Unit 7: Inference for Means

Merrick Fanning

February 22, 2026

Unit 7 Outline: Inference for Means

1. One-Sample Confidence Interval for μ (z-intervals, t-intervals)
2. One-Sample Hypothesis Test for μ (z-tests, t-tests)
3. Two-Sample Confidence Interval $\mu_1 - \mu_2$ (t-tests)
4. Two-Sample t-test for $\mu_1 - \mu_2$ (t-tests)
5. Matched-Pairs t-tests

Choosing Between Z and t Procedures for Means

Is σ known?

Yes →

Z-test: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Z-interval: $\bar{x} \pm z^* \dfrac{\sigma}{\sqrt{n}}$

No →

t-test: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

t-interval: $\bar{x} \pm t^* \dfrac{s}{\sqrt{n}}$

Which type of test do you believe is more useful in practice?

Example Scenario

Question: How much does an Oreo cookie weigh on average?

A random sample of 30 Oreo cookies is selected. The sample has:

- Sample mean: $\bar{x} = 11.1921$ grams
- Sample standard deviation: $s = 0.0817$ grams
- Sample size: $n = 30$

We will construct a 95% confidence interval for the true mean weight μ of an Oreo cookie.

Step 1: State the Problem

We are estimating the true mean weight μ of all Oreo cookies.

Given:

- Sample size: $n = 30$
- Sample mean: $\bar{x} = 11.1921$
- Sample standard deviation: $s = 0.0817$
- Confidence level: 95%

We will use a one-sample t-interval for a population mean.

Step 2: Check Conditions

1. Random: The sample was randomly selected.

2. Independence: $30 < 10\%$ of all Oreo cookies produced, so observations are approximately independent.

3. Normal/Large Sample: The sample size is $n = 30$, which is large enough to use the t-distribution by the Central Limit Theorem.

Step 3: Calculate the Interval

We use the formula:

$$\text{Confidence Interval} = \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

Where:

- $\bar{x} = 11.1921$
- $s = 0.0817$
- $n = 30$
- Degrees of freedom: $df = 29$
- t^* for 95% confidence with 29 degrees of freedom is approximately 2.045

$$\text{Margin of error} = 2.045 \cdot \frac{0.0817}{\sqrt{30}} \approx 0.0305$$

$$\text{Interval: } 11.1921 \pm 0.0305 = (11.1616,\ 11.2226)$$

Step 4: Interpretation

We are 95% confident that the true mean weight μ of all Oreo cookies is between 11.1616 grams and 11.2226 grams.

This means that if we took many random samples of 30 Oreo cookies and built a confidence interval from each, about 95% of those intervals would capture the true mean weight.

Application: Comparing to a Claim

Nabisco claims that the average Oreo weighs 11.3 grams.

Does our interval provide evidence that the true mean is less than 11.3 grams?

Yes. The value 11.3 grams is not included in the 95% confidence interval from 11.1616 to 11.2226 grams. This suggests the average weight of an Oreo is likely less than 11.3 grams.

Common Misunderstanding

Incorrect: “There is a 95% probability that the true mean is in the interval.”

Correct: The true mean μ is a fixed value. We are either capturing it or not.

What is random is the process. The confidence level means that if we repeated this process many times, 95% of those intervals would capture the true mean.

One Sample t-tests for Means

Context: Students should get 8 hours of sleep per night.

A researcher collects a **random sample** of 10 university students and records the number of hours they sleep per night:

$$6.5,\ 4.75,\ 5,\ 8,\ 7,\ 7.25,\ 4.5,\ 5.5,\ 7,\ 6$$

Do the data provide convincing evidence that students sleep **less than 8 hours per night**?
Use significance level $\alpha = 0.05$.

Step 1: State the Hypotheses

We are conducting a **one-sample t-test** for the population mean.

Let μ be the **true mean hours of sleep** for all university students.

$$H_0 : \mu = 8 \qquad H_a : \mu < 8$$

This is a **left-tailed test**.

