STAT 40001/MA59800 **Statistical Computing** **Fall 2017**

Homework 1-

Name:

Due : September 7, 2017 PUID:

Instruction: Please submit your R code along with a brief write-up of the solutions (do not submit raw output). Some of the questions below can be answered with very little or no programming. However, write code that outputs the final answer and does not require any additional paper calculations. For example, suppose I ask for how many numbers are greater then 5 in the vector, x=c(1,9,2,8,10,12). Do not simply count the number of **TRUE***s by hand, instead let the R count the number of* **TRUE***s by coding* $\boldsymbol{sum\ (x > 5)}$ *or* $\boldsymbol{length(x[x > 5])}$.

Q.N. 1) Calculate the following numerical results to the three decimal places

a) $\mathbf{(7+8)+5^3-5\div 6+\sqrt{62}}$

b) $\mathbf{\ln 3+\sqrt{2}\sin(\pi)-e^3}$

c) $\mathbf{2\times(5+3)-\sqrt{6}+9^2}$

d) $\mathbf{\ln 5-\exp(2)+2^3}$

e) $\mathbf{(9\div 2)\times 4-\sqrt{10}+\ln 6-\exp(1)}$

f) $\mathbf{\log 14+\ln 14+(47 mod\ 5)}$

Solution: We used the following R codes to perform the calculations

```
a)
> round((7+8)+5^3-5/6+sqrt(62),3)
[1] 147.041
b)
> round((log(3)+sqrt(2)*sin(pi)-exp(3)),3)
[1] -18.987
c)
> round((2*(5+3)-sqrt(6)+9^2),3)
[1] 94.551
d)
> round((log(5)-exp(2)+2^3),3)
[1] 2.22
e)
> round(((9/2)*4-sqrt(10)+log(6)-exp(1)),3)
[1] 13.911
f)
> round((log10(14)+log(14))+(47%%5),3)
[1] 5.785
```

Q.N. 2) Create the following vectors

a) The vector consisting of the decreasing sequence of consecutive integers from 50 to -5.

b) The vector of first 100 positive integers, without the perfect squares.

c) The vector of the factorial values of 0 to 10.

d) V1= 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5

e) V2= MATH, MATH, STAT, STAT, STAT, STAT, STAT, ECE,ECE,ECE, BIO,BIO

Solution:

a) We can use R code below to generate the sequence

```
> seq(50,-5)
 [1] 50 49 48 47 46 45 44 43 42 41 40 39 38 37 36 35 34 33 32 31 30 29 28 27 26
[26] 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10  9  8  7  6  5  4  3  2  1
[51]  0 -1 -2 -3 -4 -5
```

b) We use R code below to generate the desired sequence of numbers

```
> x=seq(1,100)
> y=(seq(1,10))^2
> y
 [1]   1   4   9  16  25  36  49  64  81 100
> x[!x%in% y]
 [1]  2  3  5  6  7  8 10 11 12 13 14 15 17 18 19 20 21 22 23 24 26 27 28 29 30
[26] 31 32 33 34 35 37 38 39 40 41 42 43 44 45 46 47 48 50 51 52 53 54 55 56 57
[51] 58 59 60 61 62 63 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 82 83 84
[76] 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99
```

c) We can use R code below to generate this vector

```
> x
 [1]  0  1  2  3  4  5  6  7  8  9 10
> factorial(x)
 [1]       1       1       2       6      24     120     720    5040   40320
[10]  362880 3628800
```

OR

```
> for(i in 0:10){print(factorial(i))}
[1] 1
[1] 1
[1] 2
[1] 6
[1] 24
[1] 120
[1] 720
[1] 5040
[1] 40320
[1] 362880
[1] 3628800
```

d) We use R code below to generate the required vector

```
> V1=rep(c(1,2,3,4,5), each=4)
> V1
 [1] 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5
```

e) We use R code below to generate the desired vector

```
> V2=rep(c("MATH","STAT","ECE","BIO"), c(2,5,3,2))
 > noquote(paste(V2, collapse=","))
[1] MATH,MATH,STAT,STAT,STAT,STAT,STAT,ECE,ECE,ECE,BIO,BIO
```

