Regression Models Project : Motor Trend analysis

Tanmoy Rath

31 August 2018

I. Executive Summary

This report studies the mtcars dataset and tries to explore the relationship between dependent variable **mpg** i.e "Miles/(US) gallon" and independent variable **am** i.e Automatic or Manual Transmission by answering:

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

In this report 3 models were tested, along with their residual plots and variance inflation factors in order to answer the above questions. We also performed a t-test and analysis of the p-values to know the statistical significance of our results.

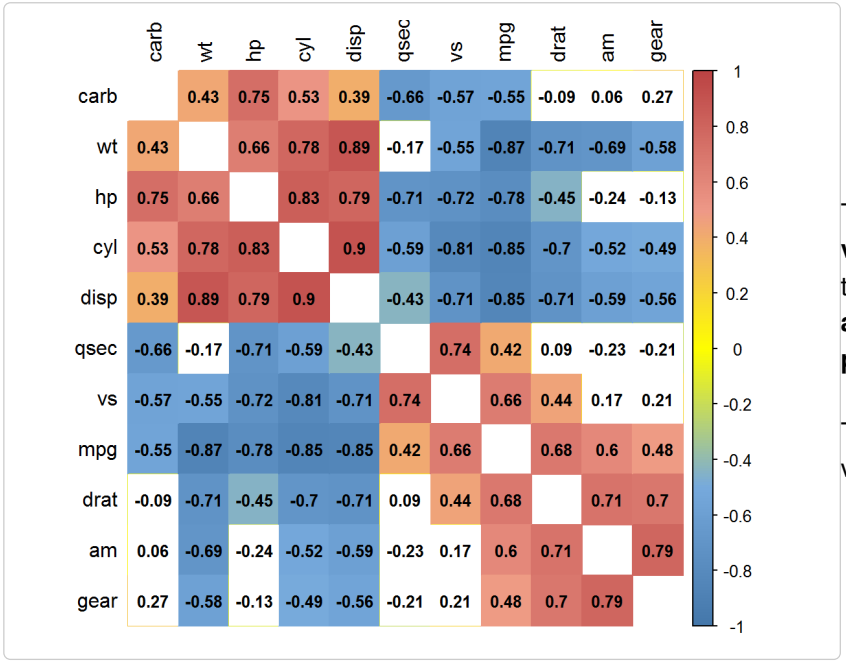
II. Exploratory Analysis

The dataset **mtcars** was found to have no NA values.

```
data("mtcars")
c(sum(is.na(mtcars$mpg)), sum(is.na(mtcars$cyl)), sum(is.na(mtcars$disp)), sum(is.na(mtcars$hp
)), sum(is.na(mtcars$drat)), sum(is.na(mtcars$wt)), sum(is.na(mtcars$qsec)), sum(is.na(mtcars$
vs)), sum(is.na(mtcars$am)), sum(is.na(mtcars$gear)), sum(is.na(mtcars$carb)))
```

```
##  [1] 0 0 0 0 0 0 0 0 0 0 0
```

Given below is the correlation matrix for the whole mtcars dataset.



The cells with **high +ve correlation values are coloured maroon**, while those with **high -ve correlation values are coloured blue**. Cells whose **p-values are < 5%, are coloured white**.

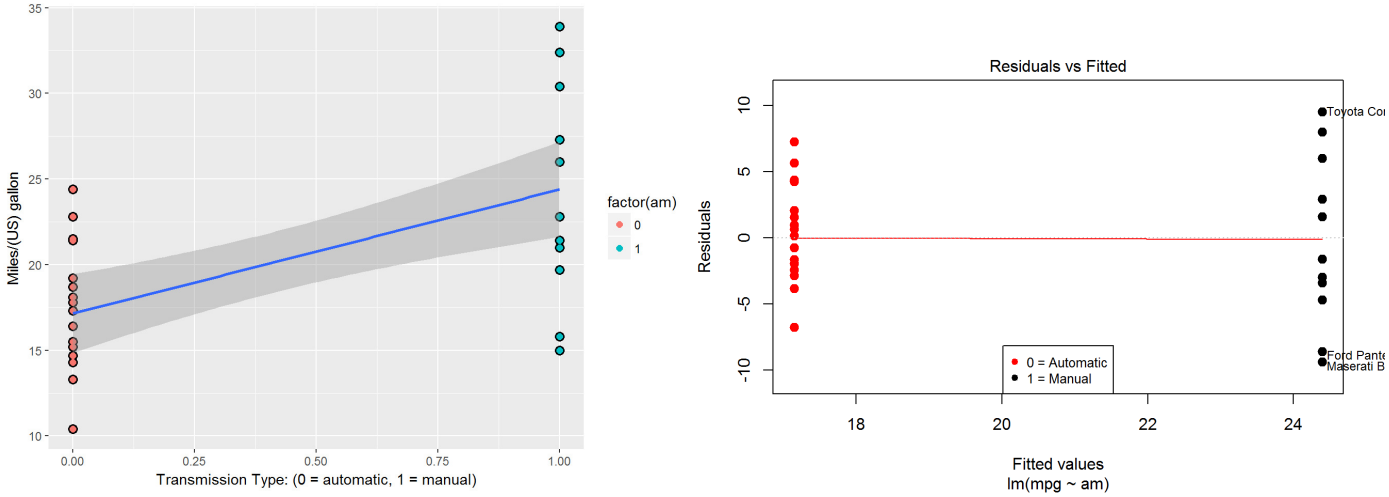
That means, only colours of correlation values that are significant are visible.

