STATISTICS · STUDY GUIDE

Exam 5 Study Guide

This guide covers chapters 9 and 10, the two core tools of inference. You will build confidence intervals for means and proportions, choose between z and t, plan a sample size and read an interval the right way. Then you will run the logic of a hypothesis test: state the hypotheses, compute a p-value, decide with α, weigh Type I and Type II errors and power, and separate statistical from practical significance. Work one chapter at a time, try each worked example with the steps covered, and finish with the ten self-test items. The cram sheet at the end holds every formula.

WHAT THIS GUIDE COVERS

- Chapter 9 · point estimates and the CI skeleton, z and t intervals for a mean, the interval for a proportion, the three dials of width, sample size planning and the correct interpretation of confidence.
- Chapter 10 · null and alternative hypotheses, one and two tails, the p-value and what it is not, α and the decision rule, Type I and Type II errors and power, and statistical versus practical significance.

CHAPTER 9

Confidence Intervals

Read it in full: **Chapter 9 in the textbook**. Practice it in the lab: **Confidence Interval for a Proportion** and **What 95 Percent Confidence Means**.

What to know

A **point estimate** is a single number from a sample offered as a guess for a parameter, and because of sampling error it is almost certainly a little off. A **confidence interval** reports a range with a stated confidence level. Every interval at this level has the same skeleton: point estimate ± margin of error, and **margin of error = critical value × standard error**. The interval works because the sampling distribution of the estimate is centered on the truth; the interval is that distribution read backwards.

An interval has a center (the estimate), a half-width (the margin of error) and two endpoints, so its full width is twice the margin of error. It describes a parameter, never an individual: an interval for mean sleep says nothing about any one person's sleep.

When σ is truly known, the **z interval** is $\bar{x} \pm z^*(\sigma \div \sqrt{n})$, with $z^* = 1.96$ at 95%. The book's bottling line with $\sigma = 4$ mL, $n = 36$ and $\bar{x} = 502$ gives 500.69 to 503.31 mL, so a true mean of 500 is not plausible at the 5% level.

When σ is unknown, which is almost always, use the **t interval** $\bar{x} \pm t^*(s \div \sqrt{n})$ with $df = n - 1$. The t distribution (Gosset, "Student," 1908) has heavier tails to pay for estimating σ, so t^* is larger than z^*; at 95% t^* is 2.776 for df = 4, 2.064 for df = 24 and 1.984 for df = 99, approaching 1.96. The rule: if you had to compute s, use t. Conditions: a random sample of independent observations (n under 10% of the population), and a roughly normal population or a large enough n (30 or more as a rule of thumb, more for skew or outliers).

For a **proportion**, $\hat{p} \pm z^*\sqrt{[\hat{p}(1 - \hat{p}) \div n]}$, using z, not t. Check random sampling, the 10% rule and the **success-failure condition**: $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$. The book's 148 of 400 gives 0.323 to 0.417. A margin of 4.7 **percentage points** is an additive move on the percent scale, not 4.7% of 37%. Polls quote the worst case $\hat{p} = 0.5$, which is why about 1,000 people gives ±3 points.

Three dials set the width. Higher confidence means a larger critical value and a wider interval ($z^* = 1.645$, 1.96 and 2.576 for 90, 95 and 99%). Larger n shrinks the standard error, but only at the rate $1 \div \sqrt{n}$: quadruple n to halve the margin. More variability means a wider interval.

