0.1 Learning Objectives

By the end of this lecture, you will be able to:

- Understand why statistical tests are needed
- Define null and alternative hypotheses
- Calculate and interpret test statistics
- Interpret p-values correctly
- Distinguish one-tailed from two-tailed tests

1. Why Statistical Tests?

1.1 Warm-Up

Two UC Davis students live the same distance from campus.

Student A bikes to class. Student B drives.

1.1 Warm-Up

Two UC Davis students live the same distance from campus.

Student A bikes to class. Student B drives.

Who has a shorter commute?

1.1 Warm-Up

Two UC Davis students live the same distance from campus.

Student A bikes to class. Student B drives.

Who has a shorter commute?

Try it yourself

Talk to your neighbor and try to solve this problem.

1.2 Warm-Up: One Data Point

We time their commutes one morning:

Student A (bike): 12 minutes

Student B (car): 18 minutes

1.2 Warm-Up: One Data Point

We time their commutes one morning:

Student A (bike): 12 minutes

Student B (car): 18 minutes

Is biking faster than driving at UC Davis?

1.2 Warm-Up: One Data Point

We time their commutes one morning:

Student A (bike): 12 minutes

Student B (car): 18 minutes

Is biking faster than driving at UC Davis?

Can we conclude anything from a single observation per group?

1.3 Warm-Up: More Data

We survey six students and time their Monday morning commutes:

Student	Commute (min)	Mode
1	11	Bike
2	14	Bike
3	9	Bike
4	16	Car
5	22	Car
6	13	Car

1.3 Warm-Up: More Data

1.3 Warm-Up: More Data

We survey six students and time their Monday morning commutes:

Student	Commute (min)	Mode
1	11	Bike
2	14	Bike
3	9	Bike
4	16	Car
5	22	Car
6	13	Car

Is biking faster than driving? Why or why not?

1.3 Warm-Up: More Data

1.3 Warm-Up: More Data

We survey six students and time their Monday morning commutes:

Student	Commute (min)	Mode
1	11	Bike
2	14	Bike
3	9	Bike
4	16	Car
5	22	Car
6	13	Car

Is biking faster than driving? Why or why not?

1.3 Warm-Up: More Data

Try it yourself

Talk to your neighbor and try to solve this problem.

1.4 The Core Problem

Our observations are the result of a *random* process.

1.4 The Core Problem

Our observations are the result of a *random* process.

Sometimes, we observe two samples that differ widely. But is this difference due to the variable of interest (commute mode), or is it due to **random chance**?

1.4 The Core Problem

Our observations are the result of a *random* process.

Sometimes, we observe two samples that differ widely. But is this difference due to the variable of interest (commute mode), or is it due to **random chance**?

Note: "Random chance" collects the influence of *all other variables* we aren't measuring: route chosen, time of departure, traffic lights, finding parking, etc.

1.5 Statistical Tests

We often want to make statements about the **population** based on a **sample**.

1.5 Statistical Tests

We often want to make statements about the **population** based on a **sample**.

But finite samples present challenges.

1.5 Statistical Tests

We often want to make statements about the **population** based on a **sample**.

But finite samples present challenges.

Statistical tests help us quantify how confident we should be in conclusions drawn from finite samples.

2. Sample vs. Population

2.1 Population Statistics

Definition: Population Parameters

For a population of size N:

Population mean: $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$

Population variance: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$

2.1 Population Statistics

Definition: Population Parameters

For a population of size N:

Population mean: $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$

Population variance: $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$

We rarely have access to the entire population.

2.2 Sample Statistics

Definition: Sample Statistics

For a sample of size n drawn from the population:

Sample mean: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2$

2.2 Sample Statistics

Definition: Sample Statistics

For a sample of size n drawn from the population:

Sample mean: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2$

Note: Why $n-1$? Dividing by $n-1$ gives an **unbiased estimate** of the population variance: $E(s^2) = \sigma^2$. Dividing by n would systematically underestimate σ^2.

