Regression Models Course Project

Analysis of Influesnce of Transmission Type (automatic vs. manual) on Fuel Consuption Based on *mtcars* Dataset

T-StrawClown

February 26, 2016

Executive Summary

This report analyses which type of transmission, manual or automatic, is better and how much of an influence it has on MPG (miles per galon). All results are based on analizing the famous *mtcars* dataset using Regression Models. Without getting into details it may look that transmission might be major factor for MPG - the average MPG for cars with automatic transmission is **17.15** and **24.39** for cars with manual transmission. However, after some additional exploration this report concludes that transmission type has no significant impact on MPG. Inclusion of additional factors into analysis diminishes the difference in MPG when comparing automatic and manual transmission and it can't be considered reliable. More data samples should be obtained in order to be able to quantify influence of transmission type on MPG.

Modelling

Regression model with transmission type as a single regressor will create just 2 points of predicted values (means of MPG for respective type of transmission) which might seem influential. 95% confidence level Student's test also confirms that intercept term **17.15** of such model (wich actually is predicted value for all cars with automatic transmission, since *am* variable is binomial) is between **14.85** and **19.44**, and change of intercept term **7.24** for cars with manual transmission is between **3.64** and **10.85**, which means that with 95% confidence we can conclude that switching from automatic to manual transmission we would get **7.24** better MPG. Distribution of residuals of such model is almost perfectly normal (Figure1 in Appendix).
This seems suspicious that transmission type alone has such a big impact on MPG, so let's explore if there are any other variables worth considering by our model. From Figure 2 in Appendix we can conclude that heavier cars with bigger and more powerful engines tend to have automatic transmissions installed and lighter and less powerful cars are equiped with manual transmissions; most of 4 cylinder cars are manual and most of 8 cylinder cars are automatic. It also can be seen that weight, power and number of cylinders have high relationship directly with MPG themselves, so we are going to include them into our model and see what results we get.
Inclusion of additional properties into the model and comparing how did that change the influence of transmission type on MPG results the following:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl + wt
## Model 4: mpg ~ am + cyl + wt + hp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 264.50  2    456.40 39.2861 1.388e-08 ***
## 3     27 182.97  1     81.53 14.0354 0.0009026 ***
## 4     26 151.03  1     31.94  5.4991 0.0269346 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It inticates that adding properties of car size and engine overall improves our model, as sum of squared residuals decreases with each new property added, while maintaining high level of significance of added properties.
Verifying coefficients of linear regression model which contains centered weight, centered power, transmission type, and discretized number of cylinders gives the following summary:

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          20.965      1.568  13.375    0.000
## am1                   1.809      1.396   1.296    0.206
## cyl6                 -3.031      1.407  -2.154    0.041
## cyl8                 -2.164      2.284  -0.947    0.352
## I(wt - mean(wt))     -2.497      0.886  -2.819    0.009
## I(hp - mean(hp))     -0.032      0.014  -2.345    0.027
```

Results and Conclusions

Since number of cylinders seems to be not too reliable (very low confidence level for 8 cylinder cars), it was removed from model. Coefficients of the model now are:

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          19.244      0.717  26.845    0.000
## am1                   2.084      1.376   1.514    0.141
## I(wt - mean(wt))     -2.879      0.905  -3.181    0.004
## I(hp - mean(hp))     -0.037      0.010  -3.902    0.001
```

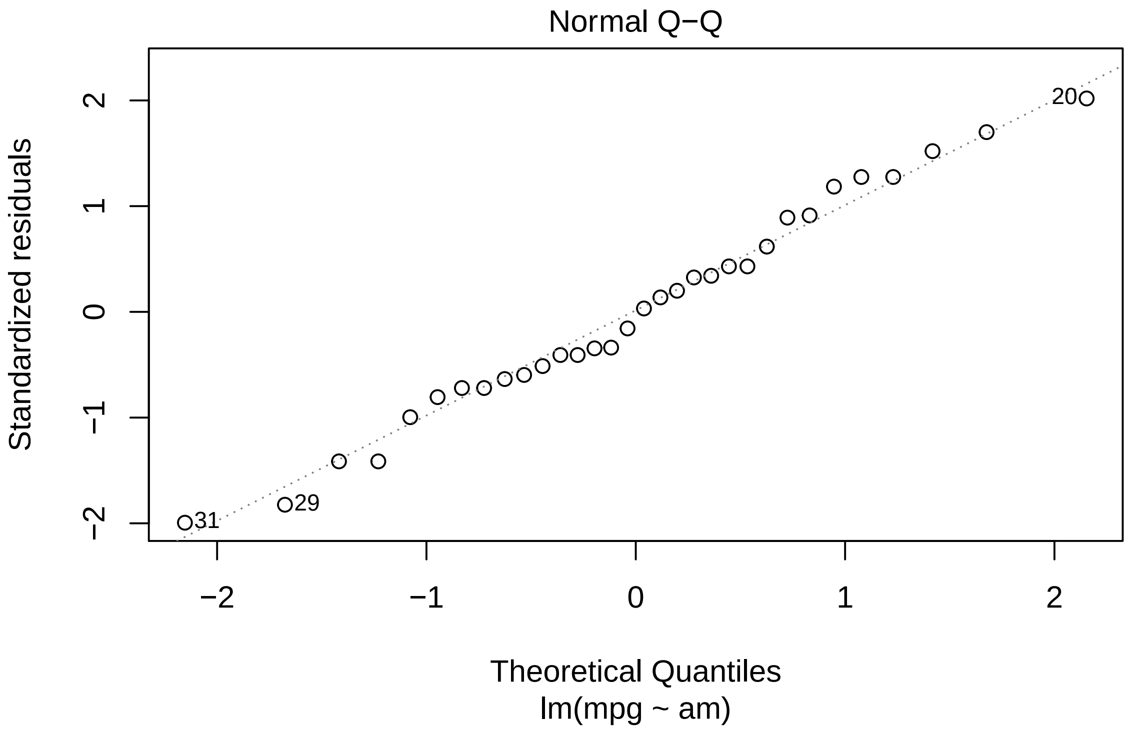
Diagnostics (Figure 3 in Appendix) of created model don't seem to have any particularly bad data points, some of them (17, 18 and 20) have maybe a bit high influence, but removing them would actually even more diminish influence of transmission type on MPG.

