UAB
Universitat Autònoma de Barcelona

Exploratory Data Analysis

Code: 104853
ECTS Credits: 6

2025/2026

Degree	Type	Year
Applied Statistics	FB	1

Contact

Name: Maria Rosa Camps Camprubi

Email: rosa.camps@uab.cat

Teachers

Montserrat Ferre Delgado

Teaching groups languages

You can view this information at the end of this document.

Prerequisites

None.

Objectives and Contextualisation

Learn the descriptive and exploratory techniques applied to summarize the information contained in theexperimental datasets. Also the interpretation of the results and the diagrams in the context of the data,

Finally it is important that the students learn to use statistical software to manipulate data, perform descriptive analysis and graphs.

Learning Outcomes

1. CM05 (Competence) Interpret situations based on sets of data, graphic representations and statistical summaries.
2. CM06 (Competence) Apply the knowledge acquired to organise data, create and show tables and work with different data representations.
3. KM07 (Knowledge) Identify behaviour patterns in data exploration.
4. SM08 (Skill) Analyse survey results.

Content

1. Preliminaries.

1.1. The goal of Exploratory Data Analysis.
1.2. Types of variables and measurement scales.
1.3. Rounding and scientific notation.

2. Summary of statistical data.

2.1. Frequency distributions: tables.
2.2. Grouping data into intervals.
2.3. Graphical representation.

3. Numerical measures of a variable.

3.1. Central position measures: mean, median, mode.
3.2. Other position measures: quartiles, deciles and percentiles.
3.3. Measures of dispersion: variance and standard deviation (sample and population), range, interquartilerange.
3.4. Measures of relative dispersion
3.5. Standard scores.
3.6. Measures of form: symmetry and curtosis

4. Extra tools for the study of a variable.

4.1. Exploratory analysis: boxplot and other diagrams.
4.2. Transformation of variables.
4.3. Other means: geometric, harmonic, quadratic.
4.4. Chebyshev's inequality.

5. Comparison of a variable in two or more groups: Exploratory analysis.

5.1. Situation of independent samples.
5.2. Situation of paired samples.

6. Tabulation and representation of the joint distribution of two categorical variables.

6.1. Contingency tables (joint, marginal and conditional frequency distributions).
6.2. Descriptive analysis of the dependence between two categorical variables.

7. Numeric description of the joint distribution of two statistical variables.

7.1. Marginal and conditionañ measures.
7.2. Regression curves and correlation coefficient.
7.3. Linear fitting and prediction.

8. Introduction to time series.

8.1. The classical decomposition.
8.2. Smoothing series: application of filters.

*Unless the requirements enforced by the health authorities demand a prioritization or reduction of these contents.dexes.

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Computer lab	30	1.2	
Lectures	18	0.72	
Problem sessions	8	0.32	
Studying theoretical concepts, solving problems by hand and using R	84	3.36	

Classroom work, theory and problems will be complemented by computer practices where the R-band packages will be used.

In the Moodle space of the course, students will find the subject's planning, the sets of exercices and computer lab sessions, as well as possible changes of classrooms, schedules, etc. It is important to keep in mind that CampusVirtual is not a static website
but will be updated throughout the course.

In the most practical part of the course, if possible and through the analysis and comparison of statistical data, we will comment on the causes and the social and cultural mechanisms that can sustain the observed inequalities.

* The proposed teaching methodology may experience some modifications depending on the restrictions to face-to-face activities enforced by health authorities.

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
End of term written exam	25%	2	0.08	
First computer lab exam	15%	2	0.08	
Mid term written exam	25%	2	0.08	CM05, KM07, SM08
Qualification of practical classes with computer	10%	2	0.08	CM05, CM06
Second computer lab exam	25%	2	0.08	

The final grade of the subject F will be obtained from:

1) The notes of the two partial exams of theory and problems, TP1 and TP2, with respective weights 25% and 25%.

2) The notes of the two computer tests, O1 and O2, with respective weights 15% and 25%.

3) In the weekly practical sessions, deliveries will be proposed. Then, every two or three weeks, they will be evaluated with small questionnaires, giving the mark PC, with a total weight of 10%. This part is not recoverable. In case of doubts in the final grade, the deliveries made by the student throughout the course will be reviewed.

The final grade of the subject is obtained by making the weighted average F = 0.25 TP1 + 0.15 O1 + 0.25 TP2 + 0.25 O2 + 0.10 PC.

A requirement to pass the subject with the previous formula is that the marks TP1, TP2 and O2 must be greater than or equal to 4 and O1 must be greater or equal to 3.5.

