TSE
Toulouse School of Economics
École d'économie de Toulouse

UNIVERSITÉ TOULOUSE 1 CAPITOLE

Université Toulouse 1 Capitole
Ecole d'économie de Toulouse

Année universitaire 2016-2017

Session 1

Semestre 1

Master 1 Economics & Statistics
Epreuve : Statistical Softwares for Data Scientists
Date de l'épreuve : 12 décembre 2016
Durée de l'épreuve : 1h30

Liste des documents autorisés : Aucun

Liste des matériels autorisés : Aucun

Nombre de pages (y compris page de garde) : 3

TSE
Toulouse School of Economics
École d'économie de Toulouse

UNIVERSITÉ TOULOUSE 1 CAPITOLE

Exercise 1 (4 points)

Let X the following matrix:

```
     [,1] [,2] [,3]
[1,]    2    1    2
[2,]    0    5    6
[3,]    4    3    5
```

What display the following instructions?

```
a) X[,2]          b) X[-1,]        c) X[2,]+X[,2]

d) X[c(1,3),]     e) X[,1:2]       f) X[-1,c(1,3)]
```

g) Give the instructions that enable to replace elements greater or equal to 4 by 1 and elements strictly lower than 4 by 0.

Exercise 2 (4 points)

Let's consider the cut() function presented as follows in the help of R:

Description

cut divides the range of x into intervals and codes the values in x according to which interval they fall. The leftmost interval corresponds to level one, the next leftmost to level two and so on.

Usage

cut (x, breaks, labels = NULL, include.lowest = FALSE, right, dig.lab = 3,
ordered_result = FALSE, ...)

Arguments

x	a numeric vector which is to be converted to a factor by cutting.
breaks	either a numeric vector of two or more cut points or a single number (greater than or equal to 2) giving the number of intervals into which x is to be cut.
labels	labels for the levels of the resulting category. By default, labels are constructed using "(a,b]" interval notation. If labels = FALSE, simple integer codes are returned instead of a factor.
include.lowest	logical, indicating if an 'x[i]' equal to the lowest (or highest, for right = FALSE) 'breaks' value should be included.
right	logical, indicating if the intervals should be closed on the right (and open on the left) or vice versa.
dig.lab	integer which is used when labels are not given. It determines the number of digits used in formatting the break numbers.
ordered_result	logical: should the result be an ordered factor?
...	further arguments passed to or from other methods.

Give (at least 6) various examples of calls for the cut() function and explain the matching between the arguments of the call and the parameters (also called 'formal arguments') of the function.

Exercise 3 (5 points)

Let's consider the following function:

```
> reverse
function(x) {
     print(x)
     if (length(x)<=1) return(x)
     return(c(x[length(x)], reverse(x[2:(length(x)-1)]),x[1]))
     print(x)
     }
```

What does the call `reverse(1:5)`?
Explain how the result returned by the function is computed.

Exercise 4 (7 points)

Propose a function that takes as input a dataframe and a list of columns names and for each one of the corresponding column of the dataframe replace the 5% lowest values and the 5% highest values by respectively the mean – the standard deviation and the mean + the standard deviation of the original distribution.