Chapter 9: Small-Sample Inference

Lecture slides: PDF pages 1–3 (open)

Probability and Statistics for the Life Sciences **November 2, 2026**

You might measure an enzyme level in a small sample of patients, compare growth under two treatments, or measure the same specimen twice. Each question asks about a population, but the evidence comes from a sample. These notes explain how we allow for uncertainty when estimating the population's spread from that sample.

The reasoning stays the same

1. **Choose the target.** Which population mean, or difference of means, answers the question?
2. **Estimate it.** Calculate a sample mean or mean difference.
3. **Measure uncertainty.** A standard error estimates how much that estimate varies from sample to sample.
4. **Compare with a model.** For a test, ask how surprising the discrepancy would be if the null hypothesis were true. For an interval, give a range using that same uncertainty.

$$\boxed{\text{test statistic} = \frac{\text{estimate} - \text{null value}}{\text{estimated standard error}}}$$

In Chapter 8 we often compared this number with a normal curve. Here, under the stated normal-population assumptions, **Student's t curve** accounts for the extra uncertainty from estimating the spread.

Words and symbols used throughout

μ, σ	Population mean and population standard deviation (SD).
n, $\bar{x}$, s	Sample size, sample mean, sample SD.
SE	Standard error: the SD of an estimator's sampling distribution, usually estimated from the data.
H_0, H_a	Null claim being tested and the alternative research claim.
α	Significance level chosen before looking at the test result.

Reading map

Topic	These notes	Slide PDF
SD, SE, t curves, degrees of freedom	2–4	2–18
One mean: intervals and tests	5–9	19–33
Two independent means: pooling and inference	10–14	34–55
Paired measurements: analyze the differences	15–16	56–62
Choosing a method	17	Review

The headings link to the current 62-page lecture PDF, including its incremental reveals. Page numbers mean PDF viewer pages, not slide-frame numbers. If your viewer ignores a page link, enter the printed page number or use the linked section index. The slide examples are retained; additional biological illustrations are labeled.

From the normal curve to Student's t

Lecture slides: PDF pages 2–6 (open)

First distinguish SD from SE

Standard deviation (SD) describes how much individual measurements vary. **Standard error (SE)** describes how much a sample estimate varies across repeated samples. For independent observations with the same population SD σ,

$$\mathrm{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}}, \qquad \text{estimated SE of the mean} = \frac{s}{\sqrt{n}}.$$

Capital $\bar{X}$ refers to a sample mean before the sample is observed; lower-case $\bar{x}$ is the value we actually calculate. Likewise, S varies across samples and s is the observed sample SD.

New biological illustration. Suppose nine independently sampled plants have a sample SD of 1.8 cm in height. The individual heights have an estimated spread of 1.8 cm; the sample mean has estimated SE $1.8/\sqrt{9} = 0.6$ cm. Averaging reduces the sampling variability of the mean. It does not make the individual plants less variable.

What changes when sigma is unknown?

If the population is normal and σ is known, the standardized sample mean is exactly normal:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Usually we do not know σ. We substitute the sample SD:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

Now both the numerator and denominator change across samples. Occasionally S is small by chance, making the ratio larger. A standard normal curve does not fully account for that extra source of variation.

The result that justifies the method

If the observations are an independent random sample from a **normal population**, then

$$\boxed{\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.}$$

The symbol t_{n-1} names Student's t distribution with $n - 1$ *degrees of freedom*. We explain that number next. The result is exact under normal sampling, for every $n \geq 2$.

Check the model before using the result. Approximate normality gives an approximate method. Strong skew, extreme outliers or dependent measurements can make small-sample t inference unreliable. A small sample does not become normal just because we use a t table. At large sample sizes, suitable normal approximations may be close; $n = 30$ is a course rule of thumb, not a guarantee for every population.

Check your understanding. With unknown σ, what changes from Chapter 8?

Reasoning. The estimate-minus-null idea stays. We estimate SE from the data and use an appropriate t reference under the normal-sampling model.

Why the degrees of freedom are n minus one

Lecture slides: PDF pages 7–13 (open)

Degrees of freedom, abbreviated *df*, count the free pieces of information used here to estimate the spread after fitting a mean.

The sample mean and variance are

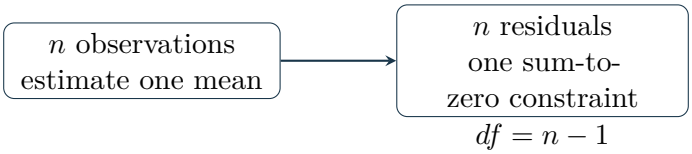
$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}, \qquad s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

A **residual** is an observation minus the sample mean: $x_i - \bar{x}$. The residuals always add to zero. Once we know $n-1$ of them, the last is determined.

A small example

For measurements $3, 5, 7$, the sample mean is 5. The residuals are $-2, 0, 2$. If the first two residuals are -2 and 0, the third must be 2 to make their sum zero. There are three residuals but only two free choices:

$$df = 3 - 1 = 2, \qquad s^2 = \frac{(-2)^2 + 0^2 + 2^2}{2} = 4, \qquad s = 2.$$



Why not divide by n?

The fitted mean lies in the middle of our particular sample. Distances from this fitted center tend to be smaller than distances from the true population mean. Dividing their squared sum by $n-1$, rather than n, corrects that tendency: across repeated independent samples with common variance, the average value of S^2 is σ^2.

Optional derivation: the correction in one identity

Split each deviation into a residual and the error in the sample mean:

$$X_i - \mu = (X_i - \bar{X}) + (\bar{X} - \mu).$$

Squaring and adding, the cross terms cancel because the residuals sum to zero:

$$\sum (X_i - \mu)^2 = \sum (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.$$

Across repeated samples, the average of the left side is $n\sigma^2$, and the average of the last term is $n(\sigma^2/n) = \sigma^2$. The average residual sum of squares is therefore $(n-1)\sigma^2$. Dividing it by $n-1$ gives the correct average variance.

Separate two claims. This average-variance argument needs independent observations with common finite variance. The exact t result additionally needs normal sampling. The first claim alone does not prove the second.

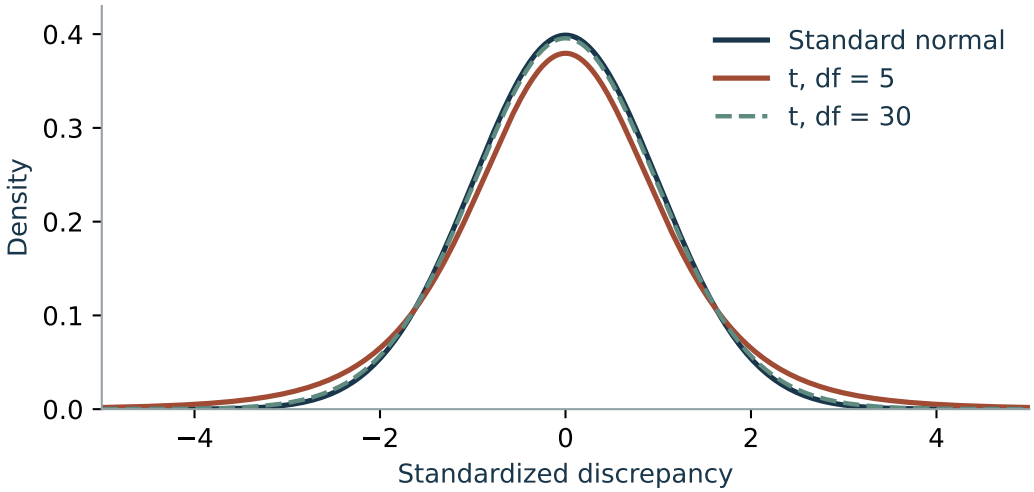
Check your understanding. If $n = 10$, which row of the one-sample t table is used?

Reasoning. The $df = 9$ row, because one mean was estimated from ten observations.

Reading a t curve and its table

Lecture slides: PDF pages 14–18; curve formula on p. 6 (open)

Each degrees-of-freedom value gives a different t curve. All are symmetric about zero. Compared with a standard normal curve, a t curve has **heavier tails**: large positive or negative standardized values occur more often. Small df means more uncertainty in the estimated SD, so the tails are heavier.



As df increases, the t curve approaches the standard normal curve. The density formula on slide p. 6 describes these curves; the practical task is to use their tail areas, not memorize that formula.

The course notation refers to an upper tail

$$\boxed{P(T_{df} > t_{df,\alpha}) = \alpha.}$$

Here df selects the curve and α is the upper-tail area. Thus $t_{9,.025} = 2.262$ means that 2.5% of the t_9 curve is to the right of 2.262. By symmetry, another 2.5% is below -2.262; 95% lies between the two cutoffs.

df	Upper 5%: $t_{df,.05}$	Upper 2.5%: $t_{df,.025}$	Upper 0.5%: $t_{df,.005}$
4	2.132	2.776	4.604
5	2.015	2.571	4.032
9	1.833	2.262	3.250
16	1.746	2.120	2.921
Normal limit	1.645	1.960	2.576

- **Two-sided 95% interval or 5% test:** split .05 into .025 per tail.
- **Right-sided 5% test:** all .05 is in the upper tail.
- **Left-sided 5% test:** use the negative of the upper 5% cutoff.

Check your understanding. Why is 2.571 larger than 1.960?

Reasoning. At $df = 5$, the t curve allows more large standardized values. A wider central interval is needed to contain 95% of its area.

A confidence interval for one population mean

Lecture slides: PDF pages 19; summary p. 28 (open)

The Chapter 7 structure still applies:

$$\boxed{\text{estimate } \pm \text{ critical value} \times \text{SE}} \qquad \boxed{\bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}.}$$

This interval estimates μ, the population mean. It uses an independent random sample from a normal population with unknown SD. If normality is only approximate, so is the coverage claim.

A calculation you can follow

1. Find n, the sample mean $\bar{x}$ and the sample SD s.
2. Divide the SD by $\sqrt{n}$ to get the estimated SE of the mean.
3. For confidence level $1-\alpha$, use $df = n-1$ and upper-tail area $\alpha/2$.
4. Multiply the SE by the critical value. This is the **margin of error**.
5. Add and subtract that margin. State the population target and units.

