Objectives

By the end of this set of slides, you should be able to:

1. Differentiate between Type I and Type II Errors
2. Describe hypothesis testing in general and in practice
3. Conduct and interpret hypothesis testing for a mean
4. Conduct and interpret hypothesis testing for a proportion
Inferential Statistics

- **Inferential Statistics**: Using sample data to make generalizations about an unknown population.
- Confidence intervals were one way to do statistical inference.
- Another way to do statistical inference is to make a decision about the value of a parameter.
- Example: A car dealer claims that its new small truck gets 35 miles per gallon on average.
- Example: A company says that women managers in their company earn an average of $60,000 per year.
- Example: A tutoring service claims that its method of tutoring helps 90% of its students get an A or B.
- How do we test the "claims" in the examples?
- Hypothesis Testing!
What is a Hypothesis?

- **Hypothesis**: A statement about the value of a population parameter developed for the purpose of testing.

- Example: The mean monthly income for computer programmers is $9,000

- Example: At least 20% of all juvenile offenders are caught and sentenced to prison

- Example: The standard deviation for an investment portfolio is no more than 10% month

- So **hypothesis testing** is going to be a way to prove (or disprove) whether or not a hypothesis is true or false
Hypothesis Testing

• **Hypothesis Testing**: A procedure, based on sample evidence and probability theory, used to determine whether the hypothesis is a reasonable statement and should not be rejected, or is unreasonable and should be rejected

• **Steps to perform a hypothesis test**
  1. Planning the test and formulating hypothesis
  2. Gather data
  3. Analyze the data
  4. Decide whether you reject or fail to reject your hypothesis based on analyzing the data
  5. Interpret the decision in the context of the problem

• The above steps are the general framework we are going to learn about

• Let’s discuss step 1 first
Null and Alternative Hypothesis

• The actual tests begins by considering two hypotheses
  ▶ The null hypothesis
  ▶ The alternative hypothesis

• The null and alternative hypothesis contain opposing viewpoints

• Null Hypothesis: A statement about the value of a population parameter that is assumed to be true for the purpose of testing

• Alternative Hypothesis: A statement about the value of a population parameter that is assumed to be true if the Null Hypothesis is rejected during testing

• The alternative hypothesis is the OPPOSITE of the null hypothesis

• Notation:
  ▶ $H_0$: The null hypothesis
  ▶ $H_a$: The alternative hypothesis
Null and Alternative Hypothesis

• Example: We want to test if college students take less than five years to graduate from college on average
  ▶ \( H_0: \mu \geq 5 \)
  ▶ \( H_a: \mu < 5 \)

• Example: We want to test whether the mean GPA of students at De Anza is different from 2.0
  ▶ \( H_0: \mu = 2.0 \)
  ▶ \( H_a: \mu \neq 2.0 \)

• After you have determined which hypothesis the sample supports, you make a decision

• There are two options for a decision
  1. Reject \( H_0 \) if the sample data evidence favors the alternative hypothesis
  2. Fail to reject \( H_0 \) if the sample data evidence is insufficient to reject the null hypothesis
Rules for Setting Up the Hypotheses

- The null hypothesis $H_0$ must contain equality of some type
- The alternative hypothesis $H_a$ will be the opposite of the null hypothesis

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>$H_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>equal ($=$)</td>
<td>not equal ($\neq$)</td>
</tr>
<tr>
<td>greater than or equal to ($\geq$)</td>
<td>less than ($&lt;$)</td>
</tr>
<tr>
<td>less than or equal to ($\leq$)</td>
<td>more than ($&gt;$)</td>
</tr>
</tbody>
</table>

- Let’s look at some examples in Handout 7
Type I and Type II Errors

- When you perform a test, there are four possible outcomes depending on the truth of $H_0$ and whether you decide to reject or not:

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ is True</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_0$ is False</td>
<td>Type II Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

- A Type I Error occurs when you reject the null hypothesis when it is in fact true.

- A Type II Error occurs when you fail to reject the null hypothesis when it is in fact false.

