of a Type II error
 (D) The probability, assuming the null hypothesis is true, of a result at least as extreme as the one observed
- A test gives a p-value of 0.03. Using $\alpha = 0.05$, which conclusion is correct?
 (A) Reject H_0 ; there is convincing evidence for H_a
 (B) Fail to reject H_0 ; there is convincing evidence that H_0 is true
 (C) Reject H_0 ; the probability that H_0 is true is 0.03
 (D) Fail to reject H_0 ; the sample was too small
- A pharmacy tests H_0 : the new batch has a mean of 500 mg of the active ingredient, against H_a : the mean is not 500 mg. A Type I error would be
 (A) concluding the mean is 500 mg when it is not
 (B) concluding that the mean differs from 500 mg when it does not
 (C) failing to detect that the mean differs from 500 mg
 (D) getting a p-value larger than α

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. Which change would increase the power of a test?
 - (A) Decreasing the sample size
 - (B) Decreasing α
 - (C) Increasing the sample size
 - (D) Using a null value closer to the true parameter
6. A two-sided test has $z = -1.80$. What is the p-value?
 - (A) 0.0719
 - (B) 0.0359
 - (C) 0.9641
 - (D) 0.1438
7. The probability of a Type I error equals
 - (A) the p-value
 - (B) the significance level α
 - (C) the power
 - (D) $1 - \alpha$
8. A test of $H_0 : p = 0.40$ versus $H_a : p > 0.40$ has a p-value of 0.12. Which is an appropriate conclusion at $\alpha = 0.05$?
 - (A) There is not convincing evidence that $p > 0.40$
 - (B) There is convincing evidence that $p = 0.40$
 - (C) There is convincing evidence that $p > 0.40$
 - (D) The null hypothesis has been proven true
9. Which statement is a valid hypothesis?
 - (A) $H_a : \hat{p} < 0.30$
 - (B) $H_0 : \bar{x} = 12$
 - (C) $H_a : p = 0.30$ and $H_0 : p < 0.30$
 - (D) $H_0 : p = 0.30$ and $H_a : p < 0.30$ with p the true proportion of all students
10. A city claims that 40% of its residents recycle. A resident believes the real percentage is lower. State the hypotheses and define the parameter.
11. A test of $H_0 : \mu = 100$ against $H_a : \mu \neq 100$ gives a p-value of 0.012. Interpret the p-value in context if μ is the mean score on a reading test, and state the conclusion at $\alpha = 0.05$.
12. A bank tests H_0 : a customer is not committing fraud, against H_a : the customer is committing fraud. Describe a Type I error and a Type II error in context and give one consequence of each.
13. A researcher runs a simulation of 200 samples under H_0 . Nine of the simulated results are as extreme as the observed sample result in the direction of H_a . Estimate the p-value and state a conclusion at $\alpha = 0.05$.
14. A test of $H_0 : p = 0.50$ versus $H_a : p > 0.50$ at $\alpha = 0.05$ has $n = 100$. Find the power when $p = 0.60$, and explain what would happen to the power if the sample size were increased.
15. A test of $H_0 : p = 0.12$ versus $H_a : p \neq 0.12$ gives a p-value of 0.34 at $\alpha = 0.05$. Write the conclusion, and explain why we do not say that we have proven the true proportion is 0.12.

34.2 Tests for Proportions and Means

Carry out one-sample z tests for proportions and one-sample t tests for means, check conditions with the correct numbers, and write complete four-step conclusions in context.

Why This Matters

Once you understand the logic of a significance test, the next job is to carry it out correctly. Two procedures cover most one-sample questions: the **one-sample z test for a proportion** (is the true percentage equal to a claimed value?) and the **one-sample t test for a mean** (is the true average equal to a claimed value?). They are used to check labels on packages, test manufacturing standards, evaluate teaching programs, and review the claims in advertisements.

Students lose points on these for the same reasons again and again: checking conditions with the wrong numbers (using $\hat{p}$ when the test calls for p_0), using $\hat{p}$ in the standard error of a z test, forgetting the degrees of freedom, and writing conclusions without context. A fixed four-step write-up (State, Plan, Do, Conclude) removes most of these problems.

Core Ideas

Both tests compute a **test statistic**: how many standard errors the sample result is from the null value. A big distance (in the direction of H_a) gives a small p-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 →

One-Sample z Test for a Proportion

$H_0 : p = p_0$. Test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Conditions: Random; 10% ($n \leq 0.10N$); Large Counts using the **null value**: $np_0 \geq 10$ and $n(1 - p_0) \geq 10$. The standard error uses p_0 because the test assumes H_0 is true. The p-value comes from the standard normal curve.

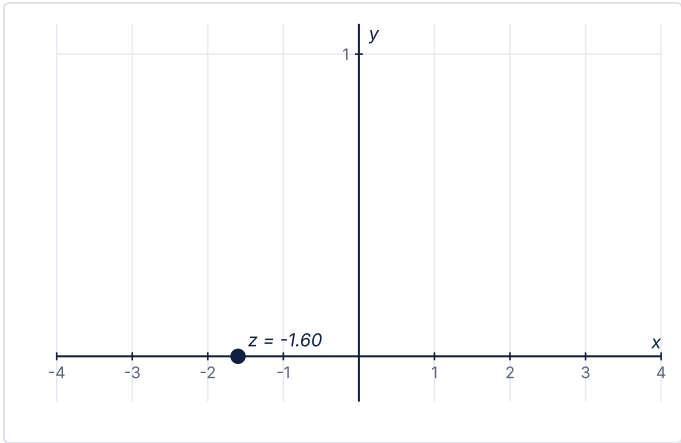
One-Sample t Test for a Mean

$H_0 : \mu = \mu_0$. Test statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \quad df = n - 1.$$

Conditions: Random; 10%; Normal/Large Sample ($n \geq 30$, or no strong skew or outliers for smaller samples, shown by a graph). The p-value comes from the t distribution with $n - 1$ degrees of freedom.

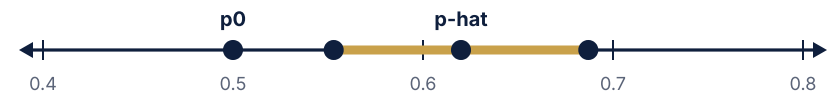
A good write-up has four parts. **State:** the hypotheses, the parameter in words, and α . **Plan:** name the test and check the conditions with numbers. **Do:** compute the test statistic and p-value. **Conclude:** compare the p-value with α and answer in context. The picture shows the p-value for a left-tailed test with $z = -1.60$: the area under the curve to the left of -1.60 .



For a left-tailed test with $z = -1.60$, the p-value is the shaded area to the left, about 0.0548.

	Proportion	Mean
Parameter	p	μ
Null value	p_0	μ_0
Standard error	$\sqrt{p_0(1 - p_0)/n}$ (uses p_0)	$s/\sqrt{n}$
Distribution	Standard normal	t with $n - 1$ df
Key condition	$np_0 \geq 10, n(1 - p_0) \geq 10$	Normal/Large Sample

Tests and intervals agree. A two-sided test at significance level α rejects $H_0 : \mu = \mu_0$ exactly when μ_0 is outside the $(1 - \alpha)$ confidence interval (this is exact for means, and close for proportions). On the number line below, the interval for a proportion does not contain the null value 0.50, so a two-sided test at $\alpha = 0.05$ would reject $H_0 : p = 0.50$.



A 95% interval (0.553, 0.687) with sample proportion 0.62. The null value 0.50 lies outside the interval.

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 →

EXAMPLE 1

A school says that 70% of its seniors apply to a four-year college. In a random sample of 150 seniors from this large school, 118 say they applied. Is there convincing evidence that the true proportion is greater than 0.70? Use $\alpha = 0.05$.

SOLUTION

State: p is the true proportion of all seniors at the school who applied. $H_0 : p = 0.70$, $H_a : p > 0.70$, $\alpha = 0.05$.

Plan: One-sample z test for a proportion. Random: given. 10%: the school is large. Large Counts: $150(0.70) = 105 \geq 10$ and $150(0.30) = 45 \geq 10$.

Do: $\hat{p} = 118/150 = 0.7867$. $z = \frac{0.7867 - 0.70}{\sqrt{0.70(0.30)/150}} = \frac{0.0867}{0.0374} \approx 2.32$. p-value = $P(Z > 2.32) \approx 0.0103$.

Conclude: Since $0.0103 < 0.05$, we reject H_0 . There is convincing evidence that the true proportion of seniors at the school who applied is greater than 0.70.

