Regression Models Course Project

Summary

In this analysis we wish to study the effect of transmission on miles per gallon.We conducted t test which alludes that there is significant difference in mpg for automatic and manual tranmission.

The fitted model suggest that estimated mpg for manual transmission is 2.9358 more than the estimated mpg for automatic transmission,provided the other variables are kept constant.

Data processing

```
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
```

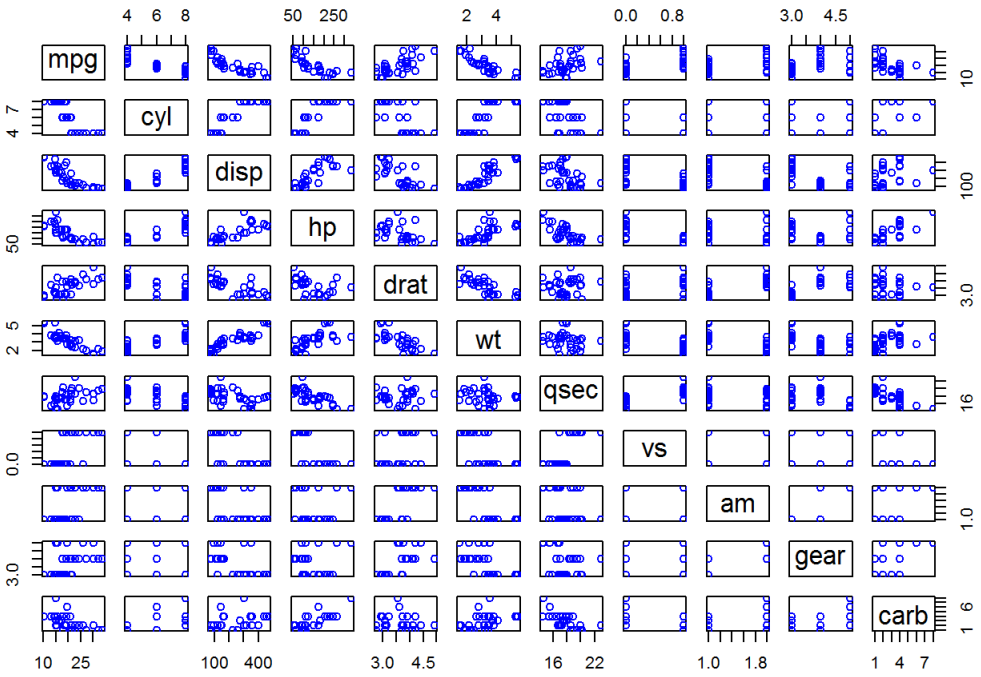
```
library(ggplot2)
data(mtcars)
mtcars$am<-as.factor(mtcars$am)
levels(mtcars$am)=c("at","mt")
```

The mtcars dataset has been loaded to be analysed.The column am(transmission) has been tranformed into factor with levels at(automatic) and mt(manual).

Exploratory Data Analysis.

In this analysis we specifically wish to answer the question-Is an automatic or manual transmission better for MPG?.Hence our exploratory analysis will revolve aroung mpg and am(transmission).

```
pairs(mtcars,col="blue")
```

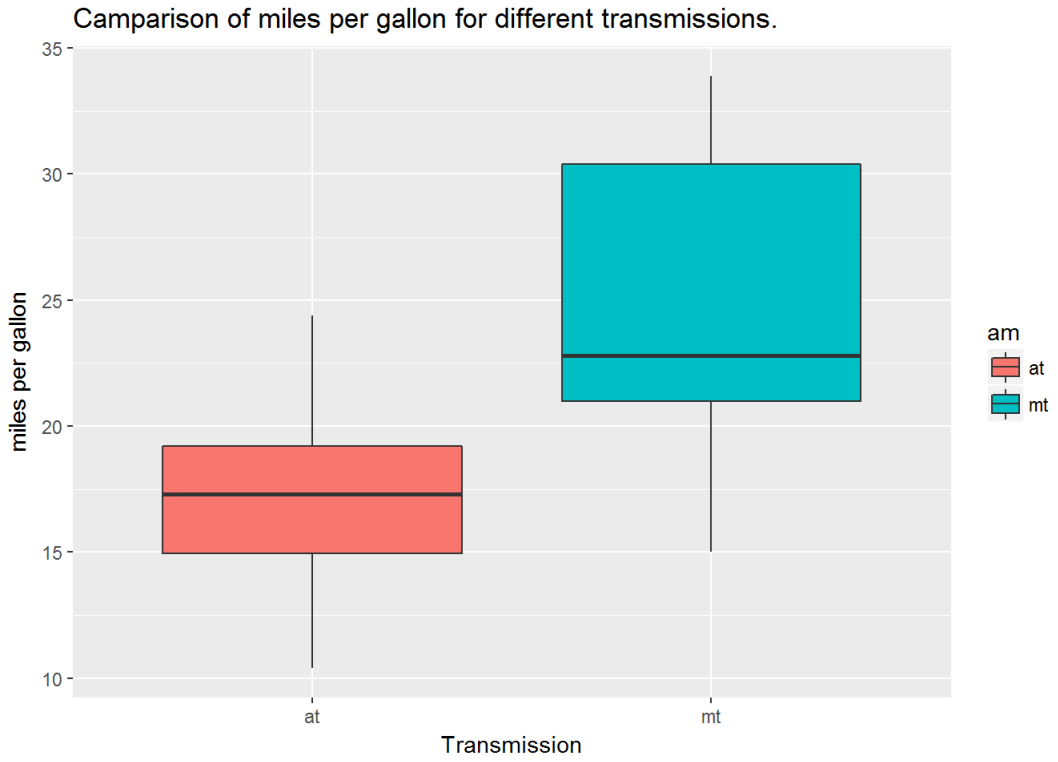


The above pair plot shows potential relationships between different variables of the mtcars dataset.

```
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  : Factor w/ 2 levels "at","mt": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
ggplot(mtcars,aes(am,mpg,fill=am))+geom_boxplot()+labs(x="Transmission")+labs(y="miles per gallon")+labs(tit
le="Camparison of miles per gallon for different transmissions.")
```



