UAB
Universitat Autònoma de Barcelona

Biostatistics

Code: 101965
Credits: 6

2026/2027

Degree programme	Type	Course
Genetics	FB	2

Contact lecturer

Name : Marta Coronado Zamora

Email : marta.coronado@uab.cat

Teaching staff

David Castellano Esteve

Group languages

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

Prerequisites

Those required for admission to the degree program.

It is highly advisable to have passed the first-year Mathematics course in order to successfully follow this subject.

Objectives

Statistical tools are fundamental to research and data analysis in Genetics and Genomics. In this Biostatistics course, we will learn to describe, analyze, and interpret experimental data, as well as to communicate results rigorously and effectively. This course introduces the fundamental concepts of statistics, data handling with R, and basic data visualization techniques for representing information clearly and reproducibly.

The course objectives are:

- Understand the fundamental concepts of descriptive and inferential statistics.
- Develop the ability to apply appropriate statistical methods to answer biological questions, particularly in the field of Genetics and Genomics.
- Learn to formulate hypotheses, interpret the results of statistical analyses, and evaluate their limitations.
- Acquire basic competencies in using R for data handling, statistical analysis, and generating visualizations.
- Develop the ability to communicate statistical results clearly, rigorously, and effectively through tables and figures.
- Introduce the principles of experimental design and reproducibility in biological data analysis.
- Introduce other modern approaches to data analysis, such as Bayesian statistics and machine learning techniques, including Deep Learning, and understand their complementary role within genetic data analysis.

Learning outcomes

- CM07 (Validate biological hypotheses through the application of statistical models and the design of significance contrasts.) Validate biological hypotheses through the application of statistical models and the design of significance contrasts.
- CM08 (Design experiments and strategies for statistical analysis suitable for the resolution of genetic problems.) Design experiments and strategies for statistical analysis suitable for the resolution of genetic problems.
- CM09 (Communicate the results and findings derived from the statistical analysis in scientific and technical reports clearly and rigorously.) Communicate the results and findings derived from the statistical analysis in scientific and technical reports clearly and rigorously.
- KM05 (Identify descriptive and inferential statistical methods and computer tools for the analysis of biological data, integrating the detection of sex and gender biases.) Identify descriptive and inferential statistical methods and computer tools for the analysis of biological data, integrating the detection of sex and gender biases.
- SM05 (Use statistical software and computational resources for calculation and experimental design in the field of biomedicine.) Use statistical software and computational resources for calculation and experimental design in the field of biomedicine.
- SM06 (Analyse quantitative data by executing statistical procedures and programming for the resolution of biological problems.) Analyse quantitative data by executing statistical procedures and programming for the resolution of biological problems.

Contents

Block 1. Introduction to biostatistics and data visualization

- Why showing data correctly matters
- Descriptive statistics
- Introduction to R
- Data visualization with ggplot2

Block 2. Probability distributions and confidence intervals

- Probability distributions: binomial and Poisson
- Normal distribution and the central limit theorem
- Tools for visualizing distributions
- Estimation: standard error and confidence intervals

Block 3. Statistical inference and hypothesis testing

- Statistical reasoning and hypothesis generation
- Inference: the logic of hypothesis testing
- Resampling-based testing: permutation and bootstrap
- Type I and II errors, effect size and power
- One-sample, two-sample, and paired t-tests, and their nonparametric alternative: the Wilcoxon test
- One-way ANOVA, post-hoc tests, and ANOVA as a linear model
- Two-way ANOVA and its nonparametric version: the Kruskal-Wallis test
- Review and integrative problem sets

Block 4. Linear models and relationships between quantitative and categorical variables

- Correlation
- Linear regression: simple and multiple
- Logistic regression and generalized linear models, GLM
- Chi-square test and Fisher's exact test

Block 5. Multiple testing correction and causal inference

- Multiple comparisons: FWER and FDR, using the Bonferroni, Holm, and Benjamini-Hochberg methods
- Causal inference: correlation, confounding, and causal graphs, or DAGs

Block 6. Advanced methods and topics

- Bayesian statistics: Bayes' theorem, priors, posteriors and credible intervals, posterior estimation and MCMC
- Deep learning and large language models, LLMs: from the perceptron to neural networks

Learning activities and methodology

Title	Hours	ECTS	Learning outcomes
Group mentoring sessions	4	0.16	
Computer practicals	12	0.48	
Practical work	20	0.8	
Studying hours	60	2.4	
Problems seminars	11	0.44	
Lecture classes	30	1.2	

The content of the Biostatistics course aims to give students a general introduction to the basic concepts of statistics, so they can understand statistical reasoning and correctly apply these tools to the design and analysis of experiments.

Lectures: Students acquire the core scientific knowledge of the course by attending lectures, which they complement with personal study of the topics covered. Classes will be dynamic and interactive, incorporating active-participation tools to encourage understanding and reflection on the concepts presented.

Problem-solving seminars: The knowledge acquired in lectures is applied to solving practical cases. These sessions are devoted to problem-solving, discussion of concepts and applications, and to student presentations, fostering active participation, critical thinking, and scientific communication skills.

Computer lab sessions: This part of the course is essential for consolidating the knowledge acquired. Students will learn to use the R language and various statistical packages to implement the analyses seen in the lectures. Basic data visualization techniques will also be covered, to communicate results clearly and effectively.

In addition, as a new feature for the 2026/27 academic year, the course will take part in the AIR-LAB teaching innovation project, working with real air quality data collected on campus. This data, generated and analyzed by the students themselves, will allow them to apply their knowledge of biostatistics and data visualization in a real-world context. In this first year the activity is set up as a non-graded pilot: students, working in groups, will analyze this data and present their results at the end of the course.

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	Weight	Hours	ECTS	Learning outcomes

Written exams. First midterm	20%	2	0.08	CM07, KM05, SM06
Seminar presentation	20%	1	0.04	CM08, CM09, SM05
Written exams. Second midterm	30%	3	0.12	CM07, CM08, KM05, SM06
Problem-based seminars and practicals	10%	5	0.2	CM07, CM08, CM09, KM05, SM05, SM06
Practical exams	20%	2	0.08	CM07, CM09, SM05, SM06

The competencies of this course will be assessed through continuous assessment, which includes written exams, practical tests, and individual assignments.

The assessment system is organized into three modules, each carrying a specific weight in the final grade.

Theory

Assessment will be carried out through two midterm exams. The first midterm will carry a weight of 20% and the second midterm a weight of 30%. The final resit exam is aimed at students who have not passed one of the midterms and will carry the same weight in the final grade as the corresponding midterm.

Problem-solving seminars and oral presentation

Assessment of the problem-solving seminars will be carried out through short in-class exercises. Students must solve and discuss the problems in front of their classmates. This component accounts for 10% of the final course grade.

There will also be a group oral presentation in which students present the statistical analysis of a dataset or case study. This activity will carry a weight of 20% of the final course grade and will be assessed through a combined system of self-assessment, peer assessment, and instructor assessment.

Labs

Assessment of this component will be carried out through a practical test in the computer lab. In this session, students must use the appropriate statistical software, enter the data from a study, propose an analysis, and answer specific questions. This test will carry a weight of 20% of the final grade.

Additional considerations

Students must notify the instructors by email, during the first day of class, whether they choose continuous assessment or single assessment.

The percentages corresponding to seminars and labs will apply provided the student has obtained a minimum grade of 4 in each theory midterm, or in the corresponding resit exam. A grade of 4 does not mean the student has passed theory outright, which requires a grade of 5 or higher, but it does allow that grade to be averaged with the seminar and lab grades.

The final course grade will be calculated by weighting the grades of each component (theory, seminars, and labs). However, a grade of 4 or higher in the first and second midterms, or in the resit exam, will be required in order to calculate the final grade.

Students who have passed the course through the midterm exams and wish to improve their grade may sit the theory resit exam. Sitting this exam means forfeiting the grade previously obtained in the corresponding midterm.

Students who, having chosen the continuous assessment track, have not taken part in the continuous assessment activities or submitted the problem-set assignments, may be assessed through the final exam. In this case, the final grade cannot exceed 70% of the maximum grade. This 70% cap applies exclusively to students on the continuous assessment track; it does not affect students who chose the single assessment option, whose exam can reach 100% of the final grade.

A final grade of 5 or higher is required to pass the course.

Repeating students must complete all assessment activities, including the problem-set submissions, participation in the seminar presentation, and the practical assessment.

Attendance at lab sessions is mandatory.

To be eligible for the resit, students must have already been assessed on activities whose combined weight amounts to at least two-thirds of the total course grade. Accordingly, students will receive a grade of "Not assessed" when the activities completed account for less than 67% of the final grade.

Single assessment

The single assessment option consists of a single exam covering theory (70%), computer labs (15%), and problem sets (15%). This exam will account for 100% of the final grade.