New biological illustration. For nine independent plants from an approximately normal height population, suppose $\bar{x} = 12.4$ cm and $s = 1.8$ cm. Then $SE = .6$ cm and $df = 8$. With $t_{8,.025} = 2.306$,

$$95\%\text{ CI for mean height} = 12.4 \pm 2.306(.6) = (11.02, 13.78)\text{ cm.}$$

This concerns the average height in the represented population, not the height of a particular plant.

Why the formula has this form

Under the normal model, the standardized mean is within the two t cutoffs with probability $1-\alpha$:

$$P\left(-t_{n-1,\alpha/2} \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le t_{n-1,\alpha/2}\right) = 1-\alpha.$$

Multiplying by the positive SE and solving the inequalities for μ gives the endpoints above. The two-sided interval includes population means within one margin of error of the sample mean.

What 95 percent confidence says

If we repeatedly sampled and constructed intervals this way, about 95% would contain the fixed true mean under the model. One observed interval either contains that mean or does not. The statement is *not* that 95% of plants lie inside the interval or that there is a 95% posterior probability for μ in it.

Reading an interval backward. Its midpoint is the sample mean; its half-width is the margin of error. If n and the confidence level are known,

$$\bar{x} = (L+U)/2, \qquad M = (U-L)/2, \qquad s = M\sqrt{n}/t_{n-1,\alpha/2}.$$

Rounded endpoints give approximate recovered values.

Example 1 Synthetic diamonds

Lecture slides: PDF pages 20–23 (open)

Question from the slides

A new process for producing synthetic diamonds is profitable only if the population mean weight exceeds 0.5 karat. Six diamonds from the process have weights

$$0.46, \quad 0.61, \quad 0.52, \quad 0.48, \quad 0.57, \quad 0.54 \quad \text{karat}.$$

There is reason to model the weights as normally distributed. Find a 95% confidence interval for the population mean weight μ. Treat the sample as independent and representative of the process.

Compute the center and uncertainty

$$\begin{aligned} n = 6, \quad df = 5, \qquad \bar{x} &= \frac{3.18}{6} = .53, \\ \sum (x_i - \bar{x})^2 &= .0156, \\ s^2 = \frac{.0156}{5} &= .00312, \qquad s = .05585696, \\ \mathrm{SE}(\bar{x}) = \frac{s}{\sqrt{6}} &= .02280351 \text{ karat}. \end{aligned}$$

With $\alpha = .05$, use $t_{5,.025} = 2.5705818 \approx 2.571$:

$$.53 \pm 2.5705818(.02280351) = .53 \pm .05861828.$$

$$\boxed{95\% \text{ CI for } \mu: \quad (.4714, .5886) \text{ karat}.}$$

The slides round s and the critical value earlier and obtain (.4713, .5887). That small difference is rounding, not a change in method.

Interpret the result in context

The sample mean is above .5, but the uncertainty is substantial: the interval includes means both below and above the profitability threshold. The two-sided 95% interval does not establish that the true average exceeds .5.

If the scientific question is specifically $\mu > .5$, the corresponding *one-sided* test should be specified in advance. As an added connection, its statistic is

$$t_{\text{obs}} = \frac{.53 - .5}{.02280351} = 1.3156 < t_{5,.05} = 2.015.$$

It also fails to reject at level .05. A two-sided 95% interval and a one-sided level .05 test are not generally matching procedures, even though both are inconclusive here.

Check your understanding. Why not use .0559 directly as the margin of error?

Reasoning. That is an estimated observation SD. First divide by $\sqrt{6}$ for the mean's SE, then multiply by the confidence critical value.

Why the wider t interval is needed

Lecture slides: PDF pages 24–28 (open)

For the six diamonds on the previous page, the estimated SE is .0228035 karat. Using 1.960 as the multiplier gives the narrower interval (.4853, .5747). Using the appropriate t_5 multiplier gives (.4714, .5886).

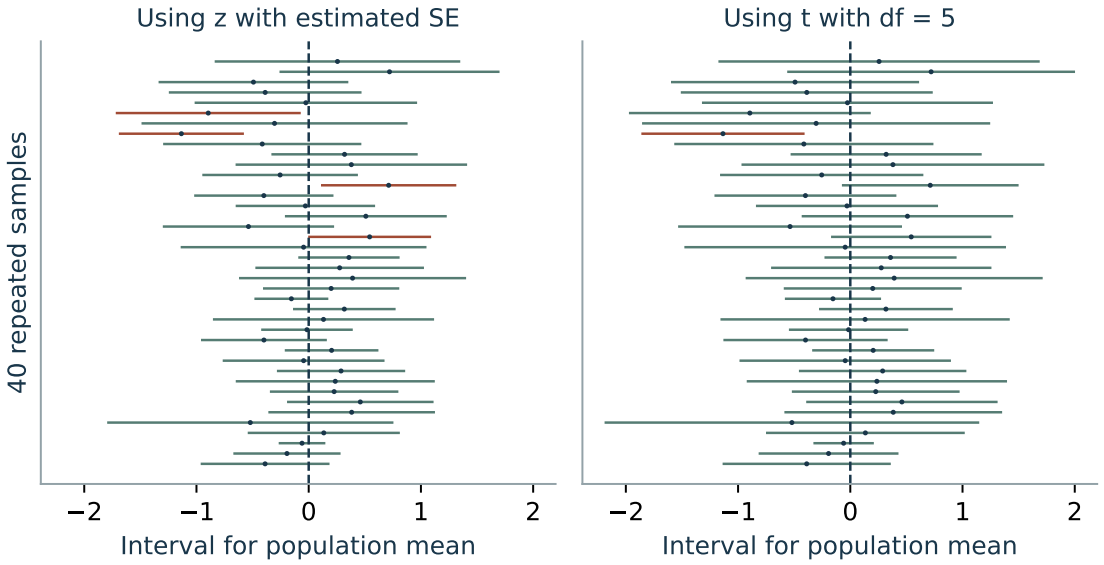
A narrower interval is not automatically better. It must also have the stated coverage. When we estimate SD from six normal observations, the standardized mean follows t_5, not a standard normal curve. The larger multiplier allows for that uncertainty.

