

## Data Analysis

Code: 100452  
ECTS Credits: 6

**2025/2026**

Degree	Type	Year
Criminology	OB	2

## Contact

Name: Marc Ajenjo Cosp

Email: marc.ajenjo@uab.cat

## Teaching groups languages

You can view this information at the [end](#) of this document.

## Prerequisites

The teaching of the subject will be taught taking into account the perspective of the Sustainable Development Goals.

It's recommended to have passed the course of quantitative methods and basically knowledge of *RStudio*. Those students who are not taking the Quantitative Methods course this year; it is recommended that they evaluate their knowledge of RStudio and that they do not reject the possibility of signing up for an introductory course in RStudio.

Teaching will be taught in Catalan. Despite this, it is possible that some of the seminars will be taught in Spanish.

## Objectives and Contextualisation

To use specific criminological methods and research techniques for analysing the data and experiences of conflict and crime control that exists on a particular social context.

In this context, the objectives of the course are:

- Understanding and consolidating the concepts of statistical inference.
- Introducing multivariate analysis techniques for both the primary and secondary data analysis.
- Applying these concepts to criminological research.
- Consolidating and going in depth into the use of data analysis tools applied to quantitative criminology.

## Competences

- Ability to analyse and summarise.
- Accessing and interpreting sources of crime data.
- Applying the quantitative and qualitative data collection techniques in the criminological field.
- Clearly explaining and arguing a carried out analysis about a conflict or crime problem and its responses in front of specialised and non-specialised audiences.

- Designing a criminological research and identifying the appropriate methodological strategy to the proposed goals.
- Drawing up an academic text.
- Working autonomously.

## Learning Outcomes

1. Ability to analyse and summarise.
2. Applying the quantitative and qualitative data collection techniques in the criminological field.
3. Choosing the appropriate research methodology in criminological works.
4. Drawing up an academic text.
5. Interpreting in a scientific way statistical data from the criminological field.
6. Transmitting in a reasoned manner the results of a criminological research.
7. Working autonomously.

## Content

The subject is structured in two parts.

First, and as a continuation of the previous course, *Quantitative Methods*, we revise the inference technique's introduction and go in depth into some of the techniques most used at criminology research. It's important the knowledge of statistical data analysis packages.

Second, we give an overview on the data treatment done when there is a significant number of variables, giving special weight to the logistic regression, using computer tools as a support.

### PART I. Bivariate inference applied to criminology

#### 1. Introduction to statistical inference: hypothesis testing

- 1.1. Descriptive statistics versus inferential statistics. The statistical tests in solving problems posed in the field of criminology
- 1.2. The approach of hypothesis testing. The null hypothesis and the alternative hypothesis. Significant differences and no significant differences
- 1.3. Test hypothesis errors. Type I error (significance level and confidence level) and Type II error (power of a test)
- 1.4. Resolution of hypothesis testing. Steps when solving hypothesis testing

#### 2. Hypothesis testing based on proportions

- 2.1. Goodness of fit tests for qualitative variables. The confidence interval compared to a ratio of observed and theoretical
- 2.2. Comparison of proportions with independent data. The contingency table. The chi-square test and some statistical coefficients: Cramer V

#### 3. Hypothesis testing based on averages or other measures of central tendency

- 3.1. Parametric and nonparametric statistical tests. The importance of application conditions when the sample size is small
- 3.2. T-test to compare theoretical and observed averages
- 3.3. T-test to compare two matched means and two independent means. The nonparametric tests

3.4. Variance analysis to compare more than two independent means. Post hoc tests. The corresponding nonparametric tests (*Kruskal-Wallis*)

#### 4. Inferential statistics in the regression

4.1. Regression line inferential level. The conditions of the model

4.2. Tests on the parameters of the line and on the coefficient of determination. Results interpretation

#### 5. Data analysis and inference based on bivariate statistical packages

5.1. Proportions comparisons. The goodness of fit tests. The chi-square test and related statistical coefficients

5.2. Comparing means. Tests parametric and nonparametric. The *Shapiro-Wilk test* to assess normality. Comparing theoretical and observed means. Paired comparison of two means. Comparison of two or more independent means

5.3. The linear regression

### PART II. Introduction to multivariate analysis. The Logistic regression

#### 6. Logistic regression

6.1. Conceptual introduction. Logistic regression and models Loglinear as a variant. The logit, odds and odds ratio

6.2. Bivariate logistic regression

6.3. The importance of the control of a third variable. Simpson's paradox

6.4. To introduce multiple variables in the regression. The selection of variables and the goodness of fit of the model

#### 7. Logistic regression on statistical packages

7.1. Logistic regression with one independent variable

7.2. Introducing a second variable. Multivariate logistic regression

7.3. The development of logistic regression models. The different methods of selection of variables and statistical goodness of fit

## Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Lectures	19.5	0.78	2, 3, 1, 6
Practical class	19.5	0.78	2, 3, 5, 1, 6, 7
Type: Supervised			
Group work preparation and development	41	1.64	2, 3, 5, 4, 1, 6
Type: Autonomous			

Mock test. Reading, understanding and synthesis of materials	60	2.4	2, 3, 5, 4, 1, 6, 7
Test	10	0.4	2, 3, 5, 1, 6, 7

Before the start of the course, a detailed schedule of sessions will be published on the virtual campus.

Two types of activities will be held in the classroom:

- Lecturers for the entire group. The concepts and content of the course will be presented using PowerPoint presentations. The final 10 minutes of each session will be reserved for a test on the specific content of the session. This test will serve as an attendance check, but also to assess student adherence to the course.
- Practical sessions in computerized classrooms. The students will work with the free software RStudio, which is installed in all social science computerized classrooms (students are encouraged to also have the software installed on their laptops). A dossier will be provided to students in each session through the virtual campus to facilitate follow-up. As with the theoretical sessions, the final 10 minutes of each session will be reserved for a test on the specific content of the session. This test will serve as an attendance check, but also to assess student progress in the course.

Outside the classroom

- Two types of exercises must be completed weekly, one for the theoretical session and one for the practical session. These exercises must be submitted via the virtual campus before the next session:
  - Theoretical session. This will consist of problems in applied statistics in criminology. The exercises will be solved at the beginning of the next session.
  - Practical session. Work exercises with RStudio software to gain autonomy in its use. The solutions to these exercises will be posted on the virtual campus for students to self-correct.
- A research project will be carried out, in which the basic concepts of logistic regression will be applied. This project will require supervision by the instructor, which will require mandatory tutorials for all group members.

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

### Continuous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Active monitoring of the sessions (Part I of the program)	10%	0	0	5, 1, 7
Individual written test (Part I of the program)	50%	0	0	2, 3, 5, 1, 7
Research work in criminology (Part II of the program)	40%	0	0	2, 5, 4, 1, 6

#### 1. Continuous assessment Model

The continuous assessment involves the active participation of the students and includes regular attendance at all sessions, as well as the delivery of weekly exercises. At the end of each session, a questionnaire will be administered with 10 short questions about the content explained in that session. Without adequate follow-up of the classes, which includes both attendance and the delivery of the weekly exercises (80% of the evidence),

the student will not be evaluated. The mandatory attendance excludes cases of illness or absence due to force majeure. On the other hand, the non-delivery of the weekly proposed exercises cannot be justified.

After the middle of the course, an assessment will be conducted to show if the student has achieved the minimum knowledge to follow the course. A test assessing knowledge of Part I of the program (Bivariate inference applied at criminology) will be performed. Student's achievement is a prerequisite to continue with the last part of the course. Students who initially do not have successfully passed this assessment will conduct an extra support class in order to achieve the skills for repeating the assessment. The students who do not reach the minimum required, will have to take a final test that will include the content of the entire course.

Part II (Introduction to multivariate analysis) is assessed through a research project which will demonstrate the concept and logic a mastery of logic of logistic regression. It will be about developing teamwork that, once completed, and within a week, will require group tutoring where the content of the work must be individually defended. In this tutorial, if necessary, the bases will be established to be able to correct the most relevant deficiencies in the work. In this sense, if students wish to correct their work they must modify and deliver it a week later. To be evaluated students must have followed the logistic regression classes (100%).

To access the calculation of the final mark students are required to have passed the individual test, as well as group work. Therefore, it is contemplated that the failed activities can be reassessed during the course.

