Homework 5

GSBA 524

Fall 2022, Term 3

Problem 1

In this problem we will be comparing four data sets (`dataset1.csv`, `dataset2.csv`, `dataset3.csv`, `dataset4.csv`) in various ways.

a) Report the sample mean, sample standard deviation, and sample size for each of the four data sets. (Hint: In Radiant, go to Data > Explore)

b) Fit a regression model of y on x for each of the four data sets and provide the model summary for each of the four regressions.

c) On the basis of parts (a) and (b), would you conclude that these four data sets are quite similar?

d) For each data set, make a scatter plot of y versus x. What do you notice from these plots?

e) Describe in a few sentences the takeaway from this problem.

Problem 2

In our first class, we discussed the (made up) example of an education researcher who visits a middle school cafeteria during lunchtime and measures the heights (in feet) of students and also gives them a math quiz. The file `education.csv` contains the (made up) data. Here are the first four rows of the data set:

math_score	height	grade_level
90.09	4.82	6th
91.14	4.97	6th
89.81	4.84	6th
92.97	4.84	6th

a) Make a scatter plot of math score versus height.

b) Fit a simple linear regression model to predict math score based on height (and provide the model summary).

c) Explain the meaning of the height coefficient in this regression model.

d) Is height a statistically significant variable in this model at the 0.05 significance level? Justify your answer.

e) What proportion of variability in math score is explained by height?

f) Fit a linear regression model with both height and grade level (and provide the model summary).

g) Explain the meaning of the height coefficient in this regression model.

h) Is height a statistically significant variable in this model at the 0.05 significance level? Justify your answer.

i) Looking back at your answers to (d) and (g), provide some context/explanation for what is happening here.

j) Suppose someone asks you to predict the math score of a 7th grader who is 5 feet tall. What would you say? Justify your answer.

k) Suppose someone asks you to predict the math score of a 7th grader who is 10 feet tall. What would you say? Justify your answer.

l) Make a scatter plot of math score versus height, color the points by grade level, and include a line through each grade level. (Hint: In Radiant, go to Data > Visualize, and look for the "Color" option and the "Line" check box.)

m) Is there a statistically significant difference (at the 5% level) in the slopes of the 6th and 7th graders' lines? Justify your answer clearly, showing relevant output in your explanation. (Hint: To answer this, you'll need to fit a regression model with a 2-way interaction between height and grade level and then interpret the relevant part of the model summary.)