

Regression Models: Motor Trends

noootsab

2014-08-16

Summary

Analyse cars data to detect if some difference exists between gas consumption of manual cars, or automatic ones.

For that, we'll use the data from mtcars which containing 11 measures for 32 cars from the years 73-74.

In the following sections we'll explain why we think that the difference between the manual and automatic transmission is statically significant. But also, we'll see that the weight has a dramatic importance when choosing a car's transmission regarding mpg.

In fact, when the car is light enough (~1000lb), manual transmission is initially at 16.8 mpg which is around 10 more than for automatic ones. However, when the weight increases, there is a decrease of the mpg of (around) -7 for the manual cars, and only -3 for the automatic ones. Thus, starting, with cars more than 3500lb, the automatic cars should be the choice of preference.

As a side note, if we had only looked at mean within group we would have always preferred manual cars to automatic, but this is influenced by the fact that the dataset has a few number of lighter cars and all are automatic! See (fig-3).

The code can be found on [github](#).

Analysis

First of all we're going to load the data, `mtcars`, available in R's package `datasets`. Then we'll head into to look at what it contains (and adapt the types if necessary).

```
##           mpg cyl disp  hp drat   wt  qsec vs      am gear
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0   Manual    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0   Manual    4
## Datsun 710      22.8   4  108  93 3.85 2.320 18.61  1   Manual    4
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44  1 Automatic    3
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0 Automatic    3
## Valiant        18.1   6  225 105 2.76 3.460 20.22  1 Automatic    3
##           carb
## Mazda RX4      4
## Mazda RX4 Wag  4
## Datsun 710      1
## Hornet 4 Drive  1
## Hornet Sportabout 2
## Valiant        1
```

So there are 32 observations of cars and 11 measures.

We're interested in the role playing by the transmission (`am`) in the evolution of miles/gallon consumption (`mpg`).

To see a visual interpretation of the relation between the two, a boxplot is available in the appendix (fig-1). Still visually, it looks like the difference between the two groups (manual and automatic transmissions) is

true, let's see the mean of each first and then perform a between two-groups t-test (assuming normality and independence).

Sounds good, the difference between both is significant (p-value 0.0014) and the 95% confidence interval doesn't contain 0, and thus, at this stage, we can say that the manual cars are better than automatic cars from 3.2097 to 11.2802 miles per gallon.

However, using the transmission alone is not enough to quantify the difference for specific cases. To see that, we can look at the prediction of mile per gallon using the single transmission independent variable in a linear regression.

So the relation between the two is rather clear, with p-value of the change being 2.8502×10^{-4} . However, the variance explained is quite low, with an R^2 at 0.3598. So something is missing in the mix.

In order to find another model explaining better the miles per gallon, we'll use the best model selection since our dataset is quite small. For the sake of sanity, models selected by the stepwise method (forward and backward) have been ran, and they were only diverging at the third variable selecting `hp` or `qsec`. To run these selections, we'll use the useful package `leaps`.

After playing around, `wt` is the first candidate to try out. A visual representation on how `mpg` is related to both `am` and `wt`, a plot has been provided in the appendix fig-2. The linear regression involving both is increasing R^2 to 0.7528, however the effect of the change of `am` is not more significant. To solve that, we'll will increase the model by adding the `qsec` (based on the model selection results above). Now the model has a R^2 of .

Nevertheless, the fig-2 is more or less showing an interaction between `wt` and `am`. So, we're going to add this interaction to the model and check if it's significant.

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.723      5.899   1.648 0.1108925
## wt           -2.937      0.666  -4.409 0.0001489
## amManual     14.079      3.435   4.099 0.0003409
## qsec          1.017      0.252   4.035 0.0004030
## wt:amManual  -4.141      1.197  -3.460 0.0018086
```

That's interesting, not only the coefficient are all significant but the new interaction term is explaining more intuitively how the weight of a car is affecting the miles per gallon consumption when it's a manual or an automatic. That's to say, a manual car is worst by a factor of -4.1414 per 1000lb increase in weight. Last but not least, this regression is showing a pretty good residual plot and doesn't present evidence of outliers - see figures after fig-4, specially the Cook's distance.

Now we can test how this new variable (interaction) in the model is significant to explain the variance, for this we can run an anova.

Great, sounds like it's fair enough to include it, since the p-value is 0.0018!

For the sake of sanity, we can have a quick look at the VIF of the models, using the `vif` function in the `car` package. Without interaction, we have this very good VIF:

