

Statistical Inference 1

Code: 104855
ECTS Credits: 6

Degree	Type	Year	Semester
2503852 Applied Statistics	FB	1	2

Contact

Name: Anna Lopez Ratera

Email: anna.lopez.ratera@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.

Teachers

Queralt Miró Catalina

Prerequisites

A good knowledge of the contents of the subjects studied during the first semester is considered very important, especially those of Introduction to Probability, Calculus 1 and Exploratory data analysis.

Objectives and Contextualisation

This subject is the first of the Degree dedicated to Statistical Inference, which is the part of the Statistics that allows to obtain, in a controlled way, information about a population based on the data of a "representative" sample. The subject has a central character within the studies, as different concepts and techniques that will be used in many of the subjects that will be studied from now on are introduced here. Specifically, an introduction to the Statistics will begin, and then the estimation of parameters, both punctual and by confidence intervals, will be treated, as well as classical parametric hypotheses testing, both for one and two normal and dichotomous populations, ending with the chi-square tests.

Competences

- Analyse data using statistical methods and techniques, working with data of different types.
- Correctly use a wide range of statistical software and programming languages, choosing the best one for each analysis, and adapting it to new necessities.
- Make efficient use of the literature and digital resources to obtain information.

- Select statistical models or techniques for application in studies and real-world problems, and know the tools for validating them.
- Select the sources and techniques for acquiring and managing data for statistical processing purposes.
- 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 communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
- Summarise and discover behaviour patterns in data exploration.
- Use quality criteria to critically assess the work done.

Learning Outcomes

1. Analyse data through different inference techniques using statistical software.
2. Analyse data through various inference techniques for one or more samples.
3. Critically assess the work done on the basis of quality criteria.
4. Describe the basic properties of point and interval estimators in classical and Bayesian statistics.
5. Determine the sample size and establish a sampling strategy for studies on parameter estimation, comparison of means, proportions, etc.
6. Identify statistical distributions.
7. Identify statistical inference as an instrument of prediction.
8. Interpret the results obtained and formulate conclusions regarding the experimental hypothesis.
9. Make effective use of references and electronic resources to obtain information.
10. Purge and store information on digital media.
11. 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.
12. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
13. Understand the concepts associated with hypothesis tests in classical and Bayesian statistics.
14. Use statistical software to obtain summary indices of the variables in the study.
15. Use the properties of the functions of distribution and density.
16. Validate and manage information for statistical processing.

Content

1. Statistical inference: introduction and basic concepts
 - 1.1. Introduction, objectives and program of the assignment
 - 1.2. Population and sample.
 - 1.3. Statistics.
 - 1.4. Distribution of the proportion, mean and sample variance: normal distribution and Central Limit Theorem.
2. Punctual forecasting
 - 2.1. The problem of punctual estimation. Parameter and estimator
 - 2.2. Properties of Estimators
 - 2.2. Estimator of a proportion
 - 2.3. Estimator of a mean and population variance.
 - 2.4. How to find a good estimator: the method of moments and the method of maximum likelihood.

3. Estimated by confidence intervals

3.1. Concept of Confidence Interval

3.2. Confidence interval for a proportion

3.3. Confidence intervals for the average (with known population variance or unknown population variance). Normal case and general case

3.4. Interval of confidence for variance. The Normal Case

3.5. Interval for the difference in averages (paired data or independent samples with known, unknown and equal population variances, or unknown and different). Normal case and general case

4. Proof of hypothesis for a population. Basic Concepts

4.1. The problem of a hypothesis test. Type of hypothesis. Type I and II errors

4.2. Level of significance and critical region. The P-value. The Power Function

4.3. Contrast to a proportion.

4.4. Contrast to the average population. The Z-test and the Student t-test

4.5. Determining sample size to ensure a given level of confidence and accuracy

4.6. Contrast to Variation

4.7. Relationship between the region of acceptance of a hypothesis test and the confidence interval

5. Contrasts of hypothesis and confidence intervals to compare two populations

5.1. Comparison of the proportions of two independent populations

5.2. Comparison of the averages of two populations from paired data

5.3. Comparison of the averages of two independent populations

5.4. Comparison of the proportions of two independent populations

5.5. Comparison of the variances of two independent normal populations. The F-Test

6. Non-parametric tests based on the law of khi squared

6.1. Pearson's square khi test for sample adjustment to a distribution

6.3. The khi squared test of independence for categorical data

6.4. The test of chi square homogeneity for categorical data

7. Comparison of three or more averages

7.1. ANOVA contrast of a factor

7.2. Multiple comparison of averages (Tuckey test)

8. The Simple Linear Regression Model

8.1. Linear relationship between two numerical variables: Pearson's correlation coefficient

8.2. The Linear Regression Model

8.3. Estimation of parameters (minimum squares) and hypotheses contrasts

8.3. Good fit with the model. Coefficient of determination and range of values.

8.4. Punctual and Interval Forecasting

8.5. The logarithm function to improve the linear relationship

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 subject is structured from theory classes, problems and practices.

In theory classes we will introduce the concepts and techniques described in the course program. Considering that the content is essentially the standard of a first course of statistical inference, one can follow the course making use of the recommended basic bibliography. The material corresponding to each topic explained in the classroom will also be posted on the Virtual Campus.

The classes of problems are intended to work and understand statistical concepts. In the Virtual Campus the lists of problems will be posted and, when they have already been solved in class, also the solutions.

The objective of the practices is the use of statistical software R, to obtain and clarify the results of the procedures that have been introduced in theory classes and problems. In the Virtual Campus the statement of each practice will be posted in advance.

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, 2, 3, 13, 10, 5, 7, 8, 12, 11, 9, 14
Problems classes	18	0.72	2, 3, 13, 10, 4, 5, 6, 7, 8, 12, 11, 9, 15
Theory classes	30	1.2	2, 3, 13, 4, 5, 6, 7, 8, 11, 9, 15

Type: Autonomous			
Exams	15	0.6	2, 3, 13, 4, 5, 6, 7, 8, 11, 9, 15
Problems resolution	25	1	2, 3, 13, 10, 4, 5, 6, 7, 8, 12, 11, 9, 15
Workshop resolution	20	0.8	1, 2, 3, 13, 10, 5, 7, 8, 12, 11, 9, 14

Assessment

The continuous evaluation note will be obtained from a control of the problems that will give a note C, and from a control of the practices of the subject that will give a note P. Note C has a weight of 20% and note P a weight of 30%. The final exam grade E1 is worth 50% of the final grade. With the notes C, P and E1 you get the grade of the subject, G, as follows:

$$G = 0.50 \times E1 + 0.20 \times C + 0.30 \times P$$

Important: The evaluation can be:

- or Continuous Evaluation where problem delivery is resolved on three different days during the course, the practice exam on a date close to the end of the course and the final exam on another date different from the date of the practice exam,
- or Single Evaluation where on the same day of the final exam of the Continuing Evaluation the student must submit to all three tests at the same time (E1, C and P). Important, to be able to submit to the Single Evaluation you must use the channel and submit the application on the deadlines established by the Faculty of Sciences.

Given the quantitative nature of the subject, the Continuous Evaluation is recommended by the teachers of the subject and by the studies.

Recovery and/or improvement of the exam grade:

The student passes the subject if N is greater than or equal to 5 and, at the same time, E1 is greater than 4. Otherwise or if the student wants to improve grade, there is a possibility to improve the part of the E1 exam through a recovery exam, whose grade will be E2. Thus, from this recovery note you get the final grade of the subject:

$$FG = 0.50 \times \max (E1, E2) + 0.20 \times C + 0.30 \times P$$

Observation 1: C and P continuous assessment grades are not recoverable.

Observation 2: It is considered that the student has submitted to the announcement of the subject if any of the two exams that give rise to the E1 or E2 notes are presented. In case otherwise, it will be a Non Presented, even if it has a continuous evaluation grade (C and / or P).

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Final exam / Reassessment (E)	0,50	10	0.4	2, 3, 13, 4, 5, 6, 7, 8, 12, 11, 15
Practical exam (P)	0,30	12	0.48	1, 2, 3, 13, 10, 4, 5, 6, 7, 8, 12, 11, 9, 15, 14, 16
Problems delivery (C)	0,20	8	0.32	2, 3, 13, 4, 5, 6, 7, 8, 12, 11, 9, 15, 16

Bibliography

Berger, R.L., Casella, G.: Statistical Inference. Duxury Advanced Series. 2002.

Daalgard, P.: Introductory Statistics with R. Springer. 2008.

Daniel, W.W.: Biostatistics. Wiley. 1974.

DeGroot, M. H.: Schervish, M.J. Probability and Statistics. Pearson Academic. 2010.

Peña, D.: Estadística. Fundamentos de estadística. Alianza Universidad. 2001.

R Tutorial. An introduction to Statistics. <https://cran.r-project.org/manuals.html>. juny 2019.

Silvey, S.D.: Statistical Inference. Chapman&Hall. 1975.

Novales, A.: Econometría Graw-Hill 2000.

Software

The software to be used to work with the data will be Excel and the statistical program R.