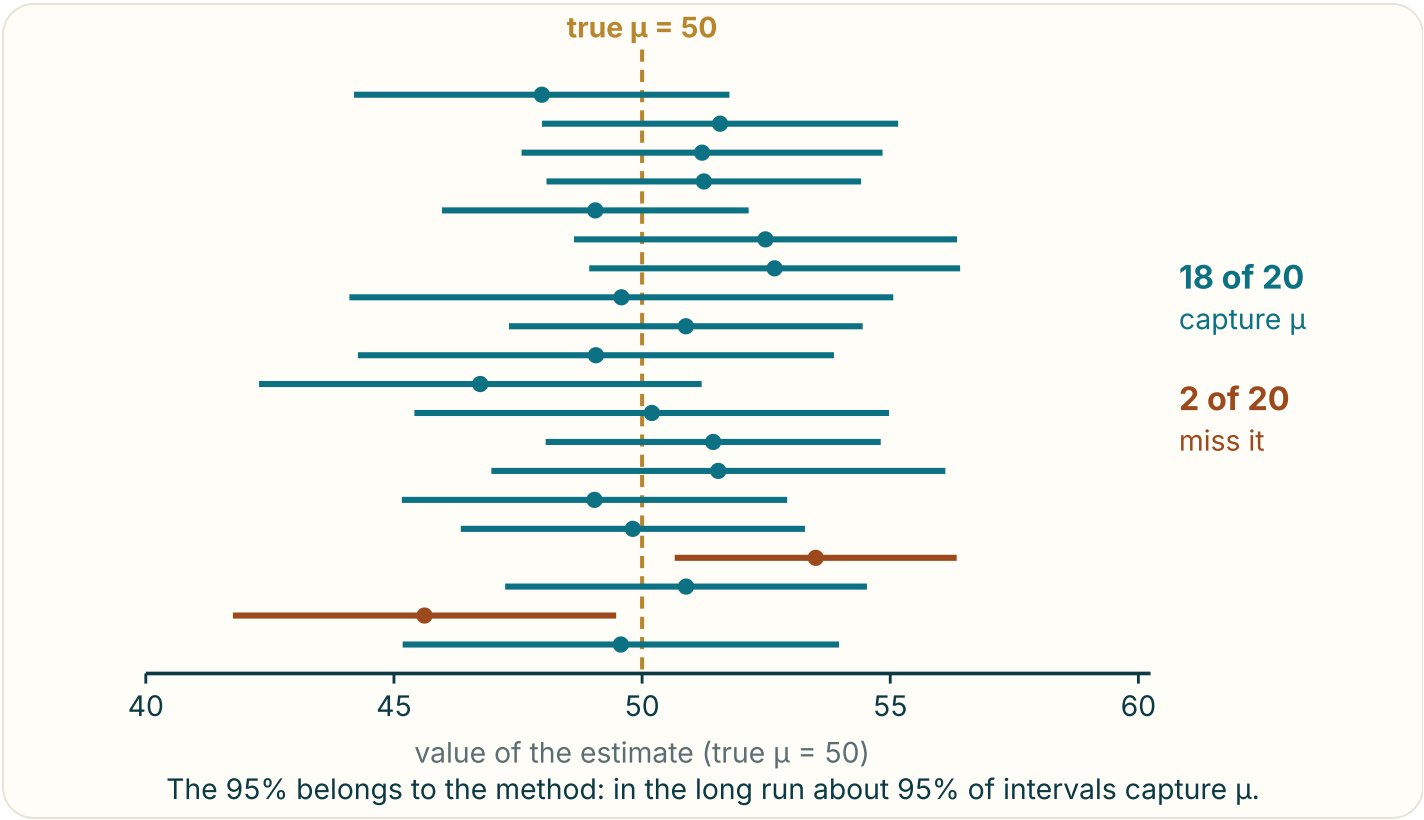
To plan a study, solve for n and **always round up**. For a mean, $n = (z^*\sigma \div E)^2$; $\sigma \approx 12$ and $E = 2$ at 95% give 138.3, so 139. For a proportion, $n = z^{*2}\, p(1 - p) \div E^2$, using $p = 0.5$ when nothing is known; $E = 0.03$ at 95% gives 1,067.1, so 1,068.

The correct sentence: "We are 95% confident that the interval from a to b contains the true [parameter in context]." The 95% belongs to the **procedure**: about 95% of intervals built this way capture the parameter. Your one interval either contains μ or it does not. It is wrong to say there is a 95% probability μ is in it, that 95% of data values fall in it, or that 95% of future sample means will. A value inside the interval is plausible, not proven, and "fail to reject" never means "accept."

If a 95% interval excludes a value, a two-sided test at $\alpha = 0.05$ rejects it. The margin of error covers random sampling only, never bias.

LOCK THESE IN

- CI = estimate ± critical value × SE; width = 2 × margin of error.
- σ known: $\bar{x} \pm z^*\sigma/\sqrt{n}$. σ unknown: $\bar{x} \pm t^*s/\sqrt{n}$ with df = n − 1.
- Proportion: $\hat{p} \pm z^*\sqrt{[\hat{p}(1-\hat{p})/n]}$; need $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.
- $z^* = 1.645$ (90%), 1.96 (95%), 2.576 (99%).
- Quadruple n to halve the margin. Higher confidence, wider interval.
- $n = (z^*\sigma/E)^2$ for a mean; $n = z^{*2}p(1-p)/E^2$ for a proportion; round up.
- "95% confident" describes the method; this interval either contains μ or not.
- A CI excluding a value matches a two-sided test rejecting it at $\alpha = 1 - C$.



