Multivariate Analysis

Academic Year: (2021 / 2022) **Review date: 03-06-2021**

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)

Mathematics for Data Science
Probability
Statistical Inference
Programming in R
Numerical Methods for Data Science

OBJECTIVES

Competences that the student acquires:

CB6 Possess and understand knowledge that provides a basis or opportunity to be original in the development and / or application of ideas, often in a research context.

CB7 Applications of the acquired knowledge and ability to solve problems in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their area of ¿¿study.

CB9 Ability to communicate their conclusions and the knowledge and ultimate reasons that sustain them to specialized and non-specialized audiences in a clear and unambiguous way.

CB10 Learn skills that allow them to continue studying in a way that will be largely self-directed or autonomous.

CG1 Ability to apply the techniques of analysis and representation of information, in order to adapt it to real problems.

CG2 Ability to identify the most appropriate statistical model for each real problem and know how to apply it for the analysis, design and solution of it.

CG3 Ability to obtain scientifically viable solutions for complex real statistical problems, both individually and as a team.

CG4 Ability to synthesize the conclusions obtained from these analyzes and present them clearly and convincingly in a bilingual environment (Spanish and English) both in writing and orally.

CG5 Being able to generate new ideas (creativity) and anticipate new situations, in the contexts of data analysis and decision making.

CG6 Apply social skills for teamwork and to relate to others autonomously.

CG7 Apply the advanced techniques of analysis and representation of information, in order to adapt it to real problems.

CE1 Apply in the development of methods of analysis of real problems, advanced knowledge of statistical inference.

CE2 Use free software such as R and Python for the implementation of statistical analysis.

CE3 Predict and represent possible associations between random phenomena, related to real problems and reflected in the collected data, applying concepts of multivariate analysis

CE9 To correctly identify the type of statistical analysis corresponding to certain objectives and data.

CE10 Apply statistical modeling in the treatment of relevant problems in the scientific field.

CE12 Apply models for unsupervised learning.

Learning results

1. Principal components analysis.
2. Factor analysis.
3. Correspondence analysis.
4. Multidimensional scaling.
5. Cluster analysis.

DESCRIPTION OF CONTENTS: PROGRAMME

1. Multidimensional data sets
   1.1 Multidimensional data sets
   1.2 Visualizing multidimensional data sets
   1.3 A brief summary of multivariate random variables
   1.4 Inference with the data matrix
   1.5 Imputation of missing values
2. Principal component analysis
   2.1 Main goals of principal component analysis
   2.2 Standard principal component analysis
   2.3 Normalized principal component analysis
3. Cluster analysis
   3.1 Cluster analysis
   3.2 Partitional clustering
   3.3 Hierarchical clustering
   3.4 Model-based clustering
4. Factor analysis
   4.1 Latent variables
   4.2 The factor model
   4.3 Principal component factor analysis
   4.4 Principal factor analysis
   4.5 Maximum likelihood estimation
5. Multidimensional scaling
   5.1 Similarities and dissimilarities
   5.2 Multidimensional scaling
6. Correspondence analysis
   6.1 Two qualitative variables
   6.2 Testing for independency between two qualitative variables
   6.3 Correspondence analysis

LEARNING ACTIVITIES AND METHODOLOGY

Learning activities:

Theoretical classes
Practical classes
Tutorials
Team work
Individual work of the student
In-person evaluation tests

Methodology to be used:

Theoretical classes with support material available on the Web.
Problem solving classes. Computational practices in computer rooms. Oral exhibitions

Tutorial regime:

Individual tutorials throughout the course.

ASSESSMENT SYSTEM

Group project (50%)
Conceptual exercises (40%)
Oral presentations (10%)

% end-of-term-examination:	50
% of continuous assessment (assigments, laboratory, practicals…):	50

BASIC BIBLIOGRAPHY

- Alan Julian Izenman Modern Multivariate Statistical Techniques, Springer, 2008
- Richard A. Johnson and Dean W. Wichern Applied Multivariate Statistical Analysis, Pearson Education, 2007