Motor Trend Car Road Analysis

Srishti Saha

EXECUTIVE SUMMARY

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions: "Is an automatic or manual transmission better for MPG" "Quantify the MPG difference between automatic and manual transmissions"

SOLUTION

Importing Dataset and EDA

```
#obtain the data

library(datasets)

data(mtcars)

#correlation

cor(mtcars$mpg,mtcars[,-1])
```

```
##            cyl       disp         hp      drat         wt     qsec

## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684

##             vs        am      gear       carb

## [1,] 0.6640389 0.5998324 0.4802848 -0.5509251
```

```
# change data type of a few variables

mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
```

1. Is an automatic or manual transmission better for MPG?

Null-hypothesis = there is no difference in MPG usage for different transmission method. Alternative hypothesis = there is difference in MPG usage for different transmission method.

```
# perform a simple t-test

t.test(mtcars$mpg~mtcars$am)
```

```
##

##  Welch Two Sample t-test

##

## data:  mtcars$mpg by mtcars$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 Automatic    mean in group Manual
##                17.14737                24.39231
```

Based on the results, p-value = 0.001374<0.05, reject the null hypothesis that there is no difference between MPG usage between the different transmission methods, and conclude, that automatic transmission cars have lower mpg compared with manual transmission cars, with assumption that all other conditions remain unchanged.

Furthermore, checking mean of mpg in Automatic (at 17.1) and Manual (at 24.4), **conclusion is that the Manual transmission is better for MPG; keeping all other factors constant.**

2. Quantify the MPG difference between automatic and manual transmissions

The aim here is to quantify the MPG difference between transmission types, and find if there are other variables that account for the MPG differences.

```r
# creating a step wise model
stepmodel = step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
summary(stepmodel)
```

```
##
## 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 ***
## amManual      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
```

About 84% (by adjusted R-squared as that penalizes for too many variables and normalizes the variance) of the variance is explained by this model. Cylinders cyl6 and cyl8 both have a negative relation with mpg (-3.03 miles and -2.16 miles for cyl6 and cyl8 respectively). Similarly, horsepower (hp) sees a negative relation wrt mpg i.e. (-0.03miles), and weight (wt) shows -2.5miles (for every 1,000lb).

However, amManual has a value of 1.80921 mpg , thus indicating the magnitude by which manual transmission is better than automatic transmission i.e 1.81mpg

Improvement in model for Q2

```r
# creating a step wise model
model_imp <- lm(mpg~ factor(am):wt + factor(am):qsec,data=mtcars)
summary(model_imp)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               13.9692     5.7756   2.419  0.02259 *
## factor(am)Automatic:wt    -3.1759     0.6362  -4.992 3.11e-05 ***
## factor(am)Manual:wt       -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)Automatic:qsec   0.8338     0.2602   3.205  0.00346 **
## factor(am)Manual:qsec      1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13
```

Observation 1: Observation is that this model has an adjusted R-squared of 87.9% (or 89.5%)

From the coeff values, we have the following conclusions:

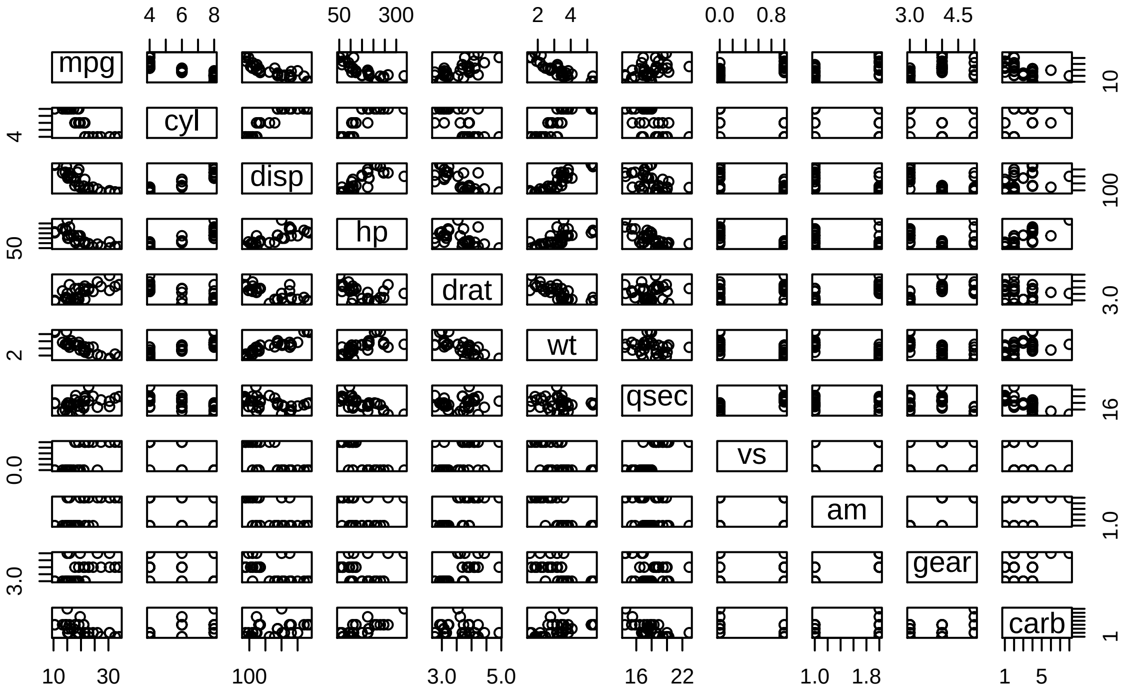
Observation 2: As the weight increased by 1000 lbs, the mpg decreased by -3.176 for automatic transmission, and -6.09 for manual transmission. Conclusion: with increasing car weight we should choose manual transmission cars

