Análises Ecológicas no R

2021-08-11

# Capa

# Prefácio

# 1 Pré-requisitos

## 1.1 Introdução

O objetivo desta seção é

## 1.2 Pacotes

## 1.3 Versão do R

## 1.4 Dados

# 2 Cap. 2 - Introdução

## 2.1 Histórico deste livro

Este livro foi estruturado a partir da apostila elaborada pelos pesquisadores Diogo B. Provete, Fernando R. da Silva e Thiago Gonçalves-Souza para ministrar o curso *Estatística aplicada à ecologia usando o R* no PPG em Biologia Animal da UNESP de São José Rio Preto/SP, em abril de 2011. Os três pesquisadores eram então alunos do PPG em Biologia Animal quando elaboraram o material disponibilizado na [apostila]((https://cran.r-project.org/doc/contrib/Provete-Estatistica_aplicada.pdf)). A proposta de transformar a apostila em livro sempre foi um tópico recorrente desde 2011, e concretizado agora, 10 anos depois.

Neste período, Diogo, Fernando e Thiago foram contratados pela Universidade Federal de Mato Grosso do Sul, Universidade Federal de São Carlos campus Sorocaba, e Universidade Federal Rural de Pernambuco, respectivamente. Nestes anos eles ofertaram diferentes versões do curso *Estatística aplicada à ecologia usando o R* para alunos de graduação e pós-graduação em diferentes instituições do Brasil. A possibilidade da oferta destes novos cursos fortaleceu a ideia de trasformar a apostila em um livro com base nas experiências dos pesquisadores em sala de aula.

Considerando que novas abordagens ecológicas vêm sendo descritas e criadas a uma taxa elevada nos últimos anos, era de se esperar que as informações disponíveis na apostila estivessem defasadas após 10 anos. Por este motivo, Diogo, Fernando e Thiago convidaram outros dois pesquisadores, Gustavo B. Paterno da Universität München e Maurício H. Vancine do PPG em Ecologia, Evolução e Biodiversidade da UNESP Câmpus de Rio Claro, que são referências no uso de estatística em ecologia usando o R. Com o time completo, passaram mais de um ano realizando reuniões, compartilhando scripts e pagando cerveja para os coautores por capítulos atrasados até chegarem neste primeira versão do livro.

## 2.2 Objetivo deste livro

Nossa proposta com este livro é de traçar o melhor caminho (pelo menos do nosso ponto de vista) entre questões ecológicas e os métodos estatísticos mais robustos para testá-las. Guiar seus passos nesse caminho (nem sempre linear) necessita que você utilize um requisito básico: o de utilizar seu esforço para caminhar. O nosso esforço, em contrapartida, será o de indicar as melhores direções para que você adquira certa independência em análises ecológicas. Um dos nossos objetivos é mostrar que o conhecimento de teorias ecológicas e a utilização de questões apropriadas são o primeiro passo na caminhada rumo à compreensão da lógica estatística. Não deixe que a estatística se torne a “pedra no seu caminho.” Em nossa opinião, programas com ambiente de programação favorecem o entendimento da lógica estatística, uma vez que cada passo (lembre-se de que você está caminhado em uma estrada desconhecida e cheia de pedras) precisa ser coordenado, ou seja, as linhas de comando (detalhes abaixo) precisam ser compreendidas para que você teste suas hipóteses.

A primeira parte deste livro pretende utilizar uma estratégia que facilita a escolha do teste estatístico apropriado, por meio da seleção de questões/hipóteses claras e da ligação dessas hipóteses com a teoria e o método (veja Figura 1 no @[cap3]). Enfatizamos que é fundamental ter em mente aonde se quer chegar, para poder escolher o que deve ser feito. Posteriormente à escolha de suas questões, é necessário transferir o contexto ecológico para um contexto meramente estatístico (hipótese nula/alternativa). A partir da definição de uma hipótese nula, partiremos para a aplicação de cada teste estatístico (de modelos lineares generalizados à análises multivariadas) utilizando a linguagem R.

Antes de detalhar cada análise estatística, apresentaremos os comandos básicos para a utilização da linguagem R e os tipos de distribuição estatística que são essenciais para a compreensão dos testes estatísticos. Para isso, organizamos um esquema que chamamos de “estrutura lógica” que facilita a compreensão dos passos necessários para testar suas hipóteses (veja Figura 2 @[cap3]).

