

Regression Models Assignment - Car data analysis

Fabien Nugier

11/17/2019

Introduction

This project uses the data from the R dataset `mtcars` and aims at exploring the relationship between a set of variables and the outcome variable of miles per gallons (MPG). More precisely, we want to know which of automatic or manual transmission is better for MPG and to quantify the difference between the two categories of cars.

Data Processing

We load the data and display its first rows:

```
data(mtcars)
head(mtcars)
```

```
##           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
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44 1  0   3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02 0  0   3    2
## Valiant         18.1   6  225 105 2.76 3.460 20.22 1  0   3    1
```

From the help (`?mtcars`), we get that the data consists of a data frame with 32 observations on 11 (numeric) variables:

- `mpg` : Miles/(US) gallon
- `cyl` : Number of cylinders
- `disp` : Displacement (cu.in.)
- `hp` : Gross horsepower
- `drat` : Rear axle ratio
- `wt` : Weight (1000 lbs)
- `qsec` : 1/4 mile time
- `vs` : Engine (0 = V-shaped, 1 = straight)
- `am` : Transmission (0 = automatic, 1 = manual)
- `gear` : Number of forward gears
- `carb` : Number of carburetors

The two columns that interest us the most are `mpg` and `am`. Let us display summary information about them:

```
summary(mtcars[,c("mpg", "am")])
```

```
##           mpg           am
##  Min.   :10.40  Min.   :0.0000
## 1st Qu.:15.43  1st Qu.:0.0000
##  Median :19.20  Median :0.0000
##   Mean   :20.09  Mean   :0.4062
## 3rd Qu.:22.80  3rd Qu.:1.0000
##   Max.   :33.90  Max.   :1.0000
```

Let us make the `am` variable a factor variable:

```
mtcars$am <- factor(mtcars$am, levels=c(0,1), labels=c("A","M"))
```

and we can plot the MPG output against the transmission type as a box plot:

```
library(ggplot2)
library(gridExtra)
g1 = ggplot(data=mtcars, aes(x=am,y=mpg, fill=am)) + geom_boxplot()
g1 = g1 + scale_x_discrete("Transmission") + scale_y_continuous("Miles / gallon")
g1 = g1 + ggtitle("Box plot of Miles / Gallon (MPG) ratio \n against Transmission Type.")
g1 = g1 + theme(legend.position=c(0.95,0.3), legend.justification=c(1,1))
g2 = ggplot(mtcars, aes(x=mpg, fill=am)) + geom_density(alpha=0.5)
g2 = g2 + coord_flip() + theme(legend.position="none") + labs(title="\n Density")
grid.arrange(g1,g2,ncol=2,nrow=1,widths=c(4,2))
```

