200601 - CEO - Software for Statistics and Optimization

Coordinating unit: 200 - FME - School of Mathematics and Statistics
Teaching unit: 715 - EIO - 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 Compulsory)
ECTS credits: 5

Teaching languages: Spanish

Teaching staff

Coordinator: KLAUS GERHARD LANGOHR
Others: Primer quadrimestre:
  DANIEL FERNÁNDEZ MARTÍNEZ - A, B
  KLAUS GERHARD LANGOHR - A, B
  ANTONIO MONLEON GETINO - A, B
  LUIS ORTIZ GRACIA - A, B
  ANA MARIA PÉREZ MARÍN - A, B

Opening hours

Timetable: At agreed times.

Prior skills

Concerning the R lectures, there will be two courses: an introductory-level course and an intermediate-level course. The first is for students with no or little experience of R, the second for students who have worked with R previously such as students with a degree in statistics. By contrast, the SAS lectures will be the same for all students.

Requirements

The intermediate-level R course requires that students have experience in working with R.

Degree competences to which the subject contributes

Specific:
3. CE-1. Ability to design and manage the collection of information and coding, handling, storing and processing it.
4. CE-5. Ability to formulate and solve real problems of decision-making in different application areas being able to choose the statistical method and the optimization algorithm more suitable in every occasion.
   Translate to English
5. CE-6. Ability to use appropriate software to perform the necessary calculations in solving a problem.
7. CE-9. Ability to implement statistical and operations research algorithms.

Transversal:
1. 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.
200601 - CEO - Software for Statistics and Optimization

2. 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.

Teaching methodology

The lectures will take place in the computer room where both statistical packages, R and SAS, will be presented. The first part of the course will be dedicated to R and the second part to SAS. To illustrate the use of functions for statistics and graphics, real data sets will be used. During the course, students will have to do exams (in class) and a final exercise (at home) with each software package.

Learning objectives of the subject

In this course, two statistical software packages are presented, R and SAS, that are widely used in the academic field as well as in business and industry.

The course aims to enable the student to use both software packages to

- read data from external files,
- carry out descriptive analysis,
- make high quality graphs to represent data,
- fit regression models to data sets,
- write own functions.

Study load

Total learning time: 125h

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

The lectures will take place in the computer room where both statistical packages, R and SAS, will be presented. The first part of the course will be dedicated to R and the second part to SAS. To illustrate the use of functions for statistics and graphics, real data sets will be used. During the course, students will have to do exams (in class) and a final exercise (at home) with each software package.
# Content

<table>
<thead>
<tr>
<th>Introduction to R [Introductory level]</th>
<th>Learning time: 1h 30m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 1h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 0h 30m</td>
</tr>
</tbody>
</table>

**Description:**
- a) The web page of R
- b) Installation of R and its contributed packages
- c) Sources of help

<table>
<thead>
<tr>
<th>R objects</th>
<th>Learning time: 6h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 4h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
</tbody>
</table>

**Description:**
- Creation and manipulation of
  - a) Numeric and alphanumeric vectors,
  - b) Matrices,
  - c) Lists,
  - d) Data frames.

<table>
<thead>
<tr>
<th>Descriptive and exploratory analysis with R</th>
<th>Learning time: 6h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 4h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
</tbody>
</table>

**Description:**
- a) Reading external data files
- b) Univariate descriptive analysis
- c) Bivariate descriptive analysis
- d) Graphical tools: histogram, box plot, scatter plot and others
<table>
<thead>
<tr>
<th><strong>200601 - CEO - Software for Statistics and Optimization</strong></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>Basic programming with R</strong></th>
<th><strong>Learning time:</strong> 6h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 4h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
</tbody>
</table>

**Description:**
- a) Basic programming: loops with for, while, if-else
- b) Functions tapply, sapply, lapply
- c) Writing your own function
- d) Working with date variables

