Course guides
200648 - SERS - Statistical Software: R and SAS

Unit in charge: School of Mathematics and Statistics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research.
1004 - UB - (ENG)Universitat de Barcelona.
Degree: MASTER’S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Compulsory subject).
Academic year: 2021 ECTS Credits: 5.0 Languages: English

LECTURER

Coordinating lecturer: KLAUS GERHARD LANGOHR

Others:
- Primer quadrimestre: KLAUS GERHARD LANGOHR - A, B
- ANTONIO MONLEON GETINO - A, B
- DAVID MORIÑA SOLER - A, B
- ANA MARIA PÉREZ MARÍN - A, B

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.
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 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

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours large group</td>
<td>30,0</td>
<td>24.00</td>
</tr>
<tr>
<td>Hours small group</td>
<td>15,0</td>
<td>12.00</td>
</tr>
<tr>
<td>Self study</td>
<td>80,0</td>
<td>64.00</td>
</tr>
</tbody>
</table>

Total learning time: 125 h

CONTENTS

Introduction to R [Introductory level]

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

Full-or-part-time: 1h 30m
Theory classes: 1h
Laboratory classes: 0h 30m

R objects

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

Full-or-part-time: 6h
Theory classes: 4h
Laboratory classes: 2h
### Descriptive and exploratory analysis with R

**Description:**
- a) Reading external data files
- b) Univariate descriptive analysis
- c) Bivariate descriptive analysis
- d) Graphical tools: histogram, box plot, scatter plot and others

**Full-or-part-time:** 6h  
Theory classes: 4h  
Laboratory classes: 2h

### Basic programming with R

**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

**Full-or-part-time:** 6h  
Theory classes: 4h  
Laboratory classes: 2h

### Statistical inference with R: hypothesis tests and regression models

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

**Full-or-part-time:** 1h 30m  
Theory classes: 1h  
Laboratory classes: 0h 30m

### Intermediate-level R

**Description:**
- a) Review of working with data frames  
- b) Reshaping data sets  
- c) Intermediate level programming with R  
- d) An introduction to Tidyverse

**Full-or-part-time:** 1h 30m  
Theory classes: 1h  
Laboratory classes: 0h 30m
### Introduction to SAS

**Description:**
- a) Structure of the SAS programs: DATA and PROC.
- b) SAS data sets and libraries.
- c) Importation and exportation of data.
- d) Creation of variables. Commands of assignment.
- e) Merging data bases.
- f) Management of data sets

**Full-or-part-time:** 1h 30m  
Theory classes: 1h  
Laboratory classes: 0h 30m

### Basic procedures with SAS

**Description:**
- a) Introduction to procedures.
- b) Statistical and graphical procedures.

**Full-or-part-time:** 5h  
Theory classes: 2h 30m  
Laboratory classes: 2h 30m

### Transformation and manipulation of data

**Description:**
- a) Use of predefined functions.
- b) Conditional transformation of variables.
- c) Data generation with DO loops.
- d) Date variables.
- e) String functions.
- f) Error diagnosis and depuration.

**Full-or-part-time:** 5h 30m  
Theory classes: 3h 30m  
Laboratory classes: 2h

### 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.

**Full-or-part-time:** 5h 30m  
Theory classes: 3h 30m  
Laboratory classes: 2h
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.

Full-or-part-time: 1h 30m
Theory classes: 1h
Laboratory classes: 0h 30m

SAS macros

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

Full-or-part-time: 3h 30m
Theory classes: 3h 30m

GRADING 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 tests: 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: