# Regression Models and EDA for the mtcars dataset

We will explore the relationship between a set of variables and MPG (outcome) to answer whether an automatic or manual transmission better for MPG and also, quantify the MPG difference between automatic and manual transmissions.

## Executive Summary

From our analysis, we can say that

1. Manual transmission is better for MPG than automtic ransmissions.
2. The cars with manual transmission have MPG 2.93 higher than the cars with automatic transmission.

## Exploratory Data Analysis (EDA)

Plot 1 (see Appendix) clearly shows that the mpg for cars with manual transmissions is better than the car with automatic transmissions.

## Regression Analysis

We will first convert "am" (transmission types: 0 = automatic, 1 = manual) into a categorical variable for analysis -

```
mtcars$am <- factor(mtcars$am)
levels(mtcars$am) <- c("auto","manual")
```

Now, we can perform regression analysis for mpg and transmission -

```
fit <- lm(mpg ~ mtcars$am, data=mtcars)
summary(fit)

##
## Call:
## lm(formula = mpg ~ mtcars$am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       17.147      1.125  15.247 1.13e-15 ***
## mtcars$ammanual    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
```

Through coefficient analysis, we can see that cars with manual transmissions have 7.24 mpg higher value than cars with automatic transmissions. We will now perform correlation analysis between mpg and transmission (am) to quantify their relationship -

```
data(mtcars)
cor(mtcars$am,mtcars$mpg)

## [1] 0.5998324
```

Correlation value 0.5998 between mpg and transmission types indicates a significant positive correlation. As our first regression model explains only 36% of the variance, we will now perform modeling based on all variables in the dataset using Stepwise model (both forward and backward) selections to determine best fit variables -

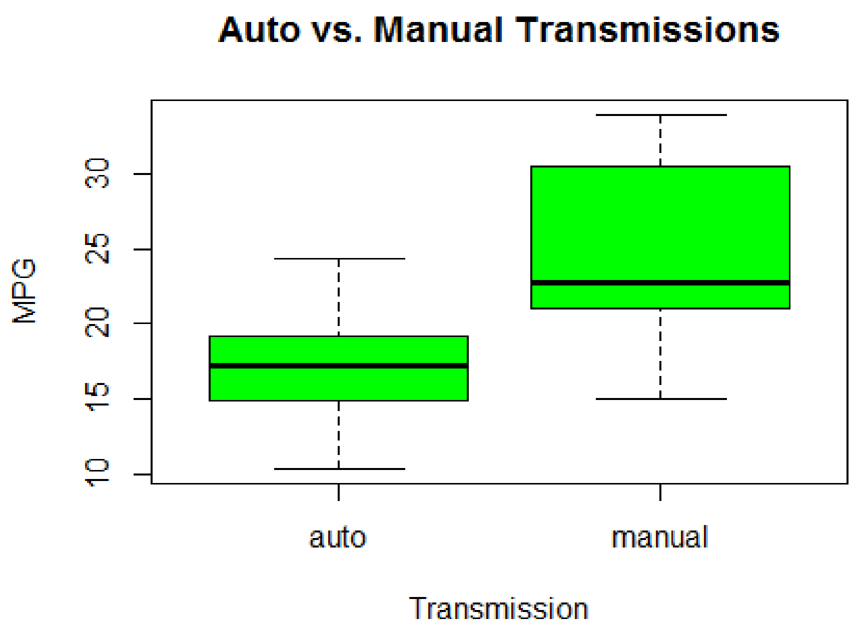
```
fit_all <- step(lm(mpg ~ ., data=mtcars), direction = "both")

summary(fit_all)

##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

This model explains 85% of the variance with predictor variables wt, qsec and am so should be best fit model for regression analysis. Coefficient analysis shows that cars with manual transmission have 2.93 mpg higher than the cars with automatic transmission.

We also performed residual analysis (see Plot 2 in Appendix) to show that no non-normal data pattern implying reasonable fit for a linear model. Residuals vs. Fitted plot shows randomly scattered points above and below the 0 line confirming the normal distribution. Normal Q-Q plot also shows residual points located mostly near the line confirming the residuals are normally distributed.

# Appendix

## Plot 1 - Comparison between Manual and Automatic Transimission




## Plot 2 - Residual Analysis