

Máster universitari en Estadística i Investigació Operativa (MESIO)
STATISTICAL SOFTWARE:R and SAS
Laboratory with R (GROUP B)

IMPORTANT: Last day to deliver the practice is on 12-31-2025 (23.59h) through the ATENEA campus.
The practice should be done in groups of 2-3 people (maximum 3 people and minimum 2 persons)

PROBLEM STATEMENT: Europe is affected by climate change and its effects are not only received on earth. European water bodies (lakes, rivers and oceans and seas of the continent) are also affected. Given that there is more water than land on the planet's surface, it is not surprising that the warming of the oceans has accounted for about 93% of the global warming since the 1950s. This warming occurs as a result of the increase in greenhouse gas emissions, especially carbon dioxide, which in turn trap more and more solar energy within the atmosphere. Most of this trapped heat is eventually stored in the oceans, which affects the temperature and the circulation of water.

Sea surface temperatures on European coasts are rising more rapidly than those in the world oceans. Water temperature represents one of the most important regulatory elements of marine life, so that temperature increases are already causing major changes under the surface of the water (www.eea.europa.eu)

In the file *data frame sea_temperature.xls* the temperature is displayed (grados centígrados) of sea water at different depths (0, -20, -50, -80m) from 2000 to 2017, measured at the Estartit observation point on the Costa Brava (1 mile east of the Medes Islands (Girona); coordinates: 4203 N, 315 E).

This table shows the temperatures of each month by depth and the temperatures of each month for a period of about 30 years. The average temperatures of the periods are also presented.

The description of this data can be found in the IDESCAT link:
<https://www.idescat.cat/pub/?id=aec&n=218&t=2000>

The objective of your work is to contribute to the study of climate change, especially that produced in the last 30 years, through the use of real data.

What is requested is to carry out a study as complete as possible (descriptive, graphical, functions, bookstore of R) on the change of the sea temperature by depth for the different years and months studied, as well as in the different depths and their comparison with the previous periods.

Different functions should be performed as described in the following different sections.

Exercise 1 (4 puntos)

- a) Import the study data (*data frame sea_temperature.xls*) and place them in a *data frame* con nombre **sea.deep**. Name the variables according to the specifications in the data set and place labels describing units and what each variable and case is (years, months, depth, etc.). You should place it properly according to the format of the data.

If you can, import the data directly from the data file in the link indicated above, without having

- to save it previously to file.
- b) Use the function `label` of package `Hmisc` to tag variables (using of data frame `sea.deep`).
 - c) Indicate the dimension of the data-frame and make a descriptive of the variables.
 - d) Represent by means of Boxplots the average temperatures by depth and year.
 - e) Calculate the mean, median, standard deviation and the interquartile range for each of the previous groups (or other statistics if necessary) . You can present other statistics suitable for this type of data, which you think are convenient.
 - f) Properly represent the data to be able to see the annual variations of the average temperature in the total depths and years.
 - g) Export *data frame* with the new variables created to a new file sheet, for example in `NUEVO.xlsx`.

Exercise 2 (2 puntos)

Make two or more graphs with the data from Exercise 1 using functions from two of the following packages:

gplots <http://cran.r-project.org/web/packages/gplots/index.html>,
plotrix <http://cran.r-project.org/web/packages/plotrix/index.html>,
vcd <http://cran.r-project.org/web/packages/vcd/index.html>.

For (at least) one of the two graphs you should use the study group variable (depth or year) and for the other one a couple of the numerical variables. Present the syntax and interpret both graphs.

Schedule a function that does the following,

Exercise 3 (4 puntos)

- a) Calculate the temperature between one month and the following month (eg January and February) for each year and depth, offer a graph per year and the average of all years studied. You should be able to indicate what years or depths it should represent.
- b) Calculate the difference between the temperature of each month and the same month of the previous 30 years (example: January 2000 and January period 1974-1999), offer a graph for each year and the average of All years studied. You should be able to indicate what year or depths it should represent.
- c) Verify the operation of the function with the *data frame* `sea.deep` and ensure that everything is correct (optionally you can invent a data frame to verify it, with simulated temperatures using a normal truncated distribution of temperatures and with the same depths).
- d) Comment on the results found and if you think they have effects on climate change. Optionally, you can build a library to project the future of seawater temperature at a certain temperature for the following years.
- e) Optionally, Look for other data (public repositories, other sources), if you think it is necessary and cross them with those already available, especially if they help to better explain the possible climate change that is taking place
- f) Optionally, create a library in R to store the previous functions, the source data and place it in a public repository such as Github (see at: <https://github.com/>)