

# Regression Models Course Project

*Sheetal*

19 October 2018

## Executive Summary'

The purpose of this report is the answer 2 questions : 1.“Is an automatic or manual transmission better for MPG” 2.“Quantify the MPG difference between automatic and manual transmission

Information on the datasets is found in the R help guides. There are 10 variables : 1 mpg: Miles/(US) gallon 2 cyl: Number of cylinders 3 disp: Displacement (cu.in.) 4 hp: Gross horsepower 5 drat: Rear axle ratio 6 wt: Weight (1000 lbs) 7 qsec: 1/4 mile time 8 vs: V/S 9 am: Transmission (0 = automatic, 1 = manual) 10 gear: Number of forward gears 11 carb: Number of carburetors

## Exploratory Data Analysis

The below sections performs some preliminary exploration of the dataset

```
mtcars <- mtcars  
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
```

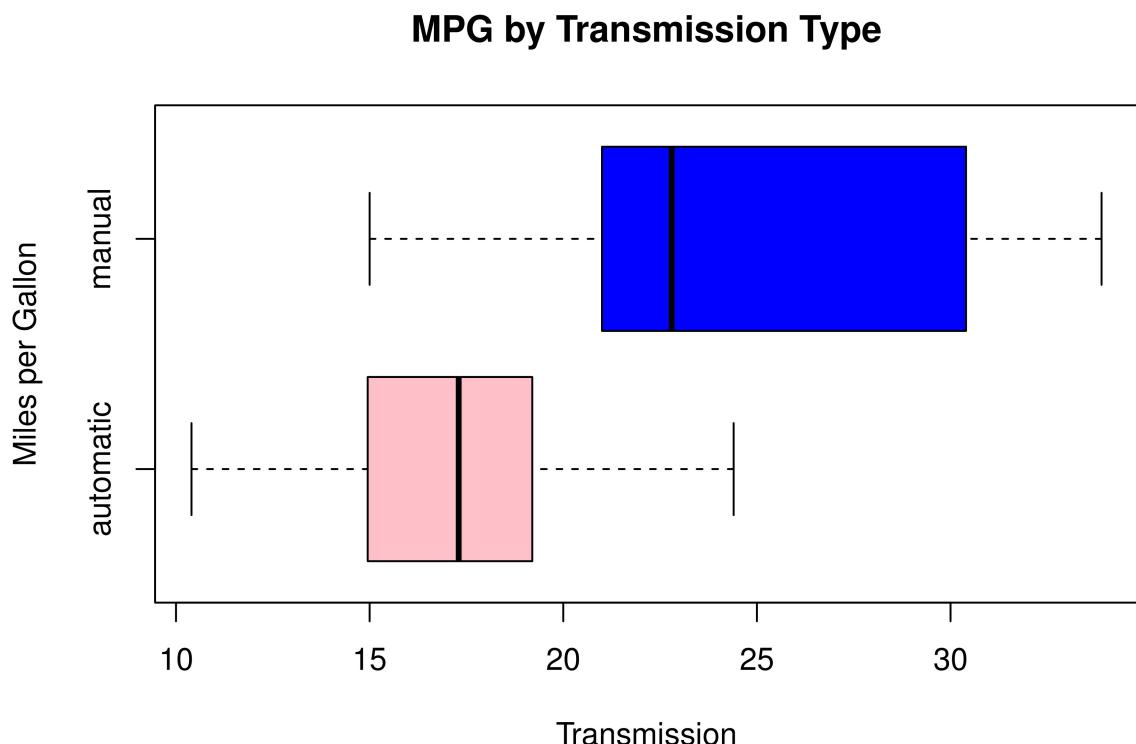
There are 32 observations.

The summary statistics shows that there is 1 dependent variable and 10 explanatory variables. VS and AM are factor variables so they need to be converted to factors.

```
mtcars$vs <- as.factor(mtcars$vs)  
mtcars$am <- as.factor(mtcars$am)
```

## 1.“Is an automatic or manual transmission better for MPG”

```
boxplot(mpg ~ am, data = mtcars,
        col = c("pink", "blue"),
        xlab = "Transmission",
        ylab = "Miles per Gallon",
        main = "MPG by Transmission Type",
        names= c("automatic","manual"),
        horizontal= T)
```



From the box plot it appears that the manual transmission has greater miles per gallon on average.

This can be double checked via hypothesis testing

H<sub>0</sub> : The transmission type has no impact on miles per gallon  
H<sub>1</sub> : The transmission type impacts mile per gallon

```
auto=subset(mtcars,select=mpg,am==0)
manual=subset(mtcars,select=mpg,am==1)
t.test(auto,manual)

##
##  Welch Two Sample t-test
##
## data: auto and manual
## 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 of x mean of y
## 17.14737 24.39231

```

As the p-value is small we can reject the null hypothesis.

## 2.“Quantify the MPG difference between automatic and manual transmission

### Linear Regressions

```

simpleregression <- lm(mpg~am,mtcars)
summary(simpleregression)

##
## Call:
## lm(formula = mpg ~ 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 ***
## am1          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

```

Doing a simple linear regression between miles per gallon (MPG) and transmission, shows that manual is better by 7.245 miles per gallon. The R squared is only 36%, so there is more variance to explain,

The next step is multiple regression

```

multipleregression <- lm(mpg~.,mtcars)
summary(multipleregression)

##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##    Min     1Q   Median     3Q    Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337  18.71788  0.657  0.5181
## cyl         -0.11144   1.04502 -0.107  0.9161

```

```

## disp      0.01334  0.01786  0.747  0.4635
## hp       -0.02148  0.02177 -0.987  0.3350
## drat      0.78711  1.63537  0.481  0.6353
## wt       -3.71530  1.89441 -1.961  0.0633 .
## qsec      0.82104  0.73084  1.123  0.2739
## vs1       0.31776  2.10451  0.151  0.8814
## am1       2.52023  2.05665  1.225  0.2340
## gear      0.65541  1.49326  0.439  0.6652
## carb     -0.19942  0.82875 -0.241  0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869, Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07

```

The R squared has now improved to 86.9%. It now looks like manual is better by 2.52 miles per gallon. The p values have all gone up, and now none of the coefficients are significant.

I am now going to try step wise regression

```

regSR=step(multipleregression,trace=0)
summary(regSR)

##
## 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 ***
## am1          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

```

The R squared is 85%, and now we are left with significant coefficients : wt: Weight (1000 lbs) qsec: 1/4 mile time Transmission (0 = automatic, 1 = manual)

We can now conclude that manual transmission allows 2.9358 miles than automatic transmission.

```
plot(regSR, which=c(1:1))
```

