

Academic Year: ( 2022 / 2023 )

Review date: 17-05-2022

Department assigned to the subject: Statistics Department

Coordinating teacher: GALEANO SAN MIGUEL, PEDRO

Type: Compulsory ECTS Credits : 3.0

Year : 1 Semester : 1

**REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)**

Statistical modelling for data analysis

Mathematics for data analysis

**OBJECTIVES****Basic competences**

To have and to include/understand knowledge that contribute to a base or opportunity of to be original at the development and/or application of ideas, to slight in an investigation context

That the students know to apply to the acquired knowledge and their capacity of resolution of problems in new surroundings or little known within ampler contexts (or you will multidiscipline) related to their area study

That the students are able to integrate knowledge and to face the complexity to formulate judgments from a information that, incomplete or being limited, includes reflections on tie the social and ethical responsibilities to the application of its knowledge and judgments

That the students have the learning abilities that allow them to continue studying of a way that will be to be in great homing or independent measurement.

**General competitions**

To apply the theoretical foundations of the collection techniques , storage, treatment and presentation of information, specially for great volumes of data, as it bases for the development and adaptation of these techniques to concrete problems

To identify different techniques to store, to talk back and to distribute great amounts of data, and to differentiate them based on its theoretical and practical characteristics

To identify the techniques of analyses of data but adapted each problem and knowledge to apply them for the analysis, design and solution of such

To obtain practical and efficient solutions for problems of treatment of great volumes of data, as much individually as in equipment

To synthesize the obtained conclusions of these analyses and to as much present/display them of clear and convincing way in bilingual surroundings (Spanish and English) in writing as orally

To be able to generate new ideas (creativity) and to anticipate new situations, in the contexts of the analysis of data and the decision making

To use abilities for the work in equipment and to be related to others of independent form

**Specific competitions**

To use the basic results of inference and regression like foundation for advanced methods of prediction and classification

To identify and to select the suitable software tools for the treatment of great amounts of data

To use advanced statistical procedures for the treatment of great volumes of data in areas like the estimation, the inference, the prediction or the classification, as well as the way to apply them of efficient form

To correctly identify the kind of statistical problem corresponding to certain objectives and data

To know how to design systems for the processing of the data, from the obtaining and initial filtrate of such, its statistical analysis, to the presentation of the final results

To use techniques and usable tools of operations research with massive data in procedures for its analysis, visualization of its results or within systems of support to decisions

## DESCRIPTION OF CONTENTS: PROGRAMME

1. Multidimensional data
  - 1.1 Multidimensional datasets.
  - 1.2 Visualizing multidimensional datasets.
  - 1.3 A brief summary of multivariate random variables.
  - 1.4 Inference with the data matrix.
  - 1.5 Imputation of missing values.
2. Dimension reduction techniques
  - 2.1 Dimension reduction techniques.
  - 2.2 Principal component analysis.
  - 2.3 Independent component analysis.
  - 2.4 Factor analysis.
3. Unsupervised classification
  - 3.1 Unsupervised classification.
  - 3.2 Partitional clustering.
  - 3.3 Hierarchical clustering.
  - 3.4 Model-based clustering.
  - 3.5 Alternatives.
4. Supervised classification
  - 4.1 Supervised classification.
  - 4.2 K-nearest neighbors (KNN).
  - 4.3 Methods based on the Bayes Theorem.
  - 4.4 Logistic regression.

## LEARNING ACTIVITIES AND METHODOLOGY

Learning activities

Theoretical classes

Practical classes

Student individual work

Methodology

Theoretical classes with support of computers and audiovisuals to present and develop the main concepts of the course. Teachers will provide students with supplementary material.

Critical reading of documents provided by the teachers: newspaper articles, reports, manuals and / or academic papers, either for later discussion in class, either to expand and consolidate the knowledge of the subject.

Resolution of practical cases, problems, etc... proposed by the teacher individually or in groups.

Preparation of projects individually or in group.

## ASSESSMENT SYSTEM

Group project (70%)

Homeworks (20%)

Class presentations (10%)

Ordinary evaluation similar to extraordinary.

<b>% end-of-term-examination:</b>	0
<b>% of continuous assessment (assignments, laboratory, practicals...):</b>	100

## BASIC BIBLIOGRAPHY

- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani An introduction to statistical learning, Springer, 2013
- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani An introduction to statistical learning, Springer, 2013
- Trevor Hastie, Robert Tibshirani and Jerome Friedman The Elements of the Statistical Learning, Springer, 2009

- Trevor Hastie, Robert Tibshirani and Martin Wainwright Statistical learning with sparsity, CRC Press, 2015