Johns Hopkins University - Motor Trends Analysis

Myr Balada

31/10/2020

Executive Summary

The objective of this Report is to analyze the fuel efficiency expressed in mpg (miles per gallon) in cars with automatic and manual transmission. The data was extracted from the Motor Trend Magazine of the year 1974 (it is an American magazine that talks about the automobile industry). For practical purposes, the data is available in R Packages "dataset". We are particularly interested in the following two questions:

1. "Is an automatic or manual transmission better for MPG"
2. "Quantify the MPG difference between automatic and manual transmissions"

To answer the questions above, we will rely on simple linear regression analysis. If we leave the variables of weight, power and displacement constant, we will see that a better performance in mpg is appreciated in a car with a manual transmission over another with an automatic transmission. The improvement observed is greater than 2 mpg. Details of the exploratory Analyzes, graphical and other information are available in the Appendix.

The source of the data: "mtcars"

Here is a look at some technical data of cars that are in "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  V  M    4    4
## Mazda RX4 Wag    21.0   6  160 110 3.90 2.875 17.02  V  M    4    4
## Datsun 710       22.8   4  108  93 3.85 2.320 18.61  S  M    4    1
## Hornet 4 Drive   21.4   6  258 110 3.08 3.215 19.44  S  A    3    1
```

```
## '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 : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
##  $ 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  : Factor w/ 2 levels "V","S": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "A","M": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Exploratory Data Analysis

A look on the range (for **automatic and manual transmission**) of the variable "mpg" and its quantiles:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.43   19.20   20.09   22.80   33.90
```

Range for **automatic transmission** of the variable "mpg" and its quantiles:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```

Range for **manual transmission** of the variable "mpg" and its quantiles:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

To get another angle in the mpg analysis, we ran a t.test (see results in *Appendix A.1*), obtaining an average of **17.14737** for the automatic transmission and an average of **24.39231** for the manual transmission. Also, you can see the results in the boxplot graph in *Appendix A.2* In addition, let's see that other variables of the cars under analysis are correlated to efficiency in mpg:

```
##   mpg    wt   cyl  disp    hp  drat    vs    am  carb  gear  qsec
##  1.00 -0.87 -0.85 -0.85 -0.78  0.68  0.66  0.60 -0.55  0.48  0.42
```

According to the correlation table, there are at least four variables with a high correlation to our outcome variable "mpg". The highest value comes from the weight variable "wt". Let's have a look to this variable separately for automatic (0) and manual (1) transmission in Appendix A.3.

Linear Models

The linear dependencies suggest to analyze linear models as follows:

```
fit1 <- lm(mpg ~ am , data = mtcars)
fit2 <- lm(mpg ~ am + wt, data = mtcars)
fit3 <- lm(mpg ~ am + wt + hp , data = mtcars)
fit4 <- lm(mpg ~ am + wt + hp+ disp, data = mtcars)
fit5 <- lm(mpg ~ ., data = mtcars)
```

The base variable of our analysis is "mpg" as a function of "am" (automatic / manual transmission). We incorporate a series of complementary variables, in order to obtain a model that allows us to explain the changes in "mpg". Given the high correlation between the variables "cyl" and "disp", we chose to leave out "cyl". The variables "wt" and "hp" interest us because of the notable improvement of the model. This corresponds to the case of the variable "fit3". Appendix A.5 shows the correlations of the variables used. On the other hand, Appendix A.6 shows the summation of the "fit3" model that bases 84% of the variability in "mpg".

By reviewing Appendix A.7, we can understand through the "Residual" graph of the "fit3" model. The fit that was made in the model works because it is the requirement of a linear model which is precisely what explains the variability in "mpg".

Final Conclusion

Regarding the first question, it is clear that when choosing between automatic or manual transmission, mpg performance is better for the manual version (keeping the rest of the parameters constant). For the second question, if it is also more difficult to answer, it is demonstrated through making a good adjustment to the model, since it allows us to obtain an improvement of 2.93 in mpg for the manual transmission. We obtain a p <0.05 and an R squared of 0.85. There is another relevant factor that has to do with the number of observations in the database (they are not so many as to obtain a better analysis). In the residual plot, the manual transmission model is seen to be better, but we are only taking 32 observations.

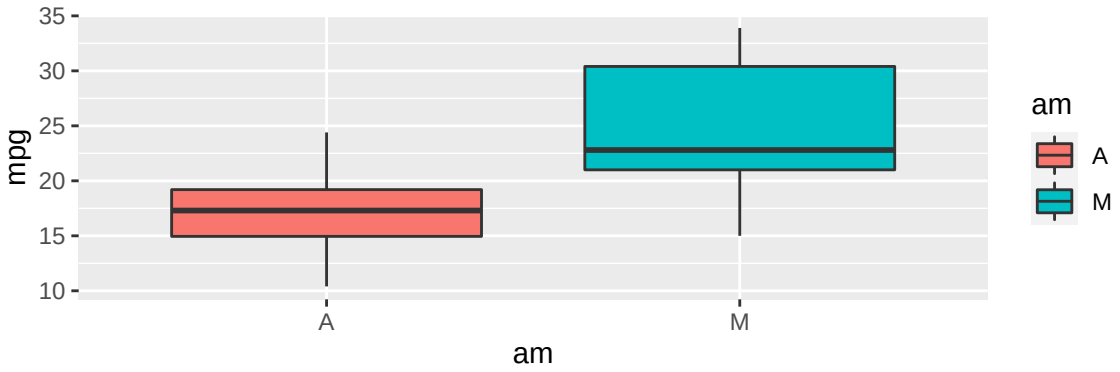
(You can see the code detail at https://github.com/myrba2win/Regression-Model-Project-Motor-Trends)

Appendix

A.1 Using t-Test with the variable "mpg" for Automatic and Manual Transmission

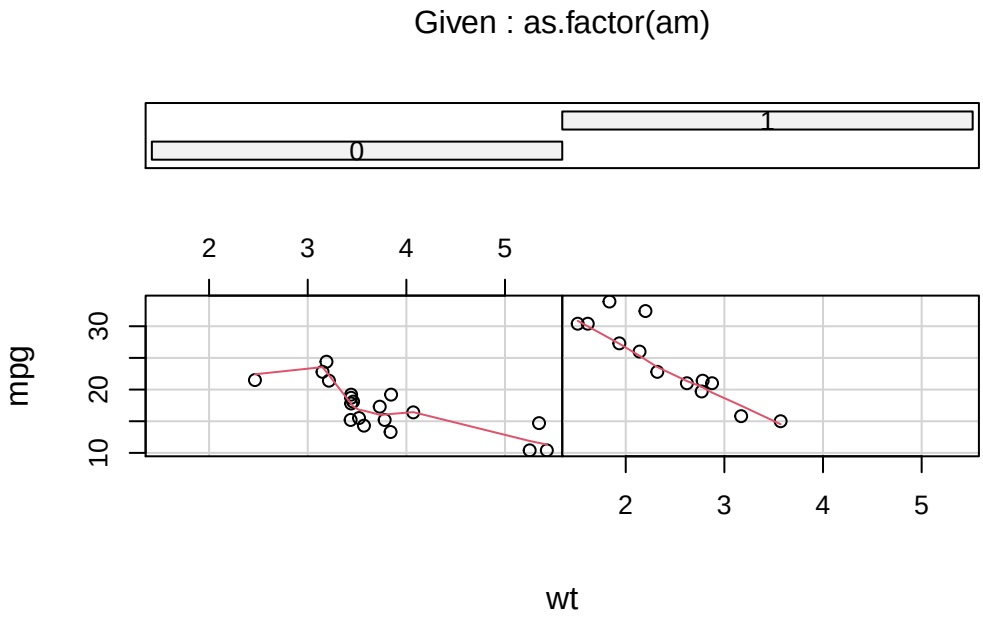
```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

