Activity 9.1a&b - Test of Hypothesis Basics and Errors

There are 2 types of statistical inference

- **Confidence Interval-CI (Chapter 8):**
  - estimates all the plausible values for a population parameter. It gives us more information. Our course covers “p” and “μ.”

- **Significance Tests-TOH (Chapters 9-12):**
  - are formal procedure for comparing observed data with a claim (hypothesis) whose truth we want to assess.
  - We express the results of a significance test in terms of a probability (p-value) that measures how well the data and the claim agree.

### 9.1 CONCEPTS YOU MUST KNOW

**DEFINE HYPOTHESES:**

- State the parameter of interest
- Null Hypothesis
- Alternative Hypothesis

**STATISTICAL INFERENCE:**

- Claim
- 2 Outcomes of a Statistical Test
- P-Value
- Significance Level
- Statistically Significant

**Error:**

- Type I
- Type II
- Power (covered next class)
Example: The Basketball Player

Setting up Significance Tests

Example:
Ben claims that he makes 80% of his free-throws.

1. What is the population parameter we want to test?
   \[ P = \text{true proportion of free-throws Ben made.} \]

2. What is our first claim that we are seeking to gather evidence against? This is the null hypothesis. Express in symbols and words:
   \[ H_0: p = .80 \ (\text{Ben's true foul shooting is 80%}) \]

3. What is our second claim that we suspect to be true instead of the null hypothesis. There are 3 possible scenarios to consider for the alternative hypothesis.

   State the alternate hypothesis and sketch the graph:

   - **Alternate Scenario #1:** We think Ben is exaggerating and can't possibly shoot that well.
     \[ H_\text{a}: p < .80 \ (\text{Left Tail Test}) \]

(next page)
**Example:** The Basketball Player

**Setting up Significance Tests (cont.)**

State the alternate hypothesis and sketch the graph:

- **Alternate Scenario #2:** We think Ben is being modest, and is the best free-throw shooter in the state.

  \[ H_A : \ p > 0.80 \ \text{(Right Tail Test)} \]

- **Alternate Scenario #3:** We simply think Ben is lying.

  \[ H_A : \ p \neq 0.80 \ \text{(2 Tail Test)} \]
Example: The Basketball Player

The Reasoning of Significance Tests

**Example:** Now we gather evidence. We do not think Ben makes 80% of his free throws. So, we have him attempt 50 free-throws. He makes 32 of them. His sample proportion of made shots is $32/50 = 0.64$. What can we conclude about the claim based on this sample data?

Option 1: What hypothesis do we want to test if we think Ben is exaggerating?

- $H_0: p = 0.8$
- $H_a: p < 0.8$

Do we have enough evidence to reject our null hypothesis?

We typically set the significance level to $\alpha = 0.05$.

**Conclusion**

Since the $p$-value (.002) is very small (extreme value) and less than our predetermined significance level ($\alpha = 0.05$), we have convincing evidence to reject $H_0$ and have sufficient evidence Ben shoots less than 80%.
**Example:** The Basketball Player

*Would a Confidence Interval Provide the Evidence?*

**Option 2:** Let's try a different hypothesis. We don't believe Ben is an 80% free-throw shooter.

1. **What hypothesis do we want to test?**
   
   \[ H_0 : p = .80 \]
   \[ : p \neq .80 \]

2. **What evidence do we have (assume conditions of random, independent and normal are met)?** [Create a 95% CI.]

   ![Graph of normal distribution with critical values]

   \[ \hat{p} = .64 \quad n = 50 \quad z^* = 1.96 \]
   \[ \hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \]
   \[ .64 \pm 1.96 \cdot \sqrt{\frac{(.64)(.36)}{50}} \]
   \[ .64 \pm 1.96 (.068) \]
   \[ .64 \pm .13 \]
   \[ LB = .64 - .13 = .51 \]
   \[ UB = .64 + .13 = .77 \]

3. **Do we have enough evidence to reject our null hypothesis?**

   **OUR CI:** WE ARE 95% CONFIDENT THAT THE TRUE FOUL SHOOTING PROPORTION IS BETWEEN 51% AND 77%.

   **DO WE HAVE CONVINCING EVIDENCE?**

   **YES** SINCE OUR CI DOES NOT INCLUDE .80, WE HAVE CONVINCING EVIDENCE THAT BEN IS NOT AN 80% FREE THROW SHOOTER.

   (WE WOULD HAVE REJECTED \( H_0 \))
**SUMMARY:**

**Stating Hypotheses** In any significance test

1) The **null hypothesis** has the form

\[ H_0 : \text{PARAMETER} = \text{VALUE} \]

2) The **alternative hypothesis** has one of 3 forms
   - To determine the correct form of \( H_a \), read the problem carefully!!
   - Determine the symbol (=, >, <, ≠)
   - Label the P-value.

<table>
<thead>
<tr>
<th>Left-sided Test -</th>
<th>Right-sided Test -</th>
<th>Two-sided Test -</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 : \text{parameter} \leq \text{value} )</td>
<td>( H_0 : \text{parameter} \geq \text{value} )</td>
<td>( H_0 : \text{parameter} \neq \text{value} )</td>
</tr>
</tbody>
</table>

* P-value combines both areas
Stating Hypotheses – Try this practice problem

Studying Job Satisfaction - see Activity 9.1A (answer key)

EXAMPLE: Does the job satisfaction of assembly-line workers differ when their work is machine-paced rather than self-paced? One study chose 18 subjects at random from a company with over 200 workers who assembled electronic devices. Half of the workers were assigned at random to each of two groups. Both groups did similar assembly work, but one group was allowed to pace themselves while the other group used an assembly line that moved at a fixed pace. After two weeks, all the workers took a test of job satisfaction. Then they switched work setups and took the test again after two more weeks. The response variable is the difference in satisfaction scores, self-paced minus machine-paced.

a) Describe the parameter of interest in this setting.

\[ \mu = \text{DIFFERENCE IN SATISFACTION SCORES} \]

b) State appropriate hypotheses for performing a significance test. (in symbols and words)

\[ H_0: \mu = 0 \] (No difference in job satisfaction scores)

\[ H_a: \mu \neq 0 \] (there is a difference in job satisfaction scores)

Note: Workers could either be more or less satisfied
Significance Tests: The Basics

✓ A significance test assesses the evidence provided by data against a null hypothesis $H_0$ in favor of an alternative hypothesis $H_a$.

✓ We use the $P$-value of a test and our predetermined $\alpha$ (alpha - the significance level), to make decisions regarding our hypothesis.

✓ The $P$-value of a test is the probability, computed supposing $H_0$ to be true, that the statistic will take a value at least as extreme as that actually observed in the direction specified by the alternate hypothesis $H_a$.

✓ Small $P$-values indicate strong evidence against $H_0$. To calculate a $P$-value, we must know the sampling distribution of the test statistic when $H_0$ is true. There is no universal rule for how small a $P$-value in a significance test provides convincing evidence against the null hypothesis.

✓ If the $P$-value is smaller than a specified value $\alpha$ (called the significance level), the data are statistically significant at level $\alpha$. In that case, we can reject $H_0$. If the $P$-value is greater than or equal to $\alpha$, we fail to reject $H_0$.

✓ General Rule:

✓ Small $P$-values we $\underline{\text{REJECT}}$ the null hypothesis.

✓ Large $P$-values we $\underline{\text{FAIL TO REJECT}}$ the null hypothesis.

