270601 - SMDE - Statistical Modelling and Design of Experiments

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
Academic year: 2019
Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Compulsory)
ECTS credits: 6
Teaching languages: English

Prior skills

Students must have sufficient knowledge of algebra and mathematical analysis to assimilate the concepts related to algebra of sets, numerical series, functions of real variables of one or more dimensions, derivation and integration.

Degree competences to which the subject contributes

Specific:
CEE2.3. Capability to understand models, problems and mathematical tools to analyze, design and evaluate computer networks and distributed systems.
CEE4.1. Capability to analyze, evaluate and design computers and to propose new techniques for improvement in its architecture.

General:
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.

Transversal:
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Teaching methodology

The course is practical and aims that students will be able, once the course is completed and from the work done in the sessions, to solve real problems similar to those developed in class.

Learning objectives of the subject

1. Applying the mathematical formalism to solve problems involving uncertainty.
2. Applying the queuing models for computer systems performance evaluation and/or configurations analysis.
3. Ability to design, conduct experiments and analyze results.
## Study load

<table>
<thead>
<tr>
<th><strong>Total learning time:</strong> 150h</th>
<th>Theory classes: 26h</th>
<th>17.33%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical classes: 0h</td>
<td></td>
<td>0.00%</td>
</tr>
<tr>
<td>Laboratory classes: 26h</td>
<td></td>
<td>17.33%</td>
</tr>
<tr>
<td>Guided activities: 2h</td>
<td></td>
<td>1.33%</td>
</tr>
<tr>
<td>Self study: 96h</td>
<td></td>
<td>64.00%</td>
</tr>
</tbody>
</table>
Content

Introduction to probability

Degree competences to which the content contributes:

Description:
Students should feel comfortable with the use of set notation and basic statistical terminology. Likewise, the student should be able to write the sample space of simple experiments, including sampling with replacement (like throwing coins or throwing dice), sampling without replacement, from Bernoulli trials and with rules of detention. Likewise, the student should be able to calculate the probabilities in simple cases of the above type of experiment.

Introduction to statistical estimation

Degree competences to which the content contributes:

Description:
Estimation, in the framework of statistical inference, is the set of techniques with the aim of give an approximate value for a parameter of a population from data provided by a sample. From the different methods that exist (point estimate, estimate intervals, or Bayesian estimation) we focus on the point estimate.

Analysis of data

Degree competences to which the content contributes:

Description:
The main objective of the section is to know the procedures associated with the analysis of variance (ANOVA terminology in English) and when is useful to be applied. This activity also introduces MANOVA, as a technique useful when there are two or more dependent variables. We also work with the techniques of linear regression and PCA, completing the repertoire of tools for data analysis.

Introduction to experimental design

Degree competences to which the content contributes:

Description:
Statistical experimental design, a.k.a. design of experiments (DoE) is the methodology of how to conduct and plan experiments in order to extract the maximum amount of information in the fewest number of runs (saving resources). In this section we describe different techniques to achieve that.

Introduction to queuing theory and simulation

Degree competences to which the content contributes:

Description:
This section will introduce the student to use the techniques of operations research for systems analysis for making quantitative decision in the presence of uncertainty through their representation in terms of queuing models and simulation.
# Planning of activities

## Introduction to probability

**Hours:** 9h  
Theory classes: 1h  
Practical classes: 1h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 5h

**Description:**  
At the end of this activity the Student must be comfortable with using basic set notation and terminology. Also the Student must be capable of write down the sample space for simple experiments, including sampling with replacement (such as tossing coins or rolling dice), sampling without replacement, and Bernoulli trials with stopping rules. Also the Student must be capable of calculate probabilities in straightforward instances of the above types of experiment.

**Specific objectives:**

1

## Introduction to statistical estimation

**Hours:** 16h  
Theory classes: 2h  
Practical classes: 2h  
Laboratory classes: 4h  
Guided activities: 0h  
Self study: 8h

**Description:**  
Estimation, in the framework of statistical inference, is the set of techniques with the aim of give an approximate value for a parameter of a population from data provided by a sample. From the different methods that exist (point estimate, estimate intervals, or Bayesian estimation) we focus on the point estimate.

**Specific objectives:**

1

## ANalysis Of VAriance

**Hours:** 9h  
Theory classes: 1h  
Practical classes: 1h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 5h

**Description:**  
The main objective of the activity is to know the procedures associated with the analysis of variance (ANOVA terminology in English) and when is useful to be applied. This activity also introduces MANOVA, as a technique useful when there are two or more dependent variables.

**Specific objectives:**

1
### Linear regression

| Hours: 10h |
| Theory classes: 1h |
| Practical classes: 1h |
| Laboratory classes: 2h |
| Guided activities: 0h |
| Self study: 6h |

**Description:**
Linear regression is a mathematical method that models the relationship between a dependent variable $Y$, independent variables $X_i$ and a random term. This section will examine this method and explain its applicability from different examples.

**Specific objectives:**
1

---

### Principal component analysis

| Hours: 10h |
| Theory classes: 1h |
| Practical classes: 1h |
| Laboratory classes: 2h |
| Guided activities: 0h |
| Self study: 6h |

**Description:**
The principal component analysis (PCA, PCA in English), in statistics, is a technique that reduces the dimensionality of a dataset. This allows us to represent them graphically in two or three dimensional graphs of various variables grouped the data into factors, or components, consisting of the grouping variables. In this section we will work this technique from a practical point of view.

**Specific objectives:**
1

---

### Factorial design

| Hours: 27h |
| Theory classes: 3h |
| Practical classes: 3h |
| Laboratory classes: 9h |
| Guided activities: 0h |
| Self study: 12h |

**Description:**
Many experiments are conducted to study the effects of two or more factors. In this case the factorial designs are more efficient, presented in this section.

**Specific objectives:**
3
### Randomized blocks, Latin squares and related designs

**Hours:** 10h  
Theory classes: 1h  
Practical classes: 1h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 6h

**Description:**  
In many research problems it is necessary to design experiments that can systematically control the variability caused by different sources. This section will consider some experimental designs for solving these situations.

**Specific objectives:**

### Incomplete block design

**Hours:** 10h  
Theory classes: 1h  
Practical classes: 1h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 6h

**Description:**  
Description. Incomplete blocks design, useful when you cannot develop all combinations of treatment within each block.

**Specific objectives:**

### General structure of queuing models

**Hours:** 9h  
Theory classes: 1h  
Practical classes: 1h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 5h

**Description:**  

**Specific objectives:**

### Queuing models based on birth and death processes

**Hours:** 9h  
Theory classes: 1h  
Practical classes: 1h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 5h
<table>
<thead>
<tr>
<th>Description</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized queuing patterns with non-exponential distributions and serial exponential queues.</td>
<td>9h Theory classes: 1h Practical classes: 1h Laboratory classes: 2h Guided activities: 0h Self study: 5h</td>
</tr>
<tr>
<td>Validation Verification and Accreditation</td>
<td>9h Theory classes: 1h Practical classes: 1h Laboratory classes: 2h Guided activities: 0h Self study: 5h</td>
</tr>
<tr>
<td>First report</td>
<td>5h Guided activities: 0h Self study: 5h</td>
</tr>
<tr>
<td>Second report</td>
<td>5h Guided activities: 0h Self study: 5h</td>
</tr>
</tbody>
</table>

Specific objectives:

1. Introduction to basic concepts and elements of the analysis of Markov processes. Markov queues.
3. Techniques to Verify, Validate and do the Accreditation of the models.
4. Validation Verification and Accreditation
5. First report
6. Second report

Hours: 9h
Theory classes: 1h
Practical classes: 1h
Laboratory classes: 2h
Guided activities: 0h
Self study: 5h
### Third report

| Hours | Guided activities: 0h  
Self study: 5h |
|-------|-------------------|

### Specific objectives:
- 3

### Final exam

| Hours | Guided activities: 2h  
Self study: 10h |
|-------|-------------------|

### Specific objectives:
- 1, 2, 3

### Qualification system

The course will have different exercises that the students must solve during the course (80% of the final grade). At the end there will be an exam that will weigh 20% of the final grade.
Bibliography

Basic:


Complementary:


Others resources:
Hyperlink
http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/p

http://cran.r-project.org/