Today:
Section 12.2, Lesson 3: What can go wrong with hypothesis testing
Section 12.4: Hypothesis tests for difference in two proportions

ANNOUNCEMENTS:
• No discussion today.
• Check your grades on eee and notify me if any of them are incorrect.
• Quiz #7 begins Wed after class and ends Friday.
• Quiz #8 begins Wed before Thanksgiving and ends on Monday after Thanksgiving. That is the last quiz.
• Jason Kramer will give the lecture this Friday.

HOMEWORK (due Fri, Nov 19): Chapter 12: #62, 83, 101

REVIEW OF HYPOTHESIS TEST FOR ONE PROPORTION
ONE-SIDED TEST, WITH PICTURE

\[ H_0: p = p_0 \quad \text{versus} \quad H_a: p > p_0 \]

Reject \( H_0 \) if \( p \)-value < .05. For what values of \( z \) does that happen? Notation: level of significance = \( \alpha \) (alpha, usually .05) \( p \)-value < .05 corresponds with \( z > 1.645 \)

AN ILLUSTRATION OF WHAT HAPPENS WHEN \( H_a \) IS TRUE
A Specific Example
Suppose the truth for the population proportion \( p \) is one standard deviation above the null value \( p_0 \). Then the mean for the standardized scores will be 1 instead of 0. How often would we (correctly) reject the null hypothesis in that case? Answer (purple region) is .2595.

Section 12.2, Lesson 3
What Can Go Wrong in Hypothesis Testing: The Two Types of Errors and Their Probabilities

Example: Case Study 1.6, aspirin and heart attacks. Found statistically significant relationship; \( p \)-value was < .00001.

Possible errors:

Type 1 error (false positive) occurs when:
• Null hypothesis is actually true, but
• Conclusion of test is to Reject \( H_0 \) and accept \( H_a \)

Type 2 error (false negative) occurs when:
• Alternative hypothesis is actually true, but
• Conclusion is that we cannot reject \( H_0 \)
Heart attack and aspirin example:

Null hypothesis: The proportion of men who would have heart attacks if taking aspirin = the proportion of men who would have heart attacks if taking placebo.

Alternative hypothesis: The heart attack proportion is lower if men were to take aspirin than if they were not to take aspirin.

Type 1 error (false positive): Occurs if there really is no relationship between taking aspirin and heart attack prevention, but we conclude that there is a relationship.

Consequence: Good for aspirin companies! Possible bad side effects from aspirin, with no redeeming value.

Type 2 error (false negative): Occurs if there really is a relationship but study failed to find it.

Consequence: Miss out on recommending something that could save lives!

Which type of error is more serious?

Probably all agree that Type 2 is more serious.

Which could have occurred?

Type 1 error could have occurred. Type 2 could not have occurred, because we did find a significant relationship.

Aspirin Example: Consequences of the decisions

<table>
<thead>
<tr>
<th>Decision:</th>
<th>Reject H₀, Conclude aspirin works</th>
<th>Don’t conclude aspirin works</th>
<th>Which error can occur?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₀: Aspirin doesn’t work</td>
<td>Type 1 error: People take aspirin needlessly; may suffer side effects</td>
<td>OK</td>
<td>Type 1 error can only occur if aspirin doesn’t work.</td>
</tr>
<tr>
<td>Ha: Aspirin works</td>
<td>Type 2 error: Aspirin could save lives but we don’t recognize benefits</td>
<td>OK</td>
<td>Type 2 error can only occur if aspirin doesn’t work.</td>
</tr>
<tr>
<td>Which error can occur?</td>
<td>Type 2 can only occur if H₀ is not rejected.</td>
<td>Type 1 can only occur if H₀ is rejected.</td>
<td></td>
</tr>
</tbody>
</table>

Note that because H₀ was rejected in this study, we could only have made a Type 1 error, not a Type 2 error.

Some analogies to hypothesis testing:

Analogy 1: Courtroom:
Null hypothesis: Defendant is innocent.
Alternative hypothesis: Defendant is guilty

Note that the two possible conclusions are “not guilty” and “guilty.” The conclusion “not guilty” is equivalent to “don’t reject H₀.” We don’t say defendant is “innocent” just like we don’t accept H₀ in hypothesis testing.

Type 1 error is when defendant is innocent but gets convicted
Type 2 error is when defendant is guilty but does not get convicted.

Which one is more serious??
Analogy 2: Medical test

Null hypothesis: You do not have the disease
Alternative hypothesis: You have the disease

Type 1 error: You don’t have disease, but test says you do; a "false positive"
Type 2 error: You do have disease, but test says you do not; a "false negative"

Which is more serious??

Notes and Definitions:

Probability related to Type 1 error:
The conditional probability of making a Type 1 error, given that $H_0$ is true, is the level of significance $\alpha$. In most cases, this is 0.05. However, it should be adjusted to be lower (0.01 is common) if a Type 1 error is more serious than a Type 2 error.

