Universitat Autònoma de Barcelona

2023/2024

Complex Data Analysis

Code: 104399
ECTS Credits: 6

Degree	Type	Year	Semester
2503740 Computational Mathematics and Data Analytics	OB	2	2

Contact

Name: Pedro Puig Casado

Email: pere.puig@uab.cat

Teaching groups languages

You can check it through this link. To consult the language you will need to enter the CODE of the subject. Please note that this information is provisional until 30 November 2023.

External teachers

Dorota Mlynarczyk

Prerequisites

It is recommended a good knowledge of the course Modelling and Inference and to have some fluency in the software R.

Objectives and Contextualisation

The main objective is to provide statistical tools for data analysis, mastering the most relevant techniques to cope with complex models.

Competences

- Calculate and reproduce certain mathematical routines and processes with ease.
- Formulate hypotheses and think up strategies to confirm or refute them.
- Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
- Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
- Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.

- Use computer applications for statistical analysis, numerical and symbolic computation, graphic visualisation, optimisation and other to experiment and solve problems.
- Using criteria of quality, critically evaluate the work carried out.
- Work cooperatively in a multidisciplinary context assuming and respecting the role of the different members of the team.

Learning Outcomes

1. Analyse data using inference techniques for one or two samples.
2. Choose the appropriate statistical software to analyse the data through inference techniques.
3. Identify statistical inference as an instrument of prognosis and prediction.
4. Identify the distinct sources of information available.
5. Interpret obtained results and provide conclusions that refer to the experimental hypothesis.
6. Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
7. Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
8. Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.
9. Understabd the distinct methods of data collection.
10. Use statistical software to manage databases.
11. Use statistical software to obtain the summary indexes of study variables.
12. Use the properties of distribution function.
13. Use the properties of the density function.
14. Using criteria of quality, critically evaluate the work carried out.
15. Validate and manage information to carry out statistical processing on this.
16. Work cooperatively in a multidisciplinary context, taking on and respecting the role of the distinct members in the team.

Content

1- Linear models: multiple regression and ANOVA.

2- Generalized linear Models: logistic and Poisson regression.

3- Resampling methods 1: permutation tests.

4- Resampling methods 2: bootstrap.

5- Resampling methods 3: jackknife.

If we have time, we will also include an introduction to Principal Component Analysis.

Methodology

Accordingly with the aims of the subject, the development of the course will be based on the following activities:

Theoretical lectures: The student acquires the scientific and technic skills of the subject assisting to the theoretical lectures and complementing them with his/her personal work on the topics explained. The theoretical lectures are the activities demanding less interactiveness: they are conceived like a fundamentally unidirectional method of transmission of knowledge of the teacher to the student. The lectures will be given

using a support of slides (PowerPoint) in English that will be uploaded also at the Virtual Campus.

Problems and practices: The workshop of exercises and practical sessions have a double mission. On the one hand the students will work with the scientifical and technical issues explained in the theoretical lectures to complete its understanding developing a variety of activities, since the typical resolution of problems until the discussion of practical cases. On the other hand, the workshop of exercise are the natural forum at which argue in common the development of the practical work.

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Theoretical lectures	24	0.96	14, 9, 3, 4, 5, 8, 6, 7, 16, 13, 12, 15
Workshop of exercises	20	0.8	14, 9, 3, 4, 5, 8, 6, 7, 16, 13, 12
Type: Supervised			
Practical sessions	20	0.8	1, 14, 2, 16, 10, 11, 15
Type: Autonomous			
Personal working	61	2.44	14, 2, 3, 4, 5, 6, 7, 13, 12, 10, 11

Assessment

The evaluation runs continuously along the course. The continued evaluation has several fundamental aims: To check the process of education and learning and to verify that the student has attained the corresponding skills of the course.

This is the method of evaluation: The practical exercises delivered by the students (30%), a partial examination of Theory in the middle of the course (35%), and another partial examination of Theory at the end of the course (35%). The second-chance examination only will be allowed to the students having a minimum score of 3 at the final mark, recovering only the part corresponding to the Theory.

The students who chose the single assessment modality must take a final test that will consist of an exam in which there may be questions of theory and problem-solving and a practice exam in front of the computer. This test will be carried out on the same day, time, and place in which the test of the second partial is carried out. Anyone who misses the test without a valid excuse will be classified as NOT EVALUABLE. If a grade of less than a 5 is received, it may be recovered on the same day, at the same time, and in the same location as the other students in the course.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Exercises	30	20	0.8	14, 2, 16, 10, 11

Partial exam 1	35	2.5	0.1	1, 9, 3, 4, 5, 8, 6, 7, 13, 12, 15
Partial exam 2	35	2.5	0.1	1, 9, 3, 4, 5, 8, 6, 7, 13, 12, 15

Bibliography

- Introduction to Linear Regression Analysis. Montgomery, D. Peck, A. Vining, G., 2001.
- An R Companion to Linear Statistical Models. Christopher Hay-Jahans, 2012.
- Generalized Linear Models. McCullagh, P. and Nelder, J., 1992.
- Resampling methods: a practical guide to data Analysis. Phillip I. Good, 2006.
- The jackknife, the bootstrap and other resampling plans. Bradley Efron, 1982.
- Bootstrap methods and their application. A.C. Davison, D.V. Hinkley, 1997.

Software

We'll utilize the R programming language.