

Data Analysis

Code: 100452
ECTS Credits: 6

Degree	Type	Year	Semester
2500257 Criminology	OB	2	2

Contact

Name: Marc Ajenjo Cosp
Email: marc.ajenjo@uab.cat

Use of Languages

Principal working language: catalan (cat)
Some groups entirely in English: No
Some groups entirely in Catalan: Yes
Some groups entirely in Spanish: No

Other comments on languages

International students, if there are will be offered the possibility to follow the subject in spanish with personal tutoring.

Prerequisites

It's recommended to have passed the course of quantitative methods and basically knowledge of *RStudio*.

Objectives and Contextualisation

To use specific criminological methods and research techniques for analyzing 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 techniques 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 *Kolmogorov-Smirnov* 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.

Methodology

Part of the activities is carried out with the teacher's support:

1. Lectures that are designed to introduce the main concepts and content of the course.
2. Following the presentation of content there will be several practical sessions. These are aimed at:
 - a) to solve simple cases without computer support and
 - b) to solve complex cases by using appropriate software (RStudio).

In order to consolidate the knowledge, these activities should be complemented outside the class by:

1. At the end of each session, a case study will be given to students in order to solve it at home. At the beginning of the next session, students will deliver their answers and the teacher will reveal the solution.
2. Sessions with statistical packages follow the same logic. In this case, however, students must upload solved exercises, and the solution will be sent. Allowing self-evaluation by students.

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.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			

Lectures	18	0.72	2, 3, 1, 6
Practical class	18	0.72	2, 3, 5, 1, 6, 7
Type: Autonomous			
Group work preparation and development	41	1.64	2, 3, 5, 4, 1, 6
Mock test. Reading, understanding and synthesis of materials	63	2.52	2, 3, 5, 4, 1, 6, 7
Test	10	0.4	2, 3, 5, 1, 6, 7

Assessment

1. Assessment Model

The course involves active participation of the student and includes regular attendance at all sessions. For this reason, at the end of each session a brief questionnaire with 10 questions about the content explained will be passed. Without an adequate attendance of the classes (80% of sessions) the student will not be evaluated.

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. A group work must be performed, a tutorial to the teacher should be done within a week from the deliver in order to correct the most important shortcomings. 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%).

The students who do not pass the first part of the course will be entitled to a final test that will include the content of the entire course.

2. Justified absences

Assistance is mandatory excepts in cases of illness or absence due to force majeure.

3. Conditions to pass the course and re-assessment

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

As an exception, the students who have not passed the first part of the course will be entitled to take a final test. Only the students with a minimum attendance of 80% will have the right to take this test.

4. Fraudulent conduct

A student that cheat or attemptto cheat in the exam will get a 0, losing the right to a second chance. Plagiarism will conduct to a fail of the essay.

5. Punctuality

Classes start on time. Late arrival is not admitted. Students are not allowed to leave before the end of the class.

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

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, James Alan, Levin, Jack A., & Forde, David R. (2013). *Elementary Statistics in Criminal Justice Research* (3a ed.). Pearson Education.

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

Sánchez Carrión, Juan Javier (1999). *Manual de análisis de datos*. Alianza Universidad Textos.

Specific readings Part II:

Cea D'Ancona, María Ángeles (2002). *Análisis multivariable. Teoría y práctica en la investigación social*. Editorial Síntesis.

Etxeberria, Juan (2007). *Regresión múltiple*. Editorial La Muralla / Hespérides, Cuadernos de Estadística 4.

Guillén, Mauro F. (2014). *Análisis de regresión múltiple*. Madrid: Centro de Investigaciones Sociológicas, Cuadernos Metodológicos 4.

Jovell, Albert J. (1995). *Análisis de regresión logística*. Madrid: Centro de Investigaciones Sociológicas, Cuadernos Metodológicos 15.

Lozares Colina, Carlos & López-Roldán, Pedro (1991). El análisis multivariado: definición, criterios y clasificación. *Papers, Revista de Sociologia*, 37, 9-29.

Specific readings on software tools for data processing:

López-Roldán, Pedro, & Fachelli, Sandra (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 can not attend classes.

Software

The free software *RStudio* will be used