CONTENTS IN DETAIL

PREFACE **xxiii**

ACKNOWLEDGMENTS **xxv**

INTRODUCTION **xxvii**

A Brief History of R xxvii
About This Book xxviii
Part I: The Language xxviii
Part II: Programming xxix
Part III: Statistics and Probability xxix
Part IV: Statistical Testing and Modeling xxix
Part V: Advanced Graphics xxx
For Students xxx
For Instructors xxxi

PART I
THE LANGUAGE

1
GETTING STARTED **3**

1.1 Obtaining and Installing R from CRAN 3
1.2 Opening R for the First Time 4
1.2.1 Console and Editor Panes 5
1.2.2 Comments 6
1.2.3 Working Directory 7
1.2.4 Installing and Loading R Packages 7
1.2.5 Help Files and Function Documentation 8
1.2.6 Third-Party Editors 11
1.3 Saving Work and Exiting R 11
1.3.1 Workspaces 11
1.3.2 Scripts 12
1.4 Conventions 12
1.4.1 Coding 13
1.4.2 Math and Equation References 14
1.4.3 Exercises 14
Exercise 1.1 14

2
NUMERICS, ARITHMETIC, ASSIGNMENT, AND VECTORS 17

2.1 R for Basic Math 17
2.1.1 Arithmetic 18
2.1.2 Logarithms and Exponentials 19
2.1.3 E-Notation 20
Exercise 2.1 21
2.2 Assigning Objects 21
Exercise 2.2 22
2.3 Vectors 23
2.3.1 Creating a Vector 23
2.3.2 Sequences, Repetition, Sorting, and Lengths 24
Exercise 2.3 27
2.3.3 Subsetting and Element Extraction 28
Exercise 2.4 32
2.3.4 Vector-Oriented Behavior 33
Exercise 2.5 36

3
MATRICES AND ARRAYS 39

3.1 Defining a Matrix 39
3.1.1 Filling Direction 40
3.1.2 Row and Column Bindings 41
3.1.3 Matrix Dimensions 42
3.2 Subsetting 42
3.2.1 Row, Column, and Diagonal Extractions 43
3.2.2 Omitting and Overwriting 44
Exercise 3.1 47
3.3 Matrix Operations and Algebra 47
3.3.1 Matrix Transpose 47
3.3.2 Identity Matrix 48
3.3.3 Scalar Multiple of a Matrix 49
3.3.4 Matrix Addition and Subtraction 49
3.3.5 Matrix Multiplication 50
3.3.6 Matrix Inversion 51
Exercise 3.2 52
3.4 Multidimensional Arrays 52
3.4.1 Definition 53
3.4.2 Subsets, Extractions, and Replacements 55
Exercise 3.3 58

4
NON-NUMERIC VALUES 59

4.1 Logical Values 59
4.1.1 TRUE or FALSE? 60
4.1.2 A Logical Outcome: Relational Operators 60
Exercise 4.1 64

4.1.3 Multiple Comparisons: Logical Operators 64
Exercise 4.2 67
4.1.4 Logicals Are Numbers! 67
4.1.5 Logical Subsetting and Extraction 68
Exercise 4.3 72
4.2 Characters 72
4.2.1 Creating a String 73
4.2.2 Concatenation 74
4.2.3 Escape Sequences 76
4.2.4 Substrings and Matching 77
Exercise 4.4 78
4.3 Factors 79
4.3.1 Identifying Categories 79
4.3.2 Defining and Ordering Levels 82
4.3.3 Combining and Cutting 83
Exercise 4.5 86

5
LISTS AND DATA FRAMES **89**

5.1 Lists of Objects 89
5.1.1 Definition and Component Access 89
5.1.2 Naming 91
5.1.3 Nesting 93
Exercise 5.1 94
5.2 Data Frames 95
5.2.1 Construction 96
5.2.2 Adding Data Columns and Combining Data Frames 98
5.2.3 Logical Record Subsets 100
Exercise 5.2 101

6
SPECIAL VALUES, CLASSES, AND COERCION **103**

6.1 Some Special Values 103
6.1.1 Infinity 104
6.1.2 NaN 106
Exercise 6.1 107
6.1.3 NA 108
6.1.4 NULL 110
Exercise 6.2 113
6.2 Understanding Types, Classes, and Coercion 114
6.2.1 Attributes 114
6.2.2 Object Class 116
6.2.3 Is-Dot Object-Checking Functions 119
6.2.4 As-Dot Coercion Functions 120
Exercise 6.3 125

