Exploring the relationship between a set of variables and miles per gallon (MPG)

Executive Summary

In this analysis I used the motor trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, I am interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome) and particularly interested in the following two questions,
So,I do linear regression on all variables and exploring relationship between miles per gallon with other variables. I choose those variables that have impact on our model and left others models. Then, I confirm my result by various measures, like regression plot and hypothesis testing. I came up with result that there is quite difference between automatic or manual transmission better for MPG.

Data Processing

Firstly I grab libraries and dataset needed for that analysis.

```r
library(ggplot2)
data(mtcars)
data<-mtcars
```

Exploratory Analysis

After that I do little bit of exploratory analysis for understanding, visually seeing and exploring relationships. And see on appendix for its figure 1.1

```r
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 ...
```

```r
colnames(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

Model design and selection

Firstly I do multivariate regression to explore relationship between all variables and then choosing those that have quite impact on mpg and using backward-elimination strategy to eliminate the unrelated variables one-at-a-time.

```
temp<-summary(lm(data=data,mpg~.-1))$coeff
temp[order(temp[,"Pr(>|t|)"]),][1:3,]
```

```
##       Estimate Std. Error   t value   Pr(>|t|)
## qsec  1.191397  0.4594232  2.593245 0.01659185
## wt   -3.826131  1.8623808 -2.054430 0.05200271
## am    2.832222  1.9751282  1.433944 0.16564985
```

and then after selection top three that have impact and choosing wt, qsec, am and neglection others. Then fitting models and seeing quite improvment in p-value

```
ImprovedData<-mtcars[,c("mpg","am","qsec","wt")]
temp<-summary(lm(data=ImprovedData,mpg~.-1))$coeff
temp[order(temp[,"Pr(>|t|)"]),]
```

```
##       Estimate Std. Error   t value     Pr(>|t|)
## qsec  1.599823  0.1021276 15.664944 1.091522e-15
## wt   -3.185455  0.4827586 -6.598442 3.128844e-07
## am    4.299519  1.0241147  4.198279 2.329423e-04
```

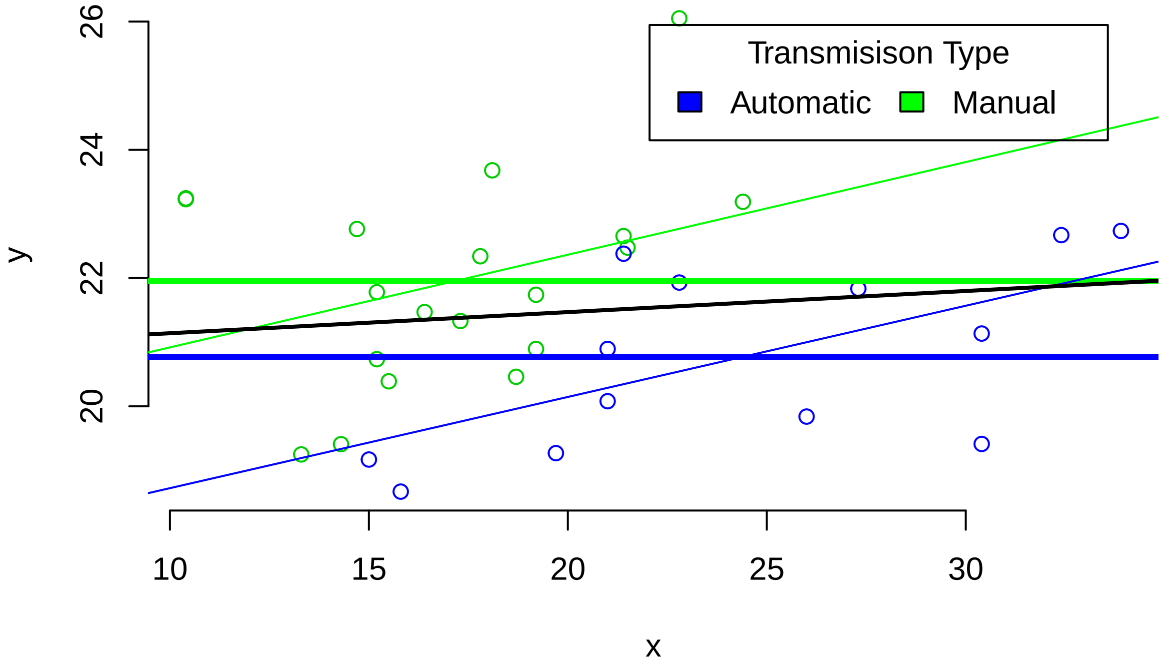
and coefficients for my model will be

```
fit<-lm(data=ImprovedData,mpg~.-1)
fit
```

```
##
## Call:
## lm(formula = mpg ~ . - 1, data = ImprovedData)
##
## Coefficients:
##     am    qsec      wt
##  4.300   1.600  -3.185
```

and then quantifying the difference by the following graph,

```
par(mfrow = c(1, 1)); x<-mtcars$mpg; y<-mtcars$am+mtcars$wt+mtcars$qsec
plot(x,y,frame = FALSE,col=mtcars$am+3,cex=1,bg = "salmon")
abline(h = mean(y[mtcars$am==0]), lwd = 3,col="green"); abline(h = mean(y[mtcars$am==1]), lwd = 3,col="
fit1<-lm(I(y[mtcars$am==0])~I(x[mtcars$am==0])); abline(fit1,col="green")
fit2<-lm(I(y[mtcars$am==1])~I(x[mtcars$am==1])); abline(fit2,col="blue")
abline(lm(y~x),lwd=2)
legend("topright", inset=.05, title="Transmisison Type",c("Automatic","Manual"), fill=c("Blue","Green")
```



for seeing relation among them as, see in appendix for figure 1.2,and then checking validation of my model, For verifying our model, see various error measures in appendix figure 1.3

Conclusion

So by looking at my model, it is quite vivid that there is quite difference between automatic or manual transmission better for MPG. And doing t test on it for further verification.

```
coeff <-summary(lm(mpg~.-1, data = ImprovedData))$coeff
CInterval <- coeff["am", 1] + c(-1, 1) * qt(0.975, df = lm(mpg~.-1, data = ImprovedData)$df) # coeff["a
CInterval
```

```
## [1] 2.254290 6.344749
```

So, looing at the 95% of our model, it suggests that changing from automatic to manual transmission give a 2.2 to 6.39 increase in miles per gallon for the cars. So, in a nutshell, manual transmission performance is quite good as compagreen to automatic transmission.

Appendix

Figure 1.1 for exploratory analysis is below

```
p <- ggplot(mtcars, aes(factor(am), mpg))
p + geom_boxplot() + geom_jitter()+ggtitle("MPG VS Transmission type")+xlab("Transmission type")
```

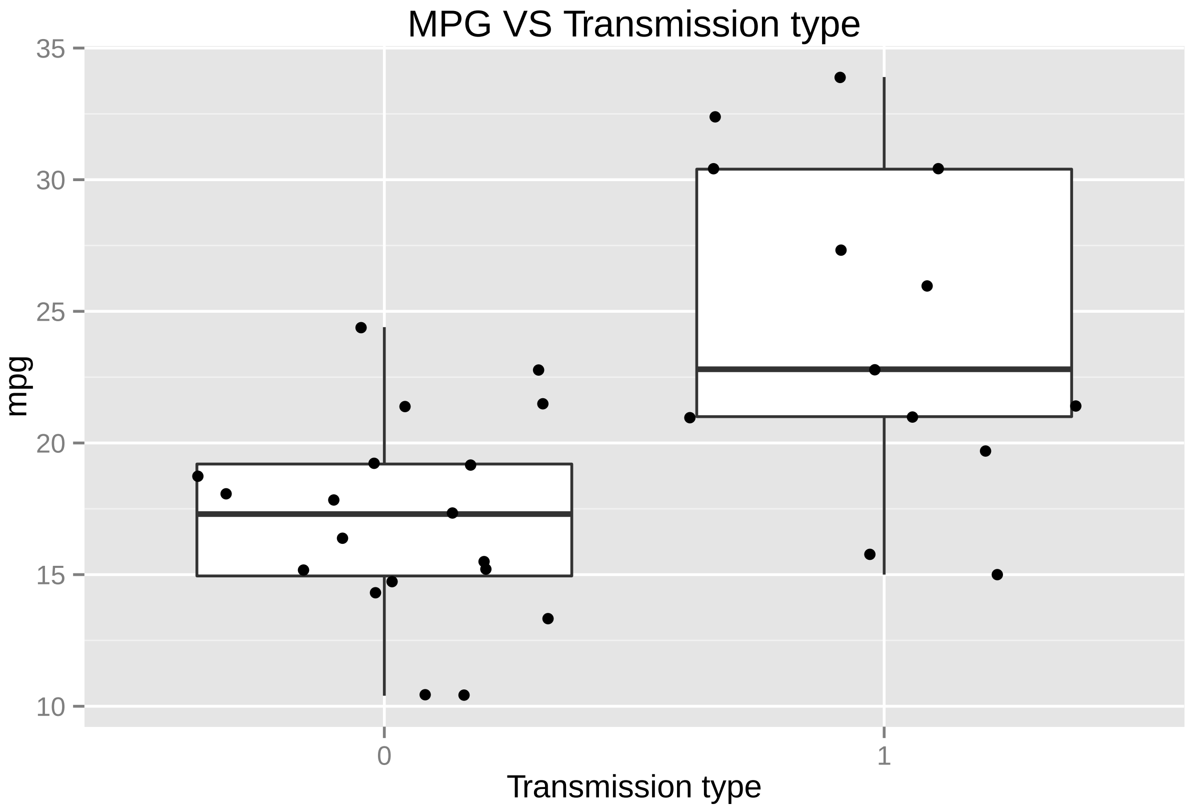


Figure 1.2 seeing relation among varialbles is below

