MSDS 6306: Introduction to Data Science - Live session Unit 01 Assignment

### 1)

1. log(10) = 2.3025851
2. ?log(): description says that log() default is the natural log. Using base 10: log(10, base = 10) = 1
3. Log(-10) = NaN Cannot take the log of a negative number. The log curve is asymptotic to negative infinity as x approaches zero(log(0) = -\infty{}
4. sqrt(4) = 2

### 2)

a.

rNormVector <- rnorm(15)  
mean(rNormVector)

## [1] 0.09415648

sd(rNormVector)

## [1] 1.346474

b.

rNormVector <- rnorm(15,mean = 10,sd = 2)  
mean(rNormVector)

## [1] 9.275331

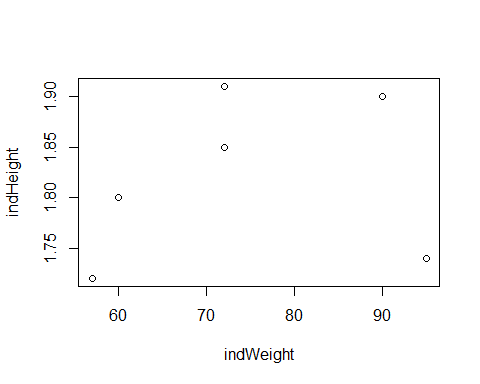
sd(rNormVector)

## [1] 1.343167

1. The values are randomly sampled fromt he distribution so the exact values will not match. As the sample size increases the mean will get closer to the actual mean.

### 3)

indWeight <- c(60, 72, 57, 90, 95, 72)  
indHeight <- c(1.80, 1.85, 1.72, 1.90, 1.74, 1.91)   
plot(indWeight, indHeight) # Weight and height would seem to be positivly related

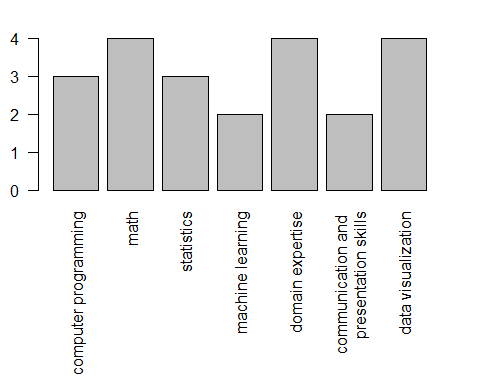


indBMI <- indWeight / sqrt(indHeight)  
sum(mean(indWeight) - indWeight)

## [1] -2.842171e-14

### 4)

brychan <- data.frame(  
 area = c("computer programming", "math", "statistics", "machine learning", "domain expertise",   
 "communication and \npresentation skills", "data visualization"),  
 rank = c(3, 4, 3, 2, 4, 2, 4)  
 )  
par(mar = c(10,2,2,2))  
barplot(brychan$rank, names.arg = brychan$area,las=2)



### 5)

rm(list = ls()) # swirl suggest removing all workspace variables

library(swirl)

swirl()

#R Programing - mod 1

5+7

x <- 5+7

x

y <- x - 3

y

z <- c(1.1, 9, 3.14)

?c

z

c(z,555,z)

z \* 2 + 100

my\_sqrt <- sqrt(z - 1)

my\_sqrt

my\_div <- z / my\_sqrt

my\_div

c(1, 2, 3, 4) + c(0, 10)

c(1, 2, 3, 4) + c(0, 10, 100)

#R Programing - mod 2

getwd()

ls()

x <- 9

ls()

list.files()

?list.files

args(list.files)

old.dir <- getwd()

dir.create("testdir")

setwd("testdir")

file.create( "mytest.R")

list.files()

file.exists("mytest.R")

file.info("mytest.R")

file.rename("mytest.R", "mytest2.R")

file.copy("mytest2.R", "mytest3.R")

file.path("folder1","folder2")

?dir.create

dir.create(file.path('testdir2', 'testdir3'), recursive = TRUE)

setwd(old.dir)

#R Programing - mod 3

1:20

pi:10

15:1

?":"

seq(1, 20)

my\_seq <- seq(5, 10, length=30)

length(my\_seq)

1:length(my\_seq)

seq(along.with = my\_seq)

seq\_along(my\_seq)

rep(0, times = 40)

rep(c(0, 1, 2), times = 10)

rep(c(0, 1, 2), each = 10)

#R Programing - mod 4

num\_vect <- c( 0.5, 55, -10, 6)

tf <- num\_vect < 1

tf

num\_vect >= 6

my\_char <- c("My", "name", "is")

my\_char

paste(my\_char, collapse = " ")

my\_name <- c(my\_char, "Brychan")

my\_name

paste(my\_name, collapse = " ")

paste("Hello", "world!", sep = " ")

paste(1:3, c("X", "Y", "Z"), sep="")

paste(LETTERS, 1:4, sep = "-")

#R Programing - mod 5

x <- c(44, NA, 5, NA)

x\*3

y <- rnorm(1000)

z <- rep(NA, 1000)

my\_data <- sample(c(y, z), 100)

is.na(my\_data)

my\_na <- is.na(my\_data)

my\_na

my\_data == NA

sum(my\_na)

my\_data

0/0

Inf-Inf

#R Programing - mod 6

x <- sample(c(rnorm(20), rep(NA,20)), 20)

x[1:10]

x[is.na(x)]

y <- x[!is.na(x)]

y

y[y > 0]

x[x > 0]

x[!is.na(x) & x > 0]

x[c(3,5,7)]

x[0]

x[3000]

x[c(-2, -10)]

x[-c(2, 10)]

vect <- c(foo = 11, bar = 2, norf = NA)

vect

names(vect)

vect2 <- c(11, 2, NA)

names(vect2) <- c("foo", "bar", "norf")

identical(vect, vect2)

vect["bar"]

vect[c("foo", "bar")]

#R Programing - mod 7

my\_vector <- 1:20

my\_vector

length(my\_vector)

dim(my\_vector) <- c(4, 5)

dim(my\_vector)

attributes(my\_vector)

my\_vector

class(my\_vector)

my\_matrix <- my\_vector

?matrix

my\_matrix2 <- matrix(1:20, 4, 5)

identical(my\_matrix, my\_matrix2)

patients <- c("Bill", "Gina", "Kelly", "Sean")

cbind(patients, my\_matrix)

my\_data <- data.frame(patients, my\_matrix)

my\_data

class(my\_data)

cnames <- c("patient", "age", "weight", "bp", "rating", "test")

colnames(my\_data) <- cnames

my\_data