Regression Models - Course Project

Brian Altman

December 25, 2015

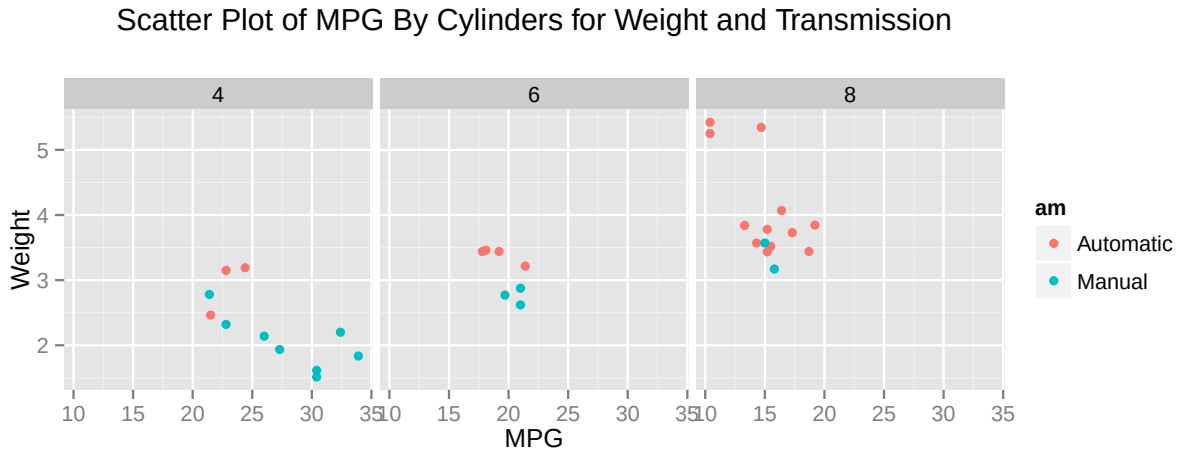
Executive Summary

Based on data from 1974 Motor Trend US magazine (mtcars in R), the following was investigated:

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

Initially, this study showed manual transmissions on average have 7.3 MPG better performance than automated transmission cars (24.392 MPG vs 17.147 MPG). With further analysis, cylinders(cyl), horsepower(hp) and weight(wt) together had impact on MPG and manuals instead added just 1.8 MPG over automatic transmissions.

Initial exploratory analysis The following visualizes weight, cylinders and mpg relationship. See the Appendix for addditional analysis.



Inference Null hypothesis: No MPG difference for automatic and manual transmissions. Using the following t-test:

```
model.AM <-t.test(mpg ~ am, data=mtcars) #See Appendix
```

The P-value of 0.00137 suggests automatic and manual transmissions are of different populations. Manual Avg MPG is 24.39 while Automated is 17.147, a difference of 7.245 MPG.

Regression Analysis A model using the full set of variables was developed. The full model:

```
model.full <- lm(mpg ~ ., data=mtcars) #See Appendix
```

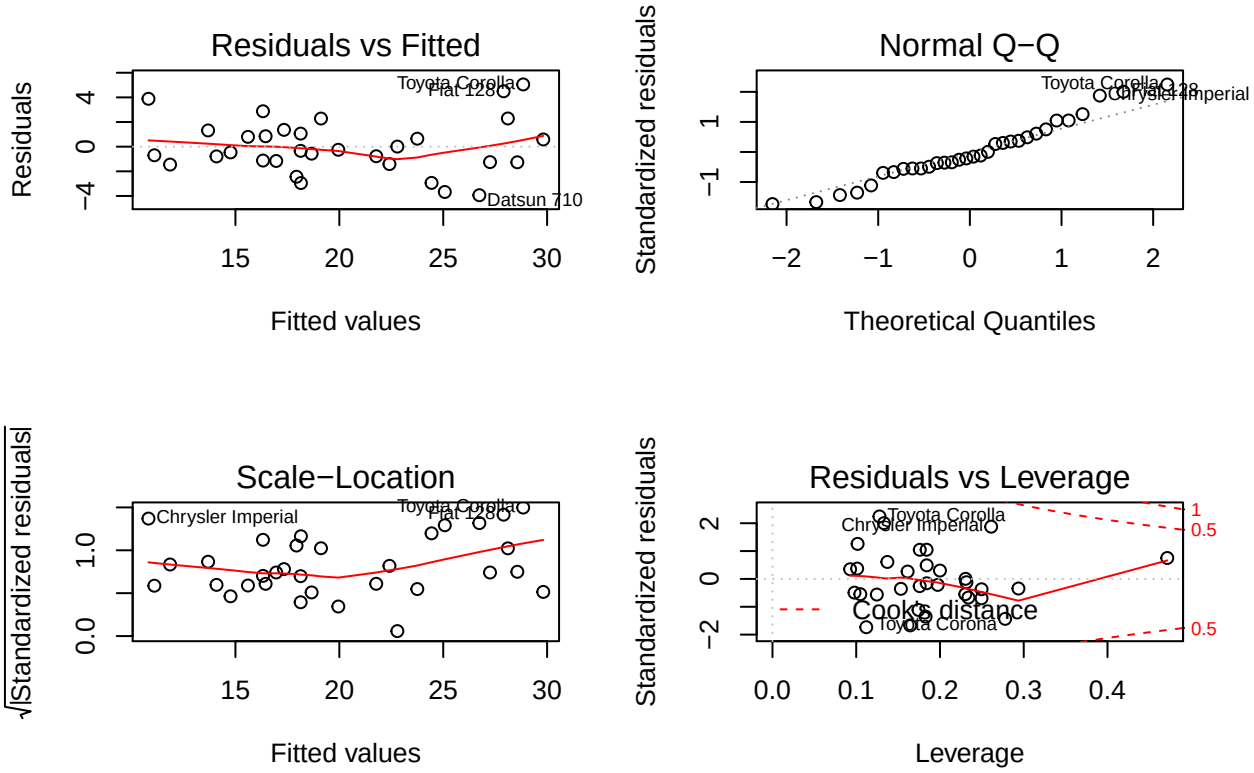
Nearly 78% of the MPG variance can be explained by all the 10 variables. Using the Full Model, the STEP function was used to help determine the most significant variables.

```
summary(model.step)  #See Appendix
```

The p-value indicates the variables cyl(6,8), hp and wt in contribute to the accuracy of the model and, with AM, help explain approximately 84% of the variability. Manual transmission achieve 1.8 more MPG compared to Automatic transmission cars when adjusting for hp, cyl, and wt. The final model:

```
model.final <- lm(mpg ~ am + cyl + wt + hp, data=mtcars) #See Appendix
```

Model Residuals and Diagnostics Residuals and other diagnostics were used to identify potential Final model problems. The top 3 observations were determined and reviewed for leverage and outliers. Please see Appendix also



Conclusions

No patterns in residual plots, leverage and outliers suggested that there were additional, significant predictors of Fuel consumption. The QQ-plot vaidates the normality of the final model.

Manual transmission cars achieve 1.8 better MPG when adjusting for hp, cyl, and wt.

Appendix

```
dim(mtcars)
```

```
## [1] 32 11
```

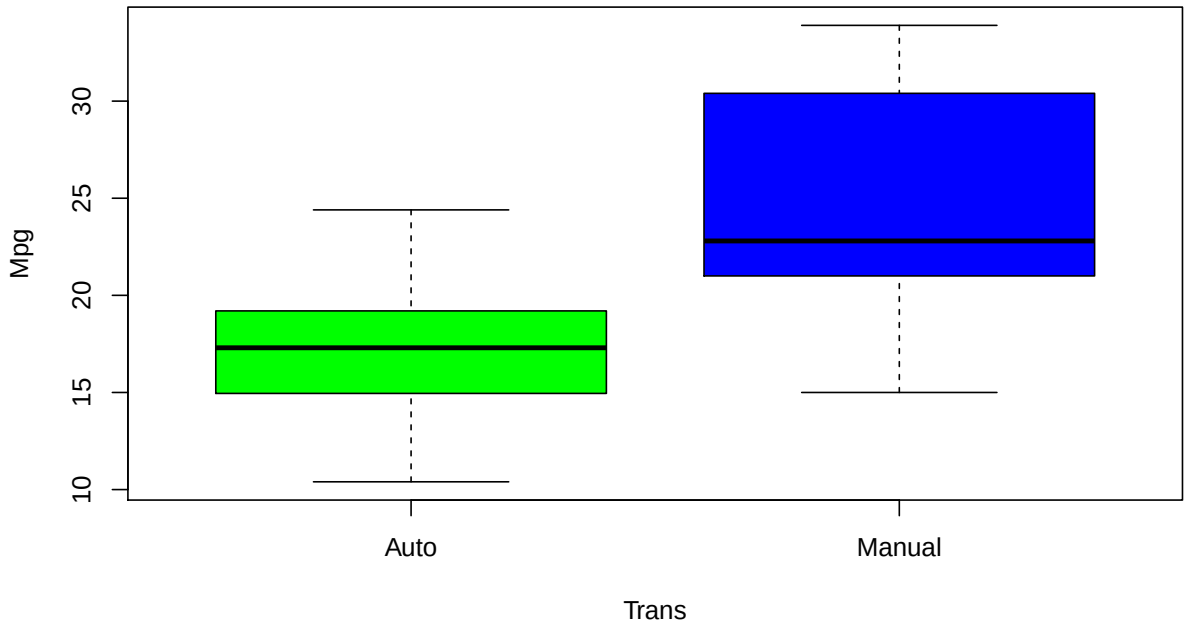
```
leverage <- hatvalues(model.final)
tail(sort(leverage),3)
```

```
##       Toyota Corona Lincoln Continental       Maserati Bora
##           0.2777872           0.2936819           0.4713671
```

```
influential <- dfbetas(model.final)
tail(sort(influential[,6]),3)
```

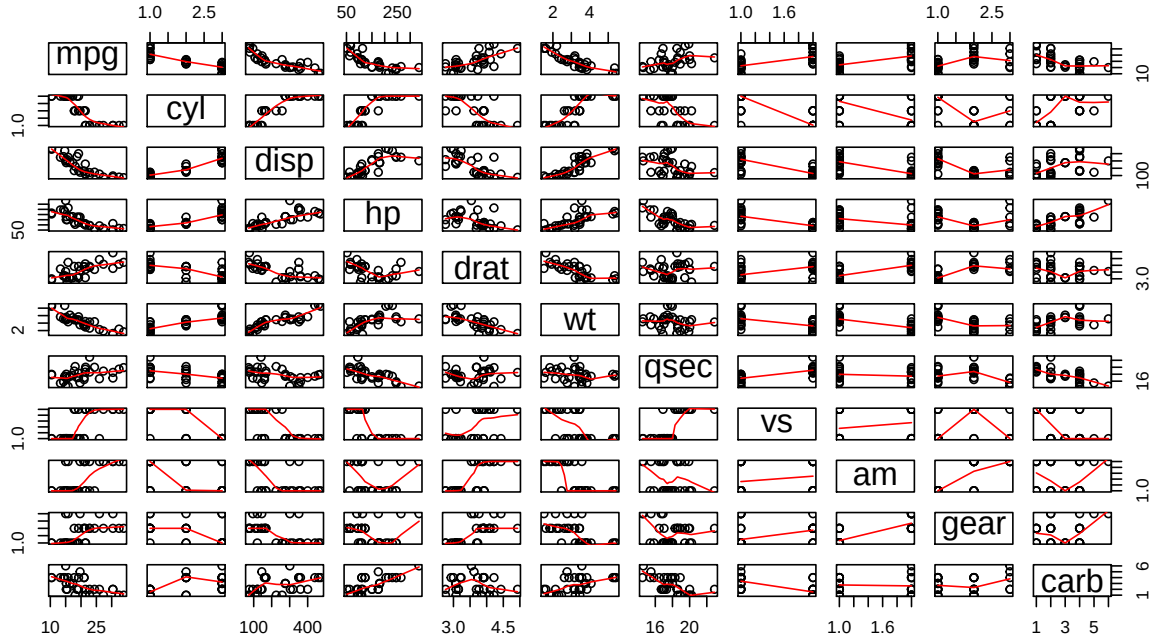
```
## Dodge Challenger      AMC Javelin    Maserati Bora
##        0.2945800        0.3495625        0.5250878
```

```
boxplot(mpg ~ am, data = mtcars, col = (c("green","blue")), ylab = "Mpg", xlab = "Trans")
```




```
pairs(mtcars, panel=panel.smooth, main="Pair Graphs")
```

Pair Graphs




```
summary(model.full)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.87913   20.06582   1.190   0.2525
## cyl6        -2.64870    3.04089  -0.871   0.3975
## cyl8        -0.33616    7.15954  -0.047   0.9632
## disp         0.03555    0.03190   1.114   0.2827
## hp          -0.07051    0.03943  -1.788   0.0939 .
## drat         1.18283    2.48348   0.476   0.6407
## wt          -4.52978    2.53875  -1.784   0.0946 .
## qsec         0.36784    0.93540   0.393   0.6997
## vs1          1.93085    2.87126   0.672   0.5115
## amManual     1.21212    3.21355   0.377   0.7113
## gear4        1.11435    3.79952   0.293   0.7733
## gear5        2.52840    3.73636   0.677   0.5089
## carb2       -0.97935    2.31797  -0.423   0.6787
## carb3        2.99964    4.29355   0.699   0.4955
## carb4        1.09142    4.44962   0.245   0.8096
## carb6        4.47757    6.38406   0.701   0.4938
```

```
## carb8        7.25041    8.36057   0.867   0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF,  p-value: 0.000124
```

```
summary(model.final)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## amManual     1.80921    1.39630   1.296  0.20646
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## wt          -2.49683    0.88559  -2.819  0.00908 **
## hp          -0.03211    0.01369  -2.345  0.02693 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```