Course guides
33104 - TAEDDPE - Techniques of Statistic Analysis of Data and Design and Planning of Experiments

Last modified: 31/05/2020

Unit in charge: Manresa School of Engineering
Teaching unit: 749 - MAT - Department of Mathematics.

Degree: MASTER'S DEGREE IN NATURAL RESOURCE ENGINEERING (Syllabus 2009). (Compulsory subject).
MASTER'S DEGREE IN NATURAL RESOURCE ENGINEERING (Syllabus 2015). (Compulsory subject).
MASTER'S DEGREE IN NATURAL RESOURCE ENGINEERING (Syllabus 2008). (Compulsory subject).

Academic year: 2020 **ECTS Credits:** 5.0 **Languages:** Spanish

LECTURER

Coordinating lecturer: FRANCISCO PALACIOS QUIÑONERO

Others: JOSEP MARIA ROSSELL GARRIGA

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
1. Have the ability to analyse field and laboratory data and design experiments using computer methods.
2. A capacity for understanding analytical techniques for the characterisation of inorganic natural resources and waste in different states, using these techniques and interpreting the results.
3. The ability to use scientific and technical information to respond efficiently to any demand for the preparation of an analytical method for characterising a material of natural or anthropogenic origin.

TEACHING METHODOLOGY

The module is based on a set of directed learning activities, which include the study of the main statistical methods and solving applied problems with a computer. In the face-to-face modality, expository and practical classes will be held to introduce the new concepts, provide help in the use of the computational resources and solving doubts. In the blended (semipresencial) modality, a proper program of activities will be proposed to allow the students cover autonomously the learning objectives. In this second case, online consultation sessions will be held to provide study guidelines and solving doubts. The use of advanced computational resources is a fundamental element in the application of statistical methods. In this module, we will use the statistical computational environment R, which provides a cross-platform and open-souce high-performance tool for professional practice, research and learning of advanced statistical methods.

LEARNING OBJECTIVES OF THE SUBJECT

The primary objective of the course is to familiarize the student with the main statistical methods, both in their theoretical and applied aspects. Specifically, we will introduce and / or review the fundamental concepts and basic tools of statistics to carry out sampling, process data, analyze the results, make statistical inferences, formulate and fit models, design experiments and work with time series. The multidisciplinary nature of the concepts and methods studied will be illustrated by using environmental, industrial, biomedical and socioeconomic data.

STUDY LOAD

Type	Hours	Percentage
Hours large group	30,0	66.67
Hours medium group	15,0	33.33

Total learning time: 45 h

CONTENTS

Unit 1: Basic concepts of statistics. Introduction to R.

Description:
Data types. Descriptive and inferential statistics. Random sampling. Types of sampling. Basics concepts of probability. Installation of R. Basic operation of R. Computation of statistical descriptors with R. Graphical representations with R.

Specific objectives:
Review of the fundamental statistics concepts. Installation and basic usage of R.

Related activities:
Preparation and submission of the exercise set E1.

Full-or-part-time: 25h
Theory classes: 5h
Laboratory classes: 5h
Self study : 15h

Unit 2. Single-variable statistics.

Description:
Random variables. Discrete distributions: uniform, binomial, geometric, hypergeometric, Poisson. Continuous distributions: exponential, gamma, normal, t-Student, chi-squared, F-Fisher. Computation of probability distributions with R. Simulation of random variables with R. Confidence intervals. Computation of confidence intervals with R. Hypothesis tests. Hypothesis testing with R. Power of a test. Nonparametric hypothesis tests.

Specific objectives:
Review of the main probability distributions and their application with R. Review of the main inferential statistical tools for single-variable problems and their implementation with R.

Related activities:
Preparation and submission of the exercise set E2.

Full-or-part-time: 20h
Theory classes: 5h
Laboratory classes: 2h
Self study : 13h

Unit 3. Fitting of single-predictor models.

Description:
Models with a single predictor. Linear regression. Pearson and determination coefficients. Fitting of polynomial models. Fitting of nonlinear models. Model fitting with R

Specific objectives:
Revision of the main concepts associated to the fitting of single-predictor models and their implementation with R.

Related activities:
Preparation and submission of the exercise set E3.

Full-or-part-time: 20h
Theory classes: 5h
Laboratory classes: 2h
Self study : 13h

Unit 4: Multivariate statistics.

