Regression Models - Course Project

Giovanni Picardi

23 dicembre 2015

Executive summary

Dataset mtcars has been explored trying to find one or more predictors for the outcome `mpg` (miles per US gallon) and in particular to investigate how the variable `am`, an indicator variable of manual transmission (0 = automatic, 1 = manual transmission), is related to `mpg`. An exploratory analysis of the dataset showed that manual transmission is associated to a higher mean mileage per gallon with respect to automatic transmission, but also that a number of other variables are in a much stronger relationship with the outcome, and in a way that could be approximated with linear models. A T-test has been used to check the hypothesis of a higher mean mileage per gallon for manual transmission and a few linear models have been analysed and compared, in terms of R^2, to a reference model based on the single predictor `am`. Finally an approximate quantitative impact of transmission type on miles per gallon has been estimated.

Exploratory data analysis

The datasets has no missing values, all of the 11 variables are numeric and they are related to a wide range of different motorcars (32 models). A comparison of boxplots of the variable `mpg` for automatic and manual transmission (Figure 1 in Appendix) shows that the mean mileage per gallon is higher for manual transmission. Although `mpg` distributions are far from being normal (Figue 2 in Appendix), a T-test can help in testing this hypothesis or, more precisely, in rejecting the hypothesis that `mpg` means for automatic and manual transmission are equal:

```
t.test(mtcars$mpg[mtcars$am == 0], mtcars$mpg[mtcars$am == 1])
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg[mtcars$am == 0] and mtcars$mpg[mtcars$am == 1]
## 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 negative extremes of the 95% confidence interval confirm the hypothesis that the `mpg` mean for manual transmission is higher.

The relationship between pairs of variables, outcome inlcuded, can be visually examined with `pairs(mtcars)` (Figure 3 in Appendix): the more "linearly" related to `mpg` are the wight `wt`, the displacement `disp` and the horse power `hp`; the most useful discrete variable seems to be the number of cylinders `cyl`.

Modelling

Results

Appendix

Figure 1 − Miles per gallon and Transmission

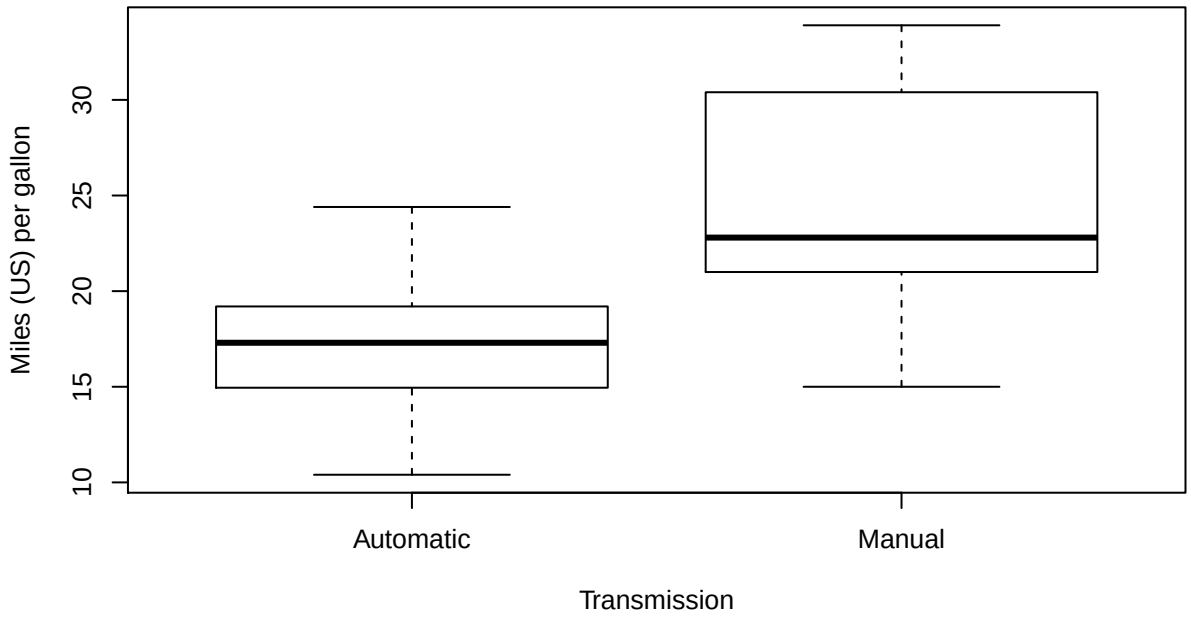

Figure 2 − Miles (US) per gallon densities

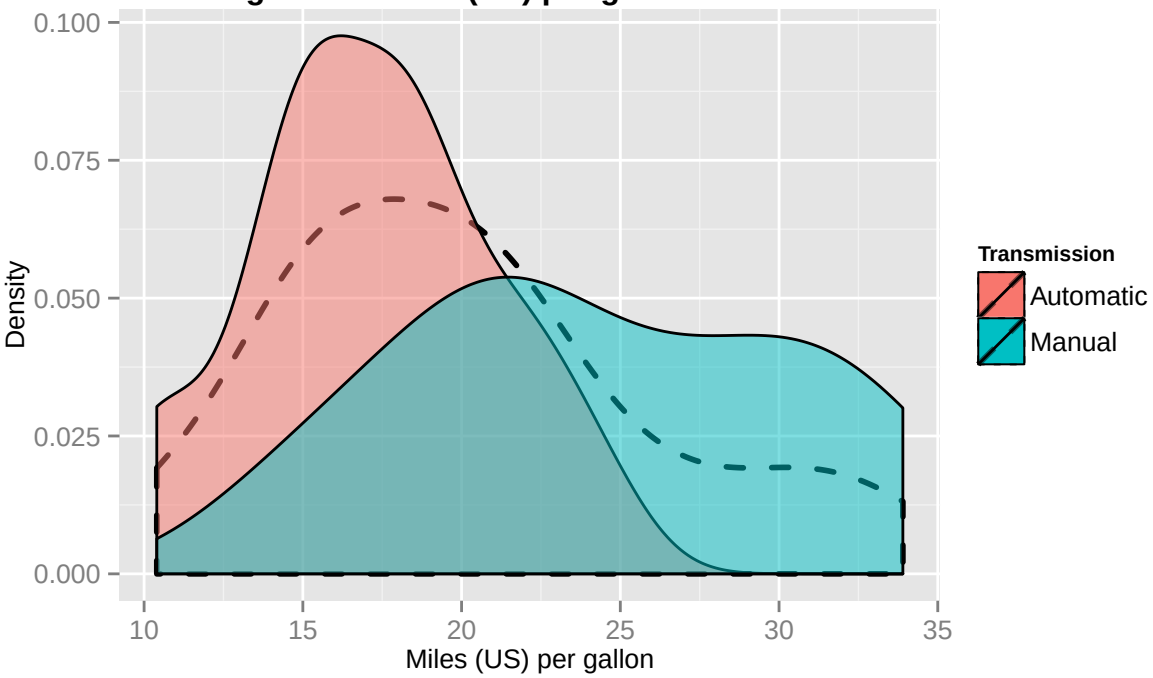

Figure 3 − Relationship between couples of variables