The single-assessment exam will coincide with the date of the second midterm of the continuous assessment track, while the practical component will coincide with the date of the continuous assessment's practical exam. The resit for the single assessment option will coincide with the resit for the continuous assessment track.

Use of AI

In this course, the use of Artificial Intelligence (AI) technologies is permitted exclusively for support tasks, such as literature or information searches, R code review, text correction, or translations. Students must clearly identify which parts were generated using these technologies, specify the tools used, and include a critical reflection on how they influenced the process and the final outcome of the activity. Lack of transparency in the use of AI will be considered a breach of academic integrity and may result in a partial or full penalty on the activity, as well as additional sanctions in more serious cases.

Fraud protocol

Any irregularity committed during an assessment activity (academic fraud, plagiarism, or improper use of AI, except in authorized cases) that could significantly alter the grade will result in that activity being graded 0. If the course syllabus establishes a minimum grade on this activity as a requirement to pass the course, or if multiple irregularities occur within the same course, the final grade will be 0. In addition, disciplinary proceedings may be initiated against the student responsible for any of these irregularities.

Bibliography

Biblioteca de Biociències:

- Faulín, Francisco Javier [i altres]. (2025). Bioestadística amigable. (5ª ed.) Elsevier Faulí (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991011225055106709)
- Grafen, Alan& Hails, Rosemary. (2002). Modern statistics for the life sciences. Oxford Graf University Press (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991008801599706709)
- Howell, David C. (1992). Statistical methods for psychology. (3rd ed.) Duxbury Press How (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991002129609706709)
- Martínez González, Miguel Ángel. (2025). Bioestadística amigable. (5th ed.) Mart (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991011198009206709)
- Pagano, Marcello & Mattie, Heather & Gauvreau, Kimberlee. (2022). Principles of Paga biostatistics. (3nd ed.) CRC Press (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991011201610106709)
- Shahbaba, Babak. (2012). Biostatistics with R : an introduction to statistics through Sha biological data. (1st ed.) Springer (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991010406543506709)
- Sokal, Robert R. & Rohlf, F. James. (2012). Biometry : the principles and practice of Soka statistics in biological research. (Extensively rev. 4th ed) WH Freeman (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991011158282906709)
- Vu, Julie & Harrington, David O. (2020). Introductory statistics for the life and biomedical Vu, sciences.

(1st ed.) OpenIntro (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1eqfv2p/alma991011220759006709)
- McElreath, R. (2020). *Statistical rethinking : a Bayesian course with examples in R and Stan* (Second Edition). CRC Press. (https://bibcercador.uab.cat/permalink/34CSUC_UAB/tidafd/alma991010621135306709)
- Pearl, J., Mackenzie, D., & García, G. (2020). *Pearl, Judea, Book of why. Castellà
El Libro del porqué : la nueva ciencia de la causa y el efecto*. Pasado & Presente. (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1fbc57r/alma991010842529906709)
- Tuffery, S. (2023). *Deep learning : from big data to artificial intelligence with R*. Wiley. (https://bibcercador.uab.cat/permalink/34CSUC_UAB/1fbc57r/alma991010729983706709)

GitHub/Internet:

- Rodenburg, F. J. (2020). Introduction to biostatistics. https://github.com/FransRodenburg/Biostatistics-Book-Series
- Rodenburg, F. J. (2021). Elements of biostatistics. https://github.com/FransRodenburg/Biostatistics-Book-Series
- Lawson, J. (2014). Design and Analysis of Experiments with R (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b17883

Software

The course will use the R software and the RStudio development environment, employing specific packages for statistical analysis and data visualization (such as *ggplot2*, *dplyr*, *tidyr*, among others). All the necessary software will be installed and available on the faculty's computers.

Course groups and languages

The information provided is provisional until November 30. After this date, you will be able to consult the language of each group through this link. To access the information, you will need to enter the course CODE

Type of teaching	Group	Language	Semester	Shift
(TE) Theory	62	Catalan/Spanish	first semester	morning-mixed
(PAUL) Classroom practices	621	English	first semester	morning-mixed
(PLAB) Practical laboratories	621	Catalan/Spanish	first semester	morning-mixed
(PAUL) Classroom practices	622	English	first semester	morning-mixed
(PLAB) Practical laboratories	622	Catalan/Spanish	first semester	morning-mixed
(PLAB) Practical laboratories	623	Catalan/Spanish	first semester	morning-mixed