Cars Consumption Analysis

Pedro A. Alonso Baigorri

1 - Executive Summary

The purpose of this analysis is exploring the relationship between a set of variables and miles per gallon (MPG) (outcome) in the mtcars dataset included in R Studio. On this sense the objective is to answer the following two questions:

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

Regression Modelling techniques are used to perform this analysis.

2 - Exploratory Data Analysis

With the following r Code I can see the main variables of the dataset.

```
data(mtcars)
head(mtcars, 1)
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4  21   6  160 110  3.9 2.62 16.46  0  1    4    4
```

The outcome of this analysis is the variable mtcars$mpg (Miles per Gallon). Higher values of this variable means lower fuel consumption.

In the annex can be found a plot showing that there is a main dependence of the consumption depending on the transmission type. Cars with automatic transmission in general consumes more than manual transmission. Other plots with the analysis of the rest of variables can be found also in the annex.

3 - Model selection

Now I wil fit a regression model to quantify the impact of Transmission in the consumption but linked also with the rest of the variables. First of all I'll try a single model using only am as a regressor since it's thekey variable for this analysis

```
fit0 <- lm(mpg ~ factor(am), data = mtcars)
summary(fit0)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

```
summary(fit0)$r.squared
```

```
## [1] 0.3597989
```

As a result we can see that the P-Values are very small, so it means that the relation between mpg and am is very relevant. However, the Adjusted R-Squared is very low, meaning that this model only explains the 35.98 % of the variance of the dataset.I will create two more models adding more variables grouping them according to the correlation index.

```
cor(mtcars)[-1, 1]
```

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

So for the second model I will use the variables with a corr index > 0.7 (cyl, disp, wt, hp), and for the third model the ones between 0.5 - 0.7 (drat, vs, carb). And then I will use the nested modelling technique to analyze what models to use.

```
fit0 <- lm(mpg ~ factor(am), data = mtcars)
fit1 <- lm(mpg ~ factor(am) + cyl + disp + wt + hp, data = mtcars)
fit2 <- lm(mpg ~ factor(am) + cyl + disp + wt + drat + vs + carb, data = mtcars)
anova(fit0, fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + cyl + disp + wt + hp
## Model 3: mpg ~ factor(am) + cyl + disp + wt + drat + vs + carb
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 163.12  4    557.78 20.222 2.112e-07 ***
## 3     24 165.50  2     -2.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to this the P-Value for the fit1 model is very low so we keep this model, however the fit2 is not necessary to include it. So the statistics for the final model are:

```
summary(fit1)$coefficients
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 38.20279869 3.66909647 10.412045 9.084987e-11
## factor(am)1  1.55649163 1.44053603  1.080495 2.898430e-01
## cyl         -1.10637984 0.67635506 -1.635797 1.139322e-01
## disp         0.01225708 0.01170645  1.047036 3.047194e-01
## wt          -3.30262301 1.13364263 -2.913284 7.256888e-03
## hp          -0.02796002 0.01392172 -2.008374 5.509659e-02
```

And in this case we have a good Adjusted R-Squared: 85.51. The 2nd coef gives the increase in MPG for Manual transmission.

4 - Model Residuals

In the annex can be found a figure with the residuals obtained with this model.According to this figure the residuals are very distributed across all the variability of the prediction.

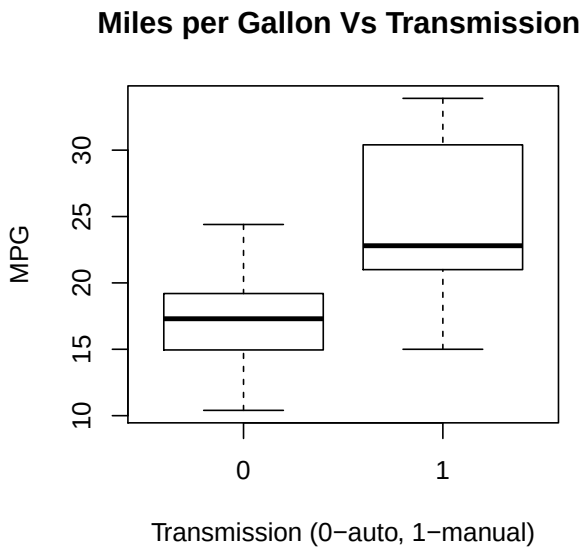
5 - Conclusions

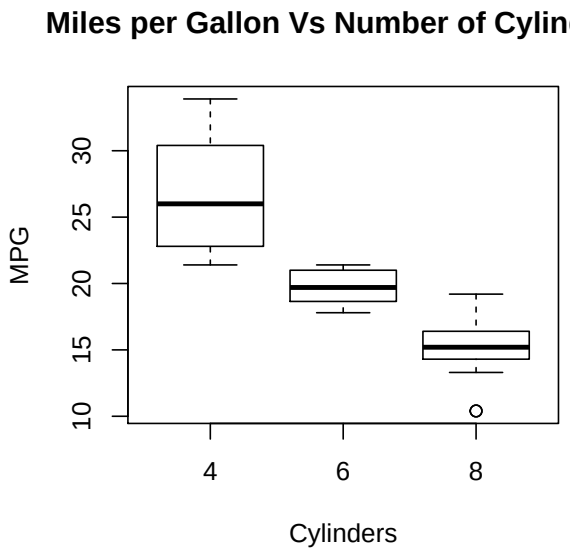
After this analysis we can conclude that:

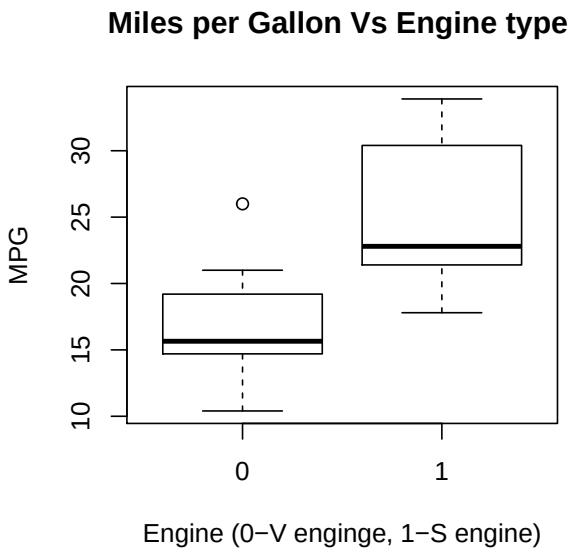
- There is a relationship between Transmission and Consumption: Manual transmission is better for MPG (MPG bigger means fuel consumption higher)
- We can quantify the difference as an estimation: According to our model cars with manual transmission have a 1.5564916 Miles per Gallon more than cars with automatic transmission.

