Course guide
200132 - EST - Statistics

Last modified: 11/04/2024

Unit in charge: School of Mathematics and Statistics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research.

Degree: BACHELOR'S DEGREE IN MATHEMATICS (Syllabus 2009). (Compulsory subject).

Academic year: 2024 **ECTS Credits:** 7.5 **Languages:** Catalan, Spanish

LECTURER

Coordinating lecturer: PEDRO FRANCISCO DELICADO USEROS

Others: Segon quadrimestre:
PEDRO FRANCISCO DELICADO USEROS - M-A, M-B
JOSEP GINEBRA MOLINS - M-A, M-B
JOSE ANTONIO SÁNCHEZ ESPIGARES - M-A, M-B

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
1. CE-2. Solve problems in Mathematics, through basic calculation skills, taking in account tools availability and the constraints of time and resources.
2. CE-3. Have the knowledge of specific programming languages and software.
3. CE-4. Have the ability to use computational tools as an aid to mathematical processes.

Generical:
5. CB-1. Demonstrate knowledge and understanding in Mathematics that is founded upon and extends that typically associated with Bachelor's level, and that provides a basis for originality in developing and applying ideas, often within a research context.
6. CB-2. Know how to apply their mathematical knowledge and understanding, and problem solving abilities in new or unfamiliar environments within broader or multidisciplinary contexts related to Mathematics.
7. CB-3. Have the ability to integrate knowledge and handle complexity, and formulate judgements with incomplete or limited information, but that include reflecting on social and ethical responsibilities linked to the application of their knowledge and judgements.
8. CG-1. Show knowledge and proficiency in the use of mathematical language.
9. CG-2. Construct rigorous proofs of some classical theorems in a variety of fields of Mathematics.
10. CG-3. Have the ability to define new mathematical objects in terms of others already know and ability to use these objects in different contexts.
11. CG-4. Translate into mathematical terms problems stated in non-mathematical language, and take advantage of this translation to solve them.
12. CG-6 Detect deficiencies in their own knowledge and pass them through critical reflection and choice of the best action to extend this knowledge.

Transversal:
4. SELF-DIRECTED LEARNING. Detecting gaps in one's knowledge and overcoming them through critical self-appraisal. Choosing the best path for broadening one's knowledge.

TEACHING METHODOLOGY

There are 5 class hours per week: 3 hours corresponding to theoretical lessons and 2 hours of problems or laboratory practicals.

Theoretical lessons:

The theoretical lessons are basically master classes given by the theory professor. Theorem proofs are developed on the blackboard, and important concepts are summarized by means of slides. Detailed examples are introduced, emphasizing on the application of Statistics in real life problems. Virtual campus Atenea will be used to circulate the class material.

Problems lessons:

The problems professor previously introduces the problems that the students have to solve. In class, the professor exposes and explains the solution of selected problems. The Atenea virtual campus is used to pose self-correcting questionnaires to students, which they must answer within a deadline. These quizzes score.

Laboratory practicals:

The practical sessions will be taught with the statistical software R. They will consist of some introductory session plus the last month of class, where the satistical modelization will be practiced.

LEARNING OBJECTIVES OF THE SUBJECT

A student who has completed this Statistics course:

1. Is able to perform and interpret basic descriptive statistics with statistical software.
2. Is able to perform statistical inference with statistical software and correctly interpret the results obtained.
3. You can formulate the difference between the two statistical schools, frequentist and Bayesian.
4. Is capable of analytically obtaining moment estimators and maximum likelihood estimators for parameters of the most usual distributions.
5. Is able to compare different estimators and choose the optimal estimator according to some optimality criterion (bias, mean square error).
6. Is able to construct confidence intervals based on pivotal quantities (exact or asymptotic).
7. Is able to design an optimal test for certain hypotheses testing problems about distribution parameters, applying the Neyman-Pearson criterion and the general likelihood ratio test.
8. Is able to formulate the difference between parametric and non-parametric tests.
9. Is able to apply the classic parametric tests (normal Z test, Student's t-test with independent samples and paired data, F for equality of variances) to data sets and correctly interpret the results.
10. Is able to apply simple non-parametric tests based on the multinomial distribution to data sets and correctly interpret the results.
11. Is able to fit a multiple linear regression model with R and correctly interpret the results.
12. Is able to fit a logistic regression model with R and correctly interpret the results.

STUDY LOAD

Type	Hours	Percentage
Self study	112,5	60.00
Hours large group	45,0	24.00
Hours small group	30,0	16.00

Total learning time: 187.5 h

CONTENTS

1. Introduction to Statistics

Description:
1.1. Population and sample. Descriptive statistics.
1.2. Parameters, statistics and estimators
1.3. Sampling distribution.
1.3.1 The empirical distribution function
1.3.2 Simulation
1.3.3 Bootstrap
1.4. Statistical models
1.4.1. Normal model. Sampling Distributions under normality
1.4.2. Binomial model
1.4.3. Location and scale models
1.4.4. Exponential families
1.5. Objectives of inference: estimation, hypothesis testing and prediction

Specific objectives:
Carry out univariate and bivariate descriptive statistical analysis.

Related activities:
Theoretical classes and sessions in a computer room.

Full-or-part-time: 12h 30m
Theory classes: 7h 30m
Laboratory classes: 5h

2. Defining point estimators

Description:
2.1. Method of moments
2.1.1. Plug-in method
2.1.2. Method of moments
2.2. Maximum likelihood estimation
2.2.1. Likellihood function
2.2.2. Maximum likellihood estimator
2.2.3. Relation with Kullback-Leibler Divergence
2.2.4. Numerical computation of the maximum likellihood estimators
2.2.5. Invariance principle of the maximum likellihood estimator
2.3. Estimation in the normal and in the binomial models

Specific objectives:
Defining estimators using different methods.

Related activities:
Theory classes and problem sessions.

Full-or-part-time: 15h
Theory classes: 9h
Practical classes: 6h

3. Evaluating estimators

Description:
3.1. Systematic error (bias) and precision of an estimator
3.2. Optimal unbiased estimators (UMVUE)
3.2.1. Fisher's Information. Cramér-Rao lower bound
3.2.2. sufficiency, completeness
3.2.3. Rao-Blackwell and Lehmann-Scheffé theorems
3.3. Asymptotic behavior
3.3.1. Consistency
3.3.2. Asymptotic normality
3.3.3. Delta method
3.3.4. Asymptotic theory for the maximum likelihood estimator

Specific objectives:
Derive properties of estimators.

Related activities:
Theory classes and problem sessions.

Full-or-part-time: 10h
Theory classes: 6h
Practical classes: 4h

4. Interval estimation

Description:
4.1. Confidence intervals
4.2. Methods for constructing confidence intervals
4.2.1. Pivotal quantities
4.2.2. Asymptotic Confidence Intervals
4.3. Estimation by confidence intervals in the normal and binomial models

Specific objectives:
Construction of confidence intervals.

Related activities:
Theory classes, problem sessions and laboratory sessions.

Full-or-part-time: 10h
Theory classes: 6h
Practical classes: 4h

4. Hypothesis testing

Description:
5.1. Basic definitions. Simple hypothesis tests
5.1.1. Types of errors
5.1.2. Neyman–Pearson Lemma
5.1.3. Conclusions of a test: the p-value
5.2. Uniformly most powerful tests
5.2.1. Neyman-Pearson Lemma for composite alternatives
5.3. Likelihood ratio test
5.4.1. Relation to the Neyman-Pearson Lemma
5.4.2. Properties of Likelihood Ratio Tests
5.4.3. Tests related to maximum likelihood: Scores and Wald
5.4. Hypothesis tests on normal and binomial models
5.5. Tests based on the multinomial distribution

Specific objectives:
Development of hypothesis tests.

Related activities:
Theory classes and problem sessions.

Full-or-part-time: 12h 30m
Theory classes: 7h 30m
Practical classes: 5h

6. Linear regression model. Logistic regression

Description:
6.1. Linear Regression
6.1.1. Least squares (OLS) and maximum likelihood estimation
6.1.2. Properties of the OLS estimator
6.1.3. Model validation. Residue analysis
6.1.4. Hypothesis tests on the parameters
6.1.5. Prediction
6.1.6. Models with categorical explanatory variables
6.3. Logistic regression

Specific objectives:
Apply linear regression and interpret the results.

Related activities:
Laboratory practicals.

Full-or-part-time: 15h
Theory classes: 9h
Laboratory classes: 6h

GRADING SYSTEM

The assessment comprises the following elements: final exam, midterm exam, deliverable questionnaires (approximately once a week). The final exam and the midterm exam consist of open theoretical questions and problems to solve. The final mark (FM) is calculated as:

FM = 0.15 * Max(Deliverables, Final) + 0.25 * Max(Midterm, Final) + 0.6 * Final

Therefore, the mark of the partial will be taken into account (with a weight of 25% of the overall) only if it is higher than the mark of the final exam. The same applies to the note of weekly deliveries (with a weight of 15% of the overall).

An extra exam will take place on July for students that failed during the regular semester, with a structure similar to that of the final exam. The mark of the extra call will be that of the extra exam.

BIBLIOGRAPHY

Basic:

- Casella, George; Berger, Roger L. Statistical inference. 2nd ed. Pacific Grove: Duxbury, cop. 2002. ISBN 0534243126.
- Evans, Michael; Rosenthal, Jeffrey S. Probability and statistics : the science of uncertainty [on line]. 2nd ed. New York: W.H. Freeman and Company, cop. 2010 [Consultation: 26/06/2023]. Available on: http://www.utstat.toronto.edu/mikevans/jeffrosenthal/. ISBN 9781429224628.
- DeGroot, Morris H.; Schervish, Mark J. Probability and statistics. 4th ed. Boston: Pearson, 2012. ISBN 9780321709707.
- Wasserman, Larry. All of statistics : a concise course in statistical inference. Pittsburgh: Springer, cop. 2010. ISBN 9781441923226.

Complementary:

- Bickel, Peter J.; Doksum, Kjell A. Mathematical statistics: basic ideas and selected topics, volume I. 2nd ed. San Francisco: Holden-Day, 2015. ISBN 0816207844.
- Dalgaard, Peter. Introductory statistics with R [on line]. 2nd ed. New York: Springer, 2008 [Consultation: 26/06/2023]. Available on: https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-79054-1. ISBN 9780387790534.
- Efron, Bradley; Hastie, Trevor. Computer age statistical inference: algorithms, evidence, and data science [on line]. First published. New York: Cambridge University Press, 2016 [Consultation: 26/06/2023]. Available on: https://hastie.su.domains/CASI/. ISBN 9781107149892.
- Fan, Jianqing; Li, Runze; Zhang, Cun-Hui; Zou, Hui. Statistical foundations of data science. Chapman and Hall/CRC, 2020. ISBN 9781466510845.
- Peck, Roxy. Statistics: a guide to the unknown. 4th ed. Belmont: Thomson Brooks/Cole, 2006. ISBN 0534372821.

RESOURCES

Hyperlink:

- R-software: www.r-project.org. Resource