Announcements:

• Quiz #6 begins at 4pm today and ends at 3pm on Wed, Feb 27.

• Midterm Friday, Chs 7 to 10 (including additional stuff done in class). Bring 2 sheets of notes and calculator. See website for sections to skip.

• Wed: Catch up and review. I will do an overview, then ask for questions, then (if time) go over review sheet (already posted on website).

Homework: (Due Wed, Feb 27)
Chapter 10: #4, 30, 78
Chapter 10

Estimating Proportions with Confidence
<table>
<thead>
<tr>
<th>Parameter name and description</th>
<th>Sampling Distribution</th>
<th>Confidence Interval</th>
<th>Hypothesis Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For Categorical Variables:</strong></td>
<td>Chapter 9</td>
<td>Chapter 10</td>
<td>Chapter 12</td>
</tr>
<tr>
<td>One population proportion or binomial probability</td>
<td>Today &amp; Fri.</td>
<td>Mon, Feb 25</td>
<td>Mon, Mar 4</td>
</tr>
<tr>
<td>Difference in two population proportions</td>
<td>Friday</td>
<td>Mon, Feb 25</td>
<td>Wed, Mar 6</td>
</tr>
<tr>
<td><strong>For Quantitative Variables:</strong></td>
<td>Chapter 9</td>
<td>Chapter 11</td>
<td>Chapter 13</td>
</tr>
<tr>
<td>One population mean</td>
<td>Fri, March 8</td>
<td>Mon, Mar 11</td>
<td>Wed, Mar 12</td>
</tr>
<tr>
<td>Population mean of paired differences (paired data)</td>
<td>Fri, March 8</td>
<td>Mon, Mar 11</td>
<td>Wed, Mar 12</td>
</tr>
<tr>
<td>Difference in two population means (independent samples)</td>
<td>Fri, March 8</td>
<td>Mon, Mar 11</td>
<td>Wed, Mar 12</td>
</tr>
</tbody>
</table>
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><strong>For Categorical Variables:</strong></td>
<td></td>
<td></td>
</tr>
<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>
<tr>
<td><strong>For Quantitative Variables:</strong></td>
<td></td>
<td></td>
</tr>
<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 we will:
√ 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
Example from last lecture

Gallup poll of $n = 1018$ adults found 39% believe in evolution. So $\hat{p} = .39$

A 95% confidence interval or interval estimate for the proportion (or percent) of all adults who believe in evolution is .36 to .42 (or 36% to 42%).

Confidence interval: an interval of estimates that is likely to capture the true population value.

Goal today: Learn to calculate and interpret confidence intervals for $p$ and for $p_1 - p_2$ and learn general format.
Remember population versus sample:

- **Population proportion:** the fraction of the *population* that has a certain trait/characteristic or the probability of success in a binomial experiment – denoted by $p$. The value of the *parameter* $p$ is fixed but *not known*.

- **Sample proportion:** the fraction of the *sample* that has a certain trait/characteristic – denoted by $\hat{p}$. The *statistic* $\hat{p}$ is an estimate of $p$.

The **Fundamental Rule for Using Data for Inference:**
Available data can be used to make inferences about a much larger group *if the data can be considered to be representative with regard to the question(s) of interest.*
Some Definitions:

• **Point estimate:** A *single number* used to estimate a population parameter. For our five situations:
  
  \[
  \text{point estimate} = \text{sample statistic} = \text{sample estimate} \\
  = \hat{p} \quad \text{for one proportion} \\
  = \hat{p}_1 - \hat{p}_2 \quad \text{for difference in two proportions}
  \]

• **Interval estimate:** An *interval* of values used to estimate a *population parameter*. Also called a *confidence interval*. For our five situations, **always** the formula is:
  
  Sample estimate ± multiplier × standard error
Details for proportions:

