COURSE DATA

DATA SUBJECT

Code: 36502
Name: Exploratory Data Analysis and Databases
Cycle: Undergraduate Studies
ECTS Credits: 6
Academic year: 2025-26

STUDY (S)

Degree	Center	Acad. year	Period
1332 - Degree in Business Intelligence and Analytics	Facultat d'Economia	1	First quarter

SUBJECT-MATTER

Degree	Subject-matter	Character
1332 - Degree in Business Intelligence and Analytics	Fundamentos del Análisis de Datos	BASIC

COORDINATION

COLL SERRANO VICENTE

SUMMARY

Exploratory analysis of data and databases is a basic training subject assigned to the area of Quantitative Methods for Economics and Business that is taught in the first term of the first year of the Degree in Business Intelligence and Analytics with a total teaching load of 6 ECTS credits.

Exploratory data analysis (E.D.A, Exploratory data analysis), developed by John W. Tukey, is the first stage of the data analysis process and consists of a set of statistical methods and graphs that help to:

organize and structure the data.

explore the distribution of the variables considered with a view to, for example, detecting the presence of missing values, anomalous values, understanding the dispersion and shape of the data, etc.

understand the relationships between the variables

check compliance with the assumptions on which a large part of multivariate methods are based.

reorganize and restructure the data for subsequent analysis procedures

To train the student in the exploratory analysis of data, the subject studies the basic concepts of descriptive statistics, both one-dimensional and multidimensional, and uses real databases with statistical analysis software R program and RStudio.

PREVIOUS KNOWLEDGE

RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

There are no specified enrollment restrictions with other subjects of the curriculum.

OTHER REQUIREMENTS

COMPETENCES / LEARNING OUTCOMES

-

Apply methods and techniques of analysis, synthesis and graphical representation by means of software tools.

Be able to access and manage information in different formats for subsequent analysis in order to obtain knowledge through data.

Be able to analyse and search for information from diverse sources.

Be able to use ICT, both in academia and in professional practice.

Demonstrate skills for analysis and synthesis.

Know the different types of data.

Reorganise and restructure variables and databases.

Students must have the ability to gather and interpret relevant data (usually in their field of study) to make judgements that take relevant social, scientific or ethical issues into consideration.

DESCRIPTION OF CONTENTS

1. Topic 1. Introduction to Exploratory Analysis of data

2. Topic 2. R for the statistical analysis of data. Introduction.

3. Topic 3. Access, cleaning and debugging of data

4. Topic 4. Introduction to Data Visualization.

5. Topic 5. Transforming data.

6. Topic 6. Analysis of multidimensional data.

7. Topic 7. Data wrangling: handling/managing data.

WORKLOAD

PRESENCIAL ACTIVITIES

Activity	Hours
Theory	15,00
Computer classroom practice	45,00
Total hours	**60,00**

NON PRESENCIAL ACTIVITIES

Activity	Hours
Attendance at other activities	0,00
Individual or group project	10,00
Independent study and work	26,00
Preparation of lessons	30,00
Preparation for assessment activities	14,00
Resolution of case studies	10,00
Total hours	**90,00**

TEACHING METHODOLOGY

The development of the subject is structured fundamentally around theoretical sessions and practical sessions. Depending on the type of session (theoretical or practical) a didactic or other teaching method will be chosen.

In the theoretical sessions, which last 1 hour, the contents will be presented in a theoretical and practical way, corresponding to the basics of descriptive statistics and exploratory data analysis.

The predominant teaching method in the theoretical classes will be the participatory master class. This methodology is effective with large groups of students, offering the advantages of a master class without limiting the participation of the students and the teacher-student interaction. It aims to encourage participation and discussion in the class in order to offer the student a direct involvement with the content.

In the practical sessions, which last 3 hours, the teacher will guide the student in the management of R and RStudio and its application to the exploratory analysis of data through the use of real databases. In order to consolidate the knowledge acquired in the subject, students will be asked to solve similar practical cases, but not totally identical, to those dealt with in class. Given the nature of the practical sessions, the weekly monitoring of the contents studied by the students is essential.

The practical sessions will be developed following different teaching strategies depending on the contents discussed in the theoretical and practical sessions.

EVALUATION

Evaluation of the subject will be completed taking into account continuous assessment and a final exam at the end of the term.

The final exam, not necessarily written, will consist of theoretical-practical questions that will assess whether the student has assimilated the key concepts of the program. This exam will be 70% of the final mark.

Continuous evaluation aims to develop the skills of students and stimulate daily work. Therefore, this part of the evaluation will be based on an assessment of the consistent participation in the subject. The continuous assessment will represent 30% of the final mark. By its very nature, tasks and/or continuous assessment activities are NOT RECOVERABLE.

The final mark will be the weighted sum of the final exam and continuous assessment. In the event that the exam is not passed (3.5 points out of 7, or alternatively 5 out of 10 points), the final mark cannot exceed the maximum of 4.5 points.

The student who does not participate in continuous assessment may be evaluated based on the final exam and may obtain a maximum final mark of 7 points. To pass the subject the student will have to obtain a minimum of 5 points out of 7 (or alternatively, 7 points out of 10) in the final exam.

REFERENCES

- Baumer, B.S.; Kaplan, D.T. y Horton, N.J. (2017) Modern Data Science with R. CRC Press.
- Esteban, J. y otros (2006). Estadística Descriptiva y nociones de probabilidad. Paraninfo, Madrid.
- Murgui, S. y otros (2002). Ejercicios de Estadística. Economía y Ciencias Sociales. Valencia: Tirant lo Blanch.
- Peng, R.D. (2015). Exploratory Data Analysis with R. Lean publishing. (bookdown: https://bookdown.org/rdpeng/exdata/)
- Teetor, P. (2011). R Cookbook. OREILLY
- Wichkam, H. y Grolemund, G. (2016). R for Data Science. OREILLY.
- Newbold, P. y otros (2008). Estadística para Administración y Economía. Madrid Pearson-Prentice Hall, (6ª Edición).
- Pearson, R. K. (2018). Exploratory Data Analysis using R. Chapman and Hall/CRC.
- Coll Serrano, V. (2024). Introducción al Análisis Exploratorio de Datos. Aplicaciones con R y datos reales. Leanpub. https://leanpub.com/analisis_exploratorio_datos_con_R