DSA Assignment

Sunny Sharma

18/12/2020

Object creation & Value assignment

a=25  
b=30

Command line Interface

b

## [1] 30

Object need not be explicitly defined

c=45  
class(c)

## [1] "numeric"

d=TRUE  
class(d)

## [1] "logical"

Simple Mathematical Operations:

1. Addition

a+b

## [1] 55

2)Subtraction

b-a

## [1] 5

1. Division

b/a

## [1] 1.2

1. Multiplication

b\*a

## [1] 750

1. Square root

sqrt(a)

## [1] 5

1. Power

a^3

## [1] 15625

7)logarithm

log(a,base=exp(1))

## [1] 3.218876

1. Factorial

factorial(a)

## [1] 1.551121e+25

Data types are as follows:

1. Numeric

a=20  
class(a)

## [1] "numeric"

2)Character

b="Hello"  
class(b)

## [1] "character"

1. Logical

c=TRUE  
class(c)

## [1] "logical"

d=FALSE  
class(d)

## [1] "logical"

1. Date

date1=as.Date("2015-04-16")  
class(date1)

## [1] "Date"

5)POSIXct - Date & Time

date2=as.POSIXct("2015-5-18 16:32")  
class(date2)

## [1] "POSIXct" "POSIXt"

Factors

rm(list=ls())  
a=c("jan","feb","mar","apr")  
a

## [1] "jan" "feb" "mar" "apr"

as.numeric(a)

## Warning: NAs introduced by coercion

## [1] NA NA NA NA

class(a)

## [1] "character"

b=as.factor(a)  
as.numeric(b)

## [1] 3 2 4 1

b

## [1] jan feb mar apr  
## Levels: apr feb jan mar

# Missing Value test  
c=c(1,2,3,NA,NA,5)  
is.na(c)

## [1] FALSE FALSE FALSE TRUE TRUE FALSE

d=c("jan",NA,"feb")  
is.na(d)

## [1] FALSE TRUE FALSE

# Null value  
e=c(NULL,3,4)  
is.null(e)

## [1] FALSE

f=NULL  
is.null(f)

## [1] TRUE

TO get a working directory

getwd()

## [1] "C:/Users/Admin/Documents/R Files"

Divider Function

divider = function(x,y)  
{  
 result = x/y  
 print(result)  
}  
divider(100,20)

## [1] 5

Multiplier Function

multiply = function(x,y)  
{  
 result = x\*y  
 print(result)  
}  
multiply(8,20)

## [1] 160

Concatenation and Arrays

d=c(1:5)  
d

## [1] 1 2 3 4 5

d+6

## [1] 7 8 9 10 11

d-2

## [1] -1 0 1 2 3

R is a vectorized language

e=c(1:10)  
e

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

f=e\*2  
f

## [1] 2 4 6 8 10 12 14 16 18 20

g=f/2  
g

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

Matrix Creation

M=matrix(1:15,nrow=3)  
M

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

a=matrix(1:10,nrow=5)  
b=matrix(11:20,nrow=2)  
a

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

b

## [,1] [,2] [,3] [,4] [,5]  
## [1,] 11 13 15 17 19  
## [2,] 12 14 16 18 20

# Matrix addition  
c=t(b)  
a+c

## [,1] [,2]  
## [1,] 12 18  
## [2,] 15 21  
## [3,] 18 24  
## [4,] 21 27  
## [5,] 24 30

# Matrix Multiplication  
a%\*%b

## [,1] [,2] [,3] [,4] [,5]  
## [1,] 83 97 111 125 139  
## [2,] 106 124 142 160 178  
## [3,] 129 151 173 195 217  
## [4,] 152 178 204 230 256  
## [5,] 175 205 235 265 295

Arrays (N dimensional)

d=array(1:16, dim=c(4,2,2))  
d

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

d[,,2]

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

d[,1,1]

## [1] 1 2 3 4

url=read.table("C:\\Users\\Admin\\Documents\\R\\Tomato First.csv",header=TRUE,sep=",")  
head(url)

## Round Tomato Price Source Sweet Acid Color Texture Overall  
## 1 1 Simpson SM 3.99 Whole Foods 2.8 2.8 3.7 3.4 3.4  
## 2 1 Tuttorosso (blue) 2.99 Pioneer 3.3 2.8 3.4 3.0 2.9  
## 3 1 Tuttorosso (green) 0.99 Pioneer 2.8 2.6 3.3 2.8 2.9  
## 4 1 La Fede SM DOP 3.99 Shop Rite 2.6 2.8 3.0 2.3 2.8  
## 5 2 Cento SM DOP 5.49 D Agostino 3.3 3.1 2.9 2.8 3.1  
## 6 2 Cento Organic 4.99 D Agostino 3.2 2.9 2.9 3.1 2.9  
## Avg.of.Totals Total.of.Avg  
## 1 16.1 16.1  
## 2 15.3 15.3  
## 3 14.3 14.3  
## 4 13.4 13.4  
## 5 14.4 15.2  
## 6 15.5 15.1

Loading Built in data sets

data("mtcars")  
head(mtcars, 7)

## 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  
## Duster 360 14.3 8 360 245 3.21 3.570 15.84 0 0 3 4

# Generate a random sample  
d=sample(x=1:25,size=50,replace=TRUE)  
d

## [1] 15 13 12 10 19 6 21 22 16 22 10 5 10 7 5 20 12 20 16 10 8 12 4 13 1  
## [26] 20 16 10 19 16 6 18 18 15 19 17 25 25 7 12 6 5 14 24 19 4 14 4 23 8

mean(d)

## [1] 13.46

# Weighted means  
grade=c(95,92,91,87)  
weights= c(1/2,1/4,1/8,1/8)  
mean(grade)

## [1] 91.25

weighted.mean(x=grade,w=weights)

## [1] 92.75

# Variance & SD  
var(d)

## [1] 41.31469

sd(x=d,na.rm=FALSE)

## [1] 6.427651

# Min & Max Functions  
min(d)

## [1] 1

max(d)

## [1] 25

median(d)

## [1] 13.5

quantile(d,probs=c(0.2,0.4,0.6,0.8))

## 20% 40% 60% 80%   
## 6.8 12.0 16.0 19.2

library("ggplot2")  
head(economics)

## # A tibble: 6 x 6  
## date pce pop psavert uempmed unemploy  
## <date> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1967-07-01 507. 198712 12.6 4.5 2944  
## 2 1967-08-01 510. 198911 12.6 4.7 2945  
## 3 1967-09-01 516. 199113 11.9 4.6 2958  
## 4 1967-10-01 512. 199311 12.9 4.9 3143  
## 5 1967-11-01 517. 199498 12.8 4.7 3066  
## 6 1967-12-01 525. 199657 11.8 4.8 3018

# Covariance  
cor(economics$pce,economics$psavert)

## [1] -0.7928546

rm(list=ls())  
data(tips, package="reshape2")  
head(tips)

## total\_bill tip sex smoker day time size  
## 1 16.99 1.01 Female No Sun Dinner 2  
## 2 10.34 1.66 Male No Sun Dinner 3  
## 3 21.01 3.50 Male No Sun Dinner 3  
## 4 23.68 3.31 Male No Sun Dinner 2  
## 5 24.59 3.61 Female No Sun Dinner 4  
## 6 25.29 4.71 Male No Sun Dinner 4

str(tips)

## 'data.frame': 244 obs. of 7 variables:  
## $ total\_bill: num 17 10.3 21 23.7 24.6 ...  
## $ tip : num 1.01 1.66 3.5 3.31 3.61 4.71 2 3.12 1.96 3.23 ...  
## $ sex : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 2 2 2 2 2 ...  
## $ smoker : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ day : Factor w/ 4 levels "Fri","Sat","Sun",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ time : Factor w/ 2 levels "Dinner","Lunch": 1 1 1 1 1 1 1 1 1 1 ...  
## $ size : int 2 3 3 2 4 4 2 4 2 2 ...

# Returns Unique values  
unique(tips$sex)

## [1] Female Male   
## Levels: Female Male

unique(tips$day)

## [1] Sun Sat Thur Fri   
## Levels: Fri Sat Sun Thur

# one sample t test (H0 = 2.5, Two Tailed)  
t.test(tips$tip, alternative="two.sided",mu=2.5)

##   
## One Sample t-test  
##   
## data: tips$tip  
## t = 5.6253, df = 243, p-value = 5.08e-08  
## alternative hypothesis: true mean is not equal to 2.5  
## 95 percent confidence interval:  
## 2.823799 3.172758  
## sample estimates:  
## mean of x   
## 2.998279

# one sample t test (H0 = 2.5, upper tail)  
t.test(tips$tip, alternative="greater",mu=2.5)

##   
## One Sample t-test  
##   
## data: tips$tip  
## t = 5.6253, df = 243, p-value = 2.54e-08  
## alternative hypothesis: true mean is greater than 2.5  
## 95 percent confidence interval:  
## 2.852023 Inf  
## sample estimates:  
## mean of x   
## 2.998279

# Two sample t test   
t.test(tip~sex, data=tips, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: tip by sex  
## t = -1.3879, df = 242, p-value = 0.1665  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.6197558 0.1074167  
## sample estimates:  
## mean in group Female mean in group Male   
## 2.833448 3.089618

# Anova  
tipAnova=aov(tip~day,tips)  
summary(tipAnova)

## Df Sum Sq Mean Sq F value Pr(>F)  
## day 3 9.5 3.175 1.672 0.174  
## Residuals 240 455.7 1.899

Vector Creation

a=c(1:15)  
a

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

Vector Operation

a+3

## [1] 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

a-2

## [1] -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13

a^2

## [1] 1 4 9 16 25 36 49 64 81 100 121 144 169 196 225

b=c(-4:0)  
a+b

## [1] -3 -1 1 3 5 2 4 6 8 10 7 9 11 13 15

d=c(-4:10)  
a

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

d

## [1] -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10

any(a<d)

## [1] FALSE

all(a<d)

## [1] FALSE

a[4]

## [1] 4

f=c("january","february","march","april","may","june")  
nchar(f)

## [1] 7 8 5 5 3 4

Naming Vectors

c(one="a",two="b",three="c")

## one two three   
## "a" "b" "c"

w=1:3  
names(w)=c("a","b","c")  
w

## a b c   
## 1 2 3