---

title: "Regression model"

output: html_document: df_print: paged pdf_document: default

Executive Summary

In the following we work with the mtcars dataset and determine the relationship between the variables within & which of automatic or manual transmission is better for mpg(miles per gallon) which we treat as response .To summarise the findings, we note that manual transmissions on average do give 5.97 miles per gallon more than automatic transmission; however this is taking into account the confounding variables of weight and cylinders.

Performing exploratory data analysis.

Loading dependencies

```
library(datasets)
library(ggplot2)
library(dplyr)
```

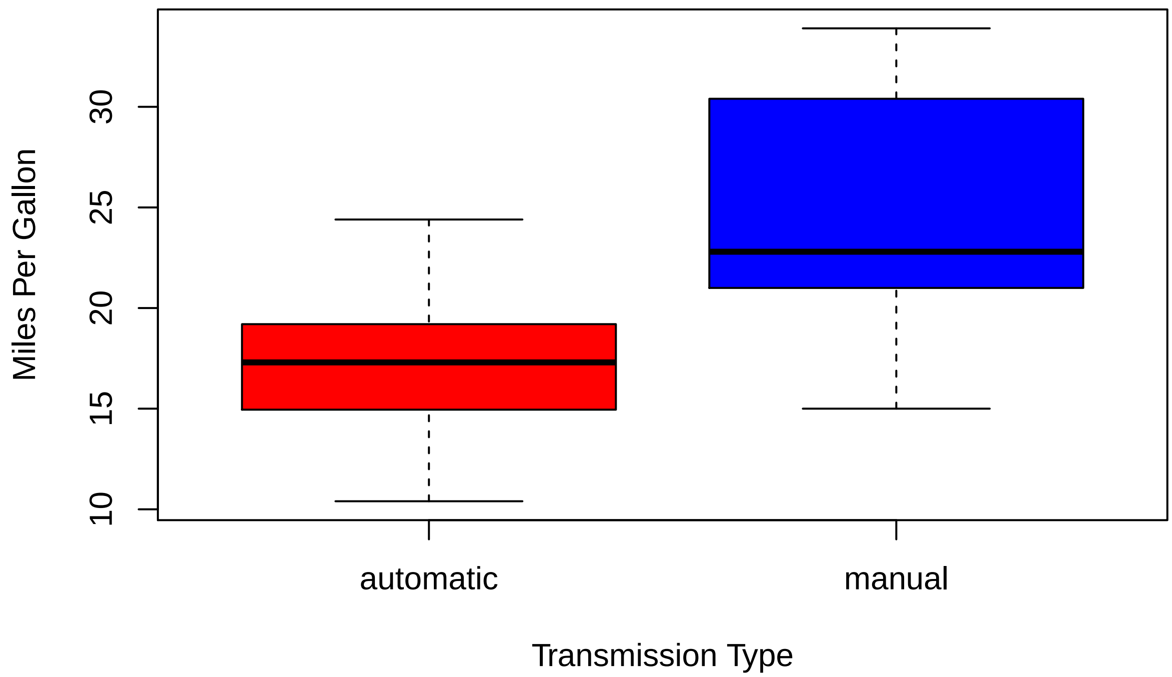
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
head(mtcars ,2)
```

```
##               mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4      21   6  160 110  3.9 2.620 16.46  0  1    4    4
## Mazda RX4 Wag  21   6  160 110  3.9 2.875 17.02  0  1    4    4
```

Making a new object called "data" such that data = mtcars Adding a new variable Transmission .

```
mtcars$am<-as.factor(mtcars$am)
data<-mtcars
data<-data%>%mutate(transmission = ifelse(data$am==1,"manual","automatic"))
boxplot(mpg ~ transmission, data, col = (c("red","blue")), ylab = "Miles Per Gallon", xlab = "Transmiss
```



Finding the means of mpg differentiated by transmission

```
auto<-filter(data,data$transmission=="automatic")
manual<-filter(data,data$transmission=="manual")
mean(auto$mpg);mean(manual$mpg)
```

```
## [1] 17.14737
```

```
## [1] 24.39231
```

It shows that on average Manual yields an average of 7 mpg more than automatic.

Hypothesis Testing and Linear Regression Model

Let the null hypothesis be that there is no difference between the mpg outcome of cars having either automatic or manual transmissions. The Alternate Hypothesis shall be that there is a difference in the values & that mean(mpg;automatic) is not equal to mean(mpg;manual) First lets create a linear model where the response mpg is solely decided by the factor variable transmission .

```
t.test(formula = 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
```

We see that the p-value is less than 0. 05 and hence we reject the null hypothesis and adopt the alternative that there is a difference between mpg for manual and automatic transmissions .

```
fit<-lm(mpg~transmission,data )
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ transmission, data = data)
##
## 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 ***
## transmissionmanual    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
```

However by the Rsquare value we see that the current model only explains approx. 36 % of the variation. Hence we have to adopt a multi-variable regression model .

```
stepModel <- step(fit, k=log(nrow(data)))
summary(stepModel)
#mpg ~ wt + qsec + am
#results hidden to make the document smaller
```

This model is "mpg ~ wt + qsec + am". It has the Residual standard error as 2.459 on 28 degrees of freedom. And the Adjusted R-squared value is 0.8336, which means that the model can explain about 83% of the variance of the MPG variable. All of the coefficients are significant at 0.05 significant level.

Please refer to the Appendix:Plots section for the plots again. According to the scatter plot, it indicates that there appear to be an interaction term between "wt" variable and "am" variable, since automatic cars tend to weigh heavier than manual cars. Thus, we have the following model including the interaction term:mpg~wt + am + qsec + wt :am

```
final<-lm(mpg~wt + am + qsec + wt :am,data)
summary(final)$coef
```

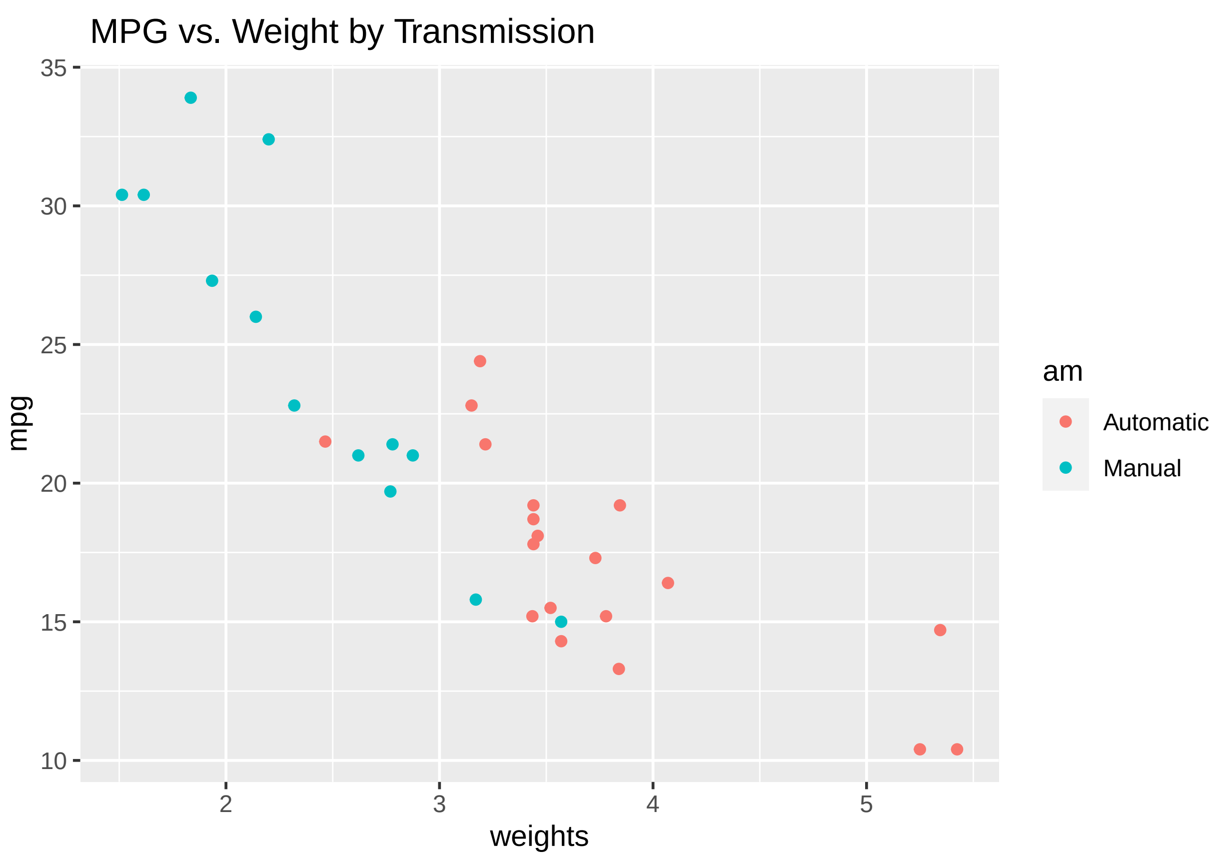
```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.723053  5.8990407  1.648243 0.1108925394
## wt          -2.936531  0.6660253 -4.409038 0.0001488947
## am1         14.079428  3.4352512  4.098515 0.0003408693
## qsec         1.016974  0.2520152  4.035366 0.0004030165
## wt:am1      -4.141376  1.1968119 -3.460340 0.0018085763
```

Conclusion:Thus ,with all others values fixed , we determine that manual cars give 5.97 MPG more than cars with automatic transmissions.

Appendix : Plots

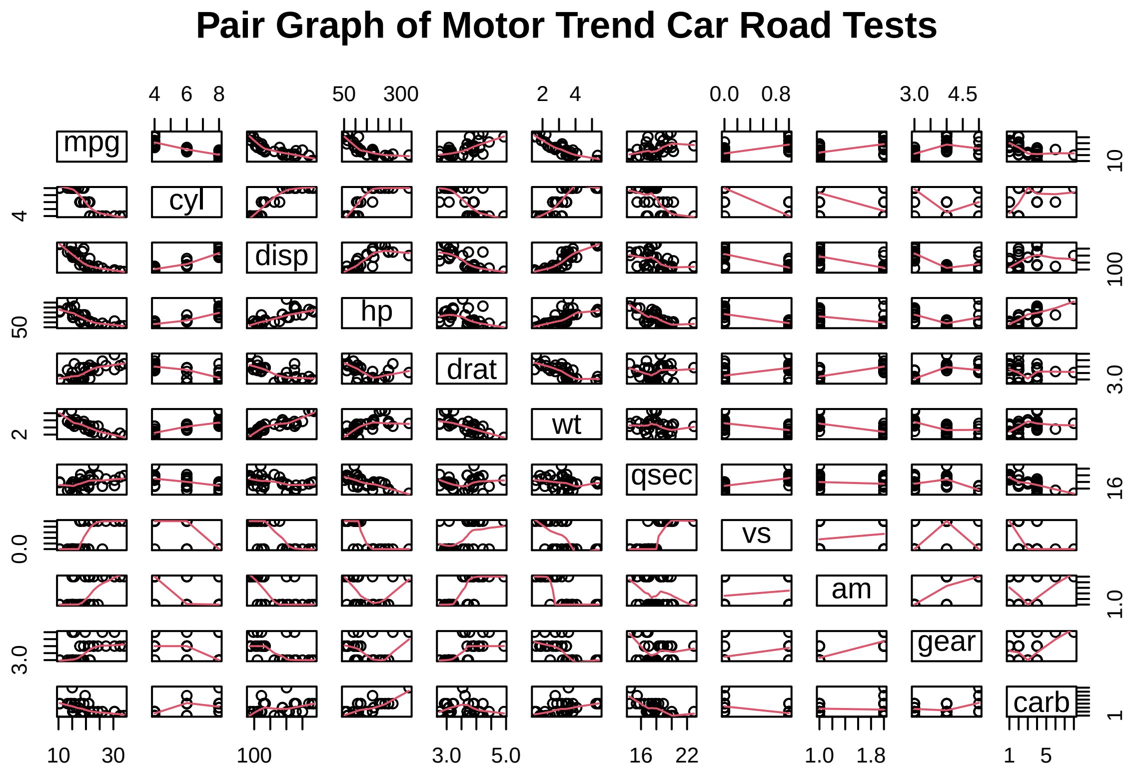
A : Scatterplot displaying the trend of MPG for auto and manual with weight .

```
ggplot(mtcars, aes(x=wt, y=mpg, group=, color=am, height=3, width=3)) + geom_point() +
scale_colour_discrete(labels=c("Automatic", "Manual")) +
xlab("weights") + ggtitle(" MPG vs. Weight by Transmission")
```



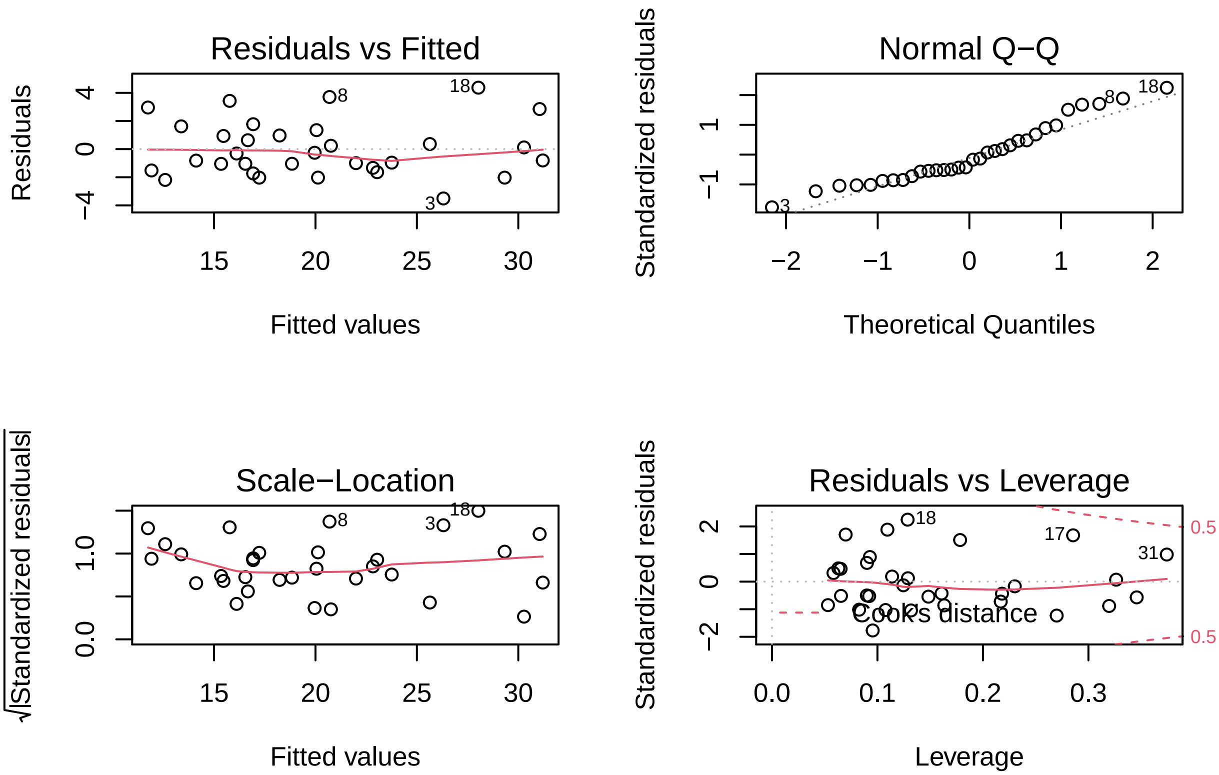
B :

```r
pairs(mtcars, panel=panel.smooth, main="Pair Graph of Motor Trend Car Road Tests")
```



C:

```r
 par(mfrow = c(2, 2))
plot(final)
```



Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standard residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage