STAT 40001/MA 59800 Statistical Computing/ Computational Statistics Fall 2013
Homework 1-Solution

Name:
Due : September 5, 2013 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* ***sum*** $(\boldsymbol{x > 5})$.

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) $\mathbf{\log 14+\ln 14}$

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

```
a)
> round((7-8)+5^3-5/6+sqrt(62),3)
[1] 131.041
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)),3)
[1] 3.785
```

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  Biology  Biology  Biology Statistics  Statistics Statistics Statistics
```

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
> rep(c("Math","Biology","Statistics"),times=2:4)
[1] "Math"       "Math"       "Biology"    "Biology"    "Biology"
[6] "Statistics" "Statistics" "Statistics" "Statistics"
>
```

Q.N. 3) 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?
f) Calculate the class average of both test 1 and test 2.

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

```
> 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
f)
> mean(Test1, na.rm=T)
[1] 84.42857
> mean(Test2, na.rm=T)
[1] 84.21429
```

Q.N. 4) 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. 5) Solve the system of equations

$$\begin{aligned} \mathbf{2x + 3y} &= \mathbf{21} \\ \mathbf{3x - 4y} &= \mathbf{23} \end{aligned}$$

Solution:

```
> A<-matrix(c(2,3,3,-4),nrow=2); Y<-matrix(c(21,23),nrow=2)
> A
     [,1] [,2]
[1,]    2    3
[2,]    3   -4
```

```
> Y
     [,1]
[1,]   21
[2,]   23
> X=solve(A)%*%Y
> X
     [,1]
[1,]    9
[2,]    1
```

Hence, the solution of the given system of equations is $\boldsymbol{(x, y) = (9, 1).}$