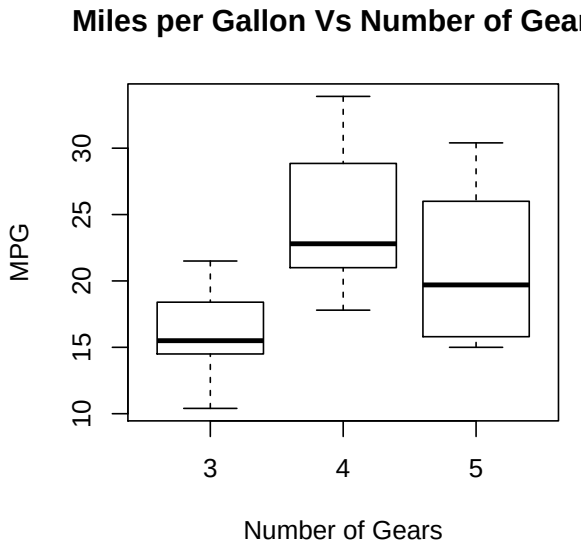
Annexes

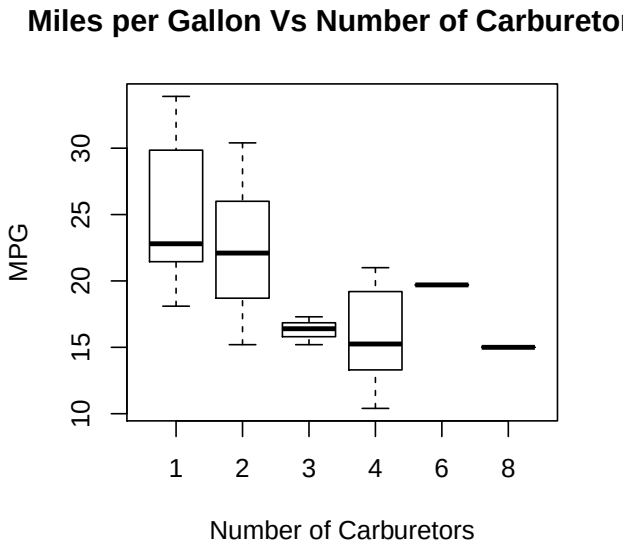
A1 - Figures of data exploration

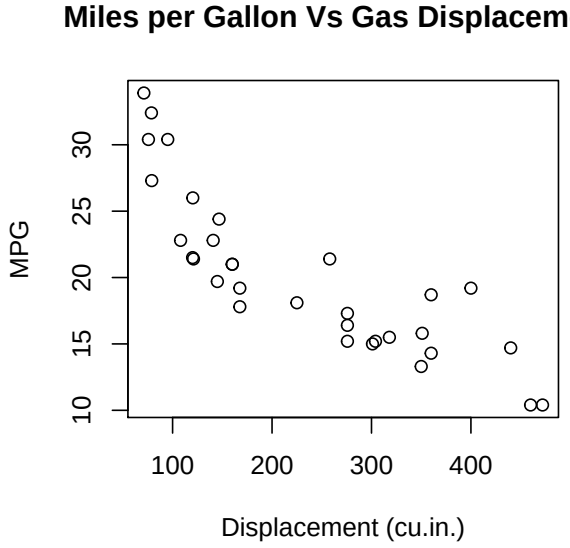


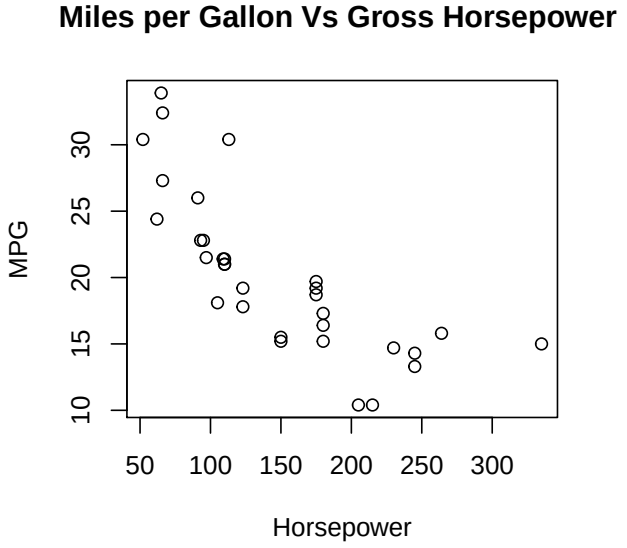


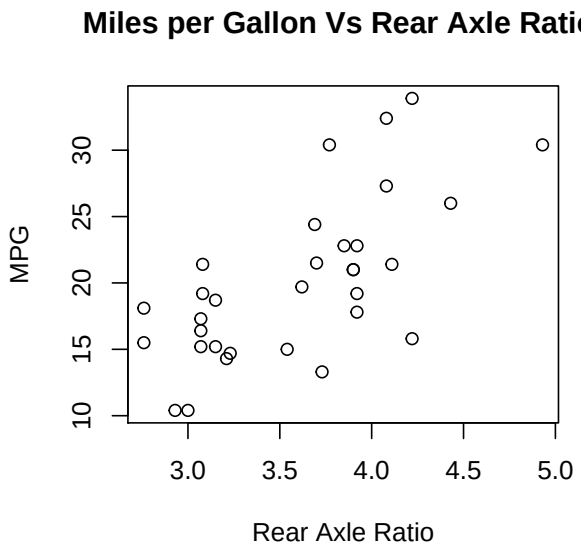


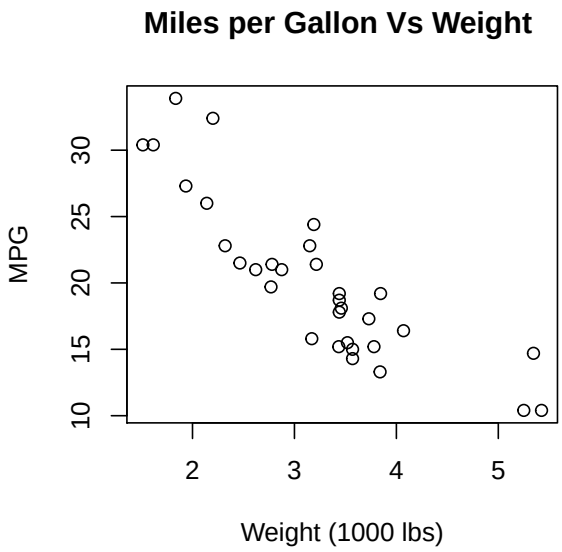

Miles per Gallon Vs Number of Gears
MPG
10
15
20
25
30
3
4
5
Number of Gears


Miles per Gallon Vs Number of Carbureto
MPG
10
15
20
25
30
1
2
3
4
6
8
Number of Carburetors


Miles per Gallon Vs Gas Displacement
MPG
10
15
20
25
30
100
200
300
400
Displacement (cu.in.)


Miles per Gallon Vs Gross Horsepower
MPG
10
15
20
25
30
50
100
150
200
250
300
Horsepower





A2 - Residuals plot