EXAMPLE 2

A game spinner is supposed to land on red 25% of the time. In 400 spins it lands on red 88 times. Test whether the spinner is fair with respect to red, using $\alpha = 0.05$.

SOLUTION

State: p is the true probability of landing on red. $H_0 : p = 0.25$, $H_a : p \neq 0.25$.

Plan: z test for a proportion. The spins are treated as a random process; Large Counts: $400(0.25) = 100$ and $400(0.75) = 300$, both at least 10.

Do: $\hat{p} = 88/400 = 0.22$. $z = \frac{0.22 - 0.25}{\sqrt{0.25(0.75)/400}} = \frac{-0.03}{0.02165} \approx -1.39$. Two-sided p-value = $2P(Z < -1.39) \approx 0.166$.

Conclude: Since $0.166 > 0.05$, we fail to reject H_0 . There is not convincing evidence that the spinner's probability of red differs from 0.25.

EXAMPLE 3

A snack company says bags contain a mean of 28 grams of chips. A random sample of 16 bags has $\bar{x} = 27.1$ g and $s = 1.8$ g, with a roughly symmetric dotplot. Is there convincing evidence that the mean is less than 28 grams? Use $\alpha = 0.05$.

SOLUTION

State: μ is the true mean weight of chips in all bags. $H_0 : \mu = 28$, $H_a : \mu < 28$.

Plan: One-sample t test for a mean. Random: given. 10%: a company produces many bags. Normal/Large Sample: $n = 16 < 30$, but the dotplot is roughly symmetric with no outliers.

Do: $t = \frac{27.1 - 28}{1.8/\sqrt{16}} = \frac{-0.9}{0.45} = -2.00$, $df = 15$. p-value = $P(t_{15} < -2.00) \approx 0.032$.

Conclude: Since $0.032 < 0.05$, we reject H_0 . There is convincing evidence that the true mean weight of chips per bag is less than 28 grams.

EXAMPLE 4

(a) A test of $H_0 : \mu = 50$ versus $H_a : \mu > 50$ with 20 observations gives $t = 2.31$. Estimate the p-value. (b) A researcher wants to test $H_0 : p = 0.05$ using a random sample of 60 people, 7 of whom had the trait. May she use a z test?

SOLUTION

(a) $df = 19$. In a t table row for 19 df, 2.31 lies between 2.093 (upper-tail area 0.025) and 2.539 (area 0.01), so the p-value is between 0.01 and 0.025. Technology gives $P(t_{19} > 2.31) \approx 0.0161$.

(b) Large Counts uses the **null value**: $np_0 = 60(0.05) = 3 < 10$. The condition fails even though 7 successes were observed, so the z test is not appropriate.

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

A town claims the mean commute time is 15 minutes. A random sample of 10 commuters has times (minutes) 16.2, 14.8, 17.5, 15.9, 18.1, 16.4, 15.2, 17.0, 16.8, 14.9, and a graph of the data is roughly symmetric. Is there convincing evidence that the mean is greater than 15? Use $\alpha = 0.01$.

SOLUTION

State: μ is the true mean commute time in the town. $H_0 : \mu = 15$, $H_a : \mu > 15$, $\alpha = 0.01$.

Plan: t test for a mean. Random sample (given), roughly symmetric with no outliers (small sample).

Do: $\bar{x} = 16.28$, $s = 1.106$. $t = \frac{16.28 - 15}{1.106/\sqrt{10}} = \frac{1.28}{0.3498} \approx 3.66$, $df = 9$, p-value ≈ 0.0026 .

Conclude: Since $0.0026 < 0.01$, we reject H_0 . There is convincing evidence that the mean commute time in the town is greater than 15 minutes.

EXAMPLE 6

A 95% confidence interval for the proportion of voters supporting a measure is (0.553, 0.687). (a) What would a two-sided test at $\alpha = 0.05$ of $H_0 : p = 0.50$ conclude? (b) What about $H_0 : p = 0.60$?

SOLUTION

(a) 0.50 is outside the interval, so the test rejects H_0 : there is convincing evidence that the true proportion differs from 0.50. Since the interval is above 0.50, the proportion is likely greater than 0.50.

(b) 0.60 is inside the interval, so the test fails to reject H_0 : there is not convincing evidence that the true proportion differs from 0.60.

This agrees with the logic of the test: the interval lists the null values that would **not** be rejected.

Coch's Playbook

- 1 **State.** Parameter in words, H_0 and H_a , and α (0.05 if not given).
- 2 **Plan.** Name the test. For a proportion, check Large Counts with p_0 . For a mean, graph or discuss the data when $n < 30$.
- 3 **Do.** Write the formula, substitute the numbers, and then give the test statistic and p-value (with df for t).
- 4 **Conclude.** Compare p-value to α , decide, and write a sentence with context and the words "convincing evidence".
- 5 **Check direction.** For a left-tailed test the p-value must be below 0.5 if the statistic is negative. If your p-value is above 0.5, you used the wrong tail.
- 6 **Use technology wisely.** You may use a calculator for the test statistic and p-value, but you still need to write the conditions and the conclusion.

A mental model: the test statistic is a **z-score or t-score for the sample result**, assuming the null is true. Anything beyond about ± 2 is rare. Also remember the contrast between **statistical significance** (an unlikely result under H_0) and **practical importance** (a difference big enough to matter).

Common traps

1. Using $\hat{p}$ in the standard error of the z test or in the Large Counts check. Both use p_0 .
2. Using a z procedure for a mean, or $df = n$ for t.
3. Doubling the p-value for a one-sided test, or forgetting to double it for a two-sided test.
4. Concluding without the context. A correct answer says what the evidence is about.

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 →

Spot the error 1

A student tests $H_0 : p = 0.30$ with $\hat{p} = 0.24$, $n = 150$ and writes $z = \frac{0.24 - 0.30}{\sqrt{0.24(0.76)/150}} = -1.72$. Where is the error? The standard error must use the null value: $\sqrt{0.30(0.70)/150} = 0.0374$, so $z = \frac{-0.06}{0.0374} = -1.60$. Using $\hat{p}$ in the denominator is the confidence interval formula, not the test formula.

Spot the error 2

A student runs a t test with $n = 16$ and writes: " $df = 16$, so the p-value is $P(t_{16} < -2.00) \approx 0.031$." Where is the error? The degrees of freedom are $n - 1 = 15$. With $df = 15$ the p-value is 0.032. The difference is small here, but on a table-based question it can change the decision, and it shows the wrong degrees of freedom.

Spot the error 3

A student concludes: "The p-value of 0.0544 is greater than 0.05, so we fail to reject H_0 . This proves that the true proportion is 0.30." Where is the error? Failing to reject does not prove H_0 . Correct: "Since $0.0544 > 0.05$, we fail to reject H_0 . There is not convincing evidence that the true proportion of customers who order the seasonal drink is less than 0.30."

Real-World Connection

Quality inspectors test whether a machine still fills cans to the target mean. They sample a few cans each hour and run a t test. A p-value below the chosen level means the machine drifted and should be adjusted. Because every false alarm costs time, the plant picks α with the cost of a Type I error in mind.

Challenge Corner

With $n = 40,000$ voters sampled at random, a poll finds $\hat{p} = 0.5075$ supporting a measure. Test $H_0 : p = 0.50$ versus $H_a : p > 0.50$ and comment on whether the result matters. Solution: $z = \frac{0.5075 - 0.50}{\sqrt{0.25/40000}} = \frac{0.0075}{0.0025} = 3.0$, p-value = $P(Z > 3.0) \approx 0.0013$. We reject H_0 at any usual level, so the result is statistically significant. But the estimated difference is only 0.75 percentage points, which may be too small to change anything in practice: with a very large sample even a tiny difference becomes statistically significant.

Practice**PRACTICE Tests for Proportions and Means**

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 200 gives $\hat{p} = 0.42$. For $H_0 : p = 0.50$, what is the test statistic?
 (A) $z \approx -2.29$ (B) $z \approx -0.08$ (C) $z \approx -2.26$ (D) $z \approx 2.26$
- To check the Large Counts condition for a z test of $H_0 : p = 0.05$ with $n = 200$, you should verify that
 (A) $200(0.05) \geq 10$ and $200(0.95) \geq 10$ (B) $200\hat{p} \geq 10$ and $200(1 - \hat{p}) \geq 10$
 (C) $n \geq 30$ (D) the population is normally distributed
- A t test for a mean is based on a random sample of 22 observations. How many degrees of freedom does it have?
 (A) 22 (B) 20 (C) 23 (D) 21
- A random sample of 25 has $\bar{x} = 48.2$ and $s = 6$. What is the test statistic for $H_0 : \mu = 50$?
 (A) $t = -0.3$ (B) $t = -1.5$ (C) $t = -7.5$ (D) $t = -1.8$

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

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

BOOK NOW →

5. A left-tailed z test has a test statistic of -2.26 . What is the p-value?
 (A) 0.0119 (B) 0.0238 (C) 0.9881 (D) 0.0226
6. When is a t procedure used instead of a z procedure for a mean?
 (A) When the sample is small (B) When the population is skewed
 (C) When the data are proportions (D) When σ is unknown and estimated by s
7. A z test of $H_0 : p = 0.70$ versus $H_a : p > 0.70$ gives a p-value of 0.0103. Which conclusion at $\alpha = 0.05$ is correct?
 (A) Fail to reject H_0 ; the true proportion is 0.70
 (B) Reject H_0 ; the probability that H_0 is true is 0.0103
 (C) Reject H_0 ; there is convincing evidence that the true proportion is greater than 0.70
 (D) Fail to reject H_0 ; there is not convincing evidence that the proportion is greater than 0.70
8. A random sample of 100 shows 12 successes. A student wants a z test of $H_0 : p = 0.05$. Which statement is correct?
 (A) The Large Counts condition is met because there are 12 successes and 88 failures
 (B) The Large Counts condition is not met because $100(0.05) = 5 < 10$
 (C) The test cannot be done because 12 is greater than 5
 (D) The condition is met because $n \geq 30$
9. A 95% confidence interval for a mean is (12.4, 15.6). A two-sided test of $H_0 : \mu = 16$ at $\alpha = 0.05$ would
 (A) fail to reject H_0 because 16 is close to the interval
 (B) not be possible without the data
 (C) fail to reject H_0 because 16 is greater than the sample mean
 (D) reject H_0 because 16 is outside the interval
10. A cafe says that 30% of its customers order a seasonal drink. In a random sample of 150 customers from a large population, 36 ordered one. Is there convincing evidence that the true proportion is less than 0.30? Use the four-step process with $\alpha = 0.05$.
11. A packaging plant claims a mean of 8 ounces per cup. In a random sample of 12 cups, $\bar{x} = 7.4$ and $s = 1.1$, and the data are roughly symmetric with no outliers. Test whether the mean differs from 8 ounces at $\alpha = 0.05$.
12. A random sample of 40 students spends a mean of $\bar{x} = 102.5$ minutes ($s = 8$) on a daily reading task, and the guideline is 100 minutes. Find the test statistic and p-value for $H_0 : \mu = 100$ versus $H_a : \mu > 100$ and state the conclusion at $\alpha = 0.05$.
13. A researcher tests $H_0 : p = 0.20$ with a random sample of 45 people and observes $\hat{p} = 0.30$. Explain why the Large Counts condition is not met for a z test and what number you checked.
14. A 99% confidence interval for the proportion of customers who prefer product A is (0.32, 0.46). Use the interval to say what a two-sided test of $H_0 : p = 0.50$ at $\alpha = 0.01$ would conclude.
15. A very large poll of 10,000 people finds $\hat{p} = 0.503$ for a measure and the test of $H_0 : p = 0.50$ versus $H_a : p \neq 0.50$ is run. Find z and the p-value (use $P(|Z| > 0.6) = 0.549$), and explain why a result like this is not evidence that the true proportion is far from 0.50.

34.3 Comparing Two Groups

Compare two proportions, two means, and paired data. Choose the right procedure from the study design, check conditions for both groups, and write conclusions that compare the groups in context.

Why This Matters

Most useful questions compare two groups. Does a new study app raise pass rates? Do students in one program score higher than students in another? Did a training plan improve race times? Comparing two groups is the heart of experiments, A/B tests, medical trials, and surveys of two populations. The AP Statistics exam asks for these procedures on almost every free-response section.

The challenge is choosing the right procedure and then writing it correctly. Students mix up **independent** samples with **paired** data, forget to pool the proportions in a two-proportion test, use the wrong degrees of freedom, and write conclusions that do not mention both groups. This section organizes everything around one question: *how were the data collected?*

Core Ideas

First decide the design. **Two independent groups** come from two separate random samples or from two groups created by random assignment in an experiment. **Paired data** are two measurements on the same individual (before and after) or on matched pairs. Paired data are analyzed by finding the differences and using a one-sample t 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 →

Situation	Procedure	Parameter
Two groups, categorical response	Two-proportion z	$p_1 - p_2$
Two independent groups, quantitative response	Two-sample t	$\mu_1 - \mu_2$
Same individuals measured twice, or matched pairs	Paired t (one-sample t on differences)	μ_d

Two-Proportion z Procedures

Test of $H_0 : p_1 = p_2$: use the **pooled** proportion $\hat{p}_c = \frac{x_1 + x_2}{n_1 + n_2}$ and

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c(1 - \hat{p}_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Interval for $p_1 - p_2$ (no pooling):

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

Conditions: independent random samples (or random assignment); 10% for each sample; Large Counts: at least 10 successes and 10 failures in each group.

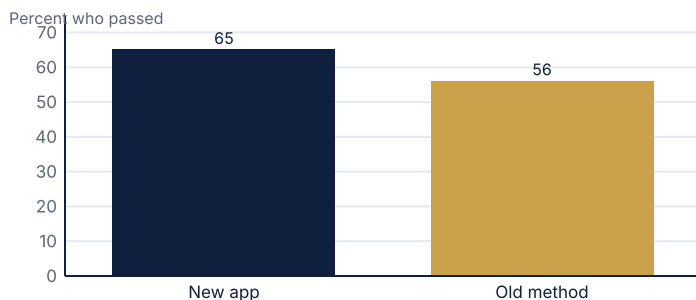
Two-Sample t Procedures

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad \text{interval: } (\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Degrees of freedom: use technology; the conservative choice is $\min(n_1 - 1, n_2 - 1)$.

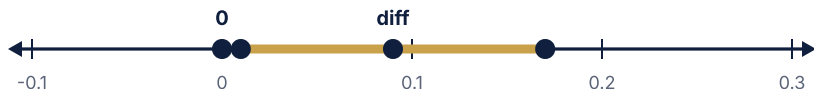
Conditions: independent random samples (or random assignment); 10% for each sample; each group is approximately normal, or has $n \geq 30$ (for smaller groups graph the data and check for no strong skew or outliers).

The bar chart shows the results of an experiment with 200 students in each group. The difference between pass rates is 9 percentage points, but the question is whether a difference this large could be due to the random assignment alone. That is exactly what the test addresses.



Pass rates in a randomized experiment with 200 students per group: 130 of 200 (65%) with the new app and 112 of 200 (56%) with the old method.

A confidence interval for a **difference** is easiest to read relative to 0. If 0 is inside the interval, then no difference is a plausible value. If the whole interval is above 0, the first group's parameter is probably larger. The number line shows a 90% interval for $p_1 - p_2$ for the experiment above.



A 90% confidence interval for $p_1 - p_2$ is (0.010, 0.170). The interval is entirely above 0.

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 →

EXAMPLE 1

In an experiment, 400 students were randomly assigned to use a new study app (200 students, 130 passed the unit exam) or the old method (200 students, 112 passed). Is there convincing evidence that the new app has a higher pass rate? Use $\alpha = 0.05$.

SOLUTION

State: p_1 = true proportion who would pass using the new app; p_2 = true proportion who would pass using the old method. $H_0 : p_1 = p_2$, $H_a : p_1 > p_2$.

Plan: Two-proportion z test. Random assignment. Large Counts: pooled $\hat{p}_c = \frac{130 + 112}{400} = 0.605$; successes and failures in both groups are above 10 (130, 70, 112, 88).

Do: $\hat{p}_1 - \hat{p}_2 = 0.65 - 0.56 = 0.09$. $SE = \sqrt{0.605(0.395) \left(\frac{1}{200} + \frac{1}{200} \right)} \approx 0.0489$. $z = \frac{0.09}{0.0489} \approx 1.84$. p-value ≈ 0.033 .

Conclude: Since $0.033 < 0.05$, reject H_0 . There is convincing evidence that the new app gives a higher pass rate than the old method. Because of random assignment, the app can be given credit for the difference.

EXAMPLE 2

Construct a 90% confidence interval for $p_1 - p_2$ using the data in Example 1 and interpret it.

