STAT 40001/MA 59800 Statistical Computing/ Computational Statistics Fall 2013
Homework 2- Solution

Name:

Due : September 17, 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.

Q.N. 1) Table 1 and Table 2 below are the test scores of 10 students in Test 1 and Test 2

Name	Test 1
Ana	56
Brian	78
Cathy	87
Dough	89
John	95
Lucas	98
Marcus	59
Nabin	78
William	87
Zoe	98

Table 1: Test 1 Scores

Name	Test2
Ana	86
Brian	67
Cathy	78
Dough	89
John	87
Lucas	67
Marcus	94
Nabin	78
William	81
Zoe	83

Table 2: test 2 scores

a) Use ***merge(.,.)*** to create a single table containing the student's test 1 and test 2 scores.

Solution: We can save both tables and import them individually and merge them.
Alternatively, we can use simply use the following codes to merge them

```
Names=c("Ana","Brian","Cathy","Dough","John","Lucas","Marcus","Nabin","William","Zoe")
Test1=c(56,78,87,89,95,98,59,78,87,98)
Test2=c(86,67,78,89,87,67,94,78,81,83)
Table1=data.frame(Names,Test1)
Table2=data.frame(Names,Test2)
```

```
> merge(Table1, Table2)
     Names Test1 Test2
1      Ana    56    86
2    Brian    78    67
3    Cathy    87    78
4    Dough    89    89
5     John    95    87
6    Lucas    98    67
7   Marcus    59    94
8    Nabin    78    78
9  William    87    81
10     Zoe    98    83
```

b) How many students did better in the second test?
Solution:

```
> sum(Test2>Test1)
[1] 2
```

Q.N. 2) The data frame ***Calculus*** from the ***PASWR*** package has the Mathematical assessment scores for 36 students enrolled in a biostatistics course according to whether or not the students had successfully completed a calculus course prior to enrolling in the biostatistics course.

a) Attach the dataframe ***Calculus*** and print the data set.
Solution: First we install the ***PASWR*** *package from the R GUI*
We can then use the following R command to attach the Calculus dataset in the PASWR package:

```
>library(PASWR)
>data(Calculus, package="PASWR")
>attach(Calculus)
>Calculus
> Calculus
   No.Calculus Yes.Calculus
1           73           82
2           39           90
3           55           85
4           72           87
5           88           86
6           64           79
7           57           85
8           58           92
9           75           89
```

```
10          44          82
11          76          92
12          68          82
13          64          85
14          55          87
15          62          92
16          61          85
17          76          95
18          40          90

> round(colMeans(Calculus),3)
 No.Calculus Yes.Calculus
      62.611       86.944
```

b)Calculate the average score of the students who didn't complete Calculus course.
Solution: From the R output above the average score of the students who didn't complete Calculus course is 62.611.

c) Calculate the average score of the students who completed the Calculus course.
Solution: From the R output above the average score of the students who completed Calculus course is 86.944.

Q.N. 3) Access the data from url ***http : //www.stat.berkeley.edu/users/statlabs/data/babies.dat*** and store the information in an object named **BABIES** using the function *read.table()*.
a) How many variables are included in the study?
b) Print first 10 lines of the data set.
c) Print last 10 lines of the data set.

Solution: The R code below will be used to answer these questions

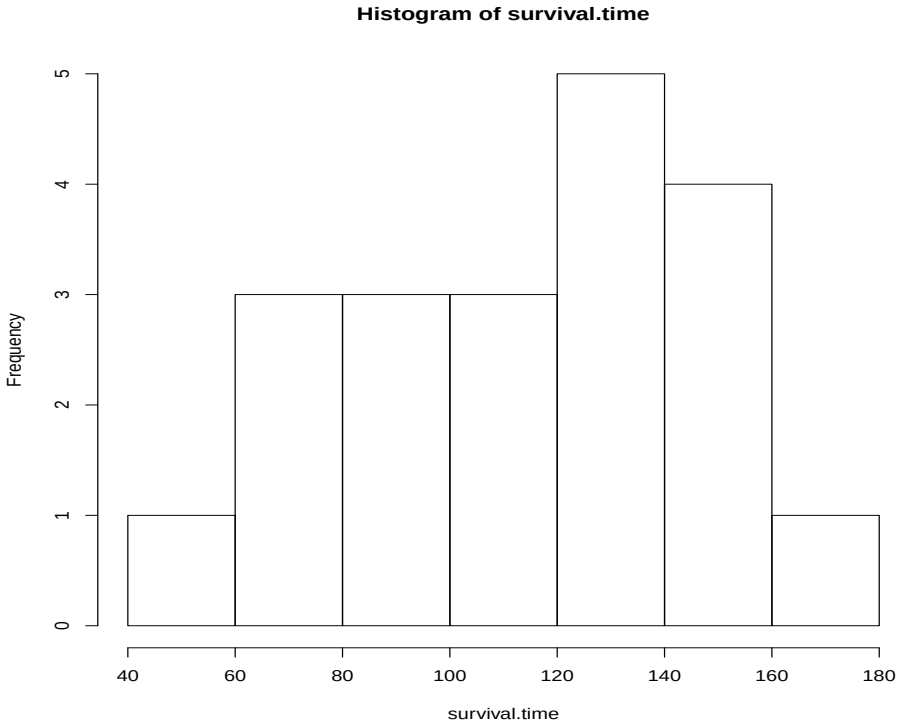
```
a)
> site<-"http://www.stat.berkeley.edu/users/statlabs/data/babies.data"
> BABIES<-read.table(site, header=T)
> length(BABIES)
[1] 7

b)
> head(BABIES,10)
   bwt gestation parity age height weight smoke
1  120       284      0  27     62    100     0
2  113       282      0  33     64    135     0
3  128       279      0  28     64    115     1
4  123       999      0  36     69    190     0
5  108       282      0  23     67    125     1
6  136       286      0  25     62     93     0
7  138       244      0  33     62    178     0
8  132       245      0  23     65    140     0
9  120       289      0  25     62    125     0
10 143       299      0  30     66    136     1
```

```
c)
> tail(BABIES,10)
     bwt gestation parity age height weight smoke
1227 109       244      1  21     63    102     1
1228 103       278      0  30     60     87     1
1229 118       276      0  34     64    116     0
1230 127       290      0  27     65    121     0
1231 132       270      0  27     65    126     0
1232 113       275      1  27     60    100     0
1233 128       265      0  24     67    120     0
```

Q.N. 4) The data frame ***Rat*** from the ***PASWR*** package has the survival time in weeks of 20 male rats exposed to high levels of radiation. Draw a histogram of the survival times of the rats.
Solution: We access the Rat dataframe using the code below

```
>library(PASWR)
>data(Rat, package="PASWR")
>attach(Rat)
> Rat
   survival.time
1            152
2            152
3            115
4            109
5            137
6             88
7             94
8             77
9            160
10           165
11           125
12            40
13           128
14           123
15           136
16           101
17            62
18           153
19            83
20            69
> hist(survival.time)
```



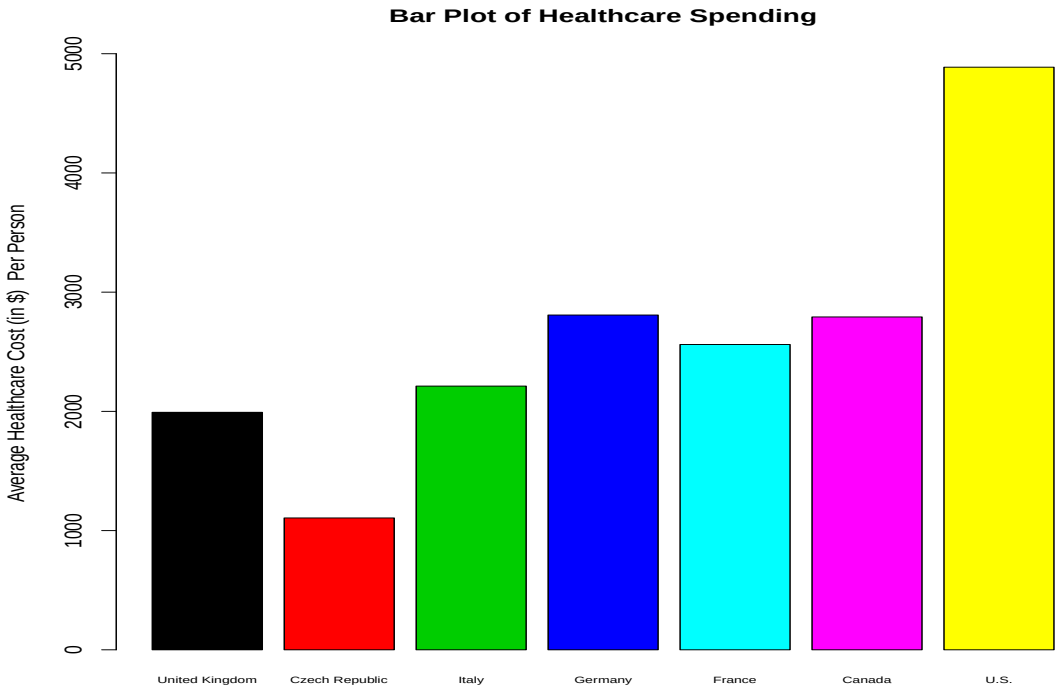
Q.N. 5) Consider the following data regarding the average spending on health care per person for various countries published in scholastic update in 2001.

Country	Amount (**$**)
United Kingdom	1992
Czech Republic	1106
Italy	2212
Germany	2808
France	2561
Canada	2792
U.S.	4887

Construct a bar graph to display the information.

Solution: We can use the R code below to construct a Bar graph. We can save the data and import it to display the bar graph
Or we can simply use the R code below:

```
>cost=c(1992,1106,2212,2808,2561,2792,4887)
>names(cost)=c("United Kingdom","Czech Republic","Italy","Germany","France", "Canada","U.S.")
>barplot(cost,col=c(1:7),ylab="Average Healthcare Cost (in $)  Per Person",ylim=c(0,5000),
main=" Bar Plot of Healthcare Spending ", cex.names=.6)
```


Bar Plot of Healthcare Spending
Average Healthcare Cost (in $) Per Person
0
1000
2000
3000
4000
5000
United Kingdom
Czech Republic
Italy
Germany
France
Canada
U.S.