Topic 7 - Hypothesis tests based on a single sample

• Sampling distribution of the sample mean

• Basics of hypothesis testing

• Hypothesis test for a population mean

• Hypothesis test for a population proportion
Hypothesis testing

• We have already discussed how you can make decisions based on data using confidence intervals.

• If someone claims a population parameter has a certain value, we only believe the claim if the value is inside our confidence interval for the parameter.

• Hypothesis testing provides a more formal method for testing claims.
Hypothesis testing

• A **hypothesis test** checks sample data against a claim or assumption about the population.

• The claim being tested is called the **null hypothesis**, $H_0$.

• The null hypothesis typically represents the status quo or no change belief.

• The **alternative hypothesis**, $H_A$, represents what we suspect is true.

• *The researcher’s goal is typically to show that the alternative hypothesis is true.*
Studying example

• A study reports that the mean time freshmen spend studying is 7.06 hours per week. A TAMU teacher feels that freshmen here spend more time on average studying.

• What are the appropriate null and alternative hypotheses?
What does it look like?
ImanAggie Redimix

• Remember the problem with the strength of the concrete and the TxDot contract? The company claims, based on information from “Joe”, their Production Foreman, that the average is 3,200 psi, with a standard deviation of 250?

  – Someone else at the company questions that that amount is overstated, because if it’s lower than the claim, it opens the company to potential lawsuits.

  – What are the null and alternate hypotheses in this case?
What does it look like?
Is the operation “status quo”?  

- Our California-based company, produces rivets for use in the aircraft industry. Historically, these rivets have a mean diameter of .025 inches, with a standard deviation of .000001 inches (pretty tight tolerance).

- A recent earthquake might have impacted the average width setting on our machines.

- If they are now too wide, they won’t fit in the holes of the airplanes. Conversely, if they’re too narrow, they “could” cause a safety issue.

- What are the null and alternate hypotheses in this case?
What does it look like?
TV Sitcom example

• The percentage of viewers that like sitcoms on average is about 35% (that’s a madeup number, but assume that’s the base-line that the execs allow to keep a show on the air).

• Now assume that you’ve developed a new sitcom that you claim is “vastly better” and which will have a larger viewing audience than the typical sitcom. Why is that good?

• You want to prove this up to the execs through focus groups.

• What are the null and alternate hypotheses in this case?
Decision making

- We look for evidence in the form of sample data against the null hypothesis and in favor of the alternative hypothesis.

- We use a **test statistic**, computed from the data, to make our decision.

- We evaluate the evidence from the sample by identifying it’s **p-value**, the probability of a more extreme test statistic than the one observed if the null is true. For example, if our Z test value was 2.0, the area to the right of that Z (the p-value) would be approximately 2.5%.

- A small **p-value** most likely indicates we should reject the null hypothesis.

- A large **p-value** most likely means there is not strong evidence against the null.

- At the end of the test, we either reject or fail to reject the null hypothesis based on the **p-value**. *Never Accept the null.*
Hypothesis testing process…..

I would suggest drawing a curve and labeling it from the word problem.

1. Establish your null and alternate hypothesis and mark the “break point” on the curve, based on the “alpha” given to you by the person that asked for the test. Alpha is the area associated with values more extreme than the critical value of Z or T, the “cutoff” value.

2. Take a sample and calculate the test statistics.

3. If the test statistic falls outside the break point on the curve, the Z or t-critical for example, or if you’re resulting p-value is less than alpha, reject the null hypothesis.

4. If not, you do not accept the null. Rather you would state that you don’t have sufficient evidence to disprove the null.
Studying example

• A random sample of 35 freshmen at TAMU report the hours they spend studying, and the sample mean is 8.43 hours with a sample standard deviation of 4.32 hours. *Remembering of course that the prior study said that the average was 7.06 hours.*

• What test statistic should we use?

