

Homework 8 Solutions

Math 150

Enrique Treviño

6.2 Young Americans, Part I.

- (a) False. It could be considered to be slightly left skewed, but the proportion of values to the left of 11 (the outliers) is less than 0.01, so the sample is very close to normal. Going by the book, the answer here would be “True”, since the sample size is not large enough to be considered “normal” and hence it should be skewed in one direction. As a nice example, if the sample size is greater than 50, then $50 \times .23 > 10$, so we can assume normality and hence say it is not skewed. If the sample size is 10, then it’s so small that it is skewed.
- (b) False. The reason is not that $n \geq 30$. For proportions we need to consider the success-fail condition. With a sample size of 40, the successes are $40 \times .77 = 30.8$ and the failures are $40 \times .23 = 9.2$, so we don’t pass that condition. However, doing an analysis of the actual outputs from samples of size 40, the distribution does behave like the normal distribution.
- (c) False. The proportion of the time a sample of size 60 would yield 85% or more under the assumption that 77% of the population think they can achieve the American dream is 0.0893191, which is greater than 0.05.
- (d) True. The proportion of the time a sample of size 120 would yield 85% or more under the assumption that 77% of the population think they can achieve the American dream is 0.020506, which is less than 0.05. Even under a two-sided test would be considered unusual.

6.6 Elderly drivers.

(a)

$$SE = \sqrt{\frac{0.66 \times 0.34}{1018}} = \sqrt{\frac{.2244}{1018}} \approx 0.014847.$$

$$ME = 1.96 \times SE \approx 1.96 \times 0.014847 \approx 0.0291 \approx 0.03.$$

Indeed, the margin of error is approximately 3%.

- (b) No. The confidence interval is (0.63, 0.69), so it doesn’t include 0.70, i.e., 70% is implausible.

6.10 Legalizations of marijuana, Part I.

(a) Sample statistic.

(b)

$$SE = \sqrt{\frac{0.48 \times 0.52}{1259}} \approx 0.0140802.$$

Then

$$ME = 1.96 \times SE \approx 0.0275972.$$

Therefore the confidence interval is

$$.48 \pm 0.0275972,$$

which translates to $(0.4524, 0.5076)$.

- (c) Yes, because the sample size is large. The only qualm would be with whether the respondents were randomly sampled. If the survey was not randomly sampled, then it is not valid to assume it is normally distributed.
- (d) It is plausible that the majority of Americans think marijuana should be legalized. But it is also plausible that a majority of Americans think marijuana should **not** be legalized. Therefore, the headline is not justified.

6.12 Is college worth it? Part I.

- (a) The hypotheses are:

Null Hypothesis: $p = 0.5$.

Alternative Hypothesis: $p < 0.5$.

Since $p = 0.5$, then

$$SE = \sqrt{\frac{0.5 \times 0.5}{331}} \approx 0.0274825.$$

Given that $\hat{p} = 0.48$, we have that

$$z = \frac{0.48 - 0.5}{0.0274825} = \frac{-0.02}{0.0274825} \approx -0.727736.$$

Using Table B.1, we see that the p -value is approximately 0.2327. Therefore we fail to reject the null hypothesis.

We do not have strong evidence supporting that only a minority of Americans who decide not to go to college do so because they cannot afford it.

- (b) Yes, because 0.5 is a plausible proportion as evidenced by the hypothesis test above.

6.14 Is college worth it? Part II.

- (a)

$$SE = \sqrt{\frac{0.48 \times 0.52}{331}} \approx 0.0274605.$$

For a 90% confidence interval we use the multiplier 1.645. Therefore

$$ME = 1.645 \times 0.0274605 \approx 0.0451725.$$

Then the 90% confidence interval is

$$0.48 \pm 0.0451725$$

Therefore we have that the confidence interval is $(0.4348, 0.5252)$.

(b) For a 90% confidence interval the z -multiplier is 1.645. Therefore

$$ME = 1.645 \times SE.$$

We want $ME \leq 0.015$. And we know

$$SE \geq \frac{1}{2\sqrt{n}}.$$

Therefore

$$\begin{aligned} 0.015 &\geq 1.645 \times SE \\ 0.015 &\geq 1.645 \times \frac{1}{2\sqrt{n}} \\ \sqrt{n} &\geq \frac{1.645}{0.015 \times 2} \\ n &\geq \left(\frac{1.645}{0.03}\right)^2 \\ n &\geq 3006.69. \end{aligned}$$

Therefore we want a sample size of at least 3007 to have that small a margin of error.

6.16 Legalize marijuana, Part II. We want $ME \leq 0.02$ and we know $ME = 1.96 \times SE$ for 95% confidence intervals and we know $SE \geq \frac{1}{2\sqrt{n}}$. Therefore

$$\begin{aligned} 0.02 &\geq 1.96 \times SE \\ 0.02 &\geq 1.96 \times \frac{1}{2\sqrt{n}} \\ \sqrt{n} &\geq \frac{1.96}{0.02 \times 2} \\ n &\geq \left(\frac{1.96}{0.04}\right)^2 \\ n &\geq 2401. \end{aligned}$$

Therefore we need a sample size of 2401 or higher.