Description:
Introduction to multivariable statistics. Multiple linear regression. Multiple linear regression with R.

Specific objectives:
Presentation of the basic concepts of multivariable statistics. Discussion of model fitting with several predictors and its implementation with R.

Related activities:
Preparation and submission of the exercise set E4.

Full-or-part-time: 20h
Theory classes: 5h
Laboratory classes: 2h
Self study : 13h

Unit 5: Design and analysis of experiments.

Description:
Observational studies and experiments. Unifactorial experiments. Multifactorial designs. Design of experiments with R.

Specific objectives:
Presentation of the main elements of the design of experiments and their implementation with R.

Related activities:
Preparation and submission of the exercise set E5.

Full-or-part-time: 20h
Theory classes: 5h
Laboratory classes: 2h
Self study : 13h

Unit 6: Time series.

Description:
Introduction to time series. Stationary time series. Non-stationary time series. Time series analysis with R.

Specific objectives:
Presentation of the main elements of time series and their implementation with R.

Related activities:
Preparation and submission of the exercise set E6.

Full-or-part-time: 20h
Theory classes: 5h
Laboratory classes: 2h
Self study : 13h

GRADING SYSTEM

The grading system will follow a continuous assessment scheme, based on the Exercise Sets E1-E6 that the students will deliver periodically through the Atenea digital platform. In each Exercise Set Ej, j = 1, ..., 6, the student will obtain a MEj mark between 0 and 10. The global mark of the module will be GM = (ME1 + ···· + ME6) / 6.

EXAMINATION RULES.

Students must individually solve the Exercise Sets E1-E6 and submit them through the Atenea digital platform within the established deadlines for their correction and grading. Not submitted exercises sets will be graded with a 0 mark. When deemed appropriate, the authorship of the submitted works will be validated by means of an additional questionnaire and / or a personal interview (online or in person).

BIBLIOGRAPHY

Basic:
- Sarkar, Deepayan. Lattice: multivariate data visualization with R [on line]. New York: Springer, 2008 [Consultation: 08/07/2020]. Available on: http://dx.doi.org/10.1007/978-0-387-75969-2. ISBN 9780387759692.
- Wickham, Hadley. ggplot2. Elegant Graphics for Data Analysis [on line]. London: Springer, 2009 [Consultation: 08/07/2020]. Available on: http://dx.doi.org/10.1007/978-0-387-98141-3. ISBN 9780387981413.
- Albert, Jim ; Rizzo, Maria. R by Example [on line]. New York: Springer, 2012 [Consultation: 28/05/2020]. Available on: https://doi.org/10.1007/978-1-4614-1365-3. ISBN 9781461413653.
- Dalgaard, Peter. Introductory Statistics with R [on line]. Second Edition. New York: Springer, 2008 [Consultation: 08/07/2020]. Available on: http://dx.doi.org/10.1007/978-0-387-79054-1. ISBN 9780387790541.
- Heumann, Christian. Schomaker Shalabh, Michael. Introduction to Statistics and Data Analysis [on line]. Switzerland: Springer, 2016 [Consultation: 28/05/2020]. Available on: https://link.springer.com/book/10.1007/978-3-319-46162-5. ISBN 978-3-319-46162-5.
- Everitt, Brian. Hothorn,Torsten. An Introduction to Applied Multivariate Analysis with R [on line]. New York: Springer, 2011 [Consultation: 08/07/2020]. Available on: https://doi.org/10.1007/978-1-4419-9650-3. ISBN 9781441996503.
- Cowpertwait, Paul S.P.; Metcalfe, Andrew V. Introductory time series with R [on line]. London: Springer, 2009 [Consultation: 09/07/2020]. Available on: http://dx.doi.org/10.1007/978-0-387-88698-5. ISBN 9780387886985.
- Verzani, John. Using R for introductory statistics. Second Edition. Boca Raton, FL: Chapman & Hall/CRC, 2005. ISBN 1584884509.
- Montgomery, Douglas C. Design and analysis of experiments. 8th ed., international student version. New York: John Wiley & Sons, 2013. ISBN 9781118097939.
- Heiberger, Richard M.; Holland, Burt. Statistical Analysis and Data Display. An Intermediate Course with Examples in R [on line]. Second Edition. New York: Springer, 2015 [Consultation: 28/05/2020]. Available on: https://link.springer.com/book/10.1007/978-1-4939-2122-5. ISBN 978-1-4939-2122-5.