Regression Model Course Project : Motor Trend

Jonathan Lok-Chuen Lo

November 21, 2015

Executive Summary

This report aims to explore the relationship between a set of variables and miles per gallon (MPG). In particular, we are interested in: 1) Is an automatic or manual transmission better for MPG: 2) Quantify the MPG difference between automatic and manual transmissions. Data used in the analysis was based on the *mtcars* dataset which was extracted from the 1974 Motor Trend US magazine for 32 automobiles (1973-74 models).

Regression of MPG against only transmission types shows that manual transmission is 7.245 MPG better than automatic transmission. However, by including weight and 1/4 mile time, the result becomes inconclusive. It is found that the dataset contains more manual transmission cars of smaller size and lower engine power, which could skew the analysis (i.e. the apparent better MPG of manual transimission could be due to other factors such as weight and engine power).

Explorary Data Analysis

Load the *mtcars* dataset and the required libraries. Transmission is converted into a factor variable. Box-plot of MPG vs transmission is shown in Appendix A.

```
library(ggplot2); library(dplyr); library(datasets)
data(mtcars)
mtcars$am = factor(mtcars$am, c(0, 1), labels = c("Automatic", "Manual"))
```

In the first glance, manual transmission seems to be more efficient in terms of MPG. However, MPG is highly dependent on other factors such as engine power and weight. Appendix B and C shows MPG vs number of cylinders and weight respectively, between automatic and manual transmission. It shows the dataset contains a larger number of smaller manual transmission cars.

Model Selection

MPG is dependent on many other factors. Two important factors are the car weight and engine power. For engine power, 1/4 mile time, displacement and horsepower are all good indicators. To avoid including unnecessary variables, a nested model testing is performed.

```
fit1 <- lm(mpg ~ am, data = mtcars)
fit2 <- update(fit1, mpg ~ am * (qsec + wt))
fit3 <- update(fit1, mpg ~ am * (qsec + wt + hp + disp))
anova(fit1, fit2, fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + qsec + wt + am:qsec + am:wt
## Model 3: mpg ~ am + qsec + wt + hp + disp + am:qsec + am:wt + am:hp +
##     am:disp
```

```
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)    
## 1     30 720.90                                  
## 2     26 116.47  4    604.42 31.8498 7.329e-09 ***
## 3     22 104.37  4     12.10  0.6376    0.6412    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA above, using 1/4 mile time and weight as regressors reduce the residual sum of squares (RSS) and the result is significant. Including horsepower and displacement only improves the RSS marginally. Hence, the regression model is chosen to use transmission , 1/4 mile time and weight.

Regression Model

First, we run a regression model between MPG and transmission types only.

```r
fitSameSlope <- lm(mpg ~ am, data = mtcars)
summary(fitSameSlope)$coef
```

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

Results above shows that the average MPG for manual transmission is 7.245 MPG higher as compared to automatic transmissions of average 17.147 MPG.

However, as discussed earlier, the dataset contains more data for manual transmission cars having smaller weight with less engine power, as compared to automatic transmission. A regression with adjustment using weight and 1/4 mile time is used.

```r
fitDifferentSlope <- lm(mpg ~ am * (wt + qsec), data = mtcars)
summary(fitDifferentSlope)$coef
```

