A. Textbook Exercises

- 8.53

The "mean" and the "standard deviation" in this exercise refer to the bell curve for comparing two proportions.

\[ \mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0.4 - 0.5 = -0.1 \] (On average we're correct!)

\[ \sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{(0.4)(0.6)}{25} + \frac{(0.5)(0.5)}{30}} = 0.1339 \]

- 8.54

(a) \[ \mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0.4 - 0.5 = -0.1 \]

\[ \sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{(0.4)(0.6)}{190} + \frac{(0.5)(0.5)}{120}} = 0.0670 \]

(b)

- Using sample sizes four times as large cuts standard deviation but only in half!
- This is like the idea of diminishing marginal returns which you learned in Microeconomics (the second ice cream cone doesn't taste quite as good as the first) applied to Statistics.
  Our stats are more accurate but they're only twice as accurate instead of four times as accurate.
8.56

\( p_1 \) = proportion of men who lie about height in online dating
\( p_2 \) = proportion of women who lie about height in online dating

Sample Data

\( n_1 = 40 \) \( x_1 \) = number of men who lie = 22
\( n_2 = 40 \) \( x_2 \) = number of women who lie = 17

(a) \( \hat{p}_1 = 22/40 = 0.55 \)
\( \hat{p}_2 = 17/40 = 0.425 \)

(b) \( \hat{p}_1 - \hat{p}_2 = 0.55 - 0.425 = 0.125 \)

(c) \[
\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1 (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 (1 - \hat{p}_2)}{n_2}}
= \sqrt{\frac{(0.55)(0.45)}{40} + \frac{(0.425)(0.575)}{40}}
= 0.1109
\]

(d) So a 95% confidence interval for \( (p_1 - p_2) \) is
\[
(\hat{p}_1 - \hat{p}_2) \pm z^* \sigma_{\hat{p}_1-\hat{p}_2} = 0.125 \pm (1.96)(0.1109)
= 0.125 \pm 0.2173
= (-0.0923, 0.3423)
\]

(e) Interpret:
We are 95% confident that between 9.23% fewer and 34.23% more men than women lie about height.

8.61

Four Steps:

1. (Define)
   (Proportions \( p_1 \) and \( p_2 \) are as defined as in exercise 8.56.)

   \( H_A: p_1 \neq p_2 \) or \( (p_1 - p_2) \neq 0 \)
   \( H_0: p_1 = p_2 \) or \( (p_1 - p_2) = 0 \)

2. (Calculate)

   \[ \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{22 + 17}{40 + 40} = \frac{39}{80} = 0.4875 \]
\[ Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}(1 - \hat{p})}} \]

\[ = \frac{(0.55 - 0.425)}{\sqrt{\left(\frac{1}{40} + \frac{1}{40}\right)(0.4875)(0.5125)}} \]

\[ = \frac{0.125}{0.1118} \]

\[ = 1.12 \]

\[ P \text{-value} = P(Z < -1.12) + P(Z > 1.12) = 2(0.1314) = 0.2628 \]

3. (Decide) Fail to Reject \( H_0 \) since \( P \text{-value} = 0.2628 > 0.05 = \alpha \).

4. (Interpret) There is not sufficient evidence to show that men and women differ in lying about height in online dating.

- Minitab 8.56: \((-0.0923429, 0.342343)\)

- Minitab 8.61: \( P \text{-value} = 0.263 \)

8.57

- \( p_1 \) = proportion of men who lie about weight in online dating
- \( p_2 \) = proportion of women who lie about weight in online dating

Sample Data

| \( n_1 = 40 \)  | \( x_1 = \text{number of men who lie} = 24 \) |
| \( n_2 = 40 \)  | \( x_2 = \text{number of women who lie} = 23 \) |

(a) \( \hat{p}_1 = 24/40 = 0.60 \)

\( \hat{p}_2 = 23/40 = 0.575 \)

(b) \( \hat{p}_1 - \hat{p}_2 = 0.025 \)

