Universitat Autònoma de Barcelona

2019/2020

Statistical Inference 1

Code: 104855
ECTS Credits: 6

Degree	Type	Year	Semester
2503852 Applied Statistics	FB	1	2

Contact

Name: David Moriña Soler
Email: David.Morina@uab.cat

Use of Languages

Principal working language: catalan (cat)
Some groups entirely in English: No
Some groups entirely in Catalan: Yes
Some groups entirely in Spanish: No

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 Exloratory 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

Preliminaries of Probability (reminder): Probability and random variables. Law concept. Discrete 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. Estimation by confidence intervals.

1. Concept of confidence 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 deviation.

    3.2. For the mean of a Normal population with unknown deviation.

    3.3. For the variance of a Normal population with unknown mean.

    3.4. For the variance of a Normal population with known mean.

    3.5. Asymptotic confidence intervals.

4. Confidence intervals using the inequality of Txevixev.

5. Confidence intervals for the parameters of two populations.

5.1. Confidence intervals with independent samples.

5.2. Confidence intervals for the difference of means of two Normal populations with paired data

Topic 3: Point estimation.

1. Point estimators: definition and "good" properties.

1.1. Bias

1.2. Comparison of estimators without bias. Relative efficiency

1.3. The Cramér-Rao bound.

1.4. Comparison of estimators with bias: the mean square error.

1.5. Consistency of an estimator.

Topic 4: Hypothesis tests.

1. Introduction.

2. Tests for the parameters of a population.

2.1. For the mean of a Normal population with known deviation.

2.2. For the mean of a Normal population with unknown deviation.

2.3. Asymptotic tests for the mean of a population when the sample is large and for the proportion.

2.4. Tests 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.

4. The chi-square tests.

4.1. Chi-square test of goodness of fit.

4.2 Chi-square test of independence.

Topic 5: Simple linear regression.

1. Objectives and hypotheses of the model.

2. Estimation by the ordinary least squares (OLS) method.

2.1 The OLS estimators.

2.2 Estimation of the variance of the errors.

2.3 Properties of the OLS estimators.

3. Inference based on the simple linear regression model.

3.1 Confidence intervals for model parameters.

3.2 Hypothesis testing for the parameters of the model.

4. Goodness of fit.

5. Forecasting based on the simple linear regression model

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.

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:

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

Recovery and / or improvement of the exam note:

The student passes the subject if G is greater than or equal to 5. Otherwise, or if the student wants to improve note, there is a possibility to improve the part of the E1 exam grade by a recovery exam, the 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.