Step 2: Check Conditions

1. Random: The sample was randomly selected.

2. Independence: The sample size $n = 10$ is less than 10% of all university students.

3. Normality: Sample size is small ($n < 30$), but the data appear roughly symmetric and free of outliers.

$\Rightarrow$ Conditions met to proceed with a t-test.

Step 3: Calculate Test Statistic

We compute the sample mean and standard deviation:

$$\bar{x} = 6.15, \quad s = 1.18$$

Then we calculate the standard error:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1.18}{\sqrt{10}} \approx 0.376$$

Now calculate the *t*-statistic:

$$t = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{6.15 - 8}{0.376} \approx -4.92$$

Step 3 (continued): p-value and Visual

Degrees of freedom: $df = 9$

p-value:

$$P(t_9 < -4.92) \approx 0.0004$$

Visual: Sketch the sampling distribution, rejection region, and statistic.

Step 4: Conclusion

Compare:

$$\text{p-value} = 0.0004 < \alpha = 0.05$$

Conclusion: Reject H_0.

There is convincing evidence that students at this university get **less than 8 hours** of sleep per night on average.

If the true mean were 8 hours, the chance of observing a sample mean as low as 6.15 (or lower) is only 0.04%.

Confidence Interval for a Difference of Means

Example: Researchers want to know if caffeine affects reaction time. Two random samples of students are tested on a reaction time task (in milliseconds). Reaction Time Test.

- Group 1: 8 students drank **coffee** before the test
- Group 2: 10 students had **no caffeine** before the test
- Sample statistics:

$$\bar{x}_1 = 245 \text{ ms}, \quad s_1 = 18 \text{ ms}, \quad n_1 = 8$$

$$\bar{x}_2 = 260 \text{ ms}, \quad s_2 = 20 \text{ ms}, \quad n_2 = 10$$

We want a 95% confidence interval for $\mu_1 - \mu_2$, the true difference in mean reaction times (coffee − no caffeine).

Check Conditions for 2-Sample t Interval

Conditions:

1. **Random:** Both groups were randomly selected.
2. **Independent:** The two groups are independent and each sample is less than 10% of the population.
3. **Normal:**
   - If $n_1, n_2 \geq 30$, the Central Limit Theorem applies automatically.
   - For smaller samples ($n_1 = 8$, $n_2 = 10$), we must check that the data are roughly symmetric and free of outliers.

Calculate the Interval

Step 1: Compute the point estimate and standard error.

$$\bar{x}_1 - \bar{x}_2 = 245 - 260 = -15 \text{ ms}$$

$$SE = \sqrt{\frac{18^2}{8} + \frac{20^2}{10}} \approx 8.57 \text{ ms}$$

Step 2: Determine the critical t. Degrees of freedom ≈ 14 (from software or conservative $\min(n_1 - 1,\ n_2 - 1)$)

$$t^* \approx 2.145 \quad \text{for } 95\%$$

Step 3: Construct the CI.

$$(-15) \pm 2.145(8.57) \approx (-33.4,\ 3.4) \text{ ms}$$

Interpretation of the Interval

- We are 95% confident that the **true difference in mean reaction times** (coffee − no caffeine) is between −33.4 ms and 3.4 ms.
- Because the interval includes 0, we do not have strong evidence that caffeine changes reaction time.
- A negative difference suggests coffee may slightly **reduce** reaction time, but it is not statistically conclusive.

Two-Sample t-Test Example: Push-ups

We want to know if older students (Div 4) can do more push-ups on average than younger students (Div 2).

- $H_0 : \mu_{Div2} = \mu_{Div4}$
- $H_a : \mu_{Div2} < \mu_{Div4}$ (older students do more)

Sample statistics:

$$\bar{x}_1 = 8.1, \quad s_1 = 2.2, \quad n_1 = 8$$

$$\bar{x}_2 = 21.4, \quad s_2 = 2.3, \quad n_2 = 10$$

Check Conditions

- **Random:** Students sampled randomly from each division.
- **Independence:** Div 2 and Div 4 students are independent groups. Each sample is less than 10% of their respective population.
- **Normality:**
  - Sample sizes are small ($n_1 = 8$, $n_2 = 10$)
  - Sample data appear roughly symmetric (no outliers). **DRAW HISTOGRAM/DOTPLOT**