## 2.3 O que você não encontrará neste livro

Aprofundamento teórico, detalhes matemáticos, e explicação dos algoritmos são informações que infelizmente não serão abordadas neste livro. O foco aqui é a explicação de como cada teste funciona (teoria e procedimentos matemáticos básicos) e sua aplicação em testes ecológicos usando scripts na linguagem R. Para tanto, o livro de Pierre e Louis Legendre ([Legendre and Legendre 2012](#ref-legendre_numerical_2012)) é uma leitura que permite o aprofundamento de cada uma das análises multivariadas propostas aqui. Além disso, são de fundamental importância para o amadurecimento em análises ecológicas as seguintes leituras: [Manly](#ref-manly_randomization_1991) ([1991](#ref-manly_randomization_1991)), [Pinheiro and Bates](#ref-pinheiro_mixed-effects_2000) ([2000](#ref-pinheiro_mixed-effects_2000)), [Scheiner and Gurevitch](#ref-scheiner_design_2001) ([2001](#ref-scheiner_design_2001)), [Burnham and Anderson](#ref-burnham_pvalues_2014) ([2014](#ref-burnham_pvalues_2014)), [Quinn and Keough](#ref-quinn_experimental_2002) ([2002](#ref-quinn_experimental_2002)), [Venables and Ripley](#ref-venables_modern_2002) ([2002](#ref-venables_modern_2002)), [Magurran and McGill](#ref-magurran_biological_2011) ([2011](#ref-magurran_biological_2011)), [Gotelli and Ellison](#ref-gotelli_primer_2013) ([2013](#ref-gotelli_primer_2013)), [Zar](#ref-zar_biostatistical_2010) ([2010](#ref-zar_biostatistical_2010)), [Zuur, Ieno, and Elphick](#ref-zuur_protocol_2009) ([2009](#ref-zuur_protocol_2009)), [Crawley](#ref-crawley_r_2012) ([2012](#ref-crawley_r_2012)) e [James et al.](#ref-james_introduction_2013) ([2013](#ref-james_introduction_2013)).

## 2.4 Por que usar o R?

Os criadores do R o chamam de uma linguagem e ambiente de programação estatística e gráfica (Venables et al. 2021). A linguagem R também é chamada de programação “orientada ao objeto” (*object oriented programming*), o que significa que utilizar o R envolve basicamente a criação e manipulação de objetos em uma tela branca, em que o usuário tem de dizer exatamente o que deseja que o programa execute, ao invés de simplesmente clicar em botões. E vem daí uma das grandes vantagens em se usar o R: o usuário tem total controle sobre o que está acontecendo e também tem de compreender o que deseja antes de executar uma análise. Além disso, o R permite integração com outros programas escritos em C++, Python e Java, permitindo que os usuários possam aplicar novas metodologias sem ter que aprender novas linguagens.

Na página pessoal do Prof. Nicolas J. Gotelli ([link](http://www.uvm.edu/~ngotelli/homepage.html)), existem vários conselhos para um estudante iniciante de ecologia. Dentre esses conselhos, o Prof. Gotelli menciona que o domínio de uma linguagem de programação é uma das habilidades mais importantes, porque dá liberdade ao ecólogo para executar tarefas que vão além daquelas disponíveis em pacotes estatísticos comerciais. Além disso, a maioria das novas análises propostas nos mais reconhecidos periódicos em ecologia normalmente são implementadas em linguagem R, e os autores incluem normalmente o código fonte no material suplementar dos artigos, tornando a análise acessível. A partir do momento que essas análises ficam disponíveis (seja por código fornecido pelo autor ou por implementação em pacotes pré-existentes), é mais simples entendermos a lógica de análises complexas, especialmente as multivariadas, utilizando nossos próprios dados, realizando-as passo a passo. Sem a utilização do R, normalmente temos que contatar os autores que nem sempre são tão acessíveis.

Uma última vantagem é que por ser um software livre, a citação do R em artigos é permitida e até aconselhável. Para saber como citar o R, digite citation() na linha de comando. Para citar um pacote específico, digite citation() com o nome do pacote entre aspas dentro dos parênteses. Neste ponto, esperamos ter convencido você leitor, de que aprender a utilizar o R tem inúmeras vantagens. Entretanto, provavelmente vai ser difícil no começo, mas continue e perceberá que o investimento vai valer à pena no futuro.

## 2.5 Indo além da linguagem de progração para a Ecologia

Um ponto em comum em que todos os autores deste livro concordaram em conversas durante sua estruturação, foi a dificuldade que todos tivemos quando estávamos aprendendo a linguagem:

1. Como transcrever os objetivos (manipulação de dados, análises e gráficos) em linguagem R
2. Como interpretar os resultados das análises estatísticas do R para os objetivos ecológicos

Num primeiro momento, quando estamos aprendendo a linguagem R é muito desafiador pensar em como estruturar nossos códigos para que eles façam o que precisamos: importar dados, selecionar linhas ou colunas, qual pacote ou função usar para uma certa análise ou como fazer um gráfico que nas nossas anotações são simples, mas no código parece impossível. Bem, não há um caminho fácil nesse sentido e ele depende muito da experiência e familiaridade adquirida com o tempo de uso da linguagem, assim como outra língua qualquer, como inglês ou espanhol. Entretanto, uma dica pode ajudar: estruture seus códigos antes de partir para o R. Num papel escreva os pontos que quer que seus códigos façam, como se estivesse explicando para alguém os passos que precisa para realizar as tarefas. Depois disso, transcreva para o script (iremos explicar esse conceito no @[cap4]) esse texto. Por fim, traduza isso em linguam R. Pode parecer massante e cansativo no começo, mas isso o ajudará a ter maior domínio da linguagem, sendo que esse passo se torna desnecessário quando se adquire bastante experiência.

