Chapter 9 Why Student's t appears

Detailed notes pp. 1–4

Lecture slides PDF pp. 1–18

The testing framework is unchanged: estimate minus null value, divided by its estimated SE. What changes is the reference distribution when population variability must be estimated.

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

For an independent random sample from a normal population, the first quantity is standard normal when σ is known, while the second has an exact t_{n-1} distribution when σ is unknown. Both numerator and denominator of T vary across samples.

Degrees of freedom and heavier tails

$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}.$$

The residuals sum to zero after estimating one mean, leaving $n-1$ degrees of freedom. Student's t curves are symmetric about zero with heavier tails than a standard normal curve; the extra tail probability reflects uncertainty in the estimated SD. As degrees of freedom increase, the curves approach the normal curve.

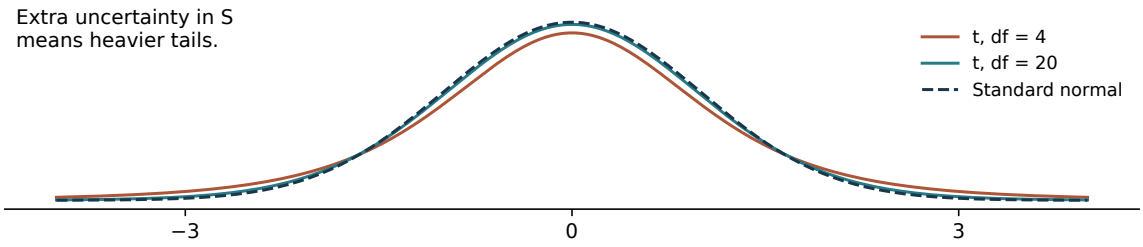
Read the upper-tail notation carefully

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

A two-sided confidence level $1-\alpha$ uses $t_{df,\alpha/2}$. A right-sided level-α test uses $t_{df,\alpha}$; a left-sided test uses its negative. For example, $t_{9,.025} = 2.262$; its normal limit is $z_{.025} = 1.960$.

The assumption is not created by the sample size

Small n alone does not justify t. Exact results need the stated normal-sampling assumptions. Approximate normality gives approximate inference; strong skew, outliers or dependence can be serious at small sample sizes. The exact one-sample t result also holds for large normal samples.



Unknown population SD introduces extra uncertainty through the random S.

Keep in mind. SD describes individuals. Estimated SE, $s/\sqrt{n}$, describes the sample mean's variability. Do not substitute one for the other.

Chapter 9 One population mean

Detailed notes pp. 5–9
Lecture slides PDF pp. 19–33

Conditions and target

Use an independent random sample from a normal population with unknown SD. The target is its mean μ; observed summaries are $n, \bar{x}, s$. Set $df = n - 1$ and estimated SE $= s/\sqrt{n}$.

Confidence interval

$$\boxed{CI_{1-\alpha}(\mu): \quad \bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}.}$$

Interpret confidence as repeated-sample coverage of the fixed population mean. The interval does not contain a stated percentage of individual observations.

Hypothesis test

For $H_0 : \mu = \mu_0$, calculate

$$\boxed{t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \qquad T \sim t_{n-1} \text{ under } H_0.}$$

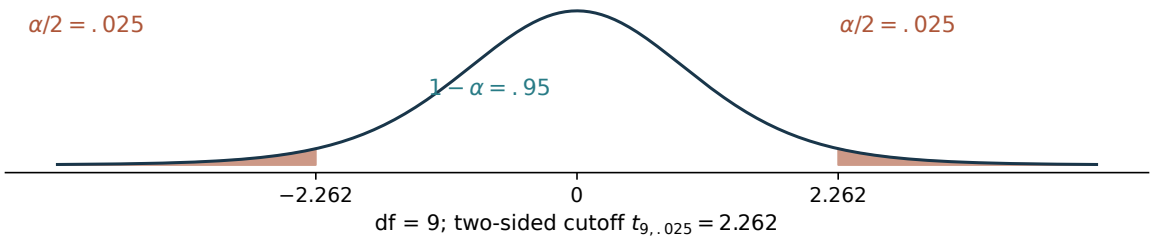
Let F_{df} denote the CDF of a t variable with df degrees of freedom:

Alternative	p-value	Rejection rule
$\mu > \mu_0$	$1 - F_{df}(t_{\text{obs}})$	$t_{\text{obs}} > t_{df,\alpha}$
$\mu < \mu_0$	$F_{df}(t_{\text{obs}})$	$t_{\text{obs}} < -t_{df,\alpha}$
$\mu \neq \mu_0$	$2[1 - F_{df}(\lvert t_{\text{obs}}\rvert)]$	$\lvert t_{\text{obs}}\rvert > t_{df,\alpha/2}$

The tail logic is the same as Chapter 8; the reference is t, not Φ.

Connect test and interval

For the matching two-sided t test and CI, reject $H_0 : \mu = \mu_0$ at level α exactly when the $1 - \alpha$ interval excludes μ_0, apart from boundary conventions. A two-sided 95% CI does not directly implement every one-sided 5% decision.



Two-sided inference splits the total error probability across two tails.

Keep in mind. A table can give a p-value range rather than a precise decimal. Do not manufacture extra precision when only bounding table entries are available.

Chapter 9 Two independent means

Detailed notes pp. 10–14
Lecture slides PDF pp. 34–55

The pooled model has an extra assumption

Two independent random samples come from normal populations with **a common unknown population variance** σ^2. The sample SDs need not be numerically equal. The course uses $1/3 < s_1^2/s_2^2 < 3$ as a screening rule of thumb (slide p. 46; detailed p. 11). This is not proof that the population variances are equal. Define $\Delta = \mu_1 - \mu_2$ with a fixed group order.

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

Here s^2 and s mean the pooled variance and SD, as in the slides. The pooled variance combines squared deviations, weighted by their degrees of freedom. It is not the simple average of SDs.

$$\widehat{\mathrm{SE}}(\bar{x}_1 - \bar{x}_2) = s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \qquad df = n_1 + n_2 - 2.$$

Interval and test

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

For $H_0 : \Delta = \Delta_0$,

$$\boxed{t_{\mathrm{obs}} = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s\sqrt{1/n_1 + 1/n_2}}.}$$

Use the t tails from p. 2 with $df = n_1 + n_2 - 2$. Under the stated normal and equal-variance assumptions, this reference is exact.

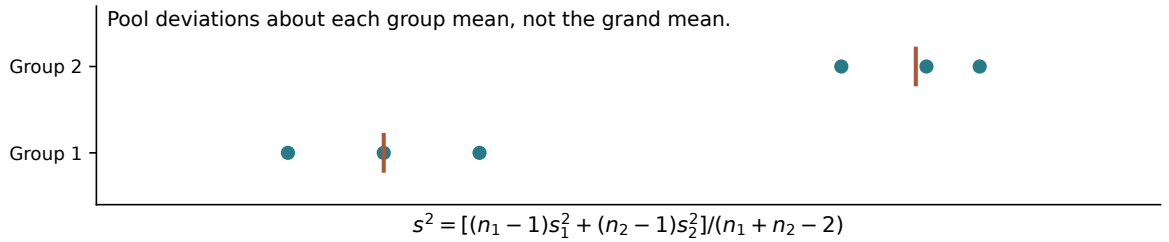
What pooling means here

Detailed notes pp. 11: pooled variance and its assumptions

This pools estimates of a *common variance*. Chapter 8's two-proportion test pools a *common success probability under the null*. They are different models; sharing the word "pooled" does not make their formulas interchangeable.

Check design and assumptions before computing

Independent individuals justify independent groups; equal sample sizes do not create pairs. Without a justified common population variance, use an unequal-variance procedure, beyond this deckś pooled formula.



Estimate common variability about each group mean; two fitted means use two degrees of freedom.

Keep in mind. The online/classroom example on detailed pp. 12–14 corrects source arithmetic while retaining the original data. Use its checked calculations when reviewing the slide example.

Chapter 9 Paired differences and method choice

Detailed notes pp. 15–17

Lecture slides PDF pp. 56–62

When the same individuals are measured twice, or observations are explicitly matched, analyze within-pair differences. Choose an order and define $D_i = X_i - Y_i$; the target is μ_D.

Reduce to one-sample t inference

For n independent pairs whose differences are normally distributed,

$$\bar{d} = \frac{1}{n}\sum d_i, \qquad s_D^2 = \frac{\sum(d_i - \bar{d})^2}{n-1}, \qquad df = n - 1.$$

$$\boxed{CI_{1-\alpha}(\mu_D): \quad \bar{d} \pm t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}}}$$

$$\boxed{t_{\text{obs}} = \frac{\bar{d} - d_0}{s_D/\sqrt{n}} \quad \text{for } H_0 : \mu_D = d_0.}$$

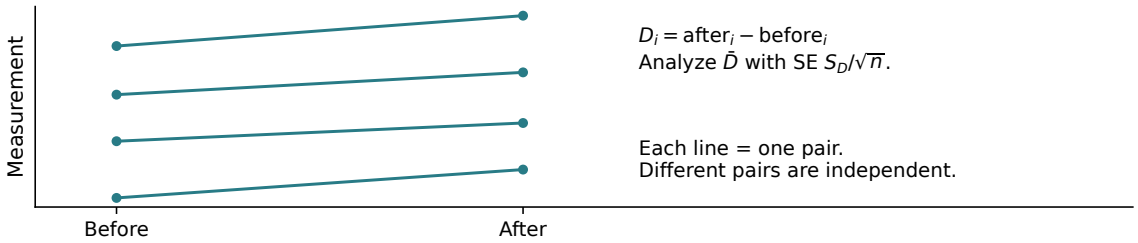
The required normality concerns the *differences*, and the required independence concerns different pairs. Within-pair dependence is expected and is retained through s_D. Reversing subtraction reverses signs and interval endpoints' order after negation; reverse the alternative consistently too.

Select by study design

Design	Course small-sample normal method
One sample; one population mean	One-sample t, $df = n - 1$.
Separate independent groups; common variance justified	Pooled two-sample t, $df = n_1 + n_2 - 2$.
Repeated or matched observations	One-sample t on differences, $df =$ number of pairs -1.

The final explanation matters

State the parameter, assumptions, estimate, SE, degrees of freedom, reference, tail and conclusion. Fail-to-reject does not prove equality. An interval for a mean difference is not an interval for every individual's change. The paired tire example on detailed p. 16 shows the workflow.



One pair gives one difference: $df = n - 1$, where n counts independent pairs.

Keep in mind. Use existing released course examples for review. Future Chapter 9 practice and review questions are selected separately; this guide does not announce a new practice set or predict the exam's questions.