

## BioSys PhD | Earthsystems PhD - Introduction to R

### EXERCISES

#### Basic Operations

1. Perform the following operations by assigning always an object to the expression:
  - (a)  $4 + 5(\log_{10}(4) - e^3)$
  - (b)  $\sqrt{37} + 4 + 5(\log_{10}(4) - e^3)$
  - (c)  $\sin(\pi \div 2) + \cos(\frac{2\pi}{3})$
2. Create a sequence of values in descending order between 17 and 26. How many elements does the sequence have?
3. Create a vector  $\mathbf{x}$  (as you wish) and consider the following conditions:
  - (a)  $x > 13$
  - (b)  $x < 13 \wedge x = 23$
  - (c)  $x \in [0, +\infty[$

What is the logical value of:

- $(a) \wedge (c)$
- $(b) \vee (c)$
- $\sim (b) \wedge (c)$

#### Vectors, Matrices and Data Frames

4. (a) Define two vectors for the variables:

<b>weight</b>	62	70	52	98	90	70
<b>height</b>	1.70	1.82	1.75	1.94	1.84	1.61

- (b) How many elements have more than 69kg and how many have less than 1.70m.

- (c) Calculate body mass index (BMI) using the previous vectors and save the results in an object. What kind of object is it and what is its size?
- (d) For the vector BMI, assign a name to each element in accordance with the following table:

BMI	Condition
< 18.5	Low weight
[18.5, 25[	Normal weight
[25, 30[	Overweight
≥ 30	Obesity

5. Build a list with the following information: An experiment conducted by the laboratory, 1000Experiments, consisted in the analysis of three kinds of drugs for the treatment of diabetes. A female volunteer, with 43 years, was given the three drugs, considering washing periods of 3 weeks between each administration. After every take was evaluated the blood sugar and for each drug the following results were obtained: 167, 245, 165. It was further evaluated the blood sugar before the treatments: 345.

6. Proceed with the following operations:

```
data(iris); plants<-iris; names(plants)
```

- (a) From the editor of R, change the name of the variables.
- (b) What is the dimension of the data frame?
- (c) Access only to the first 15 records.
- (d) List only the variable corresponding to the species.
- (e) Identify the species that have a length of petal greater than 6.
- (f) Apply operation `attach()` to access directly to the variables. For the species "setosa", how many plants have sepals width less than 3?
- (g) Apply operation `detach()`, proceed as follows and verify what the commands do:  

```
by(data=Petal.Length,INDICES=Species,FUN=summary)
boxplot(Petal.Length~Species, xlab="Species", ylab="Length",
main="Petal length by Species"),col=rainbow(3)
```

- (h) Save the data frame `plants` using function `write.table()`.

### Import Data

- 7. Access data frame `Orange` from package `datasets`.
- 8. Build a contingency table for the variables `year`, and `sex`, of data frame `Melanoma`, which is part of package `MASS`.

### Graphical Functions

- 9. Consider the data frame `Animals` from package `MASS`.
  - (a) Construct a graph (with subtitles on the axes) for the weight of the brain *vs.* the weight of the body.
  - (b) Repeat the previous step but now considering the natural logarithms of the variables. Use the labels to identify the points with higher body weight.
  - (c) Repeat the previous graphs but now side-by-side in the same device and store them in a `.jpeg` file.

### Loops and Functions

- 10. The following function calculates the mean and standard deviation of a numeric vector.

```
mean.dp <- function(x){  
  m <- mean(x)  
  dp <- sd(x)  
  c(mean=m, stand_dev = dp) }  
x<-c(2,4,6,8,10,12)  # for example  
mean.dp(x)
```

Modify this function so that:

- (a) is automatically applied to a vector of 20 numbers generated on the basis of the standard normal distribution, `rnorm()`, and calculates only the standard deviation;
- (b) if there are missing values, `NA`, the mean and the standard deviation are calculated from the remaining values.