Regression Models Course Project

Andy Tse

Motor Trend Analysis

In this report, we are going to analyze the differnces between the manual versus automatic transmission. There will be quanitifications to analyze the data in order to determine the difference on a set of variables and miles per gallon. This is a dataset that is extracted from the automotive information on 1973-1974. The dataset determines that the cars that have automatic transmissions have higher MPGs on average than the manual transmissions. Based on the information that is given in the graphs, it has been determined that the manual transmission vehicles have 7 MPG more than the automatic transmission given on average.

Exploratory Data Analysis and Model Selection

In this analysis, we are going to load the ggplot 2, and lattice in order to determine the data. The analysis is going to determine which cars are better for MPG by comparing the automatic versus the manual transmission.

```
library(ggplot2)
library(lattice)
data(mtcars)
head(mtcars)

# mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1

summary(mtcars)

mpg             cyl              disp             hp              drat
wt           qsec
Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
Min.   :2.760   Min.   :1.513   Min.   :14.50
1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5   1st
Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89
Median :19.20   Median :6.000   Median :196.3   Median :123.0
Median :3.695   Median :3.325   Median :17.71
Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
Mean   :3.597   Mean   :3.217   Mean   :17.85
3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0   3rd
Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90
```

```
Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
Max.   :4.930   Max.   :5.424   Max.   :22.90
vs               am              gear            carb
Min.   :0.0000   Min.   :0.0000   Min.   :3.000   Min.   :1.000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
Median :0.0000   Median :0.0000   Median :4.000   Median :2.000
Mean   :0.4375   Mean   :0.4062   Mean   :3.688   Mean   :2.812
3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
Max.   :1.0000   Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```r
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
basemodel <- lm(mpg ~ am, data = mtcars)
model <- lm(mpg ~ cyl + hp + wt + am, data = mtcars)
summary(basemodel)
```

```
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
```

```r
summary(model)
```

```
Call:
  lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)

#Residuals:
  Min      1Q  Median      3Q     Max
-3.9387 -1.2560 -0.4013  1.1253  5.0513

#Coefficients:
  Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
  cyl6        -3.03134    1.40728  -2.154  0.04068 *
  cyl8        -2.16368    2.28425  -0.947  0.35225
```

```
hp          -0.03211    0.01369  -2.345  0.02693 *
  wt          -2.49683    0.88559  -2.819  0.00908 **
  am           1.80921    1.39630   1.296  0.20646
---
  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.41 on 26 degrees of freedom
Multiple R-squared:  0.8659,    Adjusted R-squared:  0.8401
F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Findings:

Based on all the models below on the appendix, it has been determined that the cars with the manual transmission gets about get an additional 1.8MPG more than the automatic transmission. There is a difference in comparisons with the change in mpg that is making causes from all the other factors such as horsepower, cylinders, and weight to compensate for the data involving in cars.

Statistical Inference

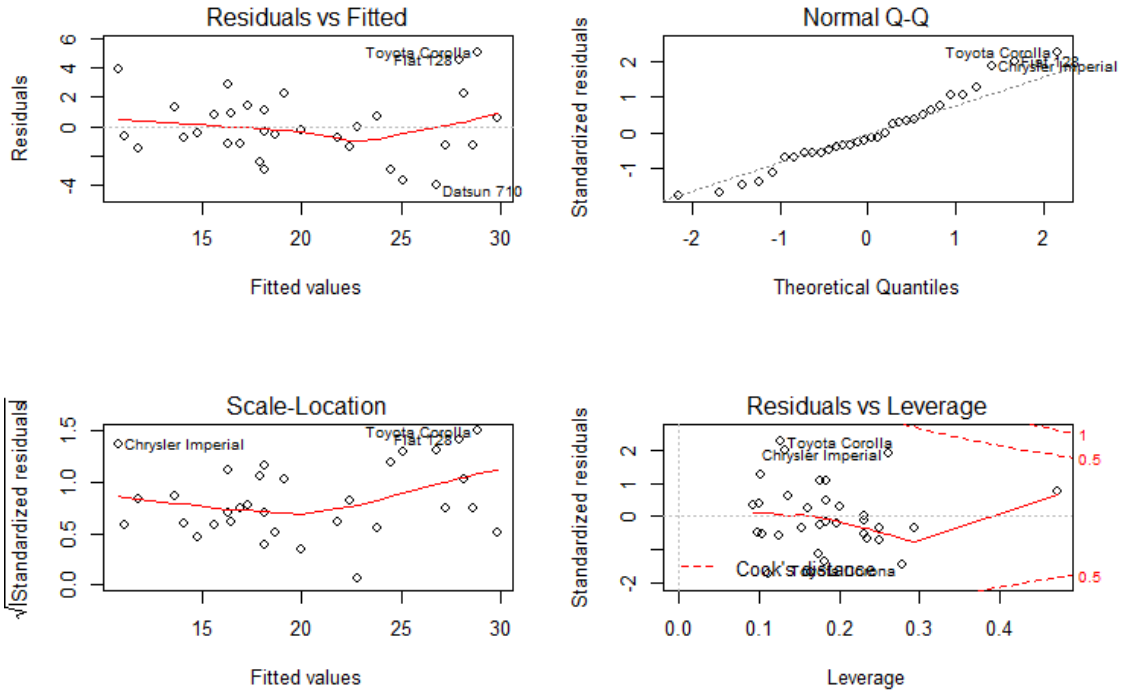
In this section, we are going to do a t-test in order to analyze the differences between the different types of transmissions in comparison the the MPG.

