Transmission vs. miles per gallon

Luis Santos

27 September 2015

Does the transmission type affect vehicle consumption?

By directly comparing the average fuel consumption (average MPG) from cars with automatic gearboxes versus cars with manuals, we could say there is a significant difference between the two: the manuals achieve an average of 24.39 MPG whereas the automatics reach a modest 17.15 - so in total, a significant 7.24 MGP more, for the manuals (Student's T test p-value of 0.001). However, this is far too simple of an approach to the question at hand, and a more evolved linear model has revealed that, although there are some differences in fuel economy between automatic and manual cars, they are not as significant: in fact, when taking into account weight, engine size and number of carburetors, there is only a gain of 1.8 MPG when using a manual gear box versus using an automatic one, which is not significant.

The detailed report

By analysing the cars dataset from the R package `mtcars`, we will try to clarify whether transmission type (automatic or manual) affect car fuel economy, also known as miles per gallon. The dataset contains a series of values for variables like vehicle weight (in lb/1000), miles per (US) gallon (mpg), engine size (displacement, in cubic inches - which we converted to litres), power, engine configuration (v shape/straight), transmission type and number of carburetors, amongst others. Figure 1 shows a series of plots of some of these variables against MPG, in order to understand some of their relationships.

It seems to be apparent that transmission alone does have an effect on MPG, but we can also confirm other suspicions: that bigger engine sizes are more thirsty, and so are heavier cars, and - apparently - cars with a V-shape engine.

Moving on to the statistics: after testing if the average MPG for manual cars was not different from the average MPg of automatics (using a T test, which yielded that they are different - p-value of 0.001), we then proceeded to use the simplest model for analysis, which also derives from the top left plot in Figure 1, and performed a simple linear regression by fitting transmission to MPG. The result (corroborating that of the T-test) was that the automatic gear boxes have an average of 17.15 MPG, and the manual ones achieve 7.24 MPG more - so 24.39 MPG. (Do note we have created a copy of the data set, called `cars`, in which the variable that describes the transmission type has had it values converted to `manual` and `automatic`.)

```
## 
## Call:
## lm(formula = mpg ~ transmission, data = cars)
## 
## 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 ***
## transmissionmanual    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
```

The fact that our value of R^2, which indicates how much variance in MPG values is explained by the model, is quite low indicates that indeed transmission alone does not account for the differences in MPG.

From the plots in Figure 1, we can determine that some of the most relevant characteristics of a vehicle affecting fuel consumption are its weight, engine capacity and number of carburetors (variables `wt` , `disp` and `carb` , respectively, in the data set). So, after properly converting the values for engine size from cubic inches to litres, we have modified our statistical model to include these variables.

First, for each of the select three variables, we fit them as confounders to transmission, using `lm(mpg ~ wt + transmission, data=cars)` for weight, and so on. The final results are:

```
##      Coef value for Transmission  R^2
## wt                        -0.024 0.75
## carb                       7.653 0.70
## disp                       1.833 0.73
```

We can see that they all explain around 70 to 75 percent of the variability of the MPG data, but their effect is not coherent: when holding weight as a fixed effect, manual transmissions seem to perform not as well as automatic ones, with a 0.024 *decrease* in average MPG, while for the number of carburetors and engine size, if they instead are the fixed effects, then average MPG increases 7.653 and 1.833, respectively.

However, if these variables are all playing their part in the results, then we can use a statistical model that includes them together. Our linear model will now be `lm(mpg ~ wt + disp + carb + transmission, data=cars)` .

This last model accounts for 84% of the variability of the MPG values, and from it, we learn that, if we hold the values of weight, carburetors and engine size, then we expect that using a manual transmission should improve the average MPG by 2.73. For diagnostic purposes, we plot the resuls of the fit in Figure 2 of the Appendix.

As a final test, we have ascertained how beneficial it was to have used these models, with more confounders and complexity. For that, using ANOVA we can test for H0 : there is no difference between models:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ transmission
## Model 2: mpg ~ wt + transmission
## Model 3: mpg ~ wt + disp + carb + transmission
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 65.0690 1.147e-08 ***
## 3     27 183.64  2     94.67  6.9597  0.003651 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results tell us that there was a significant improvement in using any of the `lma` models versus the initial one, and that the last model to be used, in which we included all variables selected, was even better.