SOLUTION

Conditions are met as in Example 1. For an interval we do not pool.

$$SE = \sqrt{\frac{0.65(0.35)}{200} + \frac{0.56(0.44)}{200}} = \sqrt{0.0011375 + 0.001232} \approx 0.0487.$$

Margin of error = $1.645(0.0487) \approx 0.0801$. Interval: $0.09 \pm 0.0801 = (0.010, 0.170)$.

We are 90% confident that the interval from 0.010 to 0.170 captures the true difference (new app minus old method) in pass rates. Since the whole interval is above 0, the new app is likely better, though the improvement could be as small as 1 point.

EXAMPLE 3

A plant biologist compares two fertilizers. Group A: $n = 15$, $\bar{x} = 12.8$ cm, $s = 2.1$. Group B: $n = 12$, $\bar{x} = 11.1$ cm, $s = 2.6$. Plants were randomly assigned, and the growth data show no strong skew or outliers. Is there convincing evidence that the mean growth differs? Use $\alpha = 0.05$.

SOLUTION

State: μ_A and μ_B are the true mean growths with fertilizers A and B. $H_0 : \mu_A = \mu_B$, $H_a : \mu_A \neq \mu_B$.

Plan: Two-sample t test. Random assignment; graphs of each group show no strong skew or outliers, which suffices for small samples.

Do: $SE = \sqrt{\frac{2.1^2}{15} + \frac{2.6^2}{12}} = \sqrt{0.294 + 0.563} \approx 0.926$. $t = \frac{12.8 - 11.1}{0.926} \approx 1.84$. Technology gives $df \approx 21.0$ and a two-sided p-value ≈ 0.081 . (Using the conservative $df = 11$ also gives a p-value above 0.05.)

Conclude: Since $0.081 > 0.05$, fail to reject H_0 . There is not convincing evidence that the true mean growth differs between the two fertilizers.

EXAMPLE 4

Construct a 95% confidence interval for $\mu_A - \mu_B$ using the data of Example 3 and compare it with the conservative df approach.

SOLUTION

Difference of means: $12.8 - 11.1 = 1.7$, $SE \approx 0.926$.

With technology $df \approx 21.0$ and $t^* \approx 2.080$: margin of error = $2.080(0.926) \approx 1.93$. Interval: $(-0.23, 3.63)$.

Conservative $df = 11$, $t^* = 2.201$: margin of error ≈ 2.04 , interval $(-0.34, 3.74)$, slightly wider.

Both intervals contain 0, which agrees with the test result: a difference of 0 is plausible.

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

Eight runners recorded their 1-mile times before and after a training program. The differences (before minus after, in minutes) were 2.1, 0.8, 1.5, 2.6, -0.4, 1.9, 1.2, and 0.9. The differences show no outliers. Is there convincing evidence that the program lowers the mean time? Use $\alpha = 0.05$.

SOLUTION

State: μ_d = true mean difference (before minus after) for all runners like these. $H_0 : \mu_d = 0$, $H_a : \mu_d > 0$.

Plan: Paired t test (one-sample t on the differences). The runners are treated as a random sample; the differences show no strong skew or outliers.

Do: $\bar{d} = 1.325$, $s_d = 0.929$. $t = \frac{1.325}{0.929/\sqrt{8}} \approx 4.04$, $df = 7$, p-value ≈ 0.0025 .

Conclude: Since $0.0025 < 0.05$, reject H_0 . There is convincing evidence that the program lowers the mean 1-mile time.

EXAMPLE 6

Choose the correct procedure for each. (a) Compare the proportion of left-handed students in two schools using random samples. (b) Compare the mean systolic blood pressure of 40 patients before and after medication. (c) Compare mean test scores of 30 students taught with method A and 28 different students taught with method B.

SOLUTION

(a) **Two-proportion z.** Two independent samples with a categorical response.

(b) **Paired t.** The same patients are measured twice, so find each patient's difference and use a one-sample t procedure.

(c) **Two-sample t.** Two independent groups with a quantitative response. A common error is treating (b) as two independent samples, which ignores that the data are linked.

Coach's Playbook

- 1 **Ask: how many groups, and are they linked?** Two separate groups means independent. The same people twice means paired.
- 2 **Identify the response.** Categorical (success or failure) uses proportions. Quantitative uses means.
- 3 **Define both parameters.** Label them with context: " p_1 is the true proportion of ___ who ___."
- 4 **Check conditions for each group.** Random, 10%, and Large Counts (or Normal/Large Sample) for both.
- 5 **Pool only for the two-proportion test.** Do not pool for an interval or for the two-sample t.
- 6 **Conclude by comparing the groups in context.** Say which direction and whether random assignment allows a cause-and-effect statement.

Mental model: both procedures ask whether the observed difference is large compared with the **standard error of the difference**. The standard error of a difference is larger than either group's own standard error because variability adds: $SE = \sqrt{SE_1^2 + SE_2^2}$. A confidence interval for a difference containing 0 means that no difference is plausible.

Common traps

1. Pooling in a confidence interval or using separate proportions in the test standard error. 2. Treating paired data as two independent samples (or the reverse). 3. Concluding cause and effect from two random samples that were not randomly assigned. 4. Writing a conclusion about only one group, instead of comparing the two parameters.

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 →

Spot the error 1

A student computes the test statistic for Example 1 as $z = \frac{0.09}{\sqrt{\frac{0.65(0.35)}{200} + \frac{0.56(0.44)}{200}}} = 1.85$ and says it is correct. Where is the error? In a test of $H_0 : p_1 = p_2$ the null hypothesis says the proportions are equal, so we pool them: $\hat{p}_c = 0.605$ and $SE = 0.0489$, giving $z = 1.84$. The unpooled standard error is for the confidence interval. The values are close here, but the method must match the procedure.

Spot the error 2

A researcher measures the resting heart rate of 12 students before and after a workout and runs a two-sample t test with $df = 11$ comparing the two columns of data. Where is the error? The measurements are paired because the same students are in both columns. The correct procedure finds the 12 differences and runs a one-sample t test with $df = 11$ on the differences. A two-sample test ignores the connection and loses power because it includes the student-to-student variability.

Spot the error 3

A student writes: "We randomly sampled 300 men and 300 women and found a significant difference in whether they like action movies, so being a woman causes the difference in preferences." Where is the error? This is an observational comparison of two random samples. Random sampling allows generalizing to the populations, but no treatment was assigned, so there is no cause-and-effect claim. Correct: "There is convincing evidence that the proportions who like action movies differ between the populations of men and women."

Real-World Connection

Technology companies run A/B tests by randomly showing visitors one of two versions of a page. They then compare the proportions of visitors who click using a two-proportion z test. A small p-value suggests that the version change caused the change in clicks, because visitors were randomly assigned.

Challenge Corner

A 95% confidence interval for $p_1 - p_2$ is built from $x_1 = 72$ of $n_1 = 120$ and $x_2 = 55$ of $n_2 = 110$. Find the interval and decide whether the data provide convincing evidence of a difference at the 5% level. Solution: $\hat{p}_1 = 0.60$, $\hat{p}_2 = 0.50$, difference 0.10. $SE = \sqrt{\frac{0.60(0.40)}{120} + \frac{0.50(0.50)}{110}} = \sqrt{0.002 + 0.002273} \approx 0.0654$. Margin of error = $1.96(0.0654) \approx 0.128$. Interval: $(-0.028, 0.228)$. Since 0 is inside the interval, a two-sided test at $\alpha = 0.05$ would fail to reject $H_0 : p_1 = p_2$: there is not 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 →