2.3 Sample vs. Population: Summary

	Sample Statistic	Population Parameter
Mean	$\overline{x} = \frac{1}{n} \sum x_i$	$\mu = \frac{1}{N} \sum X_i$
Variance	$s^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2$	$\sigma^2 = \frac{1}{N} \sum (X_i - \mu)^2$

2.3 Sample vs. Population: Summary

	Sample Statistic	**Population Parameter**
Mean	$\overline{x} = \frac{1}{n} \sum x_i$	$\mu = \frac{1}{N} \sum X_i$
Variance	$s^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2$	$\sigma^2 = \frac{1}{N} \sum (X_i - \mu)^2$

We use **sample statistics** to *estimate* population parameters.

3. Hypothesis Testing Framework

3.1 Key Terminology

Definition: Null Hypothesis (H_0)

The hypothesis that there is **no effect** or **no difference**. For example: “CS and CSEng majors have the same mean income.”

3.1 Key Terminology

Definition: Null Hypothesis (H_0)

The hypothesis that there is **no effect** or **no difference**. For example: “CS and CSEng majors have the same mean income.”

Definition: Alternative Hypothesis (H_1)

The hypothesis that the null is **not true**. For example: “CS and CSEng majors have different mean incomes.”

3.2 More Terminology

Definition: Test Statistic

A number computed from the data that **summarizes the evidence** against H_0.

3.2 More Terminology

Definition: Test Statistic

A number computed from the data that **summarizes the evidence** against H_0.

Definition: Null Distribution

The distribution of the test statistic **assuming H_0 is true**.

3.2 More Terminology

Definition: Test Statistic

A number computed from the data that **summarizes the evidence** against H_0.

Definition: Null Distribution

The distribution of the test statistic **assuming H_0 is true**.

If our observed test statistic is extreme relative to the null distribution, we have evidence against H_0.

3.3 The t-Test

Definition: t-Test

A statistical test that determines whether the difference between two sample means is likely due to chance.

3.3 The t-Test

Definition: t-Test

A statistical test that determines whether the difference between two sample means is likely due to chance.

The idea: if both samples came from the *same* population, how likely is it to see a difference this large?

3.4 The t-Statistic

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

3.4 The t-Statistic

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

where:

- $\overline{x}_1, \overline{x}_2$ are the sample means
- s_1^2, s_2^2 are the sample variances
- n_1, n_2 are the sample sizes

3.4 The t-Statistic

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

where:

- $\overline{x}_1, \overline{x}_2$ are the sample means
- s_1^2, s_2^2 are the sample variances
- n_1, n_2 are the sample sizes

A larger $|t|$ indicates a greater difference relative to the variability in the data.

3.5 The Student's t-Distribution

The **Student's t-distribution** describes the distribution of t-statistics when both samples come from the same normal distribution.

3.5 The Student’s t-Distribution

The **Student’s t-distribution** describes the distribution of t-statistics when both samples come from the same normal distribution.

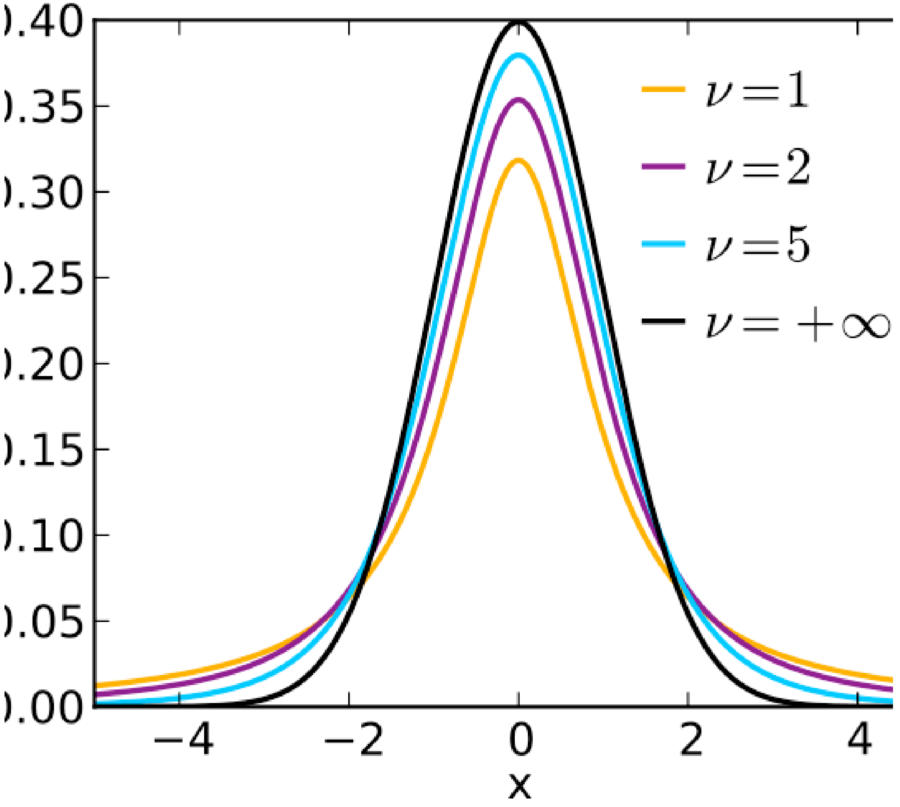
Looks like the normal distribution, but with **fatter tails** (more probability in the extremes).

3.5 The Student's t-Distribution

The **Student's t-distribution** describes the distribution of t-statistics when both samples come from the same normal distribution.

Looks like the normal distribution, but with **fatter tails** (more probability in the extremes).

Named after William Sealy Gosset, who published under the pseudonym "Student".



3.6 Try It: Reading the t-Distribution

Given

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

Where on the distribution is the region for $t \geq 2.5$?

3.6 Try It: Reading the t-Distribution

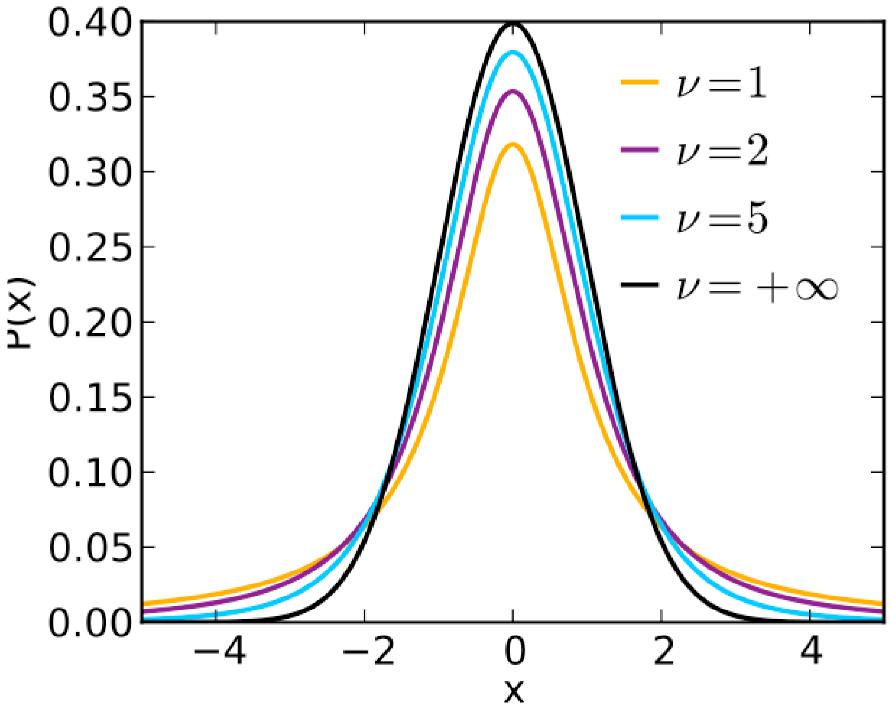
Given

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

Where on the distribution is the region for $t \geq 2.5$?

Try it yourself

Talk to your neighbor and try to solve this problem.



4. p-Values

4.1 Understanding p-Values

Definition: p-Value

The probability of observing data **as extreme as ours** (or more extreme) **if the null hypothesis is true**.