✓ We NEVER accept the null hypothesis.
**Statistically Significance at level α – Try this practice problem**

**Better Batteries – see Activity 9.1A (answer key)**

A company has developed a new deluxe AAA battery that is supposed to last longer than its regular AAA battery. However, these new batteries are more expensive to produce, so the company would like to be convinced that they really do last longer. Based on years of experience, the company knows that its regular AAA batteries last for 30 hours of continuous use, on average. The company selects an SRS of 15 new batteries and uses them continuously until they are completely drained. A significance test is performed using the hypotheses

\[ H_0 : \mu = 30 \text{ hours} \]
\[ H_a : \mu > 30 \text{ hours} \]

where \( \mu \) is the true mean lifetime of the new deluxe AAA batteries. The resulting **P-value is 0.0276.**

**For each, provide a supporting graph labeling the mean, P-value, and alpha(\( \alpha \)).**

**a) What conclusion can you make for the significance level \( \alpha = 0.05 \)?**

Since the p-value (0.0276) is less than our significance level (\( \alpha = 0.05 \)), we reject \( H_0 \) and have sufficient evidence to conclude the AAA batteries last longer than 30 hrs.

**b) What conclusion can you make for the significance level \( \alpha = 0.01 \)?**

Since the p-value (0.0276) is greater than our significance level (\( \alpha = 0.01 \)), we fail to reject \( H_0 \) and do not have enough evidence to conclude the AAA batteries last longer than 30 hrs.
**Introduction to Type I and Type II Errors:**

**EXAMPLE: Our Court System "O.J. Analogy"**

**Understand Hypothesis Testing**

- In our jury system, you are innocent until proven guilty. This is how we are going to set up our statistical test of hypothesis statements:
  
  \[ H_0 : p = \text{O.J. not guilty (innocent)} \quad \leftarrow \text{Null hypothesis. } H_0 \rightarrow "H not" \]
  
  \[ H_a : p \neq \text{O.J. guilty} \quad \leftarrow \text{Alternate hypothesis} \]

- The lawyers give evidence to prove their case (we will do the same by taking a sample).

- The jury comes back with the verdict based on whether this was a criminal or civil trial.
  
  - Criminal Trial evidence must be convincing “Beyond a reasonable doubt.”
  
  - Civil Trial evidence must be convincing "By a preponderance of the evidence"

- Which has a lower threshold? This threshold is comparable to our significance level(\(\alpha\)). We predetermine \(\alpha\) based on how much of an error we are willing to make. Typically, \(\alpha=.05\) or \(\alpha=.01\).

- The Jury decides:
  
  1. "GUILTY," if they have enough evidence. **We will do the same… If we have enough evidence, we “REJECT H_0.”**
  2. Or the jury says "NOT GUILTY." **We will do the same… If we do not have enough evidence, we “Fail to reject H_0.”**
  3. The Jury never says "INNOCENT," because OJ will never tell us the truth. We **NEVER accept** the null hypothesis because we have a chance of making a mistake.

- **Discussion Questions:**
  
  a) OJ was found “not guilty” in the criminal trial. He was found “guilty” in the civil trial. Why?
  
  b) What are 2 possible errors that could happen in our jury system?

  - **TYPE 1:** JURY FINDS OJ GUILTY, BUT HE IS INNOCENT.
  - **TYPE 2:** JURY FINDS OJ NOT GUILTY WHEN HE IS GUILTY
Type I and Type II Errors

- When we draw a conclusion from a significance test, we hope our conclusion will be correct. But sometimes it will be wrong.

- There are two types of mistakes we can make and it is very important how recognize and interpret these errors!!!

  - If we Reject $H_0$ when $H_0$ is True, we have committed a Type I error.
  - If we fail to reject $H_0$ when $H_0$ is False (OR $H_a$ is true), we have committed a Type II error.

