hw1.R

Brian Kang

2020-04-06

# Brian Kang  
# Econ 487: HW1  
  
#setwd("C:/Users/slexi/Documents/UW\_ECON/ECON483")  
  
x <- c(1,3,2,5)  
x # print x

## [1] 1 3 2 5

y = c(1,4,3)  
y # print y

## [1] 1 4 3

length(x) # 4

## [1] 4

length(y) # 3

## [1] 3

x+y # invalid

## Warning in x + y: longer object length is not a multiple of shorter object  
## length

## [1] 2 7 5 6

x <- c(1,6,2)  
x+y # (2 10 5)

## [1] 2 10 5

ls()

## [1] "x" "y"

rm(x,y) # remove x,y  
ls() # empty

## character(0)

# rm(list = ls())  
  
x <- matrix(data = c(1,2,3,4), nrow = 2, ncol=2)  
x

## [,1] [,2]  
## [1,] 1 3  
## [2,] 2 4

matrix(data = c(1,2,3,4), nrow = 2, ncol=2, byrow = T)

## [,1] [,2]  
## [1,] 1 2  
## [2,] 3 4

sqrt(x) # element wise sqrt

## [,1] [,2]  
## [1,] 1.000000 1.732051  
## [2,] 1.414214 2.000000

x^2 # element wise squaring

## [,1] [,2]  
## [1,] 1 9  
## [2,] 4 16

x <- rnorm (50) # std norm, n=50  
y <- x + rnorm (50, mean=50, sd=0.1) # make new vars~normal dist.  
cor(x,y) # correlation

## [1] 0.9965061

set.seed(123)  
rnorm(40)

## [1] -0.56047565 -0.23017749 1.55870831 0.07050839 0.12928774 1.71506499  
## [7] 0.46091621 -1.26506123 -0.68685285 -0.44566197 1.22408180 0.35981383  
## [13] 0.40077145 0.11068272 -0.55584113 1.78691314 0.49785048 -1.96661716  
## [19] 0.70135590 -0.47279141 -1.06782371 -0.21797491 -1.02600445 -0.72889123  
## [25] -0.62503927 -1.68669331 0.83778704 0.15337312 -1.13813694 1.25381492  
## [31] 0.42646422 -0.29507148 0.89512566 0.87813349 0.82158108 0.68864025  
## [37] 0.55391765 -0.06191171 -0.30596266 -0.38047100

set.seed(890) # different!  
rnorm(40)

## [1] -1.48436234 0.66541304 0.73208273 0.12172975 1.91725083 -1.14279393  
## [7] 0.31410891 0.78809968 -0.61494741 0.98738224 0.85040866 0.55030807  
## [13] 1.46288913 0.88662042 0.75484690 1.27417878 0.42242136 -1.55457583  
## [19] -0.09355673 0.46385884 1.40902761 -2.39765833 1.67184652 -0.65242864  
## [25] 0.06450384 -0.59831672 -0.32721424 -0.12329801 0.21799955 -0.86668191  
## [31] 0.55300786 0.47888673 -1.73376658 -0.84632825 -0.14808126 -0.53897628  
## [37] -0.67605776 0.83179717 -0.97401568 1.48055873

set.seed(123)  
y <- rnorm(10)  
mean(y)

## [1] 0.07462564

v <- var(y) # variance  
sqrt(v) # std dev

## [1] 0.9537841

sd(y) # std dev

## [1] 0.9537841

x <- rnorm(100)  
y <- rnorm(100)  
#plot(x,y, main = " x vs y", xlab = "x axis", ylab = "y axis")  
  
#jpeg("hw1.jpg") # start jpg output plot  
#plot(x,y, main = " x vs y",   
# xlab = "x axis", ylab = "y axis",  
# col = "red")  
#dev.off() # stop plotting  
  
x <- seq(1,10)  
y <- 1:10  
x==y # true 10 times

## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE

seq(0,1,length = 10) # 0, 0.111, 0.222, ... 0.8889, 1.

## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667  
## [8] 0.7777778 0.8888889 1.0000000

x <- seq(-pi ,pi ,length = 50)  
y <- x  
f <- outer(x,y,function (x,y)cos(y)/(1+x^2)) # cross product  
#contour(x,y,f) # contour plot  
#contour(x,y,f,nlevels = 45, add=T) # more levels  
fa <- (f-t(f))/2 # t() is transpose  
#contour(x,y,fa,nlevels = 15)  
  
#image(x,y,fa) # heatmap  
#persp(x,y,fa) # 3d  
#persp(x,y,fa ,theta =30, phi =20) # 3d different angles  
  
A <- matrix(1:16, 4,4)  
A

## [,1] [,2] [,3] [,4]  
## [1,] 1 5 9 13  
## [2,] 2 6 10 14  
## [3,] 3 7 11 15  
## [4,] 4 8 12 16

A[2,3] # 10

## [1] 10

A[c(1,3), c(2,4)] # row 1,3 and column 2,4

## [,1] [,2]  
## [1,] 5 13  
## [2,] 7 15

A[1:3, 2:4] # row 1 to 3, column 2 to 4

## [,1] [,2] [,3]  
## [1,] 5 9 13  
## [2,] 6 10 14  
## [3,] 7 11 15

A[1:2,] # everything in row 1,2

## [,1] [,2] [,3] [,4]  
## [1,] 1 5 9 13  
## [2,] 2 6 10 14

A[-(1:2) ,] # everything BUT row 1,2

## [,1] [,2] [,3] [,4]  
## [1,] 3 7 11 15  
## [2,] 4 8 12 16

dim(A) # 4x4 matrix

## [1] 4 4

#install.packages("ISLR")  
library(ISLR) # Auto dataset readily available

## Warning: package 'ISLR' was built under R version 3.6.3

#fix(Auto) # view data  
dim(Auto)

## [1] 392 9

Auto <- na.omit(Auto)  
dim(Auto)

## [1] 392 9

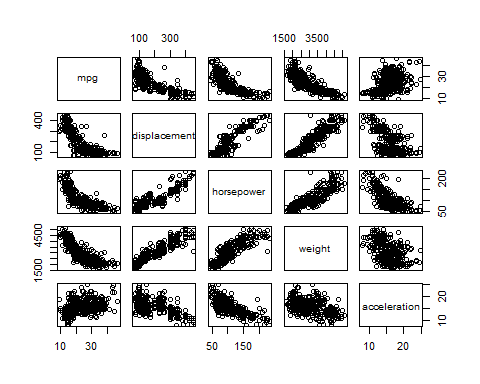
names(Auto) # variable names

## [1] "mpg" "cylinders" "displacement" "horsepower" "weight"   
## [6] "acceleration" "year" "origin" "name"

#plot(Auto$cylinders, Auto$mpg) # scatterplot  
str(Auto)

## 'data.frame': 392 obs. of 9 variables:  
## $ mpg : num 18 15 18 16 17 15 14 14 14 15 ...  
## $ cylinders : num 8 8 8 8 8 8 8 8 8 8 ...  
## $ displacement: num 307 350 318 304 302 429 454 440 455 390 ...  
## $ horsepower : num 130 165 150 150 140 198 220 215 225 190 ...  
## $ weight : num 3504 3693 3436 3433 3449 ...  
## $ acceleration: num 12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...  
## $ year : num 70 70 70 70 70 70 70 70 70 70 ...  
## $ origin : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ name : Factor w/ 304 levels "amc ambassador brougham",..: 49 36 231 14 161 141 54 223 241 2 ...

#plot(as.factor(Auto$cylinders), Auto$mpg) # gives boxplot  
attach(Auto) # make variable names as referable vectors  
cylinders <- as.factor(cylinders)  
#plot(cylinders , mpg)  
#plot(cylinders, mpg, col="red", varwidth=T, xlab="cylinders",  
# ylab ="MPG")  
#hist(mpg, col=2, breaks=15) # 2 is red  
  
#pairs(Auto) # scatterplot of all vars against all vars  
#jpeg(hw1.jpg)  
pairs(~ mpg + displacement + horsepower + weight +  
 acceleration, Auto) # scatterplot matrix of some pairs



#dev.off()  
  
#plot(horsepower, mpg)  
#identify (horsepower ,mpg ,name) # click specific points to identify  
  
summary(Auto)

## mpg cylinders displacement horsepower weight   
## Min. : 9.00 Min. :3.000 Min. : 68.0 Min. : 46.0 Min. :1613   
## 1st Qu.:17.00 1st Qu.:4.000 1st Qu.:105.0 1st Qu.: 75.0 1st Qu.:2225   
## Median :22.75 Median :4.000 Median :151.0 Median : 93.5 Median :2804   
## Mean :23.45 Mean :5.472 Mean :194.4 Mean :104.5 Mean :2978   
## 3rd Qu.:29.00 3rd Qu.:8.000 3rd Qu.:275.8 3rd Qu.:126.0 3rd Qu.:3615   
## Max. :46.60 Max. :8.000 Max. :455.0 Max. :230.0 Max. :5140   
##   
## acceleration year origin name   
## Min. : 8.00 Min. :70.00 Min. :1.000 amc matador : 5   
## 1st Qu.:13.78 1st Qu.:73.00 1st Qu.:1.000 ford pinto : 5   
## Median :15.50 Median :76.00 Median :1.000 toyota corolla : 5   
## Mean :15.54 Mean :75.98 Mean :1.577 amc gremlin : 4   
## 3rd Qu.:17.02 3rd Qu.:79.00 3rd Qu.:2.000 amc hornet : 4   
## Max. :24.80 Max. :82.00 Max. :3.000 chevrolet chevette: 4   
## (Other) :365

summary(mpg)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 9.00 17.00 22.75 23.45 29.00 46.60

#savehistory() # save commands from this session  
#loadhistory() # load commands from most recent session  
#q() # quit