In probability notation: $P(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$, usually 0.05.

Probability related to Type 2 error and Power:
The conditional probability of making a correct decision, given that $H_a$ is true is called the power of the test. This can only be calculated if a specific value in $H_a$ is specified.

Conditional probability of making a Type 2 error $= 1 - \text{power}$.

$P(\text{Reject } H_0 \mid H_a \text{ is true}) = \text{power}$.

$P(\text{Do not reject } H_0 \mid H_a \text{ is true}) = 1 - \text{power} = \beta = P(\text{Type 2 error})$.

How can we increase power and decrease $P(\text{Type 2 error})$?

Power increases if:
- Sample size is increased (because having more evidence makes it easier to show that the alternative hypothesis is true, if it really is)
- The level of significance $\alpha$ is increased (because it’s easier to reject $H_0$ when the cutoff point for the $p$-value is larger)
- The actual difference between the sample estimate and the null value increases (because it’s easier to detect a true difference if it’s large) We have no control over this one!

Trade-off must be taken into account when choosing $\alpha$. If $\alpha$ is small it’s harder to reject $H_0$. If $\alpha$ is large it’s easier to reject $H_0$:
- If Type 1 error is more serious, use smaller $\alpha$.
- If Type 2 error is more serious, use larger $\alpha$. 
Ways to picture the errors:

Truth, decisions, consequences, conditional (row) probabilities:

<table>
<thead>
<tr>
<th>Decision:</th>
<th>Don’t reject $H_0$</th>
<th>Reject $H_0$</th>
<th>Error can occur:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>Correct 1 – $\alpha$ Type 1 error $\alpha$</td>
<td>Type 1 error can only occur if $H_0$ true</td>
<td></td>
</tr>
<tr>
<td>$H_a$</td>
<td>Type 2 error $\beta$ Correct power</td>
<td>Type 2 error can only occur if $H_a$ is true.</td>
<td></td>
</tr>
</tbody>
</table>

Error can occur: $H_0$ not rejected $H_0$ rejected

SECTION 12.4: Test for difference in 2 proportions

Reminder from when we started Chapter 9

Five situations we will cover for the rest of this quarter:

<table>
<thead>
<tr>
<th>Parameter name and description</th>
<th>Population parameter</th>
<th>Sample statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>One population proportion (or probability)</td>
<td>$p$</td>
<td>$\hat{p}$</td>
</tr>
<tr>
<td>Difference in two population proportions</td>
<td>$p_1 - p_2$</td>
<td>$\hat{p}_1 - \hat{p}_2$</td>
</tr>
</tbody>
</table>

For Quantitative Variables:

<table>
<thead>
<tr>
<th>Parameter name and description</th>
<th>Population parameter</th>
<th>Sample statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>One population mean</td>
<td>$\mu$</td>
<td>$\bar{x}$</td>
</tr>
<tr>
<td>Population mean of paired differences (dependent samples, paired)</td>
<td>$\mu_d$</td>
<td>$\bar{d}$</td>
</tr>
<tr>
<td>Difference in two population means (independent samples)</td>
<td>$\mu_1 - \mu_2$</td>
<td>$\bar{x}_1 - \bar{x}_2$</td>
</tr>
</tbody>
</table>

For each situation will we:

- Learn about the sampling distribution for the sample statistic
- Learn how to find a confidence interval for the true value of the parameter
- Test hypotheses about the true value of the parameter

Comparing two proportions from independent samples

Reminder on how we get independent samples (Lecture 19):

- **Random samples** taken separately from two populations and same response variable is recorded.
  
  **Example:** Compare proportions who think global warming is a problem, in two different years.

- **One random sample** taken and a variable recorded, but units are categorized to form two populations.
  
  **Example:** Compare 21 and over with under 21 for proportion who drink alcohol.

- Participants randomly assigned to one of two treatment conditions, and same response variable is recorded.
  
  **Example:** Compare aspirin and placebo groups for proportions who had heart attacks.

Hypothesis Test for Difference in Two Proportions

**Example:** (Source: http://www.pollingreport.com/enviro.htm)

Poll taken in June 2006, just after the May release of *An Inconvenient Truth*

Asked 1500 people “In your view, is global warming a very serious problem, somewhat serious, not too serious, or not a problem?”

Results: 615/1500 = .41 or 41% answered “Very serious”

Poll taken again with different 1500 people in October, 2009.

Results: 525/1500 = .35 or 35% answered “Very serious.”

**Question:** Did the population proportion that thinks it’s very serious go down from 2006 to 2009, or is it chance fluctuation?
Notation and numbers for the Example:

Population parameter of interest is $p_1 - p_2$ where:

$p_1 =$ proportion of all US adults in May 2006 who thought global warming was a serious problem.

$p_2 =$ proportion of all US adults in Oct 2009 who thought global warming was a serious problem.

