
Academic Year: (2024 / 2025)

Review date: 30-04-2024

Department assigned to the subject: Statistics Department

Coordinating teacher: UCAR MARQUES, IÑAKI

Type: Compulsory ECTS Credits : 6.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Introduction to Programming with R (19151)

OBJECTIVES

- Knowledge of data programming structures and procedures.
- Ability to import tabular data in a variety of formats with the R programming language.
- Ability to work with remote databases.
- Ability to prepare, clean, transform and enrich tabular data for further modeling and visualization with R and SQL programming languages.

DESCRIPTION OF CONTENTS: PROGRAMME

1. R base programming
 - 1.1. Introduction to R ecosystem
 - 1.2. Introduction to RStudio. Working with projects
 - 1.3. Basic data types
 - 1.4. First uses of functions and packages. Basic operations
 - 1.5. Understanding errors
2. From cell to dataset
 - 2.1. Concatenate values: vectors (variables)
 - 2.2. Basic operations with vectors
 - 2.3. Loops vs. vectorial programming. Control flow structures
 - 2.4. First databases: matrices and data.frames
 - 2.5. Tibbles as standard type for databases. Datapasta package
3. Tidy data
 - 3.1. R base vs. tidyverse. Pipe operator
 - 3.2. Principles of tidy data: tidy vs. messy data
 - 3.3. Pivoting datasets
4. RMarkdown and quarto: report results
5. Tidyverse
 - 5.1. Operations by rows. Cleaning data: NA values and duplicates
 - 5.2. Operations by columns
 - 5.3. Aggregating and recategorizing variables
 - 5.4. Group variables: group_by and .by
 - 5.5. Summaries
 - 5.6. Joining datasets
 - 5.7. Import/export from/to different formats
 - 5.8. Use of APIs
6. Advanced data types
 - 6.1. Categorical variables: forcats package

- 6.2. Handling characters: stringr package
- 6.3. Handling dates: lubridate package
- 6.4. Handling lists: purrr package. Functional programming

- 7. Advanced data management
 - 7.1. dbplyr package: from tidyverse to SQL
 - 7.2. arrow package: handling massive databases

- 8. SQL programming
 - 8.1. Introduction to relational databases
 - 8.2. Data handling and querying
 - 8.3. Complex queries, aggregations and subqueries
 - 8.4. Joining tables

LEARNING ACTIVITIES AND METHODOLOGY

Training Activities:

- Theoretical-practical classes
- Tutorials
- Group work
- Individual student work

Teaching Methods:

- Presentations in the professor's lecture room with computer and audiovisual support, in which the main concepts of the subject are developed and a bibliography is provided to complement the students' learning.
- Resolution of practical cases, problems, etc. raised by the professor, either individually or in a group.
- Presentation and discussion in class, under the moderation of the professor, of topics related to the content of the subject, as well as practical case studies.
- Developing pieces of work and reports, individually or in group.

ASSESSMENT SYSTEM

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

- Participation in the class (10%)
- Individual work done during the course (60%)
- Group work done at the end of the course (30%)

In the extraordinary call, the evaluation system will be as follows:

- 1) Exam: 100%

BASIC BIBLIOGRAPHY

- Hadley Wickham R for Data Science, O'Reilly, 2017

ADDITIONAL BIBLIOGRAPHY

- Chester Ismay and Albert Y. Kim Statistical Inference via Data Science: a Modern Dive into R and the tidyverse, Chapman & Hall, 2022
- Steph Locke Data Manipulation in R, Locke Data, 2017