Motor Trend US 1974 Magazine Analysis

Vitor Pereira Barbosa

10/21/2020

1 - Introduction

This report bring valuable information about how manual or automatic transmission can influence the fuel efficiency from cars. The dataset comprises data from the year of 1974, with 10 different aspects of design and engineering from 32 automobiles.

2 - Summary

The analysis is performed as follows:

- Exploratory Data Analysis with visual tools to asses the difference in fuel efficienty due to the automatic or manual transmission

- Use of Hypoteshis test the show the difference statistical significance

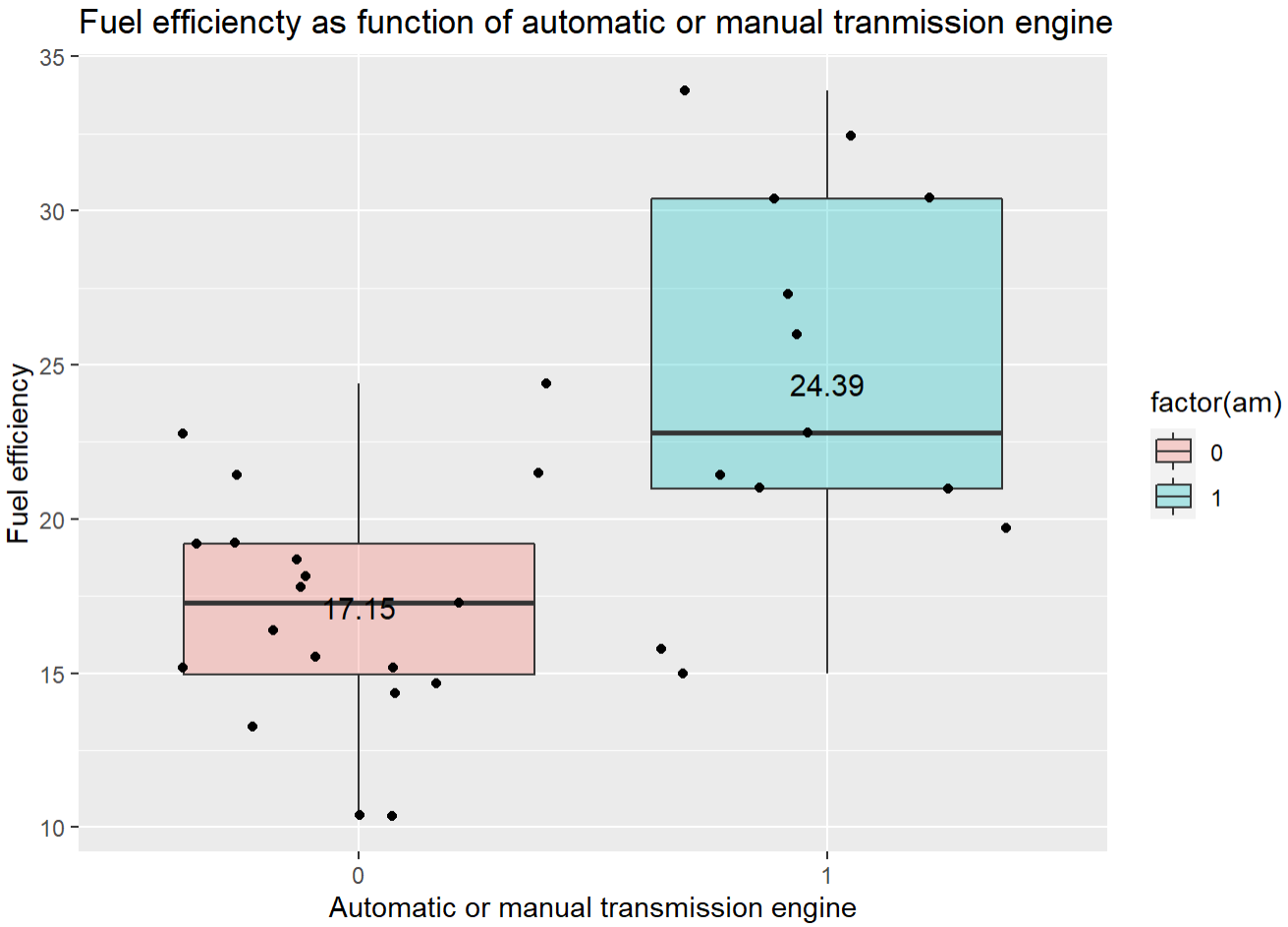
- Simple Linear Regression and Multivariate Linear Regression to quantify how much automatic or manual tranmission affects the car’s fuel efficiency

3 - Automatic or manual transmission influence over MPG

Exploratory Data Analysis

As we can see from the chart below, there is a visual difference in mpg when comparing

automatic (am = 0) and manual (am=1) transmission. The median and mean from mpg is higher in manual than in automatic transmission.



