Project regression models

Nb

29 September 2020

link for Coursera course [*Click here*](https://www.coursera.org/learn/regression-models)

### Inroduction

**Decription**

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

**Format**

A data frame 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

**Source**

Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411

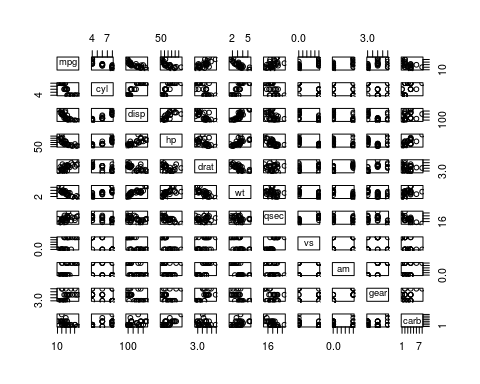
### Reading data

data(mtcars)  
# a brief view of data  
str(mtcars)

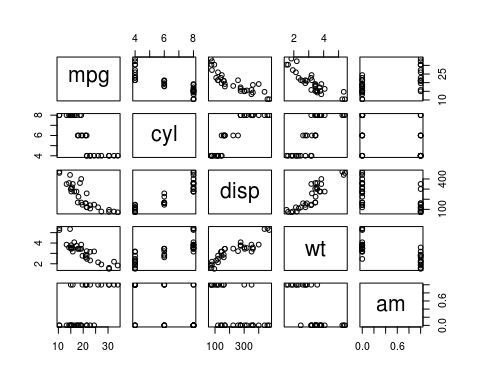
## 'data.frame': 32 obs. of 11 variables:  
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...  
## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...  
## $ disp: num 160 160 108 258 360 ...  
## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...  
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 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 ...

### Analysing Our data

pairs(mtcars)



#select some ones  
pairs(mtcars[c("mpg","cyl","disp","wt","am")])

 ### Linear Model

print("Number of cylinders")

## [1] "Number of cylinders"

fit1<-lm(cyl~., mtcars)  
fit1

##   
## Call:  
## lm(formula = cyl ~ ., data = mtcars)  
##   
## Coefficients:  
## (Intercept) mpg disp hp drat wt   
## 12.107199 -0.004857 0.004610 0.003723 -0.427435 -0.222489   
## qsec vs am gear carb   
## -0.187945 -0.644076 -0.500770 -0.500323 0.179872

print("Weight")

## [1] "Weight"

fit2<-lm(wt~., mtcars)  
fit2

##   
## Call:  
## lm(formula = wt ~ ., data = mtcars)  
##   
## Coefficients:  
## (Intercept) mpg cyl disp hp drat   
## -0.230634 -0.041666 -0.057254 0.006685 -0.003230 -0.090083   
## qsec vs am gear carb   
## 0.199541 -0.066368 0.018445 -0.093508 0.248688

print("miles / (US) gallon")

## [1] "miles / (US) gallon"

fit3<-lm(mpg~., mtcars)  
fit3

##   
## Call:  
## lm(formula = mpg ~ ., data = mtcars)  
##   
## Coefficients:  
## (Intercept) cyl disp hp drat wt   
## 12.30337 -0.11144 0.01334 -0.02148 0.78711 -3.71530   
## qsec vs am gear carb   
## 0.82104 0.31776 2.52023 0.65541 -0.19942

### Searching the most p-value

value<-names(mtcars)  
p<-NULL  
#append(fit\_list, lm(get(value[1])~get(value[2]), mtcars))  
for (i in value) {  
 if (value[1]!=i){  
 p[i]<-summary(lm(get(value[1])~get(i), mtcars))$coef[2,4]  
 }  
}  
  
# this is the minimum p-value for mpg with wich variable  
min(p)

## [1] 1.293959e-10

order(p)

## [1] 5 1 2 3 4 7 8 10 9 6

Now we will search the variable wich is more correlated to the mpg *Miles/(US) gallon*

p[p==min(p)]

## wt   
## 1.293959e-10

summary(lm(mpg~wt, mtcars))$coef

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 37.285126 1.877627 19.857575 8.241799e-19  
## wt -5.344472 0.559101 -9.559044 1.293959e-10

We conclude that there is relationsheep between **weigth** and mpg with slope equal to \*\* *-5.344* \*\*