Observation 3: when the acceleration speed dropped, and 1/4 mile time increased (by 1 sec), the mpg factor increased by 0.834 miles for automatic transmission cars, and 1.446 miles for manual transmission cars Conclusion: With lower acceleration speed, but same weight, manual transmission cars are better for mpg

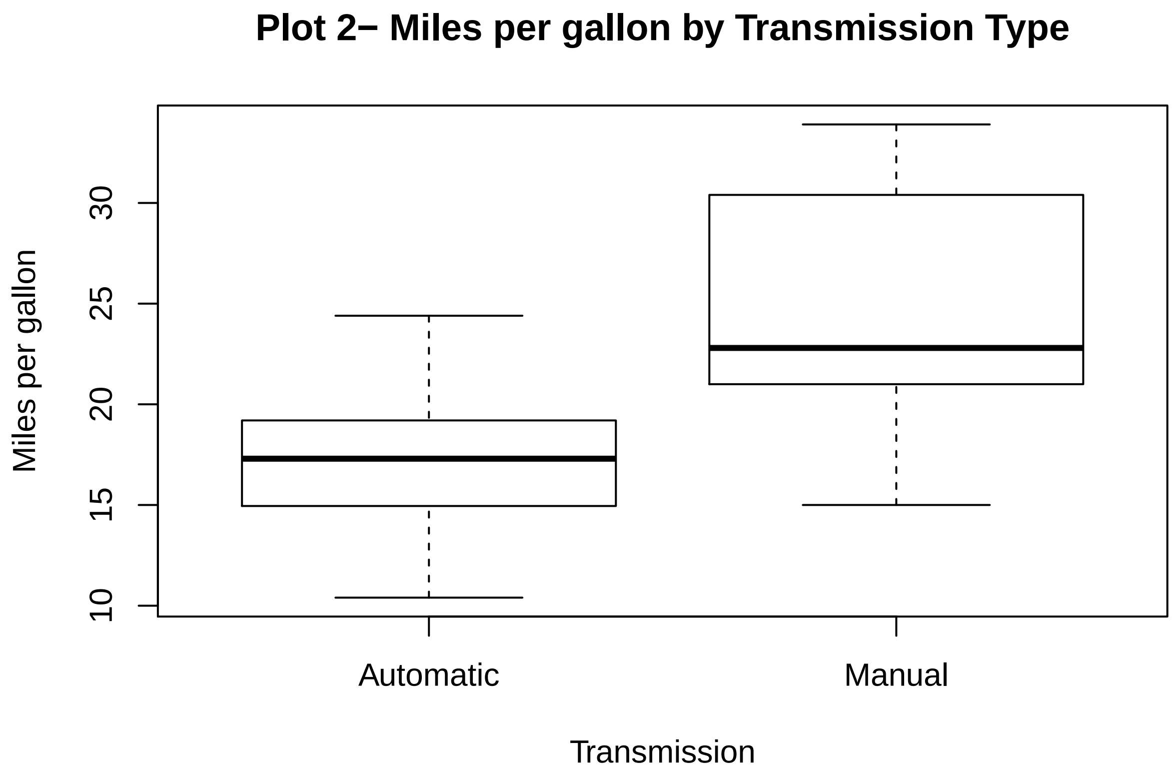
APPENDIX

```r
#checking correlation of all variables

pairs(mpg ~ ., data = mtcars, main="Plot 1 - Relationships between all the variables")
```

