Miles per Gallon

Alexander Alexandrov

Overview

Work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

- "Is an automatic or manual transmission better for MPG?"
- "Quantify the MPG difference between automatic and manual transmissions"

Its evident from my experience that *transimission type* is not the only variable that has impact on *mpg*. So several models should be analized and compared to each over. Following parameters should be considered:

- Transmission type
- Weight
- Gross horsepower
- Number of cylinders
- Number of forward gears

Some of them are correlated. To exclude injurious effect residuals's analisys should be made.

Motor Trend Cars Data Overview

Load Motor Trend cars data. And adjust some variables.

```
library(data.table)
data(mtcars)
data <- data.table(mtcars)
invisible({
    data[, vs := factor(vs)]
    data[, am := factor(am)]
    data[, cyl := factor(cyl)]
    data[, gear := factor(gear)]
});
str(data)
```

```
## Classes 'data.table' and 'data.frame':   32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

Data is clean enough.

Variable	Type	Domain	Description
mpg	numeric	positive real number	Miles/(US) gallon
cyl	factor	4, 6, 8	Number of cylinders
disp	numeric	positive real number	Displacement (cu.in.)
hp	numeric	positive real number	Gross horsepower
drat	numeric	positive real number	Rear axle ratio
wt	numeric	positive real number	Weight (lb/1000)
qsec	numeric	positive real number	1/4 mile time
vs	factor	0, 1	V/S
am	factor	0, 1	Transmission (0 = automatic, 1 = manual)
gear	factor	3, 4, 5	Number of forward gears
carb	numeric	positive integer number	Number of carburetors

Further exploratory data analyses

Some usefull plots.

```
library(ggplot2)
```

According to quantiles (boxplots) orange juice (red) is slightly better then ascorbic acid.