T test to asses the statistical difference from the two means

```
##
##  Welch Two Sample t-test
##
## data:  mpg by factor(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 the p-value is 0.001374, with 5 % significance level it is possible to reject the Null Hypothesis of means equality, proving a influence from transmission over fuel efficiency

4 - Linear Regression Model

```
##
## Call:
## lm(formula = mpg ~ factor(am), data = data)
##
## 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 ***
## factor(am)1    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 summary presented above, it is possible to verify that the beta0 coefficient (Intercept) represents the average value of mpg from automatic transmission (am = 0) and the beta1 coefficient (factor(am)1) represents the expected increase in mpg from using manual transmission in relation to automatic transmission.

The Pr(>|t|) corresponds to the coefficient p-values, in which we can verify that both are significant with 1 % significance level

Besides that, it is also necessary to verify the beta1 confidence level:

```
##    2.5 %   97.5 %
##  3.64151 10.84837
```

As printed above, with 95 % confidence level, it’s possible to verify increases in fuel efficiency from 3.64 to 10.84, due to car’s transmission type presented.

Variance Inflation Factor Analysis

The VIF analysis is executed to help in deciding which variables should be used in a multivariate regression analysis

```
##       cyl      disp        hp      drat        wt      qsec        vs        am
## 15.373833 21.620241  9.832037  3.374620 15.164887  7.527958  4.965873  4.648487
##      gear      carb
##  5.357452  7.908747
```

The Multivariate Linear Regression was executed several times, with three differents sets of regressors:

```
#am, drat and vs
mlm = lm(mpg ~ am + drat + vs, data = data)
summary(mlm)$coefficients
```

```
##             Estimate Std. Error  t value     Pr(>|t|)
## (Intercept) 8.326573   6.016660 1.383920 0.1773163811
## am          4.668725   1.837723 2.540495 0.0168984536
## drat        1.985086   1.882902 1.054270 0.3007724001
## vs          6.235194   1.421413 4.386616 0.0001479099
```

```
#qsec, vs and am
mlm2 = lm(mpg ~ am + qsec + vs, data = data)
summary(mlm2)$coefficients
```

```
##              Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -4.192306 10.3290346 -0.4058759 6.879189e-01
## am           7.502368  1.4569348  5.1494194 1.843358e-05
## qsec         1.096700  0.6007225  1.8256344 7.858926e-02
## vs           3.795142  2.1027890  1.8048137 8.187001e-02
```

```
#qec, wt and am
mlm3 = lm(mpg ~ am + wt + qsec, data = data)
summary(mlm3)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## am           2.935837  1.4109045  2.080819 4.671551e-02
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

As we can see, only in the third model, using “qsec”, “wt” and “am” as regressors the p-values are relevant with 5 % significance level. It is also possible to verify the confidence level from “am” coefficient

With “qsec” and “wt” included in the regression, the effect of “am” is reduced, but still relevant, with 0.05 to 5.83 increase in mpg with 95 % confidence level.

