Regression Models Course Project

Abhinav Agarwal

27 November 2018

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"
```

Some regression analysis was done, and the results obtained show that cylinders, horsepower, and weight are the important factors in affecting the MPG.

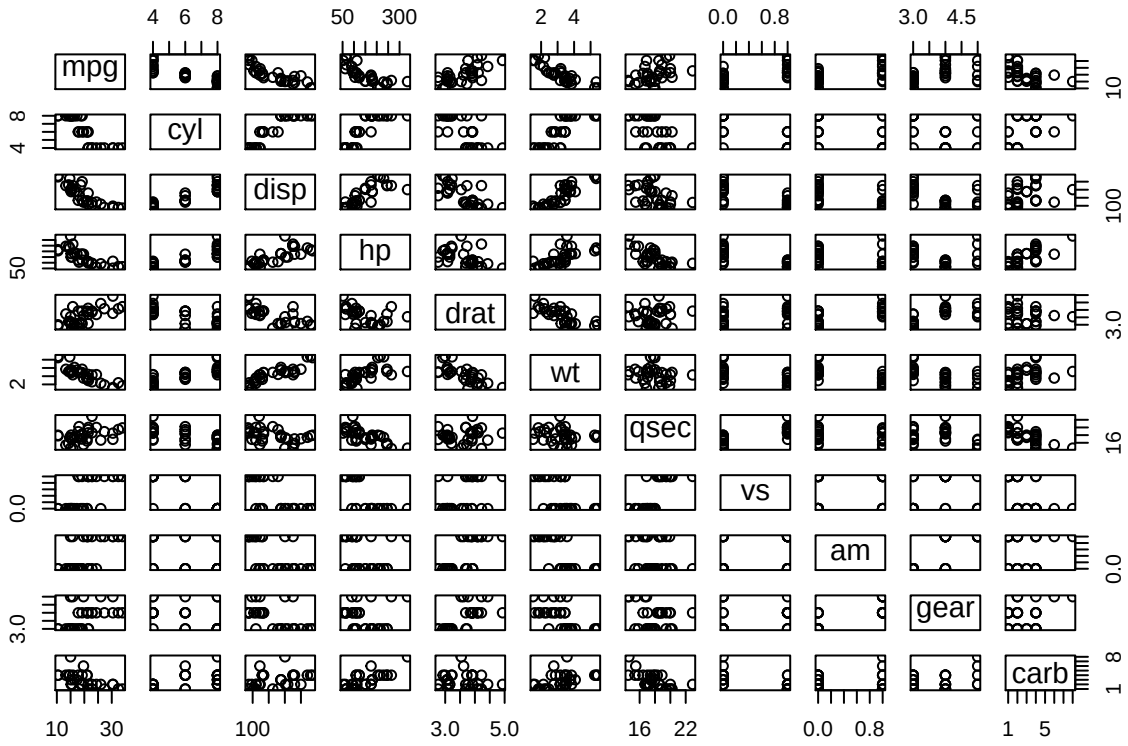
Data Processing

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

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

The pairwise scatter plot between all variables is:

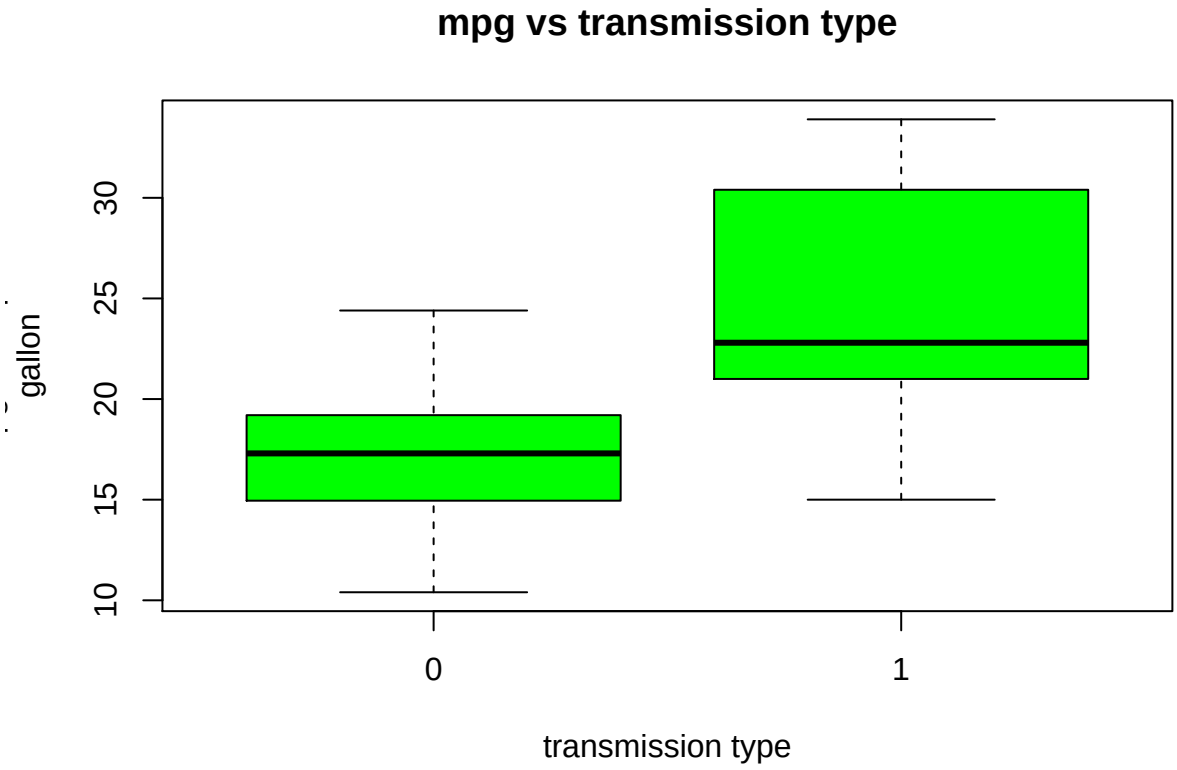
```
pairs(mtcars)
```



Is an automatic or manual transmission better for MPG?

A boxplot can be seen in, with the relationship between mpg and am type. It seems like automatic car has better mpg compared with manual cars.

