



Course guide

200648 - SERS - Statistical Software: R and SAS

Last modified: 12/06/2023

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: 2023	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 MIGUEL ANGEL SANTOLINO PRIETO - A, B

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

Lectures will be held in computer rooms. 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 two exams 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

Type	Hours	Percentage
Hours large group	30,0	24.00
Hours small group	15,0	12.00
Self study	80,0	64.00

Total learning time: 125 h

CONTENTS

Introductory R: Introduction to R [Introductory level]

Description:

- The web page of R
- Installation of R and its contributed packages
- Sources of help

Full-or-part-time: 1h 30m

Theory classes: 1h

Laboratory classes: 0h 30m

Introductory R: R objects

Description:

Creation and manipulation of

- Numeric and alphanumeric vectors,
- Matrices,
- Lists,
- Data frames.

Full-or-part-time: 6h

Theory classes: 4h

Laboratory classes: 2h



Introductory R: 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

Introductory R: 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

Introductory R: 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) The tidyverse packages

Full-or-part-time: 1h 30m

Theory classes: 1h

Laboratory classes: 0h 30m



Advanced R: Creation of R packages and web application with Shiny

Description:

- a) Preparing R code for the package
- b) Preparing the manual of the package
- c) Creating the package, checking, and error debugging
- d) Repositories and publication of packages

Full-or-part-time: 3h

Theory classes: 2h 30m

Laboratory classes: 0h 30m

Advanced R: Version control with R

Description:

- a) Systems for version control (git and subversion).
- b) Github.
- c) Version control in RStudio

Full-or-part-time: 3h

Theory classes: 2h 30m

Laboratory classes: 0h 30m

Advanced R: Automatic report generation with R

Description:

- a) Introduction to markdown
- b) Rmarkdown with RStudio
- c) Generation of presentations with RStudio
- d) Quarto documents with RStudio

Full-or-part-time: 3h

Theory classes: 2h 30m

Laboratory classes: 0h 30m

Advanced R: Parallel programming with R

Description:

- a) Objectives of parallel programming with R.
- b) R packages for parallel programming: parallel, doParallel, SNOW.
- c) Examples.

Full-or-part-time: 3h

Theory classes: 1h 30m

Laboratory classes: 1h 30m



Introduction to SAS

Description:

- a) Structure of the SAS programmes: 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.
- c) Introduction to matrix calculus with SAS/IML

Full-or-part-time: 3h

Theory classes: 1h 30m

Laboratory classes: 1h 30m

Management of large databases: SAS/SQL

Description:

- a) Introduction to the SAS/SQL module.
- b) Definition of SAS/SQL tables.
- c) Definition of SAS/SQL databases.
- d) Operations with one or more tables.

Full-or-part-time: 4h 30m

Theory classes: 3h

Laboratory classes: 1h 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: 1h 30m

Theory classes: 1h

Practical classes: 0h 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 (weight of each exam: 30%) and a final practical work at home (weight: 40%). Concerning SAS, there will be two exams in class (weight of each exam: 40%) and a final practical work at home (weight: 20%).

BIBLIOGRAPHY

Basic:

- Braun, W.J.; Murdoch, D.J. A First course in statistical programming with R. Cambridge University Press, 2007. ISBN 97805216944247.
- Kleinmann, K.; Horton, N.J. SAS and R: Data management, statistical analysis and graphics. Chapman & Hall, 2009. ISBN 978-1-4200-7057-6.
- Der, Geoff; Everitt, Brian. A Handbook of statistical analyses using SAS. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC, cop. 2009. ISBN 9781584887843.
- Crawley, Michael J. Statistics : an introduction using R. New York: John Wiley & Sons, 2005. ISBN 0470022973.
- Cody, R. Learning SAS by example: a programmer's guide [on line]. SAS Institute, 2007 Available on: <https://ebin.pub/learning-sas-by-example-a-programmers-guide-1599941651-978-1-59994-165-3.html>. ISBN 9781599941653.
- Cody, R. SAS statistics by example. SAS Institute, 2011. ISBN 9781607648000.
- Delwiche, L.D.; Slaughter, S.J. The Little SAS Book : a primer. 5th Edition. SAS Institute, 2012. ISBN 9781612903439.
- Dalgaard, P. Introductory statistics with R [on line]. 2nd Edition. Springer, 2008 [Consultation: 28/06/2023]. Available on: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-79054-1>. ISBN 9780387790541.
- Kleinman, Ken; Horton, Nicholas J. SAS and R : data management, statistical analysis, and graphics. Boca Raton, Fla.: CRC Press, 2010. ISBN 9781420070576.

Complementary:

- Murrell, P. R graphics. Chapman & Hall, 2006. ISBN 158488486X.
- Muenchen, R.A. R for SAS and SPSS users. Springer, 2011. ISBN 9781461406853.
- Base SAS® 9.2 Procedures Guide [on line]. SAS Institute, 2009 [Consultation: 28/06/2023]. Available on: <http://support.sas.com/documentation/cdl/en/proc/61895/PDF/default/proc.pdf>. ISBN 9781599947143.
- Base SAS® 9.2 Procedures Guide: Statistical Procedures [on line]. 3rd Edition. SAS Institute, 2010 [Consultation: 28/06/2023]. Available on: <http://support.sas.com/documentation/cdl/en/procstat/63104/PDF/default/procstat.pdf>. ISBN 9781607644514.
- SAS/IML® 9.2 Users guide [on line]. SAS Institute, 2008 Available on: <http://support.sas.com/documentation/cdl/en/imlug/59656/PDF/default/imlug.pdf>. ISBN 9781590479407.
- SAS/OR®9.2 User's guide mathematical programming [on line]. SAS Institute, 2008 Available on: <http://support.sas.com/documentation/cdl/en/ormpug/59679/PDF/default/ormpug.pdf>. ISBN 9781590479469.
- SAS/STAT 9.2 User's guide [on line]. 2nd Edition. SAS Institute, 2011 Available on: <http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#titlepage.htm>. ISBN 9781607648826.
- SAS 9.2. Language reference: concepts [on line]. 2nd Edition. SAS Institute, 2010 Available on: <http://support.sas.com/documentation/cdl/en/lrcon/62955/PDF/default/lrcon.pdf>. ISBN 9781607644484.
- SAS 9.2. Language reference : dictionary [on line]. 4th Edition. SAS Institute, 2011 Available on: <http://support.sas.com/documentation/cdl/en/lrdict/64316/PDF/default/lrdict.pdf>. ISBN 9781607648826.
- Wickham, Hadley; Grolemond, Garrett. R for data science : import, tidy, transform, visualize, and model data. First edition. 2016. ISBN 9781491910399.