

COURSE PROGRAM

16 hours; 23 – 26 October 2023

Session 1. 10:00 – 14:00 h (break from 11:45 to 12:15h)

- 1.1 Descriptive statistics
- 1.2 Probability distributions
- 1.3 Data exploration
- 1.4 Exercises

Session 2. 10:00 – 14:00 h (break from 11:45 to 12:15h)

- 2.1 Contrast of hypothesis: parametric and non-parametric methods
- 2.2 Goodness of fit contrasts
- 2.3 Homogeneity contrasts
- 2.4 Exercises

Session 3. 10:00 – 14:00 h (break from 11:45 to 12:15h)

- 3.1 Linear regression
- 3.2 Non-linear regression
- 3.3 Non-parametric regression
- 3.4 Exercises

Session 4. 10:00 – 14:00 h (break from 11:45 to 12:15h)

- 4.1 Analysis of variance: foundations
- 4.2 Analysis of variance: worked example
- 4.3 Analysis of variance: repeated-measures and nested designs
- 4.4 Exercises

Instructor: Aldo Barreiro Felpeto. CIIMAR.

Price: 100 € (CIIMAR/UP members) else 150 €.

Registration: after announcement, up to fill 25 available positions.

First consult by e-mail (abarreiro@ciimar.up.pt) if there are spots available. Then you can perform registration, which, together with the payment information, is available in the CIIMAR website, through the link that is sent with the e-mail announcing the course.

Proof of payment required to book the place (send proof to abarreiro@ciimar.up.pt).

Right after affective registration, a confirmatory e-mail will be sent.

Minimum audience required: 5 registrations 2 weeks before the course.

Important additional information:

- The course will be 100% through the zoom platform
- The course will be taught in English.
- The software for teaching will be R, but it is not necessary any background with R. The aim of the course is not learning R language. The material (examples, exercises) is designed in a way that the

software employed is not an obstacle.

- All the information and materials necessary for the development of the course will be made available for all the participants in the course through a link to the Open Science Framework website platform.

Contact: abarreiro@ciimar.up.pt

Syllabus

Course description

This course will focus on four aspects: descriptive statistics, distributions, regression, and hypothesis testing. Mastering these aspects will enable students to study in the future more complex statistical models that are more useful today in research, such as GLM and its extensions GLZ and GAM. In this course will be included an introductory example of GLM with an Analysis of Variance, that also includes random factors will be seen in an introductory way. Although this course uses the R programming language as the basis for the examples and exercises, which will undoubtedly serve to further familiarize the student with it, learning this language is not an objective of this course.

Course methodology

The course consists of exemplified theoretical explanations, for which the open source programming language R will be used. Students will be provided with the material for the course lessons in pdf format, which contains the theoretical explanations together with the worked examples. The examples will also be supplied as R-language scripts. In addition, students will be provided with data files, which will be used to carry out a series of exercises during class hours. These exercises will also be reviewed during class hours.

General aim of the course

Become familiar with the basic concepts of statistics and an introductory example of statistical model.

Specific aims of the course

- Understand the meaning and usefulness of the main descriptive statistics parameters and measures.
- Learn basic aspects of distributions: parameters that define them, estimation of these parameters from a series of data, usefulness of the different distributions.
- Become familiar with different basic graphic exploration options.
- Learn to differentiate the common aspects that define all regression techniques, and also those that differentiate them to apply them to each specific case.
- Learn various ways of fitting and selecting regression models.
- Understand the objective and fundamentals of hypothesis testing, as well as provide a general view of all the specific cases that exist for these tests, both in the parametric and non-parametric statistical framework.
- Understand the fundamentals of the analysis of variance.
- Recognize the differences between fixed factors and random factors, in the context of the general linear model using analysis of variance.

Aldo Barreiro Felpeto is a researcher at Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) associated to the University of Porto (Porto, Portugal). His research career has focused in plankton ecology. He defended his Ph.D. dissertation in 2007 in the Department of

Ecology at the University of Vigo (Vigo, Spain) about interactions between zooplankton and toxic phytoplankton species from the Spanish NW Atlantic coast, southern Baltic sea and southern Tirreno coast. In 2008-2010, he performed a post-doctorate in the Department of Ecology and Evolutionary Biology at Cornell University (Ithaca, New York, USA). Since 2011 he is a researcher at CIIMAR.

He developed a strong background in statistics and dynamic modelling with R software, attending 10 courses in the period 2006-2018 and since 2013, organizing 14 editions of courses about different aspects of statistics and programming with R, mostly in CIIMAR, but also in the University of Vigo (Spain) and the University of Magallanes (Chile). He co-authored two books about statistics and programming: *Tratamiento de Datos* (Ed. Díaz de Santos, Madrid, 2006) and *Tratamiento de Datos con R, SPSS y ESTATISTICA* (Ed. Díaz de Santos, Madrid, 2010).

Due to his expertise in statistics and programming, he has developed collaborations in different fields of ecology, but also environmental sciences and molecular biology. He has published 56 articles, accounting for an *h* index of 23 and an *i10* index of 40.