

Degree programme	Type	Course
Applied Statistics	OP	4

Contact lecturer

Name : Angel Gonzalez Wong

Email : angel.gonzalez@uab.cat

Teaching staff

Carolina Soriano Tarraga

Gianluigi Caltabiano

Angel Gonzalez Wong

Juan Ramón Gonzalez Ruiz

Group languages

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

Prerequisites

Basic knowledge of the English language, as a large part of the articles, tutorials, and software packages are written in English.

It is recommended to have taken the Bioinformatics course or have equivalent knowledge of:

- Basics of Molecular Biology and Genomics.
- Basic Programming with R.

Objectives

The course aims to provide an overview of the possibilities of Big Data analysis focused on Biomedicine and Bioinformatics.

The course consists of two thematic blocks:

1. Computational Methodologies Applied to Drug Discovery
2. Omics Data Analysis

At the end of the course, the student will be able to:

- Characterize and manage large-scale data from biomedical research.
- Apply statistical, bioinformatics, and machine learning algorithms to biological and biomedical data.

- Integrate and visualize multiple layers of data to interpret biological hypotheses.
- Critically evaluate results and research in the field, integrate knowledge, and communicate information clearly and appropriately within the disciplinary context.

The course is part of the Statistics Mention for Health Sciences.

Learning outcomes

- CM14 (Propose the statistical model needed to analyse data sets belonging to real studies.) Propose the statistical model needed to analyse data sets belonging to real studies.
- KM17 (Recognise the statistical models for the analysis of data with different structures and complexities that frequently appear in different fields of application.) Recognise the statistical models for the analysis of data with different structures and complexities that frequently appear in different fields of application.
- KM18 (Recognise the language of applications of economics and finances, biomedical science and engineering, provided by research and innovation in the field of statistics.) Recognise the language of applications of economics and finances, biomedical science and engineering, provided by research and innovation in the field of statistics.
- SM16 (Select appropriate sources of information for the statistical work.) Select appropriate sources of information for the statistical work.
- SM17 (Discuss scientific articles in which the analysis of a study of the different areas of application is considered.) Discuss scientific articles in which the analysis of a study of the different areas of application is considered.
- SM18 (Refine the information available for subsequent statistical processing.) Refine the information available for subsequent statistical processing.
- SM19 (Analyse complex data, whether this is due to their characteristics or their size.) Analyse complex data, whether this is due to their characteristics or their size.

Contents

MODULE 1. Big Data in Drug Discovery

- Introduction to Big Data in Biosciences, Bioconductor, and the R ecosystem
- Databases and representation of biological components and chemical compounds.
- Analysis, clustering, and visualization of chemical and pharmacological substances.
- Virtual Screening in Drug Discovery.

MODULE 2. Big Data in Omics Data Analysis

- Introduction to Bioconductor and bioinformatics tools for omics data analysis.
- Genetic association studies and GWAS (Genome-Wide Association Studies).
- Multivariate Methods for the Integration of Omics Data and Big Data.

Learning activities and methodology

Title	Hours	ECTS	Learning outcomes
Theory classes	21	0.84	
Preparation of Research Project	20	0.8	
Practical sessions	21	0.84	
Tutoring	10	0.4	
Presentation of Research Project	3	0.12	

The course is organized in sessions of 3 hours. Each session consists of a theoretical part (theory classroom) that will introduce the new concepts followed by a practical part (computer room) where the students will work on the implementation of concepts explained in the theoretical part. In each session the teacher will indicate the students some tasks to do autonomously, such as reading articles, resolution of class exercises or sending reports. The material used by the teachers will be available on the Virtual Campus of the course.

In this course, the use of Artificial Intelligence (AI) technologies is allowed as an integral part of the development of the work, provided that the final result reflects a significant contribution from the student in terms of analysis and personal reflection. The student must clearly identify which parts have been generated using this technology, specify the tools used, and include a critical reflection on how these have influenced both 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 penalty in the grade for the activity, or more serious sanctions in severe cases.

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
Practicum Reports Preparation	30	0.5	0.02	CM14, KM17, KM18, SM16, SM18, SM19
Presentation Research Project	20	2	0.08	KM18, SM16, SM17, SM18, SM19
Presentation class exercises	30	0.5	0.02	CM14, KM17, KM18, SM18, SM19
Theoretical-Practical Exam	20	2	0.08	CM14, KM17, KM18, SM19

BLOCK 1. Big Data in Drug Design (50%):

- Class exercises presentation (15%)
- Preparation of Practice Reports (15%)
- Bioinformatics Project Presentation before a committee (20%)

BLOCK 2. Big Data in Omics Data Analysis (50%):

- Class exercise presentation (15%)
- Preparation of Practice Reports (15%)
- Theoretical-Practical Test (20%)

The minimum overall grade required to pass the course will be 5 points. To calculate the average, the minimum grade for each of the assessable activities must be equal to or greater than 3,5 points.

In order to be eligible for the resit, students must have previously been assessed in a set of activities whose weight is equivalent to at least two-thirds of the total grade for the course. Students who have failed or not submitted one or more of the assessments may take the resit exam corresponding to the failed block. If the established threshold is not reached in any of the blocks during the resit, the final course grade will be the minimum of the block grades. This course does not allow for the single assessment system.

The commission of any irregularity in an assessment activity (academic fraud, plagiarism, or improper use of AI, unless such use is expressly authorized in the course guide) that may lead to a significant change in the grade will result in that activity being graded with a 0. If the course guide establishes that passing the subject

requires obtaining a minimum mark in that assessment activity, or if multiple irregularities occur in the assessment activities of the same subject, the final grade for the subject will be 0. In addition, disciplinary proceedings may be initiated against any student who incurs any of these irregularities.

Bibliography

- Attwood, T.K., Parry-Smith, D.J., *Introducción a la Bioinformática*. Pearson Education, 2002.
- Foulkes A.S. *Applied Statistical Genetics with R. For Population-based Association Studies*. Springer Dordrecht Heidelberg London New York. ISBN 978-0-387-89553-6
- Buffalo, V. *Bioinformatics Data Skills*. O'Reilly Media, 2015.
- Lesk, A. M. *Introduction to Bioinformatics*. Oxford University Press, 2019.
- González, J. R., Cáceres, A. *Omic Association Studies with R and Bioconductor*. Chapman and Hall/CRC, ISBN 9781138340565, 2019.
- Specialized readings and articles available on the course's virtual campus
- <https://www.bioconductor.org/>

Software

R: <https://www.r-project.org/>

Rstudio: <https://www.rstudio.com/>

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	1	Catalan	second semester	afternoon
(PLAB) Practical laboratories	1	Catalan	second semester	afternoon