Motor Trend Analysis

Abhijeet Sharma

Wednesday, June 17, 2015

Introduction

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”

First, We load the dataset in R and get a general feel of it.

```
dim(mtcars)
```

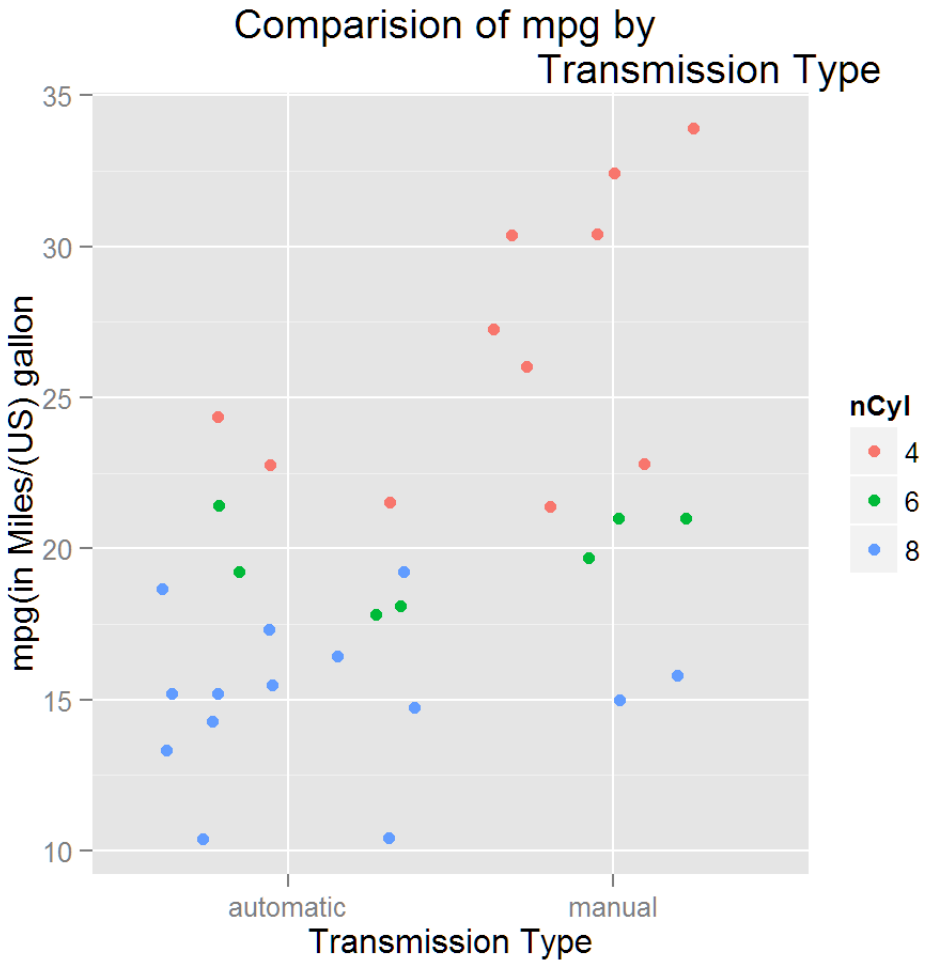
```
## [1] 32 11
```

```
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 ...
```

We are especially interested in the “am” column which represents the Transmission (0 = automatic, 1 = manual)

Plotting the mpg values with this transmission column using the number of cylinders as color factor, We investigate if there might be other factors involved too. We have used the “jitter” parameter to give us a better view of the data.



Difference in Average Means:

```
##     auto   manual      diff
## 4 22.900 28.07500 -5.175000
## 6 19.125 20.56667 -1.441667
## 8 15.050 15.40000 -0.350000
```

We observe that difference in average mpg between automatic and manual is higher for 4-cylinder cars compared to 6-cylinder and 8-cylinder cars.

