Regression Model Peer Graded Assignment

Executive Summary

Motor Trend, a magazine about the automobile industry, 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:

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

Reading and analysing car data

```
library(datasets)
data(mtcars)
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 ...
```

The variable we are interested is ,am, of numeric class into a factor class and also we can level it into Automatic and Manual for better exploratory analysis and making better model.

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

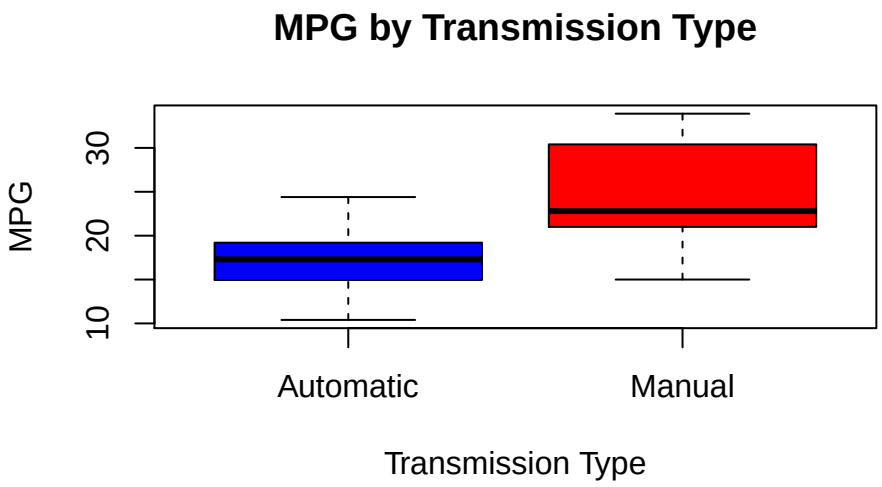
Exploraoty Data Analysis

```
## Warning: package 'ggplot2' was built under R version 3.3.1
```

For properly fitting the linear model we have to find out all the predictor variables for determining the outcome, now we donot want to omit any predictor variable in our model neither do we want to add any correlated variable.

A box plot is created to see the variation of Transmission on MPG and it seems that manual transmission has better mpg than automatic transmission.

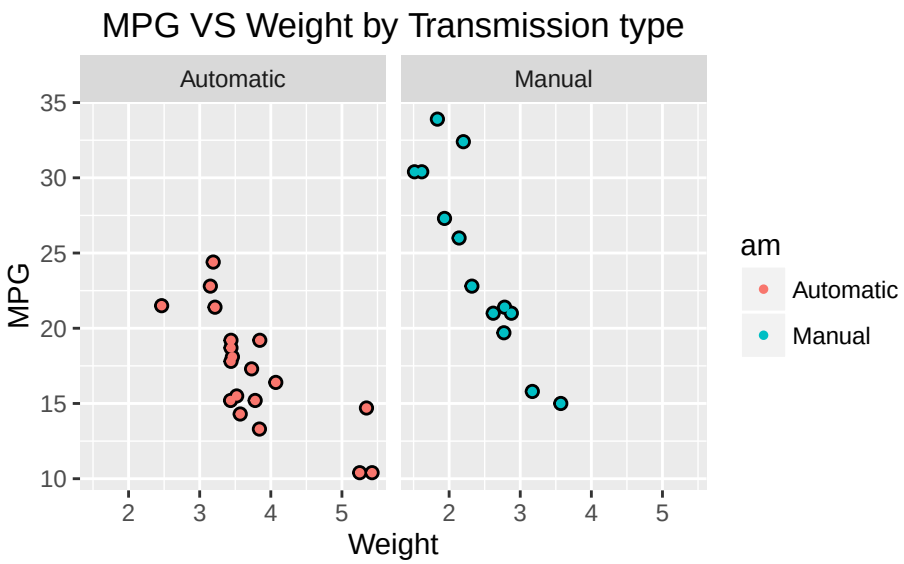
```
boxplot(mpg~am, data = mtcars,
        col = c("blue", "red"),xlab = "Transmission Type",ylab = "MPG",
        main = "MPG by Transmission Type")
```



Their might be serveral other variables that must be causing the increament of MPG with manual tansmission one of it can be weight.

We have plotted a variation of MPG with respect to weight for different transmission types.

```
ggplot(data=mtcars,aes(x=wt,y=mpg, color=am))+facet_wrap(~am)+xlab("Weight")+
  ylab("MPG")+ggtitle("MPG VS Weight by Transmission type")+
  geom_point(size=2,colour="black")+geom_point(size=1)
```



So, we can clearly see that wight has indirect relationship with MPG and manual trasmisison has less wieght than automatic trasmission, so might be because of this Mpg was showing better fuel efficiency with manual transmisison as compared to Automatic transmission.

Correlation test

It is used to indentify whether other variable that have any dependency on MPG or not.

```
data("mtcars")
sort(cor(mtcars)[1,])
```

```
##         wt        cyl       disp         hp       carb       qsec
## -0.8676594 -0.8521620 -0.8475514 -0.7761684 -0.5509251  0.4186840
##       gear         am         vs       drat        mpg
##  0.4802848  0.5998324  0.6640389  0.6811719  1.0000000
```

From the correlation data, we could see cyl, hp, wt and carb are negatively correlated with mpg. In addition to am (which by default must be included in our regression model), we see that wt, cyl, disp, and hp are highly correlated with our dependent variable mpg.

Modeling

Simple linear model

```
fit <- lm(mpg~am, data = mtcars)
summary(fit)
```

```
##
## 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 ***
## am             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
```

We do not gain much more information from our hypothesis test using this model. Interpreting the coefficient and intercepts, we say that, on average, manual transmission cars have 7.245 MPGs more than automatic transmission. In addition, we see that the R^2 value is 0.3598. This means that our model only explains 35.98% of the variance.

We need to understand the impact of transmission in conjunction with other factors to quantify the mpg difference between automatic and manual transmission.

Multivariate Modeling

For multivariate we have choose the predicatable variable that are least correlated to each other.

```
bestfit<-lm(mpg~factor(am)+qsec+wt, data=mtcars)
summary(bestfit)
```

```
##
```

```
## Call:
## lm(formula = mpg ~ factor(am) + qsec + wt, 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
## factor(am)1   2.9358     1.4109   2.081 0.046716 *
## qsec          1.2259     0.2887   4.247 0.000216 ***
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## ---
## 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
```

This shows that in adddition to transmission, weight of the vehicle as well as accelaration speed have the highest relation to explaining the variation in mpg. The adjusted R^2 is 84% which means that the model explains 84% of the variation in mpg indicating it is a robust and highly predictive model.

Conclusion

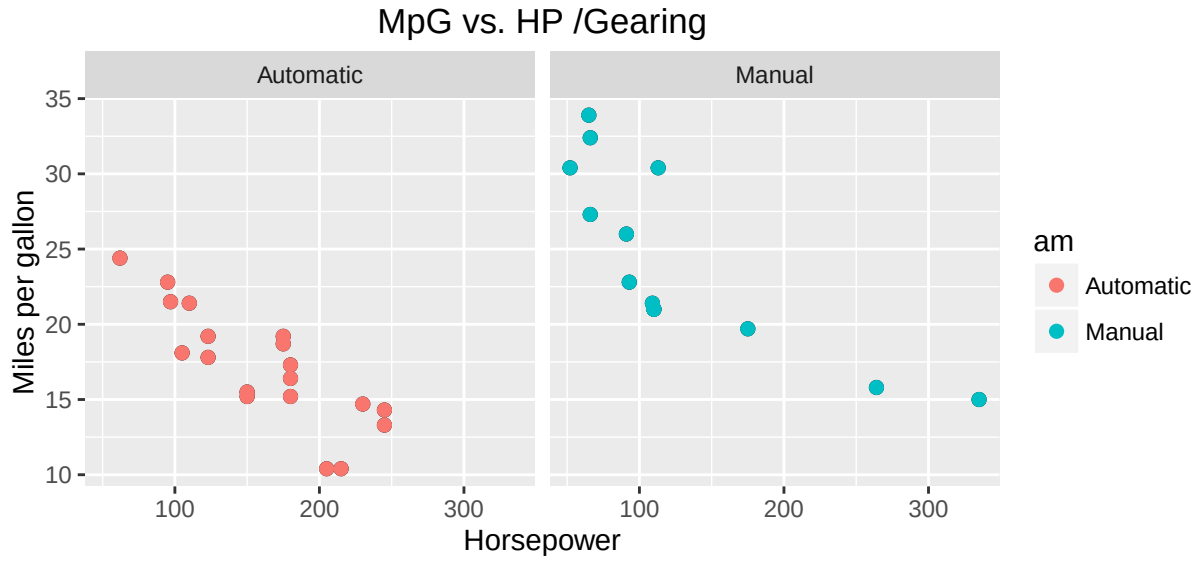
- Manual transmission has higher MPG i.e. higher fuel efficiency as compared to automatic transmisison.We say that, on average, manual transmission cars have 2.94 MPGs more than automatic transmission cars.
- MPG has indirect relationship with weight, as weight increases per 1000lb (0.5 tons), the mpg will decrease by 2.5.

So, the above analysis answers the question of auto or manual transmission but we have to keep in mind other factors too, as it is not a direct factor. So, you can't answer it.

Appendix

Variable testing

```
ggplot(data=mtcars,aes(x=hp, y=mpg, col=am))+facet_wrap(~am)+
  xlab("Horsepower")+ylab("Miles per gallon")+ ggtitle("MpG vs. HP /Gearing")+
    geom_point(size=2,colour="black")+geom_point(size=2)
```



Residual plot for best fit model

```
par(mfrow = c(2,2))
plot(bestfit)
```