Miles per gallon seems to differ according to transmission.Hence formal inference should be conducted over this idea.

Formal Hypothesis testing

```
mpg_at<-select(filter(mtcars,am=="at"),mpg)
mpg_mt<-select(filter(mtcars,am=="mt"),mpg)
t.test(mpg_at,mpg_mt)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg_at and mpg_mt
## 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
```

The p value is less than 0.05 and hence we fail to accept the null hypothesis.The T test alludes to statistically significant difference between mpg of automatic and manual transmission.

Linear Models

Model 1

```
fit1<-lm(mpg~am,data=mtcars)
summary(fit1)
```

```
##
## 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 ***
## ammt           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
```

In this models both the intercept and coefficient are statistically significant.Acoording to the model the estimated mpg for manual transmission is 7.245 more than the automatic transmission.The R^2 of the model is 0.3598,which alludes that this model only explains 35.98% of the relationship.

Model 2

```
fit2<-step(lm(mpg~.,data=mtcars),trace=0)
summary(fit2)
```

```
##
## 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 ***
## ammt          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 step algorithm picks the variables which have highest effect on the mpg for the model.Acoording to this model transmission,weight and qsec(1/4 mile time) seems to have highest effect on mpg.
This model suggests that mpg for manual transmission is 2.9358 more than mpg of automatic transmission,provided the other variables are kept constant.
The adjusted R^2 suggests that around 83.36% of the relationship is explained by the model.

Anova testing

```
anova(fit1,fit2)
```

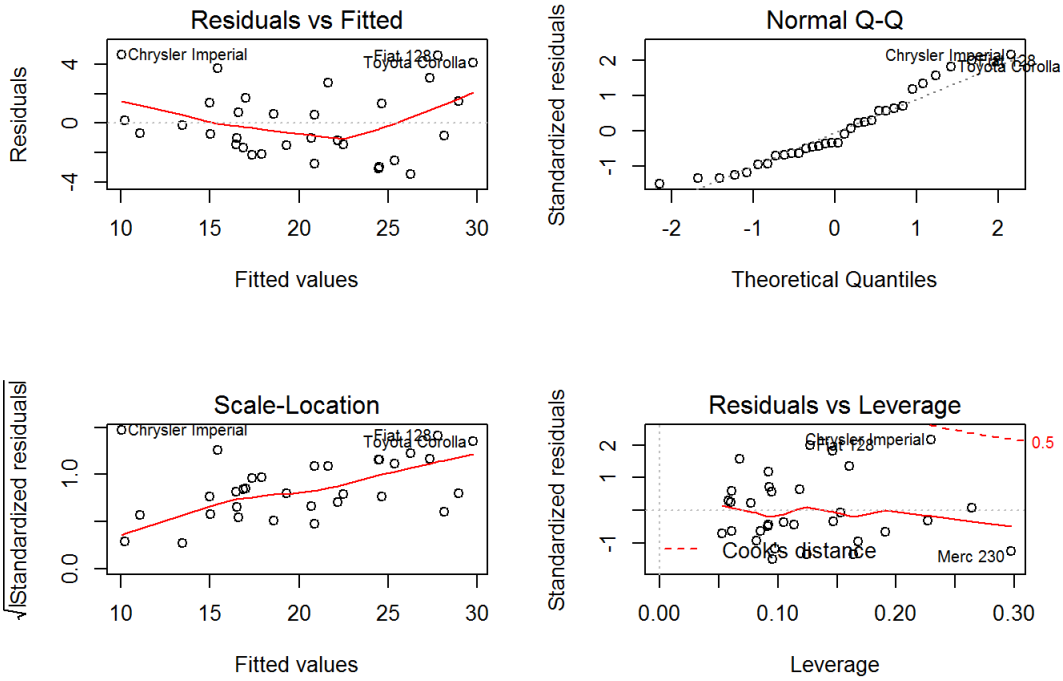
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the anova testing,fit2 is signifiantly better model tha fit1.It explains around 83.36% of the relationship.

Residual testing for the chosen model.

residual plots

```
par(mfrow=c(2,2))
plot(fit2)
```



There does not seem to be any pattern between residuals and the fitted value and residuals seems to follow normal distribution.

Shapiro wilk test

```
shapiro.test(fit2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit2$residuals
## W = 0.9411, p-value = 0.08043
```

The p value is 0.08043 and hence we fail to reject the null hypothesis.The null hypothesis suggest the normality of the residuals.

Conclusion

The fitted model suggest that estimated mpg for manual transmission is 2.9358 more than the estimated mpg for automatic transmission,when other variables are kept constant.