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. I 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

I have pr-processed so that `am` is converted to a logical variable instead of a numeric one. 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 regressionn lines seem to fit the data well).

Model 1

I begin the analysis by fitting a simple linear regression model of mpg to one of the hypothesized covariates from the exploratory stage. In this case, I have chosen to model automatic transmission, `am`. The coefficients of the model are presented here:

```
##              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
```

The interpretation of this mod

Many other simple linear models could be fit. My strategy for model selection is to

A review of the model summary (see appendices) reveals that the t-test for $H_0 : \beta_{mpg} = 0$ versus $H_a : \beta_{mpg} \neq 0$ is statistically signficant ($p < .01$).

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

The coefficient interpretation is that holding all other variables fixed, `mpg` XXX...

The multivariate model improves our regression significantly compared to the initial model (RSS of 169 vs 721).

The diagnostic plots (appendix B) reveal that there is no apparent heteroskedasticity or correlation between the standarized residuals. The Q-Q plot fits the identify line nicely, suggesting our data is normally distributed and that our parametic tests are appropriate.

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
```

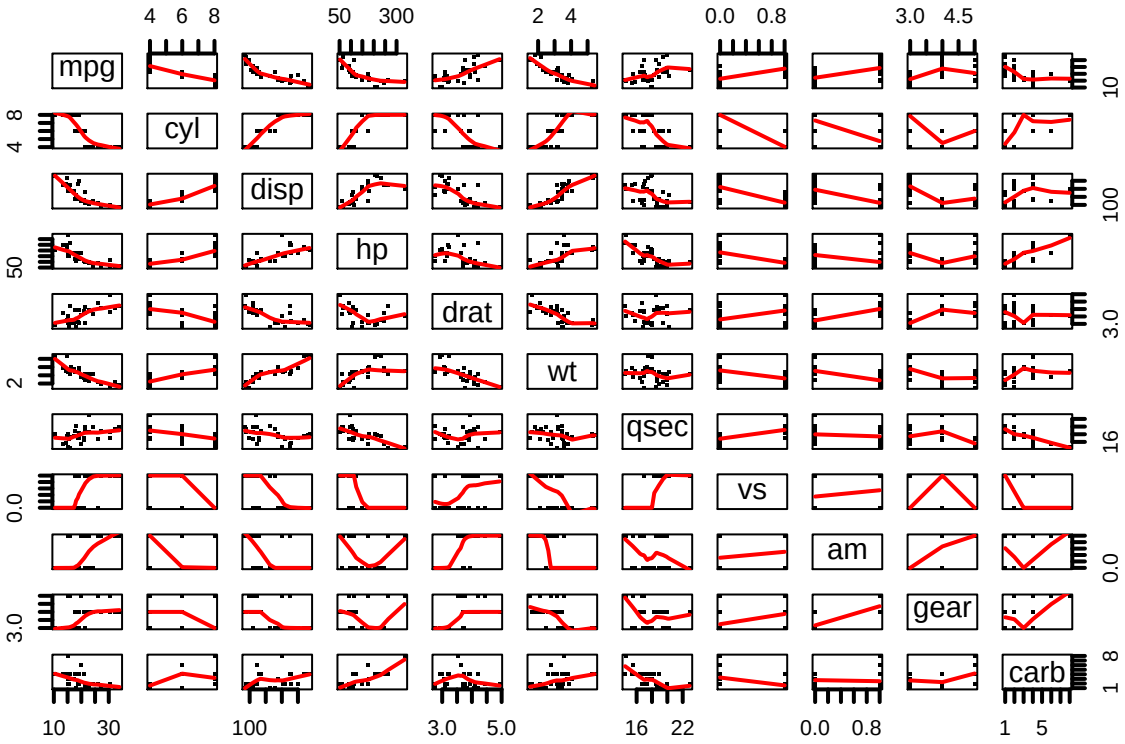
There is a significant reduction in the residual sum of squares by using the multivariate model. This difference is significant ($p < .01$).

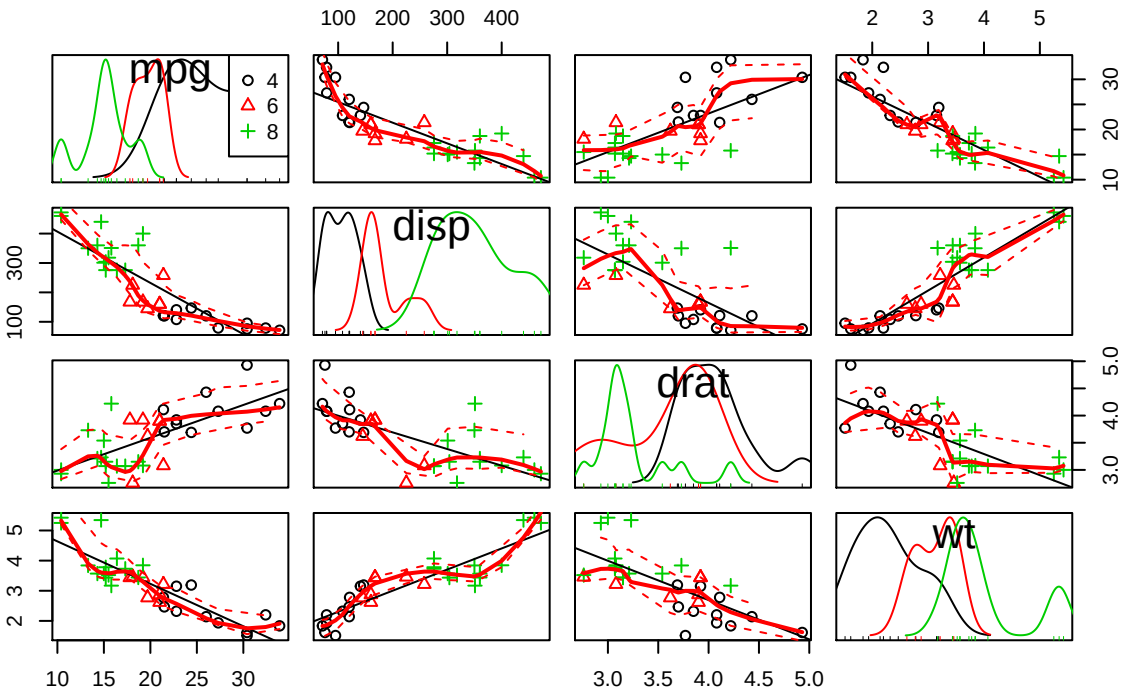
coefficient Interpretation we estimate an expected XX ## Findings

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

Appendices

Appendix A: Exploratory Charts





Appendix B: Model Diagnostic Plots