- Each error has its own probability of occurring.
Type I Errors

- Type I error occurs when you reject the null hypothesis when it is true.
- The probability that you make a type I error is equal to $\alpha$.
- $\alpha$ is a number between 0 and 1.
- $\alpha$ is called the significance level of the hypothesis test.
- $\alpha$ is pre-determined before the hypothesis test.
- The significance level $\alpha$ should be small; you want a low risk of incorrectly rejecting the null hypothesis if it is really true.
Type II Errors

- Type II error occurs when you fail to reject the null hypothesis when it is false.
- The probability that you make a type II error is equal to $\beta$.
- $\beta$ is a number between 0 and 1.
- $1 - \beta$ is called the power of the hypothesis test.
- The power of a hypothesis test should be large (close to 1); you want large probability of correctly rejecting a false null hypothesis.
- One way of increasing the power of the test is to increase the sample size.
Type I and Type II Errors

- You want $\alpha$ and $\beta$ to be as small as possible because they are probability of making errors.
- Which error is worse to make? Type I error or type II error?
- Example: Suppose the null hypothesis, $H_0$ is: The victim of an automobile accident is alive when he arrives at the emergency room of a hospital.
  - Type I Error: The emergency crew thinks that the victim is dead, when in fact they are alive.
  - Type II Error: The emergency crew does not know if the victim is alive, when in fact, the victim is dead.
- Which error is worse in this example?
- Can we think of some other examples?
Type I and II Errors

- Example: A Court Trial is a real-life example of a hypothesis test
- Null Hypothesis: Not Guilty: A person is assumed to be innocent
- Alternate Hypothesis: Guilty: A person must be proven guilty beyond a reasonable doubt

<table>
<thead>
<tr>
<th></th>
<th>Fail to Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jury Decides Not Guilty</td>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
</tbody>
</table>

- Which error is worse? Type I or Type II?
Type I and Type II Errors

- Example: A certain experimental drug claims to cure prostate cancer
  - The null hypothesis, $H_0$: The drug cures prostate cancer
  - The alternative hypothesis, $H_a$: The drug does not cure prostate cancer
    - Type I Error: A doctor decides an experimental drug doesn’t cure cancer when it actually does
    - Type II Error: A doctor decides an experimental drug cures cancer when it actually does not

- Which error is worse in this example? Type I or Type II?
Type I and Type II Errors

- So the Type I and Type II errors are controlled by $\alpha$ and $\beta$
- You get to specify the $\alpha$ level
- You can try to minimize the value of $\beta$ by collecting a large sample size
- So how do we perform a hypothesis test?
  1. Planning the test and formulating hypothesis ✓
  2. Gather data
  3. Analyze the data
  4. Decide whether you reject or fail to reject your hypothesis based on analyzing the data
  5. Interpret the decision in the context of the problem
So how do we do hypothesis testing?

Here are the steps in more detail:

1. State the problem
2. Formulate the null and alternative hypotheses
3. Choose the significance level \( \alpha \) (i.e., choose the probability of rejecting \( H_0 \) when \( H_0 \) is true)
4. Determine the appropriate test statistic
5. Compute the test statistic
6. Find critical value(s) at the appropriate level of significance \( \alpha \) (traditional method) or compute the p-value (p-value method)
7. Compare test statistics to critical value(s) or p-value to level of significance \( \alpha \)
8. Based on this comparison reject or fail to reject \( H_0 \)
9. State your conclusions

So how do we do all this?
Distributions Needed for Hypothesis Testing

- Recall the following three distributions
- If you are testing a single population mean and the standard deviation is **KNOWN** then
  \[ \bar{X} \sim N \left( \mu, \frac{\sigma}{\sqrt{n}} \right) \]
- If you are testing a single population mean and the standard deviation is **UNKNOWN** then
  \[ \bar{X} \sim t_{df} \]
- If you are testing a single population proportion then
  \[ \hat{p} \sim N \left( p, \sqrt{\frac{p(1-p)}{n}} \right) \]
Calculating the Test Statistic

- **Test Statistic**: A value computed from the sample data that is used in making the decision of rejecting $H_0$ (or not).
- When the hypothesis test is about the MEAN and the population standard is **KNOWN**, then the test statistic is
  \[ z = \frac{\bar{x} - \mu}{\sigma \sqrt{n}} \]
- When the hypothesis test is about the MEAN and the population standard is **UNKNOWN**, then the test statistic is
  \[ t = \frac{\bar{x} - \mu}{s \sqrt{n}} \]
- When the hypothesis test is about a PROPORTION, then the test statistic is
  \[ z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \]
Calculating the Test Statistic

• Let’s look at worksheet 7
Calculating the Critical Values

- The **critical region** is the set of all values of the test statistic that cause us to reject the null hypothesis.
- A critical value is any value that separates the critical region (where we reject $H_0$) from the values of the test statistic that do not lead to the rejection of $H_0$.
- Common choices for the significance level $\alpha$ are 0.10, 0.05 and 0.01.
- So what does a critical region look like?
Critical Region

- When $H_a$ contains a not equal to, i.e., $H_a : \mu \neq 0$

  Two Tail Test

  - So the red area is the critical region
If our test statistic is inside the red area, then we reject $H_0$
If our test statistic is inside the blue area, then we fail to reject $H_0$
• When $H_a$ contains a less than, i.e., $H_a : \mu < 0$

  \[ -z_{\alpha/2} \rightarrow 0 \leftarrow z_{\alpha/2} \]

  Left Tail Test

  - So the red area is the critical region
Critical Region

- When $H_a$ contains a greater than, i.e., $H_a : \mu > 0$

  Right Tail Test

![Critical Region Diagram]

- So the red area is the critical region
Critical Region

- So the critical region gives us a rule for when we should reject (or fail to reject the null hypothesis)
- If our test statistic is inside the critical region then we REJECT the null hypothesis
- If our test statistic is outside the critical region then we FAIL TO REJECT the null hypothesis
- This is equivalent to saying if the absolute value of our test statistic is bigger than the absolute value of the critical value then we REJECT the null hypothesis, i.e.,

\[ |z| > \left| z_{\alpha/2} \right| \]

if not, then we FAIL TO REJECT the null hypothesis
Hypothesis Testing Examples

- Let's take a look at worksheet 7
Null and Alternative Hypothesis
type I and type II errors

Hypothesis Testing

Rare Events

- Suppose you make an assumption about a property of the population (this assumption is the null hypothesis).
- Then you gather data from a random sample.
- If the sample has properties that would be very unlikely to occur if the assumption is true, then you would conclude that your assumption about the population is probably incorrect.
- So we need some sort of way to quantify how unlikely our assumption is if the null hypothesis is true.
- We will use the sample data to calculate the actual probability of getting the test result, called the *p-value*. 


P-values

- **p-value**: The p-value is the probability that, if the null hypothesis is true, the results from another randomly selected sample will be as extreme or more extreme as the results obtained from the given sample.

- So the p-value, in some sense, measures how likely our assumption about some property of the population is true.

- A p-value is always a number between 0 and 1.

- A large p-value indicates that we should **FAIL TO REJECT** the null hypothesis.

- A small p-value indicates that we should **REJECT** the null hypothesis.

- How do we calculate the p-value?
Calculating the P-value

- When $H_a$ contains a not equal to, i.e., $H_a : \mu \neq 0$

  Two Tail Test

- So the red area is the p-value
Calculating the P-value

- When $H_a$ contains a less than, i.e., $H_a : \mu < 0$
  
  ![Diagram showing left tail test](image)

  - So the red area is the p-value
Calculating the P-value

- When $H_a$ contains a less than, i.e., $H_a : \mu < 0$
  
  Right Tail Test

- So the red area is the p-value
P-values

- So what is a large (or small) p-value?
- Any p-value that is smaller than (or equal to) our $\alpha$ level is considered small
- Thus, if our p-value $\leq \alpha$ then we REJECT the null hypothesis
- Also, if our p-value $\geq \alpha$ then we FAIL TO REJECT the null hypothesis
- Examples: Let’s look at handout 7
Rules to Remember

- If our test statistic is inside the critical region then we REJECT the null hypothesis.
- If our test statistic is outside the critical region then we FAIL TO REJECT the null hypothesis.

**ALTERNATIVELY**

- If our p-value $\leq \alpha$ then we REJECT the null hypothesis.
- If our p-value $\geq \alpha$ then we FAIL TO REJECT the null hypothesis.

**BOTH WAYS LEAD TO THE SAME SOLUTION!**

- We never ACCEPT $H_0$, we always FAIL TO REJECT $H_0$. 