

Academic Year/course: 2022/23

## 68460 - Biostatistics & Bioinformatics

### Syllabus Information

---

**Academic Year:** 2022/23

**Subject:** 68460 - Biostatistics & Bioinformatics

**Faculty / School:** 100 - Facultad de Ciencias

**Degree:** 626 - Máster Universitario en Biofísica y Biotecnología Cuantitativa/Biophysics and Quantitative Biotechnology

**ECTS:** 6.0

**Year:** 01

**Semester:** Second semester

**Subject Type:** Optional

**Module:**

## 1. General information

### 1.1. Aims of the course

To know the theoretical basis of different tools for statistical modeling of biological data. To know when they can be applied and the types of problems that can be solved by each technique.

To know how to apply statistical tools to the analysis of biological data using adequate software and to program basic analysis.

To know how to interpret the results of statistical analysis.

To know the bioinformatic tools for the study of genomes, genes and proteins with relevant applications in Biotechnology and Biomedicine.

To be trained in the use of basic programming techniques applied to Biology.

These approaches and objectives are aligned with the following Sustainable Development Goals (SDGs) of the United Nations 2030 Agenda (<https://www.un.org/sustainabledevelopment/es/>), in such a way that the acquisition of the learning outcomes of the module provides training and competence to contribute to some extent to their achievement: (4) Quality education, (5) Gender equality, (8) Decent work and economic growth, (9) Industry, innovation and infrastructure, (17) Partnerships for the goals.

### 1.2. Context and importance of this course in the degree

This course introduces some fundamental concepts of statistical modelling for biological data and algorithms of computational biology and bioinformatics, from a practical point of view.

### 1.3. Recommendations to take this course

Students are expected to be familiar with basic statistics and molecular biology, and should have basic programming skills, particularly in R. More precisely, they should have taken courses on probability and statistical inference, and they should be familiar with the use of random variables, probability distributions, point estimation, hypothesis testing and basic statistical models.

## 2. Learning goals

### 2.1. Competences

Relevant probabilistic models and results for the analysis of biological data.

Statistical inference methods for problems in Biotechnology.

Construction and validation of predictive and classification models (supervised and unsupervised methods).

Programming in R for analysing and plotting biological data.

Main data formats in bioinformatics.

Main knowledge for sequence alignment algorithms.

Main approaches for comparing and predicting regulatory sequences and structures in proteins.

Main concepts of RNA sequencing

### 2.2. Learning goals

The student should demonstrate the following skills:

1: To be able to select the appropriate statistical tool or technique to model different types of biological data and to implement the analysis and modelling of biological data using R.

2: To be able to use fundamental computational tools for the study of genomes, genes and proteins, and their applications in biotechnology and biomedicine.

3: To perform simple programming tasks in the context of biological data.

### 2.3. Importance of learning goals

The skills developed while accomplishing these goals are valuable to understand the statistical methods and bioinformatics procedures in the literature, and often used during the course of research projects in this area.

## 3. Assessment (1st and 2nd call)

### 3.1. Assessment tasks (description of tasks, marking system and assessment criteria)

A. Solving of problems and practical cases, both individually and in team work. Students must submit a report at the end of each chapter following the guidelines and presentation format. These types of tasks are framed within the concept of continuous evaluation, which will allow monitoring of the learning process.

B. Written quizzes or exams which can include both theoretical and applied questions discussed throughout the course.

After each chapter or session, students are expected to return in time a report including different individual or team work tasks which will be marked by the lecturer in order to track the progress.

Students are expected to participate in class and might optionally be given a final exam regarding any aspect of the contents of the course.

## 4. Methodology, learning tasks, syllabus and resources

### 4.1. Methodological overview

The methodology followed in this course is oriented towards the achievement of the learning objectives. Several teaching and learning tasks are implemented, such as theory sessions and computer lab sessions. Theoretical and practical issues will be often combined in the same sessions, so that lectures will take place in a computer room.

Students are expected to participate actively in the class throughout the semester.

Classroom materials will be available via Moodle. These include a repository of the lecture notes used in class, the course syllabus, datasets as well as other course-specific learning materials, including a discussion forum.

Further information regarding the course will be provided on the first day of class.

## 4.2. Learning tasks

This is a 6 ECTS course organized as follows:

- Theory sessions. Lecture notes and examples will be available for the students.
- Laboratory sessions. They aim to carry out exercises to solve problems that appear in biological studies using adequate software. The code, and their corresponding solutions will be available for the students. Some of the exercises will be solved in class by the teacher. Students are provided in advance with task guidelines for each session.
- Assignments. The lecturer will also assign unsolved exercises, which the students will submit to be assessed.
- Autonomous work.

## 4.3. Syllabus

The course will address the following topics:

### Section 1. Biostatistics:

Probabilistic results for statistical inference. Central limit theorem and others.

Statistical Inference. Parametric and non parametric tests.

Multiple testing and error control.

Bayesian methods.

Supervised methods: Prediction models (linear regression models, ANOVA, generalised linear models). Model validation and goodness of fit measures (AIC, BIC, MSE, ROC curves and others).

Unsupervised methods: Classification techniques and dimension reduction techniques.

Markov chains and hidden models.

### Section 2. Bioinformatics

Computational tools for data processing in bioinformatics: reading, manipulation and writing of files

Common formats in bioinformatics: nucleic acid and protein sequences and their alignments (FASTA), molecular structures (PDB) and phylogenetic trees (Newick).

Algorithms of the dynamic programming for local and global alignments. Search for similar sequences in local databases by means of alignments. Multiple alignments of DNA and protein sequences.

Alignments of protein structures and calculation of RMSD.

Design, development and fundamentals of analysis of RNAseq experiments. Gene functional and structural annotation.

## 4.4. Course planning and calendar

The course is taught throughout the second semester, from February to June:

- The first 30h correspond to Biostatistics
- The following 30h correspond to Bioinformatics

Further information concerning the timetable, classroom, office hours, assessment dates and other details regarding this course will be provided on the first day of class or please refer to the "Facultad de Ciencias" website and the department website ( <https://ciencias.unizar.es> )

## 4.5. Bibliography and recommended resources

<http://psfunizar10.unizar.es/br13/egAsignaturas.php?codigo=68460>