

Academic Year: ( 2018 / 2019 )

Review date: 11-05-2018

Department assigned to the subject: Statistics Department

Coordinating teacher: MUÑOZ GARCIA, ALBERTO

Type: Electives ECTS Credits : 6.0

Year : 4 Semester :

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

This course assumes that the student knows the contents of

- a) Statistics I ([http://www3.uc3m.es/reina/Fichas/Idioma\\_2/204.13154.html](http://www3.uc3m.es/reina/Fichas/Idioma_2/204.13154.html)),
- b) Statistics II ([http://www3.uc3m.es/reina/Fichas/Idioma\\_2/204.13160.html](http://www3.uc3m.es/reina/Fichas/Idioma_2/204.13160.html)),

and the lesson of Properties of Matrices in

- c) Mathematics for Economics II ([http://www3.uc3m.es/reina/Fichas/Idioma\\_2/204.13156.html](http://www3.uc3m.es/reina/Fichas/Idioma_2/204.13156.html))

in the Business Administration degree.

Some notions in Multivariate Statistics

**OBJECTIVES****SPECIFIC COMPETENCES:**

- 1. To know and use advanced statistical techniques, with last generation software support.
- 2. To extract and analyze information from large data sets.
- 3. Learning the basic Statistical skills for the analysis of multivariate socio-economical data such as those coming from a market research.
- 4. Being able to describe and analyze real data sets using the techniques mentioned above.
- 5. Being able to elaborate reports with the results of the analysis of real case studies.

**CROSS COMPETENCES**

- 1. Information analysis and synthesis capacity on data mining problems.
- 2. Solving real problems.
- 3. Learning and training in the use of Statistical software to solve real case studies.
- 4. Critical and selective reasoning to solve real life problems.
- 5. Presentation abilities.

**DESCRIPTION OF CONTENTS: PROGRAMME**

- 1. Learning the R Statistical Language.
  - 1.1 Basic commands.
  - 1.2 Graphics in R.
  - 1.3 Statistical functions in R and basic programming.
- 2. Visualization Techniques for complex business data.
  - 2.1 Principal component analysis theory.
  - 2.2 Basic examples with R code.
  - 2.3 Case studies.
- 3. Distances in Data Mining.
  - 3.1 Main ideas and different distances in data mining.
  - 3.2 Programming distances in data mining.
- 4. Multidimensional Scaling.
  - 4.1 Metric scaling theory.
  - 4.2 Examples with R code.
  - 4.3 Perceptual mappings in R.
- 5. Cluster Analysis.
  - 5.1 Hierarchical methods.
  - 5.2 Centroid methods: k-means.
  - 5.3 Case studies.
- 6. Association Rules.

- 6.1 Main concepts and algorithms.
- 6.2 Complete example with R code.
- 6.3 Case studies.
- 7. Classification Trees.
  - 7.1 Information theory.
  - 7.2 Classification trees algorithms.
  - 7.3 Real case: credit scoring.
- 8. Real Case Studies.
  - 8.1 Comprehensive real cases involving all the studied techniques.

#### LEARNING ACTIVITIES AND METHODOLOGY

- 1. Theoretical lectures (4 ECTS)
- 2. Computer labs (2 ECTS)
- 3. Final project.

#### ASSESSMENT SYSTEM

- 50%: Final exam.
- 10%: Continuous evaluation (\*).
- 40%: Handing a final project.

(\*) Continuous evaluation consists of handing several case studies (homework) along the course.

<b>% end-of-term-examination:</b>	60
<b>% of continuous assessment (assignments, laboratory, practicals...):</b>	40

#### BASIC BIBLIOGRAPHY

- Avril Coghlan A little book of R for multivariate analysis, Internet, 2014
- Johannes Ledolter Data Mining and Business Analytics with R, Wiley, 2013

#### ADDITIONAL BIBLIOGRAPHY

- Y Zhao R and Data Mining. Examples and Case Studies, Elsevier, 2012