Regression Models - Course Project

Bui Dinh NGoc

February 18, 2016

Regression Models - Course Project

In this report we will analyze the mtcars data set and explore the relationship between a set of variables and miles per gallon (MPG) which will be our outcome .We are focus to answer two questions:

1."Is an automatic or manual transmission better for MPG" 2."Quantify the MPG difference between automatic and manual transmissions"

```
library(datasets)
data(mtcars)
#summary(mtcars)
```

Data processing and transformation

We load in the data set, perform some transformations by factoring the necessary variables and look at the data

Exploratory Data Analysis

Manual transmission has better mileage(mpg) than auto transmissions

```
aggregate(mpg~am, data = mtcars, mean)
```

```
##          am      mpg
## 1 Automatic 17.14737
## 2    Manual 24.39231
```

```
autoData <- mtcars[mtcars$am == "Automatic",]
manualData <- mtcars[mtcars$am == "Manual",]
t.test(autoData$mpg, manualData$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  autoData$mpg and manualData$mpg
## 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
```

Thus, we can conclude manual transmission has better mileage(mpg) than auto transmissions

Regression Analysis

Now I build linear regression model . ### Simple Linear Regression

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

This model is sightly bad R^2 value is 0.3598. This means that our model only explains 35.98% of the variance

Now we use Multivariate Linear Regression for find better fit

```
model <- lm(mpg ~ ., data = mtcars)
```

We can see R^2 is 0.869 mean our model can explains up to 86.90% of the variance

We can compare two model

```
anova(model, simplemodel)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 2: mpg ~ am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     21 147.49
## 2     30 720.90 -9    -573.4 9.0711 1.779e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary

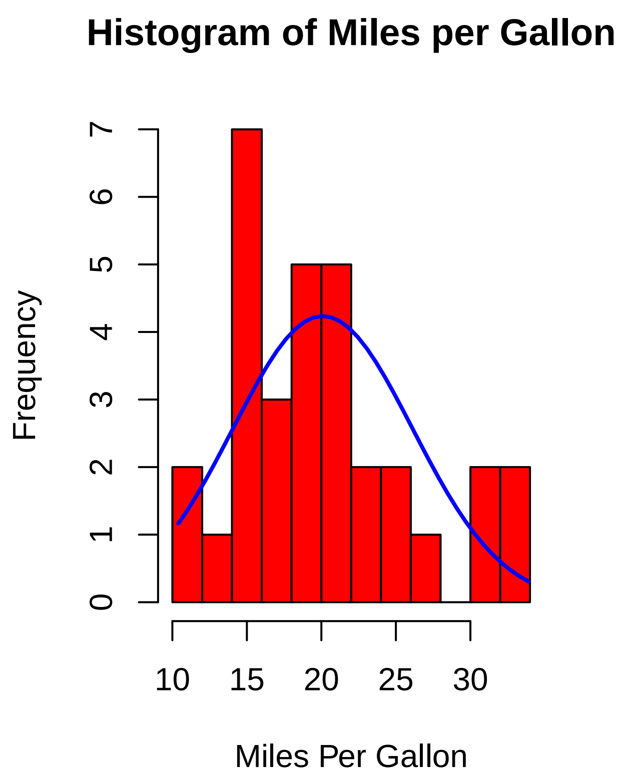
Based on the analysis from our best fit model, we can conclude:

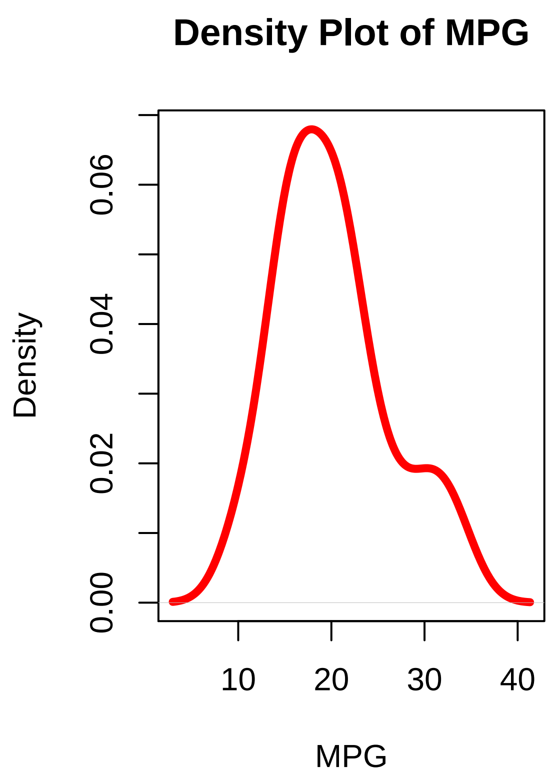
1.Cars with Manual transmission get more miles per gallon mpg compared to cars with Automatic transmission. mpg will decrease by 2.5 (adjusted by hp, cyl, and am) for every 1000 lb increase in wt. 2.In the simplest model, mpg over transmission types, shows that the manual transmission is 7.25 mpg better than automatic transmission. Taking cyliner, displaycement, weight and horsepower into account, the multivariate regression model indicates that the manual transmission is 1.81 mpg better than the automatic transmission while the goodness of fit has been reached 86%.

Appendix

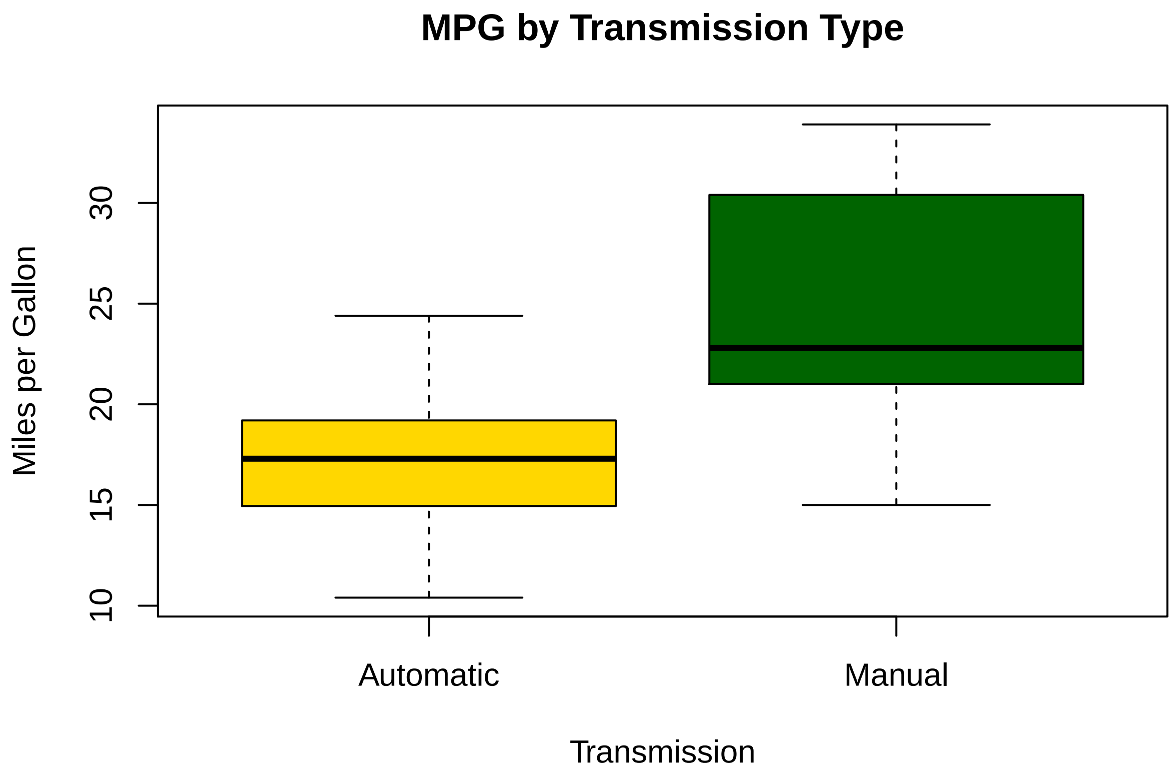
```
par(mfrow = c(1, 2))

