

ESTATÍSTICA MULTIVARIADA COM R PARA CIÊNCIAS BIOLÓGICAS

PROGRAMA DO CURSO

CIIMAR – 20 horas; 11-15 Maio 2025

Sessão 1. Contraste de hipóteses 10:00 – 14:00 h (Descanso das 11:45 às 12:15h)

- 1.1 Métodos paramétricos
- 1.2 Métodos não paramétricos
- 1.3 Exercícios

Sessão 2. Métodos de Ordenação I 10:00 – 14:00 h (Descanso das 11:45 às 12:15h)

- 2.1 Análise de Componentes Principais (PCA)
- 2.2 PCA robusto
- 2.3 PCA-env
- 2.2 Análise de Correspondências
- 2.3 Análise Factorial
- 2.5 Exercícios

Sessão 3. Métodos de Ordenação II 10:00 – 14:00 h (Descanso das 11:45 às 12:15h)

- 3.1 Análise de Correspondências Canónicas
- 3.2 Análise da Redundância
- 3.3 Escalamento Multi-Dimensional
- 3.4 Exercícios

Sessão 4. Outros métodos aplicados ao estudo do nicho ecológico 10:00 – 14:00 h (Descanso das 11:45 às 12:15h)

- 4.1 ENFA
- 4.2 Random forests
- 4.3 Exercícios

Sessão 5. Métodos de Classificação 10:00 – 14:00 h (Descanso das 11:45 às 12:15h)

- 5.1 Análise Discriminante
- 5.2 Análise de Conglomerados hierárquicos
- 5.3 Análise de conglomerados k-médias
- 5.4 Exercícios

Instrutor: Aldo Barreiro Felpeto. CIIMAR.

Preço: 240 € (120 € para membros do CIIMAR-LA/UP)

Inscrição: após anúncio, até preencher 25 vagas disponíveis.

A inscrição, juntamente com a informação de pagamento, está disponível no site do CIIMAR, através do link enviado com o e-mail de divulgação do curso.

Informação importante:

- Todo o curso será ministrado através da plataforma Teams.
- O curso será lecionado em inglês.
- Recomenda-se, no mínimo, conhecimentos básicos em R e estatística.
- Todas as informações e materiais necessários para o desenvolvimento do curso (instruções para instalação do R e pacotes R, PDF com o conteúdo das aulas, scripts com exemplos e exercícios, dados para exemplos e exercícios) serão disponibilizados a todos os participantes do curso através de um link para a plataforma Open Science Framework.

Syllabus

Descrição do curso

Este curso irá focar-se nas técnicas mais úteis de análise estatística multivariada em Ciências Biológicas. Estas serão: análise de componentes principais, análise de correspondência canónica, escalonamento multidimensional, análise discriminante e análise de clusters. Outras técnicas, como a análise de correspondência e a análise fatorial, serão abordadas de forma mais simples. Por fim, serão abordadas algumas técnicas de forma bastante introdutória, como será o caso da correlação canónica e da análise de redundância. Uma breve introdução às técnicas paramétricas e não paramétricas para contrastes de hipóteses (teste de Hotelling, MANOVA, PERMANOVA) fará também parte da primeira sessão.

Embora este curso utilize a linguagem de programação R como base para os exemplos e exercícios, o que sem dúvida ajudará o aluno a familiarizar-se com a mesma, a aprendizagem desta linguagem não é um objetivo deste curso.

Metodologia do curso

O curso consiste em explicações teóricas exemplificadas, para as quais será utilizada a linguagem de programação de código aberto R. Os alunos receberão o material didático do curso em formato PDF, que contém as explicações teóricas juntamente com exemplos práticos. Os exemplos serão também fornecidos como scripts em linguagem R. Além disso, os alunos receberão ficheiros de dados, que serão utilizados para realizar uma série de exercícios durante as aulas. Estes exercícios também serão revistos durante as aulas.

Objetivo geral do curso

Familiarizar-se com os conceitos básicos da estatística multivariada e conhecer detalhadamente as principais técnicas.

Aldo Barreiro Felpeto é investigador no Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) da Universidade do Porto (Porto, Portugal). A sua carreira de investigação tem-se centrado na ecologia do plâncton. Defendeu a sua tese de doutoramento em 2007 no Departamento de Ecologia da Universidade de Vigo (Vigo, Espanha) sobre as interações entre o zooplâncton e as espécies fitoplanctónicas tóxicas da costa atlântica noroeste espanhola, do sul do Mar Báltico e da costa sul do Tirreno. Entre 2008 e 2010, realizou um pós-doutoramento no Departamento de Ecologia e Biologia Evolutiva da Universidade de Cornell (Ithaca, Nova Iorque, EUA). Desde 2011, é investigador no CIIMAR.

Desenvolveu uma sólida formação em estatística e modelação dinâmica com o software R, tendo frequentado 10 cursos entre 2006 e 2018 e, desde 2013, organizado 14 edições de cursos sobre diferentes aspetos da estatística e programação com R, principalmente no CIIMAR, mas também na Universidade de Vigo (Espanha) e na Universidade de Magallanes (Chile). É coautor de dois livros sobre estatística e programação: Tratamiento de Datos (Ed. Díaz de Santos, Madrid, 2006) e Tratamiento de Datos con R, SPSS y ESTATISTICA (Ed. Díaz de Santos, Madrid, 2010).

Devido à sua experiência em estatística e programação, desenvolveu colaborações em diferentes áreas da ecologia, mas também das ciências ambientais e da biologia molecular. Publicou 60 artigos, com um índice h de 27 e um índice i10 de 48.