Confidence is about the procedure. Twenty samples of n = 25 from a population with μ = 50 give twenty different 95% intervals. 18 capture μ and 2 miss; you never know which kind yours is.

Formulas and when to use them

Interval or plan	Formula	When to use it
Skeleton	estimate ± (critical value)(SE)	Every interval
Mean, σ known	$\bar{x} \pm z^*(\sigma \div \sqrt{n})$	σ given from outside the data
Mean, σ unknown	$\bar{x} \pm t^*(s \div \sqrt{n})$, df = n − 1	You computed s from the sample
Proportion	$\hat{p} \pm z^*\sqrt{[\hat{p}(1-\hat{p}) \div n]}$	$n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$
Margin from an interval	E = (upper − lower) ÷ 2	Reading a reported CI
Estimate from an interval	center = (upper + lower) ÷ 2	Reading a reported CI
n for a mean	$n = (z^*\sigma \div E)^2$	Planning; round up
n for a proportion	$n = z^{*2}\, p(1-p) \div E^2$	Use p = 0.5 if unknown; round up

Critical values

Confidence	z*	t* (df = 24)
90%	1.645	1.711
95%	1.960	2.064
99%	2.576	2.797
95%, df = 4	1.960	2.776
95%, df = 9	1.960	2.262
95%, df = 99	1.960	1.984

Worked examples

WORKED EXAMPLE 1 · A Z INTERVAL WITH Σ KNOWN

A scale's long-run σ is 5 g. A random sample of n = 49 packages has $\bar{x}$ = 68.3 g. Build a 95% interval for μ.

STEP 1 · SE

$\sigma \div \sqrt{n} = 5 \div 7 = 0.7143$.

STEP 2 · MARGIN

$1.96 \times 0.7143 =$ **1.40**.

STEP 3 · INTERVAL

68.3 ± 1.40: **66.90 to 69.70 g**.

STEP 4 · SAY IT

We are 95% confident that the interval from 66.90 to 69.70 g contains the true mean package weight.

WORKED EXAMPLE 2 · A T INTERVAL WITH Σ UNKNOWN

A random sample of $n = 16$ commuters has $\bar{x} = 42.5$ minutes and $s = 6.4$. Build a 95% interval and compare it with a careless z interval.

STEP 1 · DF AND SE

$df = 15$; $SE = 6.4 \div \sqrt{16} = 6.4 \div 4 = 1.60$.

STEP 2 · CRITICAL VALUE

t^* (95%, $df = 15$) = **2.131**.

STEP 3 · MARGIN AND INTERVAL

$2.131 \times 1.60 = 3.41$, so **39.09 to 45.91** minutes.

STEP 4 · THE Z MISTAKE

$1.96 \times 1.60 = 3.14$ gives 39.36 to 45.64, too narrow and too confident.

CHECK THE CONDITIONS

Random sample, 16 out of a large population, and with $n = 16$ the commute times should look roughly normal with no strong outliers.

WORKED EXAMPLE 3 · A PROPORTION AT TWO CONFIDENCE LEVELS

In a random sample of 500 adults, 215 say they read a book last month. Build 95% and 99% intervals for p.

STEP 1 · ESTIMATE AND CHECKS

$\hat{p} = 215 \div 500 = 0.43$. $n\hat{p} = 215$ and $n(1 - \hat{p}) = 285$, both at least 10.

STEP 2 · SE

$\sqrt{(0.43 \times 0.57 \div 500)} = \sqrt{0.000490} \approx 0.0221$.

STEP 3 · 95%

$1.96 \times 0.0221 = 0.0434$, so **0.387 to 0.473**, a margin of 4.3 percentage points.

STEP 4 · 99%

2.576 × 0.0221 = 0.0570, so **0.373 to 0.487**. More confidence, wider interval.

READ IT

Both intervals sit below 0.50, so "half of adults" is not a plausible value at either level.

WORKED EXAMPLE 4 · PLANNING THE SAMPLE SIZE

A team wants a 95% interval for mean daily screen time within E = 2 minutes, with σ ≈ 15 from a pilot. A second team wants a proportion within 4 points. Find each n.

