Regression Model Course Project

Sebastian Jaroszewicz

12/10/2020

Introduction

The objective of this study is to ooking at a data set of a collection of cars and to explore the relationship between a set of variables and miles per gallon (MPG) (outcome). In particular it is of interest to answer the following questions:

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

The data set that will be used to carry out the study will be the mtcars.

Exploratory Data Analysis

```
library(ggplot2)

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
```

```
str(mtcars)
```

```
## '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 : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ 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  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

The data set is composed of with 32 observations on 11 (numeric) variables.

1. mpg Miles/(US) gallon

2. cyl Number of cylinders
3. disp Displacement (cu.in.)
4. hp Gross horsepower
5. drat Rear axle ratio
6. wt Weight (1000 lbs)
7. qsec 1/4 mile time
8. vs Engine (0 = V-shaped, 1 = straight)
9. am Transmission (0 = automatic, 1 = manual)
10. gear Number of forward gears
11. carb Number of carburetors

```
# Transform same varialbes into factor
mtcars$am <- factor(mtcars$am,labels=c("Automatic","Manual"))
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

In order to better understand the data, we made a box plot graph mpg by tansmission type.

```
#boxplot(mpg ~ am, data = mtcars, col = (c("red","blue")), ylab = "Miles Per Gallon", xlab = "Transmiss
g <- ggplot(mtcars, aes(x=am, y=mpg, color = am)) +
          geom_boxplot()
g <- g + geom_jitter(shape=16, position=position_jitter(0.2))
g
```