STATISTICS · STUDY GUIDE

Exam 6 Study Guide

This guide covers chapters 11 and 12, where the logic of testing becomes working tools. You will run the five-step test for a mean with z or t and for a proportion, check the conditions, and match each test to its confidence interval. Then you will compare two groups: independent means with Welch's t, paired data through their differences, and two proportions with a pooled z, and you will learn to pick the right test from the design. 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 11 · the five steps of a test, p-values and errors, the z-test and t-test for a mean, degrees of freedom, the one-proportion z-test, test versus interval standard errors, and the assumptions in order of importance.
- Chapter 12 · two independent means with Welch's t, paired samples and differences, two proportions with the pooled test and unpooled interval, choosing the right test, and two-sample confidence intervals.

CHAPTER 11

One-Sample Tests

Read it in full: **Chapter 11 in the textbook**. Practice it in the lab: **The Five Steps** and **Z Test for a Proportion**.

What to know

Every test follows the same **five steps**, and only the formula in step 3 changes. (1) State H0 and Ha about the parameter before seeing the data; H0 holds the equality and Ha comes from the research question. (2) Choose α and check the conditions. (3) Compute the test statistic: (estimate − claim) ÷ SE, the distance from the claim in standard errors. (4) Find the p-value or compare with a critical value. (5) Decide and interpret in context, about the original claim.

With z = 2.50 two-tailed, p = 2 × 0.0062 = 0.0124: reject at α = 0.05 but not at α = 0.01. Same evidence, different threshold, which shows α is a convention. We **fail to reject**; we never accept. A Type I error (rate α) rejects a true null; a Type II error (rate β) misses a false one; power = 1 − β rises with n, effect size, smaller variability and larger α.

The **z-test for a mean** uses $z = (\bar{x} - \mu_0) \div (\sigma \div \sqrt{n})$ and is for the rare case when σ is truly known, such as a long-studied machine. The book's bottling line (σ = 4.0, n = 25, $\bar{x}$ = 498.2) gives SE = 0.80, z = −2.25 and p = 2 × 0.0122 = 0.0244, so reject. One-sided (chosen in advance) would give 0.0122. Cutoffs at α = 0.05: ±1.96 two-tailed, 1.645 one-tailed.

A two-sided test at level α rejects H0: $\mu = \mu_0$ exactly when the 100(1 − α)% confidence interval excludes μ_0. The bottling interval 496.63 to 499.77 excludes 500. The interval also shows the plausible size of the discrepancy, which the p-value never does.

The **t-test for a mean** uses $t = (\bar{x} - \mu_0) \div (s \div \sqrt{n})$ with df = n − 1, whenever σ is estimated, which is nearly always. Replacing σ with s adds a second source of noise, so the t curve has heavier tails and larger critical values: 2.571 for df = 5, 2.131 for df = 15, 2.042 for df = 30, approaching 1.960. One degree of freedom is spent computing $\bar{x}$.

In the book's commute test (μ_0 = 30, n = 16, $\bar{x}$ = 33.4, s = 6.4), SE = 1.60, t = 2.125 with df = 15, just short of 2.131, so p ≈ 0.051 and we fail to reject. With the same numbers from n = 25, t = 2.656 and p ≈ 0.014. The estimated effect was the same 3.4 minutes; only the precision changed, so "not significant" never means "no difference."

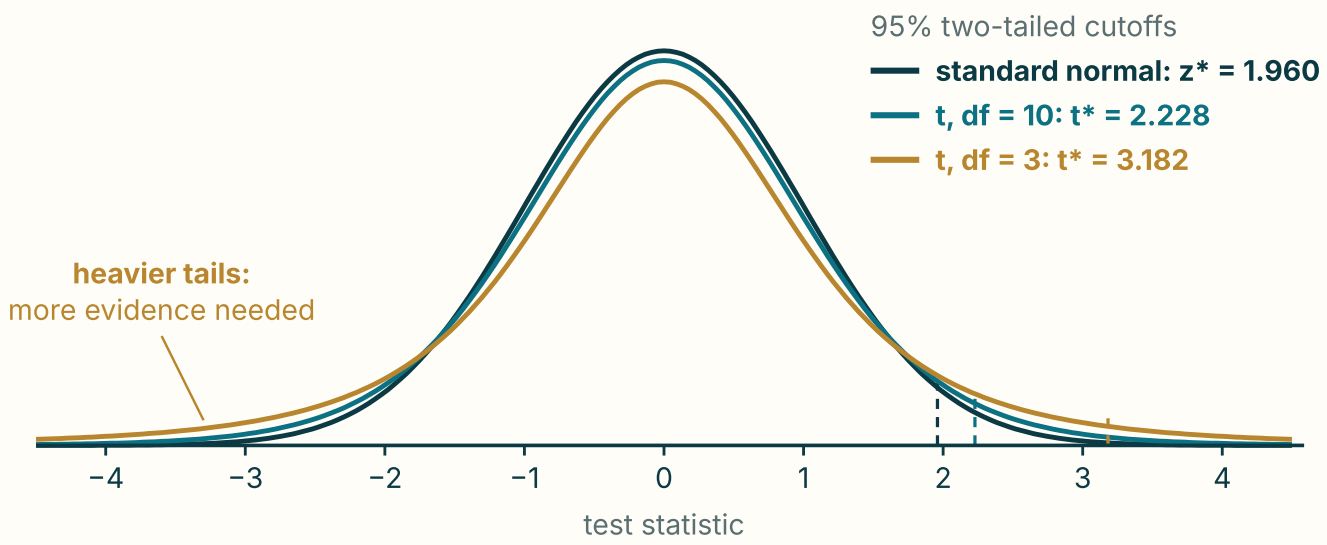
The **one-proportion z-test** uses $z = (\hat{p} - p_0) \div \sqrt{[p_0(1 - p_0) \div n]}$. There is no t-test for a proportion, because p_0 already fixes the variability. Check $np_0 \geq 10$ and $n(1 - p_0) \geq 10$, or use the exact binomial test. The **test uses p_0** in its SE; the **interval uses $\hat{p}$**. The book's 216 of 400 against p_0 = 0.60 gives z = −2.45 and p = 0.0143. Always work in decimals, not percents.

Three conditions, in order of importance. **Random sampling** cannot be rescued by any n. **Independence** holds with random sampling and n under 10% of the population; clustered data break it and make p-values too impressive. **Approximate normality** applies to the sampling distribution of $\bar{x}$, not the raw data: under 15, the data should be close to symmetric with no

outliers; 15 to 40 allows mild skew; 40 or more usually handles skew. The t-test is robust to mild non-normality but not to outliers in small samples. And the hypotheses must be fixed before the data: twenty tests at $\alpha = 0.05$ give about one false positive by chance.

LOCK THESE IN

- Five steps: hypotheses, α and conditions, statistic, p-value, conclusion in context.
- Test statistic = (estimate − claim) ÷ SE.
- z-test only when σ is truly known; otherwise t with df = n − 1.
- t has heavier tails than z; t* shrinks toward 1.96 as df grows.
- Proportion test SE uses p_0; the CI uses $\hat{p}$.
- Check $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.
- Two-sided test rejects at α exactly when the $100(1 - \alpha)\%$ CI excludes the claim.
- Random sampling, independence, normal enough: randomness cannot be repaired.



As df grows, t* shrinks toward 1.96 and the t curve becomes the normal curve.

Estimating σ costs heavier tails. With df = 3 a t statistic must pass 3.182 to reject at α = 0.05, two-tailed; with df = 10 it must pass 2.228; with σ known the cutoff is 1.96.

Formulas and when to use them

Test	Statistic	When to use it
z-test for μ	$z = (\bar{x} - \mu_0) \div (\sigma \div \sqrt{n})$	σ known from long history
t-test for μ	$t = (\bar{x} - \mu_0) \div (s \div \sqrt{n})$, $df = n - 1$	σ estimated by s
One-proportion z-test	$z = (\hat{p} - p_0) \div \sqrt{[p_0(1 - p_0) \div n]}$	Yes/no outcome; np_0, $n(1 - p_0) \geq 10$
Proportion CI	$\hat{p} \pm z^* \sqrt{[\hat{p}(1 - \hat{p}) \div n]}$	Estimating p after the test
p-value, two-sided	$2 \times P(\text{beyond } \lvert\text{statistic}\rvert)$	Ha: ≠
p-value, one-sided	One tail toward Ha	Direction set in advance
Chance of a false positive in k tests	$1 - (1 - \alpha)^k$	Multiple testing

Critical values at α = 0.05

Reference curve	Two-tailed	One-tailed
Standard normal	±1.960	1.645
t, df = 5	±2.571	2.015
t, df = 10	±2.228	1.812
t, df = 15	±2.131	1.753
t, df = 20	±2.086	1.725
t, df = 30	±2.042	1.697
t, df = 60	±2.000	1.671

Worked examples

WORKED EXAMPLE 1 · A Z-TEST WITH Σ KNOWN

A standardized test has σ = 15 and a published mean of 100. A random sample of 36 students at one district averages 104.5. Test H0: μ = 100 against Ha: μ ≠ 100 at α = 0.05.