- Fill in table:

<table>
<thead>
<tr>
<th>Conclusion based on sample</th>
<th>$H_0$ true</th>
<th>$H_0$ false ($H_a$ true)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td><strong>Type I</strong> error (α)</td>
<td>Correct conclusion</td>
</tr>
<tr>
<td>Fail to reject $H_0$</td>
<td>Correct conclusion</td>
<td><strong>Type II</strong> error (β)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simply</th>
<th>Type I: Reject $H_0$, when $H_0$ is True</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type II: Fail to Reject $H_0$, when $H_a$ is True</td>
<td></td>
</tr>
</tbody>
</table>

Truth about the population
Example "Perfect Potatoes"

A potato chip producer and its main supplier agree that each shipment of potatoes must meet certain quality standards. If the producer determines that more than 8% of the potatoes in the shipment have "blemishes," the truck will be sent away to get another load of potatoes from the supplier. Otherwise, the entire truckload will be used to make potato chips. To make the decision, a supervisor will inspect a random sample of potatoes from the shipment. The producer will then perform a significance test using the hypotheses

\[ H_0: p = 0.08 \]
\[ H_a: p > 0.08 \]

where \( p \) is the actual proportion of potatoes with blemishes in a given truckload.

**Describe Type I & Type II error in this setting; explain consequences of each:**

- **Type I error** would occur if...

\[ \text{REJECT } H_0 \text{ (concluding proportion of bad potatoes is greater than 0.08), when the actual proportion of bad potatoes is 0.08 (H}_0). \]

**Consequence:**

Trucks of potatoes are sent away with good potatoes, which results in the company losing money.

- **Type II error** would occur if...

\[ \text{Fail to reject } H_0 \text{ (concluding the proportion of bad potatoes is acceptable at the 0.08 cutoff), when the proportion of bad potatoes is actually greater than 0.08 (H}_a). \]

**Consequence:**

The company will make chips with bad potatoes and upset their customers.
Type I and II Errors—Try this practice problem

Faster fast food?" - see Activity 9.1B (answer key)

Example “Faster fast food?” The manager of a fast-food restaurant wants to reduce the proportion of drive-through customers who have to wait more than 2 minutes to receive their food once their order is placed. Based on store records, the proportion of customers who had to wait at least 2 minutes was \( p = 0.63 \). To reduce this proportion, the manager assigns an additional employee to assist with drive-through orders. During the next month, the manager will collect a random sample of drive-through times and test the following hypotheses:

\[
H_0 : p = 0.063 \\
H_a : p < 0.63
\]

where \( p \) = the true proportion of drive-through customers who have to wait more than 2 minutes after their order is placed to receive their food.

Describe Type I & Type II error in this setting; explain consequences of each:

- **A Type I error** would occur if...
  - **Consequence:**
    - REJECT \( H_0 \) (Concluding the proportion of customers wait time is less than \( 0.63 \)) WHEN THE ACTUAL PROPORTION IS \( 0.63 \) (\( H_0 \) did not improve when actually did)
    - CONSEQUENCE: THE MANAGER IS PAYING FOR AN ADDITIONAL EMPLOYEE THAT HE DOES NOT NEED

- **A Type II error** would occur if...
  - **Consequence:**
    - FAIL TO REJECT \( H_0 \), WHEN \( H_0 \) IS TRUE
      - **Consequence:**
        - FAIL TO REJECT \( H_0 \) (Concluding the proportion of customer 2min wait time IS \( 0.63 \)) WHEN THE ACTUAL PROPORTION OF 2MIN. WAIT TIME IS LESS THAN \( 0.63 \) (improved when it did not)
        - CONSEQUENCE: THE MANAGER WOULD FIRE THE ADDITIONAL DRIVE THROUGH WORKER, CUSTOMERS WOULD BE UPSET WITH THE POOR SERVICE
Quote from AP Statistics Teacher Forum

"Do not try to teach any calculations about Type II error or power. Not only is that not required, it can be confusing and it distracts students from understanding the concepts. They need to know what the two types of error are and what power is. They need to be able to explain them in the context of the questions. And they need to understand the interactions among the errors, power, sample size and effect size. But no calculations!"