<table>
<thead>
<tr>
<th><strong>Statistical inference with R: hypothesis tests and regression models</strong></th>
<th><strong>Learning time:</strong> 1h 30m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 1h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 0h 30m</td>
</tr>
</tbody>
</table>

**Description:**
- a) Hypothesis tests for one population
- b) Hypothesis tests for two or more populations
- c) Nonparametric tests
- d) Fit of general linear models

<table>
<thead>
<tr>
<th><strong>Intermediate-level R topics</strong></th>
<th><strong>Learning time:</strong> 1h 30m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 1h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 0h 30m</td>
</tr>
</tbody>
</table>

**Description:**
- a) Reshaping data sets
- b) Intermediate level programming with R
- c) An introduction to Tidyverse
- d) Integrating R code in LaTeX documents
<table>
<thead>
<tr>
<th><strong>Introduction to SAS</strong></th>
<th><strong>Learning time:</strong> 1h 30m</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td></td>
</tr>
<tr>
<td>a) Structure of the SAS programs: DATA and PROC.</td>
<td>Theory classes: 1h</td>
</tr>
<tr>
<td>b) SAS data sets and libraries.</td>
<td>Laboratory classes: 0h 30m</td>
</tr>
<tr>
<td>c) Importation and exportation of data.</td>
<td></td>
</tr>
<tr>
<td>d) Creation of variables. Commands of assignment.</td>
<td></td>
</tr>
<tr>
<td>e) Merging data bases.</td>
<td></td>
</tr>
<tr>
<td>f) Management of data sets</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Basic procedures with SAS</strong></th>
<th><strong>Learning time:</strong> 6h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td></td>
</tr>
<tr>
<td>a) Introduction to procedures.</td>
<td>Theory classes: 4h</td>
</tr>
<tr>
<td>b) Statistical and graphical procedures.</td>
<td>Laboratory classes: 2h</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Transformation and manipulation of data</strong></th>
<th><strong>Learning time:</strong> 6h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td></td>
</tr>
<tr>
<td>a) Use of predefined functions.</td>
<td>Theory classes: 4h</td>
</tr>
<tr>
<td>b) Conditional transformation of variables.</td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td>c) Data generation with DO loops.</td>
<td></td>
</tr>
<tr>
<td>d) Date variables.</td>
<td></td>
</tr>
<tr>
<td>e) String functions.</td>
<td></td>
</tr>
<tr>
<td>f) Error diagnosis and depuration.</td>
<td></td>
</tr>
</tbody>
</table>
**Introduction to matrix calculus with SAS: SAS/IML**

**Description:**
- a) Introduction to the SAS/IML module.
- b) Matrix definition.
- c) Operators and functions of SAS/IML.
- d) Importation and exportation of data bases from IML.

**Learning time:** 6h  
- Theory classes: 4h  
- Laboratory classes: 2h

---

**SAS macros**

**Description:**
- a) Introduction to the SAS macro language
- b) Definition of macro variables
- c) Creation of SAS macros

**Learning time:** 1h  
- Theory classes: 1h

---

**Advanced procedures**

**Description:**
- a) Introduction to the SAS/STAT module
- b) Parametric hypothesis tests: PROC TTEST, PROC ANOVA.
- c) Analysis of regression models: PROC REG and PROC GLM.

**Learning time:** 1h 30m  
- Theory classes: 1h  
- Laboratory classes: 0h 30m

---

**Introduction to linear programming with SAS**

**Description:**
- a) Introduction to the SAS/OR module
- b) Formulation and solution of linear programming models: PROC PL, PROC OPTLP, and PROC OPTMODEL

**Learning time:** 1h 30m  
- Theory classes: 1h  
- Laboratory classes: 0h 30m
Qualification system

The final grade will be the average of the grades obtained in the different tests
a) with R (50%),
b) with SAS (50%).
Concerning R, there will be two exams in class (weight of each test: 30%) and a final practical work at home (weight: 40%). Concerning SAS, there will be two exams in class (weight of each test: 40%) and a final practical work at home (weight: 20%).
Bibliography

Basic:


Complementary:


