STAT 40001/STAT 59800 **Statistical Computing** **Fall 2020**

Homework 1 - Solution

Name:

Due : September 10, 2020 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 than 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* ***sum*** **$(x > 5)$** *or* ***length*** **$(x[x > 5])$.**

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

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

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

c) $\ln 5 - \exp(2) + 2^3$

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

e) $\log 14 + \ln 14 + (47 mod\ 5)$

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

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

Q.N. 2) Create the following vectors using ***rep*** function:

```
V1= 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
V2= 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5
V3=  MATH,  MATH,  STAT, STAT, STAT, STAT, STAT, ECE, ECE, ECE, BIO, BIO
```

Solution:We use the following R codes to generate the desired sample

```
> rep(1:5,4)
 [1] 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
> rep(1:5,each=4)
 [1] 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5
 > noquote(rep(c("Math","STAT","ECE","BIO"),times=c(2,5,3,2)))
 [1] Math Math STAT STAT STAT STAT STAT ECE  ECE  ECE  BIO  BIO
```

Q.N. 3) An article entitled "What Does It Take to Heat a New Room? Estimating Utility Demand in a Home" by Robert Carver was published in the Journal of Statistics Education v.6, n.1 (1998). It has been observed that in a residential home, energy consumption is closely related to the outdoor temperature and size of the house. The date collected for this study are provided in the link below:
`http://jse.amstat.org/datasets/utility.dat.txt`

a)Import the data in R
b)How many variables are included in this dataset?
c)The missing values in this dataset are denoted by $*$. Please create a clean dataset by removing the missing vales.

Solution:
a) We use R code below to import the data in R

```
> data=read.table("http://jse.amstat.org/datasets/utility.dat.txt")
```

b)Based on the R output there are 81 observations with13 variables included in the given dataset

```
> dim(data)
[1] 81 13
```

c) Since the missing values are denoted by $*$ We use R code below to convert them to "NA" and count the number of missing values.

```
> data=read.table("http://jse.amstat.org/datasets/utility.dat.txt", na.strings="*")
> sum(is.na(data))
[1] 30
```

Based on the above results there is 30 missing entries in the dataset. Note that a single individual may have more than one missing entry. We use R code below to create the Clean dataset by removing all missing values

```
> Clean=na.omit(data)
> dim(Clean)
[1] 71 13
```

Q.N. 4) 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
```

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

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

Solution:

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

```
> Y=matrix(c(12,1,-2,-9,0),nrow=5)
> Y
     [,1]
[1,]   12
[2,]    1
[3,]   -2
[4,]   -9
[5,]    0
> 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) 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. 7) 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
```