Appendix

Figure 1 How different variables may influence fuel consumption, along with the transmission type

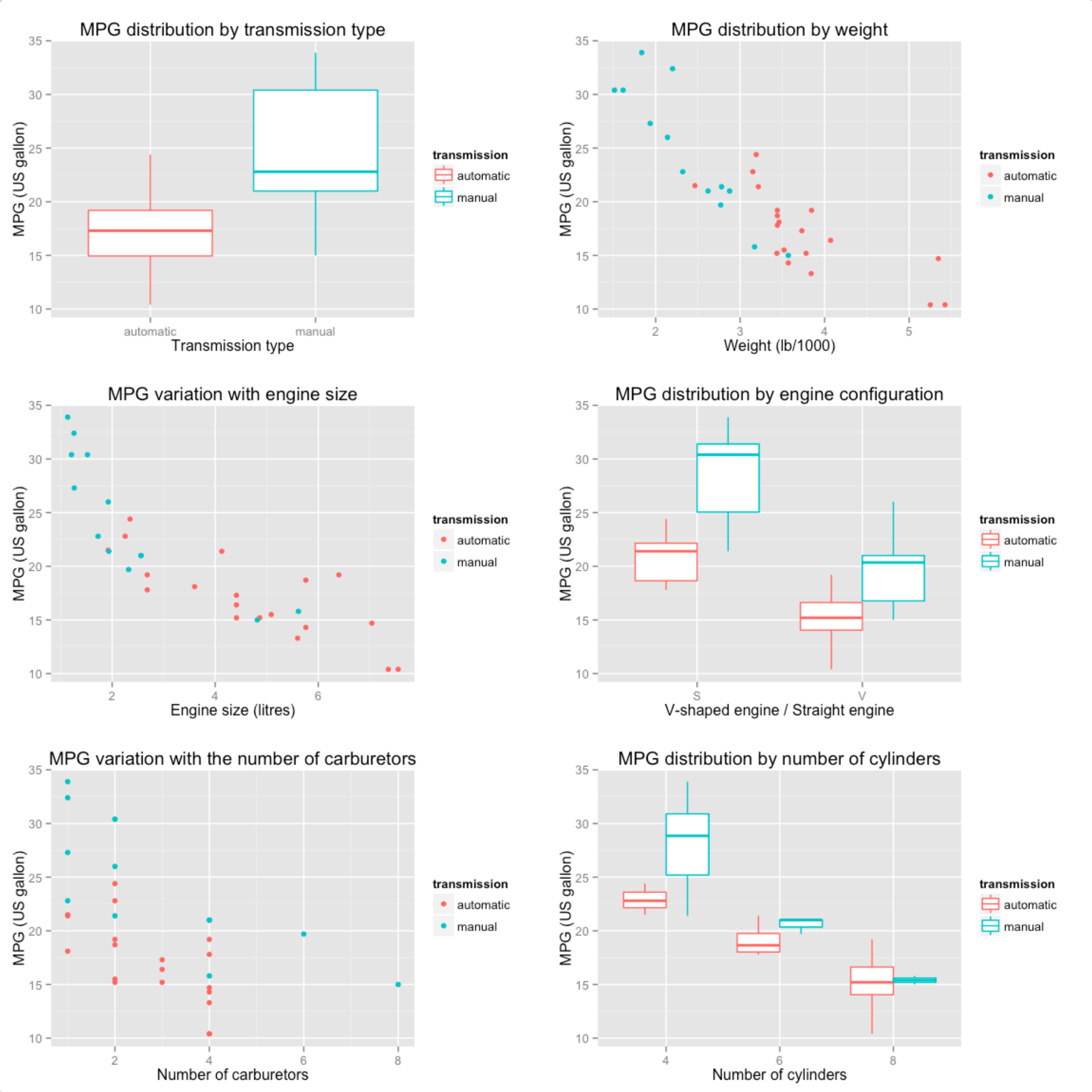

Figure 2 Diagnostic plots for the multivarible linear model