PSTAT 126

Regression Analysis

Fall 2017

Homework #4 – Due in Lab Nov 14 - 16

Note: These course materials are the sole property of Dr. Todd M. Gross. They are strictly for use by students enrolled in a course taught by Dr. Gross. They may not be altered, excerpted, distributed, or posted to any website or other document-sharing service without express permission.

*Note: Use R to perform data analyses, and provide annotated code and output.*

1. What is the primary difference between simple linear regression and multiple linear regression?
2. Name three ways in which simple and multiple linear regression are similar.
3. Write formula that solves for estimates of the regression coefficients using Matrix Notation.  
     
   **b** = ?
4. Set up the **Y** vector, **X** matrix, and **β** vector, in symbolic form, for the following regression model, assuming there are 4 observations (i.e., n = 4). In other words, write out each of the matrixes showing where each element (X11 through X43, Y1 through Y4, β0 through β3) belongs.
5. The **Auto** dataset in the **ISLR** library contains information about the gas mileage of cars (**mpg**), along with other measurements such as the weight of the car (**weight**) and the size of the engine (**horsepower**). Use this data set to predict mpg from weight and horsepower. Note: because the variable name **weight** is very common, to avoid confusion create a new variable called **autowt** using the command **autowt = Auto$weight**.
   1. Generate a scatterplot matrix for the three variables (**mpg**, **autowt**, and **horsepower**). Using the individual scatterplots, describe, in words, the relationships between **mpg** and **autowt**, between **mpg** and **horsepower** and between the predictors **autowt** and **horsepower**.
   2. Fit a linear model (name it **fitauto**) that predicts mpg from **autowt** and **horsepower**.
   3. Generate summary output for the **fitauto** linear model.
      1. Write out the linear regression equation using the estimates of intercept and slopes you obtained. Interpret, in words, the slopes for autowt and horsepower?
      2. Test the null hypothesis (H0: **1 = **2 = 0) against the alternative (H1: not all **j = 0). State the value of *F*, the corresponding p-value, and the statistical conclusion. What can you conclude about the linear model?
      3. Interpret the value for multiple R2 (not adjusted R2).
      4. Test the hypothesis that slope for **autowt** equals zero. State the value of *t*, the corresponding p-value, and the statistical conclusion.
      5. Repeat part iv for the predictor **horsepower**.
   4. Generate the following data displays:
      1. Residuals plot with fitted values on the X axis, residuals on the Y axis and a horizontal line at zero
      2. QQ normal plot, with QQ line
      3. Histogram of residuals
      4. Use these three plots to evaluate non-linearity, non-constant variance, non-normality, and outliers. Identify any potential violations.