• What is the approximate $p$-value of our test statistic?
$Z_{test} = \frac{8.43 - 7.06}{\frac{4.32}{\sqrt{35}}} = \frac{1.37}{0.7302} = 1.876$
Level of Significance

• The level of significance, $\alpha$, determines the amount of evidence we require in order to reject the null.

• The value of $\alpha$ specifies the probability of rejecting the null when it is true (type 1 error)

• The value of $\alpha$ is typically less than 0.1.

• If $p$-value $\leq \alpha$, then we reject $H_0$.

• If $p$-value $> \alpha$, then we fail to reject $H_0$, but you don’t accept the null.

• Smaller values of $\alpha$ make it more difficult to reject the null. $\alpha$ may reflect the researcher’s belief in the null hypothesis........maybe they don’t want it disproved.
What are conclusions at various alpha?

$$Z_{test} = \frac{8.43 - 7.06}{\left(\frac{4.32}{\sqrt{35}}\right)} = 1.876$$
Decision rules

- The level of significance can be used to develop decision rules based on the value of your test statistic.

- For example, to test $H_0: \mu = \mu_0$

<table>
<thead>
<tr>
<th>Test statistic for large $n$</th>
<th>$Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_A$</td>
<td>Reject $H_0$ if</td>
</tr>
<tr>
<td>$\mu &lt; \mu_0$</td>
<td>$Z &lt; -z_\alpha$</td>
</tr>
<tr>
<td>$\mu &gt; \mu_0$</td>
<td>$Z &gt; z_\alpha$</td>
</tr>
<tr>
<td>$\mu \neq \mu_0$</td>
<td>$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test statistic for small $n$</th>
<th>$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_A$</td>
<td>Reject $H_0$ if</td>
</tr>
<tr>
<td>$\mu &lt; \mu_0$</td>
<td>$T &lt; -t_{\alpha,n-1}$</td>
</tr>
<tr>
<td>$\mu &gt; \mu_0$</td>
<td>$T &gt; t_{\alpha,n-1}$</td>
</tr>
<tr>
<td>$\mu \neq \mu_0$</td>
<td>$</td>
</tr>
</tbody>
</table>
Type 2 error

- *In addition to rejecting the null when it is true, we can also fail to reject the null when it is false (type 2 error).*

- Suppose our decision rule for the studying example is to reject \( H_0 \) if \( \bar{x} \geq 8.50 \)

- Using \( \sigma = 4.32 \)
  - \( P(\text{type 1 error}) = \alpha = P(Z > \frac{8.5 - 7.06}{4.32/\sqrt{35}}) = P(Z > 1.96) \Rightarrow 0.025 \)

- If \( \mu = 7.3 \), \( P(\text{type 2 error}) = \)
  - \( P(Z < \frac{8.5 - 7.3}{4.32/\sqrt{35}}) \approx P(Z < 1.645) \approx 0.95 \)
What does it look like?
I’m an Aggie Redimix example

Remember that their production was supposedly normally distributed with a mean of 3200 and a standard deviation of 250. If we are testing BELOW that with a sample of size 36, we get the following criteria for decisions;

One tail left. Alpha was 2.5% and the corresponding Z value was 1.96, which makes the “decision point” value of our sample 1.96 standard ERRORS to the left of the mean, or

$$3200 - 1.96 \left( \frac{250}{\sqrt{36}} \right) = \bar{x}_{CRIT} = 3118.33$$
ImanAggie continued

Based on our test, it’s likely that the true mean of the process is less than 3,200 psi.

What’s the probability that we would have “accepted the null” that the mean = 3,200, if the true mean had been 3,199 psi or 3,150 psi or even 3,100 psi?