STEP 1 · HYPOTHESES

H0: μ = 100; Ha: μ ≠ 100.

STEP 2 · CONDITIONS

Random sample, independent students, σ known, and n = 36 is large enough for the CLT.

STEP 3 · STATISTIC

SE = 15 ÷ 6 = 2.50; z = (104.5 − 100) ÷ 2.50 = **1.80**.

STEP 4 · P-VALUE

Upper tail beyond 1.80 = 1 − 0.9641 = 0.0359; doubled, **p = 0.0718**.

STEP 5 · CONCLUDE

0.0718 > 0.05, so **fail to reject H0**. The 95% interval 104.5 ± 1.96(2.50) = 99.60 to 109.40 contains 100, as it must. The district may differ, but this sample cannot show it.

WORKED EXAMPLE 2 · A ONE-SIDED T-TEST SET IN ADVANCE

A wellness program's goal is that participants average 8 hours of sleep. Before collecting data, the team decided to test whether participants fall short. A random sample of 20 gives $\bar{x}$ = 7.4, s = 1.2.

STEP 1 · HYPOTHESES

H0: μ = 8; Ha: μ < 8, chosen before the data.

STEP 2 · CONDITIONS

Random sample; n = 20 sits in the 15 to 40 range, so mild skew is fine but outliers should be checked with a plot.

STEP 3 · STATISTIC

SE = 1.2 ÷ √20 = 0.2683; t = (7.4 − 8) ÷ 0.2683 = **−2.236**, df = 19.

STEP 4 · P-VALUE

Lower tail only: **p ≈ 0.0188**. The one-tailed cutoff at df = 19 is −1.729.

STEP 5 · CONCLUDE

0.0188 ≤ 0.05, so reject H0: there is evidence participants average less than 8 hours. The estimated shortfall is 0.6 hours.

WORKED EXAMPLE 3 · A ONE-PROPORTION Z-TEST AND ITS INTERVAL

A clinic says 25% of its patients are overdue for a screening. An audit of a random 500 records finds 148 overdue. Test at $\alpha = 0.05$, two-sided.

STEP 1 · HYPOTHESES

H0: $p = 0.25$; Ha: $p \neq 0.25$.

STEP 2 · CONDITIONS WITH P_0

$np_0 = 125$ and $n(1 - p_0) = 375$, both at least 10.

STEP 3 · STATISTIC

$\hat{p} = 148 \div 500 = 0.296$. $SE_0 = \sqrt{(0.25 \times 0.75 \div 500)} = 0.01936$. $z = (0.296 - 0.25) \div 0.01936 =$ **2.38**.

STEP 4 · P-VALUE

Upper tail beyond 2.38 = 0.0087; doubled, **p = 0.0174**. Reject H0.

STEP 5 · THE INTERVAL USES $\hat{P}$

$SE = \sqrt{(0.296 \times 0.704 \div 500)} = 0.02041$; $0.296 \pm 1.96(0.02041) =$ **0.256 to 0.336**, which excludes 0.25.

WORKED EXAMPLE 4 · WHY THE HYPOTHESES MUST COME FIRST

A researcher runs 20 independent tests at $\alpha = 0.05$ on data where nothing real is going on. How many "significant" results should she expect, and what is the chance of at least one?

STEP 1 · EXPECTED FALSE ALARMS

$20 \times 0.05 =$ **1** significant result by chance alone.

STEP 2 · NONE AT ALL

Each test stays quiet with probability 0.95, so all 20 do with $0.95^{20} = 0.3585$.

STEP 3 · AT LEAST ONE

$1 - 0.3585 =$ **0.6415**, about 64%.

STEP 4 · THE LESSON

Reporting only the one test that "worked" hides a real error rate near 64%, not 5%. Fix the hypotheses, the direction and the number of tests before looking.

TRAPS PEOPLE FALL INTO

- **Picking the direction from the data.** A one-sided Ha chosen after seeing the sample doubles the real error rate.
- **Using z when s came from the data.** Estimated σ means t with df = n − 1.
- **Using a t-test for a proportion.** Proportions use z; p_0 fixes the variability.