We now do a linear regression on the variables. We now build and compare multiple models:

```
fit <- lm(mpg ~ am,data=mtcars)
fit2 <- lm(mpg ~ am + cyl, data=mtcars)
fit3 <- lm(mpg ~ am + cyl * hp, data=mtcars)
fit4 <- lm(mpg ~ am + cyl * hp + wt, data=mtcars)
anova(fit, fit2, fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl * hp
## Model 4: mpg ~ am + cyl * hp + wt
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 271.36  1    449.53 89.6039 6.565e-10 ***
## 3     27 181.49  2     89.87  8.9566  0.001099 **
## 4     26 130.44  1     51.05 10.1763  0.003693 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

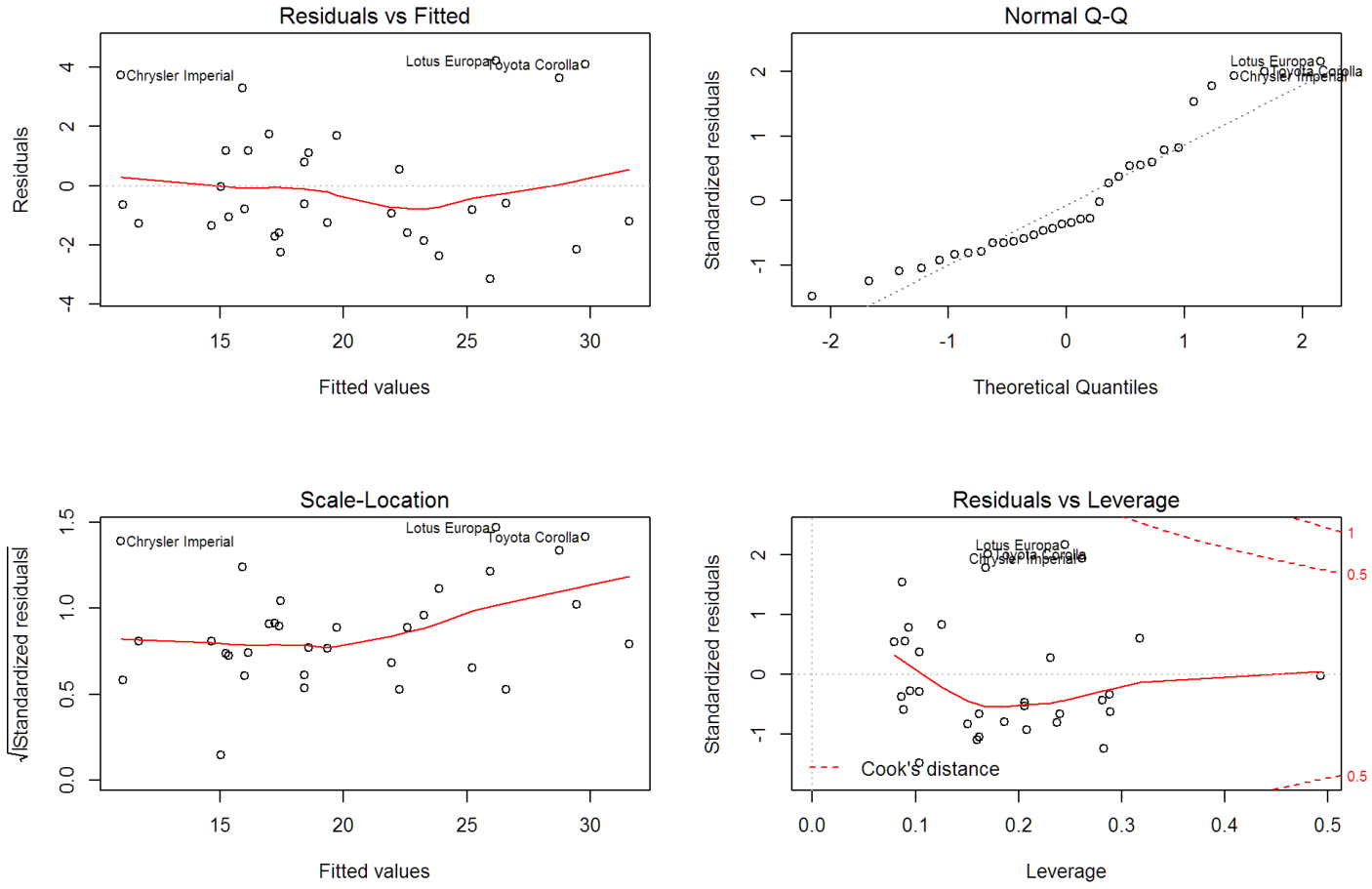
We also check the RMSE values:

```
##         sigma
## fit  4.902029
## fit2 3.058973
## fit3 2.592675
## fit4 2.239845
```

From the results above, we select fit4 as our final model as it has the lowest RMSE scores.

We now check the residual plots:

```
par(mfrow=c(2,2))
plot(fit4)
```



We observe that even though there are couple of outliers, they aren't very influential, so we keep those in the dataset(Also, the first plot is almost linear in y-axis as it should be.

Finally, we get the coefficient values for the final model.

```
round(summary(fit4)$coef[,1],3)
```

```
## (Intercept)    ammanual         cyl          hp          wt      cyl:hp
##      49.481       1.322      -2.538      -0.168      -2.619       0.019
```

The result tells us that Manual Transmission increase the value of mpg by 1.322 value more than Automatic Transmissions.

Summary

From the above results, we conclude that:
1. A Manual Transmission is generally better for MPG than an Automatic Transmission.We also observe that the difference in mpg between automatic and manual is higher for 4-cylinder cars compared to 6-cylinder and 8-cylinder cars.
2. The value of MPG increases by 1.322 miles/gallon more for a Manual transmission vehicle compared to an Automatic transmission vehicle.

We have taken the below mentioned assumptions:
1. The conclusions are based on only the mtcars dataset provided in R.
2. We have less number of observations especially for 6 and 8-cylinder manual transmission cars, so results might change if large number of new observations are used.
3. We observed from the plot that even though there are couple of outliers, they weren't very influential.So, we keep those in the dataset.