★ The correlation of am and mpg being **0.6** states that **am has moderate prediction ability over mpg**. In other words, am alone doesnot fully determine mpg. There are other and even stronger factors because of the presence of much higher correlation values.

III. Regression Models

A1. Model 1 : mpg ~ factor(am)

The plot of **mpg** ~ **am** and its **residual plot**.



★ The model plot clearly shows that **manual transmission gives more mileage** than automatic transmission. Also we find **no patterns** in our residual plot. However they don't say if the difference is significant or not. So we conduct a t-test, to know its statistical significance.

★ There seems to be a **slight amount of heteroscedasticity** in the residual plot because the points on the right, seem to have a little more variance than those on the left.

A2. T-test

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg[AT] and mtcars$mpg[MT]
## 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 of x mean of y
##  17.14737  24.39231
```

★ The **p-value(=0.001374)** at 95% confidence level states that the **difference (=7.24494)** in the averages or means of mpg for automatic and manual transmission **is statistically significant.** The individual means of mpg for automatic and manual transmission are **17.14737** and **24.39231** respectively.

A3. Coefficient Interpretation

The coefficients of the model are found to be:

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

The coefficient for automatic transmission (=**17.147368**), which **is the intercept** is the average of mpg for vehicles of automatic transmission only. The coefficient for manual transmission (=**7.244939**) **is the difference between** the averages of mpg for manual and automatic transmission. The mean for manual transmission is 24.39231.

B1. Model 2 : mpg ~ factor(am) + qsec + carb (Refer Appendix for plot & residual-plot)

★ The model plot clearly shows that **manual transmission gives more mileage** than automatic transmission. However in the residual plot, we find a rough **u-shaped pattern** which is undesirable.

B2. Variance Inflation Factors

The Variance Inflation Factors are found to be:

```
## factor(am)       qsec       carb
##   1.073110   1.878603   1.785255
```

★ The Wikipedia article here (https://en.wikipedia.org/wiki/Variance_inflation_factor#Step_three) , states that a **VIF below 5 is of negligible correlation**. Since all the VIF's are < 2, we can safely say **the regressors are uncorrelated**.

B3. Coefficient Interpretation (Refer Appendix for coefficient table)

The coefficient for automatic transmission (=**0.3252711**), which **is the intercept** is the average of mpg for vehicles of automatic transmission only. The coefficient for manual transmission (=**8.4353493**) **is the difference between** the averages of mpg for manual and automatic transmission.

★ The **p-value** for coefficient **factor(am)1**, is **statistically significant**. Hence manual transmission indeed provides more mileage than automatic transmission.

C1. Model 3 : mpg ~ (wt*factor(am)) + qsec (Refer Appendix for plot & residual-plot)

★ The model plot again clearly shows that **manual transmission gives more mileage** than automatic transmission. The residual plot has **almost no pattern**. Furthermore, we find **no sign of heteroscedasticity**.

C2. Variance Inflation Factors

The Variance Inflation Factors are found to be:

```
##            wt    factor(am)          qsec wt:factor(am)
##      3.030963     20.970925     1.447406     16.302453
```

★ Two VIF's suggest there is **very high correlation** among the regressors. The effect is compounded because of interaction terms in the model, especially when the interaction terms contain highly correlated regressors. However, since our residual plot has no patterns, this model is **better at prediction** than the previous models.

C3. Coefficient Interpretation (Refer Appendix for coefficient table)

The coefficient for automatic transmission (=**9.723053**), which **is the intercept** is the average of mpg for vehicles of automatic transmission only. The coefficient for manual transmission (=**14.079428**) **is the difference between** the averages of mpg for manual and automatic transmission.

★ The **p-value** of this difference is **statistically significant**. Hence manual transmission indeed provides more mileage than automatic transmission.

IV. Conclusion

The difference in means for automatic and manual transmission for all the 3 models are found to be **7.24494**, **8.4353493** and **14.079428** with their p-values being **2.850207e-04**, **5.423664e-08** and **0.0003408693**. We see that all are **statistically significant**. Hence we can conclude that

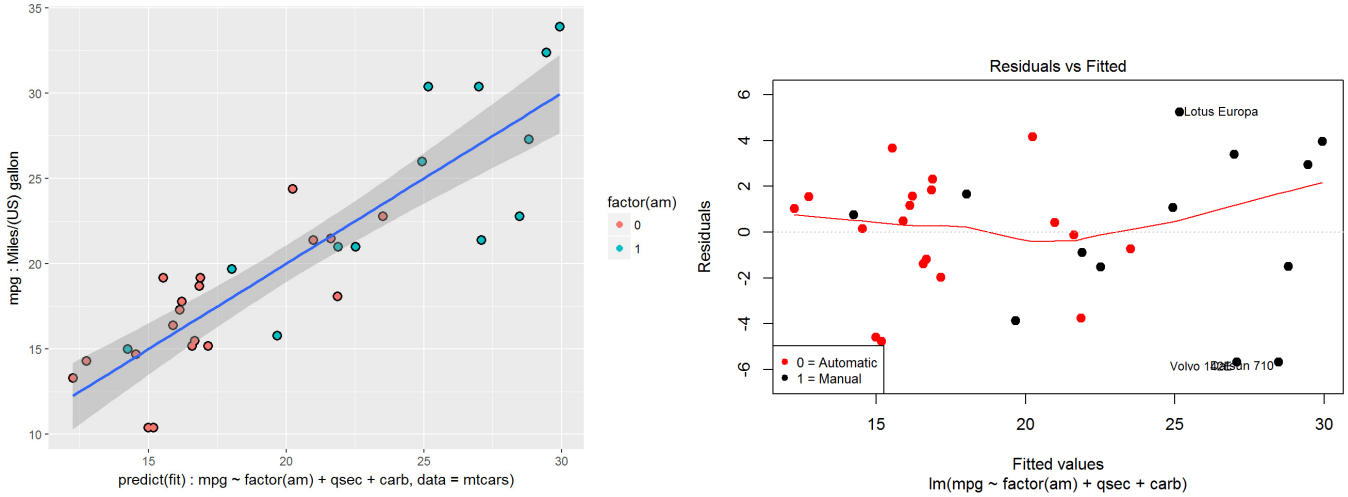
- ★ **Manual transmission is indeed better for MPG than automatic transmission**.

We also see that, the difference in means varies across the models studied. This happens because mpg, also depends on other factors such as wt, cyl, etc. Including / excluding them changes the difference in means. Hence

- ★ **The difference in means depends on the specific regression model being analysed, hence can't be quantified.**

V. Appendix

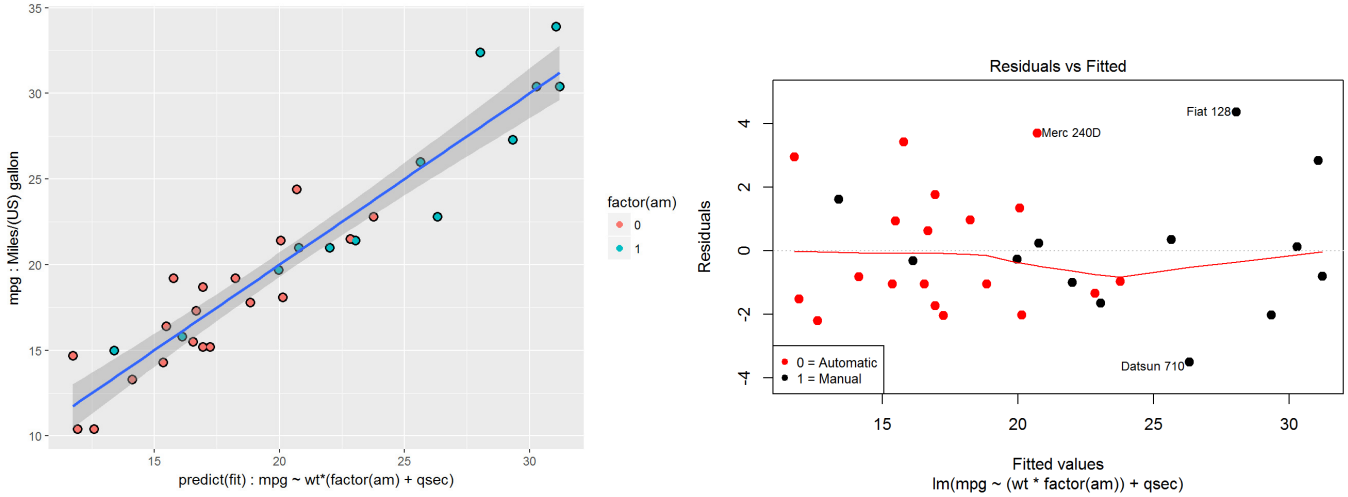
The plot of **mpg ~ factor(am) + qsec + carb** and its **residual plot**.



The coefficients of the model are:

```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)   0.3252711  8.6306964  0.0376877 9.702041e-01
## factor(am)1   8.4353493  1.1492472  7.3398913 5.423664e-08
## qsec          1.1332876  0.4246104  2.6690057 1.251399e-02
## carb         -1.3828531  0.4579391 -3.0197318 5.349452e-03
```

The plot of **mpg ~ (wt × factor(am)) + qsec** and its **residual plot**.



The coefficients of the model are:

```
##                  Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)      9.723053  5.8990407  1.648243 0.1108925394
## wt              -2.936531  0.6660253 -4.409038 0.0001488947
## factor(am)1     14.079428  3.4352512  4.098515 0.0003408693
## qsec             1.016974  0.2520152  4.035366 0.0004030165
## wt:factor(am)1  -4.141376  1.1968119 -3.460340 0.0018085763
```