At 3,199, the Type 2 is about 97.5% (1-alpha)

\[
P(Z > \frac{3118.33 - 3199}{\sqrt{36}}) \sim 97.5%
\]

At 3,150, the Type 2 is about 78% or \(P(Z > -0.76)\)
At 3,100, the Type 2 is about 33% or \(P(Z > +0.44)\)
VHS example

• A manufacturer of VHS tapes wants to make sure that the VHS tapes they sell are **120 minutes long on average**.
  – If they are too short, there is a risk of bad publicity.
  – If they are too long, then the material cost is increased.

• In a sample of 10 tapes,
  – the sample average was 120.1 minutes
  – the sample standard deviation was 0.15 minutes.

• Is there sufficient evidence at \( \alpha = .05 \) that the true mean length is different from 120?
What does it look like?

\[ T_{\text{crit}} = T_{0.025,9} = 2.2622 \]

\[ T_{\text{test}} = \frac{120.1 - 120}{0.15\sqrt{10}} = \frac{0.1}{0.047434} = 2.1082 \]
**Acid rain data**

- The EPA states that any area where the average pH of rain is less than 5.6 on average has an acid rain problem.

- Test to see if acid rain is a problem with $\alpha = 0.1$.
  - Is $\alpha = 0.1$ be easier or harder to prove than $\alpha = 0.01$ and why?

- Summary stats were: mean = 4.5779, variance = 0.08368 and the standard error was 0.03048

- What are the null and alternate hypotheses and what is the conclusion?
  
  Ho:
  
  Ha:
What does it look like?

\[ Z_{\text{test}} = \frac{4.5779 - 5.6}{\sqrt{\frac{0.28925}{0.03048}}} = -1.0221 = -33.5 \]

\[ Z_{0.0000} \approx -33.5 \]
Tests for population proportions

• Consider the nurse employment example.

• What are the appropriate hypotheses \((\text{from the boss's point of view})\)?
  – \(H_0:\)
  – \(H_a:\)

• If \(H_0\) is true, what is the probability that 32 or fewer would be handled timely in a sample of 36?
Large sample tests for a proportion

To test $H_0: p = p_0$, use the test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

<table>
<thead>
<tr>
<th>$H_A$</th>
<th>Reject $H_0$ if</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p &lt; p_0$</td>
<td>$Z &lt; -z_\alpha$</td>
</tr>
<tr>
<td>$p &gt; p_0$</td>
<td>$Z &gt; z_\alpha$</td>
</tr>
<tr>
<td>$p \neq p_0$</td>
<td>$</td>
</tr>
</tbody>
</table>
Murder case example

• What are the appropriate hypotheses?
  – Ho:
  – Ha:

• What is the $p$-value?
  – Remember the stats....n=295 jurors of which 22 were African American
    
  – Use StatCrunch>Stat>Proportions>One Sample>Summary
What does it look like?

H₀ : p = 0.13
Hₐ : p < 0.13

<table>
<thead>
<tr>
<th>Proportion</th>
<th>Count</th>
<th>Total</th>
<th>Sample Prop.</th>
<th>Std. Err.</th>
<th>Z-Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>22</td>
<td>295</td>
<td>0.074576</td>
<td>0.01958</td>
<td>-2.83058</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

\[ Z_{test} = \frac{0.07456 - 0.13}{0.01958} = \frac{-0.0554}{0.1958} = -2.83058 \]
New Product example

- We’re introducing a new product. Some folks in the company are all for it and others are adamantly against it.

- We decide to run one last large focus group before we go into production.

- If the actual percentage is > 62%, marketing will be happy.

- If it’s not, the investors will be unhappy.

- Our random sample of 256 customers results in 164 that would buy.

- Management runs the study at a 5% alpha level.
What does it look like?

\[ \hat{p} = \frac{164}{256} = 0.6406 \]

\[ Z_{test} = \frac{0.6406 - 0.62}{\sqrt{0.62(0.38)/256}} = \frac{0.0206}{0.0303367} = 0.6798 \]

p-value is approximately 0.4966

The additional file for Topic 7 contains discussion of Type II errors, along with some graphical presentation.