# Regression Models Course Project

**Name :** Jose Miguel Arrieta Ramos

**Introduction** In this project is it is supposed i work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

1. “Is an automatic or manual transmission better for MPG”
2. “Quantify the MPG difference between automatic and manual transmissions”

**Load libraries and Data**

```
library(UsingR); data(mtcars)
library(ggplot2)
```

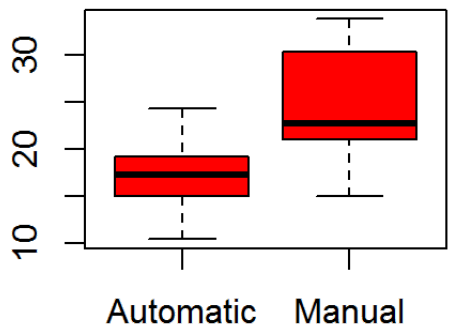
**Cleaning Data** The column named **am** is the column with the Transmission 0 for Automatic and 1 for Manual but its numeric just has 0 and 1 . I replaced with factor variable Automatic and Manual for better understanding.

```
# create factors with value labels
mtcars$am <- factor(mtcars$am,levels=c(0,1),
                    labels=c("Automatic","Manual"))
```

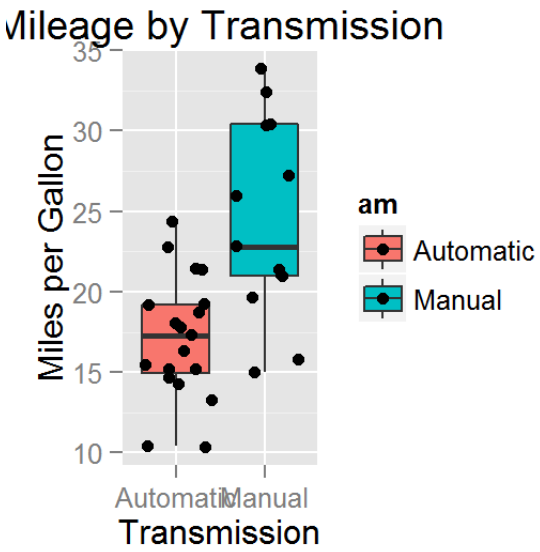
**Exploratory Analysis** The first step was to do a series of exploratories plots to see how the data behave in the two categories Automatic and manual.

```
#Mulpiple Boxplots

boxplot(mpg~am,data=mtcars,col="red")
```

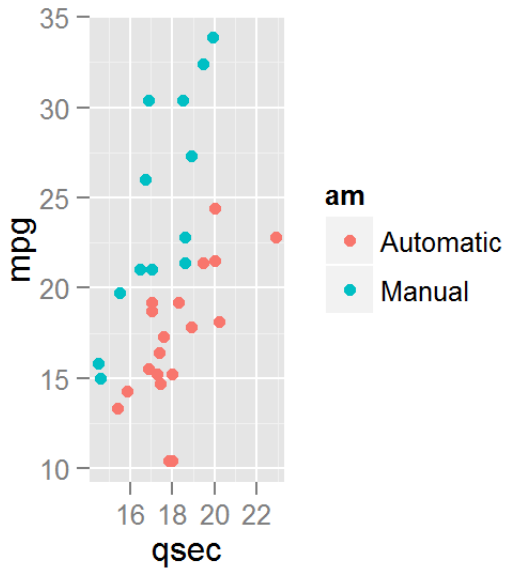


```
# Boxplots of mpg by am
# observations (points) are overlayed and jittered
qplot(am, mpg, data=mtcars, geom=c("boxplot", "jitter"),
      fill=am, main="Mileage by Transmission",
      xlab="Transmission", ylab="Miles per Gallon")
```




As you can see in the graphics, the mean MPG of automatic cars is lower than the mean of manual cars. The data in Manual cars is more disperse than automatic cars.

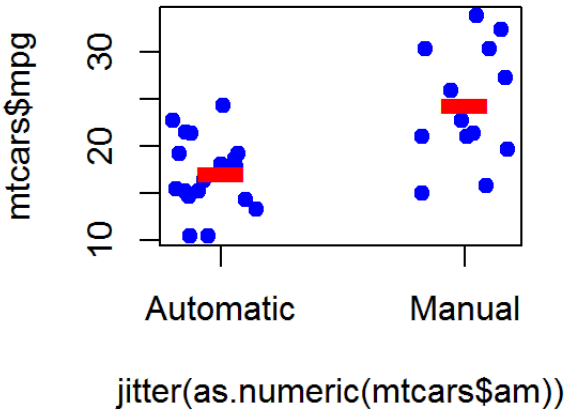
```
#ggplo2
qplot(qsec,mpg,data=mtcars,color=am)
```




```
#MPG vs Transmission

plot(mtcars$mpg~jitter(as.numeric(mtcars$am)),col="blue",xaxt="n",pch=19)
axis(side=1,at=unique(as.numeric(mtcars$am)),labels=unique(mtcars$am))

meanMPG<-tapply(mtcars$mpg,mtcars$am,mean)
points(1:2,meanMPG,col="red",pch="-",cex=5)
```




***Question 1:*** It seems than Automatic is better than Manual for mpg

**Methods** For this case is going to be used a regression models with the predictors as factor and the outcome as numeric value.

Yi=b0+b1(Ti="Manual")

*(Ti="Manual" )is a logic value that is one if the Transmission is "Manual" and zero otherwise.

```
lm1<-lm(mtcars$mpg~as.factor(mtcars$am))
summary(lm1)
```

```
##
## Call:
## lm(formula = mtcars$mpg ~ as.factor(mtcars$am))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    17.15       1.12   15.25  1.1e-15 ***
## as.factor(mtcars$am)Manual      7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

**COnfident intervals** This are the values with 95% certainty that the intercept and slope would be .

```
confint(lm1,level=0.95)
```

```
##                              2.5 % 97.5 %
## (Intercept)                 14.851  19.44
## as.factor(mtcars$am)Manual   3.642  10.85
```

***Question 2 :*** “Quantify the MPG difference between automatic and manual transmissions”

b0= ->is the average of the Automatics car .

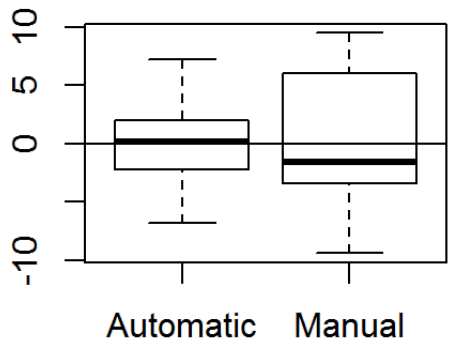
b0+b1= -> is the average of Manual cars

Dif Automatic - Manual ->b0-(b0+b1)=-b1 ->-7.24

**Results**

**Residual analysis**

```
plot(as.factor(mtcars$am), resid(lm1));
abline(h = 0)
```




Anova

```
anova(lm1)
```

```
## Analysis of Variance Table
##
## Response: mtcars$mpg
##                      Df Sum Sq Mean Sq F value  Pr(>F)
## as.factor(mtcars$am)  1    405     405    16.9 0.00029 ***
## Residuals            30    721      24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The R-squared is kind of low , in a future work maybe the anylysys should include more predictors.