```
##                 Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)   11.2489412  6.9921572  1.6087941 0.1197381605
## amManual       8.9264577 12.6661636  0.7047483 0.4872324402
## wt            -2.9962762  0.6909630 -4.3363771 0.0001936034
## qsec           0.9454396  0.3067164  3.0824553 0.0048133078
## amManual:wt   -3.7580835  1.5157747 -2.4793153 0.0199691091
## amManual:qsec  0.2355322  0.5565520  0.4231990 0.6756302398
```

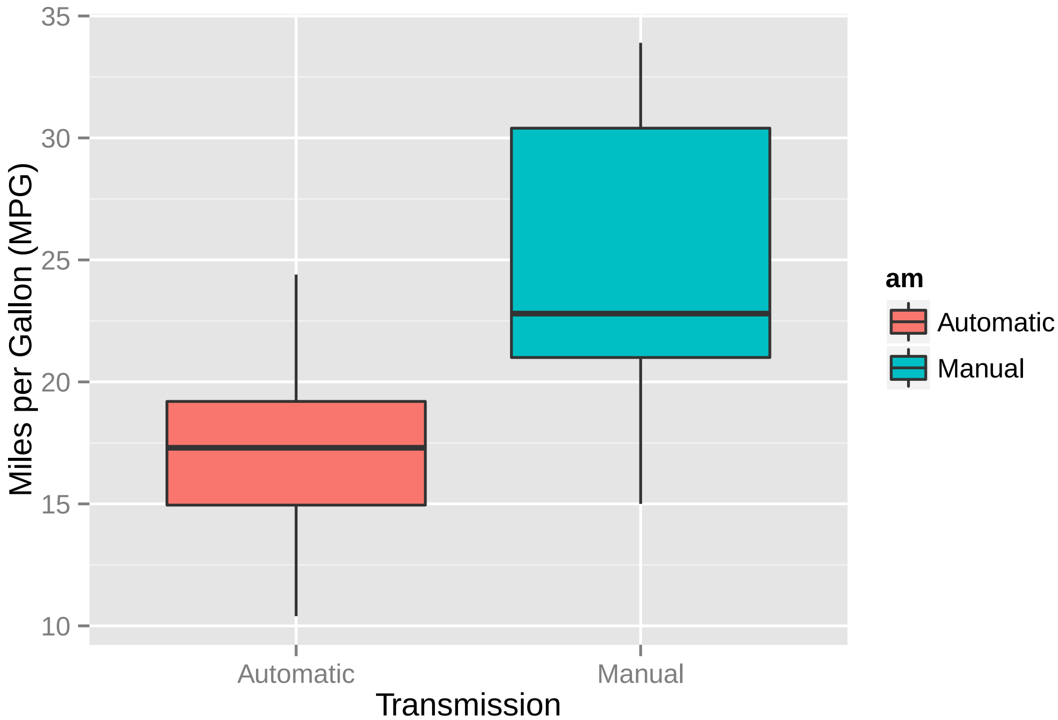
Above shows that although the average MPG for manual transmission is still higher, it is no longer statically significant. In addition, the slope of weight against MPG is significant larger for manual transmission, or slope is -3.758 MPG per 0.001lb from that of automatic transmission. This suggests that weight of cars might have different impacts on MPG as weight increases.

