Problem Set 2: Uncertainty, Holdouts, and Bootstrapping

Casey Mallon

1. Estimate the MSE of the model using the traditional approach. That is, fit the linear regression model using the entire dataset and calculate the mean squared error for the entire dataset. Present and discuss your results at a simple, high level.
2. Calculate the test MSE of the model using the simple holdout validation approach.

- Split the sample set into a training set (50%) and a holdout set (50%). Be sure to set your seed prior to this part of your code to guarantee reproducibility of results.
- Fit the linear regression model using only the training observations.
- Calculate the MSE using only the test set observations.
- How does this value compare to the training MSE from question 1? Present numeric omparison and discuss a bit.

3. Repeat the simple validation set approach from the previous question 1000 times, using 1000 different splits of the observations into a training set and a test/validation set. Visualize your results as a sampling distribution (hint: think histogram or density plots). Comment on the results obtained.
4. Compare the estimated parameters and standard errors from the original model in question 1 (the model estimated using all of the available data) to parameters and standard errors estimated using the bootstrap (B = 1000). Comparison should include, at a minimum, both numeric output as well as discussion on differences, similarities, etc. Talk also about the conceptual use and impact of bootstrapping.