Automatic or manual transmission?

Ivana Lipnerová

January 25, 2016

Introduction

The aim of this study is to investigate dataset called mtcars from R library of datasets in terms of regression models. The data are extract from 1974 *Motor Trend* magazine comprising fuel consumption and 10 aspects of automobile design and performance for 32 automobiles. The questions through I will look into this dataset are about effect of types of transmission on fuel consumption: (i) is an automatic or manual transmission better for MPG and (ii) quantify the MPG difference between automatic and manual transmissions.

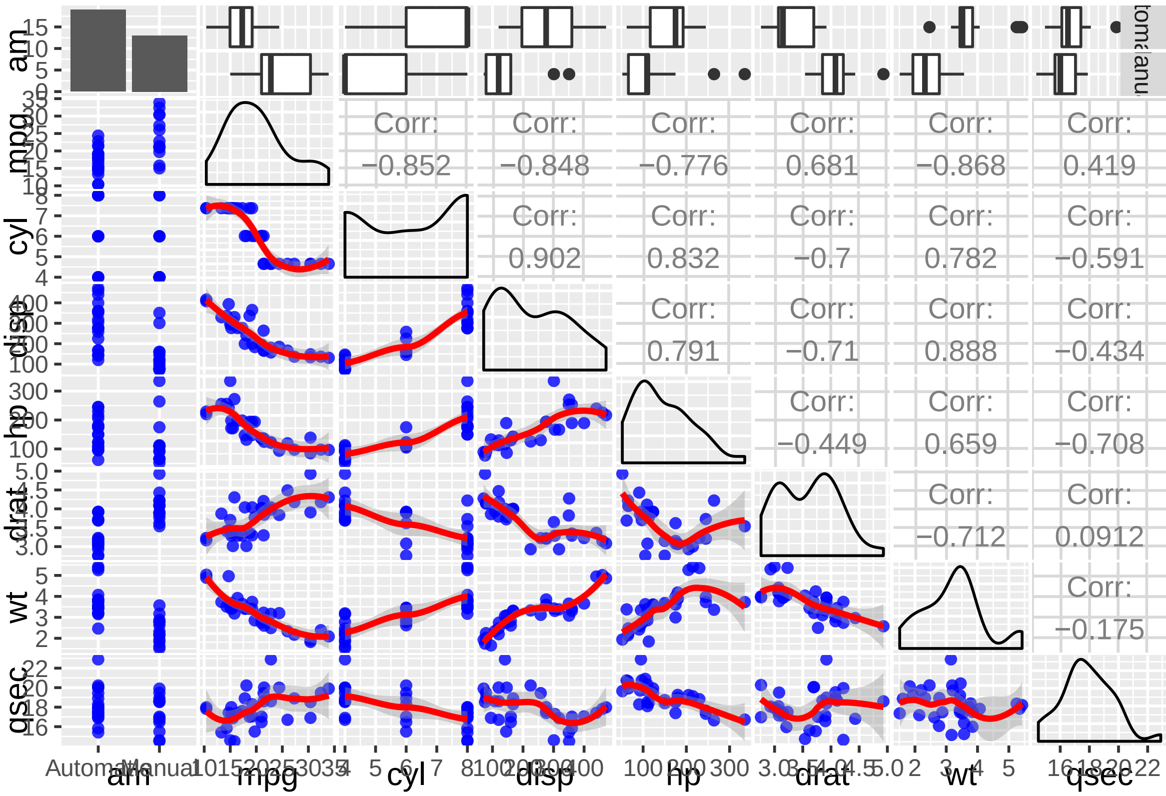
The data

```
#necessary libraries:
library(knitr); library(ggplot2); library(GGally)
#dataset:
data("mtcars"); cars<-mtcars
# Set factorial variables as factors:
cars$am[cars$am==0] <- "Automatic"; cars$am[cars$am==1] <- "Manual"
cars$vs[cars$vs==0] <- "V"; cars$vs[cars$vs==1] <- "S"
cars$am <-as.factor(cars$am); cars$vs <-as.factor(cars$vs)
```

EDA

Based on preliminary analysis of data (not shown due to length restriction), variables of impact on mpg could be all excluding vs, gear and carb, thus I decide to show just pairs plot for those included variabes. I use modified ggpairs function (originaly from GGally library, modified by Timothy A. Jenkins , alternatively on StackOverflow.

```
.ggpairs(c(9, 1:7), data=cars)
```



Fitting

The **am** seems to be related with some of other variables from the dataset, so to not omit some important effect, I will start with model with all variables and will substract nonsignificant terms step by step. I show only fit of starting and final model, all steps in between are not shown and can be found in .Rmd file on GitHub.

```
fit1<-lm(mpg ~ ., data=mtcars)
```

I step by step substracted the most unsignificant term from model until there were only significant terms and exclusion of any of them made signficant difference. The final model contains only **wt**, **qsec** and luckily, although only weakly significant, **am**. I refitted the final model so that I can compare two levels of transmissions.

```
fit10<-lm(mpg~am+wt+qsec, data=cars)
summary(fit10)$coeff
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## amManual     2.935837  1.4109045  2.080819 4.671551e-02
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

The diagnostic plots looks good (e.g. no serious problems) given binary nature of transmission variable (see Appendix for figure).

Let's calculate 95% confidence intervals of manual transmission impact.

```
sum_coeff<-summary(fit10)$coeff
sum_coeff[2,1] + c(-1, 1)*qt(.975, df=fit10$df)*sum_coeff[2,2]
```

```
## [1] 0.04573031 5.82594408
```

Summary

Cars with automatic transmission with given weight and 1/4 mile time constant are able to run on average 3 more miles per US gallon compared to the cares with manual transmission, the 95% value increase varying from 0.05 to 5.83 miles per US gallon. In internationally understandable units it is between 0.4 km/l to 2.5 km/l, with average around 1.3 km/l. So to answer the first question, the automatic transmission is better than manual transmission, if better means lower fuel consumption.

Appendix

Diagnostic plots.

```
par(mfrow=c(2,2))
plot(fit10)
```