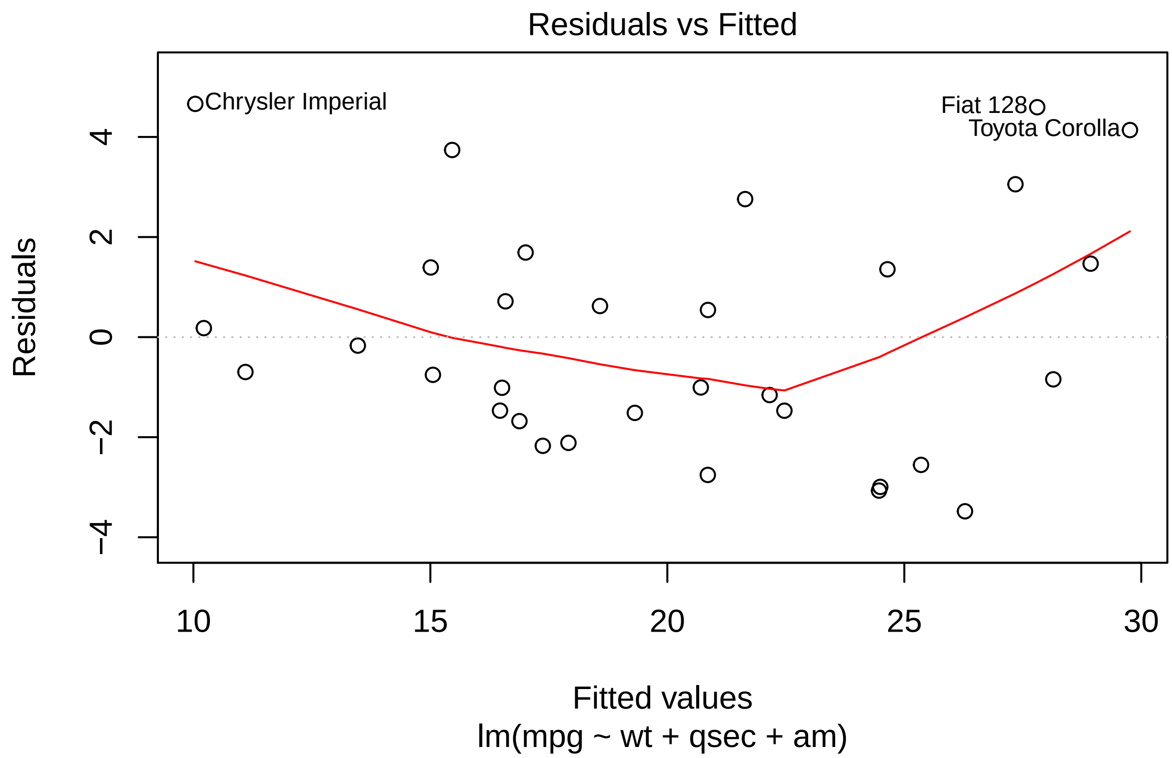
Plot 1 − Relationships between all the variables



