**Name:** *Caleb McWhorter — Solutions*
**MAT 222**
**Fall 2019**
**Homework 6**

*"[Nick] I can show you around; I can show you our world. Way up here, it's crystal clear. [Jess] Nick, you're doing Aladdin. [Nick] Again?"*

*–Nick Miller & Jessica Day, New Girl*

**Problem 1:** The Dallas Cowboys (a sports ball team) has made it to the Super Bowl (a big sports ball team event) eight times, tieing for second place in the most number of Super Bowl appearances. As of 2015 at $4 billion, the Dallas Cowboys are the most valued sports team in the world. Since 2010, the Dallas Cowboys (a sports ball team) have won 78 of their 144 games.

(a) Using this data, construct a 90% confidence interval for the win percentage of the Dallas Cowboys. Interpret the results.

*Note there are least 10 successes and failures. We have* $\hat{p} = 78/144 = 0.5417$ *and* $n = 144$*. For a 90% confidence interval, we have* $z^* = 1.645$*. Then we have*

$$0.5417 \pm 1.645\sqrt{\frac{0.5417(1-0.5417)}{144}} = 0.5417 \pm 1.645(0.04152) = 0.5417 \pm 0.0683$$

*This gives confidence interval* $(0.4734, 0.61)$*. Therefore, we are 90% sure that, on average, the Dallas Cowboys win between 47.3% and 61.0% of their games.*

(b) What if the coach of the Dallas Cowboys claimed that the team wins 75% of their games? Use the data to test this statement against the hypothesis that the winning percentage is lower than 75% at a significance level of $\alpha = 0.10$. State your conclusions.

*We use null and alternative hypotheses*

$$\begin{cases} H_0 : p = 0.75 \\ H_a : p < 0.75 \end{cases}$$

*We* $\alpha = 0.10$ *and we have test statistic*

$$z = \frac{0.5417 - 0.75}{\sqrt{\frac{0.5417(1-0.5417)}{144}}} = \frac{-0.2083}{0.03608} = -5.77 \rightsquigarrow 0.000$$

*Then we have* $p$*-value* $p \approx 0$*. Therefore, we reject the null hypothesis. There is sufficient evidence to suggest that the proportion of games the Dallas Cowboys win is less than 0.75.*

(c) Using the data stated in the problem, how many games would you have so sample to estimate the win percentage of the Dallas Cowboys within one percentage point? [Use the same significance level as in the previous parts.]

*We have $m = 0.01$, $z^* = 1.645$, and $p^* = 0.5417$. Then*

$$n \geq \left(\frac{1.645}{0.01}\right)^2 0.5417(1 - 0.5417) = 6718.01$$

*Therefore, a sample of size of at least 6,719 games would be required.*

(d) If no previous data was available to you, how many games would you have so sample to estimate the win percentage of the Dallas Cowboys within one percentage point? [Use the same significance level as in the previous parts.]

*We have $m = 0.01$ and $z^* = 1.645$. Then*

$$n \geq \frac{1}{4}\left(\frac{1.645}{0.01}\right)^2 = 6765.06$$

*Therefore, a sample of size of at least 6,766 games would be required.*

(e) Is it true that the margin of error in the confidence interval from (a) covers all sources of error?

*No. The margin of error in any confidence interval only accounts for error coming from random sampling errors.*

(f) What is the name of the $p$-method you use when there are less than 10 success and 10 failures?

*The method is called the 'Plus Four Method'.*

**Problem 2:** According to a survey by the Hill-HarrisX conduced online in the United States from June 1–2, 2019, 247 of 481 men stated they would support Trump in the 2020 election while 198 of 520 women stated they would support Trump in the 2020 election.[1]

(a) Construct a 99% confidence interval for the difference between male and female support for Trump. Interpret your answer.

*Note there are at least 10 successes and failures. For a 99% confidence interval, we have $z^* = 2.576$. We have $X_M = 247$, $n_M = 481$, $p_M = 247/481 = 0.5135$ and $X_F = 198$, $n_F = 520$, and $p_F = 198/520 = 0.3808$. We have*

$$SE_D = \sqrt{\frac{0.5135(1 - 0.5135)}{481} + \frac{0.3808(1 - 0.3808)}{520}} = 0.0312$$

*Then we have*

$$(0.5135 - 0.3808) \pm 2.576 \cdot 0.0312 = 0.1327 \pm 0.0804$$

*so that we have confidence interval $(0.0523, 0.2131)$. Therefore, we are 99% certain that on average men support Trump between 5.23% to 21.31% more than women.*

[1] https://thehill.com/hilltv/what-americas-thinking/447308-trumps-giant-gender-gap-62-percent-of-women-say-they-are

(b) Test the hypothesis that there is no difference between male and female support for Trump using a significance level of 0.1%. Interpret your answer.

*We have null and alternative hypotheses*

$$\begin{cases} H_0 : p_M = p_F \\ H_a : p_M \neq p_F \end{cases}$$

*Under the null hypothesis, the proportions are equal so we pool them:* $\hat{p} = \dfrac{247 + 198}{481 + 520} = \dfrac{445}{1001} = 0.4446$. *Then we have*

$$SE_{D_{\hat{p}}} = \sqrt{0.4446(1 - 0.4446)\left(\frac{1}{481} + \frac{1}{520}\right)} = 0.0314$$

*Therefore, the test statistic is*

$$z = \frac{0.5135 - 0.3808}{0.0314} = \frac{0.1327}{0.0314} = 4.23 \rightsquigarrow 1.00$$

*Therefore, the* $p$*-value is* $p = 2(1 - 1.00) = 0.00 < \alpha = 0.001$. *Therefore, we reject the null hypothesis. There is enough evidence to suggest that the percentage of support for Trump between men and women is difference.*

(c) What if there was already evidence suggesting that there was a 5% support difference between men and women for Trump? Is the data consistent with this evidence? Explain.

*In this case, the hypothesized difference would be* $0.05$. *Then we would have test statistic*

$$z = \frac{(0.5135 - 0.3808) - 0.05}{0.1327 - 0.05} = \frac{0.0827}{0.0314} = 2.63 \rightsquigarrow 0.9957$$

*Then we have* $p$*-value* $p = 2(1 - 0.9957) = 0.0086 > \alpha = 0.001$. *Therefore, we fail to reject the null hypothesis in this case, i.e. the data would be consistent with the fact that the difference in male and female support for Trump was thought to be 5%.*