Course guides
200132 - EST - Statistics

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: 2020 ECTS Credits: 7.5 Languages: Catalan

LECTURER
Coordinating lecturer: JOSEP ANTON SÁNCHEZ ESPIGARES
Others: Segon quadrimestre:
MARTA PÉREZ CASANY - M-B
JOSEP ANTON SÁNCHEZ ESPIGARES - M-A, M-B
JORDI VALERO BAYA - M-A

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.
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 transparencies. 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 exercises that the students have to solve. In class, the professor (or any of the students) exposes and explains the solution. The students must hand in exercises which will count towards the subject final mark. The problems professor is responsible for correcting these exercises, some of which may be solved in class.

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 statistical modelization will be practiced.

LEARNING OBJECTIVES OF THE SUBJECT

A student that has completed this statistics course:

1. Is able to carry out 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. Can express the difference between the two statistical schools: the frequentist and the bayesian approach.
4. Is able to analytically obtain moment estimators, maximum likelihood estimators and bayesian estimators for the parameters of the most common probability distributions.
5. Is able to compare different estimators and select the best estimator according to some optimality criterion (bias, variance, mean squared error).
6. Is able to design an optimal test for particular hypothesis tests regarding parameters of distributions, applying the criterion of Neyman-Pearson and the generalized likelihood ratio.
7. Is able to formulate the difference between parametric and non-parametric tests.
8. Is able to apply the classical parametric tests (Z-test for the normal distribution, Student t test for independent samples and paired observations, F test for equality of variances) to data sets and interpret correctly the results obtained.
9. Is able to apply the most common non-parametric tests (Chi-square test for independence, sign test) to data sets and correctly interpret the results.
10. Is able to read and understand the descriptive statistics and the statistical inference used in a published scientific article.

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours small group</td>
<td>30,0</td>
<td>16.00</td>
</tr>
<tr>
<td>Self study</td>
<td>112,5</td>
<td>60.00</td>
</tr>
<tr>
<td>Hours large group</td>
<td>45,0</td>
<td>24.00</td>
</tr>
</tbody>
</table>

Total learning time: 187.5 h
CONTENTS

1. INTRODUCTION

Description:
1.1. Descriptive statistics.
1.2. Population and sample.
1.3. Distributions related to the Normal distribution.

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

Related activities:
Some classes and three sessions in a computer room.

Full-or-part-time: 18h
Theory classes: 4h
Laboratory classes: 3h
Self study: 11h

2. POINT ESTIMATION

Description:
2.1. The method of moments.
2.2. The maximum likelihood method.
2.3. Bayesian estimation.

Specific objectives:
Construct estimators using various methods.

Related activities:
Theory classes and problem sessions.

Full-or-part-time: 30h
Theory classes: 6h
Practical classes: 4h
Self study: 20h

3. EVALUATION OF ESTIMATORS

Description:
3.1. Properties of estimators: bias, variance, mean squared error (MSE), sufficiency, consistency, efficiency.
3.3. Asymptotic properties of the maximum likelihood estimator.

Specific objectives:
Derive properties of estimators.

Related activities:
Theory classes and problem sessions.

Full-or-part-time: 26h
Theory classes: 6h
Practical classes: 4h
Self study: 16h
4. HYPOTHESIS TESTING

Description:
4.2. Neyman Pearson criterium for simple hypothesis.
4.3. Extension of Neyman Pearson for unilateral alternatives.
4.4. Test of the monotone likelihood function.
4.5 Generalized likelihood ration tests.
4.6 X^2 test for contingency tables and goodness of fit.

Specific objectives:
Development of hypothesis tests.

Related activities:
Theory classes and problem sessions.

Full-or-part-time: 37h
Theory classes: 7h
Practical classes: 5h
Self study: 25h

5. INTERVAL ESTIMATION

Description:
5.1. Confidence intervals.
5.2. Intervals associated to pivotal quantities.
5.3. Intervals associated to the Normal model.

Specific objectives:
Construction of confidence intervals.

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

Full-or-part-time: 8h 10m
Theory classes: 2h
Practical classes: 2h
Self study: 4h 10m
### 6. LINEAR MODEL

**Description:**
- 6.1. Simple and multiple linear regression.
- 6.2. Assumptions of the linear model. Least squares and maximum likelihood estimation.
- 6.3 Multicolinearity. Leverage. Influential observations.
- 6.4 Goodness of fit and coefficient of determination.
- 6.5 Prediction.
- 6.6 Residual analysis.
- 6.7 Anova
- 6.8 Ancova

**Specific objectives:**
Apply linear regression and interpret the results.

**Related activities:**
Laboratory practicals.

**Full-or-part-time:** 60h
- Theory classes: 13h
- Laboratory classes: 7h
- Self study: 40h

### GRADING SYSTEM

The assessment comprises the following elements: final exam, midterm exam, deliverable exercises. The final exam and the midterm exam consist of open theoretical questions and problems to solve. The deliverable exercises have to be done individually and they will be known approximately one week before. The continuous assessment mark (CAM) is calculated as:

\[
\text{CAM} = 0.5 \times \text{MFinal} + 0.25 \times (\text{MFinal, MMidterm}) + 0.25 \times \text{MDeliverables}
\]

The final subject mark (FM) is the maximum between the CAM mark and the final exam mark: \( \text{FM} = \max(\text{CAM}, \text{MFinal}) \)

An extra exam will take place on July for students that failed during the regular semester. For the students that perform the extra exam, the final mark will be equal to the maximum between the final subject mark (FM) and the one obtained in this exam.

### BIBLIOGRAPHY

**Basic:**

**Complementary:**

### RESOURCES

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