

R Statistics: for the Health, Behavioral and Social Sciences

Book Rationale

R has become one of the most widely used statistical tools, in both academia and industry. Most up-to-date R books are addressing data science techniques. Although some of these tools are applicable to other researchers, many of these tools are unnecessary, and potentially overwhelming, for researchers in public health, behavioral, and social sciences. This book, instead, addresses the up-to-date uses of R specifically applied to those fields.

Intended Audience

Beginning or intermediate researchers in the public health, behavioral, and social sciences who are familiar with statistics at a college introductory level or higher. The aim will be to be able to produce efficient, reproducible research projects with R and RStudio¹.

Extras for the Book

I've developed an R package that will go with the book. It will contain the data used in the examples (if not available already in R) and several functions that will help make the learning easier.

Outline

Part I: Data and R

Introduction

- Introduces R and RStudio, how to install and start using (Windows and Mac)
- Discusses the benefits, and weaknesses, of using R and RStudio for research
- Introduces the two main coding schemes in R:²
 - Tidy R (via piping operator “%>%”)
 - Base R

Getting Your Data into R and Where Did It Go

- Demonstrates how to import data into R
- Discusses the differences between Excel/SPSS (where you can see your data) and R (where it is in memory but not always in spreadsheet format for you to see)
 - Also discusses why this is actually a benefit
- Uses base and Tidy R functions to show importing data

How to Clean Up Your Messy Data

- Discusses how to clean up your data in both tidy and base R
 - Missingness
 - Placeholder values
 - Long vs. Wide format of data
- Subsetting your data

¹ RStudio is a useful IDE to access the power that R contains. It is also a free software.

² The two would be presented side by side on a page throughout the book so readers can sense what the similarities and differences are between the two schemes.

Exploring Your Data with ggplot2

- Benefits of exploring data with graphics (especially using flexible graphics in ggplot2)
 - Examples of descriptive statistics that mislead
- Introduces ggplot2 with both base and tidy R
- Data formatting to use ggplot2

Summarizing Your Beautiful Data

- Demonstrates simple methods for summarizing data
 - Aggregating
 - Means and SDs
- Review subsetting
 - Summaries for subsets

Part II: Modeling and R

For each topic below, we will describe the method simply, discuss the types of questions it can answer, interpretation, and then demonstrate its use in R in both base and tidy R schemes using data available online.

Linear Modeling

- Reviews T-tests, ANOVA's, χ^2 tests, correlations and the like
- Introduces linear regression and its relation to ANOVA
 - Assumptions
 - Visualizing the model
- Introduces multiple regression
 - Assumptions
 - Visualizing the model

Multilevel Modeling

- Discusses why multilevel modeling is necessary and its relation to simple and multiple regression
- Introduces common multilevel models
 - GEE
 - Mixed Effects
- Differences between the two main models and their assumptions
- Demonstrate the method in R using some useful packages

Survival Modeling (Time-to-Event)

- Discusses the benefits of survival models and when to use them
 - Assumptions and visualizing via KM plots

Structural Equation Modeling

- Introduces common SEM modeling techniques in R, including:
 - Exploratory and Confirmatory Factor Analysis
 - Latent Class Analysis
 - Mediation and Moderation Models

Statistical Learning and R

- Introduces ridge and LASSO regression techniques
- Introduces Random Forests and how it can answer questions relating to health, behavior, or social influences

- Mentions other machine learning tools and packages in R that allow for its use (but not in depth).

Part III: You as a Super Hero and R

Publishable Graphics so You Look Great

- Building on the chapter about exploratory graphics, discusses elements of good graphics (e.g., color scheme, white space, etc.)
 - Demonstrate these principles in ggplot2
 - Ways to touch up the graphic with minor adjustments that add to the aesthetic appeal

Rmarkdown and Other Simple Export Techniques to Save Time

- Demonstrates how easily tables and figures can be integrated into documents
 - Shows simple integration with Excel and Word
- Introduces Rmarkdown in RStudio
 - Shows Rmarkdown's power in producing reproducible reports

Automating Routine Procedures so You Look Like a Super Hero

- Discusses writing functions and more functions so R does the heavy lifting for you
 - Discusses elements of a good function, when to spend time writing one, and how to load it easily into R each session