

Course guide

200605 - FIE - Foundations of Statistical Inference

Last modified: 01/06/2023

Unit in charge: School of Mathematics and Statistics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research.
1004 - UB - (ENG)Universitat de Barcelona.

Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Optional subject).

Academic year: 2023 **ECTS Credits:** 5.0 **Languages:** Spanish

LECTURER

Coordinating lecturer: ANTONIO MIÑARRO ALONSO

Others: Primer quadrimestre:
ANTONIO MIÑARRO ALONSO - A
LOURDES RODERO DE LAMO - A

PRIOR SKILLS

Advanced Statistical Inference and Foundations of Statistical Inference are compulsory subjects of the MESIO UPC-UB. The former is mandatory for all graduate students in statistics or mathematics (path 1) and the latter is compulsory for all students from other degrees (path 2). Students from path 2 can choose Advanced Statistical Inference as optional. Students from path 1 can not choose Foundations of Statistical Inference.

The course assumes a basic knowledge of the concepts of probability theory. The student should know and work with major discrete and continuous probability models: Poisson, Binomial, Exponential, Uniform, Normal. In particular the student should be able to use the cumulative distribution functions and density functions or probability mass, for calculating probabilities and population parameters of the main distributions. It is also assumed the skill to work with the expectation and variance of random variables. Finally, it is important to know and understand the implications of the central limit theorem.

You can consult the following material:

Probabilidad y estadística de Evans, Michael J. (2005)
Michael J. Evans (Autor) y Jeffrey Rosenthal
Edit. Reverte
http://www.reverte.com/motor?id_pagina=catalogo/ficha&idcategoria=6&idsubcategoria=47&idlibro=664

Morris H. DeGroot and Mark J. Schervish
Probability and Statistics (4th Edition)
Addison-Wesley (2010)
ISBN 0-321-50046-6
http://www.pearsonhighered.com/pearsonhigheredus/educator/product/products_detail.page?isbn=0201524880

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:

3. CE-4. Ability to use different inference procedures to answer questions, identifying the properties of different estimation methods and their advantages and disadvantages, tailored to a specific situation and a specific context.
4. CE-6. Ability to use appropriate software to perform the necessary calculations in solving a problem.

Transversal:

1. **TEAMWORK:** Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.
2. **FOREIGN LANGUAGE:** Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

TEACHING METHODOLOGY

· Theory sessions

The teacher explains the contents of the course with the help of computer presentations. Student participation will be encouraged through some questions and examples.

· Problem sessions

By the end of each issue a session specially devoted to problems will take place. The list of problems will be available in advance on the intranet. Students should come to class with doubts related to the proposed problems in order to be solved by the teacher.

· Statistical laboratory

Several statistical analyses will be carry out with the help of some scripts of R. Students will be proposed to solve several more extensive exercises with the help of the software.

LEARNING OBJECTIVES OF THE SUBJECT

Students should achieve a good knowledge of the common language of statistical inference with both a theoretical and a practical basis. Students not only should to be able to use most of the statistical techniques but also they have to be able to learn new methodologies. Students should be able to use software R as a tool for the inferential process.

As specific goals we have the following:

- Students should know the main sample techniques and the main sample distributions based on normal law and its use in statistical inference.
- Students should be able to apply some of the usual methods of estimation. Students should know the desirable properties of an estimator and verify if they are achieved by a given statistic.
- Students should understand the concept of confidence of an interval. They have to be able to construct the most usual intervals and compute the necessary sample size to achieve a given confidence and precision.
- Students should understand the methodology underlying the testing of hypotheses including the types of errors and the importance of sample size to make decisions with a good statistical basis.
- Students should be able to obtain estimates from a linear regression model and verify the validity of the assumptions of the model in order to discuss the results of a regression study.
- Students should understand the linear model of analysis of variance together with the sum of squares variance decomposition and solve the one-way model and the two-way model both with fix and random factors.

STUDY LOAD

Type	Hours	Percentage
Hours large group	30,0	24.00
Hours small group	15,0	12.00
Self study	80,0	64.00

Total learning time: 125 h



CONTENTS

1. Introduction to inference

Description:

1.1 Basic ideas of Statistical Inference.

Specific objectives:

Basic introduction to the main concepts of statistical inference and review of the necessary ideas of the Theory of Probability

Related activities:

Theory sessions.

Full-or-part-time: 0h 30m

Theory classes: 0h 30m

2. Sampling

Description:

2.1. Definition

2.2. Sampling methods

2.3. Random sampling

2.4. Sampling distributions

2.4.1. Exact and asymptotic sampling distributions

2.4.2. The distribution in sampling from a Normal Population

2.4.3. Distributions arising from Normal sampling

2.5. Simulating random samples

Specific objectives:

Students should know the main sample techniques and the main sample distributions based on normal law and its use in statistical inference.

Related activities:

Theory sessions. Problem sessions.

Full-or-part-time: 2h 30m

Theory classes: 2h 30m

3. Parameter estimation

Description:

3.1. Introduction, concept of estimator, point and confidence estimation.

3.2. Properties of point estimates: consistency, bias, efficiency, minimal variance, sufficiency, mean square error.

3.3. Methods to obtain estimates: moments, maximum likelihood, least squares, Bayes

3.4. Resampling methods: Bootstrap, Jackknife

Specific objectives:

Students should be able to apply some of the usual methods of estimation. Students should know the desirable properties of an estimator and verify if they are achieved by a given statistic.

