Language of Instruction: English
Professor: Pere López Brosa, Ferran Carrascosa Mallafrè
Professor's Contact and Office Hours: at agreed hours
Course Contact Hours: 45 hours
Recommended Credit: 6 ECTS credits
Weeks: 4
Course Prerequisites: An introductory course on probability and statistics is basic for enrolment to this course. For the students of UPF, the compulsory requirement is the Probability and Statistics of the second year in the studies of ECO/ADE/IBE.
Language Requirements: None

Course Description:
Over the recent years, R (https://www.r-project.org) has become the leading tool for statistical computing and graphics. The basic language of R is greatly enhanced by numerous contributed packages submitted by users. The majority of computing in the leading applied statistical journals is done in R, and it is used almost exclusively in some of the leading-edge applications, such as in genetics and data sciences. This software permits data analysts to interact with their data and to design personalized protocols for statistical analysis. R is free software that can run in most of the computer platform systems (Windows, OS of Mac, Unix, etc.). The purpose of this course is to set a foundation for full exploitation and creative use of this statistical modern language for computing and graphics. The course will also include a gentle introduction to Python, a platform that has been essential in data processing in the recent years.

Much of the statistical methodology implemented in software packages is used in the form of a black box. This is advantageous for a user who is not interested in the details of the methods, but the result is often a second-rate application, because the implementation, even if of high quality, is often meant for a different context, small details in the setting of options are ignored or misunderstood, and the orientation in the output, formatted for general interest, is difficult.

The course will introduce students to the syntax and inner workings of R and Python, to become proficient in everyday computational tasks with datasets of all kinds, skilled in applications of elementary statistical methods, with an emphasis on (initial) data exploration and simple graphics. Focus will also be placed on opportunities to enhance the learning experience in other statistical courses by illustrating and applying basic statistical concepts using R.

Learning Objectives:
At the end of the course, students will have learned:

· to use fundamental tools for computing in the practice of quantitative analytical methods (the 'paper-and-pencil' tool of the 21st century), that can work for the small jobs (like a pocket calculator) as well as for the big jobs (complex statistical data analysis)

· programming, data handling, transformations, subsetting, exploratory data analysis, probability distributions and simulations, regression and linear models, summarizing data, how to handle large data sets, effective graphics.

· modern concepts of statistics based on simulations and writing a report of a quantitative analysis

Course Workload
The course is divided into lectures, discussions, practice with portable computers, and tutoring.
Methods of Instruction:
The course includes both lectures and field studies. Two-hour classroom sessions are normally divided into one-hour lecture and one-hour of practice in computing. Students are required to come to class with their own laptops. An applied project is an essential component of the classroom activities.
Both the exercises and project should be elaborated with RMarkdown and/or Python notebooks.

Method of Assessment
Assessment is composed of the following inputs:

- Continual Evaluation: contribution to class + homework (20%)
- Main Project (45%)
- Final Exam (35%) (A minimum of 15 points, out of 35, is required in the final exam to pass the course)

The main project will involve some computing in R/Python and submission of a report of up to 6 typed pages (not counting appendices).
Students will select their projects from topics of their own interest (upon the acceptance of the instructors) and will make a brief oral presentation at the end of the course.

Absence Policy

Attending class is mandatory and will be monitored daily by professors. The impact of absences on the final grade is as follows:

Absences	**Penalization**
Up to two (2) absences	No penalization.
Three (3) absences	1 point subtracted from final grade (on a 10 point scale)
Four (4) absences	2 points subtracted from final grade (on a 10 point scale)
Five (5) absences or more	The student receives an INCOMPLETE for the course

The BISS attendance policy does not distinguish between justified or unjustified absences. The student is deemed responsible to manage his/her absences.

Emergency situations (hospitalization, family emergency, etc.) will be analyzed on a case by case basis by the Academic Director of the UPF Summer School.

Classroom Norms:

· No food or drink is permitted.
· There will be a ten-minute break during the class.
· Students must come to class fully prepared.

Course Contents:

1. Introduction to R
2. Variables and data structures.
3. Data input and output
4. Basic and advanced graphics
5. Flow control and programming structures
6. Exploratory data analysis
7. Linear models, Matrix algebra
8. Simulation, permutation tests and bootstrap
9. Optimization
10. Introduction to Python

Required Readings: Handout material will be posted on the web as the course evolves.

Recommended bibliography:
Students are encouraged to consult the following sources on their own.

Dalgaard, P., *Introductory Statistics with R*, Springer, 2002.

Dennis, B., *The R Student Companion*, Taylor & Francis Group, 2013.

Matloff, N., *The Art of R Programming: A Tour of Statistical Software Design*, William Pollock, 2011.

Chihara, L. and Hesterberg, T., *Mathematical statistics with resampling and R*, Wiley, 2011.

Lander, J. P., *R for Everyone: Advanced Analytics and Graphics*, Addison-Wesley Data & Analytics Series, 2014.

J. VanderPlas, *Python Data Science Handbook*, O'Reilly Media, Inc., 2016.

Last revised, Februrary 2022