

Academic Year: (2022 / 2023)

Review date: 16-05-2022

Department assigned to the subject: Statistics Department

Coordinating teacher: GARCIA PORTUGUES, EDUARDO

Type: Electives ECTS Credits : 3.0

Year : 1 Semester : 2

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Programming in R
Advanced Programming

OBJECTIVES*** Basic competences**

- CB6: Possess and understand the knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
- CB7: Know how to apply acquired knowledge and problem-solving skills in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their area of study.
- CB8: Integrate knowledge and face the complexity of making judgments based on information that, being incomplete or limited, includes reflections on the social and ethical responsibilities linked to the application of their knowledge and judgments.
- CB9: Communicate conclusions, as well as the knowledge and the ultimate reasons that support them, to specialized and non-specialized audiences in a clear and unequivocal manner.
- CB10: Develop the learning skills that enable further study in a manner that is largely self-directed or autonomous.

*** General competences**

- CG1: Apply the techniques of analysis and representation of information, to adapt it to real problems.
- CG2: Identify the most appropriate statistical model for each real problem and know how to apply it for its analysis, design and solution.
- CG3: Obtain scientifically viable solutions to complex real statistical problems, both individually and in teams.
- CG4: Synthesize the conclusions obtained from data analysis and present them clearly and convincingly in a bilingual environment (Spanish and English), both written and oral.
- CG5: Generate new ideas (creativity) and anticipate new situations, in the contexts of data analysis and decision making.
- CG6: Apply social skills for teamwork and to relate with others in an autonomous way.

*** Specific competences**

- CE1: Apply advanced knowledge of statistical inference in the development of methods for the analysis of real problems.
- CE2: Use free software such as R and Python for the implementation of statistical analysis.
- CE5: Apply advanced statistical fundamentals for the development and analysis of real problems involving the prediction of a variable response.
- CE6: Apply nonparametric models for the interpretation and prediction of random phenomena.
- CE8: Apply and develop visualization techniques for samples collected with open source software such as R and Python.
- CE9: Correctly identify the type of statistical analysis corresponding to specific objectives and data.
- CE10: Apply statistical modeling in the treatment of relevant problems in the scientific field.
- CE11: Formalize random phenomena and model them by means of probabilistic models.

- CE12: Apply models for supervised and unsupervised learning.
- CE13: Model complex data with stochastic dependence.
- CE14: Apply advanced knowledge and skills in statistical consulting.

* Learning outcomes

Acquisition of knowledge on: 1) skills useful in a statistical consulting service; 2) techniques for automatic presentation of results in reports; 3) development of Shiny applications; 4) the tidyverse environment; 5) the tidymodels environment.

DESCRIPTION OF CONTENTS: PROGRAMME

This course covers several tools for streamlining the consulting pipeline in R: from data wrangling to presentation of results, passing through a fast statistical modeling. The focus is placed on seeing the main features of many different packages and solutions.

1. Advanced R Markdown for reporting
 - 1.1. Advanced topics on R Markdown
 - 1.2. Writing good reports
 - 1.3. Customized presentations
 - 1.4. Other documents and topics
2. Shiny applications
 - 2.1. Main paradigm
 - 2.2. Examples of simple applications
 - 2.3. Reactions and appearance
 - 2.4. More advanced applications
 - 2.5. flexdashboard
 - 2.6. Other topics
3. Data wrangling in the tidyverse I
 - 3.1. dplyr
 - 3.2. tidyr
 - 3.3. readr
 - 3.4. tibble
 - 3.5. Other packages
4. Data wrangling in the tidyverse II
 - 4.1. stringr
 - 4.2. forcats
 - 4.3. lubridate and hms
 - 4.4. glue
 - 4.5. purrr
 - 4.6. Other packages
5. Fast modeling using AutoML
 - 5.1. Introduction to AutoML
 - 5.2. Explainability
 - 5.3. Examples in regression
 - 5.4. Examples in binary classification
 - 5.5. Examples in multiclass classification
6. Fast modeling with tidymodels I
 - 6.1. broom
 - 6.2. rsample
 - 6.3. parsnip
 - 6.4. yardstick
 - 6.5. Other packages
7. Fast modeling with tidymodels II
 - 7.1. recipes
 - 7.2. workflows
 - 7.3. tune
 - 7.4. infer
 - 7.5. Other packages

The program is subject to modifications due to the course development and/or academic calendar.

LEARNING ACTIVITIES AND METHODOLOGY

The classes consist of a mixture of lectures on the software and its practical use. The statistical language R is used. 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

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

The evaluation in the ordinary call is done entirely by continuous evaluation. This is done by a mixture of:

- (a) a set of practical exercises;
- (b) active participation in lessons.

The continuous evaluation grade (in the scale 0-10) is

$$\min(A + 0.10 * B, 10),$$

where

- A (in the scale 0-10) is the grade of the practical exercises;
- B (in the scale 0-10) is the grade corresponding to (b).

Students who have not followed the continuous evaluation may take a final exam in the ordinary call with a value of 60% of the final grade.

The grade in the extraordinary call is established by the delivery of a set of practical exercises.

Further details are provided in Aula Global. The evaluation is subject to modifications due to the course development and/or academic calendar.

BASIC BIBLIOGRAPHY

- Hadley, W. and Golemund, G. R for Data Science, O'Reilly, 2017
- Xie, Y., Allaire, J.J., and Golemund, G. R Markdown, CRC Press/Chapman & Hall, 2019

BASIC ELECTRONIC RESOURCES

- Kuhn, M. and Silge, J. . Tidy Modeling with R: <https://www.tmr.org/>