Homework 6

Insert Name

Math 141, Week 6

Due: 11:59pm, Friday March 4

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

The percentage of a person's weight that is made up of body fat is often used as an indicator of health and fitness. However, accurate methods of measuring percent body fat are difficult to implement. One method involves immersing the body in water to estimate its density and then applying a formula to estimate percent body fat. An alternative is to develop a model for percent body fat that is based on body characteristics such as height and weight that are easy to measure. The dataset `BodyFat` can be loaded by running the code below, and contains such measurements for a sample of 100 men.

```
library(Lock5Data)
data(BodyFat)
```

For each subject we have the percent body fat (Bodyfat) measured by the water immersion method, Age, Weight (in pounds), Height (in inches), and circumference (in cm) measurements for the Neck, Chest, Abdomen, Ankle, Biceps, and Wrist.

a. Create a single data visualization showing the relationship between BodyFat (response) and Height, Weight (Explanatory). Comment on any trends you observe.

---

b. Compute the correlation coefficient between BodyFat and each of Height and Weight.

---

c. Use R to fit a multilinear model for BodyFat as a function of Height and Weight. Display the model coefficients using `get_regression_table`, and write the mathematical formula for the model.

---

d. Use the model to predict the bodyfat of an individual whose weight is 170 pounds and whose height is 76 inches.

e. Build another multilinear model for predicting BodyFat, that includes Abdomen, in addition to Height and Weight. Display the model coefficients using `get_regression_table` and write the mathematical formula for the model.

f. Interpret the coefficient on Weight in the context of this investigation.

g. Give one plausible explanation for why the coefficient on weight changed sign between the model in part (c) and the model in part (e).

Exercise 2

A random sample of 100 postsecondary (i.e. college) teachers is collected from the 2010 America Community Survey Public Use Microdata Sample. We are interested in determining whether the data provides evidence for salary discrimination based on gender among college teachers.

Load the dataset using the following code:

```
library(Lock5Data)
data("SalaryGender")
SalaryGender <- SalaryGender %>% mutate(Gender = ifelse(Gender == 0, "Female", "Male"))
```

The data includes Gender, yearly Salary (in thousands of dollars), Age, and whether or not the teacher has a PhD (codes as No = 0, Yes = 1.

a. Create a side-by-side boxplot comparing the distribution of salary by gender. Comment on any trends observed.

b. Create a linear model for salary as a function of gender. Use the model to determine the average salary for male teachers and for female teachers.

c. Create a scatterplot with Salary on the y axis and Age on the x-axis. Add the linear trendline to the plot, and describe any relationships you observe.

d. Fit a simple linear regression using Salary as the response variable and Age as the explanatory variable and display the coefficients of the model using `get_regression_table`. Calculate the correlation coefficient between these two variables.

e. Calculate the average and standard deviations in Age for both Male and Female teachers.

f. Create an appropriate visualization comparing the distribution of Age by Gender, and comment on any trends you see.

Exercise 3

In the previous exercise, you found that Salary was positively correlated with Age and that male teachers are older on average than Female teachers. In this exercise, you will investigate whether this explains the gap in salary between men and women in postsecondary education.

a. If differences in age were the only source of disparity in salary between Male and Female teachers, what would you expect the scatterplot of Salary vs Age, colored by Gender, to look like?

b. Create a scatterplot of Salary vs Age, colored by Gender. Add parallel slope trendlines for each gender. Comment on any trends observed.

c. Create a multilinear model predicting Salary as a function of both Age and Gender. Display the coefficients of your equation using `get_regression_table`.

d. Based on the coefficients in your regression table, does it appear that men and women earn the same salary, after age is taken into account? Explain your reasoning.

e. Create an interaction model for Salary as a function of Age and Gender. Based on the model, what is the equation for Salary as a function of Age for (i) Women, and (ii) Men?

f. Modify your scatterplot from part (b) to include non-parallel slopes instead of parallel slopes.

g. Do you think the interaction model or the parallel slopes model is more appropriate in this case? why?