A.2 Mean of the variable "mpg" for Automatic and Manual Transmission



As expected, the graph shows us that cars with a manual transmission are more efficient than other cars in automatic transmission. But also, let's explore other variables that might be correlated to understand how mpg can affect between manual and automatic transmission.

A.3 Is there a correlation between mpg and weight for Automatic and Manual Transmission?

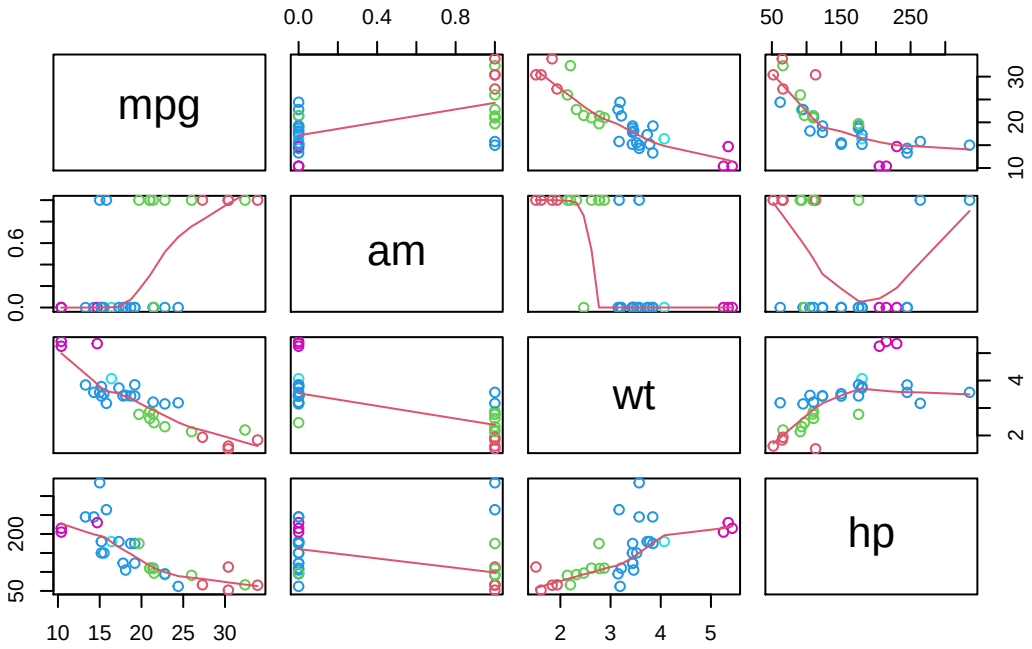


There seems to be a quite linear dependency that differs in function of the variable “am”.

A.4 Analysis of Variance (using ANOVA)

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + hp
## Model 4: mpg ~ am + wt + hp + disp
## Model 5: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 63.0133 9.325e-08 ***
## 3     28 180.29  1     98.03 13.9571  0.001219 **
## 4     27 179.91  1      0.38  0.0546  0.817510
## 5     21 147.49  6     32.41  0.7692  0.602559
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

A.5 Correlation of the variables of the Model “fit3”



A.6 Summary of the Model “fit3”

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## am           2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

A.7 Plot of the Residuals of the Model "fit3"