7
BASIC PLOTTING **127**

7.1 Using plot with Coordinate Vectors ... 127
7.2 Graphical Parameters ... 129
7.2.1 Automatic Plot Types ... 129
7.2.2 Title and Axis Labels ... 130
7.2.3 Color ... 131
7.2.4 Line and Point Appearances ... 133
7.2.5 Plotting Region Limits ... 133
7.3 Adding Points, Lines, and Text to an Existing Plot ... 134
Exercise 7.1 ... 139
7.4 The ggplot2 Package ... 139
7.4.1 A Quick Plot with qplot ... 140
7.4.2 Setting Appearance Constants with Geoms ... 141
7.4.3 Aesthetic Mapping with Geoms ... 143
Exercise 7.2 ... 146

8
READING AND WRITING FILES **147**

8.1 R-Ready Data Sets ... 147
8.1.1 Built-in Data Sets ... 148
8.1.2 Contributed Data Sets ... 149
8.2 Reading in External Data Files ... 150
8.2.1 The Table Format ... 150
8.2.2 Spreadsheet Workbooks ... 153
8.2.3 Web-Based Files ... 154
8.2.4 Other File Formats ... 155
8.3 Writing Out Data Files and Plots ... 156
8.3.1 Data Sets ... 156
8.3.2 Plots and Graphics Files ... 157
8.4 Ad Hoc Object Read/Write Operations ... 160
Exercise 8.1 ... 161

PART II
PROGRAMMING

9
CALLING FUNCTIONS **165**

9.1 Scoping ... 165
9.1.1 Environments ... 166
9.1.2 Search Path ... 168
9.1.3 Reserved and Protected Names ... 170
Exercise 9.1 ... 172

9.2 Argument Matching 172
9.2.1 Exact 172
9.2.2 Partial 173
9.2.3 Positional 174
9.2.4 Mixed 175
9.2.5 Dot-Dot-Dot: Use of Ellipses 176
Exercise 9.2 177

10
CONDITIONS AND LOOPS **179**

10.1 if Statements 179
10.1.1 Stand-Alone Statement 180
10.1.2 else Statements 183
10.1.3 Using ifelse for Element-wise Checks 184
Exercise 10.1 185
10.1.4 Nesting and Stacking Statements 186
10.1.5 The switch Function 189
Exercise 10.2 191
10.2 Coding Loops 193
10.2.1 for Loops 193
Exercise 10.3 199
10.2.2 while Loops 200
Exercise 10.4 203
10.2.3 Implicit Looping with apply 204
Exercise 10.5 208
10.3 Other Control Flow Mechanisms 209
10.3.1 Declaring break or next 209
10.3.2 The repeat Statement 211
Exercise 10.6 213

11
WRITING FUNCTIONS **215**

11.1 The function Command 215
11.1.1 Function Creation 216
11.1.2 Using return 220
Exercise 11.1 221
11.2 Arguments 222
11.2.1 Lazy Evaluation 222
11.2.2 Setting Defaults 225
11.2.3 Checking for Missing Arguments 227
11.2.4 Dealing with Ellipses 228
Exercise 11.2 231
11.3 Specialized Functions 233
11.3.1 Helper Functions 233
11.3.2 Disposable Functions 236
11.3.3 Recursive Functions 237
Exercise 11.3 238

12
EXCEPTIONS, TIMINGS, AND VISIBILITY **241**

12.1 Exception Handling 241
12.1.1 Formal Notifications: Errors and Warnings 242
12.1.2 Catching Errors with try Statements 244
Exercise 12.1 247
12.2 Progress and Timing 249
12.2.1 Textual Progress Bars: Are We There Yet? 249
12.2.2 Measuring Completion Time: How Long Did It Take? 250
Exercise 12.2 251
12.3 Masking 252
12.3.1 Function and Object Distinction 252
12.3.2 Data Frame Variable Distinction 255

