White Exam R I

Laurent Siksous

November 2, 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

Which of the following objects are not a data type of R ?

- A) Vectors and matrices
- B) Lists and arrays
- C) Factors
- D) Functions

Question 2

R functionality is divided into a number of what ?

- A) Packages
- B) Domains
- C) Classes
- D) All of the above

Question 3

The in R is a vector.

- A) Basic data structure
- B) Advance data structure
- C) Basic data types
- D) None of the Above

Question 4

To view a list of preloaded datasets in R, which command do you type into the console ?

- A) trace()
- B) data()
- C) library()
- B) summary()

Question 5

"dplyr" is one of the most popular package used in R for manipulating data and it contains 5 core functions to handle data. Which of the following is not one of the core functions of dplyr package?

- A) select()
- B) filter()
- C) arrange()
- D) summary()

Questions about interpretations of R code

Question 6

Which of the following command will remove an R object / variable named "santa" from the workspace?

- A) remove(santa)
- B) rm(santa)
- C) Both
- D) None

Question 7

A	10	Sam
B	20	Peter
C	30	Harry
D	!	?
E	50	Mark

Missing values in this csv file has been represented by an exclamation mark ("!") and a question mark ("?"). Which of the codes below will read the above csv file correctly into R ?

- A) csv('Dataframe.csv')
- B) csv('Dataframe.csv',header=FALSE, sep=',',na.strings=c('?'))
- C) csv2('Dataframe.csv',header=FALSE,sep=',',na.strings=c('?','!'))
- D) dataframe('Dataframe.csv')

Question 8

A	B
1	Right
2	Wrong
3	Wrong
4	Right
5	Right
6	Wrong
7	Wrong
8	Right

Suppose B is a categorical variable and we wish to draw a boxplot for every level of the categorical level. Which of the below commands will help us achieve that?

- A) boxplot(A,B,data=data)
- B) boxplot(A~B,data=data)
- C) boxplot(A|B,data=data)
- D) None of the above

Question 9

Consider the following function:

```
f <- function(x) {
      g <- function(y) {
             y + z
      }

      z <- 4
      x + g(x)
}
```

If we execute following commands (written below), what would be the output ?

```
z <- 10
f(4)
```

- A) 12
- B) 7
- C) 4
- D) 16

Question 10

What will be the output of following commands?

```
A <- paste(‘‘alpha’’,’’beta’’,’’gamma’’,sep=’’ ’’)
B <- paste(‘‘phi’’,’’theta’’,’’zeta’’,sep=’’’’)
parts <- strsplit(c(A,B),split=’’ ’’)
parts[[1]][2]
```

A) alpha
B) beta
C) gamma
D) phi

Question 11

One of the important phase in a Data Analytics pipeline is univariate analysis of the features which includes checking for the missing values and the distribution, etc. we wish to plot histogram for “age” variable. Which of the following commands will help us perform that task ?

- A) hist(data$age)
- B) ggplot2::qplot(data$age,geom=”Histogram”)
- C) ggplot2::ggplot(data=data,aes(data$age))+$\text{geom}_{\text{histogram}}()$
- D) All of the above

Question 12

Which of the following command will help us to rename the second column in a dataframe named “table” from alpha to beta?

- A) colnames(table)[2]=’beta’
- B) colnames(table)[which(colnames==’alpha’)]=’beta’
- C) setnames(table,’alpha’,’beta’)
- D) All of the above

Question 13

We wish to calculate the correlation between "Column2" and "Column3" of a "dataframe". Which of the below codes will achieve the purpose ?

- A) corr(dataframe$column2,dataframe$column3)
- B)

  (cov(dataframe$column2,dataframe$column3))/
  (var(dataframe$column2)*sd(dataframe$column3))

- C)

  (sum(dataframe$Column2*dataframe$Column3)-
  (sum(dataframe$Column2)*sum(dataframe$Column3)/nrow(dataframe)))/
  (sqrt((sum(dataframe$Column2*dataframe$Column2)-
  (sum(dataframe$Column2)^3)/nrow(dataframe))*
  (sum(dataframe$Column3*dataframe$Column3)-
  (sum(dataframe$Column3)^2)/nrow(dataframe))))

- D) None of the Above

Question 14

Which of the following commands will split the plotting window into 4 X 3 windows and where the plots enter the window column wise ?

- A) par(split=c(4,3))
- B) par(mfcol=c(4,3))
- C) par(mfrow=c(4,3))
- D) par(col=c(4,3))

Question 15

A Dataframe "df" has the following data:

2017-02-28
2017-02-27
...

After reading above data, we want the following output:

28 Tuesday Feb 17
27 Monday Feb 17

Which of the following commands will produce the desired output?