6.18 Heart transplant success. For the normal approximation to be valid for a confidence interval calculation, we need both the control and the treatment groups to be normal. However, the control has too few “alive” patients to be normal. In general we want at least 10 “successes” and at least 10 “failures” to be able to use the normal distribution. The control group has 4 “successes” and 30 “failures”. It doesn’t have enough successes. The distribution would be skewed. Given the skewness, the confidence interval would not be trustworthy.

6.22 Sleep deprivation, CA vs. OR, Part I. $\hat{p}_1 = 0.08$, $\hat{p}_2 = 0.088$, $n_1 = 11545$, $n_2 = 4691$. Then

$$SE = \sqrt{\frac{.08 \times 0.92}{11545} + \frac{0.088 \times 0.912}{4691}} \approx 0.00474894.$$

Our sample statistic is $\hat{p}_1 - \hat{p}_2 = 0.08 - 0.088 = -0.008$. Then the 95% confidence interval for the difference is

$$-0.008 \pm 1.96 \times 0.00474894 \approx -0.008 \pm 0.0095.$$

Therefore the 95% confidence interval is $(-0.0175, 0.0015)$.

We are 95% confident that the proportion of sleep-deprived Californians is between 1.75% less and 0.15% more than the proportion of sleep-deprived Oregonians.

6.24 Sleep deprivation, CA vs. OR, Part II.

- (a) The hypotheses are:

$$\mathbf{H}_0 : p_1 - p_2 = 0$$

$$\mathbf{H}_A : p_1 - p_2 \neq 0.$$

We have $\hat{p}_1 = 0.08$, $\hat{p}_2 = 0.088$, $n_1 = 11545$, and $n_2 = 4691$. The pooled proportion is

$$\hat{p} \approx \frac{.08 \times 11545 + .088 \times 4691}{11545 + 4691} = \frac{1336.41}{16236} = 0.0823114.$$

Then

$$SE = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n_1} + \frac{\hat{p}(1 - \hat{p})}{n_2}} \approx 0.00484598.$$

Then

$$z = \frac{.08 - .088}{SE} \approx \frac{-.008}{0.00484598} \approx -1.68.$$

Using Table B.1 we see that the area to the left is approximately 0.0465. Therefore the p -value is approximately twice that, which is 0.093. Since it is greater than 0.05, we fail to reject the null hypothesis.

- (b) It could be incorrect. In that case, it means we should have rejected the null hypothesis. This would be a Type 2 error.

6.28 Prenatal vitamins and Autism.

- (a) The hypothesis in words:

Null Hypothesis: The proportion of children with autism born to mothers that take prenatal vitamins is the same as the proportion of children with autism born to mothers that did not take prenatal vitamins.

Alternative Hypothesis: The proportion of children with autism born to mothers that take prenatal vitamins is not the same as the proportion of children with autism born to mothers that did not take prenatal vitamins.

In symbols:

$$\mathbf{H}_0 : p_1 - p_2 = 0.$$

$$\mathbf{H}_A : p_1 - p_2 \neq 0.$$

- (b) Since the children with autism were randomly sampled, the children with normal development were normally sampled, and the population of the sample sizes (respectively) are less than 10% of the respective populations, and both samples are independent of each other, then we can assume that the observations are independent. Note that the entries in the table are 111, 70, 143, and 159, therefore they are all at least 10. Therefore, we have enough “successes” and “failures” in each variable to allow us to use the normal distribution to test the hypotheses.

Now let's test the hypothesis. We have $\hat{p}_1 = \frac{111}{181}$, $\hat{p}_2 = \frac{143}{302}$, $n_1 = 181$, $n_2 = 302$. Then the pooled proportion is:

$$\hat{p} = \frac{111 + 143}{181 + 302} = \frac{254}{483} \approx 0.52588.$$

Then the standard error is

$$SE = \sqrt{\frac{\left(\frac{254}{483}\right)\left(1 - \frac{254}{483}\right)}{181} + \frac{\left(\frac{254}{483}\right)\left(1 - \frac{254}{483}\right)}{302}} \approx 0.0469373.$$

$\hat{p}_1 - \hat{p}_2 \approx 0.13975$. Then

$$z = \frac{0.13975}{0.0469373} \approx 3.02.$$

Using Table B.1 we get the the area to the right of $z = 3.02$ is $1 - 0.9987 = 0.0013$. Therefore the p -value is approximately 0.0026. Therefore we reject the null hypothesis. There is evidence that suggests that $p_1 > p_2$, i.e., that women that don't take prenatal vitamins are more likely to have children with autism.

- (c) The title does seem appropriate, as the evidence does suggest a connection. The title could be misleading if readers think it implies a cause-and-effect situation. The study is observational, so one can at most claim a correlation.

6.30 An apple a day keeps the doctor away. We cannot because the set of students is not independent from itself. One would have to do paired data analysis. We could label a response as a 1 if the student answered “yes” and as a zero if the student answered “no”. Then we would do a t -test using the set of differences.

6.48 2010 Healthcare Law.

- (a) False. We know 46% of Americans in this sample support the health care law. Therefore we are 100% confident.
- (b) True. That's what the confidence interval tells us.
- (c) False. If we considered many samples, then 95% of the confidence intervals would “capture” the parameter, that is not the same as what (c) suggests.
- (d) False. It would be lower.