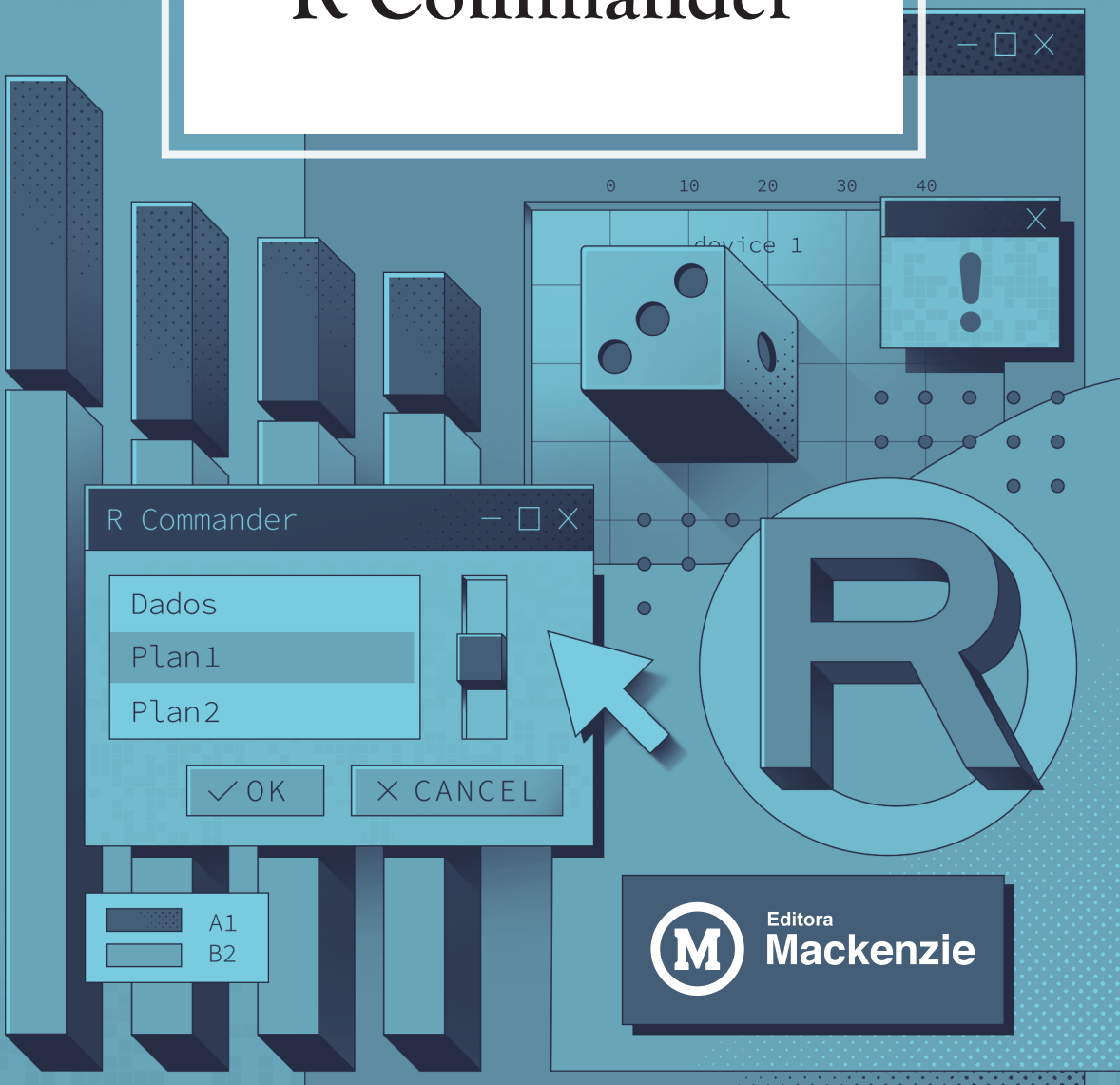
Diógenes de Souza Bido

Análise de dados quantitativos com R Commander



Editora Mackenzie

Análise de dados quantitativos com R Commander

CONEXÃO INICIAL

37

Diógenes de Souza Bido

Análise de dados quantitativos com R Commander

Editora Mackenzie

Coordenação editorial: Jéssica Dametta
Preparação de texto: Jéssica Dametta
Revisão: Pietro Menezes
Diagramação: Pedro Videira Pancheri
Capa: Pedro Videira Pancheri

Dados Internacionais de Catalogação na Publicação (CIP)

B585a Bido, Diógenes de Souza.
Análise de dados quantitativos com R Commander / Diógenes de Souza Bido. – São Paulo : Editora Mackenzie, 2021.
202 p. : il ; 23 cm. – (Coleção Conexão Inicial).

Inclui referência bibliográfica e índice.
ISBN 978-65-5545-262-4

1. Linguagem de programação. 2. Interface gráfica. 3. R Commander. 4. Dados – Análise. I. Título. II. Série.

CDD 005.13

Bibliotecária responsável: Jaqueline Bay Inacio Duarte – CRB 8/9509

EDITORA MACKENZIE
Rua da Consolação, 930
Edifício João Calvino, 6º andar
São Paulo – SP – CEP 01302-907
Tel.: (5511) 2114-8774 (editorial)
editora@mackenzie.br
www.mackenzie.br/editora

Editora afiliada:

Associação Brasileira das Editoras Universitárias

ASSOCIADO Câmara Brasileira do Livro

Àqueles que vieram antes: Argemiro e Maria José (Lia).
Àqueles que torcem e me animam: Denise e Marcos.
Àqueles que vieram depois: Enzo e sobrinhos(as).
Àquela que veio e permanece: Elaine (Naninha).
Àquele que se foi: Daniel A. Moreira (in memoriam).

Agradecimentos

Agradeço o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) na forma de bolsa de produtividade em pesquisa e da Universidade Presbiteriana Mackenzie (UPM) por propiciar a demanda e a oportunidade para a produção desta obra.

A lista de colegas a agradecer é inviável para este espaço, mas preciso citar alguns que me incentivaram (até cobraram, pois comecei a escrever este livro em 2013) e deram dicas e sugestões (às vezes, a sugestão vem como pergunta) que foram fundamentais nas decisões a respeito do conteúdo:

Adilson Aderito da Silva (Mackenzie)
Antonio Carlos de Oliveira Barroso (CNEN-IPEN/SP)
Benedito Dias Baptista Filho (*in memoriam*)
Cesar Alexandre de Souza (FEA-USP)
Darcy Mitiko Mori Hanashiro (Mackenzie)
Dirceu da Silva (Unicamp)
Fábio Frezatti (FEA-USP)
Francisco Henrique Figueiredo de Castro Junior (FEA-USP)
Herbert Kimura (UnB)
Maria Luisa Mendes Teixeira (Mackenzie)
Mario Olímpio de Menezes (CNEN-IPEN/SP e Mackenzie)
Saulo Soares de Souza (Mackenzie)

A oportunidade de participar da Coleção Conexão Inicial e o apoio dos editores Jéssica Dametta Cruz e John Marion Sydenstricker-Neto foram determinantes para que eu conseguisse concluir este livro.

Ao trabalhar nos 68 comentários do(a) avaliador(a) anônimo(a), penso que a qualidade do livro melhorou muito em relação ao material que eu tinha submetido à Editora. Um muito obrigado é pouco e o sentimento de gratidão é grande.

Sumário

Sobre o autor

Diógenes de Souza Bido é livre-docente na área de "Métodos Quantitativos e Informática: Estatística e Pesquisa Operacional" pela Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo (FEA-USP), possui pós-doutorado no Instituto de Pesquisas Energéticas e Nucleares (IPEN), é doutor e mestre em Administração de Empresas pela FEA-USP e graduado em Engenharia Química pela Escola Superior de Química Oswaldo Cruz. É Bolsista de Produtividade em Pesquisa do CNPq (PQ-2) desde 2013 e professor adjunto na Universidade Presbiteriana Mackenzie (UPM), atuando na graduação e no Programa de Pós-Graduação em Administração de Empresas com os seguintes temas: aspectos metodológicos da pesquisa em administração, modelagem de equações estruturais (LISREL e PLS-SEM), análise multivariada, análise de condições necessárias, aprendizagem individual e nas organizações e medidas em aprendizagem organizacional.

Introdução

Este livro foi preparado pensando nos alunos de graduação, mas os alunos de pós-graduação, pesquisadores e consultores que estejam iniciando nos métodos quantitativos também podem se beneficiar da sua leitura e das dicas espalhadas ao longo do texto.

1.1 DEMANDA POR *SOFTWARE* GRÁTIS

Por vários anos usei o SPSS como o *software* de referência para a disciplina de Métodos Quantitativos nos cursos de graduação e pós-graduação em Administração de Empresas. Em primeiro lugar, porque a escola onde trabalho possui mais de 100 licenças e, em segundo lugar, porque é um *software* amigável, com livros didáticos em Português e a preços acessíveis (*e.g.*, BRUNI, 2011).

Porém, eu convivia com a reclamação dos alunos de que o SPSS é pago (e caro), e eles não tinham a possibilidade de estudar em casa. Isso se agrava se lembrarmos que parte dos alunos de mestrado e doutorado são professores e seus alunos também não têm SPSS, nem as escolas onde eles lecionam.

1.2 *SOFTWARE* R COMO UMA OPÇÃO VÁLIDA E ACEITA

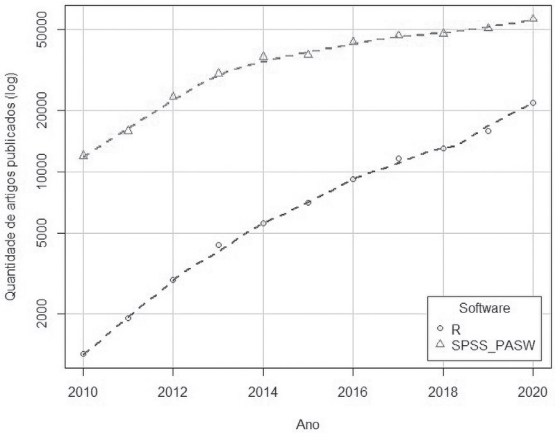
Há alguns anos (ou décadas), SPSS e SAS eram sinônimos de *softwares* para a análise de dados na área de Ciências Sociais Aplicadas, mas atualmente há uma infinidade de opções específicas para cada tipo de análise: SPSS AMOS, Eviews, LISREL, Mplus, Stata, XLSTAT etc.

Talvez você já saiba que o *software* R é grátis, válido e tem sido aceito no meio acadêmico, mas a cada ano ele tem se tornado mais presente nos artigos publicados (pense como uma amostra dos possíveis usos do *software*).

Na Figura 1.1 apresento um levantamento feito no Proquest, base escolhida por permitir a busca no texto do artigo e não apenas nos metadados. Podemos ver que, em 2010, do total de artigos que usaram SPSS ou R, 10%

tinham usado o *software* R, e essa porcentagem vem aumentando ano a ano, chegando a 28% em 2020 e a 50% em 2029, se essa tendência for mantida (SPSS com aumentos decrescentes e R com aumentos crescentes).

Portanto, as análises feitas com *software* R serão mais frequentes nos próximos anos. Se você se interessou por este livro, também deve ter percebido isso.



ano	SPSS e PASW	R	%R
2010	11970	1271	10%
2011	15880	1910	11%
2012	23321	2945	11%
2013	30381	4354	13%
2014	36580	5558	13%
2015	37573	7060	16%
2016	43348	9190	17%
2017	46592	11667	20%
2018	47645	13038	21%
2019	50608	15881	24%
2020	56207	21740	28%

Figura 1.1 – Quantidade de artigos no Proquest que usaram SPSS ou R
Nota: Levantamento feito em 02/01/2021 no Proquest com a seguinte estratégia de busca: campo = *anywhere*; tipo = *peer reviewed*; termos da 1ª busca = SPSS OR PASW; termos da 2ª busca = "software R" OR "R software" OR "R package#".

1.3 MAS NÃO SEI PROGRAMAR

Na Figura 1.2, apresento o console do *software* R e o prompt ">", que assusta muita gente e impede que qualquer conversa sobre a possibilidade de aprender a usá-lo continue. Conheço várias pessoas que chegaram até este ponto, experimentaram o "2+2" no console e pararam por aqui.

Mas veja o que acontece quando executamos esses comandos no console do R:

```
install.packages("Rcmdr", dep=T)
library(Rcmdr)
```

Agora compare o console do *software* R (Figura 1.2) com a interface do pacote R Commander (Figura 1.3). Já não temos o *prompt* inquiridor e o *menu* nos parece mais familiar. Mesmo que ainda não saibamos o que há em cada aba do *menu*, é possível ter uma expectativa bem certeira do que encontraremos: dados, estatísticas, gráficos etc.

