200605 - FIE - Foundations of Statistical Inference

Coordinating unit: 200 - FME - School of Mathematics and Statistics
Teaching unit: 715 - E1O - Department of Statistics and Operations Research
1004 - UB - (ENG)Universitat de Barcelona

Academic year: 2018
Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Teaching unit Optional)
ECTS credits: 5  
Teaching languages: Spanish

Teaching staff

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

Prior skills

The MESIO UPC-UB includes two compulsory subjects: Advanced Statistical Inference and Foundations of Statistical Inference. Advanced Statistical Inference is mandatory for all graduate students in statistics or mathematics (path 1) and Foundations of Statistical Inference 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:

Statmedia free version: http://www.ub.edu/stat/GrupsInnovacio/Statmedia/demo/Statmedia.htm

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

Morris H. DeGroot and Mark J. Schervish
Addison-Wesley (2010)

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
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.

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

<table>
<thead>
<tr>
<th>Total learning time: 125h</th>
<th>Hours large group:</th>
<th>30h</th>
<th>24.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours medium group:</td>
<td>0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Hours small group:</td>
<td>15h</td>
<td>12.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities:</td>
<td>0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Self study:</td>
<td>80h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
# Content

## 1. Introduction to inference

<table>
<thead>
<tr>
<th><strong>Description:</strong></th>
<th>Basic ideas of Statistical Inference.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Related activities:</strong></td>
<td>Theory sessions.</td>
</tr>
<tr>
<td><strong>Specific objectives:</strong></td>
<td>Basic introduction to the main concepts of statistical inference and review of the necessary ideas of the Theory of Probability</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Learning time:</strong></th>
<th>0h 30m</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theory classes:</strong></td>
<td>0h 30m</td>
</tr>
</tbody>
</table>

## 2. Sampling

<table>
<thead>
<tr>
<th><strong>Description:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Definition</td>
</tr>
<tr>
<td>2.2. Sampling methods</td>
</tr>
<tr>
<td>2.3. Random sampling</td>
</tr>
<tr>
<td>2.4. Sampling distributions</td>
</tr>
<tr>
<td>2.4.1. Exact and asymptotic sampling distributions</td>
</tr>
<tr>
<td>2.4.2. The distribution in sampling from a Normal Population</td>
</tr>
<tr>
<td>2.4.3. Distributions arising from Normal sampling</td>
</tr>
<tr>
<td>2.5. Simulating random samples</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Related activities:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory sessions. Problem sessions.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Specific objectives:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Students should know the main sample techniques and the main sample distributions based on normal law and its use in statistical inference.</td>
</tr>
</tbody>
</table>
### 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

**Related activities:**
Theory sessions. Problem sessions

**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.

#### Learning 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

**Related activities:**
Theory sessions. Problem sessions. Statistical laboratory.

**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.

#### Learning 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

**Related activities:**
- Theory sessions. Problem sessions. Statistical laboratory.

**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.
### 6. The general linear model

<table>
<thead>
<tr>
<th>Description:</th>
<th>Learning time: 9h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>6.1. Introduction</strong></td>
<td>Theory classes: 9h</td>
</tr>
<tr>
<td><strong>6.2. Parameter estimation and hypotheses testing</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.3. Simple linear regression</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.3.1. Parameter estimation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.3.2. Regression diagnostic</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.3.3. Hypotheses in regression</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.3.4. Model comparisons</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.3.5. Relationship between regression and correlation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.3.6. Smoothing</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.4. Multiple regression</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.4.1. Parameter estimation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.4.2. Regression diagnostic</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.4.3. Inference in multiple regression</strong></td>
<td></td>
</tr>
<tr>
<td><strong>6.4.4. Collinearity</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Related activities:**
- Theory sessions.
- Problem sessions.

**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.
7. **ANOVA models**

<table>
<thead>
<tr>
<th>Learning time:</th>
<th>10h 30m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes:</td>
<td>10h 30m</td>
</tr>
</tbody>
</table>

**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

**Related activities:**
- Theory sessions. Problem sessions. Statistical laboratory.

**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.

---

**Qualification 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

\[
N = 0.2 \times \text{CUEST} + 0.20 \times \text{EJER} + 0.6 \times \text{EF}
\]

---

**Bibliography**

**Basic:**