Homework 12

Insert Name

Math 141, Week 12

Due: 11:59pm, Friday April 22

Instructions

Work through the problems below and submit this document as a knitted .pdf to the Math 141 S22 Wells Lecture gradescope page.

For each problem, put your solution between the bars of red stars.

Acknowledgements

If you work with a classmate, please write a note acknowledging this.

Exercise 1

For Friday's daily survey, students in Nate's 9am and 10am sections of Math 141 were asked to estimate the population of Venezuela. However, students in the two sections were given slightly different prompts:

- Students in the 9am section were first asked to guess whether the population was greater or less than **10 million**, before providing their estimate
- Students in the 10am section were first asked to guess whether the population was greater or less than **65 million**, before providing their estimate.

We are interested in determining whether this investigation demonstrates an *anchor effect*, where the value supplied in the statement of the problem primes a reader to use that value as a starting point to estimate the population of Venezuela.

If an anchor effect is present, we would expect students in the 10am section (who were given the value **65 million**) to have larger estimates for the population than students in the 9am section (who were given the value **10 million**).

The following code chunk loads the data of student responses, including their `estimate` for the population of Venezuela (in millions), and their Math 141 `section`.

```
venezuela <- read_csv("https://reed-statistics.github.io/math141s22-wells-website/data/venezuela.csv")
```

a. What are the two implied population of this investigation? What are the two samples?

b. Specify the response and explanatory variables.

c. Is this an experiment or an observational study? Explain.

d. Was blindness used in this investigation? Explain.

e. Can the results of this investigation be generalized to the population? Can we make causal claims based on this investigation? Explain.

f. Create an appropriate visualization showing the distribution of estimates in each section. Comment on any trends observed.

g. Compute the mean and standard deviation of estimates in each section, along with the size of each sample.

Exercise 2

In this exercise, you will perform statistical inference on the data explored in Exercise 1.

a. State null and alternative hypotheses that could answer the research question based on the type of data collected in this investigation.

b. Use appropriate visualizations to assess whether each sample of estimates comes from an approximately Normal distribution.

c. Do you have any concerns about the independence of samples? (Both whether observations within each sample were independent, and whether the samples were independent of each other)

d. Regardless of your answers to the previous two questions, we will use theory-based methods to assess the hypotheses. Compute the relevant standardized test statistic. What distribution does the test statistic follow, if the null hypothesis is true?

e. Compute the p-value for the test statistic you found in the previous part.

f. Make a conclusion at the $\alpha = 0.05$ significance level.

g. Based on your answers to exercises 1 and 2, do you have any reasons to be skeptical of the conclusions you drew in part f? Explain.

Exercise 3

For each of the following statements, indicate if they are a true or false interpretation of the p-value. If false, provide a reason or correction to the misinterpretation.

You are wondering if the average amount of cereal in a 10oz cereal box is greater than 10oz. You collect 50 boxes of cereal, weigh them carefully, find a standardized test statistic, and use it to get a p-value of 0.23.

a. The probability that the average weight of all cereal boxes is 10 oz is 0.23.

b. The probability that the average weight of all cereal boxes is greater than 10 oz is 0.23.

c. Because the p-value is 0.23, the average weight of all cereal boxes is 10 oz.

d. Because the p-value is small, the population average must be just barely above 10 oz (small effect).

e. If H0 is true, the probability of observing another sample with an average as or more extreme as the data is 0.23.

Exercise 4

New York is known as "the city that never sleeps." A random sample of 25 New Yorkers were asked how much sleep they get per night. Statistical summaries of these data are shown below.

```
tibble(n = 25, mean = 7.73, sd = 0.77, min = 6.17, max = 9.78)
```

```
## # A tibble: 1 x 5
##       n  mean    sd   min   max
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    25  7.73  0.77  6.17  9.78
```

a. What is the implied population for this investigation?

b. What is the implied parameter we are estimating for this population?

c. What additional information would you need to know, either about the survey methodology or the population, in order to determine whether theory-based methods would be appropriate to use in order to estimate the parameter you specified in part b?

d. Create a 95% confidence interval for the parameter specified in part b.

e. Based on your answer to part d, is it plausible that New Yorkers sleep an average of 8 hours each night?

f. Explain why we would not be able to use the simulation based method to obtain a confidence interval just using the information provided in this problem.