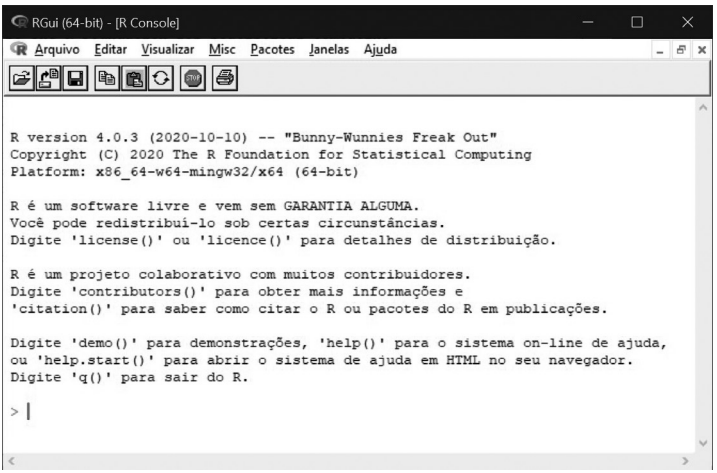


Figura 1.2 – Console do *software* R
Nota: *Command Line Interface* (CLI)

Figura 1.3 – Interface do pacote R Commander
Nota: *Graphical User Interface* (GUI)

Apesar da semelhança com outros *softwares*, o R Commander (ou Rcmdr) tem um campo (R Script) que vai sendo preenchido conforme fazemos nossas análises por meio de cliques no *menu*. Ao abrirmos um conjunto de dados (*dataset*) e gerarmos um gráfico de barras para a variável gênero, também obtemos um *script* como aquele da Figura 1.4.

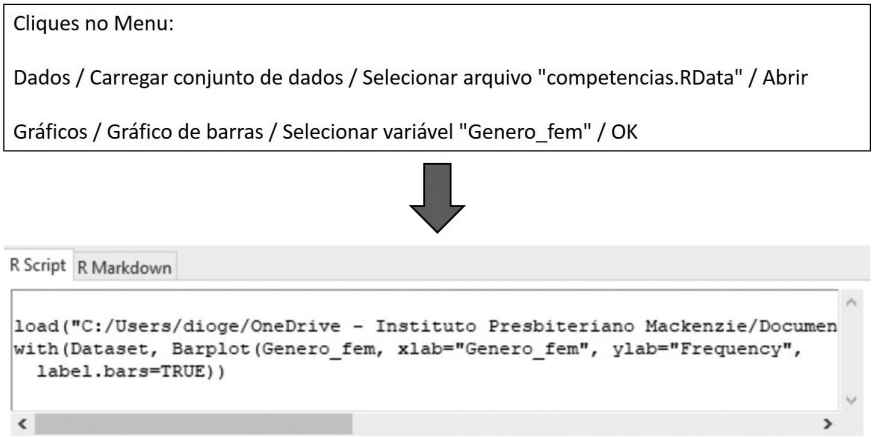


Figura 1.4 – Sequência de comandos (inferior) gerados a partir da interação com os *menus* do Rcmdr (superior)

1.4 POR QUE RCMDR?

Atualmente, há vários *softwares* bons e grátis para a análise de dados, por exemplo: BlueSky, JASP, jamovi, PSPP (é um clone do SPSS), R (https://www.r-project.org/) e a interface gráfica (Rcmdr), que torna o *software* R mais amigável. Mas ela é melhor do que os outros *softwares*?

Para responder a essa questão, no Apêndice A apresento uma comparação entre o Rcmdr e o jamovi, destacando as seguintes características:

- O jamovi é amigável e tem o mínimo necessário para se ministrar disciplinas na graduação. Além disso, possui opções para a análise de componentes principais e análise de fatores comuns, que não estão disponíveis no *menu* do Rcmdr, mas são possíveis por meio de *scripts*.
  - O Rcmdr tem muito mais opções para todas as análises, e aquelas que não estão disponíveis pelo *menu* (cliques) podem ser realizadas por meio do *script*. Conforme o aluno ou pesquisa-

dor vai usando o Rcmdr, ele vai aprendendo a revisar os *scripts* para complementar as análises que deseja e, com o tempo, pode passar a utilizar *scripts* disponibilizados na internet por outros pesquisadores. Nesse ponto, seu potencial de análises vai aumentando e não fica limitado pelo *software*. Em junho de 2021, havia 17.739 pacotes para o *software* R (https://vps.fmvz.usp.br/CRAN/). Clique do lado esquerdo em "Packages" para ter a lista por nomes ou por assuntos (melhor).

Outra vantagem de se usar o *software* R e Rcmdr em relação aos outros *softwares* pagos ou gratuitos é que há muito material de boa qualidade na internet, dezenas (ou milhares?) de cursos *on-line* (Apêndice A), além de *scripts* e pacotes específicos para cada tipo de análise, o que facilita as análises e o aprendizado.

Outra vantagem do Rcmdr é que começamos a manipular os *scripts* (copiar, colar, revisar, salvar) e, em pouco tempo, aprendemos a realizar análises que nem estavam disponíveis pelo *menu* do *software*. Esse aprendizado alavanca novos aprendizados, por exemplo, passamos a pensar nos resultados como *inputs* para novas análises e não em algo acabado.

1.5 APRESENTAÇÃO

Este livro não tem as demonstrações comuns de um livro de estatística básico, porque foi assumido que os cálculos serão feitos pelo *software*. O objetivo aqui é explicar as análises estatísticas respondendo às seguintes questões: para que serve? Quais cuidados devem ser tomados (suposições a serem atendidas)? Como fazer? Como interpretar os resultados? Como relatar os resultados?

Por isso, em cada capítulo, procuro responder a essas questões, por meio de exemplos resolvidos passo a passo, e proponho alguns exercícios, mas sugiro que você tenha ao seu lado algum livro tradicional de estatística, que ensine os cálculos feitos pelo *software*.

Nos Apêndices B e C, explico como instalar o pacote Rcmdr e seus *plugins* e apresento os *plugins* que foram mais usados neste livro.

Como os pacotes R contêm conjuntos de dados (*dataset*), que são usados como exemplos, apresento no Apêndice D como ter acesso a milhares

de conjuntos de dados, todos organizados, com descrição das variáveis e prontos para uso no R Commander. Isso pode ser útil para a preparação de exercícios para uso em aulas ou avaliações da aprendizagem.

1.6 DICAS PARA ESTUDO

O sumário já dá uma boa ideia dos conteúdos que serão tratados neste livro, mas as formas de lidar com esses conteúdos podem ser as mais variadas, dependendo do conhecimento prévio que o leitor tenha. Por exemplo: para os cursos introdutórios (Estatística descritiva): capítulos de 1 a 4; cursos intermediários (Estatística inferencial): capítulos de 1 a 8; e o livro todo para os cursos mais avançados e pessoas autodidatas.

Refaça os exemplos com o conjunto de dados que está disponível para *download* em Bido (2020): https://doi.org/10.5281/zenodo.4370760

Alguns vídeos foram disponibilizados na *playlist* "R Commander (pacote Rcmdr do R)" do meu canal no YouTube, e pretendo acrescentar novos, à medida que receber dúvidas e sugestões: https://bit.ly/3wecIyF

Espero que você goste deste livro, mas ainda há melhorias a serem feitas. Aguardo suas sugestões para a próxima edição.

Muito obrigado.

Diógenes de Souza Bido
diogenesbido@yahoo.com.br

ANÁLISE DE DADOS QUANTITATIVOS COM R COMMANDER

O livro apresenta uma sequência de conteúdos sobre o *software* devidamente acomodados para dar ao leitor a confiança necessária ao passar de um capítulo ao seguinte. Conceitos básicos, que fundamentam os tópicos mais avançados, são trabalhados gradualmente, com muitos exemplos e exercícios, além de referências que ampliam e complementam cada etapa. A onda da "ciência dos dados", resultado da necessidade de extrair conhecimento da avalanche de dados disponíveis nas organizações, tem colocado em evidência o poder das ferramentas livres, representadas aqui pelo *software* R e pelo pacote Rcmdr com seus *plugins*.

Editora Mackenzie

ISBN 978–65–5545–262–4
9 786555 452624