4.1 Understanding p-Values

Definition: p-Value

The probability of observing data **as extreme as ours** (or more extreme) **if the null hypothesis is true**.

We say a result is **significant at the 5% level** if the p-value is less than 0.05.

4.1 Understanding p-Values

Definition: p-Value

The probability of observing data **as extreme as ours** (or more extreme) **if the null hypothesis is true**.

We say a result is **significant at the 5% level** if the p-value is less than 0.05.

This means: there is only a 5% chance of seeing such an extreme result if H_0 is true.

4.2 One-Tailed vs. Two-Tailed Tests

One-tailed test: We test whether the difference is in a **specific direction**.

- H_1: $\mu_1 > \mu_2$ (or $\mu_1 < \mu_2$)

4.2 One-Tailed vs. Two-Tailed Tests

One-tailed test: We test whether the difference is in a **specific direction**.

- H_1: $\mu_1 > \mu_2$ (or $\mu_1 < \mu_2$)

Two-tailed test: We test whether there is **any** difference.

- H_1: $\mu_1 \neq \mu_2$

4.2 One-Tailed vs. Two-Tailed Tests

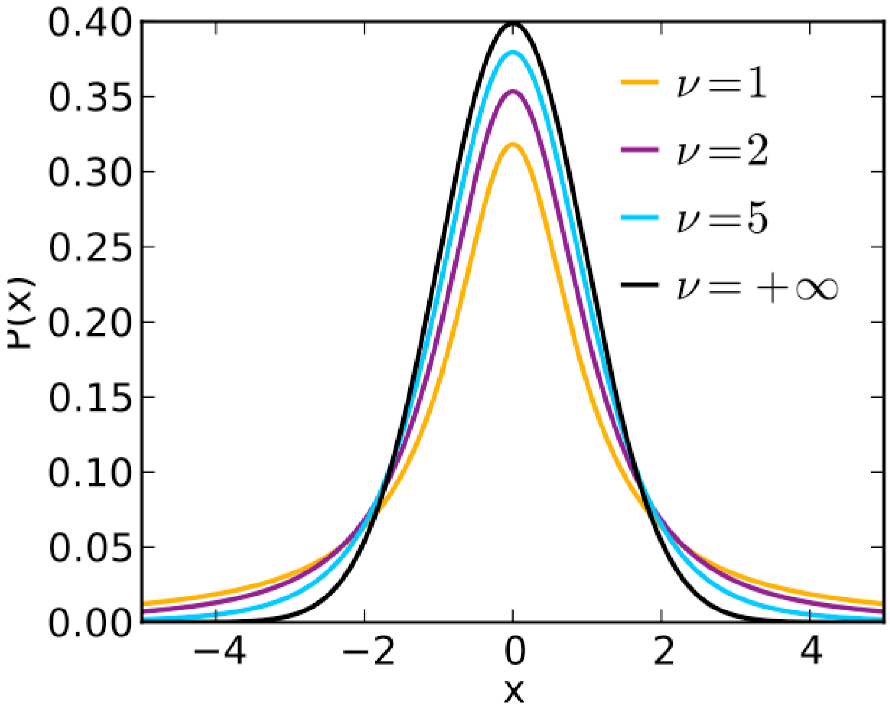
One-tailed test: We test whether the difference is in a **specific direction**.

- H_1: $\mu_1 > \mu_2$ (or $\mu_1 < \mu_2$)

Two-tailed test: We test whether there is **any** difference.

- H_1: $\mu_1 \neq \mu_2$

Two-tailed is more conservative (harder to reject H_0).



4.3 Try It: Interpreting a p-Value

Example: Commute Study

"We performed a t-test comparing mean commute times of bikers and drivers at UC Davis. We found that drivers take 5.3 minutes longer on average, with a p-value of 0.043."

In your own words, what does this mean?

4.3 Try It: Interpreting a p-Value

Example: Commute Study

“We performed a t-test comparing mean commute times of bikers and drivers at UC Davis. We found that drivers take 5.3 minutes longer on average, with a p-value of 0.043.”

In your own words, what does this mean?

Try it yourself

Talk to your neighbor and try to solve this problem.

