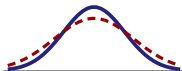


Inference for Means

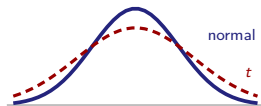
AP Statistics ■ Unit 7

AP Statistics



What we will cover

- 1 The t-distribution
- 2 Confidence interval for a mean
- 3 Test for a mean
- 4 Difference of two means
- 5 Two-mean test and paired data
- 6 Choosing a procedure





The t-distribution

We almost never know σ , so we use s_x instead. That extra guessing means the **t-distribution** t , not the normal.

- heavier tails $\Rightarrow$ a little more caution
- **degrees of freedom** $= n - 1$

Sampling variability is honest bounce – not a mistake a calculation can fix.



Two group comparison



Confidence interval for a mean

With σ unknown, use t^* with $df = n - 1$:

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

- **standard error** = $s_x / \sqrt{n}$
- **margin of error** = $t^* \cdot \text{SE}$

Check: random, $\leq 10\%$ of population,
roughly normal (or $n \geq 30$).

Batteries

$n = 25$, $\bar{x} = 50$, $s_x = 8$; $t^* = 2.064$ ($df = 24$):

$$50 \pm 2.064 \cdot \frac{8}{5} = 50 \pm 3.30$$

= (46.70, 53.30) hours.

“95% confident” describes the method



CI form: $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ · level = long-run capture rate of the method

“95% confident” describes the method, not one interval: over many samples about 95% of the intervals contain μ and about 5% miss it.



Test for a mean

A **one-sample t-test** t :

- **null** $H_0 : \mu = \mu_0$
- **alternative** $H_a : \mu >, <, \text{ or } \neq \mu_0$

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}$$

Find the p-value from the t -distribution with $df = n - 1$; reject if $p \leq \alpha$.

Batteries

Claim $\mu = 50$; sample $\bar{x} = 53$, $s_x = 8$, $n = 25$, $H_a : \mu > 50$:

$$t = \frac{53 - 50}{8/\sqrt{5}} = 1.875$$

$p \approx 0.037 < 0.05$: reject.



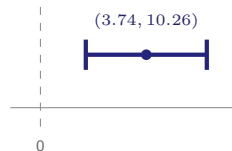
Difference of two means

The design decides the tool:

- **independent samples** $\Rightarrow$ **two-sample t-interval**
- **paired** data $\Rightarrow$ one-sample t on the differences

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Variances **add** under the root. If the interval excludes 0, the means differ.



Excludes 0 $\Rightarrow$ Brand
1 longer.



Two-mean test and paired data

A **two-sample t-test** of $H_0 : \mu_1 = \mu_2$ – do **not pool** (unlike proportions):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

Same subjects measured twice? Use a **paired t-test**: reduce each pair to a difference d , then a one-sample t .

Brands

$\bar{x}_1 = 85, \bar{x}_2 = 78; SE = \sqrt{2}$:

$$t = \frac{7}{1.414} \approx 4.95$$

$df = 8, p \approx 0.001$: reject.



Choosing a procedure

Two questions point to the right tool:

Goal?

estimate $\Rightarrow$ interval;
weigh a claim $\Rightarrow$ test

How many groups?

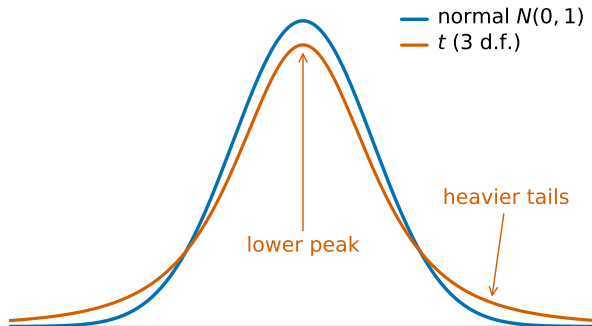
one $\Rightarrow$ one-sample; two
 $\Rightarrow$ two-sample or paired

Mean or proportion?

mean $\Rightarrow$ t ; proportion $\Rightarrow$ z

Then **select** the formula, **implement** with conditions checked, and **communicate** in context.

Confidence Interval for a Mean, continued



The t-distribution has a lower peak and heavier tails than the normal

Worked example – a one-sample t-test for a mean

Test $H_0 : \mu = 45$ against $H_a : \mu \neq 45$, given $\bar{x} = 50$, $s = 8$, $n = 25$.

- $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{50 - 45}{8/5} = \frac{5}{1.6} = 3.125$
- $df = n - 1 = 24$
- two-sided $p \approx 0.005$, so **reject** H_0 at the 5% level
- in context: the evidence says the true mean is *not* 45

As with proportions: a claimed mean **inside** the interval is plausible; **outside** it, the data give evidence against the claim. The interval and the test agree by construction.

Should I Worry About Error?

Inference for a **mean** works like inference for a proportion, with one change: we rarely know the population standard deviation σ , so we estimate it with the sample s .

That extra uncertainty means we use the t -**distribution** instead of the normal – a **distribution** that is bell-shaped but with heavier tails.

Exam tips

- Use **t-procedures** for means (population σ unknown) – the t-distribution has heavier tails than normal.
- Check conditions: random, independent, and roughly normal (or large n).
- Interpret an interval and a test in context, always tied to the parameter (the true mean).
- Match the right procedure: one-sample, two-sample, or paired (look for a natural pairing).
- State the degrees of freedom; for a two-sample t -test use technology's value (or, by hand, the conservative smaller $n - 1$).



Unit 7 – key takeaways

1

The t-model

unknown $\sigma \Rightarrow$ use s_x and t ; $df = n - 1$

2

One mean

$$\bar{x} \pm t^* s_x / \sqrt{n}; t = (\bar{x} - \mu_0) / (s_x / \sqrt{n})$$

3

Two means

add variances; don't pool; paired
 $\rightarrow$ one-sample t

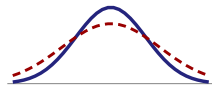
4

Choosing

goal + groups +
mean/proportion select the test

Check yourself

- 1 Why use t instead of z for a mean?
- 2 What are the degrees of freedom for a one-sample t ?
- 3 Do you pool for a two-sample t -test of means?
- 4 Before-and-after on the same subjects calls for which test?



Answers 1. σ is unknown (use s_x) 2. $n - 1$ 3. no 4. a paired t -test