Homework 8

Insert Name

Math 141, Week 8

Due: 11:59pm, Friday March 18

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

Suppose that a student is working on a statistics project using data on pulse rates collected from a random sample of 100 students from Reed college, and finds a 95% confidence interval for mean pulse rate to be 65.5 to 71.8. Discuss how each of the statements below would indicate an *improper* interpretation of this interval:

A. "I am 95% sure that all students will have pulse rates between 65.5 and 71.8 beats per minute."

B. "I am 95% sure that the mean pulse rate for this sample of 100 students will fall between 65.5 and 71.8 beats per minute."

C. "There is a 95% probability that the average pulse rate of all students at this college is in the interval from 65.5 to 71.8 beats per minute."

D. "I am sure that 95% of all students at this college will have pulse rates between 65.5 and 71.8 beats per minute."

E. "Of random samples of this size taken from students at Reed college, 95% will have mean pulse rates between 65.5 and 71.8 beats per minute."

Exercise 2

In a recent study, 23 rats showed compassion that surprised scientists. Twenty-three of the 30 rats in the study freed another trapped rat in their cage, even when chocolate served as a distraction and even when the rats would then have to share the chocolate with their freed companion. (Rats, it turns out, love chocolate.) Rats did not open the cage when it was empty or when there was a stuffed animal inside, only when a fellow rat was trapped. We wish to use the sample to estimate the proportion of rats that show empathy in this way.

a. What population parameter do researchers wish to investigate? What is the sample statistic will they use to estimate the population parameter, and what is its value?

b. Describe how you could use 30 slips of paper to create **one** bootstrap statistic. Be specific.

c. Of course, it would it would be extraodrinarily tedious to construct bootstrap distribution by hand, so let's use R. Run the following code chunk to produce the sample of 30 observations that the researchers collected.

```
# Create a dataframe from the sample data
rats <- data.frame(empathy = c(rep("Yes", 23), rep("No", 7)))
```

Now, create a bootstrap distribution of 5,000 bootstrap statistics, using `infer`.

d. Plot the bootstrap distribution and use the plot to describe the center and shape of the distribution.

e. Construct a 90% confidence interval for the paramter of interest using the percentile method.

f. Is there evidence that a majority of rats will show empathy? Justify your answer.

Exercise 3

The previous problem describes a study in which rats showed compassion by freeing a trapped rat. In the study, all six of six female rats showed compassion by freeing the trapped rat, while 17 of the 24 male rats did so. Researchers are interested in determining if it is plausible that male and female rats are equally compassionate based on this data.

To begin, run the following code chunk to produce the sample of 30 observations that the researchers collected.

```
# Create a dataframe from the sample data
mfrats <- data.frame(empathy = c(rep("Yes", 23), rep("No", 7)), gender = c( rep("F", 6), rep("M", 17), r
```

a. What population(s) are the researchers interested in? What population parameter(s) do the researchers wish to investigate? What is/are the sample statistic(s) they will use to estimate the population parameter(s)? What is/are the value(s) of these statistics?

b. Describe how you could use 30 slips of paper to create **one** bootstrap statistic. Be sure to note any differences in your method here from the method you used in problem 1(b).

c. Use `infer` to create a bootstrap distribution of 5,000 bootstrap statistics.

d. Plot the bootstrap distribution and use the plot to describe the center and shape of the distribution.

e. Based on your answer to part (d), do you have any concerns about using the standard error method bootstrap method for confidence intervals?

f. Construct a 99% confidence interval for the paramter of interest using the standard error method (use `infer` to construct your interval).

g. Is there evidence that male and female rats are not equally compassionate? Justify your answer.

Exercise 4

Fuel is expensive (now more than ever!). We want to compare the miles per gallon (MPG) of family sedans to the MPG of sports utility vehicles (SUV) to see if family sedans really do get better gas mileage (on average). I obtained the following sample of cars (in 2011) from the website http://www.fueleconomy.gov/feg/findacar.shtml#findacar.

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

a. What are the populations to be studied? What are the parameters of interest?

b. Create a useful graph of miles per gallon by type of car. Describe any trends observed.

c. Using either percentile or standard error bootstrap method, construct a 95% confidence interval for the difference in mean gas mileage.

d. Suppose we wish to study the difference in **medians** instead of the difference in means. Using the percentile method, construct a 95% confidence interval for the difference in median gas mileage.

e. As we discussed in class, the accuracy of these bootstrap methods requires that the bootstrap distribution of the statistic is fairly bell-shaped and symmetric. Graph the bootstrap distribution of the difference in sample means and the bootstrap distribution of the difference in sample medians. Are they both roughly bell-shaped and symmetric? If one is not, why do you think that is?