Homework 5

Insert Name

Math 141, Week 5

Due: 11:59pm, Friday February 25

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

A study conducted at the University of Denver investigated whether babies take longer to learn to crawl in cold months, when they are often bundled in clothes that restrict their movement, than in warmer months. Infants born during the study year were split into twelve groups, one for each birth month. We consider the average crawling age of babies in each group against the average temperature when the babies are six months old (that's when babies often begin trying to crawl). Temperature is measured in degrees Fahrenheit (F) and age is measured in weeks. (Benson 1993)

Load the data from the study by running the following code chunk.

```
library(openintro)
data("babies_crawl")
```

a. Create an appropriate data visualization showing the relationship between Temperature and Average Crawling Age. Describe this relationship.

b. Hypothesize how the relationship would change if temperature was measured in degrees Celsius (C) and age was measured in months. Then test your hypothesis by using `dplyr` verbs to create new columns for temperature in Celsius and age in months, and then creating a new scatterplot.

c. Compute the correlation between temperature in Fahrenheit and age in weeks.

d. Predict how correlation would change if temperature was measured in Celsius and age was measured in months. Then test your hypothesis by computing the new correlation using the data from from part b.

e. Create a linear model for this data, with temperature in Fahrenheit as the explanatory variable and age in weeks as the response. What are the slope and intercept of the model?

f. Predict how the slope and intercept of the model will change if temperature is measured in Celsius and age is measured in months. Test your hypothesis by creating a new linear model.

Exercise 2

A study published in the Journal of Personality and Social Psychology asked a group of 200 randomly sampled participants recruited online using Amazon's Mechanical Turk to evaluate how they felt about various subjects, such as camping, health care, architecture, taxidermy, crossword puzzles, and Japan in order to measure their attitude towards mostly independent stimuli. Then, they presented the participants with information about a new product: a microwave oven. This microwave oven does not exist, but the participants didn't know this, and were given three positive and three negative fake reviews. People who reacted positively to the subjects on the dispositional attitude measurement also tended to react positively to the microwave oven, and those who reacted negatively tended to react negatively to it. Researchers concluded that *"some people tend to like things, whereas others tend to dislike things, and a more thorough understanding of this tendency will lead to a more thorough understanding of the psychology of attitudes."* ((Hepler and Albarracin 2013))

a. What are the individual observations in this investigation?

b. What are the explanatory and response variables in this investigation?

c. Does this investigation employ random sampling? Explain. How could the investigation have obtained these participants?

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

e. Can a causal link be established between the explanatory and response variables? Why or why not?

f. Can the results of this study be generalized to the population at large?

Exercise 3

Researchers studying anthropometry collected body and skeletal diameter measurements, as well as age, weight, height and sex for 507 physically active individuals. The scatterplot below shows the relationship between height and shoulder girth (circumference of shoulders measured over deltoid muscles), both measured in centimeters. (Heinz et al. 2003)

Load the data with the following code:

```
library(openintro)
data(bdims)
```

a. Compute the mean and standard deviation for both height and shoulder girth. Then compute the correlation between the two variables.

b. Use the statistics calculated in the previous part to write down the equation of the least squares regression line for predicting height as a function of shoulder girth.

c. Verify your formula by using R to calculate the slope and intercept of the regression line.

d. Calculate R^2, and interpret in the context of this application.

e. Suppose a randomly selected student from Math 141 has shoulder girth of 100 cm. What does the model predict is the height of this student?

f. If the student actually is 160 cm tall, what is the value of the residual for this student?

g. A 4-year old child has shoulder girth of 70 cm. Would it be appropriate to use this linear model to predict the height of the child? Explain.

Exercise 4

The `FloridaLakes` dataset (loaded below) describes characteristics of water samples taken at 53 Florida Lakes. Alkalinity (concentration of calcium carbonate in mg/L) and average mercury level is recorded for a sample of fish (largemouth bass) from each lake.

```
library(Lock5Data)
data(FloridaLakes)
```

a. Create a scatterplot of Average Mercury vs. Alkalinity and describe any trends evident.

b. Find a formula for the regression line using R.

c. Interpret the slope and intercept of the line in context.

d. Use appropriate visualizations to assess whether the **LINE** conditions for inference and predictions from linear regression are met. Be sure to comment on each condition.