Exploring the Dependence of Fuel Economy on Type of Transmission

Sergey Cherkasov

18 Sep 2015

Synopsis

In this paper we are going to explore the relationship between a set of variables and miles per gallon (MPG) (outcome). We use mtcars dataset, provided by The Comprehensive R Archive Network and can code book can be found here - http://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html.

The client is 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

Original Rmd file could be found at the folloing GitHub repository - https://github.com/PestoVerde/RegModels_PeerAssessment1.

Exploratory analysis

First let us investigate the data. It describes 32 cars with 11 variables. These are mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb. The parameters we are interesting for are mpg and am.

Is an automatic or manual transmission better for MPG?

Mean for automatic transmission is 17.1 mpg whereas mean for manual transmission equals 24.4. Ranges are 10.4, 24.4 and 15, 33.9, they do not even intersect.(see also boxplot in Appendix 1).

The dependence looks significant, and Student's t-test comfirms that, with p-value:

```
## [1] 0.001373638
```

Quantify the MPG difference between automatic and manual transmissions

Since the cilents is interested in dependance of MPG on type of transmission, we build the simple linear model mpg ~ am, having am variation as a factor. The model quantify the MPG difference as:

```
## as.factor(am)1
##       7.244939
```

That is the cars with auto transmission have average 7.2 mpg worse than cars with manual one. However the quality of the model is not really good because R-squared 0.3597989, that is explain only 36% of variance (see details in Appendix 2).

From the common sence we can suppose that apart from type of transmissions there are other variables which can have an influence on fuel economy. The weight of a car (wt) and displacement of a engine (disp) look promising. Let us have quick look on the model like mpg ~ am + wt + disp. New model quantify the MPG difference as:

```
## as.factor(am)1
##      0.1777241
```

Now the cars with auto transmission have average 0.2 worse than cars with manual one. The new model have R-squared 0.7810427, now it explains 78% of variance which is much better prediction (details are in Appendix 3).

Now let us compare our first simple linear model and the multivariative model. There are signicantly different with the p-value

```
## [1] 2.996421e-07
```

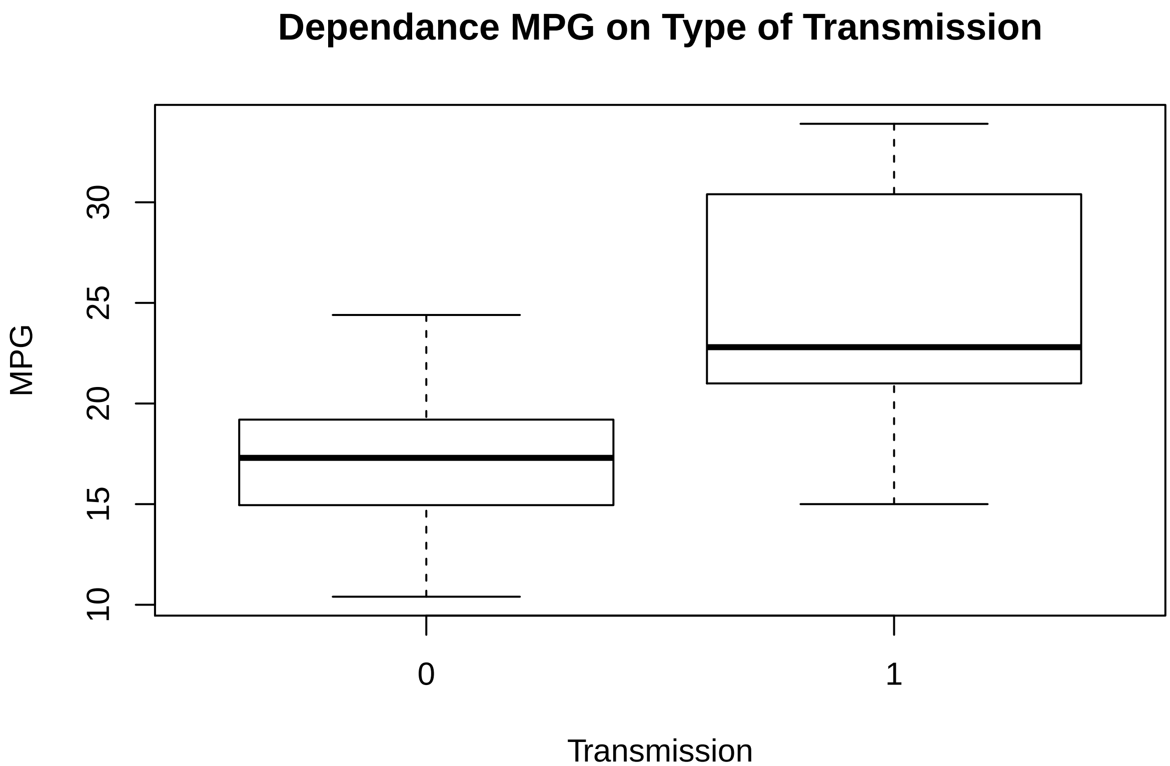
(see details in Appendix 4). Checking the model's residuals can be found in Appendix 5. It shows they are normally distributed and have a homoskedasticity.

Conclusion

We can insist that cars with manual transmission have better fuel economy. But simply dependance does not perform really good. As multivariative model shows, the fuel economy depends mostly on weight. That is we can not give a simple answer regarding two types of transmission and fuel economy. There is a necessity of more complex analysis and more complicated models.

Appendices

Appendix 1. Exploratory analysis of data. Different in mpg depending on type of transmission.



Appendix 2. Summary of simple linear model.

```
##
## Call:
## lm(formula = mpg ~ as.factor(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 ***
## as.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
```

Appendix 3. Summary of multivariative model.

```
##
## Call:
## lm(formula = mpg ~ as.factor(am) + wt + disp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4890 -2.4106 -0.7232  1.7503  6.3293
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    34.675911   3.240609  10.700 2.12e-11 ***
## as.factor(am)1  0.177724   1.484316   0.120   0.9055
## wt             -3.279044   1.327509  -2.470   0.0199 *
## disp           -0.017805   0.009375  -1.899   0.0679 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.967 on 28 degrees of freedom
## Multiple R-squared:  0.781,  Adjusted R-squared:  0.7576
## F-statistic: 33.29 on 3 and 28 DF,  p-value: 2.25e-09
```

Appendix 4. Comparing two models.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ as.factor(am)
## Model 2: mpg ~ as.factor(am) + wt + disp
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 246.56  2    474.34 26.934 2.996e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 5. Residuals and diagnostic of multivariative model.