$\hat{p}_1 =$ sample estimate from May 2006 $= \frac{X_1}{n_1} = \frac{615}{1500} = .41$

$\hat{p}_2 =$ sample estimate from Oct 2009 $= \frac{X_2}{n_2} = \frac{525}{1500} = .35$

Sample statistic is $\hat{p}_1 - \hat{p}_2 = .41 - .35 = .06$

Five steps to hypothesis testing for difference in 2 proportions: See Summary Box on pages 531-532

STEP 1: Determine the null and alternative hypotheses.

Null hypothesis is $H_0: p_1 - p_2 = 0$ (or $p_1 = p_2$); null value = 0

Alternative hypothesis is one of these, based on context:

$H_a: p_1 - p_2 \neq 0$ (or $p_1 \neq p_2$)

$H_a: p_1 - p_2 > 0$ (or $p_1 > p_2$)

$H_a: p_1 - p_2 < 0$ (or $p_1 < p_2$)

EXAMPLE:

Did the population proportion who think global warming is “very serious” drop from 2006 to 2009? This is the alternative hypothesis. (Note that it’s a one-sided test.)

$H_0: p_1 - p_2 = 0$ (no actual change in population proportions)

$H_a: p_1 - p_2 > 0$ (or $p_1 > p_2$; 2006 proportion > 2009 proportion)

STEP 2: Verify data conditions. If met, summarize data into test statistic.

For Difference in Two Proportions:

Data conditions: $n\hat{p}$ and $n(1 - \hat{p})$ are both at least 10 for both samples.

Test statistic:

$z = \frac{\text{sample statistic} - \text{null value}}{\text{(null) standard error}}$

Sample statistic: $\hat{p}_1 - \hat{p}_2$

Null value = 0

Null standard error:

- Computed assuming null hypothesis is true.
- If null hypothesis is true, then $p_1 = p_2$
- We get an estimate for the common value of $p$ using both samples, then use that in the standard error formula. Details on next page.

$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{\text{combined successes}}{\text{combined sample sizes}}$

Null standard error = estimate of $\sqrt{\frac{\hat{p}(1 - \hat{p})}{n_1} + \frac{\hat{p}(1 - \hat{p})}{n_2}}$. using combined estimate $\hat{p}$ in place of both $p_1$ and $p_2$.

So the test statistic is:

$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n_1} + \frac{\hat{p}(1 - \hat{p})}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
Step 2 for the Example:

Data conditions are met, since both sample sizes are 1500.

\[ \hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{\text{combined successes}}{\text{combined sample sizes}} = \frac{615 + 525}{1500 + 1500} = \frac{1140}{3000} = .38 \]

Null standard error = \[ \sqrt{(.38)(1-.38)(\frac{1}{1500} + \frac{1}{1500})} = .0177 \]

Test statistic:

\[ z = \frac{\text{sample statistic} - \text{null value}}{\text{(null) standard error}} = \frac{.06 - 0}{.0177} = 3.39 \]

Pictures:

On left: Sampling distribution of \( \hat{p}_1 - \hat{p}_2 \) when population proportions are equal and sample sizes are both 1500, showing where the observed value of .06 falls.

On right: The same picture, converted to z-scores.

Note that area above \( \hat{p}_1 - \hat{p}_2 = 0.06 \) is so small you can’t see it!

STEP 3:
Assuming the null hypothesis is true, find the p-value.

General: p-value = the probability of a test statistic as extreme as the one observed or more so, in the direction of \( H_a \), if the null hypothesis is true.

Difference in two proportions, same idea as one proportion. Depends on the alternative hypothesis. See pictures on p. 517

Example:
Alternative hypothesis is one-sided
\( H_a: p_1 - p_2 > 0 \)
p-value = Area above the test statistic \( z = 3.39 \)
From Table A.1, p-value = area above 3.39 = 1 – .9997 = .0003.

STEP 4:
Decide whether or not the result is statistically significant based on the p-value.

Example: Use \( \alpha \) of .05, as usual
p-value = .0003 <.05, so:
- Reject the null hypothesis.
- Accept the alternative hypothesis
- The result is statistically significant
Step 5: Report the conclusion in the context of the situation.

Example:
Conclusion: From May 2006 to October 2009 there was a statistically significant decrease in the proportion of US adults who think global warming is “very serious.”

Interpretation of the $p$-value (for this one-sided test):
It’s a conditional probability. Conditional on the null hypothesis being true (equal population proportions), what is the probability that we would observe a sample difference as large as the one observed or larger just by chance?

Specific to this example: If there really were no change in the proportion of the population who think global warming is “very serious” what is the probability of observing a sample proportion in 2009 that is .06 (6%) or more lower than the sample proportion in 2006? Answer: The probability is .0003. Therefore, we reject the idea (the hypothesis) that there was no change in the population proportion.