In case F<5 or the requirement (marks O2, TP1 and TP2 >=4 and mark O1 >= 3.5) is not satisfied the students will have the opportunity to take a resit test,

There will be two resit tests:

- STP a global exam oftheory and problems teoria i problemes, for students with marks of TP1 or TP2 less than 4 or whose bad grades in these exams cause F<5.

- SO a global exam with computer, for students with marks O1 less than 3.5 or O2 less than 4, or whose bad grades in these exams give F<5.

the final grade will be F=0,50 STP + 0,40 SO + 0,10 PC

(if only one of the tests are taken (STP or SO) then the other grade will be taken from the weighted mean of the passed partial exams grades).

Students who do not attend exam will get the qualification of "Not Evaluable".

After the second partial tests the honors qualitfications will be considered and might be given,even before the resit exam.

"Without prejudice to other disciplinary measures deemed appropriate, and in accordance withcurrent Academic regulations the irregularities committed by the student that may lead to a variation of the grade of an evaluation act will be scored with a zero. so, plagiarizing, copying or letting copy a practice or any other evaluation activity will imply suspending with a zero and cannot be recovered in the academic mateixcurs.If this activity has a minimum associated score, then the subject will be suspended. "

The use of artificial intelligence (AI) is strongly discouraged when it is done by non-expert people in the subject that is consulted, as is the case of the students of the course. In this course, the use of Artificial Intelligence (AI) technologies is not allowed to generate the deliveries of practices. Therefore, any work that includes fragments generated with AI will be considered a lack of academic honesty and may entail a partial or total penalty in the PC grade. The suspicions will be clarified with an interview with the teachers. Obviously, the AI consultation during the exams will be considered as a copy and will imply to suspend the subject with a zero.

Students who choose the single assessment modality will have to take two global exams for the subject: an exam on theory and problems and a computer test. These two exams will be held the same day, time and place as those corresponding to the second part of the subject for ordinary students. On that day, the student can hand in the computer exercise assignments that have been scheduled throughout the course, which may be found on Moodle. The weighting for the final grade will be 50% for theory and problemes exam, 40% for computer exame and 10% of the assignments. The student may subsequently be summoned to an oral review of his exams and assignments with the professors of the course. If a grade of less than 5 is obtained, the student cando the resit tests with the rest of the students (same day, same hour, same place). And the weighting will be the same.

Bibliography

Course lecture notes

X. BARDINA, M. FARRÉ, Estadística descriptiva, Manuals, 54 Servei de Publicacions, UAB

Bibliography

A.J.B. ANDERSON, Interpreting Data. A first cours in Statistics, Ed Chapman and Hall, 1989.
R Tutorial. An R introduction to statistics. (2016). www.r-tutor.com
E. CASA ARUTA, Problemas de Estadística Descriptiva, Ed. Vicens Vives.
R. JOHNSON, P. KUBY, Estadística elemental: Lo esencial, Ed Thomson, 1999.
B. PY, Statistique Descriptive, Ed Económica, 1988.
M. SPIEGEL, Estadística, Teoría y 875 problemas resueltos, Schaum-McGraw-Hill, 1990.
V. ZAIATS, M.L. CALLE i R. PRESAS, Probabilitat i Estadística. Exercicis I, Eumo Ed, 1998.
H. WICKHAM, M. CETINKAYA-RUNDEL, G. GROLEMUND, R for Data Science. https://r4ds.had.co.nz/

Complementary Bibliography

G. CALOT, Curso de Estadística Descriptiva. Ed Paraninfo, 1988.
FERNÁNDEZ, J.M. CORDERO, A. C\'ORDOBA, Estadística Descriptiva, ed ESIC 1996.
L.C HAMMILTON, Modern Data Analysis, Brooks/Cole Publishing Company, 1990.
P.G. HOEL i R.J. JESSEN, Estadística básica para negocios y economía, Compañía Editorial Continental,Mexico, 1993.
R.K. PEARSON, Exploratory Data Analysis using R. Data Mining and Knowledge Discovery Series, Chapman & Hall/CRC, 2018.
D. PEÑA SÁNCHEZ DE RIVERA, Estadística. Modelos y métodos. 1. Fundamentos i 2. Modelos lineales yseries temporales, Alianza Editorial 1995. (2 volums)

Software

R and RStudio

Groups and Languages

Please note that this information is provisional until 30 November 2025. You can check it through this link. To consult the language you will need to enter the CODE of the subject.

Name	Group	Language	Semester	Turn
(PLAB) Practical laboratories	1	Catalan	first semester	afternoon
(PLAB) Practical laboratories	2	Catalan	first semester	afternoon
(TE) Theory	1	Catalan	first semester	afternoon