

Let's practice - 1

- 1) On your computer, create a folder for this course.
- 2) Download the `course_datasets.zip` file, unzip it in the course folder.
- 3) In RStudio, create a new project in the course folder.

Let's practice - 2

1) Get familiar with the different windows of Rstudio

- Console
- Environment / history
- Script
- Files / Help

2) Prepare your first script

- Open a new script file
- Save it as ex1.R
- Paste the following code, and use Ctrl+enter
- (cmd+enter on Mac) to execute each line

#ex1.R

```
# First Steps, ex 1
w <- 3
h <- 0.5
area <- w * h
area
```

Let's practice – 3

For all exercises, feel free to use

- cheat sheets provided*
- R help (? at command prompt)*

Go back to your script file, then:

- 1) Assign the values 6.7 and 56.3 to variables **a** and **b**, respectively.
- 2) Calculate $(2*a)/b + (a*b)$ and assign the result to variable **x**. Display the content of **x**.
- 3) Find out how to compute the square root of variables. Compute the square roots of **a** and **b** and of the ratio **a/b**.
- 4) a) Calculate the logarithm to the base 2 of **x** (i.e., $\log_2 x$).
b) Calculate the natural logarithm of **x** (i.e., $\log_e x$).

Let's practice - 4

- 5) Create two vectors, **vector_a** and **vector_b**, containing the values from -5 to 5 and from 10 down to 0 , respectively.
- 6) Calculate the (element-wise) sum, difference and product between the elements of **vector_a** and **vector_b**.
- 7) a) Calculate the sum of elements in **vector_a**.
b) Calculate the overall sum of elements in both **vector_a** and **vector_b**.
- 8) Identify the smallest and the largest value among both **vector_a** and **vector_b**.
- 9) Compute the overall mean of the values among both **vector_a** and **vector_b**.

Hint: Each task in exercises 6-9 can be performed in a single statement per vector

Let's practice – 5

- 1) Install and load the package “MASS”
- 2) The following command line loads the **bacteria** data.frame present in the MASS package. Execute it:
- 3) `> data(bacteria)`
- 4) What are the names of the columns of the **bacteria** data.frame ?
- 5) Use the `[]` , to select in **bacteria** rows 100 to 119 in the “trt” column.

Optional : 5) use comparison operators to count how many rows correspond to a “placebo” treatment (“trt” column).

Let's practice - 6

A dataset from mouse experiments at 18 weeks is available in the file ***mice_data.csv*** (courtesy of F. Schutz and F. Preitner). Let's explore the dataset to see what it contains.

- 1) Open a new script file in R studio, and save it.
- 2) Have look at the csv file in R studio's file explorer. What do you need to check in order to be able to read in the file correctly?
- 3) Read the file into R, assign its content to object "mice_data". Examine the object.
- 4) How many observations and variables does the dataset have?
- 5) What is the structure of the dataset? What are the names and classes of the variables?
- 6) Which variables appear to be categorical? Convert them to factors.
- 7) Get the summary statistics of "mice_data"

Let's practice – 6bis

- 8) Use the function `table()` to compute the number of observations in different mouse groups. a) How many mice are included of each genotype (WT, KO)? b) How many mice are included per diet (HFD, CHOW)? c) Make a 2x2 table by genotype and diet crossed.
- 9) Compute the means and standard deviations for WT and KO mouse weights using `tapply()`. Then do the same for CHOW and HFD groups.
- 10) Isolate the observations for the mice on high fat diet (HFD) using `subset()`. Compute a summary statistics just for the weights of the subset. Then do the same for the mice on regular chow diet (CHOW). Export the data of each subgroup to a csv file.
- 11) Look at the results from the two previous exercises. What does this initial exploration of the data suggest about mouse weights?