Motor Trend Analysis

Executive Summary

This report is prepared for Motor Trend magazine and it focus on answering through several perspectives whether an automatic or manual transmission is better for the miles per gallon (MPG) results and how much difference are in terms of MPG between both types of transmissions. The dataset of analysis is called **mtcars**. This study has the main limitation in the number of records.

Several approaches are chosen to perform the analysis since there are multiples variables. The best model correspond to the model calculated with the Stepwise Forward Selection.

The conclusion is that manual transmission is better than automatic transmission. The range of difference depends on the other variables considered, but for the tested models the range is between 1.81 - 7.25. The best model considered the variables: Transmission (am), Number of cylinders (cyl), Weight (wt) and Gross horsepower (hp).

Analysis

Objective and Data Source

The main objective of the current analysis is to respond whether an automatic or manual transmission is better in terms of MPG and quantify the difference among them. The mtcars published by 1974 Motor Trend US magazine is used to perform the analysis. The dataset contains 11 different variables. MPG corresponds to the outcome for this analysis, then there are 5 continuous variables (disp, hp, drat, wt, qsec) and 5 factor or discrete variables (cyl, vs, am, gear, carb).R code is not presented, due to extention restriction, nonetheless it is available in this github

Simple Linear Regression

Exploratory Data

Reviewing graph 1 in Appendix, specifically the am vs mpg, it is highly noted that Manual transmission tends to have better mpg than Automatic transmissions. The numerical value is explored through a simple linear model between (mpg) and transmission (am)

Regression Results

From the results, it is easy to respond that a manual transmission is better for MPG and than the difference is 7.245 mpg. In addition, the model provides a R^2 value of 0.3598. This means that simple linear regression model only explains 35.98% of the variance.

Multiple Regression

In multiple regressions several variables are choosen as estimators of the outcome. For this analysis, three models are going to be tested: 1. With all possible variables, 2. With the highly correlated variables with mpg and finally 3. With the model founded through MASS library to perform a Stepwise Foward Selection of Variables

Exploratory Data

In graph 1 in Appendix is represented the relationship between each variable directly with mpg.

In graph 2 in Appendix is represented the correlation between each variable. Based on this analysis is suggested to take the following variables for the second multiple regression analysis: cyl, disp, hp, wt and since am is our variable of study it would be included although the correlation is less than 0.7. However, there is a high correlation between cyl and disp and considering that predictors should not exhibit collinearity, these two variables would not be included in the model.

Regression Results

```
##                                          Difference**  Adj R-2
## Simple Linear Regression - mpg ~ am          7.244939 33.84589
## All Regresors - mpg ~ .                      1.212116 77.90215
## High Corr Regresors - mpg ~ am + wt + hp     2.083710 82.27357
## SFSV - mpg ~ cyl + hp + wt + am              1.809211 84.00875
```

**Represents the amManual Coefficient in each model, expressed in Miles per Gallon assuming the rest of the variables of the model constant.

Error Analysis

As it is showed in graph 3, all errors from the different models are symmetrically distributed around the middle of the plot suggesting that the linear relationship is reasonable, nonetheless the simple linear regression shows more amplitude in y-axis due to the inaccuracy of the model that just considers one factor variable. There is not outliers since there is not "stands out" points outside of the normal pattern of each model. Generally speaking, the multiple regression models have not clear pattern and its residuals are clustered around smaller digits of y, although the best tendency is shown by the last model.

Reviewing the Q-Q plots, it is easy to observe that all the models tends to follow the line, which implies that the errors follow a normal distribution and this behaviour is desirable.

The model that follows better the line in the Q-Q Plot is the simple regression model, nonetheless when the residuals vs fitted values graph is reviewed, it is easy to note that the variance is greater due to the unaccuracy of this model. Although none of the models represent a perfect match, most of the time a decent model is better than none at all.