- **Using $\hat{p}$ in the test's SE.** The test lives in the null world: use p_0. The interval uses $\hat{p}$.
- **Entering percents.** Use 0.54, not 54.
- **Reading p = 0.051 as "no effect."** It is almost the same evidence as 0.049. Give the interval and n.
- **Checking normality on the wrong thing.** It is the sampling distribution of $\bar{x}$ that must be normal; the CLT helps as n grows.
- **Expecting large n to fix a bad sample.** The CLT fixes shape, never bias.

Chapter 11 self-test · ten items

1. Which is step 3 of every test?

A. Compute the test statistic, (estimate − claim) ÷ SE
B. State the hypotheses about the population parameter
C. Choose α and check the conditions for the test
D. Interpret the decision in the context of the claim

A. Compute the test statistic, (estimate − claim) ÷ SE Step 3 measures how many standard errors the estimate is from the claim.

2. σ = 12, n = 16, $\bar{x}$ = 56, H0: μ = 50. What is z?

A. 0.50
B. 8.00
C. 2.00
D. 1.50

C. 2.00 SE = 12 ÷ 4 = 3, and z = (56 − 50) ÷ 3 = 2.00.

3. $z = -1.80$ for Ha: $\mu \neq \mu_0$. What is p?

A. 0.0359
B. 0.9641
C. 0.0720
D. 0.0718

D. 0.0718 Lower tail $\Phi(-1.80) = 0.0359$, doubled for two sides: 0.0718.

4. n = 10, s computed from the data. What reference curve and df?

A. Standard normal
B. t with df = 9
C. t with df = 10
D. t with df = 11

B. t with df = 9 σ was estimated, so use t with df = n − 1 = 9.

5. n = 25, $\bar{x} = 41$, s = 10, H0: $\mu = 45$. What is t?

A. −2.00
B. −0.40
C. −4.00
D. −10.00

A. −2.00 SE = 10 ÷ 5 = 2, t = (41 − 45) ÷ 2 = −2.00.

6. Why does the t curve have heavier tails than the normal?

A. s varies from sample to sample, adding noise
B. Small samples are always skewed to the right
C. The t test ignores the sample mean
D. df is larger than n for small samples

A. s varies from sample to sample, adding noise Two random quantities, $\bar{x}$ and s, appear in t, so extreme values are more common.

7. H0: p = 0.40, n = 150, x = 72. What is z?

A. 1.96
B. 0.08
C. 2.00
D. 1.20

C. 2.00 $\hat{p} = 0.48$, $SE_0 = \sqrt{(0.40 \times 0.60 \div 150)} = 0.04$, $z = 0.08 \div 0.04 = 2.00$. Using $\hat{p}$ in the SE is the common slip.

8. H0: p = 0.03 with n = 200. Is the normal approximation safe?

A. Yes, n is above 30
B. No, $np_0 = 6$ is below 10
C. Yes, $n(1 - p_0)$ is 194
D. No, n must be at least 1,000

B. No, $np_0 = 6$ is below 10 Both np_0 and $n(1 - p_0)$ must reach 10. Use the exact binomial test.

9. A 95% CI for μ is 12.4 to 15.0. What does a two-sided test of μ = 14 at α = 0.05 conclude?

A. Reject, since 14 is not the center of the interval
B. Accept μ = 14 as the true population mean
C. Fail to reject, since 14 is inside the interval
D. Reject, since the interval is fairly narrow

C. Fail to reject, since 14 is inside the interval A two-sided test rejects only values outside the matching interval.

10. Which condition can no sample size repair?

A. A skewed population
B. A non-random sample
C. A t curve with heavy tails
D. An unknown σ

B. A non-random sample Bias from non-random sampling stays no matter how big n gets.

CHAPTER 12

Two-Sample Tests

Read it in full: **Chapter 12 in the textbook**. Practice it in the lab: **Paired or Independent** and **Comparing Two Proportions**.

What to know

Two samples from the same population always differ by chance. A two-sample test asks whether the observed gap is bigger than chance would comfortably produce. Samples are **independent** when no subject is in both groups and nothing links a value in one group to a value in the other. H0: $\mu_1 - \mu_2 = 0$ says the populations have equal means, not the samples.

For independent means, **variances add** (standard deviations never do): $SE = \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$ and $t = (\bar{x}_1 - \bar{x}_2) \div SE$. Use **Welch's** degrees of freedom by default: $df = (a_1 + a_2)^2 \div [a_1^2/(n_1 - 1) + a_2^2/(n_2 - 1)]$, where $a = s^2/n$. It is rarely whole; by hand, rounding down is conservative. The pooled t-test assumes equal variances, which is rarely knowable. The book's coaching example gives $t = 1.896$, $df \approx 24.9$, $p \approx 0.07$: fail to reject.

Conditions for Welch: independence within groups (random sample or assignment, under 10% of the population), independence between groups (a design fact), each population roughly normal or each sample large (30 or more per group makes mild skew harmless), and no single outlier driving the difference.

Paired data have a natural partner for each value: the same subject twice, two conditions on one subject, matched subjects or two raters scoring one item. Compute each difference d, then run a one-sample t-test on the differences: $t = \bar{d} \div (s_d \div \sqrt{n})$, $df = n - 1$, where n counts **pairs**. Conditions apply to the difference column. The book's eight reaction-time pairs give $\bar{d} = 7.00$, $s_d = 5.555$, $SE = 1.964$, $t = 3.564$ with $df = 7$ and $p \approx 0.009$. With no control group, practice effects remain a rival explanation.

Pairing removes person-to-person variation, so it is far more powerful. Analyzing the same eight pairs as independent gives $SE = 9.602$, $t = 0.73$ and $p \approx 0.48$. Pairing is decided by the design, never by how the numbers line up; forcing a pairing on independent data produces a meaningless p-value.

For **two proportions**, use z. The test pools, because H0 says the proportions are equal: $\hat{p}_{pool} = (x_1 + x_2) \div (n_1 + n_2)$ and $SE_{pool} = \sqrt{[\hat{p}_{pool}(1 - \hat{p}_{pool})(1/n_1 + 1/n_2)]}$. Check that all four expected counts are at least 10. The confidence interval uses the **unpooled** SE, $\sqrt{[\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2]}$. The book's emails, 45 of 150 versus 30 of 160, give $z = 2.31$ and $p = 0.021$. An 11.3 percentage point gap is also a 60% relative increase; say which you mean.

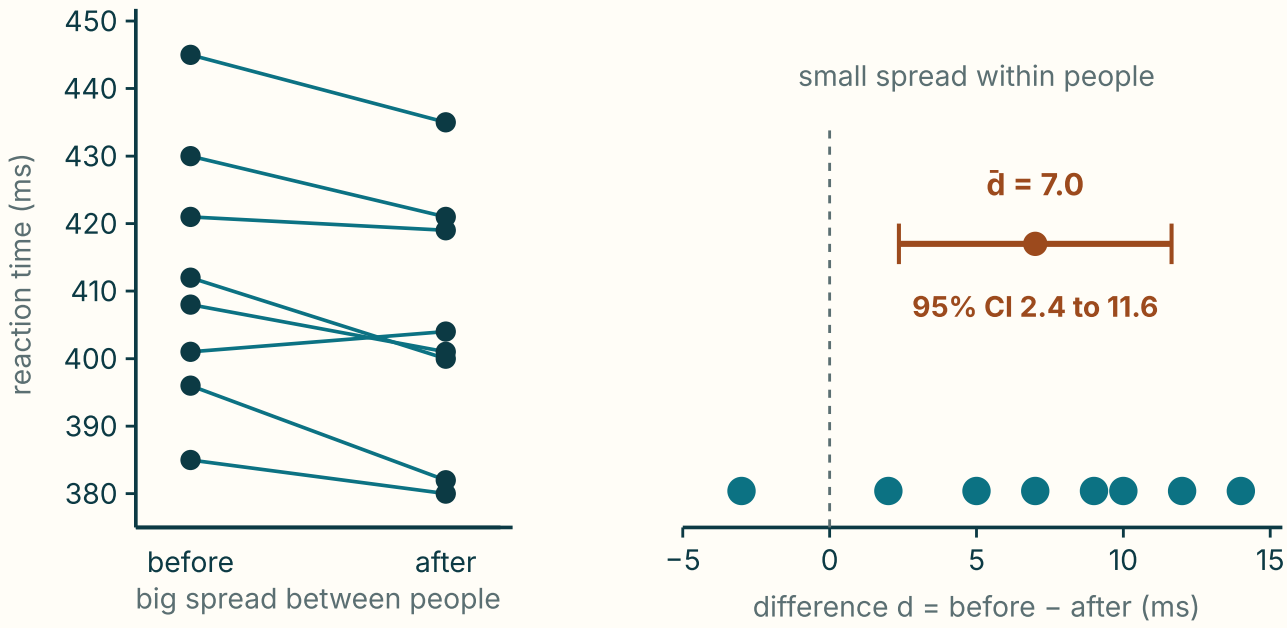
Two questions choose the test: is the outcome numeric or yes/no, and are the groups independent or paired? Numeric and independent: Welch t. Numeric and paired: paired t. Yes/no and independent: two-proportion z. Yes/no and paired: McNemar's test. Badly skewed or tiny samples:

Wilcoxon rank-sum (independent) or signed-rank (paired); small counts: Fisher's exact test. Fix the direction, α and number of comparisons in advance.