STEP 1 · MEAN

$n = (1.96 \times 15 \div 2)^2 = (14.70)^2 = 216.09$, so **217**.

STEP 2 · PROPORTION, NOTHING KNOWN

$n = 1.96^2 \times 0.25 \div 0.04^2 = 0.9604 \div 0.0016 = 600.25$, so **601**.

STEP 3 · PROPORTION WITH A PRIOR GUESS P ≈ 0.20

$n = 1.96^2 \times 0.20 \times 0.80 \div 0.0016 = 384.16$, so **385**.

WHY ROUND UP

Rounding down leaves the margin slightly larger than promised. Plan a little extra for dropouts.

TRAPS PEOPLE FALL INTO

- **Saying "95% chance μ is in this interval."** μ is fixed. The 95% describes how often the method works.
- **Reading the CI as the range of the data.** An interval for a mean is far narrower than the spread of individuals.
- **Using z after computing s.** If s came from the sample, use t with df = n − 1.
- **Using t for a proportion.** The proportion interval uses z.
- **Calling the width the margin of error.** The width is twice the margin.
- **Mixing percent and percentage points.** A 4-point margin on 37% runs 33% to 41%.
- **Rounding a planned n down.** Always round up to keep the promised margin.
- **Thinking a bigger n removes bias.** The margin of error covers chance only.

Chapter 9 self-test · ten items

1. A poll reports 46% with a margin of error of 3 points. What is the interval?

A. 44.5% to 47.5%

B. 43% to 49%

C. 45.9% to 46.1%

D. 40% to 52%

B. 43% to 49% Estimate ± margin: 46 − 3 = 43 and 46 + 3 = 49. The interval is 6 points wide.

2. A sample of n = 64 with known σ = 16 has x̄ = 120. What is the 95% interval?

A. 118.04 to 121.96

B. 116.08 to 123.92

C. 88.64 to 151.36

D. 116.71 to 123.29

B. 116.08 to 123.92 SE = 16 ÷ 8 = 2; margin = 1.96 × 2 = 3.92.

3. You compute s from a sample of n = 12. Which multiplier do you use at 95%?

A. z* = 1.96

B. t* with df = 12, about 2.179

C. t* with df = 11, about 2.201

D. z* = 1.645

C. t* with df = 11, about 2.201 σ is unknown, so use t with df = n − 1 = 11.

4. n = 9, $\bar{x}$ = 30, s = 6. What is the 95% t interval?

A. 25.39 to 34.61
B. 26.08 to 33.92
C. 16.16 to 43.84
D. 28.00 to 32.00

A. 25.39 to 34.61 SE = 6 ÷ 3 = 2, t* (df = 8) = 2.306, margin = 4.61.

5. In a sample of 200, 50 have a trait. What is the 95% margin of error?

A. 0.060
B. 0.031
C. 0.250
D. 0.050

A. 0.060 $\hat{p}$ = 0.25, SE = $\sqrt{(0.25 \times 0.75 \div 200)} \approx 0.0306$, and $1.96 \times SE \approx 0.060$.

6. Which change makes a confidence interval narrower?

A. Raising confidence from 95% to 99%
B. Increasing the sample size
C. A more spread-out population
D. Switching from z* to t*

B. Increasing the sample size A larger n shrinks the standard error. The other changes widen the interval.

7. To cut a margin of error in half, the sample size must be:

A. Doubled
B. Multiplied by √2
C. Halved
D. Multiplied by 4

D. Multiplied by 4 The margin falls as $1 \div \sqrt{n}$, so n must grow by $2^2 = 4$.

8. What sample size gives a 95% margin of 0.05 for a proportion with no prior guess?

A. 384
B. 400
C. 385
D. 196

C. 385 $1.96^2 \times 0.25 \div 0.05^2 = 384.16$; round up to 385.

9. Which is the correct reading of a 95% interval of 12.1 to 15.7?

A. There is a 95% chance the true mean is in 12.1 to 15.7
B. 95% of the data values lie between 12.1 and 15.7
C. About 95% of intervals built this way contain the true mean
D. 95% of future sample means will land in 12.1 to 15.7

C. About 95% of intervals built this way contain the true mean The confidence level describes the long-run success of the procedure, not a probability for one fixed interval.

