

Peer-graded Assignment: Regression Models Course Project

Alejandro Osorio

June 29, 2018

Choosing the Model

As

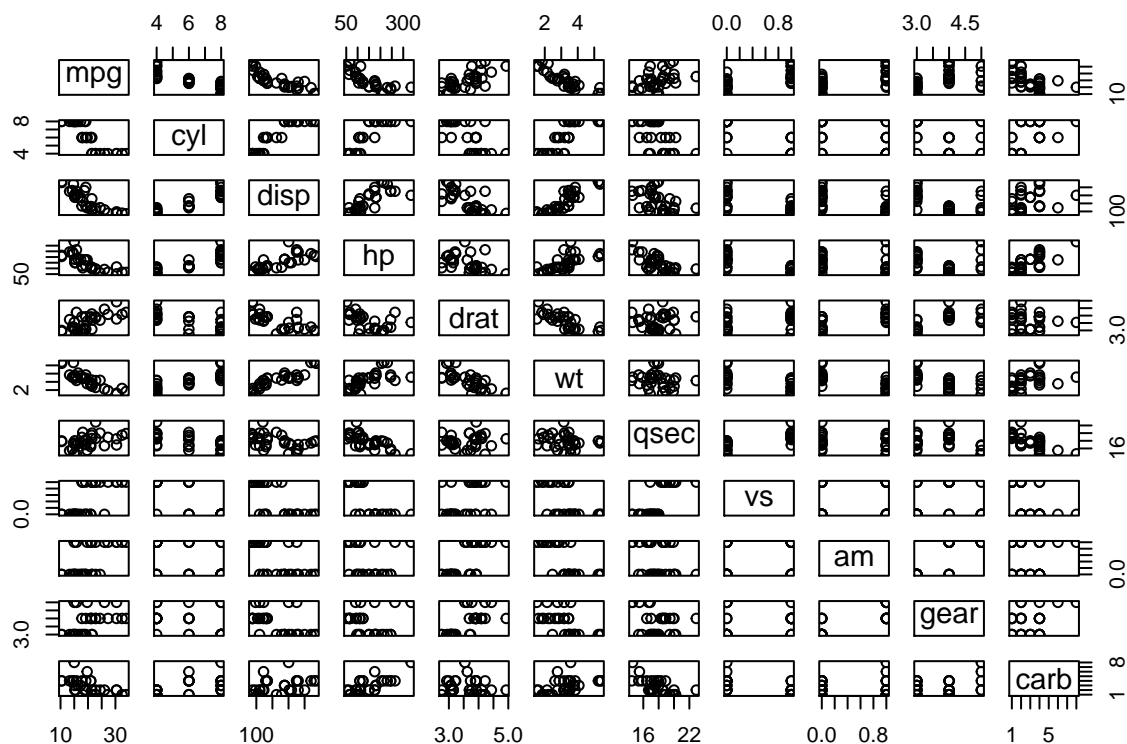
Appendix 1: Preliminary Data Analysis

Basic properties of the ‘mtcars’ dataset:

```
## 'data.frame':   32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

So, we’re talking of a small sample of only 32 observations.

Additionally, the visual correlation between variables can be seen as follows:



Appendix 2: Choosing the Predictors

After studying the meaning of each variable, which shows it's a consequence of the rest of variables 'qsec' (1/4 mile time) is eliminated from the analysis.

The rest of the variables make sense as potential predictors not only from their meanings, but also by the visual correlations between them and 'mpg', as observed in Appendix 1 (such as the inverse relation between 'mpg' and 'hp' or 'wt', among others).

Appendix 3: Choosing the Model Type

Given the nature of the analysis (continuous outcome, obtained from discrete and continuous predictors), the size of the dataset (only 32 observations), plus the visual correlations observed between 'mpg' (the outcome) and its predictors, the model to be used will be linear (lm type).