```
##      2.5 %     97.5 %
## 0.04573031 5.82594408
```

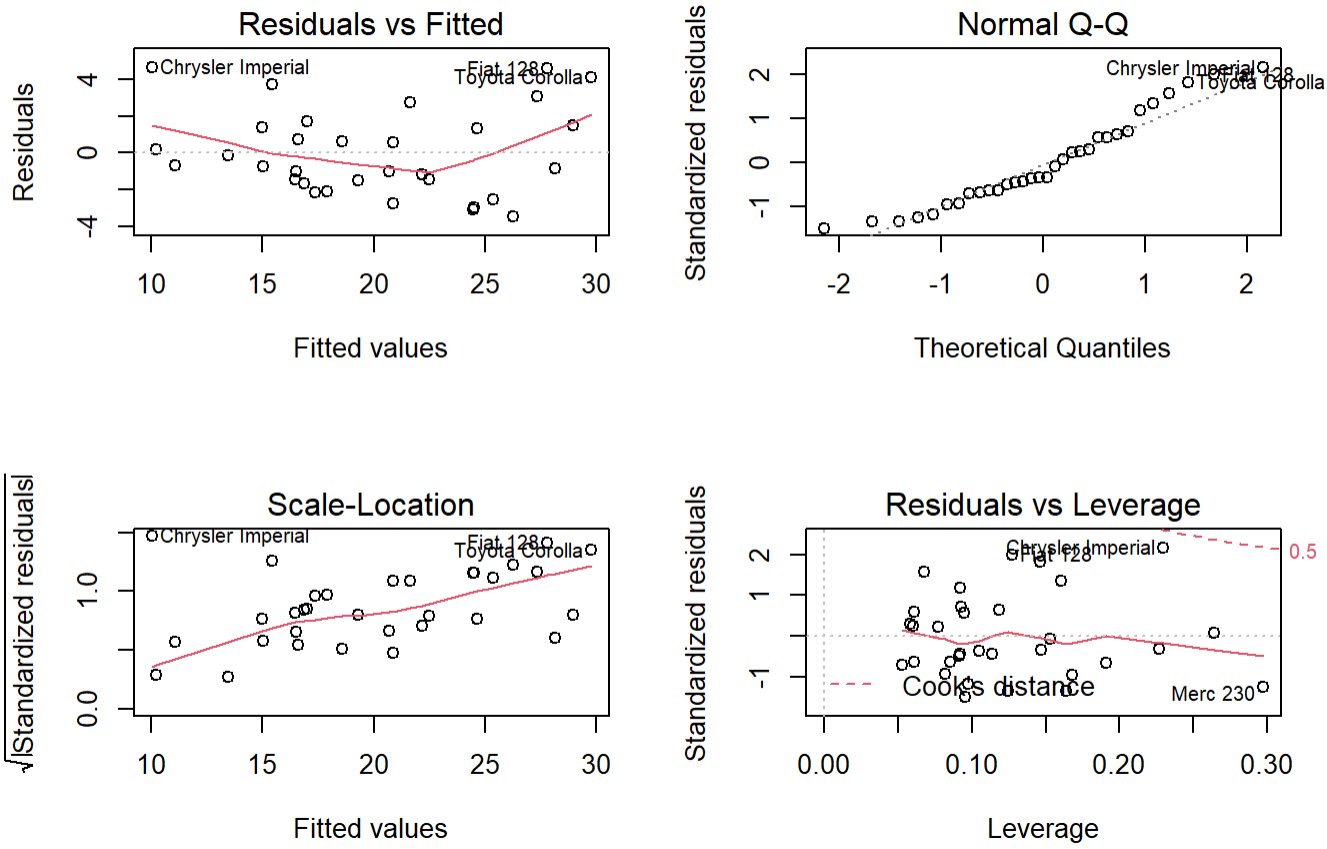
5 - Comparing models

The ANOVA analysis allows the comparison between two models, in that case to asses the difference of a new model with 3 regressors in respect of a model with only one regressor.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As presented, the second model is really different from the first one (p-value of 1.55e-09)

6 - Residual plot



Residual average value (Check Gauss-Markov assumption of linear regression)

```
## [1] 8.500145e-17
```

• Looking at the residual plots chart, the first one on the right shows a random data distribuition, and calculating the residual mean, the value is 8.8500145e-17, which respects the Gauss-Markov assumption of linear regression of residual mean value of zero.

• The Quantile-Quantile plot is satisfactory with only a slighly deviation at the far left and far right sides.

• In fact the Standardized residuals plot shows a slighly systematically distribuition which was not explained by the model

• In the Residuals vs Leverage plot, there are two cars with leverage values above 0.25

7 - Conclusion

The manual transmission is better for fuel efficiency and it is possible to quantify how much it influences the “mpg” variable, with only one variable (the “am” regressor), affecting from 3.64 to 10.85 and also with the weight (“wt”) and acceleration measure (“qsec”) affecting from 0.05 to 5.83.