10. A 95% interval for a mean is 48.2 to 53.9. What does a two-sided test of μ = 55 at α = 0.05 conclude?

A. Fail to reject, since 55 is close to the interval
B. Accept μ = 55 as the true population mean
C. No conclusion is possible without the p-value
D. Reject, since 55 lies outside the interval

D. Reject, since 55 lies outside the interval Values outside a 95% interval are rejected by the matching two-sided test at α = 0.05.

CHAPTER 10

The Logic of Hypothesis Testing

Read it in full: **Chapter 10 in the textbook**. Practice it in the lab: **State the Hypotheses** and **Type I and Type II Errors**.

What to know

A hypothesis test works like a criminal trial. The **null hypothesis** H0 is the dull, no-effect explanation, presumed true until the evidence is strong. The **alternative** Ha is the claim you care about, and it carries the burden of proof. Weak evidence does not show H0 is true; it means the case was not made.

Four rules for writing hypotheses. They are about **parameters** (μ, p, σ), never statistics such as $\bar{x}$ or $\hat{p}$. The equals sign lives in H0, because a specific value makes the probability model possible. H0 and Ha must not overlap. The direction is fixed **before** seeing the data: two-tailed (≠) unless a one-tailed direction (< or >) was chosen in advance for a real reason. When in doubt, go two-tailed.

The **p-value** is P(a test statistic at least this extreme | H0 true and the assumptions hold). Compute it in three moves: standard error, test statistic, tail area, doubling the tail for a two-sided Ha. In the book's commute study, H0: μ = 20, n = 36, $\bar{x}$ = 22.4, s = 6.0: SE = 1.00, t = 2.40 with df = 35, and

$p \approx 2 \times 0.011 = 0.022$. The one-sample t test assumes random sampling, independence and an approximately normal sampling distribution.

A p-value is not the probability the null is true, $1 - p$ is not the probability Ha is true, it does not measure how big the effect is, it is not a replication rate, and a large p does not prove no effect. A small p-value indicts the whole model, so check the sampling and assumptions first.

The **significance level** α is a false-alarm budget set in advance, usually 0.05. Decision rule: if $p \leq \alpha$, reject H0; if $p > \alpha$, **fail to reject** H0, never "accept." The critical-value route always agrees: with df = 35 and two-tailed $\alpha = 0.05$, $t^* = 2.030$, and $t = 2.40$ lands in the rejection region. Had $\bar{x}$ been 21.6, $t = 1.60$ and $p \approx 0.119$, so fail to reject. And 0.049 and 0.051 are the same result in practice.

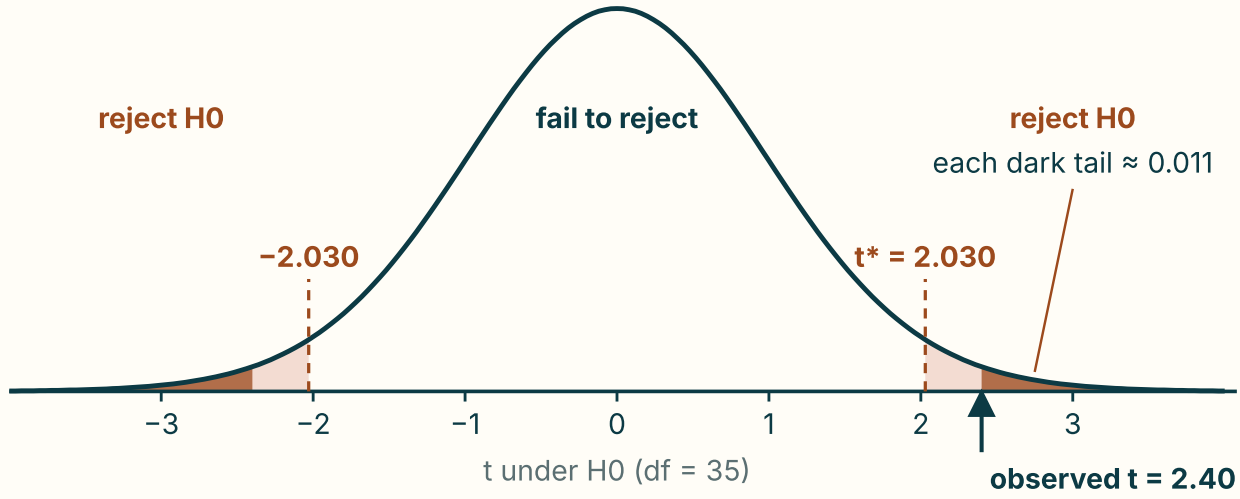
A **Type I error** rejects a true H0 (a false alarm), with probability α. A **Type II error** fails to reject a false H0 (a miss), with probability β. **Power** $= 1 - \beta$, the chance of detecting a specific effect size; 0.80 is the usual floor. With $\sigma = 6$, $n = 36$ and a true mean of 21.5, the commute test has power of only about 0.32; quadrupling n to 144 lifts it to about 0.85.