Related activities:

Theory sessions. Problem sessions

Full-or-part-time: 6h

Theory classes: 6h

4. Confidence Intervals

Description:

- 4.1. Definition
- 4.2. Construction of intervals
- 4.3. Confidence level and sample size
- 4.4. Some confidence intervals
- 4.5. Asymptotic confidence intervals

Specific objectives:

Students should understand the concept of confidence of an interval. They have to be able to construct the most usual intervals and compute the necessary sample size to achieve a given confidence and precision.

Related activities:

Theory sessions. Problem sessions. Statistical laboratory.

Full-or-part-time: 4h 30m

Theory classes: 4h 30m

5. Hypotheses testing

Description:

- 5.1. Fundamental notions of hypotheses testing
 - 5.1.1. From language to parametrical hypotheses
 - 5.1.2. Null and alternative hypotheses
 - 5.1.3. Decision rule: Critical region
- 5.2. Errors in hypotheses testing
 - 5.2.1. Type I error: level of significance
 - 5.2.2. Type II error: power of the test
 - 5.2.3. Sample size
- 5.3. P-values
- 5.4. Some hypotheses tests
 - 5.4.1. Likelihood ratio tests
 - 5.4.2. Tests for normal populations
 - 5.4.3. Tests on proportions
 - 5.4.4. Chi-squared tests
 - 5.4.5. Robust tests: tests based on ranks and permutation tests
- 5.5. Relation between confidence estimation and hypotheses testing
- 5.6. Multiple testing
- 5.7. Combining results from different tests
- 5.8. Bayesian hypothesis testing

Specific objectives:

Students should understand the methodology underlying the testing of hypotheses including the types of errors and the importance of sample size to make decisions with a good statistical basis.

Related activities:

Theory sessions. Problem sessions. Statistical laboratory.

Full-or-part-time: 12h

Theory classes: 12h

6. The general linear model

Description:

- 6.1. Introduction
- 6.2. Parameter estimation and hypotheses testing
- 6.3. Simple linear regression
 - 6.3.1. Parameter estimation
 - 6.3.2. Regression diagnostic
 - 6.3.3. Hypotheses in regression
 - 6.3.4. Model comparisons
 - 6.3.5. Relationship between regression and correlation
 - 6.3.6. Smoothing
- 6.4. Multiple regression
 - 6.4.1. Parameter estimation
 - 6.4.2. Regression diagnostic
 - 6.4.3. Inference in multiple regression
 - 6.4.4. Collinearity

Specific objectives:

Students should be able to obtain estimates from a linear regression model and verify the validity of the assumptions of the model in order to discuss the results of a regression study.

Related activities:

Theory sessions. Problem sessions.

Full-or-part-time: 9h

Theory classes: 9h

7. ANOVA models

Description:

- 7.1. One-way ANOVA
 - 7.1.1. Linear model for one-way ANOVA
 - 7.1.2. Null hypotheses
 - 7.1.3. Factor effects
 - 7.1.4. ANOVA diagnostics
 - 7.1.5. Multiple comparison of means
- 7.2. Two-way ANOVA
 - 7.2.1. Randomized blocks design
 - 7.2.2. Two fixed factors ANOVA
 - 7.2.3. Interpreting interactions
 - 7.2.4. Two random factors ANOVA
 - 7.2.5. Mixed effects model

Specific objectives:

Students should understand the linear model of analysis of variance together with the sum of squares variance decomposition and solve the one-way model and the two-way model both with fix and random factors.

Related activities:

Theory sessions. Problem sessions. Statistical laboratory.

Full-or-part-time: 10h 30m

Theory classes: 10h 30m

GRADING SYSTEM

Throughout the course students will be proposed to solve 3 small quizzes (CUEST). They will also be proposed to solve take-home exercises and deliver it within a specified period as discussed in the section on practical laboratory in teaching methodology (EJER). A final exam (EF) will take place on the date specified by the master direction. The grade of the course will be obtained as $N = 0.2 * CUEST + 0.20 * EJER + 0.6 * EF$.

BIBLIOGRAPHY

Basic:

- Rohatgi, Vijay K. Statistical inference [on line]. New York: John Wiley & Sons, 1984 [Consultation: 07/07/2023]. Available on: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=1894681>.
- Sánchez, P.; Baraza, X.; Reverter, F.; Vegas, E. Métodos estadísticos aplicados. Barcelona: UB, 2006. ISBN 8447530981.
- Peña, Daniel. Estadística : modelos y métodos. 2 vols. 2ª ed. rev. Madrid: Alianza Universidad Textos, 1986-1991. ISBN 8420689931.
- DeGroot, Morris; Schervish, Mark. Probability and statistics. 4th ed. Pearson, 2012. ISBN 0321500466.
- Evans, Michael; Rosenthal, Jeffrey S. Probability and statistics : the science of uncertainty. 2nd ed. New York: W.H. Freeman and Company, cop. 2010. ISBN 1429224622.
- Casella, G.; Berger, Roger L. Statistical inference. 2nd ed. Duxbury: Pacific Grove, 2002. ISBN 0534243126.
- Sahu, Pradip Kumar, Pal, Santi Ranjan, Das, Ajit Kumar. Estimation and Inferential Statistics [on line]. 1. New Delhi: Springer, 2015 [Consultation: 19/05/2021]. Available on: <https://link.springer.com/book/10.1007%2F978-81-322-2514-0>. ISBN 9788132225133.