```
boxplot(mtcars$mpg ~ mtcars$am, data = mtcars, outpch = 19, ylab="mpg:miles per
gallon",xlab="transmission type",main="mpg vs transmission type", col="green")
```



We conduct a t-test to test the hypothesis.

```
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 0 mean in group 1
##        17.14737        24.39231
```

Since p-value = 0.001374<0.05, we reject the null hypothesis and conclude that manual transmission is better than automatic transmission for MPG

Quantifying mpg difference

A multivariate linear regression with all variables is done.

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

Here we can adopt a stepwise algorithm, to choose the best model. We are using the step() function.

```
stepmodel = step(m,trace=0,steps=10000)
summary(stepmodel)
```

```
##
## 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
```

From the above, we may conclude that automatic run at 17.15mpg, while manual have 7.24 more mpg.

Also, R^2 is 0.36, hence the model only accounts for 36% variance.

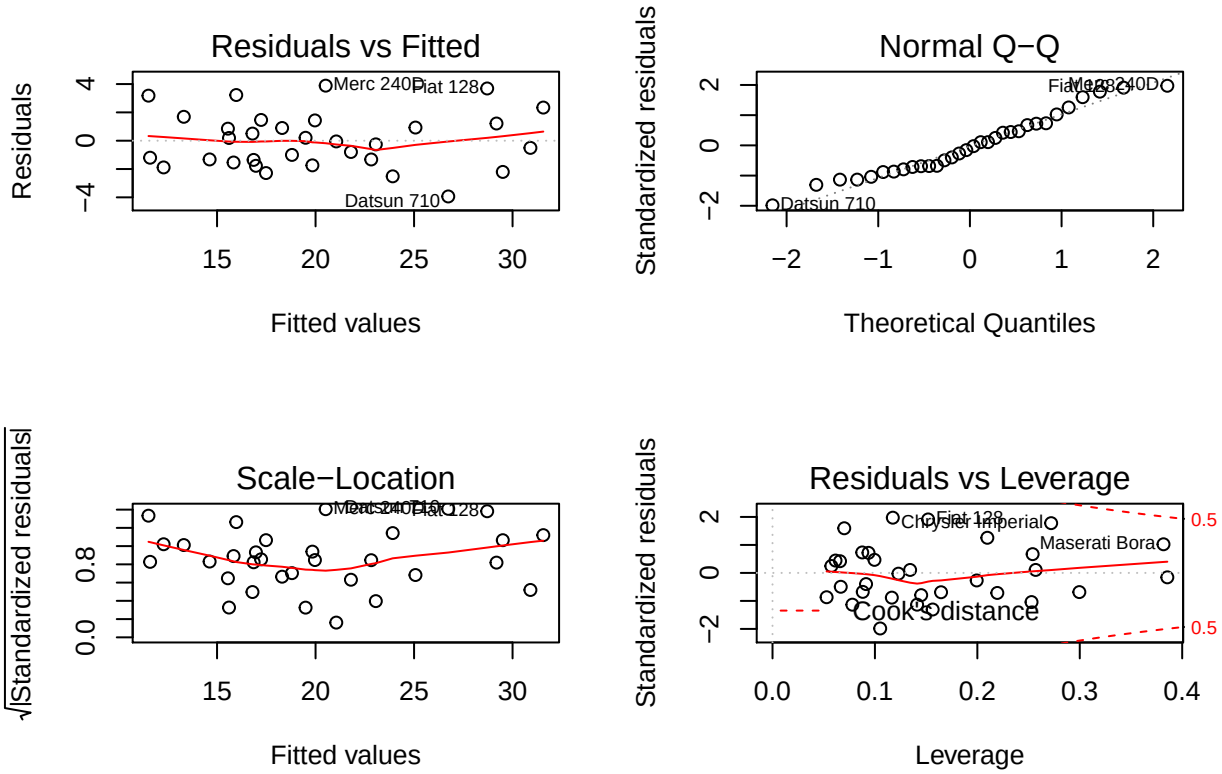
To further optimize the model, we can examine mpg ~ wt + qsec correlation with am.

```
model <- lm(mpg~ factor(am):wt + factor(am):qsec,data=mtcars)
summary(model)
```

```
##
## 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)0:wt     -3.1759     0.6362  -4.992 3.11e-05 ***
## factor(am)1:wt     -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)0:qsec    0.8338     0.2602   3.205  0.00346 **
## factor(am)1: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
```

The results suggests that the best model includes wt, and amManual variables. About 89% of the variance is explained by this model

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



Hence, the residuals are normally distributed.

Conclusion

On average, manual transmission is better than automatic transmission by 1.81mpg. However, transmission type is not the only factor accounting for MPG, cylinders, horsepower, and weight are the important factors in affecting the MPG.

References

- Motor Trend magazine: Automatic or manual transmission? by Csaba Sarkadi
- Motor Trend Analysis - Regression Model Project