UAB
Universitat Autònoma de Barcelona

Complex Data Analysis

Code: 104399
ECTS Credits: 6

2025/2026

Degree	Type	Year
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

Students should have a basic understanding of probability and statistical inference, as well as familiarity with the R software.

Objectives and Contextualisation

The course provides statistical tools for data analysis and ensures proficiency in the most relevant techniques for dealing 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. CM16 (Competence) Assess, using the data obtained, inequalities on the grounds of sex/gender.
4. KM12 (Knowledge) Identify statistical inference as a tool for forecasting and prediction.
5. KM14 (Knowledge) Identify the usefulness of Bayesian methods, applying them appropriately.
6. SM14 (Skill) Use the properties of density and distribution functions.
7. 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 linear regression and ANOVA.
Topic 2- Generalized linear models: logistic and Poisson regressions.
Topic 3- Big data in linear and generalized linear models.
Topic 4- Resampling methods: Bootstrap.
Topic 5 (subject to available time)- Regularization: Lasso and Ridge regressions.

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Lecture sessions	27	1.08	CM14, CM15, CM16, KM12, SM14, SM15, CM14
Resolution of certain laboratory problems and exercises during face-to-face sessions	14	0.56	CM14, CM15, CM16, KM12, SM14, SM15, CM14
Type: Supervised			
Resolution of laboratory problems in class	8	0.32	CM14, CM15, CM16, KM12, SM14, SM15, CM14
Type: Autonomous			
Complete all laboratory practice tasks independently	35	1.4	CM14, CM15, CM16, KM12, SM14, SM15, CM14
Resolution of theory-based problems	14	0.56	CM14, CM15, CM16, KM12, SM14, CM14
Self-directed learning to deepen understanding of lecture topics	35	1.4	CM14, CM15, CM16, KM12, SM14, CM14

The course will be developed based on the following activities, in line with its objectives:

- Theory-type sessions (lectures): Students will acquire scientific and technical knowledge of the course topics by attending these lectures and complementing them with personal study of the topics introduced. These lectures require less interactivity as they are designed to transmit knowledge unidirectionally from professor to student. The lectures are delivered with the aid of slides in English, which are also uploaded to the course Moodle page. It is essential to complement these slides with a book from the recommended bibliography of the course.
- Theory-related problems and laboratory sessions: Theory-related exercises and laboratory sessions serve two purposes. Firstly, students deepen their understanding of the scientific and technical issues introduced in the lectures by completing various activities, ranging from solving problems to discussing practical cases. Secondly, laboratory sessions provide an ideal setting for discussing the development of practical work. Although both the problem sets and laboratory practical exercises will be addressed during the corresponding in-person sessions, they must also be completed independently at home. The time available during problem-solving sessions and laboratory classes is limited and does not allow for all the necessary exercises and practical cases to be covered in full to successfully achieve the course objectives. Therefore, independent work on both problems and practical exercises is essential to ensure that the sessions are conducted as efficiently and effectively as possible.

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
Bootstrap project (BP)	14	6	0.24	CM14, CM15, CM16, KM12, KM14, SM14, SM15
Evaluable theory-based and practical problems (PP)	30	6	0.24	CM14, CM15, KM12, SM14, SM15
Exam 1 (E1)	35	3	0.12	CM14, CM15, KM12, SM14
Exam 2 (E2)	21	2	0.08	CM14, CM15, KM12, SM14

The assessment process takes place throughout the course and has several key objectives: (i) To monitor the teaching and learning process, enabling both students and instructors competency to address any issues that arise and achieve the desired level of competency, (ii) To encourage students to make continuous efforts and avoid unproductive last-minute overexertion, and (iii) To verify that students have achieved the competencies established in the curriculum.

Continuous evaluation modality

For this course, the continuous evaluation modality will consist of a first exam (E1, 35%), a second exam (E2, 21%), a final project on bootstrap resampling techniques (BP, 14%, non-recoverable) and evaluable theory-related and practical problems (PP, 30%, non-recoverable). The evaluable theory-related and practical problems (PP) will be solved during the corresponding in-person class sessions and must be submitted at the end of each session. These activities may be completed in pairs, which should change throughout the course. Late submission of evaluable activities without justification will incur a penalty for the corresponding activity. In addition, plagiarism will result in an automatic grade of 0 for the affected activity. The final grade (F) will be calculated as follows:

$F = E1 \times 0.35 + E2 \times 0.21 + PP \times 0.30 + BP \times 0.14.$

If a student does not achieve a grade of at least 5 for the course and wishes to pass, they must take the resit exam (R), which includes the opportunity to retake activities E1 and E2, but not activities BP and PP. For students taking the resit exam, the final course grade will be:

$F = \min(R \times 0.56 + PP \times 0.30 + BP \times 0.14, 5).$

It is not possible to take the resit exam to improve apassing grade.

Single evaluation modality

Students who opt for the single evaluation modality must complete a final evaluation consisting of a final exam with theoretical questions and problem-solving tasks (E, 56%). In addition, they must submit the results of a set of theory-based and practical problems (different from those required in the continuous evaluation modality, but covering similar content) (PP, 30%, non-recoverable), as well as the final project on bootstrap resampling techniques (BP, 14%, non-recoverable). This evaluation will take place on the same day, time, and location as the second exam of the continuous evaluation modality (E2). The final grade (F) will be calculated as follows:

$F=E \times 0.56 + PP \times 0.30 + BP \times 0.14.$

If a student does not achieve a final grade of at least 5 and wishes to pass, they must take the resit exam, where only the exam component (E) can be recovered. Activities PP and BP cannot be recovered. For these students, the final grade (F) will be calculated as follows:

$F = \min(R \times 0.56 + PP \times 0.30 + BP \times 0.14, 5).$

The resit exam will take place on the same day, time, and location as the resit exam of the continuous evaluation modality (R). Taking the resit exam to improve an already passing grade is not permitted.

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 will use the R programming language.

Groups and Languages

Please note that this information is provisional until 30 November 2025. You can check it through this link. To consult the language you will need to enter the CODE of the subject.

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