Practice

PRACTICE Comparing Two Groups

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

- In sample 1, 45 of 100 are successes. In sample 2, 30 of 80 are successes. What is the pooled proportion $\hat{p}_c$ used in the two-proportion z test?
 (A) 0.4125 (B) 0.4167 (C) 0.075 (D) 0.450
- Which situation calls for a paired t procedure?
 (A) Comparing mean heights of random samples of 9th and 12th graders
 (B) Comparing the proportions of two cities that support a measure
 (C) Comparing mean salaries of two unrelated companies
 (D) Comparing the pulse rates of 20 students before and after exercise
- What is the standard error of $\bar{x}_1 - \bar{x}_2$ for independent samples?
 (A) $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ (B) $\frac{s_1 + s_2}{\sqrt{n_1 + n_2}}$ (C) $\sqrt{\frac{s_1^2 + s_2^2}{n_1 + n_2}}$
 (D) $\frac{s_1}{\sqrt{n_1}} - \frac{s_2}{\sqrt{n_2}}$
- Which condition is required for a two-sample t test comparing two means?
 (A) Both sample sizes must equal each other
 (B) The two samples must be paired
 (C) The two groups are independent random samples (or randomized groups) and each is approximately normal or has $n \geq 30$
 (D) The population standard deviations must be known
- A 95% confidence interval for $p_1 - p_2$ is $(-0.02, 0.10)$. Which conclusion is best?
 (A) There is convincing evidence that $p_1 > p_2$
 (B) Zero is a plausible value, so there is not convincing evidence that the proportions differ
 (C) There is a 95% chance that $p_1 - p_2$ is positive
 (D) The two sample proportions are equal
- In sample 1, 56 of 100 succeed. In sample 2, 40 of 100 succeed. What is the test statistic for $H_0 : p_1 = p_2$?
 (A) $z \approx 0.16$ (B) $z \approx 1.13$ (C) $z \approx 2.26$ (D) $z \approx 3.2$
- A two-sample t test has groups of size 15 and 12. What is the conservative degrees of freedom?
 (A) 11 (B) 14 (C) 26 (D) 27
- Why does random assignment in a two-group experiment matter for the conclusion?
 (A) It guarantees the two groups have the same mean
 (B) It makes the data paired
 (C) It lets us generalize to all people
 (D) It makes a cause-and-effect conclusion possible if the difference is significant
- A paired t test uses 14 pairs of measurements. How many degrees of freedom does it have?
 (A) 14 (B) 28 (C) 13 (D) 12
- In city A, 135 of 300 randomly selected residents support a proposal. In city B, 90 of 250 randomly selected residents support it. Is there convincing evidence that the proportions differ? Use $\alpha = 0.05$ and the four-step process.
- A sample of 30 students taught with method A has $\bar{x} = 78.4$, $s = 9.1$. An independent sample of 28 taught with method B has $\bar{x} = 73.9$, $s = 8.2$. Both samples are random and show no extreme skew. Find the test statistic and p-value for $H_a : \mu_A > \mu_B$ and state a conclusion at $\alpha = 0.05$.
- Nine students' scores improved by an average of 3.2 points with standard deviation 2.4 after a tutoring program (differences after minus before, no outliers). Construct and interpret a 90% confidence interval for the mean improvement.
- Choose the procedure (two-proportion z, two-sample t, or paired t) and explain: (a) mean battery life of 25 phones of brand X and 30 phones of brand Y; (b) the proportion of 200 patients cured with drug A versus 200 with drug B; (c) 12 plants measured before and after a month of 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 →

14. A 95% confidence interval for $\mu_1 - \mu_2$ is $(-1.2, 5.6)$ based on independent random samples. Interpret the interval and say what it suggests about the difference between the means.
15. Of 120 randomly chosen adults in a community, 72 exercise weekly. Of 110 randomly chosen adults in a second community, 55 exercise weekly. Construct a 95% confidence interval for the difference in proportions and interpret it.

34.4 Chi-Square Tests and Regression Inference

Use chi-square tests for goodness-of-fit, homogeneity, and independence, and use t procedures for the slope of a regression line, including conditions, degrees of freedom, and complete conclusions.

Why This Matters

Not every question is about one proportion or one mean. How do we test whether a die is fair, whether the mix of colors in a bag of candy matches what the company claims, whether three teaching formats lead to different pass rates, or whether more study time really goes with higher scores? The **chi-square tests** handle questions about categorical data, and **inference for the slope of a regression line** handles two quantitative variables. These are the final tools of AP Statistics.

Students struggle with choosing among the three chi-square tests (they share a formula but differ in design), with expected counts and degrees of freedom, and with the language of the conclusion. For regression inference, the main troubles are reading computer output, remembering that $df = n - 2$, and stating the regression conditions. This section gives a way to decide which test to use and a template for each.

Core Ideas

The Chi-Square Statistic

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O is an observed count and E an expected count under H_0 . Large values of χ^2 mean the observed counts are far from what H_0 predicts. The p-value is the area to the **right** of χ^2 under a chi-square curve.

Conditions: Random (sample or assignment); 10%; **Large Counts:** every expected count is at least 5 (check expected counts, not observed counts).

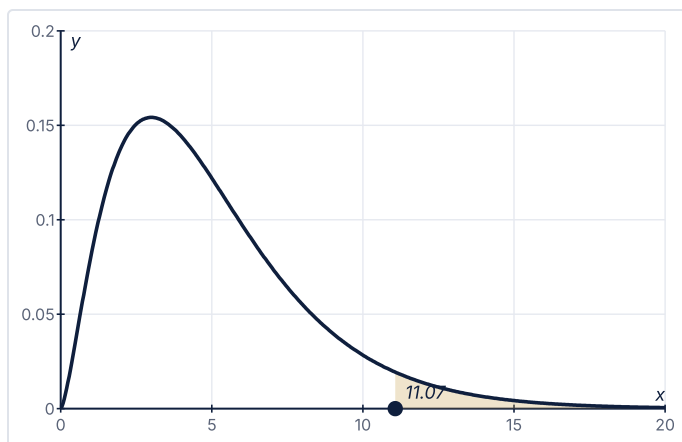
Test	Question	Data	Expected count	df
Goodness-of-fit	Does one categorical variable follow a claimed distribution?	One sample, one variable	$n \cdot p_i$ from the claim	categories $- 1$
Homogeneity	Is the distribution of a variable the same for several populations or treatments?	Separate samples from each group	$\frac{\text{row total} \times \text{column total}}{\text{table total}}$	$(r - 1)(c - 1)$
Independence	Are two categorical variables associated?	One sample, two variables	$\frac{\text{row total} \times \text{column total}}{\text{table total}}$	$(r - 1)(c - 1)$

The chi-square distribution is skewed right and takes only positive values. Its shape depends on the degrees of freedom. The curve below has 5 degrees of freedom; the shaded tail starts at the 5% critical value 11.07. A test statistic beyond 11.07 would lead to rejecting H_0 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 →



A chi-square curve with 5 degrees of freedom. The shaded right tail (area 0.05) begins at 11.07.

Inference for the Slope of a Regression Line

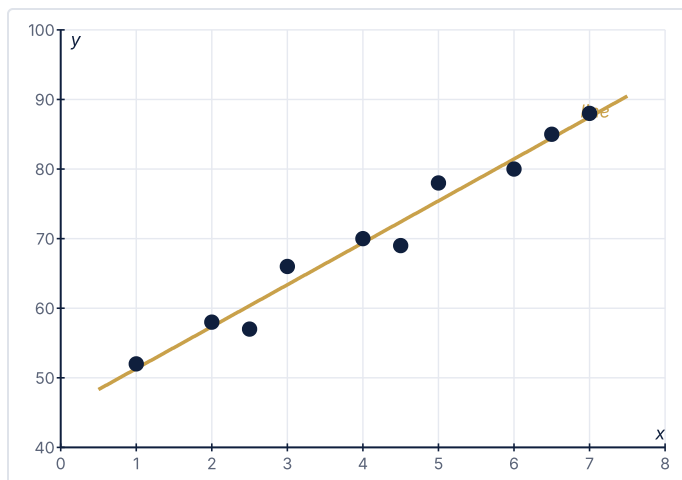
The population model is $y = \alpha + \beta x + \varepsilon$, and b estimates the true slope β . Test $H_0 : \beta = 0$ (no linear relationship) using

$$t = \frac{b - 0}{SE_b}, \quad df = n - 2.$$

Interval: $b \pm t^* SE_b$ with $df = n - 2$.

Conditions (LINER): Linear relationship, Independent observations (and 10%), **N**ormal residuals (about the line at each x), **E**qual standard deviation of residuals, **R**andom sample or experiment.

Computer output gives b and SE_b for you. The scatterplot below shows 10 students: hours studied versus exam score, with the least-squares line $\hat{y} = 45.30 + 6.02x$. The slope $b = 6.02$ means the predicted score increases by about 6 points for each extra hour studied. The test asks whether the true slope β could be 0.



Hours studied (x) and exam score (y) for 10 students with the least-squares line.

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 →

EXAMPLE 1

A die is rolled 120 times with counts for faces 1 to 6 of 18, 25, 16, 22, 14, 25. Is there convincing evidence that the die is not fair? Use $\alpha = 0.05$.

SOLUTION

State: H_0 : the die is fair (each face has probability $1/6$). H_a : at least one face has a different probability.

Plan: Chi-square goodness-of-fit. Rolls are random. Expected count for each face = $120/6 = 20 \geq 5$.

Do: $\chi^2 = \frac{(18-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(25-20)^2}{20} = \frac{110}{20} = 5.5$. $df = 6 - 1 = 5$, p-value ≈ 0.358 .

Conclude: Since $0.358 > 0.05$, fail to reject H_0 . There is not convincing evidence that the die is unfair. (Also $5.5 < 11.07$, the critical value.)

EXAMPLE 2

A candy company claims its mix is 30% red, 25% blue, 20% green, 15% yellow, and 10% orange. In a random bag of 200 candies the counts are 52, 55, 36, 38, and 19. Test the claim at $\alpha = 0.05$ and find which color contributes most to χ^2 .

SOLUTION

Expected counts: $200(0.30) = 60, 50, 40, 30, 20$. All are at least 5.

Contributions $\frac{(O-E)^2}{E}$: red $\frac{64}{60} = 1.067$; blue $\frac{25}{50} = 0.5$; green $\frac{16}{40} = 0.4$; yellow $\frac{64}{30} = 2.133$; orange $\frac{1}{20} = 0.05$.

$\chi^2 = 4.15$, $df = 4$, p-value $\approx 0.386 > 0.05$. Fail to reject H_0 : there is not convincing evidence that the color distribution differs from the claim.

Yellow contributes the most (2.133): there were more yellow candies than expected (38 versus 30), but not enough to be convincing.

EXAMPLE 3

A college offers a course in three formats. Pass and fail counts from separate random samples: online 42 pass, 18 fail; hybrid 55 pass, 25 fail; in person 38 pass, 42 fail. Is there convincing evidence that the pass rate differs among formats? Use $\alpha = 0.05$.

SOLUTION

State: H_0 : the distribution of pass/fail is the same for all three formats. H_a : the distribution differs for at least one format.

Plan: Chi-square test for homogeneity (separate samples from each format). Expected counts, for example online-pass = $\frac{135 \times 60}{220} = 36.8$; the smallest expected count is 23.2, all at least 5.

Do: Expected pass counts: 36.8, 49.1, 49.1; fail counts: 23.2, 30.9, 30.9. $\chi^2 \approx 10.21$, $df = (2-1)(3-1) = 2$, p-value ≈ 0.006 .

Conclude: Since $0.006 < 0.05$, reject H_0 . There is convincing evidence that the pass rate is not the same for all three formats. The largest contribution is from the in-person format (fewer passes than expected).

EXAMPLE 4

A random sample of 150 students is classified by whether they play a sport and whether they take music lessons. The counts: sport and music 30, sport and no music 45, no sport and music 50, no sport and no music 25. Is there convincing evidence of an association? Use $\alpha = 0.05$.

SOLUTION

State: H_0 : playing a sport and taking music lessons are independent in the population of students. H_a : they are associated.

Plan: Chi-square test for independence (one sample, two variables). Row totals 75, 75 and column totals 80, 70. Expected counts: $\frac{75 \times 80}{150} = 40$ and $\frac{75 \times 70}{150} = 35$, all at least 5.

Do: Each cell is 10 away from its expected count: $\chi^2 = \frac{100}{40} + \frac{100}{35} + \frac{100}{40} + \frac{100}{35} \approx 10.71$. $df = (2-1)(2-1) = 1$, p-value ≈ 0.001 .

Conclude: Since $0.001 < 0.05$, reject H_0 . There is convincing evidence of an association between playing a sport and taking music lessons for students. In this sample the sport players are less likely to take music.

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

Use the study-time data of the figure ($n = 10$, $b = 6.02$, $SE_b = 0.372$) to test whether there is a positive linear relationship between hours studied and exam score. Use $\alpha = 0.05$.

SOLUTION

State: β is the true slope of the regression line for predicting exam score from hours studied. $H_0 : \beta = 0$, $H_a : \beta > 0$.

Plan: t test for the slope. Linear (the scatterplot is linear), independent, residuals roughly normal with no pattern, equal spread, and the students are a random sample.

Do: $t = \frac{6.02 - 0}{0.372} \approx 16.2$, $df = 10 - 2 = 8$. p-value < 0.0001 (about 10^{-7} for one side).

Conclude: Since the p-value is less than 0.05, reject H_0 . There is convincing evidence of a positive linear relationship between hours studied and exam score for students like these.

EXAMPLE 6

Construct and interpret a 95% confidence interval for the true slope in Example 5. Also interpret $r^2 = 0.970$.

SOLUTION

$df = 8$, $t^* = 2.306$. Margin of error $= 2.306(0.372) \approx 0.858$.

Interval: $6.02 \pm 0.858 = (5.17, 6.88)$.

We are 95% confident that the interval from 5.17 to 6.88 captures the true slope: for each additional hour studied, the mean exam score increases by between about 5.2 and 6.9 points.

Since 0 is not in the interval, it agrees with the test result.

$r^2 = 0.970$: about 97% of the variation in exam scores is explained by the linear relationship with hours studied.

Coch's Playbook

- 1 Decide the test from the design.** One variable compared with a claim: goodness-of-fit. Several groups compared on one variable: homogeneity. One sample cross-classified two ways: independence. Two quantitative variables: slope t test.
- 2 Write hypotheses in words.** For chi-square tests, H_0 is "the distribution is as claimed / the same / independent" and H_a is "not" (no symbols are needed).
- 3 Compute and display expected counts.** Name the smallest one and compare with 5.
- 4 State df and the statistic.** Chi-square: categories minus 1, or $(r - 1)(c - 1)$. Slope: $n - 2$.
- 5 Conclude in context.** Link the p-value to α and use the words "convincing evidence".
- 6 For output,** find the row for the explanatory variable and read the coefficient, standard error, t, and p-value. The printed p-value is usually two-sided; halve it for a one-sided alternative when the sign of b matches.

Mental model: chi-square measures a total squared surprise. Each cell contributes how far it is from expected, relative to the size of the expected count. Looking at the largest contributions tells you **where** the difference is. A chi-square test only says that a difference exists, not which direction is larger.

Common traps

1. Checking the Large Counts condition with observed counts. The condition is that every expected count is at least 5.
2. Using percentages instead of counts in the chi-square formula.
3. Choosing homogeneity when the data came from one sample, or independence when separate samples were drawn from each group. The arithmetic is the same, but the hypotheses are worded differently.
4. Using $df = n - 1$ for the slope. Regression has $df = n - 2$.

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 →

Spot the error 1

A student tests a die with counts 18, 25, 16, 22, 14, 25 and computes χ^2 using percentages: $\frac{(15 - 16.67)^2}{16.67} + \dots$. Where is the error? The chi-square formula requires counts, not percentages. With percentages the statistic loses the sample size and is much too small. Correct expected counts are 20 each, giving $\chi^2 = 5.5$.

Spot the error 2

A student writes for the sport/music table: " $H_0 : p_{\text{sport}} = p_{\text{music}}$. Since $\chi^2 = 10.71$ and the p-value is 0.001, we reject and conclude that sports cause fewer music lessons." Where is the error? The hypotheses for independence are about association: H_0 : the two variables are independent. Also, this is an observational sample, so there is no cause-and-effect conclusion. Correct: "There is convincing evidence of an association between playing a sport and taking music lessons."

Spot the error 3

A student writes: "With 12 data points, the slope test has $df = 12 - 1 = 11$; the t statistic is $b/SE_b = 2.4/0.8 = 3.0$, so the p-value is 0.0121." Where is the error? For regression the degrees of freedom are $n - 2 = 10$ (two parameters, the slope and intercept, are estimated). With $df = 10$ the two-sided p-value is 0.0133. Always use $n - 2$ for slope inference.

Real-World Connection

Public health researchers ask whether vaccination status and the severity of an illness are independent, using a chi-square test on a table of counts. Engineers use the slope test to check whether adding a coating reduces wear in proportion to its thickness. In each case the test decides whether a pattern is real or just chance.

Challenge Corner

In Example 4, suppose every count was doubled (60, 90, 100, 50, from a sample of 300). Without recomputing from scratch, how does χ^2 change, and what is the new p-value size? Solution: Expected counts double, and each $(O - E)^2/E$ doubles (the numerator is multiplied by 4, the denominator by 2). So χ^2 doubles to about 21.43. With $df = 1$ the p-value is about 0.000004. The proportions are unchanged, but a larger sample gives stronger evidence for the same pattern.

Practice**PRACTICE Chi-Square Tests and Regression Inference**

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

- A goodness-of-fit test is run on a variable with 6 categories. How many degrees of freedom does the chi-square statistic have?
(A) 6 (B) 5 (C) 4 (D) 7
- A chi-square test uses a two-way table with 3 rows and 4 columns. What are the degrees of freedom?
(A) 12 (B) 7 (C) 11 (D) 6
- A two-way table has a total of 220, a row total of 85, and a column total of 80. What is the expected count for the cell in that row and column?
(A) 30.9 (B) 36.8 (C) 25.3 (D) 42.5
- A researcher takes one random sample of 400 voters and classifies each by age group and party preference. Which test is appropriate to see whether age group and preference are related?
(A) Goodness-of-fit (B) Two-sample t
(C) Chi-square test for independence (D) Paired t

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. Three random samples are taken, one from each of three schools, and each student is asked to choose a favorite subject. To see whether the distribution of favorite subjects is the same at all three schools, use a
 - (A) chi-square test for independence
 - (B) chi-square test for homogeneity
 - (C) t test for a slope
 - (D) goodness-of-fit test
6. Which condition must be checked for a chi-square test?
 - (A) All observed counts are at least 5
 - (B) The population is normally distributed
 - (C) The sample size is at least 30
 - (D) All expected counts are at least 5
7. A regression slope test is based on 12 data points. How many degrees of freedom are used?
 - (A) 10
 - (B) 11
 - (C) 12
 - (D) 13
8. Computer output for a regression shows $b = 2.4$ and $SE_b = 0.8$. What is the t statistic for $H_0 : \beta = 0$?
 - (A) 0.33
 - (B) 1.6
 - (C) 3.0
 - (D) 19.2
9. A residual plot for a regression shows a clear U shape. Which condition is most clearly violated?
 - (A) Equal standard deviation of residuals
 - (B) Independence
 - (C) Randomness
 - (D) Linearity of the relationship
10. A school claims that lunch choices are 40% pizza, 25% pasta, 20% salad, and 15% wraps. In a random sample of 150 students, 70 chose pizza, 32 pasta, 26 salad, and 22 wraps. Do the data provide convincing evidence that the claim is wrong? Use the four-step process with $\alpha = 0.05$.
11. A survey of separate random samples gives these counts for how people watch videos. Ages 18 to 29: 60 streaming, 20 cable. Ages 30 to 49: 45 streaming, 40 cable. Ages 50 and over: 25 streaming, 50 cable. Test whether the distributions are the same for the three age groups at $\alpha = 0.05$ (give χ^2 , df , and a conclusion).
12. A regression of weight lost on weeks in a program for 12 participants gives $b = 2.4$ and $SE_b = 0.8$. Test $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$ at $\alpha = 0.05$.
13. Using the output from the previous problem ($b = 2.4$, $SE_b = 0.8$, $n = 12$), construct and interpret a 95% confidence interval for the true slope.
14. A residual plot for a regression has a funnel shape: the residuals are tightly clustered for small x values and widely spread for large x values. Which regression condition is violated, and why does it matter for inference?
15. For the sport/music table (counts 30, 45, 50, 25), find the contribution of the cell with 30 to χ^2 . Then find the value of χ^2 if every count is doubled, and explain the change.

Answers**Answers and Explanations**

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

34.1 The Logic of Significance Tests**PRACTICE · THE LOGIC OF SIGNIFICANCE TESTS**

1. C. Hypotheses are about the parameter μ , and H_0 always has equality. Using $\bar{x}$ is the tempting error, and H_a must say greater than 50 for this claim.
2. D. The p-value is computed assuming H_0 is true. The first two options give probabilities of hypotheses, which a p-value is not.
3. A. $0.03 < 0.05$, so reject H_0 and conclude there is convincing evidence for H_a . The p-value is not the probability that H_0 is true.
4. B. Type I means rejecting a true null: concluding the mean differs from 500 mg when it really equals 500 mg. The first and third options describe a Type II error.
5. C. Larger samples reduce variability and increase power. Decreasing α lowers power, and a null value close to the truth makes the difference harder to detect.
6. A. The one-tail area for $|z| = 1.80$ is 0.0359, and it is doubled for a two-sided test: 0.0719. The value 0.0359 is only one tail.
7. B. By definition the Type I error probability equals α , chosen before the data are collected. The p-value is computed from the data.
8. A. $0.12 > 0.05$, so we fail to reject H_0 and say there is not convincing evidence for H_a . We never claim to have proven the null.
9. D. Hypotheses use parameters, state H_0 with equality, and give context. The choices with $\hat{p}$ or $\bar{x}$ use statistics, and the last reverses the roles.
10. p = the true proportion of all residents of the city who recycle. $H_0 : p = 0.40$, $H_a : p < 0.40$.
11. Interpretation: if the true mean reading score is 100, there is a 0.012 probability of getting a sample mean at least this far from 100 (in either direction) by chance alone. Since $0.012 < 0.05$, reject H_0 : there is convincing evidence that the true mean reading score differs from 100.
12. Type I: concluding that a customer is committing fraud when the customer is not; the consequence is that an innocent customer's account may be frozen. Type II: failing to find convincing evidence of fraud when the customer is committing fraud; the consequence is that fraud goes undetected and the bank loses money.
13. Estimated p-value = $9/200 = 0.045$. Since $0.045 < 0.05$, reject H_0 : there is convincing evidence for the alternative hypothesis (in context). The decision is close to the boundary, so the evidence is not overwhelming.

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

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

BOOK NOW →

14. Cutoff = $0.50 + 1.645(0.05) = 0.5822$. When $p = 0.60$: $SD = \sqrt{0.6(0.4)/100} = 0.049$. Power = $P(Z > (0.5822 - 0.60)/0.049) = P(Z > -0.36) \approx 0.64$. Increasing n shrinks the standard deviation of $\hat{p}$, so power increases (for $n = 200$ it is about 0.89).
15. Because 0.34 is greater than 0.05, we fail to reject H_0 . There is not convincing evidence that the true proportion differs from 0.12. We do not say that it is proven because many nearby values (such as 0.11 or 0.13) would also be consistent with the data; a sample can never prove an exact value.

34.2 Tests for Proportions and Means

PRACTICE · TESTS FOR PROPORTIONS AND MEANS

- C. $z = (0.42 - 0.50)/\sqrt{0.5(0.5)/200} = -0.08/0.03536 \approx -2.26$. The value -2.29 comes from using $\hat{p}$ in the standard error, which is the interval formula.
- A. The Large Counts condition for a test uses the null value p_0 : $200(0.05) = 10$ and $200(0.95) = 190$. The choice using $\hat{p}$ is for intervals.
- D. $df = n - 1 = 21$. Using $n = 22$ is the tempting mistake.
- B. $t = (48.2 - 50)/(6/\sqrt{25}) = -1.8/1.2 = -1.5$. The value -1.8 forgets to divide by the standard error.
- A. The p-value for a left-tailed test is $P(Z < -2.26) = 0.0119$. The value 0.0238 doubles it, which is only for a two-sided test.
- D. We use t when the population standard deviation is unknown and replaced by s , which is nearly always. Sample size and skew affect the condition check, not the choice of z versus t .
- C. Since $0.0103 < 0.05$, reject H_0 and conclude in context. A p-value is not the probability that H_0 is true.
- B. Large Counts uses p_0 : $np_0 = 100(0.05) = 5 < 10$. The count of 12 successes is the tempting wrong check.
- D. 16 is outside the interval, so 16 is not a plausible value of μ , and the two-sided test at $\alpha = 0.05$ rejects $H_0 : \mu = 16$.
- State: p is the true proportion of all customers who order the seasonal drink. $H_0 : p = 0.30$, $H_a : p < 0.30$, $\alpha = 0.05$. Plan: z test for a proportion; random sample (given); 10% condition met (large population); $150(0.30) = 45 \geq 10$ and $150(0.70) = 105 \geq 10$. Do: $\hat{p} = 0.24$, $z = (0.24 - 0.30)/\sqrt{0.3(0.7)/150} = -1.60$, p-value ≈ 0.054 . Conclude: since $0.054 > 0.05$, fail to reject H_0 ; there is not convincing evidence that the true proportion of customers who order the seasonal drink is less than 0.30.
- State: μ is the true mean amount per cup. $H_0 : \mu = 8$, $H_a : \mu \neq 8$. Plan: t test; random (given); data roughly symmetric with no outliers for $n = 12$. Do: $t = (7.4 - 8)/(1.1/\sqrt{12}) \approx -1.89$, $df = 11$, two-sided p-value ≈ 0.086 . Conclude: since $0.086 > 0.05$, fail to reject H_0 ; there is not convincing evidence that the true mean amount per cup differs from 8 ounces.
- $t = (102.5 - 100)/(8/\sqrt{40}) = 2.5/1.265 \approx 1.98$, $df = 39$, p-value ≈ 0.028 . Since $0.028 < 0.05$, reject H_0 ; there is convincing evidence that the true mean is greater than 100 minutes.
- Large Counts must be checked with the null value: $np_0 = 45(0.20) = 9 < 10$. The condition fails even though $\hat{p} = 0.30$ gives $45(0.30) = 13.5$, because the sampling distribution assumes H_0 is true.
- The null value 0.50 is outside the interval (0.32, 0.46), so a two-sided test at $\alpha = 0.01$ would reject H_0 : there is convincing evidence that the true proportion differs from 0.50, and it is probably less than 0.50 because the interval lies below it.
- $z = (0.503 - 0.50)/\sqrt{0.25/10000} = 0.003/0.005 = 0.6$. The two-sided p-value is 0.549, so we fail to reject H_0 . Even with a very large sample, a 0.3-point difference is small and not statistically distinguishable from 0.50; and even a significant result of this size would be practically unimportant.

34.3 Comparing Two Groups

PRACTICE · COMPARING TWO GROUPS

- B. $\hat{p}_c = (45 + 30)/(100 + 80) = 75/180 \approx 0.4167$. The value 0.4125 is the average of the two sample proportions, which is not the pooled value.
- D. The same students are measured twice, so the data are paired. The other situations involve two separate groups.
- A. Variances of independent estimates add: $SE = \sqrt{s_1^2/n_1 + s_2^2/n_2}$. A difference of standard errors is the tempting wrong form.
- C. The procedure needs independence between the groups (random samples or random assignment) and approximate normality or large samples. The groups need not be the same size.
- B. Since 0 is in the interval, a difference of 0 is plausible, so there is not convincing evidence of a difference. A confidence level is not a probability about a single interval.
- C. $\hat{p}_c = 0.48$, $SE = \sqrt{0.48(0.52)(2/100)} = 0.0707$, $z = 0.16/0.0707 \approx 2.26$. The value 0.16 is the difference only.
- A. The conservative choice is $\min(15 - 1, 12 - 1) = 11$.
- D. Random assignment makes the groups comparable at the start, so a significant difference can be attributed to the treatment. It does not guarantee equal means, and it does not allow generalizing to people not in the experiment.
- C. Paired procedures use the n differences with $df = n - 1 = 13$. The value 28 would arise from treating the data as two samples.
- State: p_A and p_B are the true proportions of residents who support the proposal in the two cities. $H_0 : p_A = p_B$, $H_a : p_A \neq p_B$. Plan: two-proportion z test; independent random samples; pooled $\hat{p}_c = 225/550 = 0.409$; all counts above 10. Do: $\hat{p}_A = 0.45$, $\hat{p}_B = 0.36$, $SE = \sqrt{0.409(0.591)(1/300 + 1/250)} = 0.0421$, $z = 0.09/0.0421 \approx 2.14$, p-value ≈ 0.033 . Conclude: since $0.033 < 0.05$, reject H_0 ; there is convincing evidence that the proportion of residents who support the proposal differs between the two cities.
- $SE = \sqrt{9.1^2/30 + 8.2^2/28} = \sqrt{2.760 + 2.401} = 2.272$. $t = (78.4 - 73.9)/2.272 \approx 1.98$, $df \approx 55.9$, one-sided p-value ≈ 0.026 . Since $0.026 < 0.05$, reject H_0 ; there is convincing evidence that the true mean score with method A is higher than with method B.
- μ_d is the true mean improvement. $df = 8$, $t^* = 1.860$, $SE = 2.4/\sqrt{9} = 0.8$, margin of error = $1.860(0.8) = 1.49$. Interval: (1.71, 4.69). We are 90% confident that the interval from 1.71 to 4.69 points captures the true mean improvement for all such students. The interval is entirely above 0, so the program appears to raise scores.

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

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

BOOK NOW →

13. (a) Two-sample t: two independent groups of phones with a quantitative response. (b) Two-proportion z: two independent groups with a categorical response (cured or not). (c) Paired t: the same plants are measured twice, so use the differences.
14. We are 95% confident that the interval from -1.2 to 5.6 captures the true difference $\mu_1 - \mu_2$ in means. Since 0 is inside the interval, there is not convincing evidence of a difference between the population means.
15. $\hat{p}_1 = 0.60$, $\hat{p}_2 = 0.50$, $SE = \sqrt{0.6(0.4)/120 + 0.5(0.5)/110} = 0.0654$, $m = 1.96(0.0654) = 0.128$. Interval $(-0.028, 0.228)$. We are 95% confident that this interval captures the true difference (first community minus second) in proportions who exercise weekly. Since 0 is in the interval, there is not convincing evidence of a difference.

34.4 Chi-Square Tests and Regression Inference

PRACTICE · CHI-SQUARE TESTS AND REGRESSION INFERENCE

1. B. For goodness-of-fit, $df = \text{categories} - 1 = 5$. The value 6 forgets to subtract 1.
2. D. $df = (r - 1)(c - 1) = (3 - 1)(4 - 1) = 6$. The product 12 is the number of cells, a tempting mistake.
3. A. Expected count $= \frac{85 \times 80}{220} = 30.9$. The other values come from using the wrong totals.
4. C. One sample with two categorical variables is tested for association with a chi-square test for independence. Goodness-of-fit uses one variable.
5. B. Separate samples from several populations compared on one variable require a test for homogeneity. Independence is for one sample classified two ways.
6. D. The Large Counts condition is on the expected counts, not the observed counts and not a sample size of 30.
7. A. For slope inference, $df = n - 2 = 10$. The value 11 is the tempting $n - 1$.
8. C. $t = b/SE_b = 2.4/0.8 = 3.0$. The value 0.33 inverts the ratio.
9. D. A curved (U-shaped) pattern in the residuals means the relationship is not linear, so a straight-line model is not appropriate. A funnel shape would signal unequal standard deviation.
10. State: H_0 : the distribution of lunch choices is 40% pizza, 25% pasta, 20% salad, 15% wraps. H_a : at least one proportion differs. Plan: chi-square goodness-of-fit; random sample (given); expected counts 60, 37.5, 30, 22.5 are all at least 5. Do: contributions 1.667, 0.807, 0.533, 0.011 give $\chi^2 \approx 3.02$, $df = 3$, p-value ≈ 0.389 . Conclude: since $0.389 > 0.05$, fail to reject H_0 ; there is not convincing evidence that the claimed distribution of lunch choices is wrong.
11. Homogeneity test. Row totals 130 (streaming), 110 (cable); column totals 80, 85, 75; total 240. Expected streaming: 43.3, 46.0, 40.6; cable: 36.7, 39.0, 34.4 (all at least 5). $\chi^2 \approx 27.15$, $df = 2$, p-value < 0.001 . Reject H_0 : there is convincing evidence that the distribution of how people watch videos differs among the three age groups. Younger groups prefer streaming more than expected.
12. $t = 2.4/0.8 = 3.0$, $df = 12 - 2 = 10$, two-sided p-value ≈ 0.013 . Since $0.013 < 0.05$, reject H_0 : there is convincing evidence of a linear relationship between weeks in the program and weight lost.
13. $df = 10$, $t^* = 2.228$, margin of error $= 2.228(0.8) = 1.78$. Interval $(0.62, 4.18)$. We are 95% confident that the interval from 0.62 to 4.18 captures the true slope: the mean weight lost increases by between 0.62 and 4.18 units per additional week. Since 0 is not in the interval, the slope is significantly different from 0.
14. The equal standard deviation condition is violated: the spread of the residuals changes with x . Because the standard error of the slope assumes a constant spread, the t procedures (tests and intervals) may be unreliable. A transformation of the data may help.
15. Expected count for that cell is $\frac{75 \times 80}{150} = 40$, so the contribution is $\frac{(30 - 40)^2}{40} = 2.5$. Doubling every count doubles every expected count and multiplies every squared difference by 4, so each contribution doubles and χ^2 doubles to about 21.43. The pattern is the same, but a larger sample gives stronger evidence.

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

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

BOOK NOW →