

Regression Models: Course Project

rtaph

December 21, 2014

Executive Summary

The goal of this assignment is to find a relationship between a set of variables and miles per gallon (MPG) from the `mtcars` dataset. In particular, I am looking at the following two questions:

- Is an automatic or manual transmission better for MPG?
- What is the MPG difference between automatic and manual transmissions?

Exploratory Data Analysis

A number of exploratory charts are presented in the appendices. From initial inspection, there appears to be strong correlation between `mpg` and the `cyl`, `disp`, `wt` variables (the regression lines seem to fit the data well).

Model Fitting

My strategy for model selection is to

```
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amTRUE      7.244939   1.764422  4.106127 2.850207e-04
```

A review of the model summary (see appendices) reveals that it is statistically significant ($p < .01$).

```
## (Intercept)          wt          qsec          amTRUE
##      9.617781      -3.916504      1.225886      2.935837
```

The multivariate model improves our regression significantly compared to the initial model (RSS of 169 vs 721). This difference is statistically significant ($p < .01$). ##### Model 2

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + qsec + am
## Model 2: mpg ~ am
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      28 169.29
## 2      30 720.90 -2    -551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a significant reduction in the residual sum of squares by using the multivariate model. This difference is significant ($p < .01$).

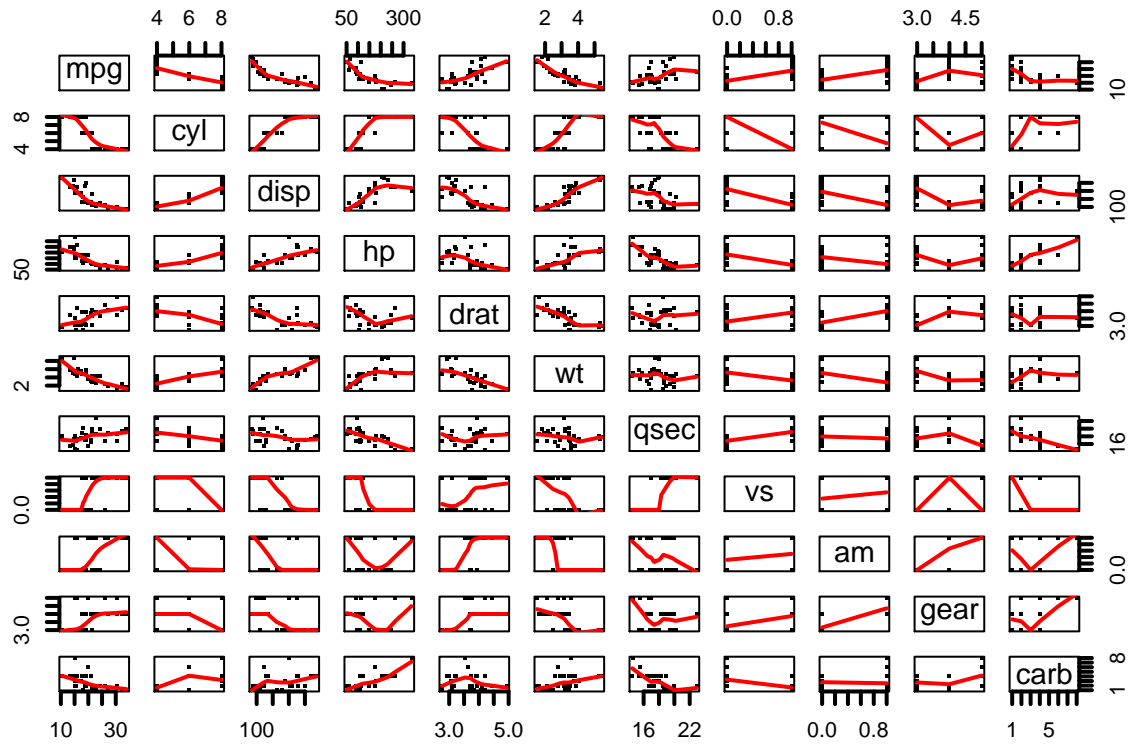
Coefficient Interpretation

we estimate an expected XX ### Findings

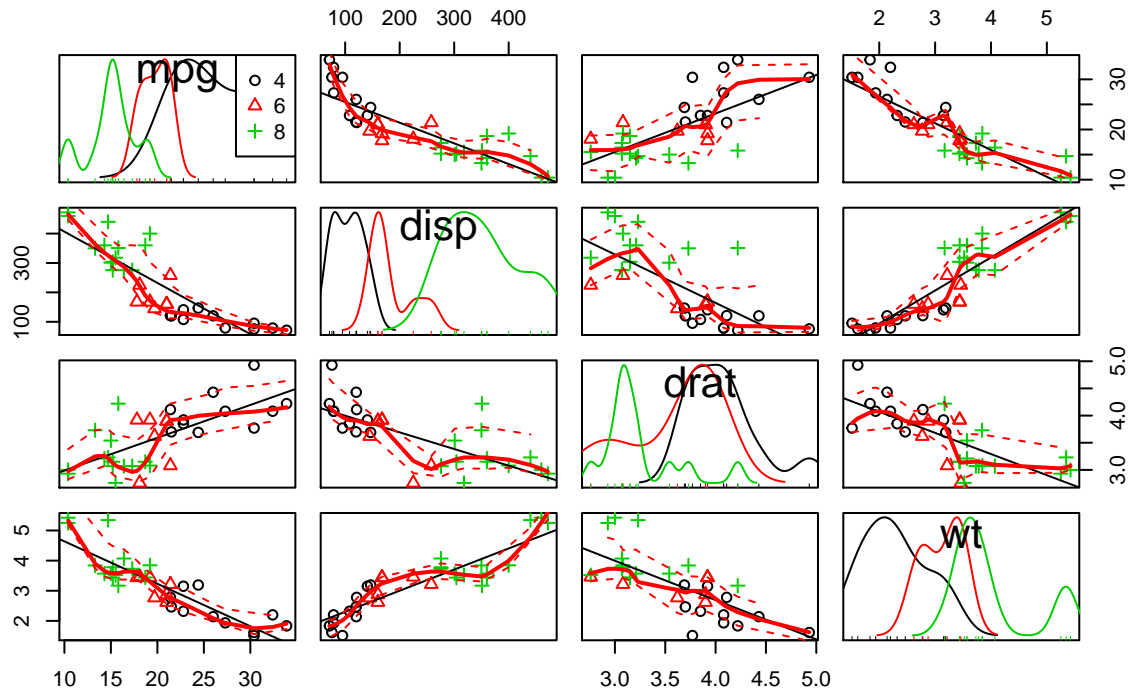
Uncertainty in the models can be quantified with the confidence intervals.

Appendices

Appendix A: Exploratory Charts



Exploratory Chart



Appendix B: Model Diagnostic Plots