Power rises with larger n, a larger true effect, less variability and a larger α. Lowering α raises β, so you cannot shrink both errors by moving α; only a bigger or cleaner sample improves both. Which error is worse is a judgment about consequences: screening favors power, courts favor avoiding false convictions. A nonsignificant result from an underpowered study means very little.

Because $SE = s \div \sqrt{n}$, a big enough sample makes any nonzero difference **statistically significant**. That answers "not zero?" but never "big enough to matter?" The book's Study A ($n = 40{,}000$) found a 0.20 pound change with $p < 0.0001$, CI 0.10 to 0.30 and $d = 0.02$: significant and trivial. Study B ($n = 25$) found 4.0 pounds with $p \approx 0.057$ and CI -0.13 to 8.13: too small to say. Report the estimate, the CI and an effect size such as Cohen's d = difference ÷ s, along with the exact p. Significance never shows causation; only the design can.

LOCK THESE IN

- H0 has the equals sign; hypotheses are about parameters.
- Choose one or two tails before seeing the data; default to two.
- p-value = P(result this extreme or more | H0 true).
- $p \le \alpha$: reject H0. $p > \alpha$: fail to reject. Never "accept."
- Type I = false alarm (α). Type II = miss (β). Power = $1 - \beta$.
- Power rises with n, effect size and α, and falls with σ.
- Test statistic = (estimate − null value) ÷ SE.
- Statistical significance is not practical importance, and not causation.



p-value = 2 × 0.0109 ≈ 0.022 ≤ α = 0.05, so reject H0.

Two routes to one verdict. For the commute test, t = 2.40 lies beyond the critical value 2.030, and the two dark tails add to p ≈ 0.022, below α = 0.05.

Formulas and when to use them

Quantity	Formula	When to use it
Test statistic (mean)	$t = (\bar{x} - \mu_0) \div (s \div \sqrt{n})$, df = n − 1	σ unknown
Test statistic (σ known)	$z = (\bar{x} - \mu_0) \div (\sigma \div \sqrt{n})$	σ given
Two-sided p-value	2 × (tail area beyond $\lvert t \rvert$)	Ha: ≠
One-sided p-value	One tail in the direction of Ha	Ha: < or >, chosen in advance
Decision	Reject H0 if $p \le \alpha$	α fixed beforehand
Rejection cutoffs (power)	$\mu_0 \pm z^*(\sigma \div \sqrt{n})$	Power calculations
Power	P(reject H0 | specific true μ)	Planning; aim for 0.80 or more
Cohen's d	$d = (\bar{x} - \mu_0) \div s$	Effect size beside the p-value

The decision table

Decision	H0 actually true	H0 actually false
Reject H0	Type I error (probability α)	Correct (power = 1 − β)
Fail to reject H0	Correct (1 − α)	Type II error (probability β)

Worked examples

WORKED EXAMPLE 1 · A ONE-SAMPLE T TEST FROM START TO FINISH

A coffee chain says its drinks average 50 mg of caffeine per cup. A lab tests a random sample of n = 25 cups: $\bar{x}$ = 53.1 mg, s = 7.5. Test at α = 0.05, two-tailed.

STEP 1 · HYPOTHESES

H0: μ = 50; Ha: μ ≠ 50. Two-tailed, chosen before the data.

STEP 2 · SE

$s \div \sqrt{n} = 7.5 \div 5 = 1.50$.

STEP 3 · TEST STATISTIC

t = (53.1 − 50) ÷ 1.50 = 3.1 ÷ 1.50 = **2.067**, df = 24.

STEP 4 · P-VALUE

One tail beyond 2.07 is 0.0249; doubled, **p ≈ 0.050**.

STEP 5 · DECIDE

0.050 is at or below 0.05, so **reject H0**. Critical-value route: t* = 2.064, and 2.067 is beyond it.

STEP 6 · IN CONTEXT

If the true mean really were 50 mg, a sample mean this far away would happen only about 5.0% of the time. The data suggest the mean differs from 50 mg.

