

62236 - Advanced statistical data analysis

Información del Plan Docente

Academic Year	2016/17
Academic center	110 - Escuela de Ingeniería y Arquitectura
Degree	534 - Master's in IT Engineering
ECTS	3.0
Course	2
Period	First semester
Subject Type	Optional
Module	---

1.Basic info

1.1.Recommendations to take this course

1.2.Activities and key dates for the course

2.Initiation

2.1.Learning outcomes that define the subject

2.2.Introduction

3.Context and competences

3.1.Goals

3.2.Context and meaning of the subject in the degree

3.3.Competences

3.4.Importance of learning outcomes

4.Evaluation

5.Activities and resources

5.1.General methodological presentation

The proposed methodology encourages at students for daily work during the course.

The general principles of the course are presented in sessions where a formal description is carried out with applications to appropriate examples. Classes in the computer lab deal with both data analysis and modelling of real events. In these seasons the students learn to use statistical free software R.

62236 - Advanced statistical data analysis

Each student develops an individual task concerning the use of statistical procedures in big data cases. Students can choose data-bases they are particularly interested in or, alternatively, data-bases provided by the instructor. In both cases a written report is mandatory.

5.2.Learning activities

Interactive lectures

Regular work sessions are developed in the computer lab. In these sessions real situations that promote interest in a wide range of statistical techniques are presented. The associated concepts and statistical procedures are shown from a practical viewpoint. The students are encouraged to model and to solve real problems by means of free software.

The R programming language is used and, in this regard, functions, standard libraries available in R Project are introduced to address different techniques.

Development of a project

Each student has to develop a project for the statistical analysis of a collection of data, with high-dimension, using appropriate techniques to draw conclusions.

Student activities

The course corresponds to 3 ECTS equivalent to 75 hours for students developed in both classroom activities and non-classroom activities whose individual weightings are as follows:

Classroom activities:

30 hours (2 hours/week) of practical classes in group sessions. These classes involve theoretical aspects, problem sets and data analysis.

Non-classroom activities:

40 hours for individual study of general principles or ideas and devoted to practical tasks.

5 hours for student appraisal.

5.3.Program

- Introduction
- o Statistical learning.

62236 - Advanced statistical data analysis

- o Exploratory data analysis.
- o Sampling and statistical inference: point and interval estimation, hypothesis testing.
- o Likelihood: Estimation by maximum likelihood, likelihood ratio test.
- o Statistical decision theory. Bayesian methods.
- o The EM algorithm. The MCMC method.
- o Statistical simulation
 - Recognition of explicit relationships: Regression Models
- o Simple linear regression, review and validation of the model Box-Cox transformation, prediction.
- o General linear model, covariates and factor analysis of variance.
- o Automatic modeling procedures: best subset, stepwise.
- o Validation, cross validation, bootstrap methods.
- o Regression with high dimensionality.
- o or models with non-Gaussian response: GLM and GAM.
 - Supervised pattern recognition: Logistic Regression.
- o Binary logistic regression models.
- o Multinomial logistic regression models.
- o Crosstabulation, log-linear models.
 - Unsupervised pattern recognition.
- o Cluster analysis, k-means method.
- o Hierarchical cluster.

5.4.Planning and scheduling

The course is organized in 2 hours of class per week, that hours take place in a computer lab.

In the Web page of the course is available all the documentation, the schedule of all sessions and a timetable for completion of the work of the subject.

The course schedule will be determined by the academic calendar. The timetable and deadlines for individual reports are announced at the Web page of Engineers School.

Every student is supposed to complete several tasks periodically. These tasks are associated to each learning block and are part of the student's appraisal.

5.5.Bibliography and recommended resources