

Tópicos Especiais em Ecologia Aplicada: Exploração de dados ecológicos

Disciplina: Tópicos Especiais em Ecologia Aplicada: Exploração de dados ecológicos

Professor: Alan M. Tonin

Carga Horária: 2 créditos

Período/hora das aulas: condensada em uma semana das 9h às 18h

Local das aulas: Laboratório de Informática 1 do IB

Objetivos

Essa disciplina tem como objetivo principal proporcionar aos estudantes noções básicas da importância da exploração de dados (*a priori* da execução de uma análise) a fim de evitar erros estatísticos muito comuns em estudos ecológicos, os quais podem inclusive alterar as conclusões do estudo. Nesta disciplina será enfatizado a utilização de diferentes ferramentas (a maioria gráficas) para verificar a adequação/violação das premissas das análises mais comumente utilizadas em Ecologia (p.ex., regressão linear, análise de variância, teste-t).

Ementa

Estudos observacionais vs. experimentais; variável resposta e preditores; variáveis discretas e contínuas; análises uni ou multivariadas; conceitos básicos de estatística (independência, replicação, erro amostral); distribuição de probabilidades (Normal, Poisson, Binomial, Binomial negativa, Gamma); introdução à exploração de dados (ferramentas gráficas; *Cleveland dotplots*, *pairplots*, *boxplots*, *scatterplots*, *coplots* e *multipanel scatterplots*); detecção de *outliers* (exploração visual e testes); normalidade e homogeneidade (exploração visual e testes); relação entre as variáveis (linear, não-linear); colinearidade (visualização e quantificação); independência das observações; transformação de dados.

Programa*

AULA 1 – SEGUNDA-FEIRA

- Apresentação da disciplina
- Conceitos básicos (tipos de estudos, variáveis e análises)
- Conceitos básicos de estatística

AULA 2 – TERÇA-FEIRA

- Distribuição de probabilidades
- Noções gerais para uso do R
- Introdução à exploração de dados (com exercícios em aula)

AULA 3 – QUARTA-FEIRA

- Detecção de *outliers*
- Normalidade
- Homogeneidade

AULA 4 – QUINTA-FEIRA

- Relação entre as variáveis
- Colinearidade
- Independência das observações
- Transformação dos dados (vantagens e desvantagens)

AULA 5 – SEXTA-FEIRA

- Apresentação e discussão dos trabalhos finais
- Revisão dos tópicos mais importantes da disciplina
- Fechamento da disciplina

* O programa pode sofrer modificações na ordem de apresentação dos conteúdos.

Metodologia de ensino

A disciplina consistirá em aulas expositivas com a base teórica e exemplificação dos conteúdos programados. Os exemplos serão principalmente apresentados em R. Discussão e resolução de exercícios e dos trabalhos individuais da aula anterior. Discussão após a apresentação de seminários (referente ao trabalho final; ver item 2 abaixo).

Avaliação

- (1) Exercícios e trabalhos individuais para serem executados em horário de aula ou extraclasse e entregues na aula seguinte;
- (2) Trabalho individual final com a exploração dos dados dos projetos de mestrado/doutorado. Entregar uma versão escrita sucinta e apresentação de seminário.

Bibliografia básica

a) Exploração de dados – *protocolos com exemplos e códigos de R

- Ieno, E. N., & Zuur, A. F. (2015). A Beginner's guide to data exploration and visualisation with R. Highland Statistics Limited.
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution*, 1: 3-14.
- *Chapter 2 - Limitations of Linear Regression Applied on Ecological Data*: Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. 2009. Mixed effects models and extensions in ecology with R. Springer, New York, NY.
- Krzywinski, M., & Altman, N. (2014). Points of significance: visualizing samples with box plots.

b) Ferramenta para exploração de dados online

- BoxPlotR: a web-tool for generation of box plots: <http://shiny.chemgrid.org/boxplotr/>
- Spitzer, M., Wildenhain, J., Rappsilber, J., & Tyers, M. (2014). BoxPlotR: a web tool for generation of box plots. *Nature methods*, 11(2), 121.

c) Independência e pseudoreplicação

- Colegrave, N., & Ruxton, G. D. (2017). Using Biological Insight and Pragmatism When Thinking about Pseudoreplication. *Trends in ecology & evolution*.

d) Distribuição de probabilidades

- *Chapter 8 - Meet the exponential family*: Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. 2009. Mixed effects models and extensions in ecology with R. Springer, New York, NY.

e) Outliers

- Altman, N., & Krzywinski, M. (2016). Analyzing outliers: influential or nuisance? *Nature Methods*, 13: 535-535.

f) Introdução ao uso do R

- Zuur, A., Ieno, E. N., & Meesters, E. 2009. A Beginner's Guide to R. Springer Science & Business Media.

g) Conceitos básicos de estatística e design experimental

- Altman, N., & Krzywinski, M. 2015. Points of Significance: Association, correlation and causation. *Nature Methods* 12: 899–900.
- Altman, N., & Krzywinski, M. 2014. Points of significance: Sources of variation. *Nature Methods* 12: 5–6.
- Altman, N., & Krzywinski, M. 2015. Points of Significance: Simple linear regression. *Nature Methods* 12: 999–1000.
- Crawley, M.J. 2015. *Statistics: An introduction using R*. Second edition, Wiley, UK.

Tópicos Especiais em Ecologia Aplicada: Exploração de dados ecológicos

- Krzywinski, M., & Altman, N. 2014. Points of significance: Designing comparative experiments. *Nature Methods* 11: 597–598.
- Vaux D.L. 2012. Research methods: Know when your numbers are significant. *Nature* 492: 180–181 (Comment)