Exercises to practice

1. Given a vector x <- c(2, 5, 120, 5, 8, 11), create a vector i that contains the indices of x.
2. Given the previous vector x, create a vector inv\_i that contains the indices of x, from highest to lowest.
3. Given the previous vector x, create a vector inv\_x that contains the values of x starting by the last one, then the previous to the last…
4. Suppose I made a mistake when creating vector x. I forgot the number 7, between 5 and 8 (indexes 4 and 5). How to fix this? I also forgot the last number: after 11 there should be a 3. Finally, the value 120 is a typographical error, that should be 12. Do the corrections.
5. Create two vectors x and y so that x contains the values from 0 to 5 with a step of 0.01, and y is equal to sin(x). Then make the plot of y vs x using the function plot(x, y).
6. Create a vector containing 12 values, the first 4 values should be equal to 0.5, the next 4 values should be 1.5, and the last 4 values should be 3.
7. Create the following vectors:

x <- c(1,3,5,7,9)

y <- c(2,3,5,7,11,13)

 Try to guess the result of the following commands before executing them:

1. x + 1
2. y \* 2
3. c(length(x), length(y))
4. x + y
5. sum(x>5)
6. sum(x[x>5])
7. sum(x[x>5 | x>3])
8. y[y>7]
9. Compute the mean of a vector x (for example, the vector x from the previous exercise) whithout using the function “mean()”. Compute, without using the function “var()”, the variance of the vector x, using the corrected sample variance formula . Compare the results with those given by the functions “mean()” and “var()”.
10. During a year, the amounts of the monthly invoices of your cell phone have been: 23, 33, 25, 45, 10, 28, 39, 27, 15, 38, 34, 29. How much have you spent during that year? Which has been the minimum? And the maximum? Which months have you made a payment higher that the average? Which percentage of months have you had an expense higher than the average expense?
11. Use a text editor to look at the data contained in the file “anscombe.txt”. There is no header with the names of the variables. The “;” character is used as a column (field) delimiter. The character “,” is used as a decimal separator. There are no row names. We are going to read those data into a data.frame called “ans”. The column names will be x1, y1, x2, y2, x3, y3, x4, y4. Use the R help for “read.table()” to understand the following command that read the text file:

ans <- read.table(file="anscombe.txt", sep=";", dec = ",", header=F, col.names=c ("x1","y1","x2","y2","x3","y3","x4","y4"))

Use the function “write.table()” to write the contents of the data.frame to a text file. Read the help to know its parameters and options. Can you tell what will do the following command?

write.table(ans, file="ans.txt", sep=";", dec=".")

1. Use “write.table()” to create a text file “ans13.txt” that contains just the data of the variables x1, y1, x3, y3. Make individual plots of (x1, y1), (x2, y2), (x3, y3), (x4, y4).
2. Given a normal population with mean 10 and standard deviation 2 (the usual way of denoting such a distribution is N(10, 2)):
3. Represent in a plot of its cumulative probability density function (you can do it by hand) the values A, B, C y D, such that:
   * A is higher than 90% of the population.
   * Between A and B there is 60% of the population.
   * C is lower than 80% of the population.
   * Between C y D there is 30% of the population
4. Which percentage of the population is to the left and to the right of each one of these points (A, B, C, D)?
5. A given population is distributed as a normal N(6, 2).
6. Determine an interval where the probability density is the highest and contains 70% of the population.
7. If a value is chosen randomly, which is the probability that it is in the interval (6, 7)? And in the interval (9, 10)?
8. Which is the probability that a value from this population, chosen randomly, be greater than 8? And lower than 4? What does the symmetry of the distribution have to do with these results?
9. If a sample value from this distribution is bigger than, at least, the 95% of the population, which are the possible values of that sample? And if it is lower than, at least, the 30%? Represent the results as an interval.