An Organic R Textbook

Ian Cero, PhD MStat

2021

Contents

1 About **7**

2 Getting Started **9**

2.1 Overview of the R ecosystem 11

2.2 Installation 11

2.3 A tour of RStudio 11

2.4 Your very first analysis 11

3 Basic R **13**

3.1 Writing in R and Rmarkdown 15

3.2 Variables 15

3.3 Vectors 15

3.4 Lists 15

3.5 Dataframes 15

3.6 Functions 15

3.7 Packages 15

3.8 Error messages 15

3.9 Coding Conventions 15

4 Working with ABCD **17**

4.1 Importing ABCD datafiles 17

4.2 A general-purpose ABCD dataset import function 17

4.3 Importing a whole folder of ABCD datasets 17

5 Visualization **19**

5.1 Overview . . . 19
5.2 `ggplot` Mappings . . . 19
5.3 Adding geoms to the plot . . . 19
5.4 Geom Options and Mappings . . . 19
5.5 Saving plots . . . 19
5.6 Other features to look for . . . 19

6 Data Wrangling **21**

6.1 Base R . . . 21
6.2 The new way: `dplyr` . . . 21
6.3 The only data verbs you'll ever need . . . 21

7 Basic hypthothesis tests **23**

7.1 Not your grandfather's statistical tests . . . 23
7.2 Steps 1 & 2 - Load and clean data . . . 23
7.3 Steps 3 & 4 - Create and unpack analysis objects . . . 23

8 Linear Regression **25**

8.1 Steps 1 & 2 - Load and clean the data . . . 25
8.2 Steps 3 & 4 - Fit the model and summarize it . . . 25

9 Logistic Regression **27**

9.1 Steps 1 & 2 - Import and clean data . . . 27
9.2 Steps 3 & 4 - Fit the model and summarize it . . . 27
9.3 Plotting predicted values . . . 27

10 Introduction to machine learning **29**

10.1 Wait, aren't those the same thing? . . . 30
10.2 The `tidymodels` framework . . . 30
10.3 Steps 1 & 2 - Load data and do some very basic cleaning . . . 30
10.4 Step 3 - Split training and test samples . . . 30
10.5 Step 4 - Build a data processing recipe . . . 30

10.6 Step 5 - Extract the preprocessed data 30
10.7 Step 6 - Fit a model 30
10.8 Step 7 - Evaluate the model 30
10.9 Step 8 - Plot your data 30

11 Multilevel Models **31**
11.1 Steps 1 & 2 - Import and clean the data 31
11.2 Steps 3 & 4 - Fit the model and summarize it 31
11.3 Logistic multilevel models 31

12 Structural Equation Modeling **33**
12.1 Step 1 and 2 - Load data and clean 33
12.2 EFA 33
12.3 CFA 33
12.4 SEM 33

13 Strings, should you care? **35**
13.1 A recurring example 35
13.2 Join the datasets 35
13.3 Extract some information 35
13.4 Phone numbers 35
13.5 String interpolation 35

14 Simulations **37**
14.1 Generating fake data 37
14.2 How is this useful? 37
14.3 Re-testing the t-test 37

Chapter 1

About

This is textbook created during live discussion with the Squeglia Research Group of the Medical University of South Carolina.