```
t.test(mpg ~ am, data = mtcars)

Welch Two Sample t-test

data:  mpg by am
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 in group 0 mean in group 1
17.14737        24.39231
```

Conclusions: With the differences on the p-values being less than one for the base model and the regular model vehicles, it has been determined that the p-values are almost close to zero. So, based on the sample tests, it can be indicated that we can reject the null hypothesis in order to show that both the manual and automatic transmissions are almost close to identical. It has been determined that the cars data have come from different sets of data.

Appendix and Graphs

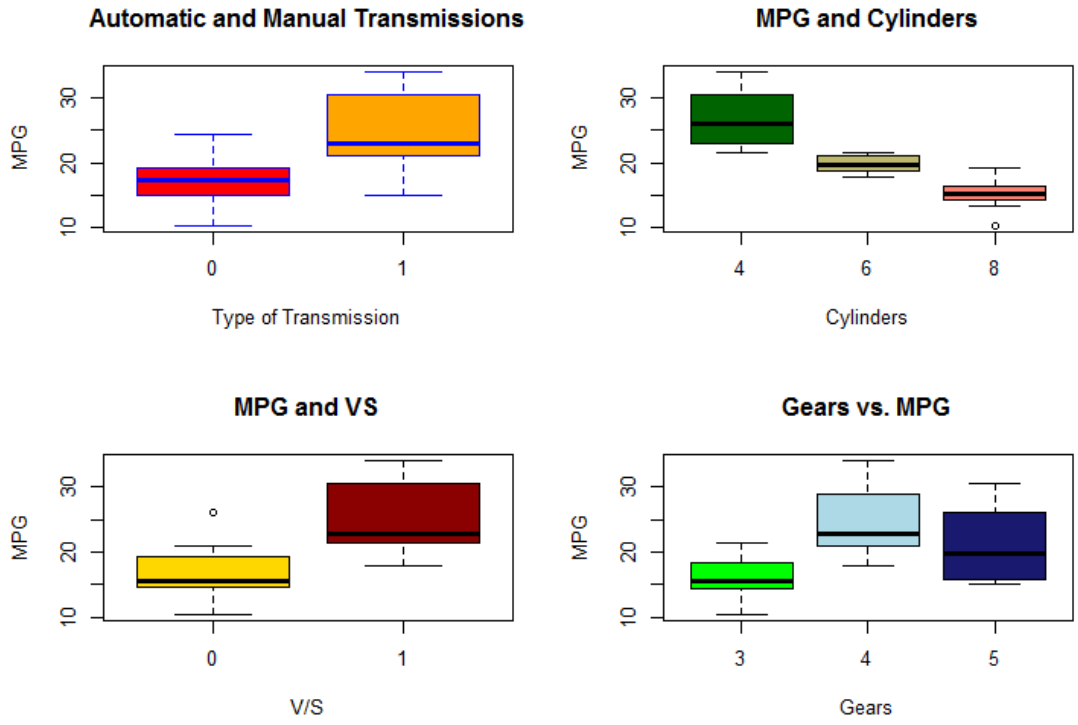
```
Residual Plot
par(mfrow = c(2, 2))
plot(model)
```



Based on the graphs that are given, it has been determined that the normal QQ graph has the normal distribution that is fitting to the intervals between [-1,1]. For all the points on the graphs, it has been determined that they are all the differences on leverage points being distributed between the fitted values and quantities on all four graphs.

