Course Syllabus- MODERN STATISTICAL COMPUTING

Language of Instruction: English
Professor: Vladimir Zaiats Protchenko
Professor's Contact and Office Hours: vladimir.zaiats@upf.edu, send an e-mail to get an appointment
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 UPF students, the compulsory requirement is the Probability and Statistics of the second year in the studies of ECO/ADE/IBE.

Language Requirements: Recommended level in the European Framework B2 (or equivalent: Cambridge Certificate)

Course Description:
Statistical computing is a highly sought-after analytical data analysis skill both in professional and research environments. This course presents modern graphical displays and data manipulation methods, interactive and reproducible reporting, emphasizing statistical methods and computation related to regression and classification methods such as linear, generalized linear and non-linear models. The concepts are introduced in R (https://www.rproject.org) one of the leading programming languages for statistical computing. Knowledge of R is highly-valued by companies in many sectors for positions related to data science, quantitative analysis or finance. R is also an indispensable tool in most research fields, including Economics, Finance, Marketing, Biomedicine, etc. R provides a rich set of off-theshelf data analysis tools, and the possibility to design our own data processing and analyses. R runs in all operating systems (Windows, Mac, Unix-like) and is a free open-source language that is enhanced by an extensive list of user-contributed packages. The purpose of this course is to introduce students to statistical computing, including flexible regression data analysis methods, and to advanced R skills. The idea is that students learn by doing. Therefore, there is a strong applied emphasis, all concepts are driven by examples discussed in class, where students are given the code to reproduce them. Students will become skilled in applications of elementary statistical methods, with an emphasis on data exploration, graphics and programming. Focus will also be placed on opportunities to enhance the learning experience in other statistical courses.

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

· To use a fundamental data analysis tool for quantitative analytical methods

· Programming, data handling, exploratory data analysis, linear / generalised linear/ non-linear regression, summarising data, effective graphics, model-free computational methods (bootstrap, permutation tests, cross-validation)

· Preparing notebooks to automatically perform quantitative analyses and create reports in formats such as pdf and html, with interactive elements

Course Workload
The course is constituted by lectures and practices. Students should be prepared to work with computer codes on a permanent basis.

Methods of Instruction:
The course includes both lectures and practice with laptop computers. The teaching philosophy is that students learn by doing.
Classroom sessions are normally split into a lecture and a practice part.
Students are required to attend classes with their own laptops.

Method of Assessment
Homework + Class contribution: 20 percent
Controls, in-class exercises: 40 percent
Final project: 40 percent

The Final Project (in groups of 2 students) requires a report to be submitted of up to 10 typed pages (not counting appendices). Students will select their projects from topics of their own interest (accepted by the course 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 course instructors. The impact of absences on the final grade is as follows:

Absences	**Penality**
Up to two (2) absences	No penalty.
Three (3) absences	1 point subtracted from the final grade (on a 10 point scale)
Four (4) absences	2 points subtracted from the final grade (on a 10 point scale)
Five (5) absences or more	The student receives an INCOMPLETE grade 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 analysed on a case-by-case basis by the Academic Director of the UPF Summer School.

Classroom Norms:

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

Course Contents:

Week 1
Introduction to R
Graphical displays (ggplot)
Data manipulation

Week 2
Programming basics
Statistical models: linear, generalized linear models. Non-linear regression

Optimisation

Week 3

Computational inference methods: bootstrap, permutation tests, cross-validation
Model comparison techniques

Week 4

Advanced reports: interactive plots, dashboards
Methods of flexible data analysis

Required Readings: The instructor will assemble a coursepack/or indicate mandatory textbooks.

Recommended bibliography:

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

Wickam, H., Grolemund, G. *R for the Data Science*. O'Reilly, https://r4ds.had.co.nz.

Lander, Jared P. *R for Everyone: Advanced Analysis and Graphics*. Boston etc.: Pearson Education, Inc, 2017.

Last updated, January 2023