# Effect of Transmission Type on MPG of Cars

*August 16, 2016*

## *Executive Summary*

In this report, we are using data collected by Moto Trend Magazine. Looking into the data we are going to analyse and explore the relationship between a set of variables and miles per gallon of automobiles. Basically we are going to address the following issue:

- "Is an automatic or manual transmission better for MPG."
- "How different is the MPG between automatic and manual transmission."

## *Method*

The data set "mtcars" include the 32 observable and 11 variables:

- mpg: Miles per US gallon
- cyl: Number of cylinders
- disp: Displacement(cubic inches)
- hp: Gross Horsepower
- drat: Rear axle ratio
- wt: Weight (lb/1000)
- vs: V/S
- am: Transmission (0 = automatic, 1 = manual)
- carb: Number of carburetors

### Exploratory analysis

Boxplot is produced to see the difference between automatic and manual transimission into the MPG. On seeing figure 1 in Appendix, it is clear that manual transmission produces more MPG. Pairwise graph (figure 2) is plotted to get intuition about the relation between different variables. There is a linear relationship between MPG and each of cyl, disp, hp, drat, wt, qsec, vs, am. The correlation is also compute between every variables and make an idea to set up model. Then a linear model is develop using all variables. Moreover, on seeing the correlation matrix table we found wt, cyl, disp and hp are highly correlated with the dependent variables mpg.However, if we look at the correlation matrix, we also see that cyl and disp are highly correlated with each other. Since predictors should not exhibit collinearity, we should not have cyl and disp in in our model.

### Model

Here we build number of model and analyse using the data set. While looking into the correlation table in appendix, we found higher correlation between hp, am, wt with mpg.

**Model 1:**

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## am           2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

Again fit a model including null hypothesis that transmission is highly affect into MPG.

**Model 2:**

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Since we have two models of the same data, we run an ANOVA to compare the two models and see if they are significantly different.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + wt + hp
```

```
## Model 2: mpg ~ am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     28 180.29
## 2     30 720.90 -2   -540.61 41.979 3.745e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a p-value of 3.745e-09, we reject the null hypothesis and claim that our multivariate model(Model 1) is significantly different from our simple model(Model 2) .

Before we report the details of our model(Model 1), choosen based on seeing p-value, it is important to check the residuals for any signs of non-normality and examine the residuals vs. fitted values plot to spot for any signs of heteroskedasticity.

- From figure 3, the residuals vs fitted plot shows no pattern between the residuals and fitted value indicating that this regression model is well fit.
- The QQ plot shows that the points line up as expected meaning that the disribution is normal and out model predictions are accurate.
- In both the Scale-Location and the Residuals vs Leverage plots, the points are in a group with none too far from the center indicating no point had too much leverage.

## *Conclusions*

On visualising the coefficient and careful study of residual plot, we choose model 1. The transmission type of a car has a significance effect on its fuel efficiency. Model's p-value is less than 0.05 also indicates that this model likely is significant - R-squared as a measure of model's quality, represents a fraction of outcome's variance explained by the model. In this case the model explains 0.8227 (adjusteed value) or 82.27% of variance – Model is significant if any of coeficientts are non-zero. Now when we read the coefficient for am, we say that, on average, manual transmission cars have 2.084 MPGs more than automatic transmission cars.
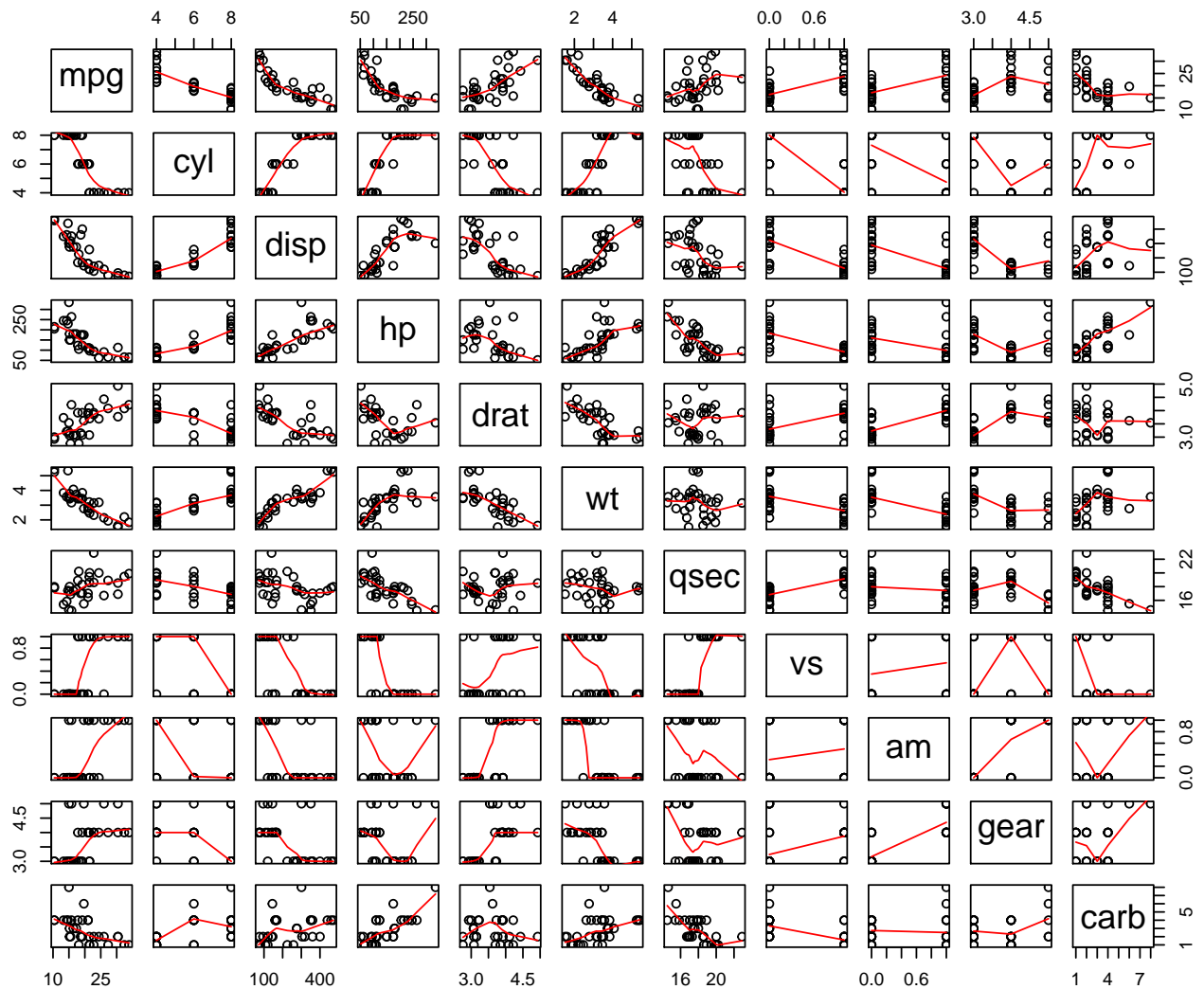
*Appendix*

**Box plot of MPG vs Transmission**



Transmission on (0= automatic, 1= manual)
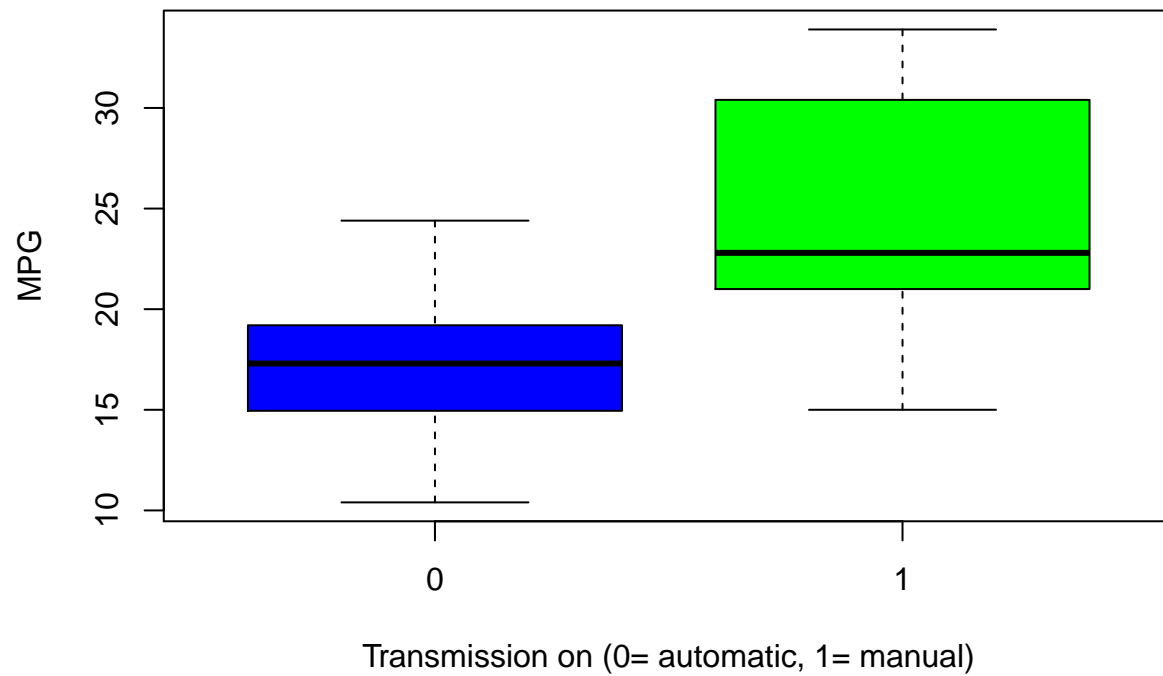
Pairgraph between variables on mtcars data

## Pair Graph on mtcars data set



Correlation of MPG with other variables in the mtcars data set.

```
##          wt         cyl        disp          hp        carb        qsec
## -0.8676594 -0.8521620 -0.8475514 -0.7761684 -0.5509251   0.4186840
##        gear          am          vs        drat         mpg
##  0.4802848   0.5998324   0.6640389   0.6811719   1.0000000
```

## Residuals vs Fitted

## Normal Q–Q

## Scale–Location

## Residuals vs Leverage