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          19.184      0.520  36.887    0.000
## am1                   0.506      1.040   0.486    0.631
## I(wt - mean(wt))     -3.750      0.712  -5.265    0.000
## I(hp - mean(hp))     -0.028      0.007  -3.920    0.001
```

Removing those records is not necessary (it's an overkill anyway), since even with them included, we are confident that:
* increase of wheight by 1000 lbs of average power car equiped with an automatic transmission, results 2.879 decrease in MPG;
* increase of 1 hp of average weight car equiped with an automatic transmission, results 0.037 decrease in MPG.
It might look that changing transmission type of average wheight, average power car from automatic to manual results 2.084 increase of MPG, however 95% confidence interval of this coefficient is from -0.736 to 4.903 and, as it contains 0, we have to conclude that such result is unreliable (we are only ~86% confident). Thus we can't confirm that transmission type has impact on MPG and, of course, we can't quantify it.

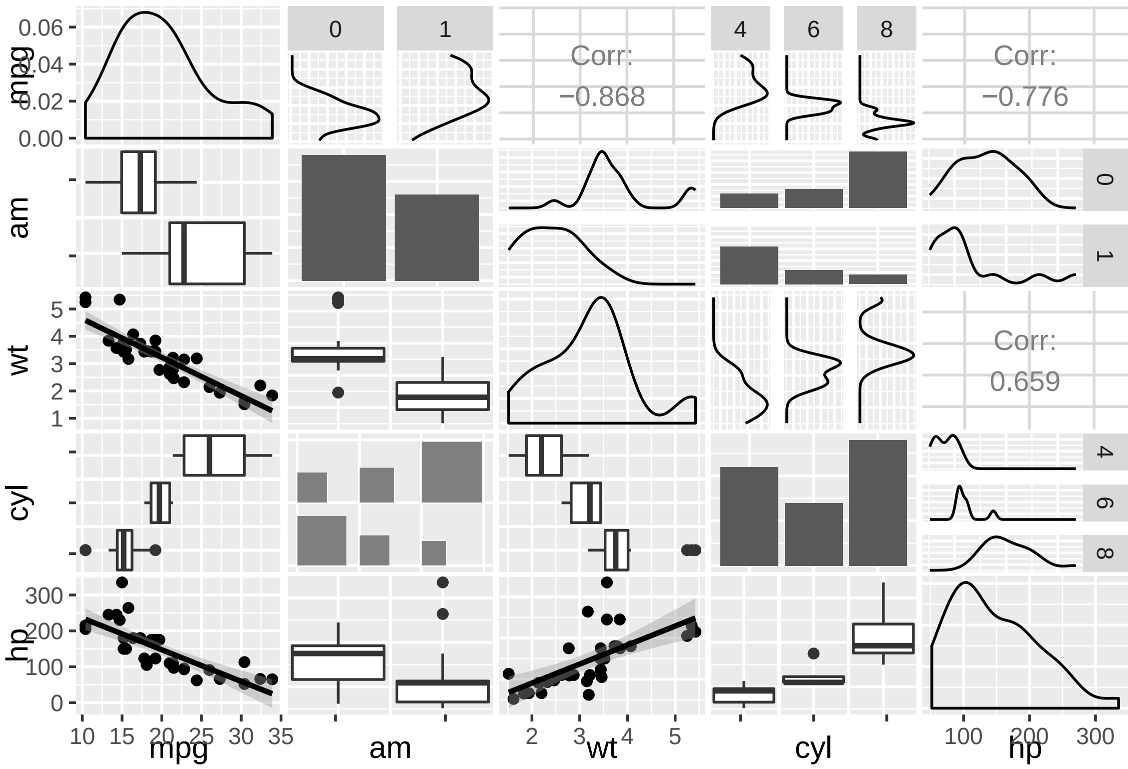
Appendix

Figure 1: Residuals Normality



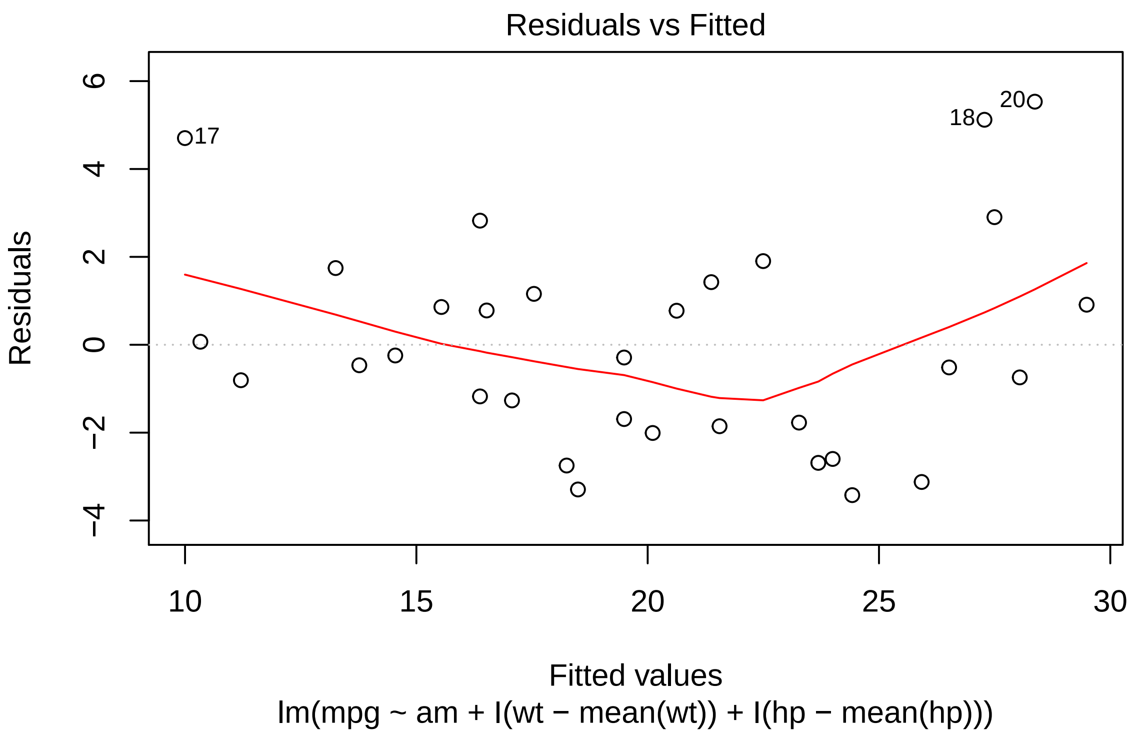
.

Figure 2: Correlation Chart of Extended Model



.

Figure 3: Diagnostics of Final Model



!!!

R Markdown file reproducing this report can be found here https://github.com/T-StrawClown/RegressionModels/blob/master/course_project.Rmd