

APRESENTAÇÃO

Este manual de Introdução à Estatística para Linguistas¹ foi criado originalmente como um curso a ser oferecido no Programa de Pós-Graduação em Linguística do Instituto de Estudos da Linguagem da UNICAMP em 2017. Sua primeira versão foi escrita no pacote *swirl*, um ambiente didático, dentro da plataforma R, para a aprendizagem dessa linguagem de programação a partir de sua aplicação prática. Desde então, o curso também tem sido ofertado como curso de extensão na UNICAMP, em dezenas de minicursos em programas de pós-graduação em todo o Brasil e como curso *on-line* assíncrono na Plataforma EAD da ABRALIN.²

A apresentação deste tutorial em formato de *e-book* tem o objetivo de torná-lo acessível a um público mais amplo, assim como facilitar sua consulta sempre que necessário. Aos interessados e às interessadas em se iniciar na análise quantitativa de dados linguísticos, recomendo seguir o curso em seu formato *swirl*, dentro da plataforma R (ver Lição 0). Dentro do ambiente didático do *swirl*, o usuário digita as respostas às questões e imediatamente tem um retorno sobre sua precisão – se a resposta está incorreta, o *swirl* pede que se tente novamente. Neste *e-book*, o texto vem acompanhado do código que pode ser reproduzido dentro do R, de modo que também é possível segui-lo fora do ambiente *swirl*.

Tenho testemunhado, nos últimos anos, uma procura cada vez maior por materiais e discussões sobre tratamento estatístico de dados linguísticos. Contudo, até o momento, a formação em Estatística não é componente obrigatório da maior parte dos cursos de Letras e de Linguística, e tem de ser buscada por iniciativa própria de

¹ Esta versão do manual corresponde à versão 2.0.3 do curso *Introdução à Estatística para Linguistas*, de 12 mai. 2021 (disponível em <https://doi.org/10.5281/zenodo.4755739>). Este *e-book* foi escrito em formato *.Rmd*, na versão 4.1.1 “Kick Things” (10/08/2021) do R e 2021.09.0+351 “Ghost Orchid” (20/09/2021) da interface RStudio.

² Disponível em <https://ead.abralin.org/>. Último acesso em 14 mai. 2022.

pesquisadores, professores e estudantes de pós-graduação ou graduação. O presente tutorial foi pensado para que possa ser útil mesmo em situação de estudo autônomo, mas seu aproveitamento certamente será melhor se o estudo for realizado em grupos de estudo ou em cursos regulares.

Linguistas precisam de Estatística?

Sempre que se lida com variação, o tratamento estatístico de dados é bem-vindo. Diversos aspectos linguísticos são variáveis, na Fonética, Fonologia, Morfologia, Sintaxe, Léxico, Discurso etc. Por exemplo, nenhum falante produz uma vogal /e/ exatamente do mesmo modo que outros falantes – com efeito, nem o mesmo indivíduo o faz da mesma maneira em todas as ocasiões. Até as intuições dos falantes são variáveis quanto à gramaticalidade, adequação ou interpretação de certas estruturas em determinados contextos. O linguista que trabalha com dados empíricos (de usos, avaliações, julgamentos de adequação, gramaticalidade, tempos de reação etc.) inevitavelmente se depara com variação.

Análises estatísticas não concorrem com, nem substituem análises qualitativas: ambas são importantes para uma compreensão mais global do funcionamento das línguas naturais. O objetivo da análise estatística não é – e nunca deveria ser! – a quantificação de dados por si só. As análises estatísticas têm três propósitos principais: resumir, explicar e prever.

Quando se lida com variação, é desejável ter uma amostra representativa do que se está analisando, o que muitas vezes implica um conjunto de dados maior do que aquilo de que somos capazes de fazer sentido “a olho (ou ouvido?) nu”, digamos. Com mais de 50, 100 dados, dificilmente é possível depreender um padrão mais geral que não seja impressionístico, sob forte risco de viés de acordo com nossas próprias expectativas ou previsões. Por meio de gráficos ou tabelas, é possível resumir um conjunto extenso de

dados para dimensões inteligíveis, sobre as quais podemos fazer alguma afirmação mais segura.

Tabelas e gráficos permitem descrever o que ocorre nos dados, mas isoladamente não permitem explicá-los: a explicação sobre certos padrões deve sempre estar alinhada às teorias e aos modelos linguísticos com que se trabalha. Análises estatísticas, no entanto, podem auxiliar na interpretação de padrões ao revelar correlações entre aspectos concomitantes. É claro que encontrar uma correlação não é sinônimo de ter encontrado a causa motivadora para determinado fenômeno, mas perceber que certas variáveis coocorrem sistematicamente pode fornecer pistas valiosas para explicar o que está por trás das estruturas observadas. Constatar, por exemplo, que a vogal pretônica /e/ é realizada mais frequentemente como [i] quando a sílaba seguinte contém uma vogal [+alta] (como *menino*) do que quando contém uma vogal [-alta] (como em *metade*), mesmo que essa relação não seja categórica (também é possível dizer *t[i]atro* para *teatro*), diz-nos algo sobre os fatores possivelmente em operação na articulação de vogais pretônicas.

Prever, por sua vez, não significa (necessariamente) determinar o que vai acontecer no futuro, próximo ou distante, mas ser capaz de extrapolar as conclusões para além do conjunto de dados específicos observados. Afinal, na maior parte das vezes, não temos interesse em fazer afirmações apenas sobre a língua falada pelas poucas dezenas de voluntários de um estudo, mas sobre a língua de um conjunto muito maior de falantes, e possivelmente – em comparação com outros trabalhos – sobre o funcionamento das línguas de modo geral. Para fazer previsões, é necessário criar modelos plausíveis, que simulam o funcionamento do que está sendo investigado numa escala manejável para nossa compreensão. Por exemplo, em Cinemática, a fórmula para o movimento retilíneo uniforme ($v = \Delta S / \Delta t$, ou seja, a velocidade média é igual à razão entre o deslocamento espacial e a diferença entre tempo final e inicial) permite prever que um objeto em movimento a 100 km/h, em uma hora, terá se deslocado 100 km; ou que minha viagem de São Paulo a Campinas deve levar cerca de uma hora, se eu mantiver a velocidade média de 100 km/h, ainda que, na prática, eu dificilmente terei mantido a velocidade constante

durante todo o percurso, tampouco terei me deslocado apenas retilmente do ponto inicial ao final. Só é possível fazer previsões a respeito de fenômenos que não são totalmente aleatórios – e a língua, certamente, é um sistema estruturado.

Os manuais básicos de Estatística costumam fazer uma divisão entre Estatística Descritiva e Estatística Inferencial, estrutura que o presente livro também segue, mas vale reforçar que não se trata de análises separadas quando se trata de um mesmo conjunto de dados. Todas as análises devem convergir para que se possa entender o que está acontecendo nos dados e se possa fazer afirmações relevantes – em nosso caso, sobre o funcionamento linguístico. Análises estatísticas, quando feitas conscientemente, permitem detectar padrões mais ou menos robustos, estimar o peso de diferentes fatores sobre determinado fenômeno variável (ou seja, não exagerar o papel de certos fatores frente a outros), e evitar generalizações equivocadas. Não é raro ainda hoje encontrar trabalhos que insistem em afirmar que, por exemplo, as mulheres favoreceram o uso da forma padrão apenas com base em proporções, mesmo quando não se verificou uma diferença significativa a partir de modelo de regressão. Aprender a desenvolver análises estatísticas implica entender e respeitar os passos da análise científica, mesmo que isso contrarie as expectativas iniciais.

Além do devido manejo de dados empíricos, conhecer diferentes tipos de representação gráfica, sistematização de dados, testes e modelos estatísticos fornece ao pesquisador maior liberdade para condução de suas pesquisas, buscando novas questões. Em muitas pesquisas verifica-se a reprodução de uma mesma pergunta que é feita repetidas vezes, apenas com pequenos ajustes. Para dar um exemplo de minha própria área de pesquisa, muitos estudos sociolinguísticos se voltam à descrição de padrões de uma variável “X” (p.ex., concordância nominal, pronúncia de /r/ em coda, vogais médias pretônicas etc.) em uma comunidade “Y” (em geral, definida como uma cidade, mas também por agrupamentos menores como comunidades de práticas); sociolinguistas interessados em outros tipos de análises, como de avaliações ou percepções, muitas vezes acabam não as realizando por desconhecimento de como coletar e tratar os dados. Desse modo, é de grande valia buscar novos tipos de análise, mesmo que não se apliquem no

atual conjunto de dados com que se está trabalhando, com vistas ao alargamento de possibilidades de análises futuras e ao desenvolvimento do campo.

Antes das análises estatísticas

As lições nesta obra pressupõem que o linguista já está pronto para realizar a análise estatística de seus dados. Entretanto, antes de chegar a esse ponto, o pesquisador certamente terá percorrido uma lista de tarefas, que inclui a coleta de dados e sua organização em uma planilha. Nesse sentido, embora tratar dessas tarefas não seja o intuito da obra, vale aqui fazer alguns breves apontamentos sobre aspectos dessas etapas que são pertinentes às posteriores análises quantitativas.

Análises quantitativas quase sempre operam sobre *amostras* de dados, ou seja, uma parcela de um universo bastante maior do que poderia ter sido coletado: a *população*. Em Estatística, *população* é um termo técnico, que se refere ao conjunto total de itens pertinentes ao objeto sob estudo – ou seja, não se trata necessariamente de pessoas. Por exemplo, em um estudo sobre a realização variável de /r/ em posição de coda silábica na fala de paulistanos, a população é o conjunto de todas as ocorrências de /r/ em final de sílaba – *porta, caderno, mulher, celular, porque* etc. – produzidas por pessoas que nasceram na cidade de São Paulo. Em um caso como este, é evidente que seria impossível trabalhar com toda a população de dados, pois os paulistanos produziram, produzem e continuarão produzindo outras tantas ocorrências de /r/ em coda para além da janela de observação do pesquisador.

As perguntas que surgem, então, são “de quantos dados preciso, no mínimo?” ou “de quantos participantes preciso, no mínimo?”. Não há resposta única para essas perguntas, pois o cálculo depende do grau de variabilidade do objeto de estudo, do número de hipóteses (operacionalizadas em variáveis previsoras) que se deseja testar, e também de questões práticas à pesquisa, como o tempo de que se dispõe para recrutamento de participantes e coleta de dados, capacidade de processamento dos dados

coletados etc. No entanto, um caminho para estimar o tamanho mínimo de uma amostra é realizar um teste de *poder estatístico* (ver p.ex., Crawley 2013).

Por outro lado, também é importante considerar a *representatividade* da amostra. Se se coletam dados apenas com falantes paulistanos, ou apenas com falantes universitários, há que se ter cautela ao generalizar os resultados para o “Português Brasileiro”, a depender do que se está examinando. Ao mesmo tempo, a possibilidade de se chegar a generalizações, ainda que mais modestas, sobre o “português paulistano” ou sobre o “português culto” depende do modo de recrutamento dos participantes: a amostra será tanto mais representativa se tiver sido aleatória, ou seja, se todos os itens da população tiveram a mesma chance de ser selecionados. Ora, sabemos que este dificilmente é o caso em estudos linguísticos: em geral, coletam-se dados de voluntários razoavelmente próximos de nossas redes sociais.

A dificuldade em se obter amostras verdadeiramente aleatórias e representativas, contudo, não implica que não podemos confiar em nossos resultados. A *replicabilidade* e a *comparação* entre resultados de diferentes estudos são os aspectos acadêmicos que permitem uma avaliação constante da validade das pesquisas. Para esse fim, é imprescindível que os pesquisadores reportem os passos e as decisões tomadas, desde a coleta de dados até seu processamento e análise, para que se possa avaliar a comparabilidade entre estudos e reproduzir procedimentos.

Uma vez que os dados estejam coletados, é necessário sistematizá-los para que possam ser facilmente manipulados e processados. A melhor maneira de organizar os dados digitalmente é em planilhas eletrônicas, em programas como o Excel ou o Calc, que permitem que os dados sejam exportados em diversos formatos (.xlsx, .csv, .txt etc.). Recomendo seguir o formato “caso por variável”³ apresentado por Gries (2019, p. 29–30) – também chamado de formato *long table* –, cujas convenções são as seguintes:

- A primeira linha contém os nomes das variáveis;

³ Ver Lição 3 para os conceitos de variável, variantes, níveis de variáveis, tipos de variáveis.

- Cada linha representa um e apenas um dado (uma observação da variável dependente);
- A primeira coluna enumera as n observações de 1 a n ;
- Cada uma das colunas representa uma e apenas uma variável;
- Dados/observações faltantes são indicados por “NA”, e não por células vazias.

Além disso, Gries (2019) sugere que se utilize caixa alta para o nome das variáveis; que variáveis nominais nunca sejam codificadas com números (p.ex., para *primeira faixa etária*, é preferível usar “1a” em vez de “1”); e que nunca se utilizem caracteres especiais como espaço, vírgula, marca de tabulação, #, aspas, acentos para codificar variáveis ou variantes/níveis de variáveis.

Cabe notar que quando se coletam dados por meio de questionários *on-line*, o formato acima descrito normalmente não é aquele gerado automaticamente pelas plataformas; neles, cada linha representa os dados de cada formulário respondido (em geral, um para cada participante) – formato conhecido como *wide table*. É possível fazer a conversão do formato *wide* para *long* por meio do pacote *tidyr* do R e, para isso, há vários tutoriais na Internet.⁴

A plataforma R e o ambiente swirl

O R é uma linguagem de programação voltada para computações gráficas e estatísticas.⁵ É uma plataforma livre e gratuita, de código aberto, disponível para Linux, MacOS e Windows, e seus usuários se dispersam pelas mais variadas áreas de conhecimento e aplicação de análise de dados. Especificamente para a Linguística, por meio do R é

⁴ Por exemplo, o *Cookbook for R*: http://www.cookbook-r.com/Manipulating_data/Converting_data_between_wide_and_long_format/. Último acesso em 14 mai. 2022.

⁵ Disponível em <https://cran.r-project.org/>. Último acesso em 14 mai. 2022.

possível, por exemplo, processar dados textuais, compilar e anotar *corpora* de dados, criar concordâncias e listas de frequências, converter arquivos em formato .wav para .mp3, além de realizar análises estatísticas e criar representações gráficas dos dados e dos padrões encontrados pelo pesquisador.

Para utilizar os recursos dessa linguagem, pode-se usar a própria interface do R ou a interface RStudio, também gratuita.⁶ Independentemente da interface que se usar, no entanto, a plataforma é o R, pois não se trata de um programa, mas de uma linguagem de programação.

Em um primeiro contato com o R, o usuário pode estranhar a falta de botões e opções em que clicar. No R, a maior parte daquilo que se quer executar é feito por meio de *linhas de comando*, algo que, em geral, pode parecer mais trabalhoso do que simplesmente clicar sobre comandos pré-programados em botões. No entanto, a execução de comandos por meio de linhas de códigos traz mais vantagens do que desvantagens. Primeiro, justamente por ser uma linguagem de programação, o usuário não está restrito a opções pré-programadas por um *software* (como o GoldVarb X, SPSS, Excel etc.), cujos programadores dificilmente podem ter previsto tudo o que um usuário pode ter interesse em executar. Segundo, ainda que digitar linhas de comando possa parecer trabalhoso, normalmente isso é feito apenas na primeira vez; na maior parte das vezes, trabalha-se com *scripts*, que são conjuntos de linhas de comando salvas em um arquivo, que podem ser reutilizadas ou adaptadas, parcial ou integralmente, em novos usos posteriores. Além disso, programas com botões pré-definidos podem ser antes um risco do que uma vantagem: é grande a chance de que um usuário comece a clicar sobre as opções disponíveis, “para ver no que dá”, e acabe desenvolvendo análises que nem sempre fazem sentido. Embora análises absurdas também possam ser realizadas em linguagem de programação, é menor a chance de que um usuário chegue a certo

⁶ Disponível em <https://www.rstudio.com/products/rstudio/download/>. Último acesso em 14 mai. 2022.

resultado aleatoriamente, pois será necessário conscientizar-se da análise adequada a ser desenvolvida antes de executá-la.

Como sociolinguista, posso relatar o que tenho observado com frequência entre colegas da área, que até a década de 1990 ou a primeira década de 2000 costumavam usar o programa Varbrul, em suas diversas implementações – o GoldVarb X sendo a última. Este é um programa de fácil manuseio e de interface (relativamente) amigável, o que permite começar a usá-lo em poucas horas de treinamento. É inegável que a facilidade de uso do GoldVarb X trouxe muitas vantagens à área da Sociolinguística Variacionista, sendo a principal delas o uso de uma ferramenta unificada de tratamento de dados que torna os resultados, em diferentes estudos, razoavelmente comparáveis. Por outro lado, a maior parte dos usuários do GoldVarb X tem apenas um conhecimento técnico de como manejá-lo: sabe que deve preparar os dados de certo modo, inseri-los no programa em determinada janela, e depois selecionar uma série de opções para que o programa “rode” os dados; no outro extremo, saem os resultados, sem garantias de que sejam confiáveis – qualquer sociolinguista experiente tinha ciência da máxima: *garbage in, garbage out*. O GoldVarb X é uma ferramenta útil, contanto que se saiba o que ela é capaz de fazer e o que não é capaz de fazer. É um programa feito para realizar apenas um tipo de análise: *modelos multivariados fixos de regressão logística com variáveis nominais* – ou seja, a variável dependente deve, necessariamente, ser binária (variante A vs. variante B, como “concordância nominal padrão” vs. “concordância nominal não padrão”),⁷ e as variáveis independentes não podem ser contínuas (como, p.ex., idade do falante: 20, 22, 27 anos), mas devem ser variáveis nominais (como, p.ex., faixas etárias: de 20 a 35, de 36 a 50 etc.). No GoldVarb X, não é possível modelar, adequadamente, variáveis contínuas, como a altura das vogais de acordo com suas medidas de F1; variáveis dependentes/resposta enébricas, com mais de dois níveis de variantes; o papel de efeitos aleatórios, próprios de cada amostra, como falantes individuais; não se pode visualizar

⁷ Embora implementações anteriores do Varbrul permitissem a realização de análises enébricas com o programa MVarb.

os dados em diferentes tipos de gráficos; e não se pode comparar as medidas geradas com resultados de outros estudos, em outros programas, pois os pesos relativos são uma medida usada apenas por sociolinguistas.

No R, é possível realizar todo e qualquer tipo de análise estatística (modelos de regressão linear, análises de componentes principais, testes de qui-quadrado, anova, teste-t, árvores de inferências condicionais etc.). Isso porque, sendo uma linguagem de programação com código aberto, usuários em todo o globo constantemente produzem novas funções, para novos comandos. A instalação base do R vem com um conjunto extenso de funções e várias outras estão disponíveis por meio de pacotes (também chamados de bibliotecas) que contêm conjuntos de funções para aplicações específicas. Vários pacotes são utilizados neste manual de Estatística.

Um deles é o pacote `swirl`, um ambiente didático dentro da plataforma R, criado justamente para que se possa aprender R na prática, usando-o. Este curso de Estatística foi elaborado dentro desse ambiente, de modo que, caso o leitor decida segui-lo no formato `swirl`, é importante saber algumas de suas especificidades.

Após ter instalado o R e os pacotes necessários para este curso (ver Lição 0), o ambiente `swirl` pode ser aberto com os comandos abaixo:

```
library(swirl)
swirl()
```

No início de cada sessão, o `swirl` sempre pedirá para o usuário se identificar; se é o primeiro acesso – que o `swirl` reconhece pela digitação de um nome não usado anteriormente – o `swirl` fornecerá uma série de informações básicas: as reticências na tela indicam que o usuário deve pressionar `ENTER` para continuar; quando há opções numeradas, o usuário deve escolher uma delas (1, 2 etc.); quando o usuário quiser sair temporariamente do ambiente `swirl`, deve digitar `play()`; se quiser voltar ao menu principal, deve digitar `main()`; quando quiser pular uma questão, deve digitar `skip()`. Esta última opção pode ser usada apenas em perguntas para as quais o usuário deve digitar uma linha de comando para prosseguir (não pode ser usada em questões de múltipla escolha). Vale notar que o comando `skip()` dentro do `swirl` fornece a resposta

da pergunta, de modo que é bom tê-lo em mente quando tiver dúvidas sobre como prosseguir na lição. Por esse motivo, este manual não contém uma seção de respostas das lições e dos exercícios. O mais importante, claro, é tomar nota da resposta correta ou esperada, e buscar entendê-la, não apenas copiá-la.