(c) \( \sigma_{\hat{p}_1 - \hat{p}_2} = 0.1100 \)

(d) \((-0.1906, 0.2406)\)

(e) Interpret:

We are 95% confident that between \( 19.06\% \) fewer and \( 24.06\% \) more men than women lie about weight.
Four Steps:

1. (Proportions \( p_1 \) and \( p_2 \) are as defined in the previous exercise.)

   \[
   H_A: p_1 \neq p_2 \quad \text{or} \quad (p_1 - p_2) \neq 0 \\
   H_0: p_1 = p_2 \quad \text{or} \quad (p_1 - p_2) = 0 
   \]

2. \( P \)-value = 0.8180

3. Fail to Reject \( H_0 \) since \( P \)-value = 0.8180 > .05 = \( \alpha \).

4. There is insufficient evidence to show that men and women differ in lying about weight in online dating.

- **MINITAB 8.57:** \((-0.190680, 0.240680)\)

- **MINITAB 8.62:** \( P \)-value = 0.820

**8.76**

(a) 1. \( p_1 \) = proportion of men employed in summer

\( p_2 \) = proportion of women employed in summer

\[
H_A: p_1 \neq p_2 \\
H_0: p_1 = p_2 
\]

2. From MINITAB, \( Z = 2.59 \), \( P \)-value = 0.010

3. Reject \( H_0 \) since \( P \)-value = 0.010 < 0.05 = \( \alpha \)

4. There is enough evidence to confirm a difference in the summer employment rate between men and women.

(b) From MINITAB, a 95% CI (rounded to 4 decimal places) is

\((0.0100, 0.0730)\)

(c) With 95% confidence, between 1% and 7.30% more undergrad men than women work during the summer.

(d)

* If the low figure of 1% is correct:

\[1\% \times 15,000 = 150 \text{ fewer women}\]

* If the high figure of 7.3% is correct:

\[7.3\% \times 15,000 = 1095 \text{ fewer women}\]

So between 150 and 1095 fewer UI women than UI men undergrads work during the summer, with 95% confidence.
B. Additional Practice

(1) This isn’t correct for the same reason that comparing men to women in Topic 3 Example 1 (Gender differences in political support) with two separate CI’s isn’t correct: The standard deviations for the two groups need to be merged into a single standard deviation. (See Topic 3 Notes.)

(2)

(a) Credit card users make between 8.79% fewer and 28.10% more impulse buys than non-credit-card users.

(b) A careful look at the output shows that the friend apparently entered the number of planned purchases (not impulse buys) into MINITAB. (See the screenshot below.) Therefore the interpretation in (a) is incorrect!

(c) The friend analyzed the proportion of planned purchases:

\[
(1 - p_1) = \text{proportion of planned purchases by credit-card users}
\]
\[
(1 - p_2) = \text{proportion of planned purchases by non-credit-card users}
\]

so MINITAB actually calculates a CI for:

\[
\left[(1 - p_1) - (1 - p_2)\right] = (p_2 - p_1) :
\]
\[
-0.0879 < p_2 - p_1 < 0.2810
\]

Interpret:

Non-credit-card users make between 8.79% fewer and 28.10% more impulse buys than credit-card users.

This interpretation doesn’t directly answer the supervisor’s question! We can get a direct answer by multiplying all three sides by (-1):

\[
-0.2810 < p_1 - p_2 < 0.0879
\]

Final Interpretation:

Credit card users make between 28.10% fewer and 8.79% more impulse buys than non-credit-card users.

(continued)
(3) The correct answer is (a).

Only one population (of men) is sampled.

The same group of 40 men is asked two questions: whether they lie about height and whether they lie about weight.

Presumably whether or not a man lies about height affects the chances that he’ll also lie about weight so the two sample lying rates $\hat{p}_1$ and $\hat{p}_2$ are highly dependent, not independent!

Since the Topic 3 CI formula assumes independence (see Notebook p. 62), any answer that we calculate from a single sample such as this is nonsense.