

Modelling and Inference

Code: 104392
ECTS Credits: 6

Degree	Type	Year	Semester
2503740 Computational Mathematics and Data Analytics	OB	2	1

Contact

Name: Amanda Fernandez Fontelo
Email: amanda.fernandez@uab.cat

Use of Languages

Principal working language: spanish (spa)
Some groups entirely in English: No
Some groups entirely in Catalan: No
Some groups entirely in Spanish: No

Other comments on languages

Evaluation, including exam questions, will be given in Catalan. Lecture slides as well as exercise sheets and lab tutorials will all be in English. In-person lectures will be offered in Spanish.

Prerequisites

Good knowledge of the contents of the courses taken during the first year of the bachelor's degree is considered very important, especially those of probability and calculus.

Objectives and Contextualisation

This is the first course in the Bachelor's degree that focuses on Statistical Inference, a branch of statistics that uses data from a "representative" sample to acquire information about a population. The course is required throughout the Bachelor's degree, as it covers different concepts and techniques that serve as the basis for many of the topics introduced in upcoming courses within the Bachelor's. In particular, the course will start with a brief introduction to statistics, followed by a chapter on parameter estimation (both point and based on confidence intervals), and finally chapters on frequentist-based significance tests and an introduction to classical linear regression models.

To protect everyone's safety, in-person teaching and evaluable activities will be adjusted in accordance with health authority recommendations.

Competences

- Calculate and reproduce certain mathematical routines and processes with ease.
- Formulate hypotheses and think up strategies to confirm or refute them.
- Relate new mathematical objects with other known objects and deduce their properties.
- 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.

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. Describe the basic properties of timestamp and interval estimators.
4. Identify statistical distributions.
5. Identify statistical inference as an instrument of prognosis and prediction.
6. Interpret obtained results and provide conclusions that refer to the experimental hypothesis.
7. Recognise the usefulness of Bayesian methods and apply these appropriately.
8. 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.
9. 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.
10. 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.
11. Understand the distinct methods of data collection.
12. Use statistical software to manage databases.
13. Use statistical software to obtain the summary indexes of study variables.
14. Use the properties of distribution function.
15. Use the properties of the density function.
16. Using criteria of quality, critically evaluate the work carried out.
17. Validate and manage information to carry out statistical processing on this.

Content

Preliminaries of Probability (reminder): Probability and random variables. Law concept. Discrete-valued distributions. Density and probability functions. Expectation and variance. Moment generating function. Examples.

Topic 1. Introduction to Statistics.

1. Descriptive statistics and inferential statistics.

1.1. Basic concepts in inference: statistical population and sample; parameters, statistics and estimators.

1.2. Statistical models: parametric and non-parametric.

2. Most common statistics: the sample moments. The order statistics.

3. Distribution of some statistics.

3.1. From a sample of a Normal population: Fisher's theorem.

3.2. The Central Limit Theorem: asymptotic normality of sample moments and proportion.

Topic 2: Point estimation.

1. Point estimators: definition and properties.

1.1. Bias

- 1.2. Comparison of estimators without bias. Relative efficiency
- 1.3. Comparison of estimators with bias: the mean square error.
- 1.4. Consistency of an estimator.

2. Methods to obtain estimators.

- 2.1. Method of moments.
- 2.2. Method of maximum likelihood (MLE)
 - 2.2.1. Invariance of the likelihood.
 - 2.2.2. Score function and Fisher information.
 - 2.2.3. Cramer-Rao inequality.
 - 2.2.4. Properties of the MLE.
 - 2.2.5. Delta method.
 - 2.2.6. Numerical procedures for determining MLE.

Topic3. Estimation by confidence intervals.

- 1. Concept of confidence region and interval.
- 2. The "pivot" method for the construction of confidence intervals.
- 3. Confidence intervals for the parameters of a population.
 - 3.1. For the mean of a Normal population with known and unknown deviations.
 - 3.2. For the variance of a Normal population with known and unknown means.
 - 3.3. Asymptotic confidence intervals: Wald, Score and LRT.
- 4. Confidence intervals for the parameters of two populations.
 - 4.1. Confidence intervals with independent samples.
 - 4.2. Confidence intervals for the difference of means of two Normal populations with paired data.

Topic 4: Significance tests.

- 1. Introduction.
 - 1.1. Type I and II errors.
 - 1.2. Power function.
 - 1.3. Tests consistency.
 - 1.4. p-values.
 - 1.5. Duality between confidence intervals and significance tests.
- 2. Tests for the parameters of a population.
 - 2.1. For the mean of a Normal population with known and unknown deviations.
 - 2.2. Asymptotic tests for the mean of a population when the sample is large.

2.3. For the variance of a Normal population.

3. Tests for the parameters of two populations.

3.1. Hypothesis tests with independent samples.

3.2. Tests of hypotheses with paired data.

Topic 5. Simple linear regression model.

1. Purpose of the model.

2. Ordinary least squares (OLS) estimators.

3. Inference based on the linear regression model.

4. Predictions.

IMPORTANT: In teaching, the gender perspective involves reviewing androcentric biases and questioning the assumptions and hidden gender stereotypes. This revision involves including the contents of the subject, the knowledge produced by scientific women, often forgotten, seeking the recognition of their contributions, as well as that of their works in the bibliographical references. Efforts will also be made to introduce the most practical part of the subject, the analysis and comparison of statistical data by sex, commenting on the classroom causes and the social and cultural mechanisms that can sustain the observed inequalities.

Methodology

The course is organized into lecture, exercise and lab sessions.

In lectures, we will introduce the concepts and techniques outlined in the course program. Given that the content is mostly based on the standard topics of an introduction to statistical inference course, the recommended bibliography can be used to follow the course. Lecture slides and related material will be available in Moodle. The exercise sessions are intended to work through and understand statistical concepts. Each exercise will be available in Moodle, as will the solutions (after they have been solved in sessions). The goal of the lab sessions is to learn how to apply the methods given in lectures using the statistical software R, as well as how to evaluate the findings. Outlines for lab sessions will be accessible in Moodle as well.

IMPORTANT: To work more comfortably with R, it is recommended to use the RStudio interface: it is free, "Open source" and works with Windows, Mac and Linux. <https://www.rstudio.com/>

OBSERVATION: The gender perspective in teaching goes beyond the contents of the subjects, since it also implies a revision of the teaching methodologies and of the interactions between the students and the teaching staff, both in the classroom and outside. In this sense, participatory teaching methodologies, where an egalitarian, less hierarchical environment is generated in the classroom, avoiding stereotyped examples in gender and sexist vocabulary, with the aim of developing critical reasoning and respect for the diversity and plurality of ideas, people and situations, tend to be more favorable to the integration and full participation of the students in the classroom, and therefore their effective implementation in this subject will be sought.

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			
Practical classes	12	0.48	1, 16, 2, 5, 6, 10, 8, 9, 12, 13, 17

Problems class	18	0.72	16, 11, 3, 4, 5, 6, 10, 8, 9, 15, 14
Theory classes	30	1.2	1, 16, 11, 3, 4, 5, 6, 10, 8, 9, 15, 14
Type: Autonomous			
Exams	15	0.6	1, 16, 11, 3, 4, 5, 6, 10, 8, 9, 15, 14
Problems resolution	25	1	1, 16, 11, 3, 4, 5, 6, 10, 8, 9, 15, 14
Workshop resolution	20	0.8	1, 16, 11, 3, 2, 4, 5, 6, 10, 8, 9, 7, 15, 14, 12, 13, 17

Assessment

During the lecture sessions (2 hours/week), we will introduce the basic concepts of the course as well as show a broad collection of different examples. We will solve theoretical problems and R lab tutorials during the lab sessions. Only some solutions (up to the professor) for both theoretical exercises and lab tutorials will be shared by Moodle. It is strongly recommended to regularly assist lectures and lab sessions to have a general idea of the course contents and the exercises to be solved.

To be eligible for the resit exam, you must get a minimum of 3/10 in practice and theory. The resit exam can only be taken by students who have not passed the course, and thus it cannot be taken to improve course qualifications.

Student's assessment may experience some modifications depending on the restrictions on face-to-face activities enforced by health authorities

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Final exam	0,50	10	0.4	1, 16, 11, 3, 4, 5, 6, 10, 8, 9, 15, 14
Grading exercises	0,20	8	0.32	1, 16, 11, 3, 4, 5, 6, 10, 8, 9, 7, 15, 14
Mid-term exam	0,30	12	0.48	1, 16, 11, 3, 2, 4, 5, 6, 10, 8, 9, 15, 14, 12, 13, 17

Bibliography

1. Berger, R.L., Casella, G.: Statistical Inference. Duxury Advanced Series. 2002.
2. Daalgard, P.: Introductory Statistics with R. Springer. 2008.
3. Daniel, W.W.: Biostatistics. Wiley. 1974.
4. DeGroot, M. H.: Schervish, M.J. Probability and Statistics. Pearson Academic. 2010.
5. Peña, D.: Estadística. Fundamentos de estadística. Alianza Universidad. 2001.
6. R Tutorial. An introduction to Statistics. <https://cran.r-project.org/manuals.html>. juny 2019.
7. Silvey, S.D.: Statistical Inference. Chapman&Hall. 1975.
8. Held, Sabanés and Bové (2013): Applied Statistical Inference: Likelihood and Bayes. Springer
9. Pawitan (2013): In all Likelihood: Statistical Modelling and Inference using Likelihood. Oxford University Press
10. Young, Smith (2005): Essentials of Statistical Inference. Cambridge University Press
11. Cox, D.R. and Hinkley, D.V. (1979). Theoretical Statistics. 1st Edition, Chapman and Hall/CRC

Software

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.