

### **IDENTIFICATION DETAILS**

Degree:	Biomedicine			
Scope	Biology and Genetics			
Faculty/School:	Experimental Sciences			
Course:	ADVANCED BIOSTATISTICS AND DATA MANAGEMENT			
Туре:	Optional		ECTS credits:	3
Year:	4		Code:	2178
Teaching period:	Seventh semester			
Subject:	Statistics			
Module:	Experimental Methodology in Biomedicine			
Teaching type:	Classroom-based			
Language:	Spanish			
Total number of student study hours:	75			

### SUBJECT DESCRIPTION

The course of advanced biostatistics and data management introduces students to multivariate statistical analysis. In the course of this course, the student will learn to design databases with multiple variables of different types. You will also learn to use different analysis techniques depending on the objective of the study and the type of data being handled.

In a biomedical study, in addition to posing the questions and the appropriate design to answer them, the correct collection and manipulation of data is of vital importance. The errors made in this process can have dire consequences, and can lead to false security in conclusions that, in fact, are erroneous. That is why the first part of this course will be dedicated to the design of databases and their proper management during the analysis process, as well as the use of descriptive statistics and data visualization techniques to assess their quality.

Once a true and accurate database is available, it is crucial to choose the right statistical techniques to be able to draw relevant conclusions in the research. Multivariable techniques will allow us to go beyond what is seen in the subject of Basic Biostatistics, allowing us to build models that allow us to evaluate the relationship between two or more variables when explaining a third one, analyzing how an event evolves over time and what variables such evolution depends on, or trying to find patterns among the data that allow us to differentiate groups.

## GOAL

At the end of taking this course, the student should be familiar with the main multivariate analysis techniques and their implementation in statistical software such as R. They should be able to design and carry out a statistical analysis with multiple variables as well as to interpret the results published and analyzed by third parties. The specific objectives of the course are:

- Know the different advanced statistical techniques and in what contexts they are applicable.

- Be able to design a database for the collection of observations on multiple variables in a systematic and functional way.

- Know how to interpret the results obtained after the application of multivariable techniques.

### PRIOR KNOWLEDGE

The student must have a solid background in the concepts of basic statistics as well as in the design of biomedical studies. This knowledge is acquired in the subjects of the subject Statistics of previous courses.

### **COURSE SYLLABUS**

- Teaching Unit I: Programming with R
- 1. Introduction to the R programming language.
- 2. Data management and manipulation.
- 3. Database cleaning and data visualization.

Teaching Unit II: Advanced Statistical Methods

- 4. Correlation and multiple linear regression.
- 5. Logistic regression.
- 6. Survival analysis.
- 7. Classification methods.
- 8. Dimensionality reduction and association methods.

In-person activities:

AF1 Expository class: the teacher will present the theoretical contents of the subject.

AF2 Practical class: the resolution of practical cases will be carried out in a cooperative manner.

AF3 Tutoring: at the student's request, the teacher will answer questions and guide the student to better learn the subject.

AF4 Evaluation: carrying out evaluation tests.

Non-face-to-face activities:

AF5 Theoretical study: the student must internalize the contents presented in the classes.

AF6 Preparation of practical classes: in addition to practical classes, the student will have exercises to practice independently. This type of work is of great importance for the assimilation of content and the raising of new questions.

Warning:

The teachers of the subject do not authorize the publication by the student of the material provided by the teachers of the subject in the virtual classroom, or by any other means.

## DISTRIBUTION OF WORK TIME

TEACHER-LED TRAINING ACTIVITIES	INDIVIDUAL WORK
30 Horas	45 Horas

### LEARNING RESULTS

Critically evaluate and use sources of clinical and biomedical information to obtain, organize, interpret and communicate information from the various areas that constitute biomedicine.

Understand the design of experiments based on statistical criteria and the various tools available for data processing in the area of life and health sciences.

# SPECIFIC LEARNING RESULTS

Design databases in a clear, organized and functional way for the collection of experimental data in the context of a biomedical study.

Identify when and for what purpose the different multivariable analysis techniques are applicable.

Apply the different techniques used correctly through the use of statistical software.

Develop the capacity to critically interpret the results presented after a multivariable analysis.

Use the results of a multivariable analysis to generate conclusions or guide decision-making.

## LEARNING APPRAISAL SYSTEM

The learning evaluation system will be through continuous evaluation, and will consist of the following aspects:

SE1 Evaluation of the theoretical content (70%). It is essential to get at least a 4 on the exam to pass the subject.

SE2 Realization and resolution of exercises and practical cases (20%).

SE3 Carrying out and evaluating experimental work (10%).

Alternative evaluation system (students with second or subsequent enrollment must contact the teacher to request to take advantage of this system):

SE1 Theory exam (70%). It is essential to get at least a 4 on the exam to pass the subject.

SE2 Realization and resolution of exercises and practical cases (30%). A series of exercises and practical cases will be proposed that the student can carry out independently.

Warning:

Plagiarism, as well as the use of illegitimate means in evaluation tests, will be sanctioned in accordance with the university's Evaluation Regulations and Coexistence Regulations.

# ETHICAL AND RESPONSIBLE USE OF ARTIFICIAL INTELLIGENCE

1.- The use of any Artificial Intelligence (AI) system or service shall be determined by the lecturer, and may only be used in the manner and under the conditions indicated by them. In all cases, its use must comply with the following principles:

a) The use of AI systems or services must be accompanied by critical reflection on the part of the student regarding their impact and/or limitations in the development of the assigned task or project.

b) The selection of AI systems or services must be justified, explaining their advantages over other tools or methods of obtaining information. The chosen model and the version of AI used must be described in as much detail as possible.

c) The student must appropriately cite the use of AI systems or services, specifying the parts of the work where they were used and describing the creative process followed. The use of citation formats and usage examples may be consulted on the Library website(<u>https://www.ufv.es/gestion-de-la-informacion\_biblioteca/</u>).

d) The results obtained through AI systems or services must always be verified. As the author, the student is responsible for their work and for the legitimacy of the sources used.

2.- In all cases, the use of AI systems or services must always respect the principles of responsible and ethical use upheld by the university, as outlined in the <u>Guide for the Responsible Use of Artificial Intelligence in Studies at UFV</u>. Additionally, the lecturer may request other types of individual commitments from the student when deemed necessary.

3.- Without prejudice to the above, in cases of doubt regarding the ethical and responsible use of any AI system or service, the lecturer may require an oral presentation of any assignment or partial submission. This oral evaluation shall take precedence over any other form of assessment outlined in the Teaching Guide. In this oral defense, the student must demonstrate knowledge of the subject, justify their decisions, and explain the development of their work.

### **BIBLIOGRAPHY AND OTHER RESOURCES**

# Basic

Daniel Pena. Regression and design of experiments./2nd ed. 3rd reimp. Madrid: Editorial Alliance, 2017.

Daniel Peña. Multivariate data analysis/Madrid:McGraw-Hill.

### Additional

Gareth James, Daniela Witen, Trevor Hastie, Robert Tibshirani An Introduction to Statistical Learning 2 (Gareth James, Daniela Witen, Trevor Hastie, Robert Tibshirani An Introduction to Statistical Learning 2, https://www.statlearning.com/)

Álvarez Cáceres, Rafael. Multivariate and non-parametric statistics with SPSS [electronic resource] application to health sciences/Madrid:Ediciones Díaz de Santos, 1995.

Joseph F. Hair, Jr.... [et al.]; Diego Cano translation; technical review and compilation of complementary readings by Mónica Gómez Suárez. Multivariate Analysis/5th ed. Madrid: Pearson, 2007.