White Exam R II

Laurent Siksous

November 3, 2020

Description

The exam will take the form of MCQs. There will be 3 types of questions.

- questions about R and programming (5p)
- questions about understanding of R scripts (10p)
- questions asking you to use R in order to make calculations (5p)

Questions about R

Question 1

How missing values and impossible values are represented in R language?

a) NA
b) NAN
c) NA & NAN
d) Not Exist

Question 2

What is the best way for communicating the results of data analysis using the R language.

a) Single document
b) Files
c) Structures
d) Loadings

Question 3

The tapply() function takes 3 main arguments. What is the second ?

- A) The numeric variable to be resumed
- B) The classification factor
- C) The R command to use
- D) A function to be applied

Question 4

You can check to see whether an R object is NULL with the function :

- a) is.null()
- b) is.nullobj()
- c) null()
- d) as.nullobj()

Question 5

If you want to save a plot to a PDF file, which of the following is a correct way of doing that?

A) Construct the plot on the screen device and then copy it to a PDF file with dev.copy2pdf().

B) Construct the plot on the PNG device with png(), then copy it to a PDF with dev.copy2pdf().

C) Open the PostScript device with postscript(), construct the plot, then close the device with dev.off().

D) Open the screen device with quartz(), construct the plot, and then close the device with dev.off().

Question 6

Which of the following is not an object of R?

a) calls
b) expressions
c) strings
d) names

Question 7

What is the meaning of "<-"?

a) Functions
b) Loops
c) Addition
d) Assignment

Question 8

Which of the following can be used to display the names of (most of) the objects which are currently stored within R?

a) object()
b) objects()
c) list()
d) class()

Question 9

Which of the following truncates real x to integers ?

- A) as.numeric(x)
- B) as.integer(x)
- C) as.order(x)
- D) All of the above

Question 10

Point out the correct statement :

- A) The value NaN represents undefined value
- B) Number Inf represents infinity in R
- C) NaN can also be thought of as a missing value
- D) None of the above

Questions about interpretations of R code

Question 11

Suppose there are 2 dataframes "A" and "B". A has 34 rows and B has 46 rows. What will be the number of rows in the resultant dataframe after running the following command?

- A) 46
- B) 12
- C) 34
- D) 80

Question 12

To display the variable X distribution by the levels of a factor z (the two are stored in a dataframe d), what command do we use ?:

- A) boxplot(d[,c(x, z)])
- B) boxplot(dz, dx)
- C) boxplot(x ~ z, data=d)
- D) boxplot(z ~ x, data=d)

Question 13

What will be the output of the following R code snippet?

```
f <- function(num = 1) {
        hello <- "Hello, world!\n"
        for(i in seq_len(num)) {
                cat(hello)
         }
         chars <- nchar(hello) * num
         chars
}

f()
```

```
Hello, world!
[1] 14
```

a)
Hello, world! [1] 14 b)
Hello, world! [1] 15 c)
Hello, world! [1] 16
d) Error

Question 14

A dataset has been read in R and stored in a variable "dataframe". Missing values have been read as NA.

A	10	Sam
B	NA	Peter
C	30	Harry
D	40	NA
E	50	Mark

Which of the following codes will not give the number of missing values in each column?
A) colSums(is.na(dataframe))
B) apply(is.na(dataframe),2,sum)
C) sapply(dataframe,function(x) sum(is.na(x))
D) table(is.na(dataframe))

Question 15

Which of the following statement changes column name to h and f ?
A. colnames(m) <- c("h", "f")
B. columnnames(m) <- c("h", "f")
C. rownames(m) <- c("h", "f")
D. None of the above

Question 16

Which of the following code will sort the dataframe based on "Column2" in ascending order and "Column3" in descending order?

- A) dplyr::arrange(table,desc(Column3),Column2)

- B) table[order(-Column3,Column2),]
- C) Both of the above
- D) None of the above

Question 17

If I have two vectors x<- c(1,3, 5) and y<-c(3, 2), what is produced by the expression cbind(x, y)?

- A) a matrix with 2 columns and 3 rows
- B) a matrix with 3 columns and 2 rows
- C) a data frame with 2 columns and 3 rows
- D) a data frame with 3 columns and 2 rows

Questions about production of R code

You will find in data/ directory a file called hills.txt. This file contains the Scottish hill races (`https://dasl.datadescription.com/datafile/scottish-hill-races/`) You will work from now on on this data set and on a computational document that you will save as a pdf file and deliver at the end of the next 5 questions.

Question R1

How many hills of 2000 meters or more do we have ? Create a new column that will code 1 if true and 0 otherwise.

- A) 12
- B) 14
- C) 21
- D) 23

Question R2

Compute a linear regression of time against distance. What is a good estimation of the average speed of racers?

- A) 4
- B) 6
- C) 7
- D) 8

Question R3

Create a scatter plot of time against distance. Rename the x and why axis nicely and give title to your graph. Is there a correlation between these two features ? Add the line of regression to the graph.

Question R4

Create 2 box plots in the same graph for distance by the the two categories of hills (inferior or superior 2000 meters).

Question R5

Present a t test of student to compare the distance of each group of hills (superior of inferior 2000 meters). Is it significant ? What is the difference of means between the two groups ?