Uma vez transposta esse barreira inicial e temos os resultados de nossas análises (valores de estatísticas, parâmetros estimados, valores de p e R², etc.), com gráficos e outras figuras que precisamos, como interpretamos à luz da teoria ecológica? Esse ponto é talvez um dos mais complicados. Com o tempo, ter um valor final de uma estatística ou gráfico à partir da linaguagem R é relativamente simples, mas o que esse valor ou gráfico significam para nossa hipótese ecológica é o ponto mais complicado. Essa dificuldade por ser por inexperiência teórica (ainda não lemos muito sobre um aspecto ecológico) ou inexperiência científica (ainda temos dificuldade para expandir nossos argumentos de forma indutiva). Destacamos esse ponto porque ele é fundamental no processo científico e talvez seja o principal aspecto que diferencia os cientistas de outros profissionais: sua capacidade de entendimento dos padrões à partir dos processos e mecanismos atrelados. Nesse ponto, quase sempre recorremos à nossos orientadores ou colegas mais experientes para nos ajudar, mas é natural e faz parte do processo de aprendizado de uso da linguagem R junto à Ecologia como ciência. Entretanto, contrapomos a importância dessa extrapolação para não nos tornarmos apenas especialistas em linguagem R sem a fundamental capacidade de entendimento do sistema ecológico que estamos estudando.

## 2.6 Como usar este livro

Os conteúdos apresentados em cada capítulo são independentes entre si. Portanto, você pode utilizar este livro de duas formas. A primeira é seguir uma ordem sequencial (capítulos 1, 2, 3, …) que recomendamos, principalmente, para as pessoas que não possuem familiaridade com a linguagem R. A segunda forma, é selecionar o capítulo que contém a análise de seu interesse e mudar de um capítulo para outro sem seguir a sequência apresentada no livro.

Com exceção dos capítulos 3, 4, 5 e 15, os outros capítulos foram elaborados seguindo a mesma estrutura contendo uma descrição da análise estatística (aspecto teóricos) e exemplos relacionados com questões ecólogicas que podem ser respondidas por esta análise. Todos os exemplos são compostos por: i) uma descrição dos dados utilizados, ii) pergunta e predição do trabalho, iii) descrição das variáveis resposta(s) e preditora(s), e iv) descrição e explicação das linhas de comando do R necessárias para realização das análises. Os exemplos utilizados são baseados em dados reais que já foram publicados em artigos científicos ou são dados coletados por um dos autores deste livro. Nós recomendamos que primeiro você utilize estes exemplos para se familiarizar com as análises e a formatação das linhas e colunas das planilhas. Em seguida, utilize seus próprios dados para realizar as análises. Esta é a melhor maneira de se familiarizar com as linhas de comando do R.

Muitas das métricas ou índices apresentados neste livro não foram traduzidas para o português, porque seus acrônimos são clássicos e bem estabelecidos na literatura ecológica. Nestes casos, consideramos que a tradução poderia confundir as pessoas que estão começando a se familiarizar com a literatura específica. Realçamos que não estamos abordando todas as possibilidades disponíveis, e existem muito outros pacotes e funções no R que realizam as mesmas análises. Contudo, esperamos que o conteúdo apresentado permita que os leitores adquiram independência e segurança, para que possam caminhar sozinhos na exploração de novos pacotes e funções para responderem suas perguntas biológicas e ecológicas.

# Referências

Burnham, K. P., and D. R. Anderson. 2014. “Pvalues Are Only an Index to Evidence: 20th- Vs. 21st-Century Statistical Science.” *Ecology* 95 (3): 627–30. <https://doi.org/10.1890/13-1066.1>.

Crawley, M. J. 2012. *The R Book*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118448908>.

Gotelli, N. J., and A. M. Ellison. 2013. *A Primer of Ecological Statistics*. Sinauer Associates.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An Introduction to Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>.

Legendre, P., and L. Legendre. 2012. *Numerical Ecology*. Elsevier. <https://doi.org/10.1016/c2010-0-66470-4>.

Magurran, A. E., and B. J. McGill. 2011. *Biological Diversity: Frontiers in Measurement and Assessment*. Oxford: Oxford University Press.

Manly, B. F. J. 1991. *Randomization and Monte Carlo Methods in Biology*. Chapman & Hall.

Pinheiro, J. C., and D. M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag. <https://doi.org/10.1007/b98882>.

Quinn, G. P., and M. J. Keough. 2002. “Experimental Design and Data Analysis for Biologists.” <https://doi.org/10.1017/cbo9780511806384>.

Scheiner, S. M., and J. Gurevitch. 2001. *Design and Analysis of Ecological Experiments*. Oxford University Press.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Springer-Verlag.

Zar, J. H. 2010. *Biostatistical Analysis*. Pearson.

Zuur, A. F., E. N. Ieno, and C. S. Elphick. 2009. “A Protocol for Data Exploration to Avoid Common Statistical Problems.” *Methods in Ecology and Evolution* 1 (1): 3–14. <https://doi.org/10.1111/j.2041-210x.2009.00001.x>.