

Exercises Introduction to R

Create a new script called "intro_R_XXX.R" (replace XXX by your surname) that contains the code of the following exercises. Upload the script in the campus virtual.

1. Code to execute a script called "myscript.R"
2. Code to assign the value A to a variable x
3. Code to generate a sequence from 7 to 30 with increment 3
4. Code to obtain information about function glm
5. Code to list all the objects in the current environment
6. Code to remove all objects
7. Code to specify the following path to the working directory:
C:\Desktop\Biostatistics
8. Create a vector x containing the numbers 1, 2, 1, 1, 1, 2
9. Create a vector y containing the words yes, no, no, yes, no
10. Compute the number of elements in vector y
11. Code to obtain the sequence of integer numbers from 10 to 25
12. Use the function rep() to generate the sequence 1, 2, 1, 2, 1, 2
13. Code to generate the sequence 1, 1, 1, 2, 2, 2
14. Code to generate a sequence containing 7 yes and 5 no
15. Code to obtain the sequence 40, 35, 30, 25, 20, 15, 10
16. Retrieve the fourth element in vector $x=(3, -1, 0, 2, -5, 7, 1)$
16. Retrieve the first three elements in vector $x=(3, -1, 0, 2, -5, 7, 1)$
17. Retrieve the first, second and fifth elements in vector $x=(3, -1, 0, 2, -5, 7, 1)$
18. Retrieve all the elements in vector $x=(3, -1, 0, 2, -5, 7, 1)$ except the second one
19. Change the value of the first and second elements in $x=(3, -1, 0, 2, -5, 7, 1)$ by 0
19. Assign the value 0 to the elements in $x=(3, -1, 0, 2, -5, 7, 1)$ that are larger than 2
20. Create a matrix M with 4 rows and 3 columns and fill it by rows with even numbers from 2 to 24
21. Obtain the number of rows and columns of matrix M
22. Retrieve the element in the first row and third column of matrix M

23. Retrieve all the elements in the third column of matrix M
24. Retrieve the third and fourth elements in the second column of matrix M
25. Retrieve a matrix containing all files in M except the first one
26. Add a new column at the beginning of matrix M with the integers from 1 to 4
27. Add a row at the end of matrix M with values 2, 4, 8
28. Compute the element-wise product between matrices `A<-matrix(1:9, nrow=3)` and `B<-matrix(1:9, nrow=3, byrow=T)`
29. Compute the matrix product between A and B
30. Compute the inverse matrix of A
31. Generate a data frame called "chol" (for cholesterol) containing the following variables (columns):
`id=(1, 2, 3, 4, 5)`, `gender=(1, 1, 2, 1, 2)`, `LDL=(237, 256, 198, 287, 212)`
32. In the previous data frame "chol" use the function "rownames()" to assign to each row the name of the patient: John, Peter, Hellen, Mat and Mary
33. Show the first 3 rows in data frame "chol"
34. Retrieve the LDL cholesterol levels of Peter his position in the data frame
35. Retrieve the LDL cholesterol levels of Peter using his name in the code
36. Save the LDL cholesterol levels of the 5 individuals in a new vector called "ldl_chol"
37. Create a new data frame named "chol_high" including only those individuals with LDL levels above 240
38. Let's consider the vector `x=(0.6, -1.3, 0.98, -0.4, 0.16)` and perform a `t.test` on x for the null hypothesis that the mean is equal to 0 and save the output in an object called `ttestx`
39. Show the attributes of object "ttestx"
40. From the output of "ttestx" retrieve the confidence interval of the mean
41. Check the data type of gender in data frame "chol"
42. Transform variable "gender" from data frame "chol" into a factor variable called `gender1` with 1=male and 2=female and with males as the reference group:
43. Transform variable "gender" from data frame "chol" into a factor variable called `gender2` with 1=male and 2=female and with females as the reference group:
44. Write the data frame "chol" into a text file called "cholesterol.txt"

45. Write the data frame "chol" into a csv file called "cholesterol.csv"
46. Write the code to install and load the R package "mbmdr" from CRAN
47. Write the code to install and load the R package "DESeq2" from Bioconductor
48. Get a numerical summary of $x=(1.5, 2.3, 4, 5.6, 2.1)$
49. Obtain the 40% percentile of $x=(1.5, 2.3, 4, 5.6, 2.1)$
50. Obtain the percentages of males and females in $gender=(1, 1, 2, 1, 2)$
51. Obtain the Pearson and Spearman correlation coefficient between $x=1:10$ and $y=x^2$
52. Test for the equality of variances in LDL cholesterol levels between males and females, assuming that LDL levels are normally distributed
53. Test for differences in LDL cholesterol mean levels between males and females, assuming that LDL levels are normally distributed
54. Test for differences in LDL cholesterol mean levels between males and females, without the assumption of normality
55. Plot a histogram of $x\leftarrow rnorm(100, 10, 4)$
56. Save the previous histogram in a pdf file called histogram.pdf
57. Write a function that for any number >0 returns $x*\log(x)$ and 0 for non positive numbers
58. Use a for loop to obtain the sum of $3^1, 4^2, 5^3, 6^4$ and 7^5
59. Use `apply()` to obtain the maximum of the rows in: `matrix(seq(1:20), ncol=4)`
60. Use `tapply()` to obtain the mean LDL levels of males and females in data frame "chol"
61. Assign a 0 to the elements in $x=(2.3, 1.4, -1.5, 1.2, 0.3)$ that are negative
62. Assign a 0 to the elements in $x=(2.3, 1.4, -1.5, 1.2, 0.3)$ that are negative and 1 otherwise
63. Sort decreasingly the elements in $x=(2, 7, 5, 1)$ with function `sort()`
64. Sort decreasingly the elements in $x=(2, 7, 5, 1)$ with function `order()`
65. Replace all NA values in a data frame "data" by a random number from a standard normal distribution ($mean=0, sd=1$)
66. Use function `grep()` to obtain the subset of ("Mary", "Peter", "Paul", "John", "Hellen") with names starting with P
67. Transform the vector ("Mary", "Peter", "Paul", "John", "Hellen") into ("Mary.1", "Peter.2", "Paul.3", "John.4", "Hellen.5")

68. Provide the indices in $x=(2.3, 1.4, -1.5, 1.2, 0.3)$ corresponding to positive values