Every two-sample interval is estimate ± critical value × SE: Welch t* for independent means, t* with df = n − 1 for pairs and z* with the unpooled SE for proportions. The coaching interval is −0.56 to 13.56, which contains 0 and is too wide to be useful; the paired interval is about 2.4 to 11.6 ms; the email interval is 0.017 to 0.208. Intervals and tests agree exactly for the t procedures and only approximately for proportions. "Zero is in the interval" does not mean "the groups are the same": the width tells you how much you learned.

LOCK THESE IN

- Independent: SE = $\sqrt{(s_1^2/n_1 + s_2^2/n_2)}$; variances add.
- Use Welch df by default; rounding df down by hand is conservative.
- Paired: differences first, then a one-sample t with df = number of pairs − 1.
- Pairing is a design fact, not something you read off the numbers.
- Two proportions: z, pooled $\hat{p}$ in the test, unpooled SE in the CI.
- Pooled $\hat{p} = (x_1 + x_2)/(n_1 + n_2)$; check four expected counts ≥ 10.
- Numeric or yes/no × independent or paired picks the test.
- Percentage points are not percent change.



Pairing removes the person-to-person noise. The raw times differ by up to 60 ms between people, but each person's change is small and consistent. The differences have mean 7.0 ms and SE 1.964, so the 95% interval sits above zero.

Formulas and when to use them

Comparison	Statistic	Interval
Two independent means (Welch)	$t = (\bar{x}_1 - \bar{x}_2) \div \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$	$(\bar{x}_1 - \bar{x}_2) \pm t^* \text{ SE}$, Welch df
Welch df	$(a_1 + a_2)^2 \div [a_1^2/(n_1 - 1) + a_2^2/(n_2 - 1)]$, $a = s^2/n$	Round down by hand
Paired means	$t = \bar{d} \div (s_d/\sqrt{n})$, $df = n - 1$	$\bar{d} \pm t^* s_d/\sqrt{n}$
Pooled proportion	$\hat{p}_{pool} = (x_1 + x_2)/(n_1 + n_2)$	Test only
Two proportions	$z = (\hat{p}_1 - \hat{p}_2) \div \sqrt{[\hat{p}_{pool}(1 - \hat{p}_{pool})(1/n_1 + 1/n_2)]}$	$(\hat{p}_1 - \hat{p}_2) \pm z^*\sqrt{[\hat{p}_1\hat{q}_1/n_1 + \hat{p}_2\hat{q}_2/n_2]}$
Relative change	$(\hat{p}_1 - \hat{p}_2) \div \hat{p}_2$	Not the same as percentage points

Which test?

Outcome and design	Test	Reference curve
Numeric, independent	Two-sample (Welch) t	t, Welch df
Numeric, paired	Paired t	t, df = pairs − 1
Yes/no, independent	Two-proportion z	Standard normal
Yes/no, paired	McNemar's test	Chi-square, df = 1
Numeric, skewed or tiny, independent	Wilcoxon rank-sum (Mann-Whitney)	Its own
Numeric, skewed or tiny, paired	Wilcoxon signed-rank	Its own
Yes/no, small counts, independent	Fisher's exact test	Exact

Worked examples

WORKED EXAMPLE 1 · A WELCH TWO-SAMPLE T-TEST WITH ITS INTERVAL

Students randomly assigned to a flipped class ($n = 20$, $\bar{x} = 24.6$, $s = 5.1$) or a lecture class ($n = 18$, $\bar{x} = 21.2$, $s = 6.3$) take the same quiz. Test for a difference at $\alpha = 0.05$ and build a 95% interval.

STEP 1 · DIFFERENCE

$24.6 - 21.2 = 3.4$ points.

STEP 2 · SE

$5.1^2/20 = 1.3005$; $6.3^2/18 = 2.2050$; SE = $\sqrt{3.5055}$ = **1.8723**.

STEP 3 · T

$t = 3.4 \div 1.8723$ = **1.816**.

STEP 4 · WELCH DF

$(3.5055)^2 \div [1.3005^2/19 + 2.2050^2/17] = 12.2885 \div 0.3750$ = **32.8**.

STEP 5 · P AND DECISION

Two-sided $p \approx 0.079$. Fail to reject H0 at $\alpha = 0.05$.

STEP 6 · INTERVAL

With df rounded down to 32, $t^* = 2.037$: $3.4 \pm 2.037(1.8723) = 3.4 \pm 3.81$, so **−0.41 to 7.21**. It contains 0 and is wide: the study is too small to say much.

WORKED EXAMPLE 2 · A PAIRED T-TEST

Six patients' systolic pressure before and after a program: before 148, 152, 139, 160, 145, 155; after 140, 147, 138, 151, 142, 146. Test at $\alpha = 0.05$, two-sided, with d = before − after.

STEP 1 · DIFFERENCES

d = 8, 5, 1, 9, 3, 9; $\bar{d} = 35 \div 6 =$ **5.83**.

STEP 2 · s_D

Squared deviations sum to 56.83; s_d^2 = that ÷ 5 = 11.367; $s_d = 3.371$.

STEP 3 · SE AND T

SE = 3.371 ÷ √6 = 1.376; t = 5.83 ÷ 1.376 = **4.238**, df = 5.

STEP 4 · DECIDE

Two-sided p ≈ 0.0082; cutoff t* = 2.571. Reject H0.

STEP 5 · INTERVAL AND CAUTION

5.83 ± 2.571(1.376) = **2.29 to 9.37 mmHg**. With no control group, the drop cannot be credited to the program alone.

WORKED EXAMPLE 3 · TWO PROPORTIONS: POOLED TEST, UNPOOLED INTERVAL

A randomized trial of reminder texts: 64 of 200 texted patients kept their appointment early versus 42 of 200 controls. Test at $\alpha = 0.05$ and give a 95% interval.

STEP 1 · PROPORTIONS

$\hat{p}_1 = 0.32$, $\hat{p}_2 = 0.21$; difference 0.11.

STEP 2 · POOLED

$\hat{p}_{pool} = 106 \div 400 = 0.265$. Expected counts: 53 and 147 per group, all at least 10.

STEP 3 · SE_{POOL} AND Z

√[0.265 × 0.735 × (1/200 + 1/200)] = 0.04413; z = 0.11 ÷ 0.04413 = **2.49**.

STEP 4 · P-VALUE

2 × (1 − 0.9936) = **0.0128**. Reject H0.

STEP 5 · UNPOOLED INTERVAL

SE = √(0.32 × 0.68/200 + 0.21 × 0.79/200) = 0.04379; 0.11 ± 1.96(0.04379) = **0.024 to 0.196**.

READ IT

An 11 percentage point gain, a 52% relative increase. Random assignment supports a causal reading.

WORKED EXAMPLE 4 · CHOOSING THE TEST FROM THE DESIGN

Name the test for each study: (a) 30 tires, one new and one old tread on each car, measuring wear; (b) a random 120 nurses and 140 doctors asked yes or no about a policy; (c) 40 patients randomly split, recovery time in days; (d) 50 people asked the same yes or no question before and after a video.

STEP 1 · (A)

Numeric outcome, each car gives a natural pair: **paired t** with df = 29.

STEP 2 · (B)

Yes/no outcome, two separate groups: **two-proportion z**.

STEP 3 · (C)

Numeric outcome, independent arms: **Welch t**; if the times are badly skewed with this few patients, Wilcoxon rank-sum.

STEP 4 · (D)

Yes/no outcome, same people twice: **McNemar's test**.

TRAPS PEOPLE FALL INTO

- **Adding standard deviations.** Variances add: $\sqrt{(s_1^2/n_1 + s_2^2/n_2)}$.
- **Analyzing paired data as independent.** Person-to-person noise swamps the effect; p went from 0.009 to 0.48.
- **Counting measurements instead of pairs.** Eight people measured twice give n = 8 and df = 7.
- **Forcing a pairing on independent groups.** Lining up rows by order invents a relationship.
- **Using the pooled SE in the interval.** Pool only for the test, where H0 says the proportions are equal.
- **Saying the means are equal after failing to reject.** A wide interval that includes 0 means the study was too small.
- **Mixing percentage points and percent.** A rise from 18.8% to 30.0% is 11.3 points and 60%.
- **Switching tests after seeing p.** The design picks the test before any computing.

Chapter 12 self-test · ten items

1. Group 1: s = 6, n = 9. Group 2: s = 8, n = 16. What is the SE of $\bar{x}_1 - \bar{x}_2$?

A. 2.828
B. 4.000
C. 8.000
D. 1.414

A. 2.828 √(36/9 + 64/16) = √(4 + 4) = √8 ≈ 2.828. Adding the separate SEs (2 + 2 = 4) is the classic mistake, and 8 forgets the square root.