4.4 Solution: Interpreting the p-Value

There is a 4.3% probability of observing a difference in means of 5.3 minutes or more **if biking and driving commute times actually follow the same distribution**.

4.4 Solution: Interpreting the p-Value

There is a 4.3% probability of observing a difference in means of 5.3 minutes or more **if biking and driving commute times actually follow the same distribution**.

At the 5% significance level ($\alpha = 0.05$): we **reject** H_0 since $0.043 < 0.05$.

4.4 Solution: Interpreting the p-Value

There is a 4.3% probability of observing a difference in means of 5.3 minutes or more **if biking and driving commute times actually follow the same distribution**.

At the 5% significance level ($\alpha = 0.05$): we **reject** H_0 since $0.043 < 0.05$.

At the 1% significance level ($\alpha = 0.01$): we **fail to reject** H_0 since $0.043 > 0.01$.

4.4 Solution: Interpreting the p-Value

There is a 4.3% probability of observing a difference in means of 5.3 minutes or more **if biking and driving commute times actually follow the same distribution.**

At the 5% significance level ($\alpha = 0.05$): we **reject** H_0 since $0.043 < 0.05$.

At the 1% significance level ($\alpha = 0.01$): we **fail to reject** H_0 since $0.043 > 0.01$.

Note: Rejecting H_0 does **not** prove H_1. It means the data is unlikely under H_0.

5. Significance Levels and Errors

5.1 Choosing the Significance Level

The **significance level** α is our tolerance for making a mistake.

5.1 Choosing the Significance Level

The **significance level** α is our tolerance for making a mistake.

Definition: Type I Error (False Positive)

Rejecting H_0 when it is actually true. Probability = α.

5.1 Choosing the Significance Level

The **significance level** α is our tolerance for making a mistake.

Definition: Type I Error (False Positive)

Rejecting H_0 when it is actually true. Probability = α.

Definition: Type II Error (False Negative)

Failing to reject H_0 when it is actually false. Probability = β.

5.2 The Error Trade-Off

	H_0 True	H_0 False
Reject H_0	Type I Error (α)	Correct!
Fail to reject H_0	Correct!	Type II Error (β)

5.2 The Error Trade-Off

	H_0 True	H_0 False
Reject H_0	Type I Error (α)	Correct!
Fail to reject H_0	Correct!	Type II Error (β)

- Decreasing α (stricter threshold) $\rightarrow$ fewer false positives, but more false negatives
- Common choices: $\alpha = 0.05$, $\alpha = 0.01$, $\alpha = 0.001$

5.3 Other Statistical Tests

The t-test compares two means. Other tests exist for different situations:

5.3 Other Statistical Tests

The t-test compares two means. Other tests exist for different situations:

- **Chi-Square Test:** Compares observed vs. expected categorical distributions

5.3 Other Statistical Tests

The t-test compares two means. Other tests exist for different situations:

- **Chi-Square Test:** Compares observed vs. expected categorical distributions
- **ANOVA:** Compares means of **multiple** groups (generalization of t-test)

5.3 Other Statistical Tests

The t-test compares two means. Other tests exist for different situations:

- **Chi-Square Test:** Compares observed vs. expected categorical distributions
- **ANOVA:** Compares means of **multiple** groups (generalization of t-test)
- **Correlation test:** Tests whether a linear relationship exists between two variables

5.3 Other Statistical Tests

The t-test compares two means. Other tests exist for different situations:

- **Chi-Square Test:** Compares observed vs. expected categorical distributions
- **ANOVA:** Compares means of **multiple** groups (generalization of t-test)
- **Correlation test:** Tests whether a linear relationship exists between two variables

In this course, we focus on the **concepts**

5.3 Other Statistical Tests

the specific test depends on your data and question.

5.4 Recap

Today we covered:

- **Statistical tests** quantify uncertainty when drawing conclusions from finite samples
- **Null hypothesis** (H_0): assume no effect; **alternative** (H_1): there is an effect
- The **t-statistic** measures how different two means are, relative to variability
- The **p-value** is the probability of seeing data this extreme if H_0 is true
- **Type I error** (false positive) rate is controlled by the significance level α