Estrutura da obra

Este livro se encontra dividido em 15 capítulos, organizados em lições temáticas. A Lição 0 contém as instruções iniciais necessárias para o acompanhamento do curso (instalações necessárias e explicações básicas sobre a plataforma R, RStudio e o ambiente didático swirl). As três lições seguintes introduzem os fundamentos da linguagem de programação R (p.ex., como manipular vetores e dataframes, importação de dados ao R) e conceitos básicos de Estatística (o que são variáveis, variantes; tipos de variáveis; sua manipulação dentro do R). As Lições 4 a 7 apresentam os fundamentos da estatística descritiva: a feitura de tabelas; cálculo de medidas estatísticas como proporções, médias e desvio padrão; e elaboração de visualização gráfica de dados em gráficos de barras, de linhas, histogramas, boxplots e de dispersão. A Lição 8 apresenta os conceitos básicos de estatística inferencial: formulação de hipóteses (H_0 e H_1); distribuição normal; significância; erros do Tipo I e do Tipo II; nível- α e intervalo de confiança. As Lições 9 a 11 apresentam testes estatísticos univariados (que correlacionam uma variável dependente a uma variável independente): teste-t, teste de proporção, teste de qui-quadrado e teste de correlação de Pearson, em suas versões paramétricas e não-paramétricas. Por fim, as Lições 12 a 15 discutem a realização de análises de regressão linear e logística, em modelos de efeitos fixos e mistos.

Quadro 1: Estrutura dos capítulos e conteúdos do livro. (Ln): número da lição. VD: variável dependente. VI: variável independente.

| (L1) Fundamentos de R | | | |
|--|--|-------------------|---|
| (L2) Estruturas básicas no R | | | |
| (L3) Tipos de variáveis | VD nominal | VD ordinal | VD numérica |
| Estatística descritiva | (L4) | | (L6) |
| | Tabela de frequências | | Média, mediana, moda |
| | Tabela de proporções | | Desvio padrão, variância |
| | (L5) | (L5) | (L7) |
| | Gráfico de barras | Gráfico de linhas | Gráfico de linhas |
| | ... | ... | Boxplots |
| | | | Histogramas |
| | | | ... |
| (L8) Estatística inferencial | | | |
| Análises univariadas (VD ~ VI) | (L9) | | (L10) |
| | Teste de proporção (VDnom ~ VInom) | | Teste-t / Teste de Wilcoxon (VDnum ~ VInom) |
| | Teste de qui-quadrado (VDnom ~ VInom) | | (L11) |
| | | | Teste de correlação (VDnum ~ VInum) |
| Análises multivariadas (VD ~ VI + VI + VI...) | (L14-L15) | | (L12-L13) |
| | Regressão logística (VDbinária) | | Regressão linear (VDnum) |

O Quadro 1 mostra essa estrutura, que organiza as possibilidades de análises (estatística descritiva ou inferencial, análises uni e multivariadas) fundamentalmente de acordo com o tipo de variável dependente (VD) com que se está trabalhando: nominal, ordinal ou numérica, e secundariamente de acordo com o tipo de variável independente (VI): nominal/nominal binária ou numérica.

Não se trata, evidentemente, de um manual completo de estatística. O objetivo é o de proporcionar ao leitor as ferramentas mínimas necessárias para que possa aplicar o conteúdo das lições a seus próprios dados e para que possa prosseguir os estudos em Estatística em outras obras mais avançadas. Como livro-base, pode ser aplicado em cursos semestrais de introdução à estatística em nível de graduação ou de pós-graduação.

Cada lição se conclui com a indicação de novas leituras para aprofundamento dos conteúdos apresentados e com uma lista de exercícios para reforço e checagem do aprendizado. O texto é escrito de modo a simular uma conversa com o leitor, para que o assunto introduzido – algo que normalmente é objeto de receio dentre nossos alunos de Letras e de Linguística – seja acessível ao público pretendido. Também de modo consciente, evitam-se fórmulas matemáticas em excesso, ao mesmo tempo em que se apresentam referências complementares para o leitor interessado.

Todos os *scripts* estão disponíveis para *download* – você encontrará os links na seção “Links para *scripts* das lições”, nas referências, junto a uma lista de indicações selecionadas para o aprofundamento dos estudos sobre Estatística e sobre a linguagem R. Há também, para cada lição, uma versão em videoaula disponível na playlist <https://www.youtube.com/playlist?list=PLS6gbp0lYlXM5trsXb7RBUG2JR0zkc8Og> em meu canal no YouTube.

Cabe destacar, por fim, que, embora o leitor possa ir direto a um capítulo específico, sua organização pressupõe a compreensão básica das lições que o precedem. A compreensão dessas lições também requer a aplicação do conteúdo na prática: dificilmente se chegará a um entendimento somente através da leitura deste livro; é necessário abrir a plataforma R, reproduzir os exemplos e aplicar as funções aos próprios dados. Boa programação!

Livia Oushiro