Calculate Test Statistic

Two-sample t-statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Plug in sample values:

$$t = \frac{8.1 - 21.4}{\sqrt{\frac{2.2^2}{8} + \frac{2.3^2}{10}}} \approx -13.0$$

Degrees of freedom (Welch's Approximation):

$$df \approx 15$$

Conclusion

Test statistic: $t \approx -13.0$
Degrees of freedom: $df \approx 15$
Two-tailed p-value: $p < 0.0001$

Since $p < \alpha = 0.05$ we reject H_0 in favour of the alternative hypothesis. Assuming the null hypothesis is true (that Division 2 and Division 4 students can do the same number of push-ups on average), the probability of observing a difference this extreme or more extreme is essentially 0.

This is extremely strong evidence that Division 2 students can do fewer push-ups on average than Division 4 students.

Paired t-Test: Sprint Times

Scenario:

- 10 Division 4 students run a 200m sprint.
- They drink a sports drink and rest 10 minutes.
- They run the 200m sprint again.

We want to know if the sports drink improves sprint times.

Paired Design: Same student before and after, so we use the differences.

Conditions for Paired t-Test

- Random Sample: Students selected randomly from Division 4 PE class.
- Independence: Each student's times are independent of others.
- Normality of Differences:
    - Compute $d = \text{Before} - \text{After}$
    - Histogram of d roughly symmetric OR sample size $n \geq 30$ for CLT

Hypothesis Test Setup

Let d =Before − After
Hypotheses:

$$H_0 : \mu_d = 0 \quad H_a : \mu_d > 0$$

(sports drink decreases sprint times)
Sample Data (seconds):

Before	38	40	36	37	39	41	42	40	38	37
After	36	39	35	35	37	39	40	39	37	35

Differences d = [2, 1, 1, 2, 2, 2, 2, 1, 1, 2]

$$\bar{d} = 1.6, \quad s_d = 0.52, \quad n = 10$$

$$t = \frac{1.6}{0.52/\sqrt{10}} \approx 9.7$$

Conclusion

Test statistic: $t \approx 9.7$
Degrees of freedom: $df = 9$
One-tailed p-value < 0.0001
Decision: Reject H_0
Plain Language:
Assuming the sports drink has no effect, the probability of observing an average improvement of 1.6 seconds or more is essentially zero.
There is very strong evidence that the sports drink improves sprint times.

Confidence Interval for Mean Difference

We can also construct a 95% confidence interval for the mean difference:

$$\bar{d} \pm t^* \frac{s_d}{\sqrt{n}}$$

- $\bar{d} = 1.6$
- $s_d = 0.52$
- $n = 10$
- $df = 9$, $t^* \approx 2.262$ for 95% CI

$$1.6 \pm 2.262 \cdot \frac{0.52}{\sqrt{10}} \approx 1.6 \pm 0.37$$

95% CI for μ_d: (1.23, 1.97)

Interpretation: We are 95% confident the sports drink improves sprint times by between 1.23 and 1.97 seconds on average.

Beyond AP Statistics: Inference for Means in College

AP Statistics provides a strong foundation, but college courses will go deeper. Here are common topics you **don't see in Unit 7** that you may encounter later:

- **Pooled t-tests** (Assume equal variances; different formula for SE and df.)
- **Nonparametric tests** for medians or ranks when normality is questionable:
  - Wilcoxon Signed-Rank Test (paired data)
  - Mann–Whitney U Test (two independent samples)
- **Matched designs with blocking** and more complex ANOVA models.
- **Power and sample size calculations** for means.
- **Effect size measures** (Cohen's d and others) to quantify magnitude of differences.
- **Robust methods** for outlier-heavy or skewed data.
- **Multiple comparisons** adjustments (Bonferroni, Tukey, etc.).
- **Bootstrapping** and **permutation tests** for means without relying on t-procedures.

AP Statistics focuses on understanding core inference logic and conditions. College courses expand to handle more situations, more assumptions, and more robust tools.