Regression Models Course Project

Christian Ibanez

October 22, 2018

Executive Summary

This report answers the following questions: “Is an automatic or manual transmission better for MPG?” “Quantify the MPG difference between automatic and manual transmissions”

The result of the analysis are the following: The difference of 7.24 between the Automatic and Manual in terms of mpg is statistically significant.

Base on the analysis transmission, disp, hp and wt has high correlation to mpg and can be the major predictor variable for mpg

Exploratory Data Analysis

Show the data structure of 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 ...
```

Convert to factors the “vs” and “am”

```
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- as.factor(mtcars$am)
num <- length(mtcars$am)
```

Set 1 = Automatic and 0 = Manual

```
levels(mtcars$am) <- c("Automatic", "Manual")
names(mtcars)[9]<-paste("Transmission")
```

The interquartile range of Manual Transmission is higher than Automatic Transmission when it comes to Miles per Gallon (As shown in the Appendix I). Let’s test this hypothesis.

Inferential Data Analysis

Perform a t.test between the mean of Automatic and Manual for MPG (Assuming equal variance,Assuming unequal variance)

First we need to split the data

```
Auto.Transmission <- subset(mtcars,select=mpg, Transmission == "Automatic")
Manual.Transmission <- subset(mtcars,select=mpg,Transmission == "Manual")
```

Assuming equal variance between the two groups

```
t.test(Auto.Transmission,Manual.Transmission, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Auto.Transmission and Manual.Transmission
## t = -4.1061, df = 30, p-value = 0.000285
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.84837  -3.64151
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

Assuming unequal variance between the two groups

```
t.test(Auto.Transmission,Manual.Transmission, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  Auto.Transmission and Manual.Transmission
## 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
```

We showed that the difference between the mpg of Automatic and Manual is significant Calculate the significant diffenrence

```
Signif.Diff <- mean(Manual.Transmission$mpg) - mean(Auto.Transmission$mpg)
Signif.Diff
```

```
## [1] 7.244939
```

Regression Analysis

Do a regression analysis to verify the result of our t.test

```
LinReg <- lm(mpg ~ Transmission, data = mtcars)
```

Explore the data using the pairs plots. Based on the graph (Appendix II) we notice that disp, hp,wt are correlated to mpg. We will create a Multi Linear Regression for mpg that accounts the said 3 variable

```
MultLinReg1 <- lm(mpg ~ disp + hp + wt, data = mtcars)
```

And another Multi Linear Regression that includes Transmission

```
MultLinReg2 <- lm(mpg ~ Transmission + disp + hp + wt, data = mtcars)
```

Use ANOVA to compare the 3 regression models

```
anova(LinReg, MultLinReg1, MultLinReg2)
```

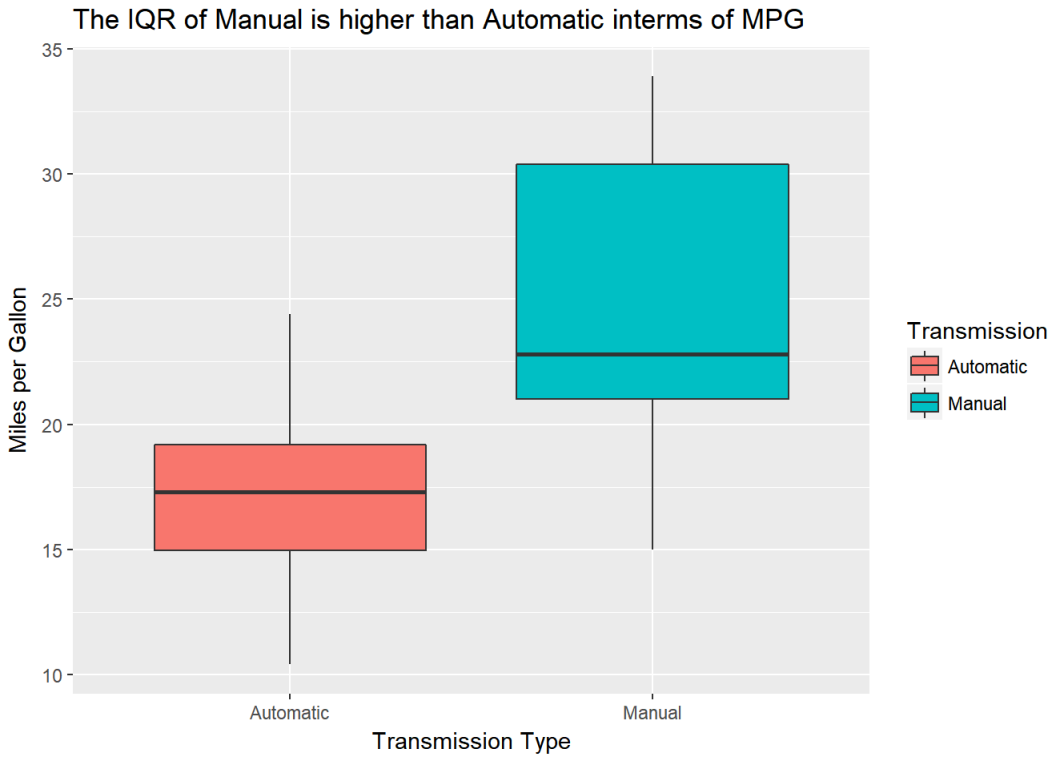
```
## Analysis of Variance Table
##
## Model 1: mpg ~ Transmission
## Model 2: mpg ~ disp + hp + wt
## Model 3: mpg ~ Transmission + disp + hp + wt
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 720.90
## 2     28 194.99  2    525.91 39.4632 9.68e-09 ***
## 3     27 179.91  1     15.08  2.2636   0.1441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our Analysis of Variance shows that MultiLinReg1 is better than the other 2.

Appendix

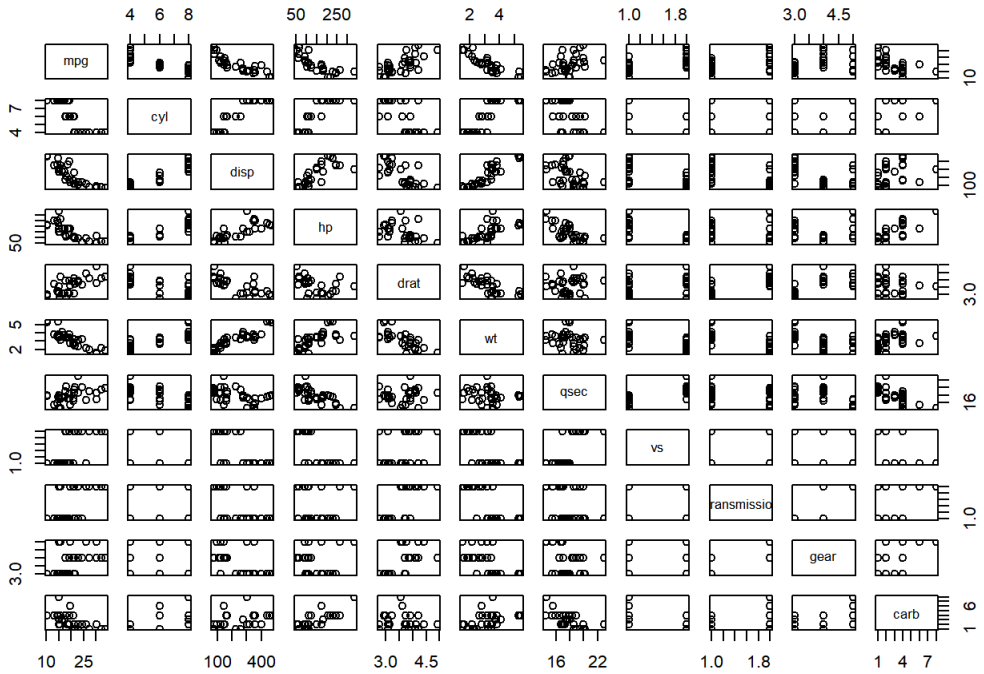
I.

```
library("ggplot2")
ggplot(aes(x = Transmission , y = mpg), data = mtcars) + geom_boxplot(aes(fill = Transmission)) + xlab("Transmission Type") + ylab("Miles per Gallon") + ggtitle("The IQR of Manual is higher than Automatic interms of MPG")
```



II.

```
pairs(mpg ~ ., data = mtcars)
```


mpg
cyl
disp
hp
drat
wt
qsec
vs
ransmissio
gear
carb