2. Which design produces paired data?

A. 25 men and 30 women surveyed at random in one city
B. Two randomly assigned treatment arms of 40 patients
C. Two cities sampled at random for household income
D. The same 25 drivers tested with and without a phone

D. The same 25 drivers tested with and without a phone Each driver is measured under both conditions, so every value has a specific partner.

3. Ten pairs give $\bar{d} = 4$ and $s_d = 5$. What is t?

A. 0.80
B. 3.58
C. 2.00
D. 2.53

D. 2.53 SE = 5 ÷ √10 = 1.581, t = 4 ÷ 1.581 = 2.53, df = 9.

4. Twelve subjects are each measured twice. What df does the paired t-test use?

A. 23
B. 11
C. 22
D. 12

B. 11 df = number of pairs − 1 = 11.

5. $x_1 = 30$ of 100 and $x_2 = 20$ of 100. What is $\hat{p}_{pool}$?

A. 0.25
B. 0.50
C. 0.10
D. 0.30

A. 0.25 (30 + 20) ÷ (100 + 100) = 50 ÷ 200 = 0.25.

6. In a two-proportion analysis, which SE goes in the confidence interval?

A. The pooled SE using the combined $\hat{p}_{pool}$
B. The larger of the two separate group SEs
C. The unpooled SE using each group's $\hat{p}$
D. The SE of the differences, $s_d \div \sqrt{n}$

C. The unpooled SE using each group's $\hat{p}$ The interval does not assume H0, so it uses each group's own proportion.

7. Why is a paired design often more powerful?

A. It doubles the number of independent subjects
B. It removes person-to-person variation from the SE
C. It uses z instead of t
D. It makes the population normal

B. It removes person-to-person variation from the SE Each subject serves as their own control, so the differences are much less variable.

8. A rate rises from 20% to 25%. Which statement is correct?

A. A 5% relative rise and a 25 point absolute rise
B. A 25 percentage point rise and a 5% relative rise
C. A 5 percentage point rise and a 25% relative rise
D. A 5 point rise and a 5% relative rise

C. A 5 percentage point rise and a 25% relative rise 25 − 20 = 5 points; 5 ÷ 20 = 25% relative.

9. Outcome yes/no, same 60 people asked before and after. Which test?

A. Two-proportion z-test
B. Paired t-test
C. McNemar's test
D. Welch t-test

C. McNemar's test Paired categorical data use McNemar's test, built from the discordant pairs.

10. A 95% CI for $\mu_1 - \mu_2$ is −0.4 to 12.8. What is the best reading?

A. The two means are shown to be exactly equal

B. Not significant, but a large difference is not ruled out

C. Group 1 is shown to be larger than group 2 here

D. There is a 95% chance the difference is near 6

B. Not significant, but a large difference is not ruled out The interval contains 0, so fail to reject, but its width allows a sizable difference.

ONE PAGE

Cram sheet

Everything from chapters 11 and 12 in one place. Cover the right column and see if you can rebuild each row.

Formulas

Name	Formula	Note
z for μ	$(\bar{x} - \mu_0) \div (\sigma/\sqrt{n})$	σ known only
t for μ	$(\bar{x} - \mu_0) \div (s/\sqrt{n})$, df = n − 1	σ estimated
One proportion	$(\hat{p} - p_0) \div \sqrt{[p_0(1 - p_0)/n]}$	CI uses $\hat{p}$
Welch t	$(\bar{x}_1 - \bar{x}_2) \div \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$	Welch df
Paired t	$\bar{d} \div (s_d/\sqrt{n})$	df = pairs − 1
Pooled $\hat{p}$	$(x_1 + x_2)/(n_1 + n_2)$	Test only
Two proportions	$(\hat{p}_1 - \hat{p}_2) \div SE_{pool}$	CI uses unpooled SE
False positives	$1 - (1 - \alpha)^k$	k tests

Rules that are not formulas

Topic	The rule
Five steps	Hypotheses, α and conditions, statistic, p-value, conclusion in context.
Direction	Chosen from the question before the data.
z or t	Estimated σ means t; proportions use z.
Duality	Two-sided test at α rejects exactly when the 100(1 − α)% CI excludes the claim.
Conditions	Random (unfixable), independent (10%), normal enough (by n).
Pairing	Decided by design; n counts pairs.
Test choice	Numeric or yes/no × independent or paired.
Zero in the CI	Fail to reject; width tells how much you learned.

Dr. Garg Learning · CC BY-SA 4.0