Programming in R

Academic Year: (2023 / 2024)   **Review date: 22-07-2023**

Department assigned to the subject: Statistics Department

Coordinating teacher: MARIN DIAZARAQUE, JUAN MIGUEL

Type: Compulsory ECTS Credits : 3.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

None.

OBJECTIVES

The student will acquire the following knowledge:
1. Proficiency in the R programming language and the R-studio working environment.
2. Mastering the different types of data structures.
3. Exploratory data analysis techniques and presentation of results through data visualization techniques.
4. Familiarity with the main data analysis packages of R.
5. Be able to perform a simulation properly.
6. Accelerate the programs implemented by means of parallel programming.
7. Find errors and bottlenecks in their code and generate reports.

DESCRIPTION OF CONTENTS: PROGRAMME

1. Basics of Programming I.

    The R-studio environment. Types of data (Arrays, Lists, Factors, Data Frames,...) and their operations. Control structures. Functions.

2. Basics of Programming II.

    Advanced data structures. Reading and storage of data.

3. Data visualization.

    The ggplot2 package.

4. Introduction to some useful packages in R.

    MASS, Caret, dplyr and data.table packages.

5. Simulations.

6. Parallel programming.

7. Debugging, Profiling and presentation of results with Rmarkdown.

LEARNING ACTIVITIES AND METHODOLOGY

The classes consist of a mixture of presentations on the fundamental concepts of the subject and the presentation of practical cases through the use of R software. Students are expected to bring their own laptops to experiment with the code during the lectures.

* Training activities

  - AF1: Theoretical lesson.
  - AF2: Practical lesson.
  - AF5: Tutorials.
  - AF6: Group work.

- AF7: Individual work.
- AF8: On-site evaluation tests.

* Teaching methodologies

- MD1: Class lectures by the professor with the support of computer and audiovisual media, in which the main concepts of the subject are developed and the bibliography is provided to complement the students' learning.
- MD2: Critical reading of texts recommended by the professor of the subject: press articles, reports, manuals and/or academic articles, either for later discussion in class, or to expand and consolidate the knowledge of the subject.
- MD3: Resolution of practical cases, problems, etc. posed by the teacher individually or in groups.
- MD4: Presentation and discussion in class, under the moderation of the professor of topics related to the content of the subject, as well as case studies.
- MD5: Preparation of papers and reports individually or in groups.

ASSESSMENT SYSTEM

The subject will be evaluated through the delivery of four short practices that will be handed in throughout the weeks of the course. They will also have to deliver a work based on a guide accessible in Aula Global, consisting of the analysis of a real database using the programming techniques learnt during the course.

% end-of-term-examination:	0
% of continuous assessment (assigments, laboratory, practicals…):	100

BASIC BIBLIOGRAPHY

- Felicidad Marques Asension R en profundidad. Programación, gráficos y estadística, RC, 2017

- Fox, J. Using the R Commander: A Point-and-click Interface for R, CRC Press., 2016

- Irizarry, R.A. Introduction to data science: data analysis and prediction algorithms with R, Boca Raton, Florida. CRC Press, 2020
- Wickham, H., & Grolemund, G. (2016) R for data science: import, tidy, transform, visualize, and model data, O'Reilly Media, Inc., 2016

BASIC ELECTRONIC RESOURCES

- . Introduction to Data Science Data Analysis and Prediction Algorithms with R: https://rafalab.github.io/dsbook/
- . R for Data Science: https://r4ds.had.co.nz/index.html

- . R Programming for Data Science: https://bookdown.org/rdpeng/rprogdatascience