

Degree	Type	Year
Data Engineering	FB	2

## Contact

Name: Joan Porti Pique

Email: joan.porti@uab.cat

## Teachers

Joan Porti Pique

## Teaching groups languages

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

## Prerequisites

There are no prerequisites. It is recommended to have followed the courses in Algebra and Calculus.

## Objectives and Contextualisation

The goal of the course is to introduce the basic tools of probability and statistics to analyze data from natural, social or economic phenomena or experiments, focusing on its correct use and the interpretation of the results. The theory and problem sessions are going to be complemented with practice classes, with the aim that students learn to use computer tools for statistical analysis.

## Competences

- Design efficient algorithmic solutions to computational problems, implement them in the form of robust software developments which are structured and easy to maintain, and verify their validity.
- Develop critical thinking and reasoning and know how to communicate it effectively in both your own language and in English.
- Search, select and manage information and knowledge responsibly.
- Use techniques of probability and statistics to analyse and model complex phenomena and solve optimisation problems.

- Work cooperatively in complex and uncertain environments and with limited resources in a multidisciplinary context, assuming and respecting the role of the different members of the group.

## Learning Outcomes

1. Acquire and consolidate the basic concepts of probability theory, mathematical expectation and conditional probability.
2. Choose and apply the most suitable data structures and strategies to solve an informatics problem efficiently.
3. Develop critical thinking and reasoning and know how to communicate it effectively in both your own language and in English.
4. Develop programmes that are well documented, using a good programming style and be able to debug, test and correct them.
5. Identify the most common descriptors for a data set and assess their applicability to a known data set.
6. Identify the statistical distributions and their application to engineering problems.
7. Search, select and manage information and knowledge responsibly.
8. Work cooperatively in complex and uncertain environments and with limited resources in a multidisciplinary context, assuming and respecting the role of the different members of the group.

## Content

### Topic 1. Descriptive statistics.

Descriptive study in one variable: frequency distributions, graphic representations, numerical summaries (position, dispersion and shape measurements).

Descriptive study in two variables: correlation and regression line, tables of contingency.

### Topic 2. Probability.

Notion of probability and main properties. Conditional probability. Formula of total probabilities. Bayes Formula.

Independence of events.

Expectation and variance of a random variable.

Discrete random variables. Bernoulli, Binomial and Poisson distributions.

Continuous random variables. Normal and Exponential distributions. Approximation of the Binomial by the Normal distribution.

Independence of random variables.

Central limit theorem.

### Topic 3. Statistical inference.

Sample and population. Most frequent statistics.

Confidence intervals for the mean and for the variance of a Normal population and for the proportion.

Hypothesis test concept. Test for the mean and for the variance of a Normal population.

Test for the proportion. Test of comparison of means and variances for two Normal populations. Comparison test of proportions.

Chi-square test for the goodness of fit and independence.

## Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
-------	-------	------	-------------------

Type: Directed			
Practices in the computer room	12	0.48	1, 7, 4, 3, 5, 6, 2, 8
Problem sessions	12	0.48	1, 7, 4, 3, 5, 6, 2, 8
Theoretical classes / lectures	26	1.04	1, 7, 4, 3, 5, 6, 2, 8
Type: Supervised			
Tutoring and consultations	10	0.4	1, 7, 4, 3, 5, 6, 2, 8
Type: Autonomous			
Independent study and preparation	60	2.4	1, 7, 4, 3, 5, 6, 2, 8

The center of the learning process is the work of the student. The student learns working, being the mission of the teaching staff help him/her in this task by providing information or showing him/her the sources where one can get it and directing his/her steps in a way that the learning process can be carried out effectively. In line with these ideas, and in accordance with the aims of the subject, there are theory classes (lectures), problem sessions, and practices sessions. In these sessions and with individual work of the student, the specific skills are achieved.

New material will be mainly introduced in the lectures, but explanations must be complemented with the autonomous study and personal work of the student, with the help of the references and the material in the CV.

The problems sessions will be devoted to the oriented resolution of some proposed problems. Attention will be paid to correctness and rigor, as well as to vocabulary, mathematical expression and clarity in writing. There will be a partial test of theory and problems.

In the practice sessions, software for statistical analysis will be introduced. Descriptive and inferential methodologies are introduced. These tools will be used to solve problems and to do some assignments that students will have to hand in.