A residual plot is shown in Appendix D. The fitted values against residual does not contain any obvious pattern. However, the Q-Q plot seems to suggest the data might not be normal, which seems to coincide our observations that dependency of MPG against weight is skewed.

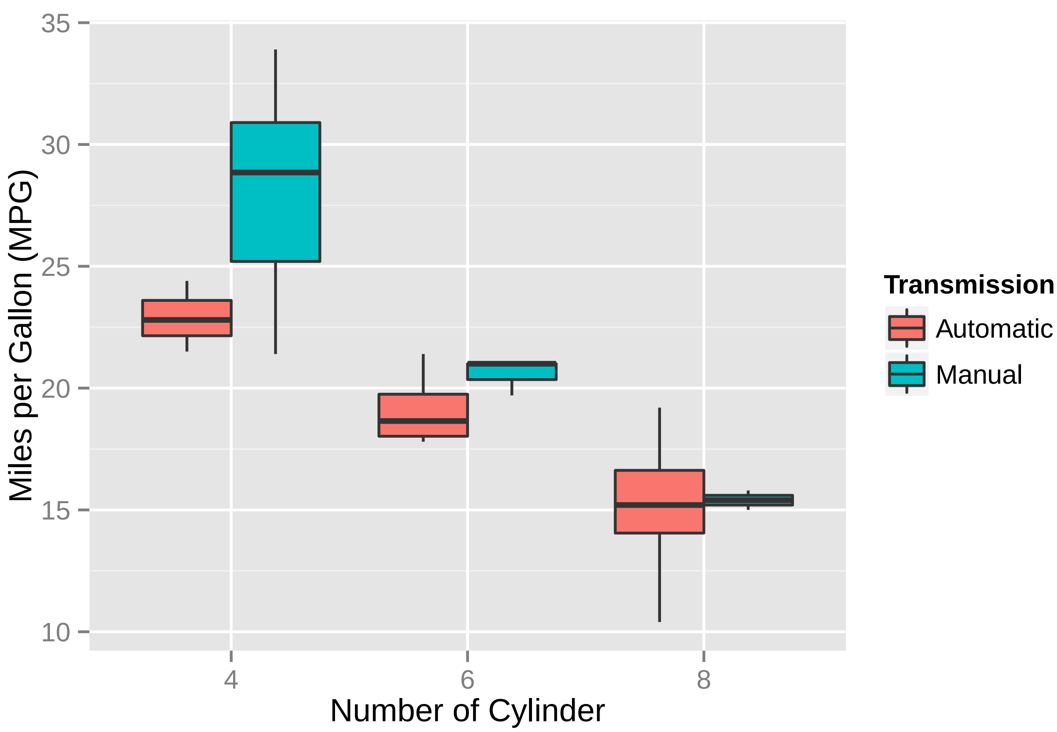
Conclusion

In summary, the dataset does not provide an conclusive answer to whether automatic or manual transmission is better for MPG. This is because the dataset contains more manual transmission cars of small size and lower engine power. To properly answer the question, additional randomized data is required.

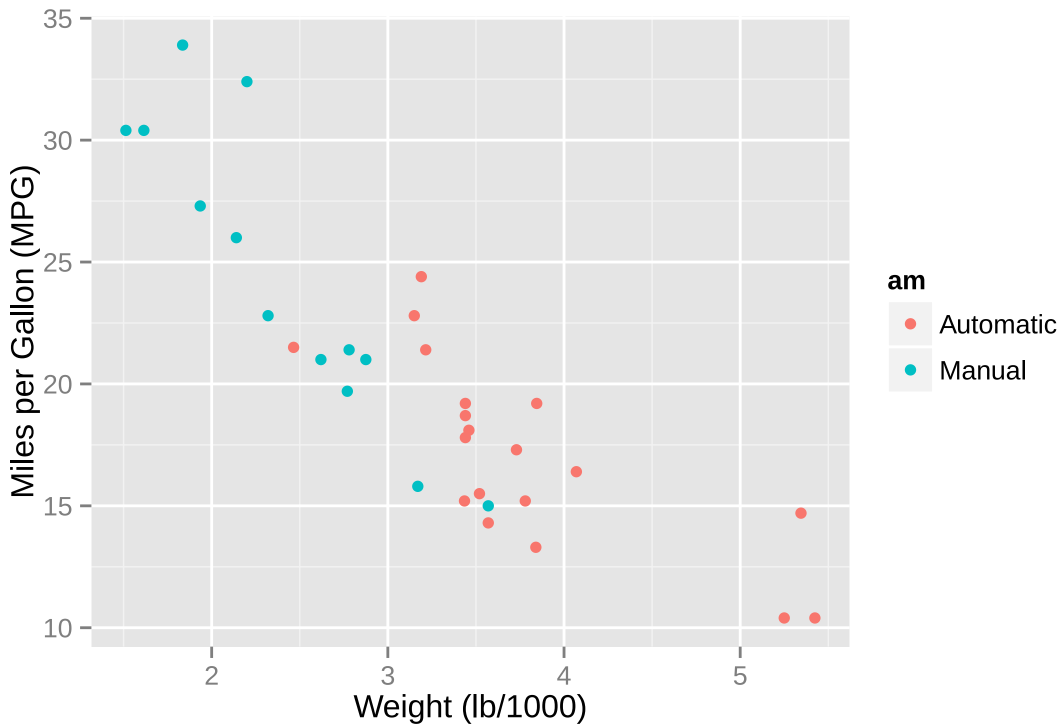
Appendix A - Plot of MPG vs Transmission



Appendix B - Plot of MPG vs Number of Cylinders for Different Transmission



Appendix C - Plot of MPG vs Weight for Different Transmission



Appendix D - Residual Plot of the Proposed Regression Model