```
pairs(ImprovedData, panel = panel.smooth, main = "comparisons",col=3+ImprovedData$am)
```

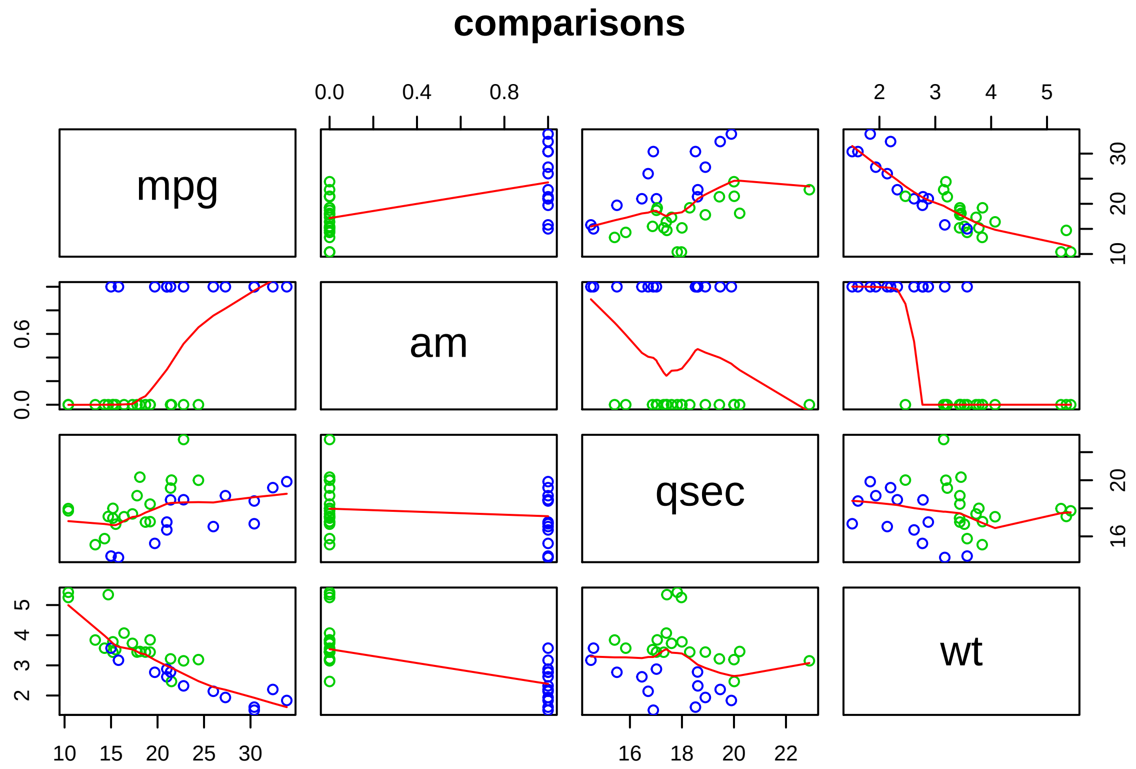
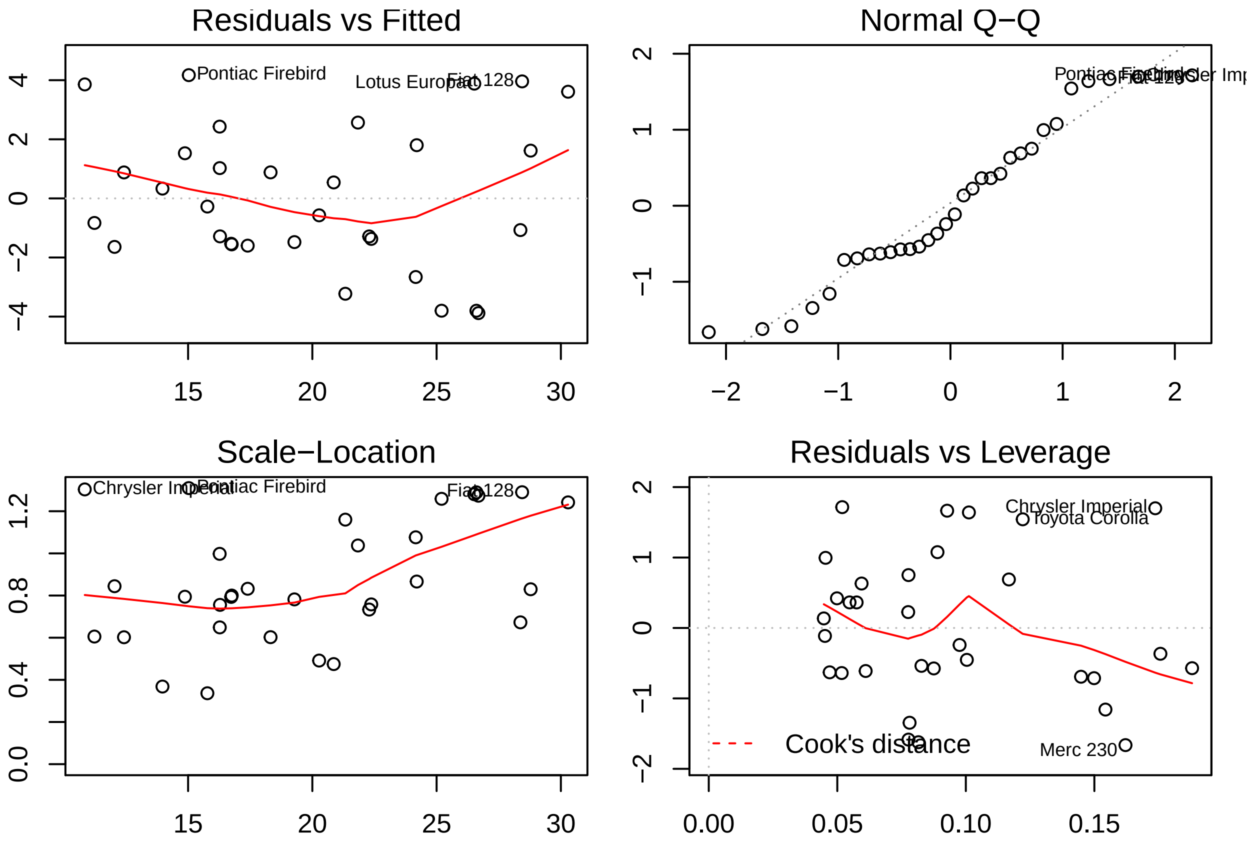


Figure 1.3 for various error measures is below

```
par(mfrow = c(2, 2)); par(mar=c(3.1,2.1,1.1,1.1))
fit <- lm(mpg ~ . -1, data = ImprovedData); plot(fit)
```


Residuals vs Fitted
Pontiac Firebird
Lotus Europa
Fiat 128
4
2
0
−2
−4
15
20
25
30
Normal Q−Q
Pontiac Firebird
Fiat 128
Chrysler Imp
2
1
0
−1
−2
−1
0
1
2
Scale−Location
Chrysler Imperial
Pontiac Firebird
Fiat 128
1.2
0.8
0.4
0.0
15
20
25
30
Residuals vs Leverage
Chrysler Imperial
Toyota Corolla
Merc 230
Cook's distance
2
1
0
−1
−2
0.00
0.05
0.10
0.15