

Contents

Foreword xiii

Preface xv

Acknowledgments xix

About the Author xxiii

1 Getting R 1

- 1.1 Downloading R 1
- 1.2 R Version 2
- 1.3 32-bit versus 64-bit 2
- 1.4 Installing 2
- 1.5 Revolution R Community Edition 10
- 1.6 Conclusion 11

2 The R Environment 13

- 2.1 Command Line Interface 14
- 2.2 RStudio 15
- 2.3 Revolution Analytics RPE 26
- 2.4 Conclusion 27

3 R Packages 29

- 3.1 Installing Packages 29
- 3.2 Loading Packages 32
- 3.3 Building a Package 33
- 3.4 Conclusion 33

4 Basics of R 35

- 4.1 Basic Math 35
- 4.2 Variables 36
- 4.3 Data Types 38
- 4.4 Vectors 43
- 4.5 Calling Functions 49
- 4.6 Function Documentation 49
- 4.7 Missing Data 50
- 4.8 Conclusion 51

5	Advanced Data Structures	53
5.1	<code>data.frames</code>	53
5.2	<code>Lists</code>	61
5.3	<code>Matrices</code>	68
5.4	<code>Arrays</code>	71
5.5	<code>Conclusion</code>	72
6	Reading Data into R	73
6.1	<code>Reading CSVs</code>	73
6.2	<code>Excel Data</code>	74
6.3	<code>Reading from Databases</code>	75
6.4	<code>Data from Other Statistical Tools</code>	77
6.5	<code>R Binary Files</code>	77
6.6	<code>Data Included with R</code>	79
6.7	<code>Extract Data from Web Sites</code>	80
6.8	<code>Conclusion</code>	81
7	Statistical Graphics	83
7.1	<code>Base Graphics</code>	83
7.2	<code>ggplot2</code>	86
7.3	<code>Conclusion</code>	98
8	Writing R Functions	99
8.1	<code>Hello, World!</code>	99
8.2	<code>Function Arguments</code>	100
8.3	<code>Return Values</code>	103
8.4	<code>do.call</code>	104
8.5	<code>Conclusion</code>	104
9	Control Statements	105
9.1	<code>if and else</code>	105
9.2	<code>switch</code>	108
9.3	<code>ifelse</code>	109
9.4	<code>Compound Tests</code>	111
9.5	<code>Conclusion</code>	112
10	Loops, the Un-R Way to Iterate	113
10.1	<code>for Loops</code>	113
10.2	<code>while Loops</code>	115

10.3 Controlling Loops 115

10.4 Conclusion 116

11 Group Manipulation 117

11.1 Apply Family 117

11.2 aggregate 120

11.3 plyr 124

11.4 data.table 129

11.5 Conclusion 139

12 Data Reshaping 141

12.1 cbind and rbind 141

12.2 Joins 142

12.3 reshape2 149

12.4 Conclusion 153

13 Manipulating Strings 155

13.1 paste 155

13.2 sprintf 156

13.3 Extracting Text 157

13.4 Regular Expressions 161

13.5 Conclusion 169

14 Probability Distributions 171

14.1 Normal Distribution 171

14.2 Binomial Distribution 176

14.3 Poisson Distribution 182

14.4 Other Distributions 185

14.5 Conclusion 186

15 Basic Statistics 187

15.1 Summary Statistics 187

15.2 Correlation and Covariance 191

15.3 T-Tests 200

15.4 ANOVA 207

15.5 Conclusion 210

16 Linear Models	211
16.1 Simple Linear Regression	211
16.2 Multiple Regression	216
16.3 Conclusion	232
17 Generalized Linear Models	233
17.1 Logistic Regression	233
17.2 Poisson Regression	237
17.3 Other Generalized Linear Models	240
17.4 Survival Analysis	240
17.5 Conclusion	245
18 Model Diagnostics	247
18.1 Residuals	247
18.2 Comparing Models	253
18.3 Cross-Validation	257
18.4 Bootstrap	262
18.5 Stepwise Variable Selection	265
18.6 Conclusion	269
19 Regularization and Shrinkage	271
19.1 Elastic Net	271
19.2 Bayesian Shrinkage	290
19.3 Conclusion	295
20 Nonlinear Models	297
20.1 Nonlinear Least Squares	297
20.2 Splines	300
20.3 Generalized Additive Models	304
20.4 Decision Trees	310
20.5 Random Forests	312
20.6 Conclusion	313
21 Time Series and Autocorrelation	315
21.1 Autoregressive Moving Average	315
21.2 VAR	322

- 21.3 GARCH 327
- 21.4 Conclusion 336

22 Clustering 337

- 22.1 K-means 337
- 22.2 PAM 345
- 22.3 Hierarchical Clustering 352
- 22.4 Conclusion 357

23 Reproducibility, Reports and Slide Shows with knitr 359

- 23.1 Installing a \LaTeX Program 359
- 23.2 \LaTeX Primer 360
- 23.3 Using knitr with \LaTeX 362
- 23.4 Markdown Tips 367
- 23.5 Using knitr and Markdown 368
- 23.6 pandoc 369
- 23.7 Conclusion 371

24 Building R Packages 373

- 24.1 Folder Structure 373
- 24.2 Package Files 373
- 24.3 Package Documentation 380
- 24.4 Checking, Building and Installing 383
- 24.5 Submitting to CRAN 384
- 24.6 C++ Code 384
- 24.7 Conclusion 390

A Real-Life Resources 391

- A.1 Meetups 391
- A.2 Stackoverflow 392
- A.3 Twitter 393
- A.4 Conferences 393
- A.5 Web Sites 393
- A.6 Documents 394
- A.7 Books 394
- A.8 Conclusion 394

xii Contents

B Glossary	395
List of Figures	409
List of Tables	417
General Index	419
Index of Functions	427
Index of Packages	431
Index of People	433
Data Index	435