DIRECTIONS

• Textbook directions for some exercises have been altered in this assignment.

• Please write answers to hypothesis tests using the Four Steps of Hypothesis Testing as shown on page 29 in the (review portion of) the Notebook. We’ll also do this on quizzes and exams.

• Text exercises marked with the icon are accessed from the Supplementary Exercises (SE) file posted on the Stats website, not the textbook itself.

• See How To Succeed With Stats Homework in the STATS GUIDE. A great way to s-q-u-e-e-z-e as much knowledge and understanding as possible from the Homework for great quiz and exam preparation!

A. Textbook Exercises for Two Proportions (Chapter 8)

• Exercise 8.53 (p. 473)
  Tips:
  ○ We compare two pop proportions $p_1$ and $p_2$ by estimating how much larger $p_1$ is than $p_2$.
  ○ So we estimate their difference $(p_1 - p_2)$ by the corresponding difference $(\hat{p}_1 - \hat{p}_2)$ in sample proportions.
  ○ The mean of $(\hat{p}_1 - \hat{p}_2)$ over many samples which we could measure is written as $\mu(\hat{p}_1 - \hat{p}_2)$ and shows whether our comparisons are correct on average.
  ○ The standard deviation of $(\hat{p}_1 - \hat{p}_2)$ is written as $\sigma(\hat{p}_1 - \hat{p}_2)$ and shows how accurate our comparisons are. (A larger number indicates more variability and less accuracy.)
  ○ To answer Exercise 8.53 look at the bell curve at the top of textbook p. 473. Use those formulas to calculate the answers. (Then check your answers with the HW 2 Solution.)

• Exercise 8.54
  ○ What happens to standard deviation if we quadruple the sample sizes (from $n_1 = 25$ and $n_2 = 30$ to $n_1 = 100$ and $n_2 = 120$)? Do our stats becomes four times as accurate?

(continued)
• Exercise 8.56

Directions:
- Use calculator and table only.
- Use at least four decimal places in calculations.
- Define parameters \( p_1 \) and \( p_2 \) before starting parts (a)–(d).
- Add part (e):
  Interpret the answer from (d).

• Exercise 8.61 (p. 479)

Directions:
- Write out the Four Steps in place of parts (a), (b), (c). Use calculator and table only.
- Use at least four decimal places in calculations. Use 5% significance.
- Ignore the (confusing) exercise title “Do men lie more often about their height than women?” and test for a difference between men and women instead.

• Exercise 8.56 Part 2

Use MINITAB to find the CI in Exercise 8.56. (Show all decimal places provided.)

• Exercise 8.61 Part 2

Use MINITAB to find the \( P \)-value for Step 2 in Exercise 8.61. (Show the exact answer, with all decimal places. Refer to Topic 2 Notes for MINITAB steps.)

• Exercise 8.57

Directions: Same as we used for Exercise 8.56 (Use calculator and table only.)

• Exercise 8.62

Directions: Same as we used for Exercise 8.61 (Use calculator and table only.)

• Exercise 8.57 Part 2

Use MINITAB to find the CI in Exercise 8.57. (Show all decimal places provided.)

• Exercise 8.62 Part 2

Use MINITAB to find the \( P \)-value for Step 2 in Exercise 8.62. (Show the exact MINITAB answer, with all decimal places.)
Exercise 8.76 (p. 4.84)

Directions:
- The textbook’s reference to “SRS” stands for simple random sample.
- Use either calculator and table or MINITAB (your choice.)
- Write out and clearly label Four Steps. Use 5% significance.
- Add two new parts:
  (c) Interpret the CI.
  (d) Translate the percentages in the CI to actual numbers of UI men and women, under the simplifying assumption that there are 15,000 undergrad men and 15,000 undergrad women (for a total of 30,000 undergrads) at UI.

Additional Practice

(1) A friend and fellow student in Marketing class says:

I looked up survey data on the internet which compares the market share for Coca-Cola Classic by age: younger than 40 years old and 40 or older. I ran MINITAB twice to get 95% confidence intervals for both groups. The proportions which drink Coca-Cola Classic every day are:

- Younger than 40: (0.1104, 0.1789)
- 40 or older: (0.0405, 0.0780)

So with 95% confidence the younger market share exceeds the older market share by at least

$$11.04 - 7.80 = 3.24\%$$

Briefly explain to your friend why the above analysis is not correct!

(Additional Practice continued next page)
(2) Suppose that you work as a college intern at the downtown Iowa City location of Hills Bank. (Many Tippie students do!)

At a 10:00 am meeting with your supervisor you’re scheduled to report recent data findings about how much more likely “impulse buys” are for consumers who use credit cards than for consumers who don’t use credit cards. Refer to the data table in Supplementary Exercise 8.29 (SE file p. 98) but don’t answer questions for that exercise.

Remembering your Stats for Strategy course, you define

\[ p_1 = \text{proportion of impulse buys made by credit-card users} \]
\[ p_2 = \text{proportion of impulse buys made by non-credit-card users} \]

A friend of yours at the bank offered to do a quick MINITAB analysis for you because she knows you’re busy. At 9:55 AM she hands you the following printout:

<table>
<thead>
<tr>
<th>Sample</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>48</td>
<td>0.729167</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>49</td>
<td>0.632653</td>
</tr>
</tbody>
</table>

Difference = p (1) - p (2)
Estimate for difference: 0.0965136
95% CI for difference: (-0.0879433, 0.280971)
Test for difference = 0 (vs not = 0): Z = 1.02 P-Value = 0.308

(a) What’s the printout’s interpretation (with 95% confidence), assuming that it matches your definitions for \( p_1 \) and \( p_2 \)?

(b) What data-entry mistake did your friend apparently make when using MINITAB? Does this affect the interpretation in (a)?

(c) You don’t have time to re-run MINITAB or use a calculator before your meeting, which begins in five minutes! 😊

But you can nevertheless “rescue” the correct interpretation from your friend’s printout by careful reasoning. Do this! (Tip: Which proportions does the printout compare?)

(see next page for last homework problem)
Suppose that we want to consider men only. In particular suppose we’d like to compare men’s rates of lying about height and weight. Our parameters would then be:

\[ p_1 = \text{proportion of men who lie about height in online dating} \]
\[ p_2 = \text{proportion of men who lie about weight in online dating} \]

But the CI formula which we used to compare men to women in Exercises 8.56 and 8.57 cannot be used with \( \hat{p}_1 = 22/40 \) and \( \hat{p}_2 = 24/40 \) from the study to produce a valid 95% CI which compares lying rates for men alone. Why not?

**Multiple choice** (Choose the single best answer.)

(a) Only one population is sampled and the sample lying rates \( \hat{p}_1 \) and \( \hat{p}_2 \) are *not* independent.
(b) Only one population is sampled and the sample lying rates \( \hat{p}_1 \) and \( \hat{p}_2 \) are *are* independent.
(c) Two populations are sampled but the sample lying rates \( \hat{p}_1 \) and \( \hat{p}_2 \) are *not* independent.
(d) Two populations are sampled but the sample lying rates \( \hat{p}_1 \) and \( \hat{p}_2 \) are *are* independent.
(e) None of the answers is correct.

*(end of assignment)*