WORKED EXAMPLE 2 · THE SAME DATA WITH A ONE-TAILED ALTERNATIVE

Suppose, before collecting data, a regulator worried only that cups contain more than 50 mg. Redo the decision with Ha: μ > 50.

STEP 1 · HYPOTHESES

H0: μ ≤ 50; Ha: μ > 50, set in advance for a stated reason.

STEP 2 · P-VALUE

Same t = 2.067, but only the upper tail counts: **p ≈ 0.0249**.

STEP 3 · DECIDE

0.0249 ≤ 0.05: reject H0.

THE WARNING

Choosing one tail after seeing that $\bar{x}$ came out high would quietly double the false-alarm rate. The direction has to come first.

WORKED EXAMPLE 3 · POWER BEFORE AND AFTER QUADRUPLING N

Test H0: μ = 100 against Ha: μ ≠ 100 at α = 0.05 with known σ = 10. If the true mean is 104, find the power for n = 25 and n = 100.

STEP 1 · N = 25 CUTOFFS

SE = 10 ÷ 5 = 2. Reject if $\bar{x} > 100 + 1.96(2) = 103.92$ or $\bar{x} < 96.08$.

STEP 2 · IN THE TRUE WORLD

Centered at 104: z = (103.92 − 104) ÷ 2 = −0.04, so P(Z > −0.04) = 0.5160. Lower side: z = −3.96, area 0.0000.

STEP 3 · POWER

0.5160 + 0.0000 ≈ **0.52**, so β ≈ 0.48. A coin flip.

STEP 4 · N = 100

SE = 1. Reject if $\bar{x} > 101.96$. $z = (101.96 - 104) \div 1 = -2.04$, $P = 0.9793$; power ≈ **0.98**.

READ IT

Four times the sample halved the SE and lifted power from about half to nearly certain detection of a 4-point shift.

WORKED EXAMPLE 4 · SIGNIFICANT BUT TRIVIAL

An app study of n = 10,000 users finds mean daily use rose by 0.3 minutes, s = 12. Test H0: $\mu = 0$ and judge the result.

STEP 1 · SE

$12 \div \sqrt{10{,}000} = 12 \div 100 = 0.12$.

STEP 2 · TEST STATISTIC

$t = 0.3 \div 0.12 =$ **2.50**; two-sided $p \approx 0.0124$. Significant at 0.05.

STEP 3 · INTERVAL

$0.3 \pm 1.96(0.12) = 0.3 \pm 0.235$: **0.065 to 0.535 minutes**.

STEP 4 · EFFECT SIZE

$d = 0.3 \div 12 =$ **0.025**, negligible.

STEP 5 · VERDICT

Real, but at most about half a minute a day: statistically significant and practically trivial. Report "0.3 min, 95% CI 0.06 to 0.54, d = 0.025, p = 0.012."

TRAPS PEOPLE FALL INTO

- **Writing hypotheses about $\bar{x}$ or $\hat{p}$.** You can see the statistic. Hypotheses are about μ or p.
- **Putting the inequality in H0.** The equals sign belongs in the null.
- **Choosing one tail after peeking.** It doubles the false-alarm rate. Decide first.
- **Reading p as P(H0 true).** p is P(data | H0), not P(H0 | data).
- **Saying "accept H0."** Say "fail to reject." The test cannot confirm the null.
- **Treating p = 0.051 as "no effect."** It is weak evidence; look at the interval.
- **Ignoring power.** A null result from a low-power study tells you little.
- **Equating significant with important or causal.** Check the effect size, and remember cause comes from the design.

Chapter 10 self-test · ten items

1. Which pair of hypotheses is written correctly?

A. H0: $\bar{x} = 12$; Ha: $\bar{x} \neq 12$

B. H0: $\mu \neq 12$; Ha: $\mu = 12$

C. H0: $\mu = 12$; Ha: $\mu \neq 12$

D. H0: $\mu > 12$; Ha: $\mu < 12$

C. H0: $\mu = 12$; Ha: $\mu \neq 12$ Hypotheses concern the parameter, and the equality belongs in H0.

2. n = 49, $\bar{x}$ = 104, s = 14, H0: $\mu = 100$. What is the test statistic?

A. 2.00

B. 0.29

C. 14.00

D. 0.57

A. 2.00 SE = 14 ÷ 7 = 2, so t = (104 − 100) ÷ 2 = 2.00.

3. A test gives p = 0.03. Which statement is correct?

A. Under H0, results this extreme occur about 3% of the time

B. There is a 3% chance that H0 is true for this population

C. There is a 97% chance that Ha is the correct hypothesis

D. The effect must be large, because the p-value is small

A. Under H0, results this extreme occur about 3% of the time The p-value is computed assuming H0; it is not a probability about H0 itself.

4. α = 0.05 and p = 0.08. What is the decision?

A. Reject H0 at 0.05
B. Accept H0 as true
C. Accept Ha as true
D. Fail to reject H0

D. Fail to reject H0 0.08 > 0.05, so the evidence is not strong enough. Never say "accept."

5. What is a Type I error?

A. Failing to reject H0 when it is false
B. Rejecting H0 when H0 is true
C. Choosing the wrong number of tails
D. Using z when t is needed

B. Rejecting H0 when H0 is true A Type I error is a false alarm, and its probability is α.

6. Power is the probability of:

A. Rejecting H0 when H0 is actually true
B. Failing to reject H0 when it is actually false
C. Rejecting H0 when a specific alternative is true
D. Getting a p-value under 0.05 purely by chance

C. Rejecting H0 when a specific alternative is true Power = 1 − β, stated against a particular true effect.

7. Which change raises power without raising the Type I error rate?

A. Raising α from 0.01 to 0.05
B. Using a two-tailed test
C. A noisier measurement tool
D. A larger sample size

D. A larger sample size More data shrinks the SE. Raising α raises power but also raises false alarms.

8. A study with n = 20 finds p = 0.30. What can you conclude?

A. There is definitely no effect at all
B. Weak evidence; a real effect may still exist
C. H0 is true with a probability of 0.30
D. Ha is true with a probability of 0.70

B. Weak evidence; a real effect may still exist A large p-value from a small study is a failure to detect, not proof of absence.

9. With df = 20, two-tailed α = 0.05 has t* = 2.086. A test gives t = −2.30. Decision?

A. Reject H0, because |t| exceeds 2.086
B. Fail to reject, because t is negative
C. Reject only with a one-tailed test
D. Fail to reject, because 2.30 < 2.086

A. Reject H0, because |t| exceeds 2.086 In a two-tailed test, either tail counts. |−2.30| = 2.30 > 2.086.

10. A study of 100,000 people finds a 0.1 point difference with p < 0.001. What is the best reading?

A. A large effect that is clearly important
B. Proof that one factor causes the other
C. A Type I error, since n is very large
D. Probably real but possibly too small to matter

D. Probably real but possibly too small to matter Huge samples make tiny differences significant. Check the effect size and CI.

ONE PAGE

Cram sheet

Everything from chapters 9 and 10 in one place. Cover the right column and see if you can rebuild each row.

Formulas

Name	Formula	Note
CI skeleton	estimate ± critical value × SE	Width = 2 × margin
Mean, σ known	$\bar{x} \pm z^* \sigma/\sqrt{n}$	z^* 1.645, 1.96, 2.576
Mean, σ unknown	$\bar{x} \pm t^* s/\sqrt{n}$	df = n − 1
Proportion	$\hat{p} \pm z^* \sqrt{[\hat{p}(1-\hat{p})/n]}$	$n\hat{p}$, $n(1-\hat{p}) \geq 10$
Sample size	$n = (z^* \sigma/E)^2$; $n = z^{*2} p(1-p)/E^2$	Round up; p = 0.5 if unknown
Test statistic	(estimate − null value) ÷ SE	t with df = n − 1 for means
Two-sided p	2 × one tail	One-sided only if chosen first
Power	1 − β	Aim for 0.80 or more
Cohen's d	difference ÷ s	Effect size

Rules that are not formulas

Topic	The rule
z or t	If you computed s, use t. Proportions use z.
Dials	More confidence widens; 4 × n halves the margin.
CI meaning	About C% of intervals from this method capture the parameter.
CI and tests	A value outside a 95% CI is rejected at α = 0.05, two-sided.
Hypotheses	About parameters; equals sign in H0; tails chosen first.
p-value	P(data this extreme | H0); not P(H0 true).
Decision	p ≤ α reject; p > α fail to reject; never accept.
Errors	Type I false alarm (α); Type II miss (β).
Significance	Not size, not importance, not cause.

Dr. Garg Learning · CC BY-SA 4.0