Students who have not passed either part will have the right to a single final exam. This right is only contemplated for those who have a minimum attendance of 80%.

## 2. Single assessment Model

Students who choose to take a single assessment will do so based on a final test where they must demonstrate that they have acquired all the skills of the subject. Although the exam content will be eminently practical, there will be a theory section corresponding to Part I of the program. If the student doesn't pass the examen, they will have the right to a make-up test.

In broad strokes, the logic of the single evaluation will be the same as the continuous evaluations: 60% will correspond to Part I and 40% to Part II.

To prepare for the final test, it is encouraged to use all the didactic materials for the subject available on the virtual campus.

To pass the subject, a minimum grade of 5 is required in the exam as a whole.

## 3. Non-assessable grade

Students will be assessed as long as they have completed a set of activities whose weight is equivalent to a minimum of 2/3 of the total grade of the subject. If the value of the activities carried out does not reach this threshold, the teacher of the subject can consider the student as not evaluable.

## 4. Fraudulent conduct

A student that cheat or attempt to cheat in the exam will get a 0, losing the right to a second chance.

In the specific case of the essay, signs of plagiarism will mean suspending the coursework. Likewise, those who cannot justify the arguments developed in the essay will have a mark of 0. Human or technological help in writing the results of the work will also be considered plagiarism.

## 5. Attitudes during the course

The UAB has a diverse and inclusive environment for students, teachers and the entire university community. In this class, a policy of zero tolerance will be applied towards any attitude of discrimination or harassment based on age, ancestry, functional diversity, gender identity, national origin, religious beliefs or sexual orientation, as well as no tolerance for attitudes that generate a hostile climate for any of the reasons cited. These attitudes will be reported, following the University's harassment prevention policy.

## 6. Punctuality

Punctuality in class is required. Any unexcused tardiness of more than 5 minutes is considered an absence.

## Bibliography

For the whole of the subject:

"Material bàsic i complementari de seguiment de les classes" available at the Virtual Campus

"Tutorials pas a pas, i exercicis (amb solucions)" available at the virtual campus

Specific readings Part I:

- Fox, J. A., Levin, J. A., & Forde, D. R. (2013). *Elementary Statistics in Criminal Justice Research* (3a ed.). Pearson Education
- López-Roldán, P., & Fachelli, S. (2015). *Metodología de la Investigación Social Cuantitativa* (1a ed.). Universitat Autònoma de Barcelona. <http://ddd.uab.cat/record/129382>
- Sánchez Carrión, J. J. (1999). *Manual de análisis de datos*. Alianza Universidad Textos

Specific readings Part II:

- Cea D'Ancona, M. Á. (2002). *Análisis multivariable. Teoría y práctica en la investigación social*. Editorial Síntesis
- Etxeberria, J. (2007). *Regresión múltiple*. Editorial La Muralla
- Guillén, M. F. (2014). *Análisis de regresión múltiple*. Centro de Investigaciones Sociológicas, Cuadernos Metodológicos 4
- Jovell, A. J. (1995). *Análisis de regresión logística*. Centro de Investigaciones Sociológicas, Cuadernos Metodológicos 15
- Lozares Colina, C., & López-Roldán, P. (1991). El análisis multivariado: definición, criterios y clasificación. *Papers, Revista de Sociología*, 37, 9-29

Specific readings on software tools for data processing:

- López-Roldán, P., & Fachelli, S. (2015). *Metodología de la Investigación Social Cuantitativa* (1a ed.). Universitat Autònoma de Barcelona. <http://ddd.uab.cat/record/129382>

Note

Materials and bibliography of the different parts of the program will be available on the Virtual Campus bibliography.

Given the eminently practical nature of the course readings that appear in these references are not compulsory, but to consult for complementing the classes explanations and clarify any queries that arise in the same explanation. They can be very useful for those students for some reason someday cannot attend classes.

## Software

The free software *RStudio* will be used

## 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
(SEM30) Seminaris (30 estudiants per grup)	11	Catalan	second semester	morning-mixed
(SEM30) Seminaris (30 estudiants per grup)	12	Catalan	second semester	morning-mixed
(SEM30) Seminaris (30 estudiants per grup)	13	Spanish	second semester	morning-mixed
(TE) Theory	1	Catalan	second semester	morning-mixed