

MODELOS ESTATÍSTICOS COM R PARA CIÊNCIAS BIOLÓGICAS

PROGRAMA DO CURSO

CIIMAR – 30 horas; 9-13 Março 2026

Sessão 1. 10:00 – 17:00 h (Descanso das 14h às 15h)

- 1.1 Relembrar regressão linear
- 1.2 Exercício
- 1.3 Relembrar o ANOVA
- 1.4 Exercício

Sessão 2. 10:00 – 17:00 h (Descanso das 14h às 15h)

- 2.1 Modelo Linear Geral (GLM). Introdução e exemplo 1
- 2.2 GLM, Exemplo 2
- 2.3 Exercício
- 2.4 GLM com fatores aleatórios (modelos de efeitos mistos). Exemplo

Sessão 3. 10:00 – 17:00 h (Descanso das 14 às 15h)

- 3.1 GLM com fatores aleatórios. Desenhos hierárquicos (aninhados)
- 3.2 GLM com fatores aleatórios. Desenhos de declives aleatórios
- 3.3 Exercício
- 3.4 Exemplo de GLM com método de Monte-Carlo

Sessão 4. 10:00 – 17:00 h (Descanso das 14h às 15h)

- 4.1 Modelos Lineares Generalizados (GLZ). Introdução e exemplo com distribuição binomial
- 4.2 GLZ, exemplo com distribuição de Poisson
- 4.3 GLZ, exemplo com sobre-dispersão (distribuição binomial negativa)
- 4.4 GLZ com fatores aleatórios. Exemplo
- 4.5 Exercício

Sessão 5. 10:00 – 17:00 h (Descanso das 14h às 15h)

- 5.1 Exercício
- 5.2 Modelos Aditivos Generalizados (GAM). Introdução e exemplo
- 5.3 GAM, exemplo com fatores aleatórios
- 5.4 Exercício
- 5.5 Apêndice com desenhos avançados
- 5.6 Exercício

Instrutor: Aldo Barreiro Felpeto. CIIMAR.

Preço: 300 € (150 € 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 adicional 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

A primeira sessão do curso é um lembrete da regressão linear e da análise de variância, as duas técnicas estatísticas que constituem a base das estruturas mais importantes dos modelos estatísticos. Estes modelos estatísticos (GLM, GLZ e GAM) serão o conteúdo das próximas quatro sessões.

O GLM (Modelo Linear Geral) será estudado na segunda e terceira sessões. Devido à sua robustez e relativa simplicidade, este é o modelo estatístico mais utilizado. Por este motivo, no curso, dedicamos mais tempo ao GLM do que aos outros frameworks de modelação (GLZ, GAM). Através de exemplos práticos, serão explicados os seguintes tópicos: I) Pressupostos do GLM, como testá-los e como resolver problemas quando alguns desses pressupostos não são cumpridos (transformação de dados para não normalidade, ajuste alternativo por mínimos quadrados generalizados para heterocedasticidade de variância e autocorrelação residual); II) contraste de hipóteses para os efeitos principais e testes post-hoc para os efeitos simples; III) visualização das previsões do modelo.

No GLM será explicada com particular detalhe a inclusão de fatores aleatórios (modelos simples de efeitos mistos, delineamentos hierárquicos e delineamentos de declives aleatórios).

Os GLZ (Generalized Linear Models) são uma extensão do GLM para distribuições de erro que podem ser diferentes das gaussianas. Serão apresentados três exemplos práticos para diferentes distribuições (Poisson binomial e binomial negativa, para dados sobredispersos) e um exemplo simples incluindo fatores aleatórios.

O GAM é uma estrutura de modelo estatístico que pode incorporar quaisquer características de GLM e GLZ. Além disso, e considerando restrições adicionais, podem incorporar funções suavizadas (não lineares, mas não paramétricas) para descrever a relação entre a variável resposta e uma ou várias variáveis preditoras. Esta estrutura de modelo estatístico tem vindo a ganhar popularidade em diferentes campos da ciência devido ao aumento do poder computacional nos últimos 20 anos.

Recomenda-se para este curso um nível básico na linguagem R, para além de alguma familiaridade com técnicas estatísticas básicas (regressão, contraste de hipóteses, ANOVA).

O curso está aberto a todos os níveis, desde estudantes de licenciatura a investigadores seniores. É considerado a "segunda parte" de um outro curso regularmente ministrado no CIIMAR, intitulado Estatística Introdutória para Ciências Biológicas.

Metodologia do curso

O conteúdo do curso consiste em breves introduções teóricas a cada tema, seguidas de exemplos práticos em linguagem R, ao longo dos quais são apresentadas todas as características do tema específico. No final de cada secção, os alunos receberão dados para realizar exercícios que serão revistos e corrigidos durante as aulas.

As explicações teóricas, bem como os exemplos práticos, são totalmente desenvolvidos em PDF, com scripts em R com os exemplos como suporte adicional. O acesso a todo este material será disponibilizado dias antes do curso através de um link para a plataforma Open Science Framework. As soluções para os exercícios serão também entregues aos alunos no final do curso.

Objetivo geral do curso

Visão geral dos principais modelos estatísticos para compreender as diferenças entre eles e, conseqüentemente, os seus contextos específicos de aplicabilidade.

Objetivos específicos do curso

- Compreender os princípios dos frameworks de modelos estatísticos para os aplicar de acordo com as características dos dados.
- Testar os pressupostos do modelo com gráficos de contraste de hipóteses e de diagnóstico, e aprender a resolver problemas associados a este processo (transformação de dados, técnicas de ajuste alternativas, distribuições alternativas de erros, utilização de ajustes não lineares ou não paramétricos).
- Compreender a diferença entre fatores aleatórios e fixos e as possíveis formas de incorporar fatores aleatórios no projeto do modelo.
- Aprender sobre os diferentes contrastes de hipóteses que podem ser realizados num modelo estatístico (análise de variância e análise de desvio para os efeitos principais, Tukey, Dunnett, diferenças menos significativas, etc. para o post-hoc).
- Aprender os métodos mais comuns para selecionar o melhor modelo de entre uma série de ajustes candidatos aos mesmos dados.

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 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 ESTADÍSTICA (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 28 e um índice i10 de 48.