

# ISyE 6203 :Homework 3

Mahadevan Vaidyanathan

March 5, 2016

## 1 PROBLEMS WITH DATA AND DEALING WITH THEM

- First and foremost, almost 80% of the miles data has zeroes. These are truncated to zero but are somewhere between 0 and 1, since there is generally a real non zero time associated with these zeroes.

The Solution : I created a new data table with all non zero miles data and the "Act Time" data associated with each one of those. Once this table was created, I ran a simple linear model in R with no intercept on this data; with the dependent variable being Miles travelled and the input variable as "Act Time". Once this model was ready, I used this model on all rows with zero miles and found the best estimate for the miles travelled.