Sample estimate ± multiplier × standard error

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sample estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>$\hat{p}$</td>
<td>$s.e.(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$</td>
</tr>
<tr>
<td>$p_1 - p_2$</td>
<td>$\hat{p}_1 - \hat{p}_2$</td>
<td>See p. 386 for formula</td>
</tr>
</tbody>
</table>
Multiplier and Confidence Level

- The **multiplier** is determined by the desired confidence level.
- The **confidence level** is the probability that the procedure used to determine the interval *will* provide an interval that includes the population parameter. Most common is .95 (95%).
- If we consider *all possible* randomly selected samples of same size from a population, the *confidence level* is the fraction or percent of those samples for which the confidence interval includes the population parameter.
- Often express the confidence level as a percent. Common levels are 90%, 95%, 98%, and 99%.
More about the Multiplier

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Multiplier</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>1.645 or 1.65</td>
<td>( \hat{p} \pm 1.65 ) standard errors</td>
</tr>
<tr>
<td>95</td>
<td>1.96, often</td>
<td>( \hat{p} \pm 2 ) standard errors</td>
</tr>
<tr>
<td></td>
<td>rounded to 2</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>2.33</td>
<td>( \hat{p} \pm 2.33 ) standard errors</td>
</tr>
<tr>
<td>99</td>
<td>2.58</td>
<td>( \hat{p} \pm 2.58 ) standard errors</td>
</tr>
</tbody>
</table>

**Note:** Increase confidence level \( \Rightarrow \) larger multiplier.

Multiplier, denoted as \( z^* \), is the standardized score such that the area between \(-z^*\) and \(+z^*\) under the standard normal curve corresponds to the desired confidence level.
For one proportion: A confidence interval for a population proportion $p$, based on a sample of size $n$ from that population, with sample proportion $\hat{p}$ is:

$$\hat{p} \pm (z^*) \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$
Example of different confidence levels

Poll on belief in evolution:
\( n = 1018 \)
Sample proportion = .39
Standard error = \[ \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{.39(1-.39)}{1018}} = .0153 } \]

90% confidence interval
\[ .39 \pm 1.65 (0.0153) \text{ or } .39 \pm 0.025 \text{ or } .365 \text{ to } .415 \]

95% confidence interval (approximate):
\[ .39 \pm 2 (0.0153) \text{ or } .39 \pm 0.031 \text{ or } .359 \text{ to } .421 \]

99% confidence interval
\[ .39 \pm 2.58 (0.0153) \text{ or } .39 \pm 0.039 \text{ or } .351 \text{ to } .429 \]
Interpreting the confidence interval:

- We are **90% confident** that the proportion of *all* adults in the US who believe in evolution is between .365 and .415.
- We are **95% confident** that the proportion of *all* adults in the US who believe in evolution is between .359 and .421.
- We are **99% confident** that the proportion of *all* adults in the US who believe in evolution is between .351 and .429.

Interpreting the confidence level of 95%:
The interval .359 to .421 *may or may not* capture the true proportion of adult Americans who believe in evolution. But, *in the long run* this procedure will produce intervals that capture the unknown population values about 95% of the time. So, we are **95% confident** that it worked this time.
Notes about interval width

- Lower confidence $\leftrightarrow$ more narrow interval
- Larger $n$ (sample size) $\leftrightarrow$ more narrow interval, because $n$ is in the denominator of the standard error.
- So, if you want a more narrow interval you can either reduce your confidence, or increase your sample size.
• Applet to demonstrate confidence interval concepts
http://www.rossmanchance.com/applets/NewConfsim/Confsim.html

• Note that on average, about 19 out of 20 of all 95% confidence intervals should cover the true population value.
Reconciling with Chapter 5
formula for 95% confidence interval

Sample estimate ± Margin of error where (conservative) margin of error was \( \frac{1}{\sqrt{n}} \)

Now, “margin of error” is \( 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)

These are the same when \( \hat{p} = .5 \). The new margin of error is smaller for any other value of \( \hat{p} \) so we say the old version is conservative. It will give a wider interval.
Comparing three versions (when sample proportion is close to .5)

For the evolution example, \( n = 1018, \hat{p} = .39 \)

- **Conservative** margin of error \( = .0313 \approx .03 \)
- **Approximate** margin of error using \( z^* = 2 \)
  \[
  2 \times .0153 = .0306 \approx .03
  \]
- **Exact** margin of error using \( z^* = 1.96 \)
  \[
  1.96 \times .0153 = .029988 \approx .03
  \]

All very close to .03, and it really doesn’t make much difference which one we use!
Compare methods when sample proportion is close to 0 or 1

Marist Poll in Oct 2009 asked “How often do you text while driving?”  \( n = 1026 \)

Nine percent answered “Often” or “sometimes” so

\[
\hat{p} = .09 \quad s.e.(\hat{p}) = \sqrt{\frac{.09(.91)}{1026}} = .009
\]

- **Conservative** margin of error = .0312
- **Approximate** margin of error = \( 2 \times .009 = .018 \).

This time, they are quite different!

The conservative one is too conservative, it’s double the approximate one! Interval is too wide.
Comparing margin of error

\[
\frac{1}{\sqrt{n}}
\]

- Conservative margin of error \( \frac{1}{\sqrt{n}} \) will be okay for sample proportions near 0.5.
- For sample proportions far from 0.5, closer to 0 or 1, don’t use the conservative margin of error. Resulting interval is wider than needed.
- Note that using a multiplier of 2 is called the approximate margin of error; the exact one uses multiplier of 1.96. It will rarely matter if we use 2 instead of 1.96.
Factors that Determine Margin of Error

1. **The sample size, \( n \).**
   When sample size increases, margin of error decreases.

2. **The sample proportion, \( \hat{p} \).**
   If the proportion is close to either 1 or 0 most individuals have the same trait or opinion, so there is little natural variability and the margin of error is smaller than if the proportion is near 0.5.

3. **The “multiplier” 2 or 1.96.**
   Connected to the “95%” aspect of the margin of error. Usually the term “margin of error” is used only when the confidence level is 95%.
Summary of the Approximate 95% CI for a Proportion

\[ \hat{p} \pm 2 \text{ standard errors} \]

The standard error is \[ s.e.(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \]

**Interpretation**: For about 95% of all randomly selected samples from the population, the confidence interval computed in this manner captures the population proportion.

**Necessary Conditions**: \( n\hat{p} \) and \( n(1 - \hat{p}) \) are both at least 10, and the sample is randomly selected (or representative).
Finding the Formula for a 95% CI for a Proportion – use Empirical Rule:

For 95% of all samples, $\hat{p}$ is within 2 st.dev. of $p$

For 95% of all samples:

$-2$ standard deviations $< \hat{p} - p < 2$ standard deviations

Don’t know true standard deviation, so use standard error.

For approximately 95% of all samples,

$-2$ standard errors $< \hat{p} - p < 2$ standard errors

which implies that for approximately 95% of all samples,

$\hat{p} - 2$ standard errors $< p < \hat{p} + 2$ standard errors

This is the approximate 95% confidence interval formula.
Same holds for *any* confidence level; replace 2 with $z^*$

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where:

- $\hat{p}$ is the sample proportion
- $z^*$ denotes the multiplier.
- $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is the standard error of $\hat{p}$. 
Example 10.3 *Intelligent Life Elsewhere?*

**Poll:** Random sample of 935 Americans
Do you think there is intelligent life on other planets?

**Results:** 60% of the sample said “yes”, $\hat{p} = .60$

\[
s.e.(\hat{p}) = \sqrt{\frac{.6(1-.6)}{935}} = .016
\]

90% Confidence Interval: $.60 \pm 1.65(.016)$, or $.60 \pm .026$

.574 to .626  or  57.4% to 62.6%

98% Confidence Interval: $.60 \pm 2.33(.016)$, or $.60 \pm .037$

.563 to .637  or  56.3% to 63.7%

**Note:** entire interval is above 50% => high confidence that a majority believe there is intelligent life.
Confidence intervals and “plausible” values

• Remember that a confidence interval is an *interval estimate* for a population parameter.
• Therefore, any value that is covered by the confidence interval is a *plausible value* for the parameter.
• Values *not* covered by the interval are still possible, but not very likely (depending on the confidence level).
Example of plausible values

• 98% Confidence interval for proportion (or percent) who believe intelligent life exists elsewhere is:
  
  .563 to .637 or 56.3% to 63.7%

• Therefore, 56.3% is a plausible value for the population percent, but 50% is not very likely to be the population percent.
• Entire interval is above 50% => high confidence that a majority believe there is intelligent life.
New multiplier: To illustrate, let’s do a confidence level of 50%

Poll: Random sample of 935 Americans “Do you think there is intelligent life on other planets?”

Results: 60% of the sample said “yes”, \( \hat{p} = .60 \)

We want a 50% confidence interval.
Want area between \(-z^*\) and \(z^*\) to be .50, so the area to the left of \(z^*\) is .75.
From Table A.1 we have \(z^* \approx .67\). (See next page for Table A.1)

50% Confidence Interval: \(.60 \pm .67(.016)\), or \(.60 \pm .011\)

\(.589\) to \(.611\) or 58.9% to 61.1%

Note: Lower confidence level results in a narrower interval.
Finding $z^*$ Multiplier for 50% confidence interval:
Relevant part of Table A.1. Want $z^*$ with area 0.7500 below it.
Closest is 0.7486, with $z^* = 0.67$. Next is 0.7517, with $z^* = 0.68$.
Could use average, 0.675, but use $z^* = 0.67$ (close enough).

<table>
<thead>
<tr>
<th>$z$</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.5000</td>
<td>.5040</td>
<td>.5080</td>
<td>.5120</td>
<td>.5160</td>
<td>.5199</td>
<td>.5239</td>
<td>.5279</td>
<td>.5319</td>
<td>.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>.5398</td>
<td>.5438</td>
<td>.5478</td>
<td>.5517</td>
<td>.5557</td>
<td>.5596</td>
<td>.5636</td>
<td>.5675</td>
<td>.5714</td>
<td>.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>.5793</td>
<td>.5832</td>
<td>.5871</td>
<td>.5910</td>
<td>.5948</td>
<td>.5987</td>
<td>.6026</td>
<td>.6064</td>
<td>.6103</td>
<td>.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>.6179</td>
<td>.6217</td>
<td>.6255</td>
<td>.6293</td>
<td>.6331</td>
<td>.6368</td>
<td>.6406</td>
<td>.6443</td>
<td>.6480</td>
<td>.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>.6554</td>
<td>.6591</td>
<td>.6628</td>
<td>.6664</td>
<td>.6700</td>
<td>.6736</td>
<td>.6772</td>
<td>.6808</td>
<td>.6844</td>
<td>.6879</td>
</tr>
<tr>
<td>0.5</td>
<td>.6915</td>
<td>.6950</td>
<td>.6985</td>
<td>.7019</td>
<td>.7054</td>
<td>.7088</td>
<td>.7123</td>
<td>.7157</td>
<td>.7190</td>
<td>.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>.7257</td>
<td>.7291</td>
<td>.7324</td>
<td>.7357</td>
<td>.7398</td>
<td>.7422</td>
<td>.7454</td>
<td>.7486</td>
<td>.7517</td>
<td>.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>.7580</td>
<td>.7611</td>
<td>.7642</td>
<td>.7673</td>
<td>.7704</td>
<td>.7734</td>
<td>.7764</td>
<td>.7794</td>
<td>.7823</td>
<td>.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>.7881</td>
<td>.7910</td>
<td>.7939</td>
<td>.7967</td>
<td>.7995</td>
<td>.8023</td>
<td>.8051</td>
<td>.8078</td>
<td>.8106</td>
<td>.8133</td>
</tr>
</tbody>
</table>
Remember conditions for using the formula:

1. Sample is **randomly selected** from the population or at least is representative. 
   **Note**: Available data can be used to make inferences about a much larger group *if the data can be considered to be representative with regard to the question(s) of interest.*

2. Normal curve approximation to the distribution of possible sample proportions assumes a “large” sample size. Both \( n\hat{p} \) and \( n(1 - \hat{p}) \) should be at least 10.
In Summary: Confidence Interval for a Population Proportion $p$ (see page 393)

Exact CI for $p$, any confidence level:

$$
\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}
$$

Approximate 95% CI for $p$:

$$
\hat{p} \pm 2 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}
$$

Conservative 95% CI for $p$:

$$
\hat{p} \pm \frac{1}{\sqrt{n}}
$$
Section 10.3: Comparing two population proportions

• Independent samples of size $n_1$ and $n_2$
• Use the two *sample* proportions as data.
• Could compute separate confidence intervals for the two population proportions and see if they overlap.
• Better to find a confidence interval for the *difference* in the two population proportions.
Case Study 10.3 Comparing proportions

Would you date someone with a great personality even though you did not find them attractive?

Women: \(0.611\) (61.1\%) of 131 answered “yes.”
95\% confidence interval is \(0.527\) to \(0.694\).

Men: \(0.426\) (42.6\%) of 61 answered “yes.”
95\% confidence interval is \(0.302\) to \(0.55\).

Conclusions:
- Higher proportion of women would say yes.
- CIs slightly overlap
- Women CI narrower than men CI due to larger sample size
More accurate to compare the two proportions by finding a CI for the difference

C.I. for difference in two population proportions:

\[
(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
\]
Case Study 10.3 Comparing proportions

Would you date someone with a great personality even though you did not find them attractive?

Women: .611 of 131 answered “yes.”
95% confidence interval is .527 to .694.

Men: .426 of 61 answered “yes.”
95% confidence interval is .302 to .55.

Confidence interval for the difference in population proportions of women and men who would say yes.

\[
(0.611 - 0.426) \pm z* \sqrt{\frac{0.611(1 - 0.611)}{131} + \frac{0.426(1 - 0.426)}{61}}
\]
Case Study 10.3, continued

• A 95% confidence interval for the difference is .035 to .334 or 3.5% to 33.4%.

• We are 95% confident that the population proportions of men and women who would date someone they didn’t find attractive differ by between .035 and .334, with a lower proportion for men than for women.

• We can conclude that the two population proportions differ because 0 is not in the interval.
Section 10.4: Using confidence intervals to guide decisions

• A value *not* in a confidence interval can be rejected as a likely value for the population parameter.
• When a confidence interval for $p_1 - p_2$ does not cover 0 it is reasonable to conclude that the two population proportions differ.
• When confidence intervals for $p_1$ and $p_2$ do not overlap it is reasonable to conclude they differ, but if they do overlap, no conclusion can be made. In that case, find a confidence interval for the difference.
Summary Tables

Sampling distributions, end of Chapter 9, p. 353.
Confidence intervals, end of Chapter 11, p. 439

These include:

• General formula for all 5 situations
• Specifics for each situation, including:
  – Parameter
  – Statistic
  – Relevant choices of these: standard deviation, standard error, standardized statistic, multiplier.
From the Midterm 2 review sheet for Chapter 10 - you should know these now

1. Understand how to interpret the confidence level
2. Understand how to interpret a confidence interval
3. Understand how the sampling distribution for \( \hat{p} \) leads to the confidence interval formula (pg. 380-381)
4. Know how to compute a confidence interval for one proportion, including conditions needed.
5. Know how to compute a confidence interval for the difference in two proportions, including conditions needed.
6. Understand how to find the multiplier for a specified confidence level.
7. Understand how margin of error from Chapter 5 relates to the 95% confidence interval formula in Chapter 10
8. Know the general format for a confidence interval for the 5 situations defined in Chapter 9 (see summary on page 439).