Boxplot of MPG Versus Transmission

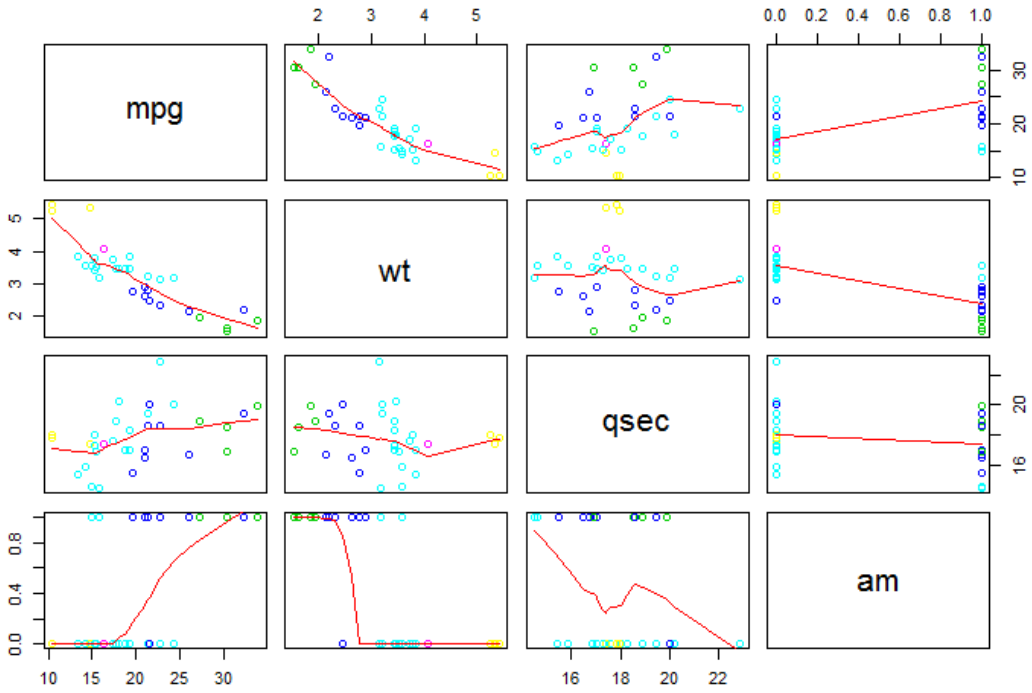
```
boxplot(mpg ~ am, data = mtcars, xlab = "Type of Transmission", ylab =
"MPG",col=(c("red","orange")), border="blue",
main = "Automatic and Manual Transmissions")
boxplot(mpg ~ cyl, data = mtcars, xlab = "Cylinders", ylab = "MPG",
main = "MPG and Cylinders", col=(c("darkgreen", "darkkhaki",
"salmon")))
boxplot(mpg ~ vs, data = mtcars, xlab = "V/S", ylab = "MPG", main =
"MPG and VS", col=(c("gold", "darkred")))
boxplot(mpg ~ gear, data = mtcars, xlab = "Gears", ylab = "MPG", main =
"Gears vs. MPG", col=(c("green", "lightblue", "midnightblue")))
```



Correlations Between Automatic and Manual Transmissions

In this graph, there will be a difference of interpretations with the diffferences between the mpg, wt, and am transmissions.

```
mtcars_vars <- mtcars[, c(1, 6, 7, 9)]
mar.orig <- par()$mar
par(mar = c(1, 1, 1, 1))
pairs(mtcars_vars, panel = panel.smooth, col = 10 + mtcars$wt)
```


mpg
wt
qsec
am
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
10
20
30
2
3
4
5
16
20
0.0
0.4
0.8
10
15
20
25
30
16
18
20
22