Q.N. 3) Create the following matrix (M) with the column and row names (Note that the numbers are in sequence from 1 to 20)

```
>M
             column-1 column-2 column-3 column-4 column-5
Experiment.1        1        5        9       13       17
Experiment.2        2        6       10       14       18
Experiment.3        3        7       11       15       19
Experiment.4        4        8       12       16       20
```

a) Determine the dimension of the matrix M
b) Select the first two row of the matrix M
c) Calculate the sum of all columns of the matrix M
d) Calculate the sum of all rows of the matrix M
e) Use "sample" to shuffle the elements of each row of the matrix M

Solution: We used the code below to create the desired matrix

```
> M<-matrix(1:20,nrow=4)
> rownames(M)<-rownames(M,do.NULL=FALSE,prefix="Experiment.")
> colnames(M)<-colnames(M,do.NULL=FALSE,prefix="Column-")
> M
             Column-1 Column-2 Column-3 Column-4 Column-5
Experiment.1        1        5        9       13       17
Experiment.2        2        6       10       14       18
Experiment.3        3        7       11       15       19
Experiment.4        4        8       12       16       20

a)
> dim(M)
[1] 4 5
b)
> M[c(1,2),]
             Column-1 Column-2 Column-3 Column-4 Column-5
Experiment.1        1        5        9       13       17
Experiment.2        2        6       10       14       18
c)
> colSums(M)
Column-1 Column-2 Column-3 Column-4 Column-5
      10       26       42       58       74
```

```
d)
> rowSums(M)
Experiment.1 Experiment.2 Experiment.3 Experiment.4
          45           50           55           60


e)
> t(apply(M,1,sample))
             [,1] [,2] [,3] [,4] [,5]
Experiment.1   17   13    9    5    1
Experiment.2   18   14   10    2    6
Experiment.3   15    7   19    3   11
Experiment.4   20   16    4   12    8
```

Q.N. 4) Test scores of Fifteen students in Test 1 and Test 2 are presented below

S.N.	Test 1	Test2
1	56	86
2	78	67
3	87	78
4	89	89
5	95	87
6	98	67
7	NA	94
8	78	78
9	87	81
10	98	83
11	54	78
12	89	NA
13	78	93
14	98	98
15	97	100

a) How many students have their test 1 score greater than 80 ?
b) How many students have their test 2 score greater than 85 ?
c) Did all fifteen students take both tests?
d) How many students did better in the second test than the first test?
e) How many students have the same score in the first and second test?

Solution: We have used the following R code to answer (a)- (e)

```
> Test1
 [1] 56 78 87 89 95 98 NA 78 87 98 54 89 78 98 97
>  Test2=data$Test2
> Test2
 [1]  86  67  78  89  87  67  94  78  81  83  78  NA  93  98 100
```

```
 a)
>  sum(Test1>80, na.rm=T)
[1] 9
b)
>  sum(Test2>85,na.rm=T)
[1] 7
c)
> any(is.na(Test1))
[1] TRUE
> any(is.na(Test2))
[1] TRUE
> which (is.na(Test1))
[1] 7
> which (is.na(Test2))
[1] 12
No, not all students took both tests.
d)
> sum(Test2>Test1, na.rm=T)
[1] 4
e)
> sum(Test2==Test1, na.rm=T)
[1] 3
> which(Test2==Test1)
[1]  4  8 14
```

Q.N. 5) Use R to solve the following system of equations:

$$\begin{aligned} \boldsymbol{x_1 + 2x_2 - x_3 + 3x_4 - x_5} &= \boldsymbol{0} \\ \boldsymbol{x_1 - 3x_2 + x_3 + 2x_4 - x_5} &= \boldsymbol{-9} \\ \boldsymbol{2x_1 + x_2 + x_3 - 3x_4 + x_5} &= \boldsymbol{12} \\ \boldsymbol{x_1 - x_2 + 2x_3 + x_4 - x_5} &= \boldsymbol{1} \\ \boldsymbol{2x_1 + x_2 - x_3 + 2x_4 + x_5} &= \boldsymbol{-2} \end{aligned}$$

Solution:

```
> A=matrix(c(1,2,-1,3,-1,1,-3,1,2,-1,2,1,1,-3,1,1,-1,2,1,-1,2,1,-1,2,1),nrow=5,byrow=T)
> A
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    2   -1    3   -1
[2,]    1   -3    1    2   -1
[3,]    2    1    1   -3    1
[4,]    1   -1    2    1   -1
[5,]    2    1   -1    2    1
```

```
> Y = matrix(c(0,-9,12,1,-2),nrow=5)
```

```
> Y
     [,1]
[1,]    0
[2,]   -9
[3,]   12
[4,]    1
[5,]   -2
> X=solve(A,Y)
> X
     [,1]
[1,]    1
[2,]    3
[3,]    2
[4,]   -2
[5,]   -1
```

Hence, the solution of the given system of equations is $(x_1, x_2, x_3, x_4, x_5) = (1, 3, 2, -2, -1)$.

Q.N. 6) The Fibonacci sequence is a famous sequence in mathematics. The first two elements are defined as [1, 1]. Subsequent elements are defined as the sum of the preceding two elements. For example, the third element is 2 (= 1+1), the fourth element is 3 (= 1+2), the fifth element is 5 (= 2+3), and so on. Print first 50 Fibonacci numbers.

Hint: To obtain the first 10 Fibonacci numbers in R, we can use

```
> Fibonacci <- numeric(10)
> Fibonacci[1] <- Fibonacci[2] <- 1
> for (i in 3:10) Fibonacci[i] <- Fibonacci[i - 2] + Fibonacci[i - 1]
> Fibonacci
 [1]  1  1  2  3  5  8 13 21 34 55
```

Solution: We use R code below to generate the first 50 Fibonacci numbers.

```
> Fibonacci <- numeric(50)
> Fibonacci[1] <- Fibonacci[2] <- 1
> for (i in 3:50) Fibonacci[i] <- Fibonacci[i - 2] + Fibonacci[i - 1]
> Fibonacci
 [1]           1           1           2           3           5           8
 [7]          13          21          34          55          89         144
[13]         233         377         610         987        1597        2584
[19]        4181        6765       10946       17711       28657       46368
[25]       75025      121393      196418      317811      514229      832040
[31]     1346269     2178309     3524578     5702887     9227465    14930352
[37]    24157817    39088169    63245986   102334155   165580141   267914296
[43]   433494437   701408733  1134903170  1836311903  2971215073  4807526976
[49]  7778742049 12586269025
```

Q.N. 7) *Create the following data frame*

```
       Age    Height Weight Sex
Joe      26    175     157   F
Nina     31    165     139   M
Mark     23    190     163   F
Sonia    52    179     155   M
Martha   76    163     170   M
Andrew   39    183     183   F
Marcie   26    164     153   M
```

a) It appears that the categorical variable Sex has been recorded incorrectly. Invert Sex for all individuals.
b) Sort the data in increasing order of Age.
c) Sort the data in descending order of Age.
Solution: We will use the following R codes a)

```
> Name=c("Joe","Nina","Mark","Sonia", "Martha","Andrew", "Marcie")
> Age=c(26,31,23,52,76,39,26)
> Height=c(175,165,190,179,163,183,164)
> Weight=c(157,139,163,155,170,183,153)
> Sex=c("F","M","F","M","M","F","M")
> data=data.frame(row.names=Name, Age, Height, Weight, Sex)
> data
       Age Height Weight Sex
Joe     26    175    157   F
Nina    31    165    139   M
Mark    23    190    163   F
Sonia   52    179    155   M
Martha  76    163    170   M
Andrew  39    183    183   F
Marcie  26    164    153   M
```

We use R code below to revert the gender

```
> levels(data$Sex)
[1] "F" "M"
> levels(data$Sex)=c("M","F")
> data
       Age Height Weight Sex
Joe     26    175    157   M
Nina    31    165    139   F
Mark    23    190    163   M
Sonia   52    179    155   F
Martha  76    163    170   F
Andrew  39    183    183   M
Marcie  26    164    153   F
```

b) We can use R code below to display the data in increasing order of the age

```
> newdata=data[order(data[,"Age"]),]
> newdata
       Age Height Weight Sex
Mark    23    190    163   M
Joe     26    175    157   M
Marcie  26    164    153   F
Nina    31    165    139   F
Andrew  39    183    183   M
Sonia   52    179    155   F
Martha  76    163    170   F
```

c) We can use R code below to display the data in decreasing order of the age

```
> newdata1=data[rev(order(data[,"Age"])),]
> newdata1
       Age Height Weight Sex
Martha  76    163    170   F
Sonia   52    179    155   F
Andrew  39    183    183   M
Nina    31    165    139   F
Marcie  26    164    153   F
Joe     26    175    157   M
Mark    23    190    163   M
```