Exercises for Lecture 2

Statistical Computing with R, 2023-24

Exercise 1: create an R script for your solutions

During today's lecture you have seen how to create an `R` scripts, and to add code and comments to it.

1. If you haven't done so yet, create a folder in your laptop where you can save all the files that you will download and create during this course
2. Create a subfolder for the material from lecture 2, and save today's slides and material from the practical therein
3. Create an `R` script and save it in the dedicated subfolder
4. Write your solutions to the exercises below in that `R` script. Don't forget to save your progress regularly!
5. Remember to use comments to separate the different exercises and questions, as well as to comment the code itself

Exercise 2

Let

$$A = \begin{bmatrix} -2 & 5 & 3 \\ 1 & -4 & 2 \end{bmatrix}, B = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, C = \begin{bmatrix} 3 & 4 \\ 1 & -4 \end{bmatrix}.$$

Compute the following quantities:

1. $B + C$
2. $B - 2C$
3. $B \cdot A$
4. $A^T \cdot C$
5. C^{-1}

Exercise 3

Use R to compute:

1. $\sin(e^5)$
2. $\sqrt{x+49}$ for $x = 0, 5, 17$
3. Round the results from point (2) to the second digit
4. $\sqrt[3]{x+7}$ for $x = 0, 1, 4$
5. $\ln(e^7 + 12.5^2)$
6. $\log_4(e^7 + 12.5^2)$
7. $|x^2 - 4|$ for $x = 0, 1, 2, 3$
8. $\sum_{x=6}^{20} \log_3(\sqrt{x})$
9. $\sum_{x=6}^{20} \frac{\log_3(x)}{2}$

Exercise 4

In this session we will use two datasets from the R package **brolgar**:

- the dataset **heights** contains information on the average height of male adults in (a selection of) countries across the world from 1500 to 2000;
- the dataset **wages** contains measurements from a survey that collected data on hourly wages, years in the workforce, education and race.

To be able to work with these datasets, you first need to import them in R. To do so, follow these steps:

1. install the R packages **remotes**, which is a requirement to install **brolgar**:

```
# Install and load `remotes`
install.packages('remotes')
library(remotes)

# install `brolgar`
install_github("njtierney/brolgar")
```

2. load the package:

```
library(brolgar)
```

3. execute the following lines of code:

```
data(heights)
heights = as.data.frame(heights)
data(wages)
wages = as.data.frame(wages)
```

Exercise 5

Use the functions `head( )` and `tail( )` to start looking into the `heights` dataset.

1. Print in the console the first 10 rows of the data frame
2. Print the last 15 rows of the data frame
3. How many variables does this data frame contain? What are their names? Can you create a vector that contains these variable names?

Exercise 6

1. How many rows does `heights` contain?
2. Which country does the record at row 1245 refer to?
3. Which rows contain observations from Portugal?
4. Create a data frame that contains only the observations from Portugal. How many observations are available? For which years?
5. Has the mean height in Portugal constantly increased over time?

Exercise 7

Now, let's focus on the data on heights in 1970 and 2000.

1. Create two data frames, one with the 1970 data and one with the 2000 data
2. For how many countries is the 1970 mean height available? And the 2000 mean height?
3. Use the function `intersect( )` to identify the countries for which data from both 1970 and 2000 are available
4. Did the mean height of men in Vietnam increase or decrease from 1970 to 2000?

Exercise 8

In this exercise we compare the distribution of mean height in Asian and American countries in 1980.

1. Create two data frames, one with data from American countries in 1980, and one with data from Asian countries in 1980.
2. What was the range of mean heights across Asian countries in 1980? And in the Americas?
3. Compute the median of the mean heights in Asia and in the Americas
4. Which American country is the country where men are on average the tallest?

Exercise 9

In this exercise we will explore the `wages` data frame that we loaded in exercise 4. Before you start, type `?wages` in the console, and read the help page associated to this data frame.

1. How many rows and columns does this data frame contain?
2. The variable `ln_wages` contains the natural logarithm of hourly wages. Create a new variable, called `hourly_wage`, that contains the value of the hourly wages
3. What is the value of the lowest hourly wage? And of the highest?
4. Compute the mean and median hourly wage across all observations

In the next points, we will check whether the mean hourly wage increases with the number of years in the workforce.

5. Split observations from workers with ≤ 2 years of experience, between 2 and 5 years of experience, and > 5 years of experience into three separate data frames
6. Compute the mean wage for workers with at most 2 years of work experience
7. Did workers with 2 to 5 years of experience earn more, *on average*, than workers with at most 2 years of experience? What about workers with more than 5 years of experience?