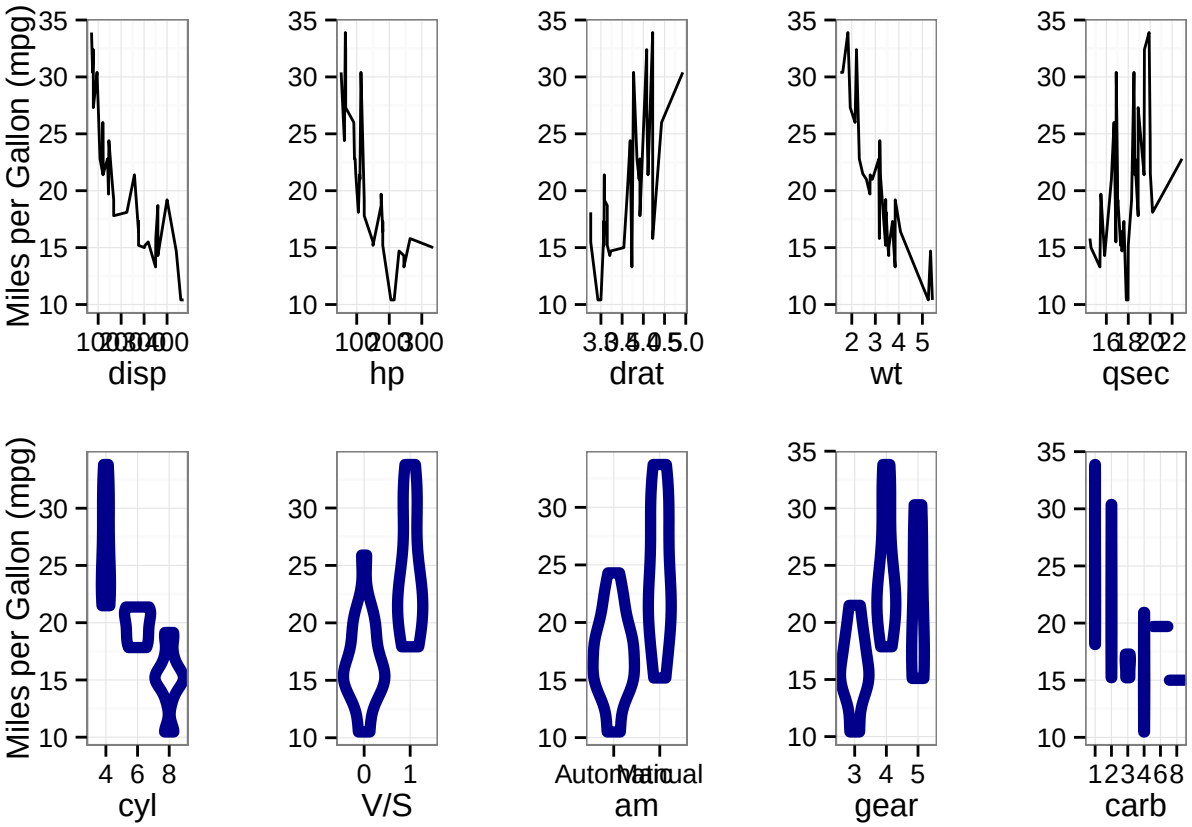
Considering the R square adjusted and the error analysis, it is concluded that the best fit model from these is the Stepwise Foward Selection of Variables

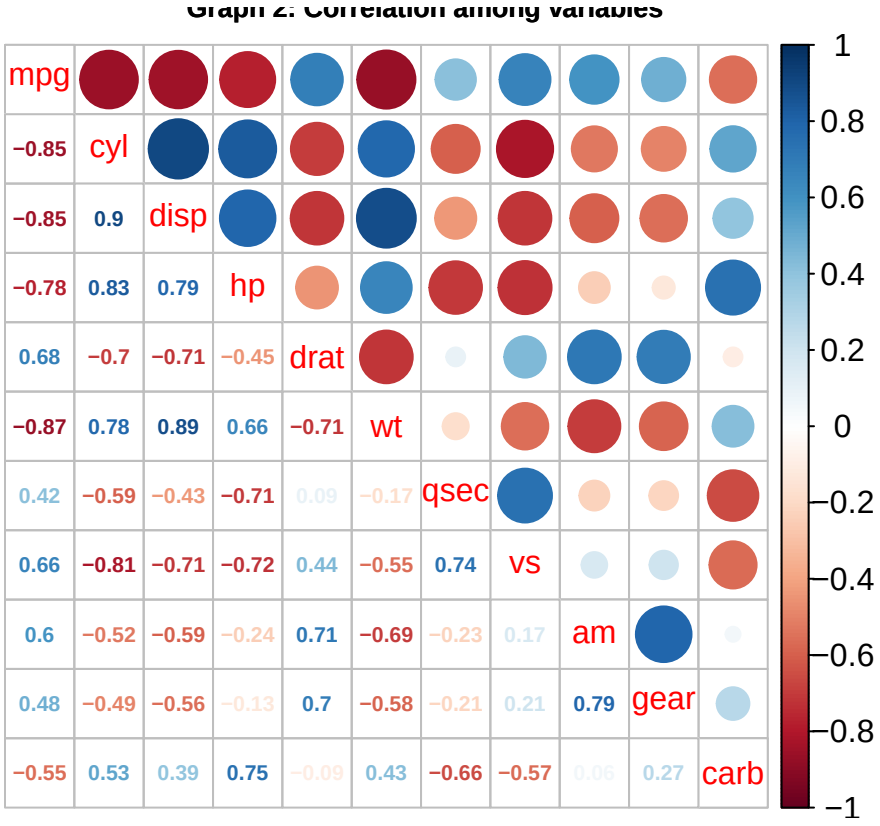
Conclusions

1. Manual transmission is better than automatic transmission in terms of mpg.
2. The best model is the Stepwise Foward Selection of Variable based on R Squared Adjusted.
3. The better model with 84.01% R2 corresponds to mpg = 33.70832 -3.03134 cyl6 - 2.16368 cyl8 - 0.03211 hp -2.49683 wt + 1.80921 am. Maintaining the rest of variables constant the manual transmission is better than automatic transmission by 1.80921 miles per gallon.

Apendix

```
## [1] "Graph 1: Variables of mtcars vs Miles per Gallon (mpg)"
```




mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

Graph 3: Fitted Values vs. Residuals and Normal Q−Q