# Factors affecting Miles per Gallon of gas for a car

*MALLESH SUBRAMANI M*

## Overview

The basic ideology used in here is to find the relation between mpg of a car, factors affecting it and to fit a model based on the variables used.

## Basic data analysis

```
dim(mtcars)         #finding the dimension of the data
```

```
## [1] 32 11
```

```
names(mtcars)       #getting the names used for the columns
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

```
head(mtcars)        #peeking at first few data
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
tail(mtcars)        #peeking at last few data
```

```
##                 mpg cyl  disp  hp drat    wt qsec vs am gear carb
## Porsche 914-2  26.0   4 120.3  91 4.43 2.140 16.7  0  1    5    2
## Lotus Europa   30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino   19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora  15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E     21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2
```

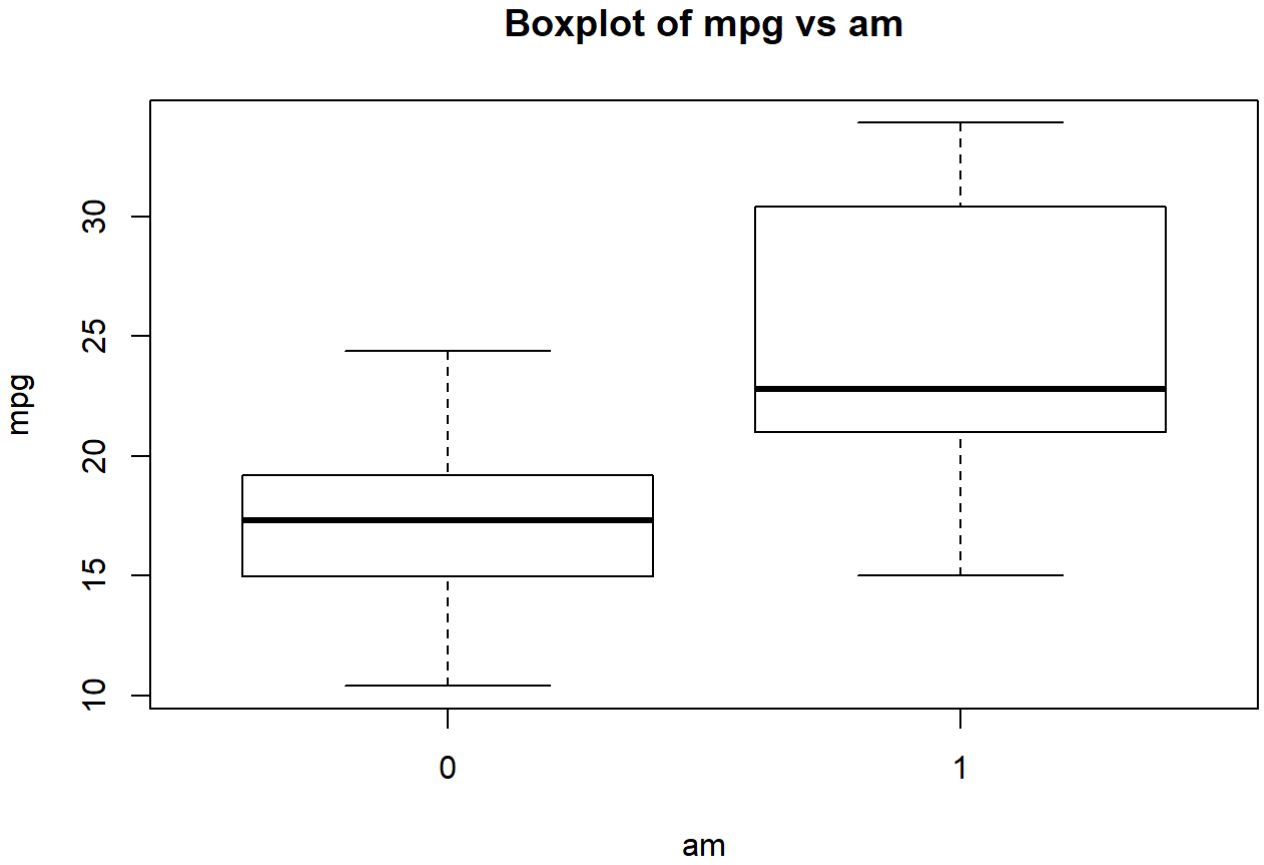
```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

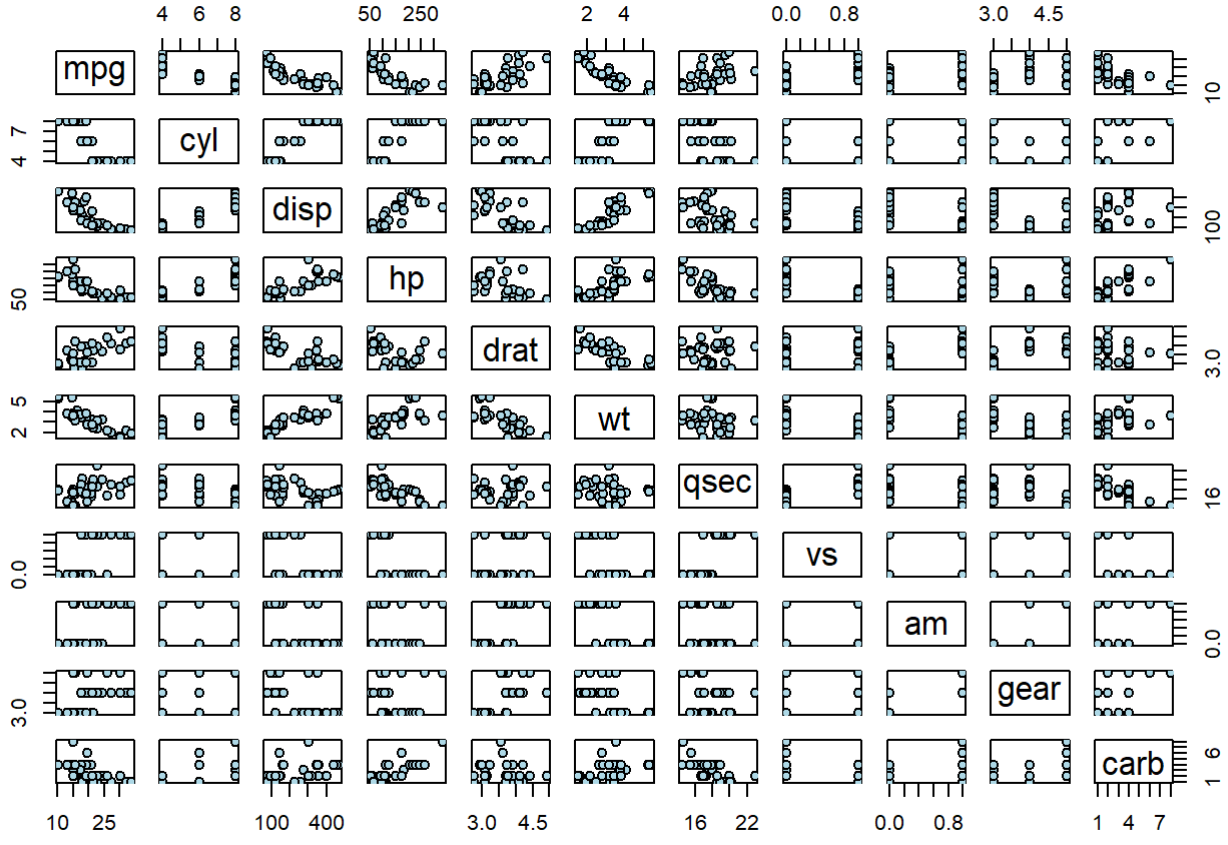
```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
boxplot(mpg~am,data=mtcars,xlab='am',ylab='mpg',main='Boxplot of mpg vs am')
```




```
#plot(mpg~hp+wt+am+wt:am,pch = 21, bg = "lightblue", cex = 2,data=mtcars)
pairs(mtcars,pch = 21, bg = "lightblue", cex = 1)
```



From the basic analysis of data we got to know that it consists of 32 observations and 11 variables recorded.

From the above box plot we can see that cars with automatic transmission has less mpg compared to cars without automatic transmission,that is to say automatic transmission cars consume more fuel than manual transmission cars.

Considering the null hypothesis as type of transmission does not depend on the mpg of the car,considering 95% confidence interval.

Above scatter plot also reveals that mpg not only depends on weight but also on many other variables

```
#Preforming a basic ttest on the mtcars data set
t.test(mpg~am,data=mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

The above ttest reveals that mpg is very much dependent on type of transmission because of the very low p value therefore we will have to reject the null hypothesis.

```
fit1<-lm(mpg~am,data=mtcars)
fit2<-lm(mpg~disp+wt,data=mtcars)
fit3<-lm(mpg~hp+wt+am+wt:am,data=mtcars)
fit4<-lm(mpg~disp+wt+hp+qsec+am,data=mtcars)
fit5<-lm(mpg~disp+wt+hp+qsec+am,data=mtcars)
##fit1
summary(fit1)$r.squared
```

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

```
##fit2
summary(fit2)$r.squared
```

```
## [1] 0.7809306
```

```
##fit3
summary(fit3)$r.squared
```

```
## [1] 0.8695925
```

```
##fit4
summary(fit4)$r.squared
```

```
## [1] 0.8637377
```

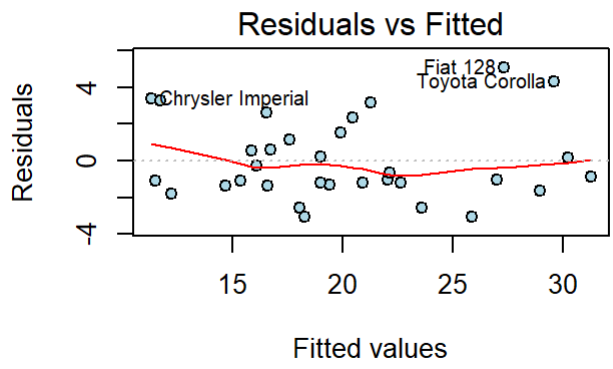
```
##fit5
summary(fit5)$r.squared
```

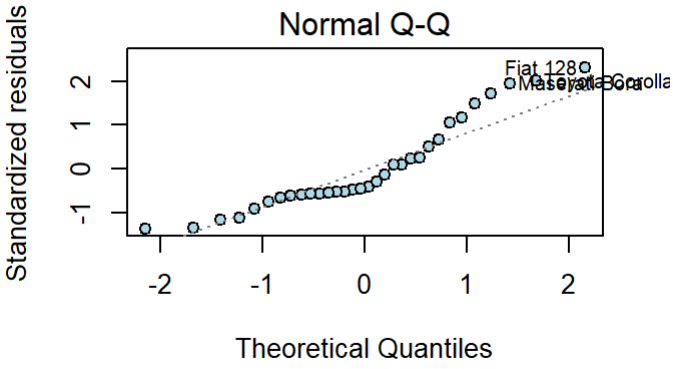
```
## [1] 0.8637377
```

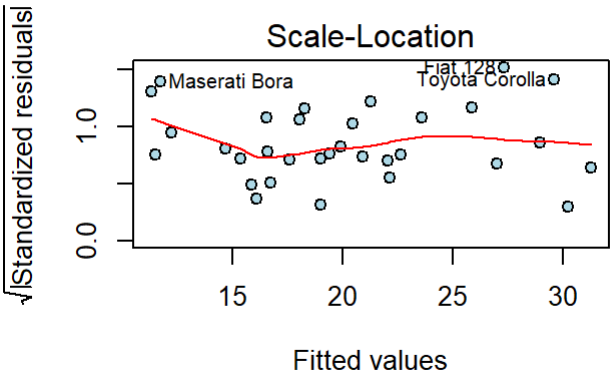
```
anova(fit1,fit2,fit3,fit4,fit5)
```

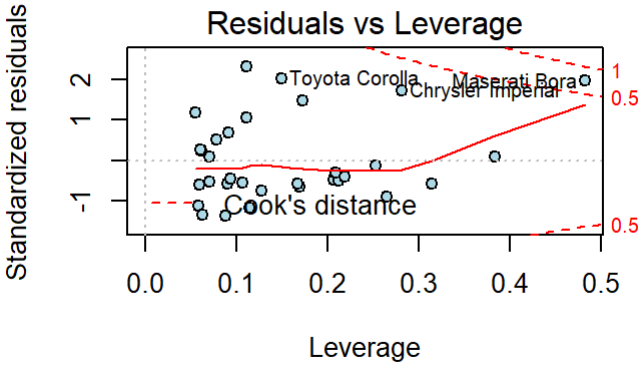
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ disp + wt
## Model 3: mpg ~ hp + wt + am + wt:am
## Model 4: mpg ~ disp + wt + hp + qsec + am
## Model 5: mpg ~ disp + wt + hp + qsec + am
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 246.68  1    474.21 80.3555 1.964e-09 ***
## 3     27 146.85  2     99.84  8.4587  0.001481 **
## 4     26 153.44  1     -6.59
## 5     26 153.44  0      0.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence it can be seen from the different fitted models fit3 has the highest r squared value that is to say it covers more region of values that is also significant from the anova test that residual sum of squares for fit3 is lowest hence we select fit3 for our modeling purpose.




# Conclusion:

-> From the ttest we can say that mpg depends on the kind of transmission and the difference in mpg for automatic and manual transmission is about 7.25 more for manual transmission cars. -> From the plot of fit we can see that residual vs. fitted values lie with a particular range. quantiles vs. residuals plot is also in a given range hence we can conclude that fit3 is the optimized fit for the given data.