Imagine repeating the whole experiment

The slides describe 500 independent samples of size 6 from a normal population with mean 0 and SD 1. For each sample:

1. Calculate its own mean and sample SD.
2. Construct one interval with multiplier 1.960 and another with 2.571.
3. Record whether each interval includes the fixed population mean, 0.

The picture below illustrates the same idea with 40 newly simulated samples. A row uses the same sample in both panels; only the multiplier changes.



Green intervals contain 0; rust intervals miss. This is a new illustration, not a reproduction of the Canvas workbook's results.

What repeated sampling shows

Under the normal model, using 1.960 with the estimated SE at $n = 6$ covers the mean only about 89.3% of the time. Using the t_5 cutoff gives 95% coverage. One finite simulation may differ from these long-run rates.

Keep the reason straight. If the population SD were genuinely known and used in the SE, the normal cutoff would be appropriate for normal data. The difference here comes from estimating SD, not simply from the word "small."

Testing a claim about one mean

Lecture slides: PDF pages 29–30 (open)

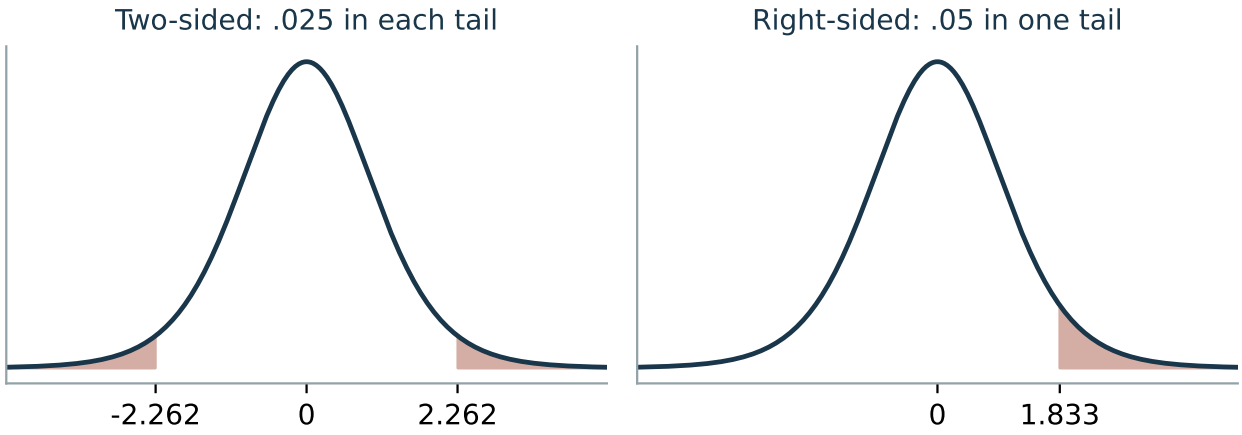
Choose the population target μ, the null value μ_0, and the scientific alternative before looking at the result. Under $H_0 : \mu = \mu_0$ and independent normal sampling,

$$\boxed{t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \qquad df = n - 1.}$$

Read this as "how many estimated SEs is the sample mean from the null value?" A positive value puts the sample mean above the null; a negative value puts it below.

The alternative tells us where to look

Alternative	Evidence against H_0	Reject at level α if
$\mu > \mu_0$	Large positive t	$t_{\text{obs}} \geq t_{df,\alpha}$
$\mu < \mu_0$	Large negative t	$t_{\text{obs}} \leq -t_{df,\alpha}$
$\mu \neq \mu_0$	Large magnitude in either direction	$\lvert t_{\text{obs}} \rvert \geq t_{df,\alpha/2}$



The shaded regions show rejection regions for $df = 9$ and $\alpha = .05$.

The p-value uses the observed statistic

A p-value is the probability, under the null model, of a test statistic at least as extreme in the direction(s) specified by H_a. For a right-sided test use $P(T_{df} \geq t_{\text{obs}})$; for a left-sided test use $P(T_{df} \leq t_{\text{obs}})$; for a two-sided test use $2P(T_{df} \geq |t_{\text{obs}}|)$. It is not the probability that H_0 is true.

Connect the test to an interval

A two-sided level α test rejects exactly when the null value lies outside the matching $(1 - \alpha)$ interval, apart from the boundary convention:

$$|\bar{x} - \mu_0| > t_{df,\alpha/2}\, SE. \quad \text{The discrepancy exceeds the interval's margin.}$$

Use the same model and SE. A 95% interval matches a two-sided 5% test; a 90% interval matches a two-sided 10% test.

Example 3 Paint coverage

Lecture slides: PDF pages 31–33 (open)

Question from the slides

A manufacturer claims that one gallon of paint covers 400 ft^2. Ten randomly selected one-gallon cans are tested on identical surfaces with the same equipment. The recorded coverages are

$$310,\ 311,\ 412,\ 368,\ 447,\ 376,\ 303,\ 365,\ 410,\ 350 \quad \text{ft}^2.$$

Is there sufficient evidence that the population mean coverage differs from 400 ft^2? Use $\alpha = .05$.

Specify the population and model

Let μ be the mean coverage of this paint under the stated testing conditions. We assume independent observations from an approximately normal coverage distribution, as in the course's small-sample exercises. Normality is a modeling assumption, not a consequence of ten observations.

$$H_0 : \mu = 400, \qquad H_a : \mu \neq 400.$$

Calculate without early rounding

$$\begin{aligned}
\bar{x} &= 365.2, \quad \sum (x_i - \bar{x})^2 = 21097.6, \\
s^2 &= \frac{21097.6}{9} = 2344.1778, \quad s = 48.4167, \\
\text{SE} &= \frac{48.4167}{\sqrt{10}} = 15.3107 \text{ ft}^2, \\
t_{\text{obs}} &= \frac{365.2 - 400}{15.3107} = -2.2729, \qquad df = 9.
\end{aligned}$$

The two-sided rejection rule is $|t_{\text{obs}}| \geq t_{9,.025} = 2.2622$. Since $2.2729 > 2.2622$, reject H_0 at the 5% level. The table bounds the two-sided p-value by $.02 < p < .05$; numerical evaluation gives $p \approx .04913$.

Conclusion and scale of the effect

There is evidence that mean coverage differs from 400 ft^2, with the sample indicating a lower mean. The estimated shortfall is 34.8 ft^2. As an added interval connection,

$$365.2 \pm 2.2621572(15.3107079) = (330.5648, 399.8352) \text{ ft}^2.$$

The interval just excludes 400. The decision is close to the .05 cutoff; keeping adequate precision matters. The evidence would not meet the more stringent .01 level. Statistical significance does not alone decide the practical consequences of a shortfall.

Check your understanding. Would changing the alternative to "less than" after seeing the data preserve the planned two-sided test?

Reasoning. No. Choose the scientific direction before examining the result. The question asks whether the mean differs, so use both tails.

Comparing two independent population means

Lecture slides: PDF pages 34–40 (open)

New biological setting. Suppose different, independently sampled plants receive two growth conditions. We want to estimate $\mu_1 - \mu_2$, the difference in population mean heights. Each plant belongs to one group. This is different from measuring the same plant twice.

The target and its estimate

Write $\Delta = \mu_1 - \mu_2$ and estimate it by $\bar{x}_1 - \bar{x}_2$. Define the subtraction order first; it determines what a positive answer means.

Both sample means vary across repeated samples. For independent groups, their variances add even though we subtract their means:

$$\mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

The uncertainty comes from both samples; subtraction does not cancel unrelated sampling errors.

The extra assumption used by the pooled method

This chapter's small-sample method assumes normal populations with the **same population variance**: $\sigma_1^2 = \sigma_2^2 = \sigma^2$. The common variance is unknown. It describes within-group spread, not equality of the two means.

$$\mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

Both samples help estimate this common spread. We combine their variances, weighting by their degrees of freedom:

$$\boxed{s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.}$$

Following the slides, s^2 denotes the pooled variance here. Take its square root to get the pooled SD s, then calculate

$$\boxed{SE(\bar{x}_1 - \bar{x}_2) = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.}$$

Why these weights?

Each $(n_i - 1)s_i^2$ is a group's sum of squared residuals. Add those two sums and divide by their total degrees of freedom. Larger groups contribute more information. For equal sample sizes, s^2 is the ordinary average of the two sample variances.

Pool *variances*, not SDs. Keep each group's own mean when computing its residuals; using a single grand mean would mix the between-group difference into the spread estimate.

Connection to Chapter 8. Pooling proportions there imposed the null equality $p_1 = p_2$. Pooling variances here uses a common-variance model for both the confidence interval and the test. These are distinct reasons for pooling.

A pooled confidence interval: conditions first

Lecture slides: PDF pages 41–46 (open)

Here s^2 denotes the pooled variance and s its square root, following the slides; s_1, s_2 are the separate group SDs. There are $n_1 + n_2$ observations, and we estimate a mean in each group. The remaining degrees of freedom are

$$df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2.$$

With independent normal populations and a common variance,

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2}.$$

This has the same estimate-minus-target divided by SE structure as the one-sample result. Only the target, SE and degrees of freedom change.

Build the interval from the same three pieces

$$\boxed{(\bar{x}_1 - \bar{x}_2) \ \pm \ t_{n_1+n_2-2,\alpha/2} \ s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.}$$

1. **Estimate:** the observed mean difference, in the stated order.
2. **SE:** the pooled SD times $\sqrt{1/n_1 + 1/n_2}$.
3. **Multiplier:** a t cutoff using $n_1 + n_2 - 2$ degrees of freedom.

