CSE2047 – Data Analytics - Lab Sheet : 5

Academic year: 2020-2021   **Branch/ Class:** B.Tech/M.Tech
Semester: Fall   **Date:**
Faculty Name: Prof. S.Gopikrishnan   **School:** SCOPE
Student name: Valiveti Manikanta bhuvanesh   **Reg. no.: 19BCD7088**

LAB 5

Questions:

1. Use a nested for loop (a for loop inside a for loop) that produces the following matrix, pre-allocate the matrix with NA values.
   0 1 2 3 4
   1 0 1 2 3
   2 1 0 1 2
   3 2 1 0 1
   4 3 2 1 0

```
> m = matrix(NA, nrow = 5, ncol = 5)
> m
     [,1] [,2] [,3] [,4] [,5]
[1,]  NA   NA   NA   NA   NA
[2,]  NA   NA   NA   NA   NA
[3,]  NA   NA   NA   NA   NA
[4,]  NA   NA   NA   NA   NA
[5,]  NA   NA   NA   NA   NA
> for (i in 1:5) {
+   for (j in 1:5) {
+     m[i, j] <- abs(i - j)
+   }
+ }
> m
```

2. Using the following variable:
   x=100
   y=50
   i=1
   For this exercise, write a repeat () loop that incrementing i computes x=x-i and y=y+i until x<y.

```
> x=100
> y=50
> i=1
> repeat
+ {
+   x=x-i
+   y=y+i
+   if (x<y)
+     break
+ }
> y
[1] ?
> x
[1] 74
> y
[1] ?n
>
```

3. Find out the river length less than 400, if so, identify it as short river. count the number of short rivers. if the river length is > 800, then it is long. count the number of long rivers. (Use revier dataset)

```
> df<-read.csv("River_Dataset.csv")
> df<-data.frame(df)
> v=c(1:dim(df)[1])
> x=0
> y=0
> for (i in v) {
+     if(df$length[i]>800){
+       y=y+1
+     }
+     else if(df$length[i]<400){
+       x=x+1
+     }
+     else{
+     }
+ }
> print(paste("Number of short rivers ",y))
[1] "Number of short rivers   394"
> print(paste("Number of long rivers ",x))
[1] "Number of long rivers   140"
>
```

4. Using an if -else find out the long living animal based on lifespan. (Use animal dataset)

5. Count how many males are in underweight and overweight. you need to measure with ounce. 1 ounce=28.34 gm. (Use BMI Dataset)

```
df<-read.csv("bmi_data.csv")
dim(subset(df,Sex=="Male" & BMI<18.5))[1]
dim(subset(df,Sex=="Male" & BMI<29.9 & BMI>25))[1]
```

6. create a dummy dataset with numeric values. a) write a function for calculating the mean. b) Write a function to compute standard deviation.

```
> df<-read.csv("mycsv.csv")
> df
  numbers1    numbers2
1       80 0.303222656
2       19 0.402372524
3       22 0.564776202
4       12 0.508582586
5       79 0.177009262
6       78 0.709175202
7       42 0.119810218
8       59 0.774755903
9       74 0.780253744
```

```
> mean1<-function(a){
+   sum=0
+   count=0
+   for (i in a){
+     sum=sum+i
+     count=count+1
+   }
+   return (sum/count)
+ }
> p=mean1(df$numbers1)
> p
[1] 50.71888
> sd<-function(a){
+   sum=0
+   count=-1
+   p=mean1(a)
+   for (i in a){
+     sum=sum+((i-p)*(i-p))
+     count=count+1
+   }
+   return(sqrt(sum/count))
+ }
>
> l=sd(df$numbers2)
> l
[1] 0.2914786
```