Homework Assignment: Week 5

Assignment Due: February 6, 2022 by Midnight CT

Submit electronic solutions to drop box in BrightSpace

**Turn in all code and plots requested. If a plot will help with an explanation, include that plot as well.**

1. After reading the article and examining the analysis of Kobe and the potential for ball hogging, should Kobe have shared the ball more when the Lakers were behind? Explain.

2. Using the dataset in the file bball3pt.csv, conduct of linear fit of percent of 3 pointers attempted as the predictor variable and percent of 3 pointers made as the response variable. Provide an interpretation of your results. The dataset came from <https://www.basketball-reference.com/> .

3. A production facility needs to develop a model for the relationship between run time in minutes and run size in units. The dataset is in the file Proddata.csv. Prepare the model using linear regression and provide an interpretation of the beta hats. Suggest one way to decrease the mean production time for a given run. Is a linear model justified? Explain.

4. Using the dataset in the autos.csv file, complete a multivariable regression to predict the price of the cars. Include four predictor variables, your choice, but all must be numeric. Try to choose variables that will produce the best fit, and explain your rational for thinking it is the best fit. Explain the meaning of each of the beta hats and provide a confidence interval for each. Which of the variables has the most impact on the price of the car? Explain

5. Using the dataset in the autos.csv file, make the following models which relate car length to curb.weight:

a) a linear model, and report the beta hats and the residual standard error.

b) a log-linear model, and report the beta hats and the residual standard error.

c) a loess model, report the beta hats and the residual standard error.

Using the residual standard error as an indicator, the smaller the error the better, which of the models best fits the data? Comment on why you think one model was better than others.

6. In this problem we will do a combination of a log-linear and polynomial fit. Using the Boston data set which is the MASS library (the MASS package is already installed, so you just need to load it with the library() command):

a) Using the age variable as the predictor and the crime (crim) as the response, make a plot of the data. Describe the data.

b) Fit the log(crim) to a quadratic function in age and plot the results.

Comment on how well the fit models the shape of the data. Information about the variable in the Boston dataset can be found at <https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html>.

7. Train and test a multivariate model which attempts to predict the median value from the rooms, crime, tax and age variables in the Boston dataset. Use a linear model, and make the training set 65% and testing set 35% of the dataset. Compare the predicted median values of the model to the median values in the test set. (You will need to define your method of comparison.) Provide a written assessment of the model.

8) Using the dataset for the class project prepare two regression models that attempt to predict the incidence of a major disease in City X. You may pick/identify the major disease. One model must be a multivariate regression using three variables. You can choose the type of the second model. Provide a description of the model and any information the models provided on the incidence of the major disease in City X.