The interval estimates a difference of population means, not a difference between two individual measurements.

Three assumptions to state

1. **Independence:** independent observations within each sample and independent groups. No participant or specimen appears in both groups.
2. **Population shape:** normal populations give exact inference. Approximately normal populations give an approximation; small samples with strong skew or outliers need care.
3. **Equal population variances:** both groups share a common spread. The sample variances need not be numerically identical.

How the slides check the spread assumption

Slide p. 46 gives the rough rule

$$\frac{1}{3} < \frac{s_1^2}{s_2^2} < 3.$$

It screens for a large observed discrepancy. It does *not* prove that population variances are equal: sample variances themselves vary. Use a stipulated equal-variance model explicitly, and remember the limitations of the rule. If equal variances are not defensible, this pooled method is not justified; an unequal-variance method is beyond this deck.

Check your understanding. Two independent groups each have nine observations. What is df?

Reasoning. There are 18 observations and two fitted means, so $df = 16$. Equal sample sizes alone do not establish equal population variances.

Example 4 Online and classroom scores

Lecture slides: PDF pages 47–48 (open)

Question from the slides

A course can be taken online or through classroom lectures. Independent random samples of nine scores from each population are listed below. Assume equal population variances and, for the small-sample method, normal population models. Let μ_1 be the online mean score and μ_2 the classroom mean. Find a 90% confidence interval for $\mu_1 - \mu_2$.

Online	32	37	25	28	41	44	35	31	32
Classroom	35	31	29	25	34	40	27	32	31

These are independent lists, not matched students. The original two-column layout does not imply pairing.

Recompute from the displayed observations

$$\sum x_{1i} = 305, \quad \bar{x}_1 = 305/9 = 33.8888889,$$
$$\sum x_{2i} = 284, \quad \bar{x}_2 = 284/9 = 31.5555556,$$
$$\bar{x}_1 - \bar{x}_2 = 2.3333333 \text{ score points.}$$

Using the within-group residual sums,

$$\sum (x_{1i} - \bar{x}_1)^2 = 292.8888889, \qquad s_1^2 = 36.6111111,$$
$$\sum (x_{2i} - \bar{x}_2)^2 = 160.2222222, \qquad s_2^2 = 20.0277778.$$

The variance ratio is about 1.828, within the course's heuristic range. Equality remains a stated model assumption.

Correction to the slide arithmetic

For the raw data printed on p. 47, the online mean is 33.8889, not 35.22, and the online variance is 36.6111, not 24.4441. Therefore the pooled variance, confidence interval and later test statistic on pp. 48–49 and 54 must be recomputed. This is a correction to the numerical work; the inference method and the final fail-to-reject decision in Example 5 remain the same.

A useful verification habit. Recompute a sample mean directly from the total before using a printed summary. Small rounding differences are expected; a total inconsistent with the listed observations is a different issue.

Example 4 Constructing the interval

Lecture slides: PDF pages 48–49 (open)

The target is online minus classroom. The estimate is 2.3333333 points, and both sample sizes are 9. Pool the corrected variances:

$$s^2 = \frac{8(36.6111111) + 8(20.0277778)}{16} = 28.3194444,$$
$$s = 5.3216017, \qquad df = 16,$$
$$\mathrm{SE}(\bar{x}_1 - \bar{x}_2) = 5.3216017\sqrt{\frac{1}{9} + \frac{1}{9}} = 2.5086271 \text{ points}.$$

For 90% confidence, $\alpha = .10$ and the upper-tail area is .05:

$$t_{16,.05} = 1.7458837 \approx 1.746.$$

Therefore

$$(\bar{x}_1 - \bar{x}_2) \pm t_{16,.05}\,\mathrm{SE} = 2.3333333 \pm 1.7458837(2.5086271)$$
$$= 2.3333333 \pm 4.3797711,$$

$$\boxed{90\%\text{ CI for } \mu_1 - \mu_2 : \quad (-2.0464,\ 6.7131).}$$

If the rounded slide table value 1.746 is used throughout, the endpoints are $(-2.0467, 6.7134)$. Both reflect the same corrected calculation.

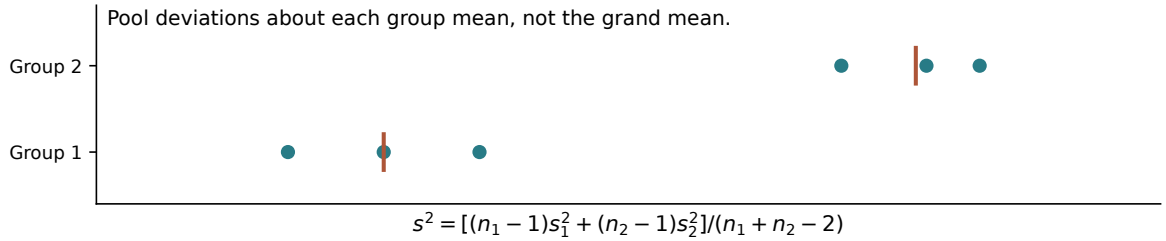
What the interval supports

The sample average is higher online by about 2.33 points, but the interval includes zero and both signs. Under the model, the data are compatible with a modest classroom advantage, no difference, or an online advantage in this range. Failure to distinguish the means is not proof that the delivery methods are equivalent.

Connecting estimation to testing

This 90% interval inverts a two-sided level .10 test under the same pooled model. The next slide example instead asks for a level .05 test. Use its own cutoff, or construct a 95% interval, rather than claiming that every confidence level matches every significance level.

Scope of the conclusion. Random samples support inference to the represented populations. Without a randomized assignment or another defensible causal design, a difference in group means alone does not establish that the delivery method caused it.



Pooling sketch (not the score data): s^2 is the pooled variance; s is its square root.

Check your understanding. What would reversing the subtraction order do?

Reasoning. It negates the estimate and swaps the negated endpoints: classroom minus online has interval $(-6.7131, 2.0464)$. A two-sided testing decision is unchanged.

Independent two-sample tests and Example 5

Lecture slides: PDF pages 50–55 (open)

The general pooled test

For $H_0 : \mu_1 - \mu_2 = \Delta_0$ under the same normal, independent, common-variance model,

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{S\sqrt{1/n_1 + 1/n_2}} \sim t_{df}, \qquad df = n_1 + n_2 - 2.$$

The null difference need not be zero. Right, left and two-sided alternatives use exactly the same tail rules as one-sample tests, with the correct df. A matching $(1 - \alpha)$ pooled interval excludes Δ_0 precisely when the two-sided level α test rejects, apart from boundary convention.

Question from the slides

For the online/classroom samples in Example 4, is there sufficient evidence of a difference in population mean exam scores at $\alpha = .05$?

$$H_0 : \mu_1 - \mu_2 = 0, \qquad H_a : \mu_1 - \mu_2 \neq 0.$$

Using the corrected raw-data summaries from the preceding pages,

$$t_{\text{obs}} = \frac{2.3333333 - 0}{2.5086271} = .9301236, \qquad df = 16.$$

The rejection region is

$$(-\infty, -t_{16,.025}] \cup [t_{16,.025}, \infty) \approx (-\infty, -2.120] \cup [2.120, \infty).$$

Because $|.9301| < 2.120$, we fail to reject H_0. The two-sided p-value is approximately .3661. The data do not provide sufficient evidence of a population mean difference at the 5% level.

Matching 95 percent interval

As an added consistency check,

$$2.3333333 \pm 2.1199053(2.5086271) = (-2.9847,\ 7.6514).$$

Zero is inside, as the level .05 decision requires. This is wider than the 90% interval, because higher confidence uses a larger critical value.

Correction in context. The slide's test statistic 1.65 follows its inconsistent earlier summaries. The displayed data give .9301. Both values lie inside the nonrejection region, so the original qualitative decision survives; its numerical evidence changes.

Check your understanding. Can we conclude the means are equal because $p > .05$?

Reasoning. No. We have insufficient evidence of a difference under this test, not evidence proving exact equality or practical equivalence.

Paired data: compare each unit with itself

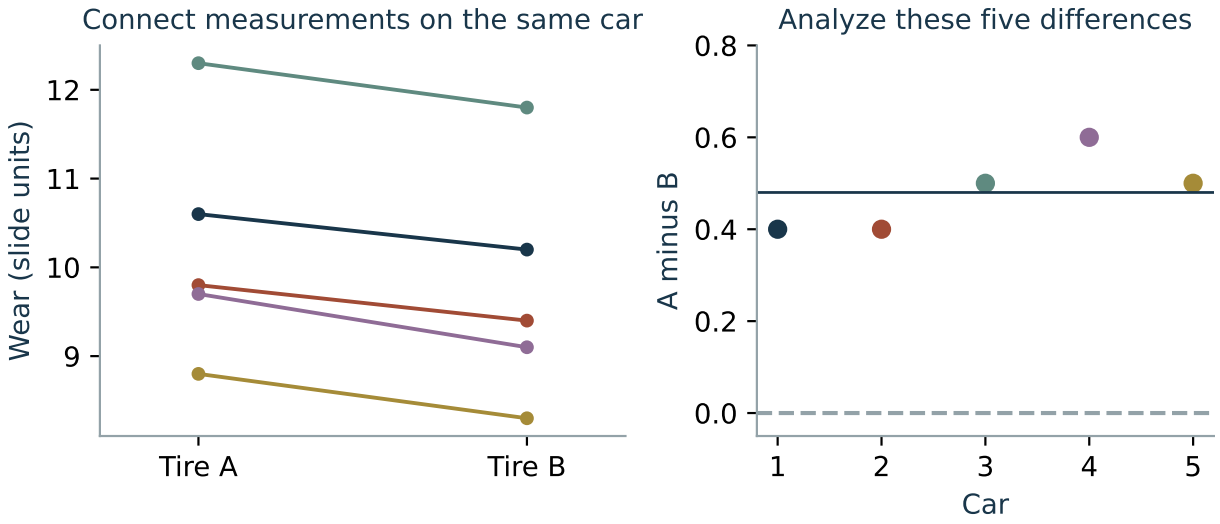
Lecture slides: PDF pages 56–60 (open)

Two measurements are **paired** when the study design links them: before and after on the same patient, two methods on the same specimen, or deliberately matched units. Equal-length columns alone do not establish a pair.

New biological illustration. If the same patient's blood pressure is measured before and after treatment, a naturally high baseline may make both readings high. Subtracting within that patient focuses on the change. Comparing unrelated patients would be a different design.