```
##      wt  qsec   am
## 2.483 1.364 2.541
```

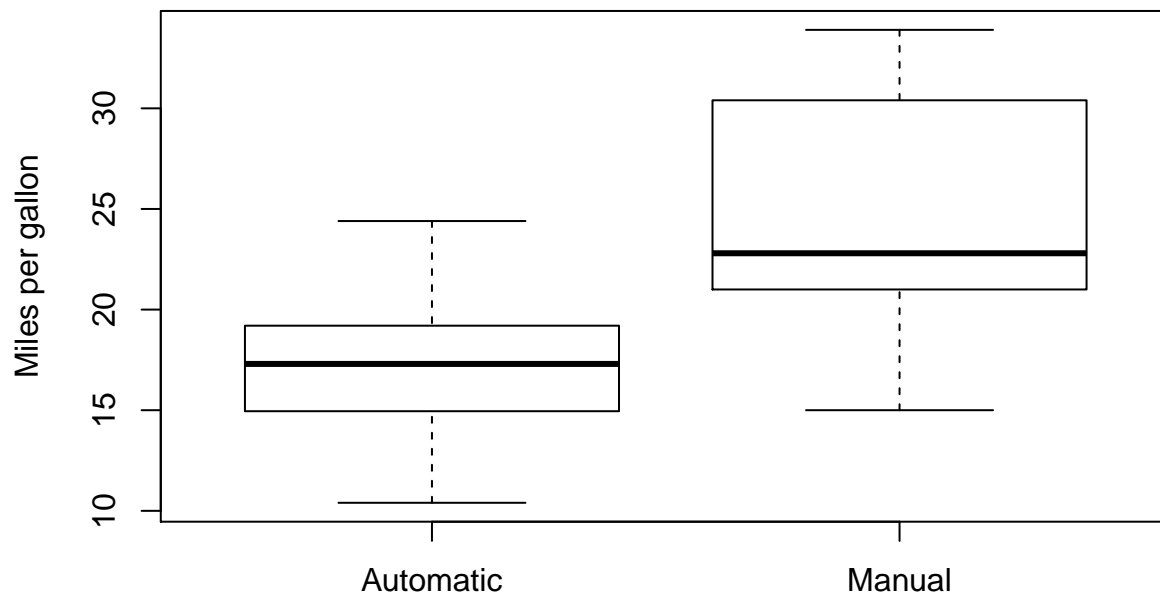
However, with the interaction we have this one:

```
##      wt      am    qsec  wt:am
## 3.031 20.971  1.447 16.302
```

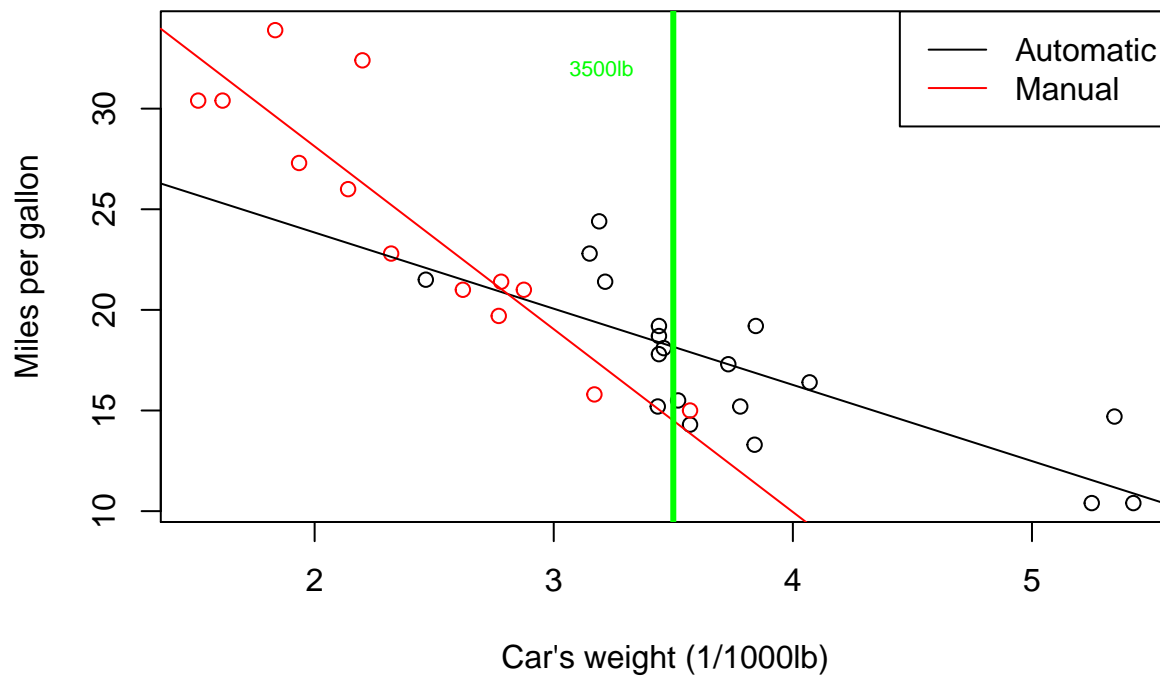
Okay, looks like the model including the interaction inflates the variance due to colinearity, but we could have foreseen it regarding fig-2. However, still it's inclusion allow better explanation of the difference between the groups.

Appendix

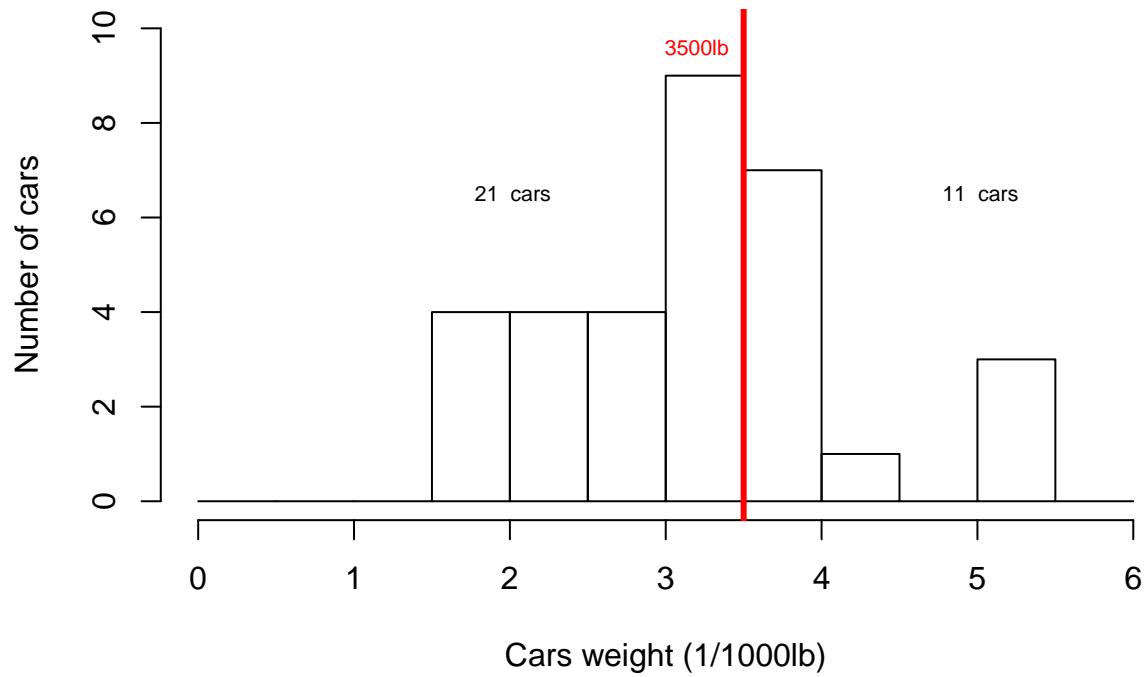
Miles per gallon by transmission type



Miles per gallon explained by the car's weight and transmission



Number of cars per weighth slots



Linear regression mpg ~ am*wt + qsec

Residuals vs Fitted