The Campus Virtual UAB is a key tool for following the subject: access to material, consultation of deadlines and monitoring of the pace of the course.

The lectures, in which mathematical models are discussed, with the problem sessions, in which several solutions to problems will be proposed, and with the individual work of the student allow to reach the transversal skills. ( T01.01, T01.02 and T01.03).

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
Parcial exams	70%	5	0.2	1, 7, 4, 3, 5, 6, 2, 8
Practices assignments	30%	20	0.8	1, 7, 4, 3, 5, 6, 2, 8

Continuous assessment of the theoretical and problem-solving components consists of two midterm exams, the first weighted at 30% and the second at 40%. These two midterms constitute the recoverable part of the course.

The assessment of computer-based practical work will count for 30% of the final grade. To evaluate this component, students must submit their work and complete a test at the end of each session. This part is not recoverable.

To be eligible for the resit exam, students must have previously been assessed in a set of activities that represent at least two-thirds of the total course grade.

The single assessment consists of one comprehensive exam covering the entire theoretical and problem-solving content of the course. The grade obtained in this final exam will account for 70% of the final mark. This exam will be held on the same date as the final continuous assessment test, according to the official calendar, and the same resit policy will apply. The assessment of practical work and assignments will follow the same procedure as in continuous assessment. The grade for this component will represent the remaining 30% of the final mark.

A weighted average minimum of 4 out of 10 is required in the midterm, final exam, or resit. A minimum grade of 4 out of 10 is also required for the average of the practical assignments. If these minimums are met, the final grade will be the weighted average. If not, the final grade will be the lower of the weighted average or 4.5 (all on a 10-point scale).

Students will be marked as Not Assessed if they have not completed activities that account for at least 50% of the course.

The distinction grade "Matrícula d'Honor" (Honours) will be awarded to the best-performing students who have achieved a final grade above 9 and who, in the opinion of the teaching staff, have met all the course objectives with excellence.

Use of AI technologies is strictly prohibited in this course, at any stage of the learning process. Any work that includes AI-generated content will be considered academic misconduct and may result in partial or total loss of marks for the activity, or more severe sanctions in serious cases.

The dates for continuous assessment activities and assignment submissions will be published on the virtual campus and may be subject to changes due to unforeseen circumstances. Any such changes will be communicated via the virtual campus, which is considered the official channel of communication between faculty and students.

Without prejudice to other disciplinary actions deemed appropriate, and in accordance with current academic regulations, any irregularities committed by a student that could affect the assessment of an activity will result in a grade of zero (0) for the corresponding part. This mark will not be recoverable and will result in failing the course, with a maximum possible final grade of 4.5 out of 10. Such irregularities include, but are not limited to, plagiarism, copying, or enabling others to copy. Possession of communication devices during in-person assessment activities, whether used or not, will also be considered a serious offense.

There is no differentiated treatment for repeating students.

## Bibliography

Bardina, X. Farré, M. *Estadística descriptiva*. Manuals UAB, 2009.

Besalú, M. Rovira C. *Probabilitats i estadística*. Publicacions i Edicions de la Universitat de Barcelona, 2013.

Delgado, R. *Probabilidad y Estadística para ciencias e ingenierías*. Delta, Publicaciones Universitarias. 2008.

Devore, J. L. *Probabilidad y estadística para ingeniería y ciencias*. Thomson, 2005.

Montgomery, D. C. Runger, G. C. *Probabilidad y estadística aplicadas a la ingeniería*. Limusa Wiley, 2002.

Walpole, R. Myers, R. H. Myers, S. L. *Probabilidad y estadística para ingenieros*. Prentice Hall, 1999.

## Software

In the computer practice sessions, the student will learn to use the free software R with the integrated development environment RStudio, in order to apply the statistical tools for the descriptive analysis of data sets and statistical inference.

## 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
(PAUL) Classroom practices	811	Catalan/Spanish	first semester	morning-mixed
(PAUL) Classroom practices	812	Catalan/Spanish	first semester	morning-mixed
(PLAB) Practical laboratories	811	Catalan/Spanish	first semester	morning-mixed
(PLAB) Practical laboratories	812	Catalan/Spanish	first semester	morning-mixed
(TE) Theory	81	Catalan/Spanish	first semester	morning-mixed