UAB
Universitat Autònoma de Barcelona

Complex Data Analysis

Code: 104399
ECTS Credits: 6

2024/2025

Degree	Type	Year
2503740 Computational Mathematics and Data Analytics	OB	2

Contact

Name: Amanda Fernandez Fontelo

Email: amanda.fernandez@uab.cat

Teaching groups languages

You can view this information at the end of this document.

Prerequisites

It is recommended that students have some knowledge of probability and statistical inference, as well as some knowledge of the R software.

Objectives and Contextualisation

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

Learning Outcomes

1. CM14 (Competence) Implement strategies to confirm or refute hypotheses.
2. CM15 (Competence) Manage the information for validation through statistical processing.
3. CM15 (Competence) Manage the information for validation through statistical processing.
4. CM16 (Competence) Assess, using the data obtained, inequalities on the grounds of sex/gender.
5. KM12 (Knowledge) Identify statistical inference as a tool for forecasting and prediction.
6. KM12 (Knowledge) Identify statistical inference as a tool for forecasting and prediction.
7. KM12 (Knowledge) Identify statistical inference as a tool for forecasting and prediction.
8. KM12 (Knowledge) Identify statistical inference as a tool for forecasting and prediction.
9. KM14 (Knowledge) Identify the usefulness of Bayesian methods, applying them appropriately.
10. KM14 (Knowledge) Identify the usefulness of Bayesian methods, applying them appropriately.
11. SM14 (Skill) Use the properties of density and distribution functions.
12. SM14 (Skill) Use the properties of density and distribution functions.
13. SM15 (Skill) Use suitable statistical software to manage databases, to obtain summary indices of the study variables and to analyse data using inference techniques.
14. SM15 (Skill) Use suitable statistical software to manage databases, to obtain summary indices of the study variables and to analyse data using inference techniques.
15. SM15 (Skill) Use suitable statistical software to manage databases, to obtain summary indices of the study variables and to analyse data using inference techniques.

Content

Topic 1- Linear models: multiple regression and ANOVA.
Topic 2- Generalized linear models: logistic and Poisson regression.
Topic 3- Regularization: Lasso and Ridge regressions.
Topic 4- Big data in linear and generalized linear models.
Topic 5- Resampling methods: Permutation tests and bootstrap.

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Theoretical lectures	24	0.96	
Workshop of exercises	20	0.8	
Type: Supervised			
Practical sessions	20	0.8	
Type: Autonomous			
Personal working	61	2.44	

Following the objectives of the course, its development will be based on the following activities

- Theoretical lectures: The student acquires scientific and technical knowledge of the subject by supporting the theoretical lectures and complementing them with personal work on the topics introduced. Theoretical lectures require less interactivity: they are conceived as a unidirectional method of transmitting knowledge from the teacher to the student. The lectures are given with the help of slides in English, which are also uploaded on the Virtual Campus.
- Problems and exercises: Sessions of exercises and practical sessions have a double mission. On the one hand, the students will work with the scientific and technical issues explained in the theoretical lectures to complete their understanding, developing a number of activities, from the typical solution of problems to the discussion of practical cases. On the other hand, the exercise sessions are a natural forum to 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.

Assessment

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Exercises	30	20	0.8	CM14, CM15, CM16, KM12, KM14, SM14, SM15
Partial exam 1	35	2.5	0.1	CM15, KM12, SM14

Partial exam 2	35	2.5	0.1	CM15, KM12, SM14

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 s k i l l s   o f   t h e   c o u r s e .

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.

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.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Hastie T., Tibshirani, R., Friedman, J. 2009.
- 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.

Language list

Name	Group	Language	Semester	Turn
(PLAB) Practical laboratories	1	Catalan/Spanish	second semester	morning-mixed
(SEM) Seminars	1	Catalan/Spanish	second semester	morning-mixed
(TE) Theory	1	Catalan/Spanish	second semester	morning-mixed