240EO011 - Applied Statistics

Coordinating unit: 240 - ETSEIB - Barcelona School of Industrial Engineering
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
Academic year: 2019
Degree: MASTER'S DEGREE IN MANAGEMENT ENGINEERING (Syllabus 2012). (Teaching unit Compulsory)
ECTS credits: 6
Teaching languages: English

Teaching staff
Coordinator: Ginebra Molins, Josep
Others: Puig Oriol, Xavier
Ginebra Molins, Josep

Degree competences to which the subject contributes

Basic:
CB7. (ENG) Que els estudiants sàpiguin aplicar els coneixements adquirits i la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinàris) relacionats amb la seva àrea d'estudi.
CB8. (ENG) Que els estudiants siguin capaços de d'integrar coneixements i enfrentar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'ús dels seus coneixements i judicis.
CB9. (ENG) Que els estudiants sàpiguen comunicar les seves conclusions i coneixements (i darrers raonaments que els sostenint), a públics especialitzats i no especialitzats de manera clara i sense ambigüïtats.

Specific:
2. Acquire concepts and techniques related to descriptive and inferential statistics.
3. Apply concepts and techniques of descriptive and inferential statistics.
CEO3. Acquire concepts and techniques related to quantitative and experimental methods for analysis and decision making.

Generical:
1. Learn and master the analytical tools necessary for decision making in the organizational context more efficient.

Transversal:
CT3. 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.
CT4. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.
240EO011 - Applied Statistics

**Teaching methodology**

There are two types of sessions: main lectures and sessions in the computer lab. In the lectures (2 hours per week) the teacher explains the basics of the subject using examples.

In the hands-on sessions (2 hours per week), practical problems will be solved using statistical packages.

Students will need to do a team project in which they will have to build a model to a set of data selected by them, and several smaller individual project that will be collected.

**Learning objectives of the subject**

After the course the student will be able to:
1. Design how to collect data and how to convert these data into useful information for decision making in environments where there is variability.
2. Understand the concept of variability and how it is measured.
3. Know and apply some of the most common techniques of data collection and analysis.
4. Learn how to build statistical models to summarize information, make predictions, dimensionality reduction and classification.
5. Learn the use of statistical software to solve problems as close as possible to those in their future professional work.

**Study load**

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Hours medium group: 27h</th>
<th>18.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours small group: 27h</td>
<td>18.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
# 240EO011 - Applied Statistics

## Content

### Chapter 1: Probability and Statistics

<table>
<thead>
<tr>
<th>Learning time: 9h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td>Self study: 5h</td>
</tr>
</tbody>
</table>

**Description:**
1. Sample, population and the difference between statistics and probability. 2. Discrete probability models; Binomial and Poisson. 3. Continuous probability models; Normal model. 4. Point estimation of the population mean. 5. Confidence intervals and hypotheses tests for the population mean.

**Specific objectives:**
Understand the concept of variability and how it can be modeled. Know how to use some probability models. Difference between probability and statistics. The concepts of confidence intervals and hypotheses tests.

### Chapter 2: Linear model

<table>
<thead>
<tr>
<th>Learning time: 45h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 10h</td>
</tr>
<tr>
<td>Laboratory classes: 10h</td>
</tr>
<tr>
<td>Self study: 25h</td>
</tr>
</tbody>
</table>

**Description:**
1. Theoretical versus fitted model. 2. Model fit by least squares and other model fit criteria. 3. ANOVA table and goodness of fit measures. 4. Statistical inference on the parameters of the model. 5. Prediction. 6. Model checking through residual analysis. 7. Cross validation and model selection. 8. Interpretation of the fitted model; Bias, collinearity and causality. 9. Use of categorical explanatory variables. 10. Comparison of means.

**Specific objectives:**
To learn how to build a linear model that relates one continuous response variable to a list of explanatory variables, to learn how to interpret that model and make predictions with it, and to learn how to design experiments statistically efficient.

### Chapter 3: Non-linear models

<table>
<thead>
<tr>
<th>Learning time: 9h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td>Self study: 5h</td>
</tr>
</tbody>
</table>

**Description:**

**Specific objectives:**
To learn how to build a non-linear model that relates one continuous response variable to a list of explanatory variables, to learn how to interpret that model and make predictions with it, and to check that the model is a good enough approximation of reality.
### Chapter 4: Categorical response models

**Learning time:** 18h  
**Theory classes:** 4h  
**Laboratory classes:** 4h  
**Self study:** 10h

**Description:**  
1. Binary response models  
2. Model fit  
3. Statistical inference  
4. Model checking  
5. Interpretation of the fitted model  

**Specific objectives:**  
To learn how to build a linear model that relates one binary response variable to a list of explanatory variables and to learn how to interpret that model and make predictions with it.

### Chapter 5: Time series models

**Learning time:** 18h  
**Theory classes:** 4h  
**Laboratory classes:** 4h  
**Self study:** 10h

**Description:**  
1. Description of a time series  
2. AR models  
3. MA models  
4. ARIMA models  
5. Seasonal ARIMA models.

**Specific objectives:**  
To learn how to build a time series model.

### Chapter 6: Dimension reduction and visualization of multivariate data

**Learning time:** 9h  
**Theory classes:** 2h  
**Laboratory classes:** 2h  
**Self study:** 5h

**Description:**  
1. Principal components analysis  
2. Correspondence Analysis

**Specific objectives:**  
Learn about dimensionality reduction techniques both for continuous as well as discrete data.
240EO011 - Applied Statistics

Chapter 7: Unsupervised classification

<table>
<thead>
<tr>
<th>Learning time: 9h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td>Self study: 5h</td>
</tr>
</tbody>
</table>

Description:

Specific objectives:
To learn about unsupervised and supervised classification (learning).

Tema 8: Supervised classification

<table>
<thead>
<tr>
<th>Learning time: 9h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td>Self study: 5h</td>
</tr>
</tbody>
</table>

Description:
1. Linear discriminant analysis. 2. Quadratic discriminant analysis. 3. Logistic discriminant analysis. 4. Nearest neighbors

Qualification system

The final course grade (NF) will be computed starting from four 'inputs':

1) Grade in small projects: NP
2) Grade in the team project: NT
3) Midterm exam: EP
4) Final exam: EF

The final course grade will be obtained through: $NF = 0.10*NP + 0.20*NT + 0.20*EP + 0.50*EF$

Bibliography

Basic: