

Lab 1 Exercise

Stats 413: Applied Regression Analysis

Sep 12, 2019

1. **Auto data**

- (a) Load the **Auto** dataset from **Auto.csv** using the **read.csv()** function. How many rows and columns are there in this dataset? What do the rows and columns represent?
- (b) Is there correlation between the number of cylinders and gas mileage? Make a boxplot of **cylinders** versus **mpg**. Is the apparent correlation intuitive?
- (c) Plot a histogram of **mpg**. Is the distribution of **mpg** symmetric?
- (d) Make a scatterplot matrix of the variables in the dataset. R should automatically make boxplots if there are discrete-valued variables. Are the signs of the apparent correlations intuitive?
- (e) Consider the task of predicting gas mileage from the other variables in the dataset. Which variables seem informative? Justify your answer.