Chapter 2

Getting Started

Placeholder

2.1 Overview of the R ecosystem

2.2 Installation

2.2.1 Step 1 - Download and install the R language

2.2.2 Step 2 - Download and install the RStudio IDE

2.2.3 Step 3 - Install the `tidyverse` package (optional)

2.3 A tour of RStudio

2.3.1 The console

2.3.2 The environment

2.3.3 The lower right pane

2.4 Your very first analysis

2.4.1 Step 1 - download the data

2.4.2 Step 2 - Make an RStudio Project

2.4.3 Step 3 - Get the data into your project folder

2.4.4 Step 4 - Open an Rmarkdown Notebook

2.4.5 Step 5 - Create a space to code

2.4.6 Step 6 - Write a command to import the data

2.4.7 Step 7 - Look inside the data

2.4.8 Step 8 - Make a histogram

2.4.9 Step 9 - Run a regression

2.4.10 Step 10 - Get a summary of your results

2.4.11 A Look forward

Chapter 3

Basic R

Placeholder

3.1 Writing in R and Rmarkdown

3.1.1 Chatting with R

3.1.2 Rmarkdown tricks

3.1.3 Code blocks

3.2 Variables

3.2.1 The assignment operator

3.2.2 Numerics

3.2.3 Characters

3.2.4 Booleans

3.2.5 Special types

3.3 Vectors

3.4 Lists

3.5 Dataframes

3.5.1 Construction

3.5.2 Built-in dataframes

3.6 Functions

3.6.1 Getting help

3.6.2 Function parameters

3.7 Packages

3.8 Error messages

3.9 Coding Conventions

3.9.1 What should my code look like?

3.9.1.1 Common coding conventions in R

3.9.2 Official coding conventions

Chapter 4

Working with ABCD

Placeholder

4.1 Importing ABCD datafiles

4.1.1 What's different about ABCD?

4.1.2 Walking through ABCD data importation

4.1.3 Copy/paste-able code

4.2 A general-purpose ABCD dataset import function

4.2.1 Getting rid of import messages

4.3 Importing a whole folder of ABCD datasets

4.3.1 As easy strategy that unfortunately won't scale

Chapter 5

Visualization

Placeholder

5.1 Overview

5.2 `ggplot` Mappings

5.3 Adding geoms to the plot

5.4 Geom Options and Mappings

5.5 Saving plots

5.6 Other features to look for

Chapter 6

Data Wrangling

Placeholder

6.1 Base R

6.2 The new way: `dplyr`

6.3 The only data verbs you'll ever need

6.3.1 `rename()`

6.3.2 `select()`

6.3.3 `mutate()`

6.3.4 `group_by()` and `summarize()`

Chapter 7

Basic hypthothesis tests

Placeholder

7.1 Not your grandfather's statistical tests

7.2 Steps 1 & 2 - Load and clean data

7.3 Steps 3 & 4 - Create and unpack analysis objects

7.3.1 Independant samples t-test

7.3.2 Paired t-tests

7.3.3 Correlation

Chapter 8

Linear Regression

Placeholder

8.1 Steps 1 & 2 - Load and clean the data

8.2 Steps 3 & 4 - Fit the model and summarize it

8.2.1 Broom can help export the results

Chapter 9

Logistic Regression

Placeholder

9.1 Steps 1 & 2 - Import and clean data

9.2 Steps 3 & 4 - Fit the model and summarize it

9.2.1 Where are my odds ratios?

9.3 Plotting predicted values

Chapter 10

Introduction to machine learning

Placeholder

10.1 Wait, aren't those the same thing?

10.2 The `tidymodels` framework

10.3 Steps 1 & 2 - Load data and do some very basic cleaning

10.4 Step 3 - Split training and test samples

10.5 Step 4 - Build a data processing recipe

10.6 Step 5 - Extract the preprocessed data

10.7 Step 6 - Fit a model

10.8 Step 7 - Evaluate the model

10.9 Step 8 - Plot your data

Chapter 11

Multilevel Models

Placeholder

11.1 Steps 1 & 2 - Import and clean the data

11.2 Steps 3 & 4 - Fit the model and summarize it

11.2.1 Where are my p-values?

11.2.2 Where are my intraclass correlations (ICCs)?

11.3 Logistic multilevel models

Chapter 12

Structural Equation Modeling

Placeholder

12.1 Step 1 and 2 - Load data and clean

12.2 EFA

12.2.1 Steps 3 & 4

12.3 CFA

12.3.1 Step 3 & 4

12.4 SEM

12.4.1 Steps 3 & 4

12.4.2 Modification indices

Chapter 13

Strings, should you care?

Placeholder

13.1 A recurring example

13.2 Join the datasets

13.2.1 Direction / order of operations

13.3 Extract some information

13.4 Phone numbers

13.4.1 What about using `str_extract_all()`

13.5 String interpolation

Chapter 14

Simulations

Placeholder

14.1 Generating fake data

14.2 How is this useful?

14.3 Re-testing the t-test