LESSON 35: HYPOTHESIS TESTS FOR \( \mu \)

Is the Population Mean Here or There?

PART A: ASSUMPTIONS

In this Lesson, we will perform hypothesis tests for a population mean \( \mu \) under these assumptions:

- The population standard deviation (SD) \( \sigma \) is unknown.
- The Central Limit Theorem (CLT) applies.

That is, one or both of these is true:

\[ X \sim \text{Normal} \]

(In fact, the distribution only has to be roughly normal if the sample size is not too small.)

\[ \text{OR} \]

§ The sample size is large \( n > 30 \).

Actually, there is more theory beyond the CLT involved in this Lesson.

PART B: THE \( z \) TEST STATISTIC \( \text{(IF } \sigma \text{ IS KNOWN) } \)

If the population standard deviation (SD) \( \sigma \) were presumed known, then we would use a \( z \) test statistic. In fact, we use the Formula for \( z \) Scores for Sample Means from Lesson 22 on the Central Limit Theorem (CLT).

\[
\begin{align*}
  z &= \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \\
  &= \frac{\bar{x} - \mu}{\sigma / \sigma_n}, \text{ or } \frac{\bar{x} - \mu}{\bar{x} / \sqrt{n}}
\end{align*}
\]

where \( \mu \) is obtained under \( H_0 \).

Instead, we will focus on the more practical case where \( \sigma \) is presumed unknown.

PART C: THE \( t \) TEST STATISTIC \( \text{(IF } \sigma \text{ IS UNKNOWN) } \)

If \( \sigma \) is presumed unknown and we replace \( \sigma \) with \( s \), then we must use a \( t \) test statistic.

\[
\begin{align*}
t &= \frac{\bar{x} - \mu}{s / \sqrt{n}} \\
n &= \frac{\bar{x} - \mu}{\bar{x} / \sqrt{n}}
\end{align*}
\]

where \( \mu \) is obtained under \( H_0 \).

We use the \( t \) distribution on \( n-1 \) degrees of freedom (df).
PART D: EXAMPLE

Example 1 (Hypothesis Test for a Population Mean; Population SD Unknown)

A large class takes an exam. Assume that the scores are approximately normally distributed. We randomly sample seven exams. The sample mean is 67.5 points and the sample standard deviation (SD) is 3.2 points. Test the claim that the average exam score for the class is less than 70.0 points at the 0.05 significance level.

• Rounding instructions vary, but here, round off to five significant digits, except round off the t test statistic to three decimal places.

Use these hints about the t distribution on 6 degrees of freedom (df):

\[ t_{0.05} = 2.067 \]

§ Solution

Let \( \mu \) be the population mean exam score for the class.

The methods of this Lesson apply because:
- We are conducting a hypothesis test for a population mean \( \mu \).
- We are assuming that the population SD \( \sigma \) is unknown.
- We are assuming that the exam scores approximately normally distributed, so the CLT applies.

Here is the setup for this left-tailed test:

- \( H_0 \): \( \mu = 70.0 \) points
- \( H_a \): \( \mu < 70.0 \) points (Claim)
- \( \alpha = 0.05 \)

We now identify the sample statistics.

The sample size \( n = 7 \) exams.

We want to use the t distribution on 6 df, since:
\[ n - 1 = 7 - 1 = 6 \]

The sample mean \( \bar{x} = 67.5 \) points.

The sample SD \( s = 3.2 \) points.

We now compute the t test statistic for our sample.

\[ t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{67.5 - 70.0}{3.2 / \sqrt{7}} = -2.067 \]

We now find the corresponding P-value.

- This is a left-tailed test, so we need a left-tailed P-value.
- Use the hints.
- Left-tailed P-value \( \approx 0.0421 \)

We now decide whether or not to reject \( H_0 \).

0.0421 < 0.05, so P-value < \( \alpha \). The P-value is “low.”

We reject \( H_0 \).

Finally, we write our conclusion relative to the claim.

- Part 1: We reject \( H_0 \). That is a “strong result.”
  Start with: There is sufficient evidence …
- Part 2: \( H_a \) is the claim. Can we “prove” \( H_a \)?
  Continue with: … for the claim that … (state the claim)

CONCLUSION: There is sufficient evidence for the claim that the average exam score for the class is less than 70.0 points.
LESSON 36: TRADITIONAL (CLASSICAL) METHOD FOR HYPOTHESIS TESTING

How Can We Use Critical Values (CVs) to Do Hypothesis Tests?

PART A: TRADITIONAL (CLASSICAL) METHOD

Before computers were in common use, people had to use statistical tables and the traditional (classical) method of hypothesis testing. We will use the method for some hypothesis tests, particularly two-tailed tests in Lesson 37, where we lose symmetry.

Example 1 (Magician’s Coin: Traditional (Classical) Method; Two-tailed Test; Revisiting Lesson 33, Example 1)

As in Lesson 33, Example 1, we want to test the claim that a magician’s coin is fair at the 0.05 significance level.

We could let \( p \) = the probability that the coin comes up heads.

Here is the setup for this two-tailed test:

\[ H_0: p = \frac{1}{2}, \text{ or } 0.5 \text{ (Claim)} \]

\[ H_1: p \neq \frac{1}{2}, \text{ or } 0.5 \]

\( \alpha = 0.05 \)

We now gather sample data. The coin is flipped 100 times, and it comes up heads 51 times.

- The sample proportion \( \hat{p} = \frac{x}{n} = \frac{51}{100} = 0.5100 \)

- Verify that normal approximations are appropriate in this problem.

Under \( H_0 \), \( np = (100) \left( \frac{1}{2} \right) = 50 \geq 5 \)

Under \( H_0 \), \( nq = (100) \left( \frac{1}{2} \right) = 50 \geq 5 \)

We now compute the \( z \) test statistic for our sample.

\[
z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.5100 - 0.5}{\sqrt{\frac{0.5(0.5)}{100}}} = 0.20
\]

- Note that a \( z \) score of 0.20 seems very close to 0. Getting 51 heads out of 100 flips seems to present little evidence against \( H_0 \).

Instead of using the \( P \)-value method, we will use the traditional (classical) method to make a decision about \( H_0 \). Both methods will yield the same decisions and conclusions.

- This is a two-tailed test, so we need two critical values (CVs), \( \pm z_{\alpha/2} \). These were the CVs that helped us with \( 1 - \alpha \) confidence intervals (CIs) in Lessons 27 and 28.

- The \( z \) values that are more extreme than the CVs make up the critical region. They correspond to the left and right tails of the \( z \) number line; see the shaded regions in the figures below. These are the \( z \) values that would be too unusual under \( H_0 \) for \( H_0 \) to survive. Think of the critical region as the “danger zone” where \( H_0 \) dies.
Decision Rule for a Hypothesis Test: Traditional (Classical) Method

- If the value of the test statistic lies in the critical region, then reject $H_0$.
- If the value of the test statistic does not lie in the critical region, then do not reject $H_0$.

Test statistic $z = 0.20$, which does not lie in the critical region, so we do not reject $H_0$.

Finally, we write our conclusion relative to the claim.

- Part 1: We do not reject $H_0$. That is a “weak result.”
  
  **Start with:** There is insufficient evidence …

- Part 2: $H_a$ is the claim. $H_a$ is “on trial.”
  
  **Continue with:** … against the claim that … (state the claim)

  **CONCLUSION:** There is insufficient evidence against the claim that the magician’s coin is fair.
(Lesson 37: Hypothesis Tests for $\sigma$ or $\sigma^2$) 37.02

PART C: EXAMPLE

**Example 1 (Hypothesis Test for a Population SD: Traditional (Classical) Method)**

A women’s college goes coed. We assume that women’s heights there have a standard deviation (SD) of 2.5 inches. Assume that the men’s heights are approximately normally distributed. We randomly sample 20 men. The sample standard deviation (SD) is 3.4 inches. Test the claim that men’s heights and women’s heights at the college have different standard deviations (SDs) at the 0.05 significance level.

- **Rounding** instructions vary, but here, round off to five significant digits, except round off the $\chi^2$ test statistic to three decimal places.

Use the **traditional (classical) method** of hypothesis testing; the loss of symmetry in the $\chi^2$ distributions works against the $P$-value method for two-tailed tests here.

Use these hints about the $\chi^2$ distribution on 19 degrees of freedom (df):

![Chi-Square Distribution](image)

**Solution**

Let $\sigma$ be the population SD for men’s heights at the college.

The methods of this Lesson apply because:

- We are assuming that the men’s heights at the college are approximately normally distributed. $X \sim \text{Normal}$

Here is the setup for this **two-tailed test**:

- $H_0: \sigma = 2.5$ inches
- $H_1: \sigma \neq 2.5$ inches (Claim)
- $\alpha = 0.05$

We now identify the **sample statistics**.

The **sample size** $n = 20$ men’s heights.

We want to use the $\chi^2$ distribution on 19 df, since:

$\chi^2 = \frac{(n - 1)s^2}{\sigma^2} = \frac{(20 - 1)(3.4)^2}{(2.5)^2} = 35.142$

We use the **traditional (classical) method** to make a decision about $H_0$:

- This is a **two-tailed** test, so we need **two critical values (CVs)**.
- $\chi^2_\text{R}$ is the right (greater) CV, and $\chi^2_\text{L}$ is the left (lesser) CV.

- The $\chi^2$ values that are **more extreme than the CVs** make up the **critical region**. They correspond to the left and right tails of the $\chi^2$ number line; see the **shaded regions** in the hints.

(Lesson 37: Hypothesis Tests for $\sigma$ or $\sigma^2$) 37.03
• Test statistic $\chi^2 = 35.142$, which lies in the critical region, so we reject $H_0$.

Finally, we write our conclusion relative to the claim.

• Part 1: We reject $H_0$. That is a “strong result.”
  Start with: There is sufficient evidence …

• Part 2: $H_1$ is the claim. Can we “prove” $H_1$?
  Continue with: … for the claim that … (state the claim)

CONCLUSION: There is sufficient evidence for the claim that men’s heights and women’s heights at the college have different standard deviations (SDs).

---

**LESSON 38: TYPE I and TYPE II ERRORS**

What Kinds of Mistakes Do We Make?

**PART A: JURY TRIAL ANALOGY**

Type I and Type II errors may be described by the table below in the context of a jury trial.

<table>
<thead>
<tr>
<th>Verdict is “not guilty”</th>
<th>Defendant is innocent</th>
<th>Defendant is guilty</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Correct verdict)</td>
<td>Type II error</td>
<td>(Correct verdict)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Verdict is “guilty”</th>
<th>Defendant is innocent</th>
<th>Defendant is guilty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error</td>
<td>(Correct decision)</td>
<td>(Correct decision)</td>
</tr>
</tbody>
</table>

**PART B: $H_0$ AS THE DEFENDANT**

We may think of $H_0$ as the defendant.

We commit a **Type I error** when we reject $H_0$ even though $H_0$ is true.

Think: Convicting the innocent.

• In fact, $\alpha = P(\text{we will reject } H_0 \mid H_0 \text{ is true})$. See Part C.

We commit a **Type II error** when we do not reject $H_0$ even though $H_0$ is false.

Think: Letting the guilty go free.
PART C: INTERPRETING $\alpha$

In Part B, we noted that $\alpha = P(\text{we will reject } H_0 \mid H_0 \text{ is true})$.

**Example 1 (Magician’s Coin: Interpreting the Significance Level; Revisiting Lesson 36, Example 1)**

In Lesson 36, Example 1, we tested the claim that a magician’s coin is fair at the 0.05 significance level.

We could let $p =$ the probability that the coin comes up heads.

Here is the setup for this two-tailed test:

$\ H_0 \colon p = \frac{1}{2}, \text{ or } 0.5 \ (\text{Claim})$

$\ H_1 \colon p \neq \frac{1}{2}, \text{ or } 0.5$

$\ \alpha = 0.05$

Let’s assume that $H_0$ is true. We could still get unusual sample results by chance … to the point where we could end up rejecting $H_0$, thus committing a Type I error. For example, a fair coin could conceivably come up heads 100 times in 100 flips. There is as 5% (or 0.05) chance that our test statistic will be in the critical region corresponding to the shaded regions below: