Fuel consumption versus transmission type

Ilya Maltsev

23 Apr 2017

Synopsis

Fuel consumption seems to be one of the most important aspects of car usage. Usually the less fuel is expended by a car the more usable the car is.

The study is focused on relationship between fuel consumption (measured as miles per gallon, MPG) and transmission type (automatic or manual).

The main questions are:

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions.

The data is used from *Motor Trend* US magazine data.

Exploratory data analysis

```
require(datasets)
require(ggplot2)
require(corrplot)
```

About the data

The *mtcars* dataset from *datasets* package is used.

```
data("mtcars")
head(mtcars, 3)
```

```
##                mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

The dataset contains 32 observations.

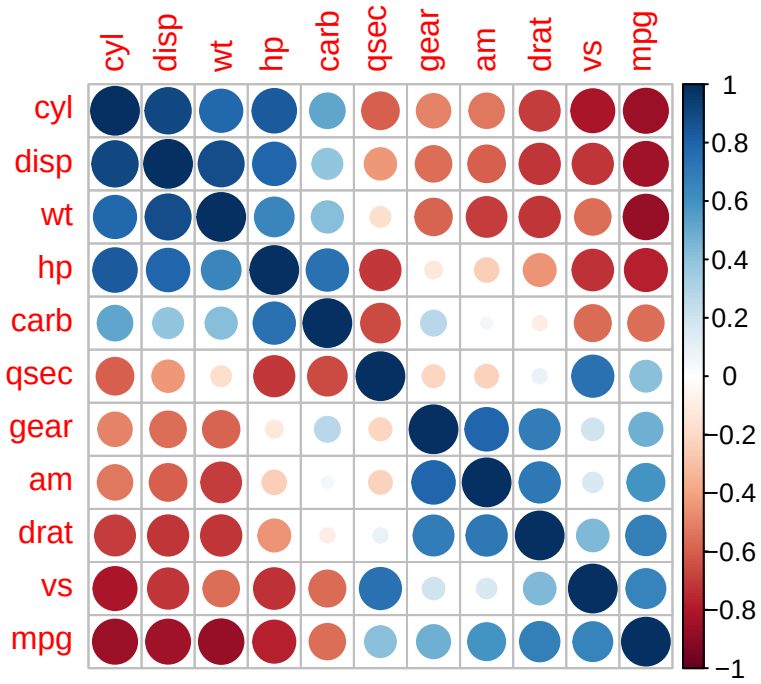
MPG distribution

The *mpg* data does not follow normal distribution (see Appendix). This means that Students t-test cannot be used to get statistical evidence of *mpg* to *am* dependency.

Correlation matrix

Some of the *mtcars* variables are physically related, e.g. displacement and horse power. Ideally, linear models should include only independent variables.

```
corrmatrix = cor(mtcars)
corrplot(corrmatrix, method = "circle", order ="FPC")
```



Correlation matrix shows that the *mpg* correlates with the *cyl*, *disp*, *hp*, *wt* the most. However the *cyl*, *disp*, *hp* are highly correlated with each other. The *wt* correlates with the *hp* less than with the *cyl* and *disp*.

Modelling

Consider the linear model $mpg = wt + hp + am$ according to the correlation matrix.

```
fit.mpg_wt_hp_am = lm(formula = mpg ~ wt + hp + I(factor(am)), data = mtcars)
round(summary(fit.mpg_wt_hp_am)$coef, 4)
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     34.0029     2.6427 12.8669   0.0000
## wt              -2.8786     0.9050 -3.1808   0.0036
## hp              -0.0375     0.0096 -3.9018   0.0005
## I(factor(am))1   2.0837     1.3764  1.5139   0.1413
```

The value for *am* in the last column shows that the we failed to reject the null hypothesis: the *am* does not affect the *mpg*. The analysis of variance table (see Appendix) confirms this result.

Conclusions

The *mtcars* data analysis does not confirm that there is a causal relationship between fuel consumption and transmission type.

Appendix

Normality of the *MPG* data

Shapiro-Wilk test is used to verify the normality of the *mpg* data. Level of statistical significance is taken to be 0.05. Results for the *mpg* and with distinction by cylinders count or by transmission type are shown below:

```
rbind(
    shapiro.test(mtcars$mpg[mtcars$cyl == 4]),
    shapiro.test(mtcars$mpg[mtcars$cyl == 6]),
    shapiro.test(mtcars$mpg[mtcars$cyl == 8]),
    shapiro.test(mtcars$mpg[mtcars$am == 0]),
    shapiro.test(mtcars$mpg[mtcars$am == 1]),
    shapiro.test(mtcars$mpg)
)[, c("data.name", "p.value")]
```

```
##      data.name                     p.value
## [1,] "mtcars$mpg[mtcars$cyl == 4]" 0.2605931
## [2,] "mtcars$mpg[mtcars$cyl == 6]" 0.3251776
## [3,] "mtcars$mpg[mtcars$cyl == 8]" 0.3228563
## [4,] "mtcars$mpg[mtcars$am == 0]"  0.8987358
## [5,] "mtcars$mpg[mtcars$am == 1]"  0.5362729
## [6,] "mtcars$mpg"                  0.1228814
```

All the p-values are greater than 0.05 so we reject the null hypothesis that the data are normal.

Analysis of variance table

```
fit.mpg_wt = lm(formula = mpg ~ wt, data = mtcars)
fit.mpg_wt_hp = lm(formula = mpg ~ wt + hp, data = mtcars)
anova(fit.mpg_wt, fit.mpg_wt_hp, fit.mpg_wt_hp_am)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + hp
## Model 3: mpg ~ wt + hp + I(factor(am))
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 278.32
## 2     29 195.05  1    83.274 12.9328 0.001226 **
## 3     28 180.29  1    14.757  2.2918 0.141268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The analysis shows that the *wt* and the *hp* are important for predicting the *mpg* and the *am* has low importance.