x <- mtcars$mpg
h<-hist(x, breaks=10, col="red", xlab="Miles Per Gallon",
   main="Histogram of Miles per Gallon")
xfit<-seq(min(x),max(x),length=40)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
```

```r
d <- density(mtcars$mpg)
plot(d, xlab = "MPG", main ="Density Plot of MPG",col="red",lwd=4)
```





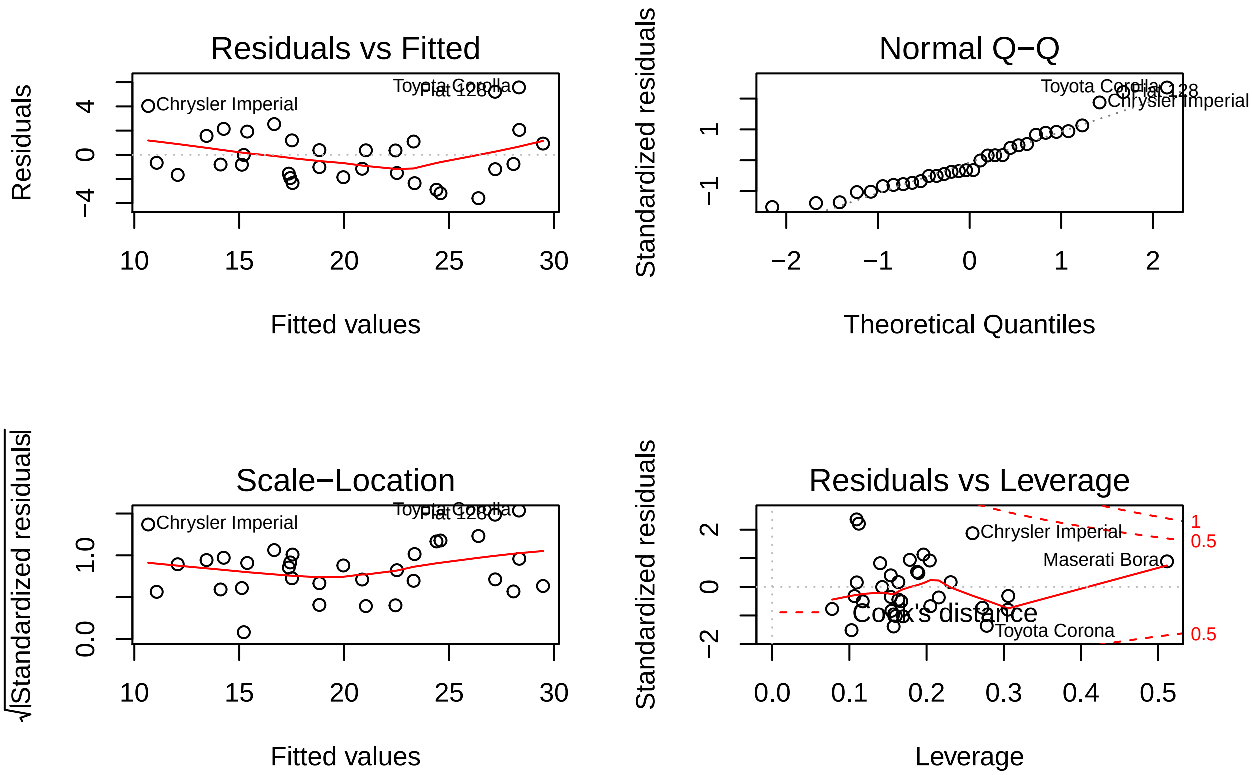

```r
boxplot(mpg~am, data = mtcars,
      col=(c("gold","darkgreen")),
      xlab = "Transmission",
      ylab = "Miles per Gallon",
      main = "MPG by Transmission Type" )
```



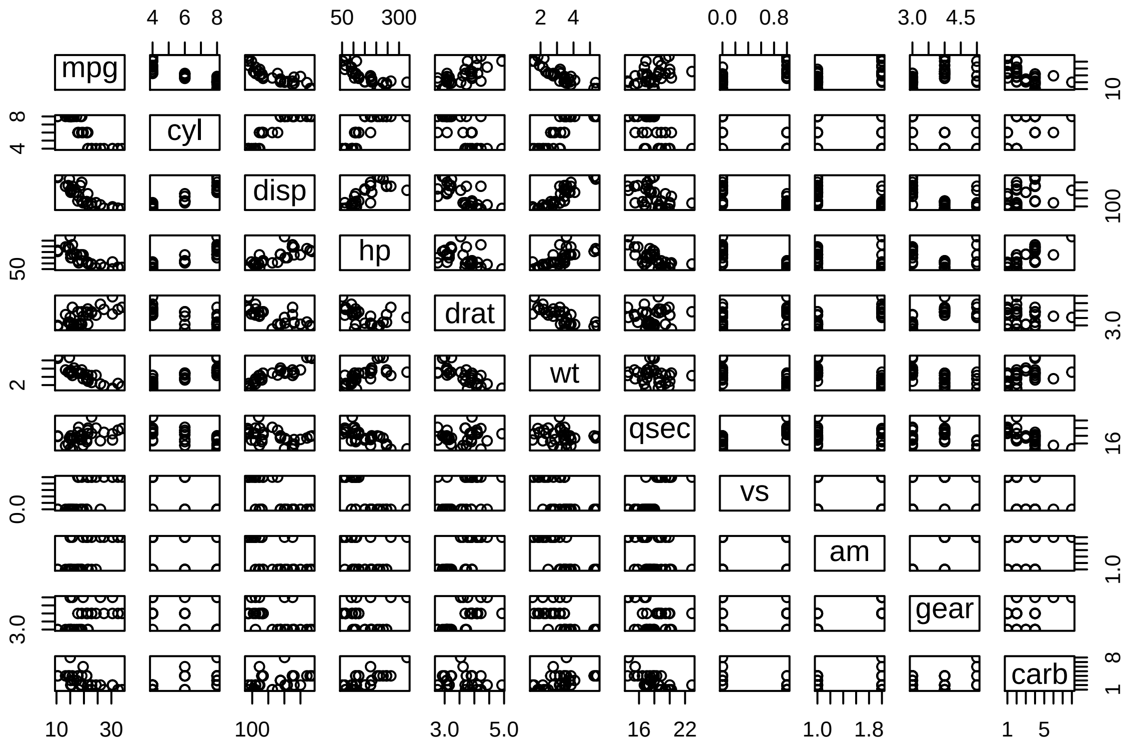

```r
m1<-lm(mpg~am,data=mtcars)
m2<-lm(mpg~am+cyl,data=mtcars)
m3<-lm(mpg~am+cyl+disp,data=mtcars)
m4<-lm(mpg~am+cyl+disp+wt,data=mtcars)
m5<-lm(mpg~am+cyl+disp+wt+hp,data=mtcars)
anova(m1,m2,m3,m4,m5)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl + disp
## Model 4: mpg ~ am + cyl + disp + wt
## Model 5: mpg ~ am + cyl + disp + wt + hp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 271.36  1    449.53 71.6522 6.037e-09 ***
## 3     28 252.08  1     19.28  3.0732  0.091376 .
## 4     27 188.43  1     63.66 10.1461  0.003736 **
## 5     26 163.12  1     25.31  4.0336  0.055097 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow=c(2,2))
plot(m5)
```




```
pairs(mtcars)
```


mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb