

R for Beginners

Day 1: Exercises 1

Exercise 1.1. The use of the c and sum functions.

This exercise uses epidemiological data. Vicente et al. (2006) analysed data from observations of wild boar and red deer reared on a number of estates in Spain. The dataset contains information on tuberculosis (Tb) in both species, and on the parasite *Elaphostrongylus cervi*, which only infects red deer.

In Zuur et al. (2009), Tb was modelled as a function of the continuous explanatory variable, length of the animal, denoted by LengthCT (CT is an abbreviation of cabeza-tronco, which is Spanish for head-body). Tb and Ecervi are shown as a vector of zeros and ones representing absence or presence of Tb and E. cervi larvae. Below, the first seven rows of the spreadsheet containing the deer data are given.

Farm	Month	Year	Sex	LengthClass	LengthCT	Ecervi	Tb
MD	11	00	1	1	75	0	0
MD	07	00	2	1	85	0	0
MD	07	01	2	1	91.6	0	1
MD	NA	NA	2	1	95	NA	NA
LN	09	03	1	1	NA	0	0
SE	09	03	2	1	105.5	0	0
QM	11	02	2	1	106	0	0

Using the c function, create a variable that contains the length values of the seven animals. Also create a variable that contains the Tb values. Include the NAs. What is the average length of the seven animals?

Exercise 1.2. The use of the cbind function using epidemiological data.

We continue with the deer from Exercise 1. First create variables Farm and Month that contain the relevant information. Note that Farm is a string of characters. Use the cbind command to combine month, length, and Tb data, and store the results in the variable, Boar. Make sure that you can extract rows, columns, and elements of Boar. Use the dim, nrow, and ncol functions to determine the number of animals and variables in Boar.

Exercise 1.3. The use of the vector function using epidemiological data.

We continue with the deer from Exercise 1. Instead of the cfunction that you used in Exercise 2 to combine the Tb data, can you do the same with the vector function? Give the vector a different name, for example, Tb2.

Exercise 1.4. Working with a matrix.

Create the following matrix in R and determine its transpose, its inverse, and multiple D with its inverse (the outcome should be the identity matrix).

$$D = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 2 & 1 \\ 2 & 3 & 0 \end{pmatrix}$$

Exercise 1.5. The use of the data.frame and list functions using epidemiological data.

We continue with the deer from Exercises 1 to 3. Make a data frame that contains all the data presented in the table in Exercise 1. Suppose that you decide to square root transform the length data. Add the square root transformed data to the data frame. Do the same exercise with a list instead of a data.frame. What are the differences?

Exercise 1.6. The use of the read.table and scan functions using deep sea research data.

The file ISIT.xls contains bioluminescent data. Prepare the spreadsheet (there are 4–5 problems you will need to solve) and export the data to an ascii file. Import the data into R using first the read.table function and then the scan function. Use two different names under which to store the data. What is the difference between them? Use the is.matrix and is.data.frame functions to answer this question.

Exercise 1.7. The use of the read.table or scan function using epidemiological data.

The file Deer.xls contains the deer data discussed in Exercise 1, but includes all animals. Export the data in the Excel file to an ascii file, and import it into R.