The tire example from the slides

Five cars are measured on tire A and then tire B. Each line below joins the two measurements from one car. Some cars have high values under both tire types; the within-car differences are much more consistent.



Make one observation per independent pair

Choose an order, for example $d_i = A_i - B_i$. Calculate the mean and SD of the *differences*, $\bar{d}$ and s_d. The parameter is μ_d, the population mean difference. A paired procedure is then a one-sample t procedure:

$$t_{\text{obs}} = \frac{\bar{d} - \mu_{d,0}}{s_d/\sqrt{n}}, \qquad df = n - 1.$$

Here n counts pairs, not individual readings. Exact inference requires independent pairs with normally distributed differences. It does not require equal marginal variances for A and B. Approximate normal differences give approximate inference.

Clarification to slide p. 60. Dependence is between the A and B measurements *within each car*; different cars are modeled as independent.

Optional mathematical connection. The variance of a difference is

$$\text{Var}(A - B) = \text{Var}(A) + \text{Var}(B) - 2\,\text{Cov}(A, B).$$

Positive covariance means the two readings tend to move together. Subtracting removes some shared variation. Pairing can improve precision, but its benefit depends on the design and relationship; it is not automatic.

Example 6 Tire wear

Lecture slides: PDF pages 56–62 (open)

Question and observations from the slides

Using the five cars described on the previous page, test whether mean wear differs between the tire types. The slide does not specify a level; we use $\alpha = .05$ for the worked decision. Wear units were not specified in the slide table.

Car	A	B	$d = A - B$
1	10.6	10.2	.4
2	9.8	9.4	.4
3	12.3	11.8	.5
4	9.7	9.1	.6
5	8.8	8.3	.5

Assume independent cars and normal differences. Then

$$\bar{d} = .48, \qquad \sum (d_i - \bar{d})^2 = .028,$$
$$s_d^2 = .028/4 = .007, \qquad s_d = .0836660,$$
$$\mathrm{SE}(\bar{d}) = .0836660/\sqrt{5} = .0374166, \qquad df = 4.$$

For $H_0 : \mu_d = 0$ versus $H_a : \mu_d \neq 0$,

$$t_{\text{obs}} = \frac{.48 - 0}{.0374166} = 12.8285 > t_{4,.025} = 2.7764.$$

Reject H_0; the two-sided p-value is about .0002128. There is evidence of a positive mean A-minus-B wear difference under this model. A matching added 95% interval is

$$.48 \pm 2.7764451(.0374166) = (.3761, .5839).$$

The slide's 12.8233 results from its rounded $s_d = .0837$; the difference is harmless rounding.

Optional clarification of the motivating slide

Treating the data as independent gives the slide's .5740 statistic, but violates the design. That hypothetical independent procedure would have $df = 8$, not 4. Its large SE discards the within-car information.

The probability 2^{-5} for all five differences to be positive requires independent, symmetric continuous null differences (or an equivalent fair-sign model); zero mean alone does not suffice. It is a one-direction event. A two-sided all-same-sign calculation is $2/32 = .0625$, not .03125. The paired t test also uses the magnitudes and their variability; it is not that sign calculation.

The fixed A-then-B order could confound tire type with period effects. The statistical calculation assumes a suitable model; randomized or balanced ordering would strengthen a causal comparison.

Choose the design, then choose the calculation

Lecture slides: PDF pages 2–62, review (open)

Three questions to ask before using a formula

1. Am I studying one population mean or comparing two means?
2. For two means, are the groups independent or are observations linked in pairs?
3. Do the population shape and spread assumptions justify this t procedure?

Design	Estimate	Estimated SE	df
One mean	$\bar{x}$	$s/\sqrt{n}$	$n-1$
Independent means, equal variances	$\bar{x}_1 - \bar{x}_2$	$s\sqrt{1/n_1 + 1/n_2}$	$n_1 + n_2 - 2$
Paired measurements	$\bar{d}$	$s_d/\sqrt{n}$	$n-1$ pairs

$$\boxed{\text{test statistic} = \frac{\text{estimate} - \text{null value}}{SE}} \qquad \boxed{\text{CI} = \text{estimate} \pm t_{df,\alpha/2}\, SE.}$$

Do not choose a method from sample size alone. These are mean-inference procedures under stated assumptions; a small binary sample does not automatically qualify for a t test.

A complete test in six lines

1. Define the population parameter, units and subtraction order.
2. State H_0 and H_a in terms of that parameter; choose α.
3. Identify the independent units and state the model assumptions.
4. Calculate the estimate, SD, SE, df, and observed test statistic.
5. Use the correct tail(s) to find a cutoff or p-value.
6. State reject or fail to reject, then interpret the evidence in context. Include the estimated effect and units.

Short checks

1. Independent or paired? Different seedlings in each treatment group are independent groups. The same seedlings measured twice supply pairs. Matching comes from the design.

2. A large p-value? Insufficient evidence against this null under the model. It does not establish equality, no biological effect, or correct assumptions.

3. A small p-value means an important effect? No. The size of the difference, its units, interval and biological context matter too.

4. A 95% CI? A procedure with 95% repeated-sampling coverage of the parameter, not a range containing 95% of individual observations.