3.4.15

Problem 1 – *Shocks Data Set*

1. (A) Find the R^2 of the regression model for expensive test's shock's rebound predicted by cheap test's shock's rebound
   (B) Find out if the slope could be equal to 1 by using the 95% confidence interval
   (C) Check if the assumptions of the residuals are valid
      1. Examine the histogram of residuals to see if it is normal
      2. Are the residuals independent of each other and have a constant variance?
      3. Are the residuals conveying any information about other residuals?

Problem 2 and 3– *Options for Making a Model*

1. Exploratory Data Analysis -Make a lot of plots to generate preconceived ideas
   a. Look for confounders and variables that are highly correlated with the response variable
2. Generate Initial Conclusion to start off somewhere with your model
   a. Decide what your judgment of the best predictor/dummy/interaction variables are
   b. Look at summary of linear model
      i. Check intercept and slope coefficient
      ii. Pay attention to the parameter's standard deviation and R^2
   c. Look at ANOVA (place the variable you are most concerned with last in your linear model)
      i. Does the addition of more variables improve the residuals sum of squares?
      ii. What does the sum of squares say about the variables effect on the model?
3. Check the robustness of your model by playing around with the variables
   a. Contemplate additional variables or omitting variables
   b. Gauge if the magnitude of the effect of deleted or additional variables is important and effective

Permutation Test –*Titanic Revisited*

Permutation test – allows you to reshuffle your values for a statistical measurement and refit the model for each shuffled data set.

Steps for Permutation Test

1. Specify H_o – the hypothesis that there is no effect
2. Use a test statistic, T - something that measures the departures from H_o)
3. Simulate $(T|H_o)$ - generate a sampling distribution (like in the form of a histogram) to show the probability of your test statistic when the null hypothesis is true
4. Check whether you test statistic is consistent with $P(T| H_o)$ – see if your initial measurement lies in the distribution and to see how significant it is

For Next Time – Use a permutation test to see if there is a significant effect of sex on the wage gap (use salary data set)