Stat 103 – Due 3-18-20 Homework Six Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Instructions: All answers must use complete sentences and express complete thoughts. All calculations should be done from within software, and any plots or graphs that you decide to include should be referenced in your discussion and should be referenced in your response.

1. A developer of vacation homes is considering purchasing a tract of land near a lake. From previous experience, she knows that the price of a lot is affected by the lot size, number of mature trees, and distance to the lake. From a nearby area, she gathers data on 60 recently sold lots. These data are stored in the file VACATION. Use the data to construct a regression model to predict the value of a lot.
2. Do the model assumptions appear to be satisfied? If not, which ones are violated?
3. What is *R*2? What does it tell you?
4. Which of the explanatory variables is linearly related to the response variable *in this* (the original) *model*?
5. If necessary, create a new model by removing insignificant variables.
6. Interpret the slopes in the new model.
7. Predict with 95% confidence the selling price of a 40,000-square foot lot with 50 mature trees that is located 75 feet from the lake.
8. Estimate with 95% confidence the average selling price of all such lots.
9. The manager of an amusement park would like to be able to predict daily attendance. After some consideration, he decided that the following three factors are critical, yesterday’s attendance, whether it’s a weekday or weekend, and the predicted weather. He then took a random sample of 40 days and recorded the data in the file AMUSEMENT. Since two of the variables are qualitative, he created the following sets if dummy variables:

Weekend = 1 (if weekend)

= 0 (if not)

Sunny = 1 (if mostly sunny is predicted)

= 0 (if not)

Rain = 1 (if rain is predicted)

= 0 (if not)

1. Construct a regression model to predict attendance. Is the model likely to be useful? Include all relevant computer output, organized so that I can follow it.
2. Can we conclude that weather is a factor in determining attendance?
3. Determine the best model using the Akaike Information Criterion (AIC) and Mallow’s Cp statistics. How does this affect your choice of final model? Does it change your answer to part (b)?
4. Does this data provide sufficient evidence that weekend attendance is, on average, larger than weekday attendance? Support your answer.
5. The general manager of a chain of catalog stores wanted to determine the factors that affect how long it takes to unload a truck delivering orders. A random sample of 50 deliveries to a store was observed. The times (in minutes) to unload the truck, the total number of boxes, and the total weight (in hundreds of pounds) were recorded in the file CATALOG.
6. Determine the multiple regression equation.
7. How well does the model fit the data?
8. Perform diagnostics on the model and report your findings.
9. Is multicollinearity a problem? If so, propose a solution.

The manager realized that time of day may affect unloading time. He recorded the following codes: 1 = morning, 2 = early afternoon, and 3 = late afternoon.

1. Construct a regression model that includes the information for the time of day.
2. Does the time of day affect the unloading time? Explain.