MULTIVARIATE STATISTICS WITH *R* FOR BIOLOGICAL SCIENCES

COURSE PROGRAM

CIIMAR – 20 hours; 11-15 May 2025

Session 1. Contrast of hypothesis 10:00 – 14:00 h (break from 11:45 to 12:15h)
      1.1 Parametric methods
      1.2 Non-parametric methods
      1.3 Exercises

Session 2. Ordination methods I 10:00 – 14:00 h (break from 11:45 to 12:15h)
      2.1 Principal Components Analysis
      2.2 Robust PCA
      2.3 PCA-env
      2.2 Correspondences Analysis
      2.3 Factorial analysis
      2.5 Exercises

Session 3. Ordination Methods II 10:00 – 14:00 h (break from 11:45 to 12:15h)
      3.1 Canonical correspondence Analysis
      3.2 Redundancy Analysis
      3.3 Multi-dimensional Scaling
      3.4 Exercises

Session 4. Other methods applied to the study of Ecological niche. 10:00 – 14:00 h (break from 11:45 to 12:15h)
      4.1 ENFA
      4.2 Random forests
      4.3 Exercises.

Session 5. Classification methods 10:00 – 14:00 h (break from 11:45 to 12:15h)
      5.1 Discriminant Analysis
      5.2 Hierarchical Clustering Analysis
      5.3 k-means Cluster Analysis
      5.4 Exercises

Instructor: Aldo Barreiro Felpeto. CIIMAR.

Price: 240 € (120 € for CIIMAR-LA/UP members)

Registration: after announcement, up to fill 25 available positions.
Registration, together with the payment information, is available in the CIIMAR website, through the link that is sent with the e-mail announcing the course.

Important information:

- All the course will be taught through a TEAMS platform.
- The course will be taught in English.
- At least a beginner's background in R and basic statistics is recommended.
- All the information and materials necessary for the development of the course (instructions to install R and R packages, pdf with lessons content, scripts with examples and exercises, data for examples and exercises) will be made available for all the participants in the course through a link to the *Open Science Framework* platform.

Syllabus

Course description
This course will focus on the most useful techniques of multivariate statistical analysis in Biological sciences. These will be principal component analysis, canonical correspondence analysis, multidimensional scaling, discriminant analysis and cluster analysis. Other techniques, such as correspondence analysis and factor analysis will be approached in a simpler way. Lastly, some techniques will be approached in a very introductory way, as it will be the case of canonical correlation, and redundancy analysis. A brief introduction to parametron and non-parametric techniques for hypothesis contrasts (Hoteling test, MANOVA, PERMANOVA) will also be part of the first session.

Although this course uses the R programming language as the basis for the examples and exercises, which will undoubtedly help the student to become more familiar 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 receive the course lesson material in pdf format, which contains the theoretical explanations together with worked examples. The examples will also be provided as scripts in R language. In addition, students will be provided with data files, which will be used to perform a series of exercises during class hours. These exercises will also be reviewed during class hours.

General aim of the course
To become familiar with the basic concepts of multivariate statistics and to know in detail the main techniques.

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 60 articles, accounting for an *h* index of 27 and an *i10* index of 48.