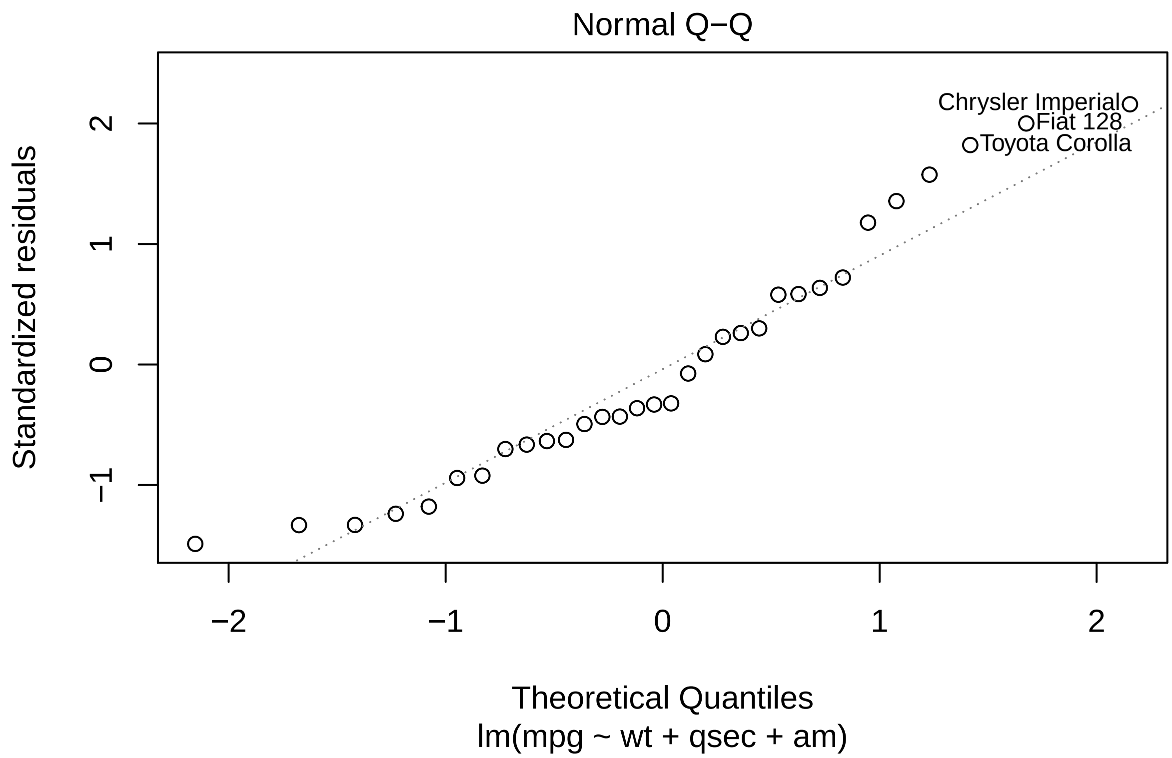

```r
#Checking the boxplots for mpg by transmission type


boxplot(mpg ~ am, data = mtcars, xlab = "Transmission", ylab = "Miles per gallon", main="Plot 2- Miles per gallon by Transmissic
```

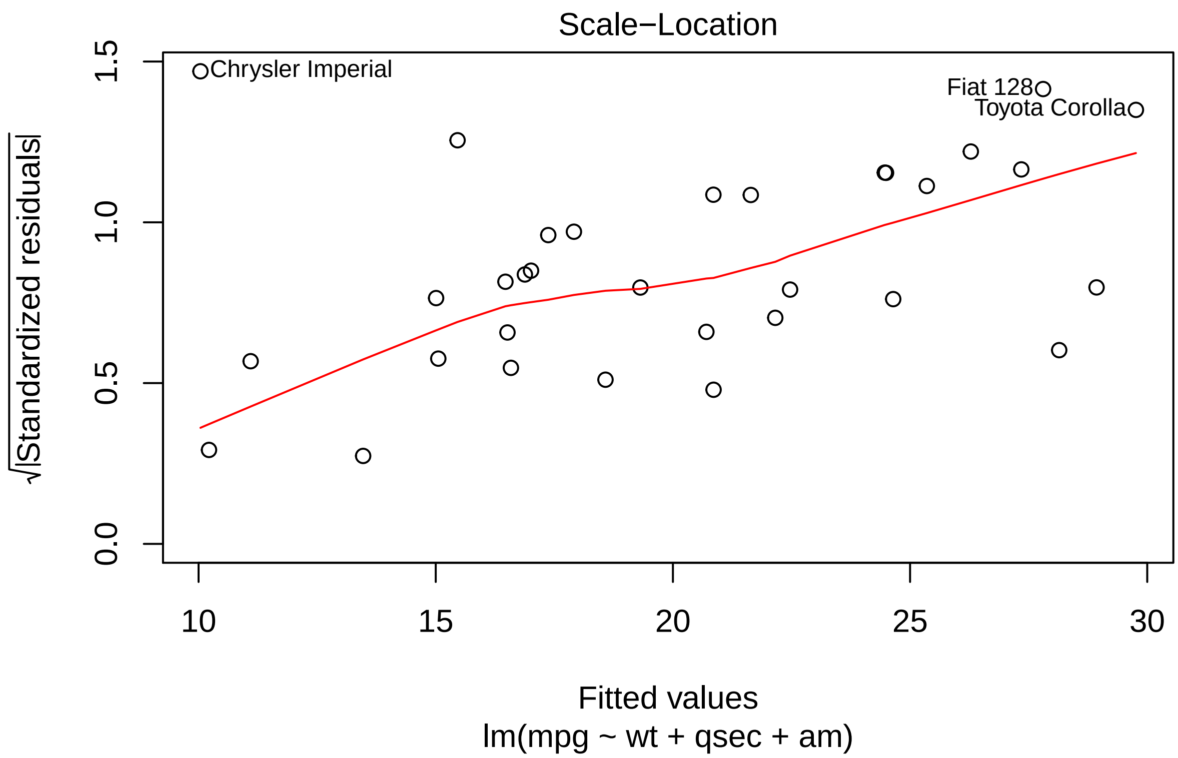


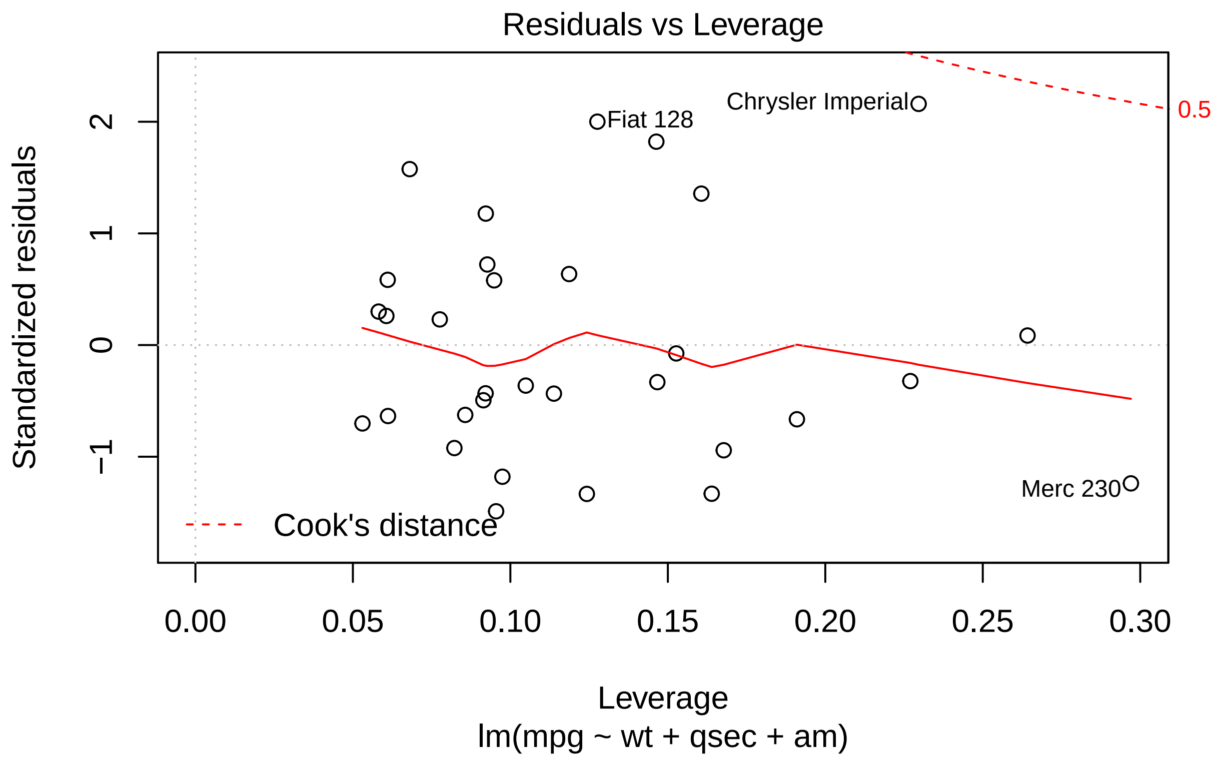

```r
#plotting the model results

plot(stepmodel)
```



Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ wt + qsec + am)
Chrysler Imperial
Fiat 128
Toyota Corolla
4
2
0
−2
−4
10
15
20
25
30


Normal Q−Q
Chrysler Imperial
Fiat 128
Toyota Corolla
Standardized residuals
−1
0
1
2
−2
−1
0
1
2
Theoretical Quantiles
lm(mpg ~ wt + qsec + am)


Scale−Location
Chrysler Imperial
Fiat 128
Toyota Corolla
1.5
1.0
0.5
0.0
√|Standardized residuals|
10
15
20
25
30
Fitted values
lm(mpg ~ wt + qsec + am)



Residuals vs Leverage
Standardized residuals
Leverage
lm(mpg ~ wt + qsec + am)
Cook's distance
0.5
Fiat 128
Chrysler Imperial
Merc 230
2
1
0
−1
0.00
0.05
0.10
0.15
0.20
0.25
0.30