- A) format(df,"%d %A %b %y")
- B) format(df,"%D %A %b %y")
- C) format(df,"%D %a %B %Y")
- D) None of above

Questions about production of R code

We use a data set containing data regarding mental health in prison. Every observation corresponds to an interview conducted with an inmate. The data set is available in the file : "smp2.csv".

Explications des features de ce dataset

- l'âge, age
- la profession, prof
- la durée de la peine, quand elle a été prononcée, duree
- est-ce que le détenu est sous mesure disciplinaire, discip
- le nombre d'enfants, n.enfant
- la taille de la fratrie, n.fratrie
- la variable relative à la scolarisation du détenu qui va de 1 à 5, ecole
- Est-ce que le détenu a été séparé de sa famille quand il était enfant oui/non, separation
- Est-ce qu'il a bénéficié de l'aide d'un juge pour enfants quand il était enfant, juge.enfant
- Est-ce qu'il a été placé, place

- Est-ce qu'il a été victime d'abus, abus
- la gravité consensuelle, grav.cons

Nous retrouvons ensuite les variables diagnostiques,

- l'existence d'une dépression par le consensus du clinicien, dep.cons
- un trouble agoraphobique, ago.cons
- le syndrôme de stress post-traumatique, ptsd.cons

L'existence

- d'un abus d'alcool, alc.cons
- d'un abus de substances, subst.cons
- d'une schizophrénie, scs.cons

Ensuite nous avons,

- la variable char qui correspond à un score semi-quantitatif qui évalue l'importance, l'intensité d'un trouble de la personnalité sous-jacent, char

Nous retrouvons les trois dimensions de personnalité

- recherche de sensation, rs
- évitement du danger, ed
- dépendance à la récompense, dr

Puis trois variables relatives au risque suicidaire,

- d'abord un score de risque suicidaire, suicide.s
- ensuite l'existence d'un haut risque suicidaire, c'est une binarisation de la variable score suicidaire, suicide.hr
- et enfin l'existence d'antécédents de tentative de suicide, suicide.past

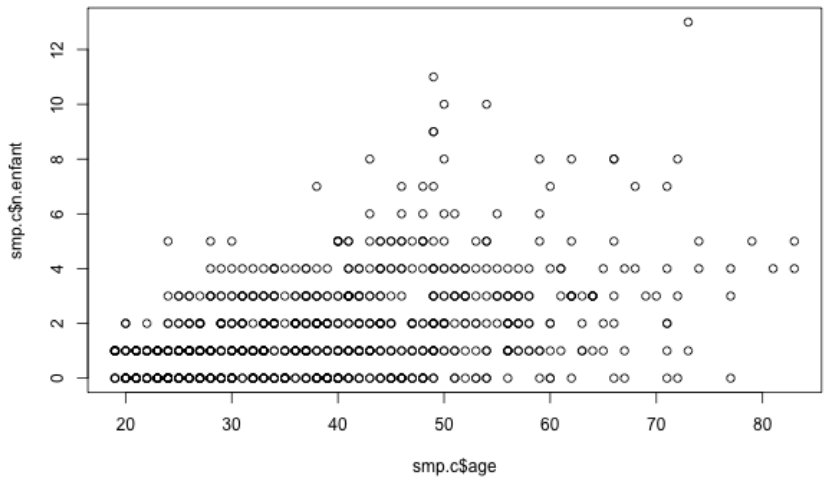
Et puis, comme dernière variable nous avons

- la durée de l'entretien que les enquêteurs ont passée avec le détenu, dur.interv

Question 16

How many inmates have more than 3 children ?

```
plot(smp.c$age,smp.c$n.enfant)
```



- A) 222
- B) 55
- C) 113
- D) 660

Question 17

Estimate the correlation coefficient between the age and the number of children of detainees ?

- A) 0.87
- B) 0.43
- C) -0.65
- D) 0.14

Question 18

Is there a significant correlation between age and sensation research ? Use a test to show if an older detainee has a lower score of sensation research. What is the lower end of confidence interval for this test ?

- A) -0.15
- B) 0.22
- C) -0.29
- D) No significant correlation

Question 19

We wish to verify if the interview duration varies whether the detainees have already tried to kill themselves or not, with the help of a Wilcoxon test. What is the degree of significance ?

- A) $p < 0.10$
- B) $p < 0.05$
- C) $p < 0.01$
- D) $p < 0.001$

Question 20

We'd like to predict how the interview duration varies considering several factors maybe interdependant. To do that, we build a multiple linar regression model on 4 traits : age, depression, drugs addiction and schizophrenia. How many minutes should we add or substract to the duration interview for a detainee with schizophrenia compared to an other ? (regardless of her age, state of depression or drug consumption)

- A) 2
- B) 7
- C) 5
- D) I don't know