PART III
STATISTICS AND PROBABILITY

13
ELEMENTARY STATISTICS **261**

13.1 Describing Raw Data 261
13.1.1 Numeric Variables 262
13.1.2 Categorical Variables 262
13.1.3 Univariate and Multivariate Data 264
13.1.4 Parameter or Statistic? 265
Exercise 13.1 266
13.2 Summary Statistics 267
13.2.1 Centrality: Mean, Median, Mode 267
13.2.2 Counts, Percentages, and Proportions 271
Exercise 13.2 273
13.2.3 Quantiles, Percentiles, and the Five-Number Summary 274
13.2.4 Spread: Variance, Standard Deviation, and the Interquartile Range . . 275
Exercise 13.3 279
13.2.5 Covariance and Correlation 280
13.2.6 Outliers 285
Exercise 13.4 287

14
BASIC DATA VISUALIZATION **289**

14.1 Barplots and Pie Charts 289
14.1.1 Building a Barplot 289
14.1.2 A Quick Pie Chart 293
14.2 Histograms 294

14.3 Box-and-Whisker Plots 298
14.3.1 Stand-Alone Boxplots 298
14.3.2 Side-by-Side Boxplots 299
14.4 Scatterplots 300
14.4.1 Single Plot 301
14.4.2 Matrix of Plots 302
Exercise 14.1 306

15
PROBABILITY **309**

15.1 What Is a Probability? 309
15.1.1 Events and Probability 310
15.1.2 Conditional Probability 311
15.1.3 Intersection 311
15.1.4 Union 312
15.1.5 Complement 312
Exercise 15.1 313
15.2 Random Variables and Probability Distributions 314
15.2.1 Realizations 314
15.2.2 Discrete Random Variables 315
15.2.3 Continuous Random Variables 318
15.2.4 Shape, Skew, and Modality 326
Exercise 15.2 327

16
COMMON PROBABILITY DISTRIBUTIONS **331**

16.1 Common Probability Mass Functions 332
16.1.1 Bernoulli Distribution 332
16.1.2 Binomial Distribution 333
Exercise 16.1 338
16.1.3 Poisson Distribution 338
Exercise 16.2 341
16.1.4 Other Mass Functions 342
16.2 Common Probability Density Functions 342
16.2.1 Uniform 343
Exercise 16.3 347
16.2.2 Normal 348
Exercise 16.4 356
16.2.3 Student's t-distribution 357
16.2.4 Exponential 359
Exercise 16.5 362
16.2.5 Other Density Functions 362

PART IV
STATISTICAL TESTING AND MODELING

17
SAMPLING DISTRIBUTIONS AND CONFIDENCE 367

17.1 Sampling Distributions 367
17.1.1 Distribution for a Sample Mean 368
17.1.2 Distribution for a Sample Proportion 373
Exercise 17.1 376
17.1.3 Sampling Distributions for Other Statistics 377
17.2 Confidence Intervals 378
17.2.1 An Interval for a Mean 378
17.2.2 An Interval for a Proportion 381
17.2.3 Other Intervals 382
17.2.4 Comments on Interpretation of a CI 382
Exercise 17.2 383

18
HYPOTHESIS TESTING 385

18.1 Components of a Hypothesis Test 385
18.1.1 Hypotheses 386
18.1.2 Test Statistic 387
18.1.3 p-value 387
18.1.4 Significance Level 387
18.1.5 Criticisms of Hypothesis Testing 388
18.2 Testing Means 388
18.2.1 Single Mean 389
Exercise 18.1 392
18.2.2 Two Means 392
Exercise 18.2 401
18.3 Testing Proportions 402
18.3.1 Single Proportion 402
18.3.2 Two Proportions 405
Exercise 18.3 408
18.4 Testing Categorical Variables 410
18.4.1 Single Categorical Variable 410
18.4.2 Two Categorical Variables 415
Exercise 18.4 419
18.5 Errors and Power 420
18.5.1 Hypothesis Test Errors 420
18.5.2 Type I Errors 421
18.5.3 Type II Errors 424
Exercise 18.5 427
18.5.4 Statistical Power 428
Exercise 18.6 433

19
ANALYSIS OF VARIANCE 435

19.1 One-Way ANOVA 435
19.1.1 Hypotheses and Diagnostic Checking 436
19.1.2 One-Way ANOVA Table Construction 439
19.1.3 Building ANOVA Tables with the aov Function 440
Exercise 19.1 442
19.2 Two-Way ANOVA 443
19.2.1 A Suite of Hypotheses 443
19.2.2 Main Effects and Interactions 444
19.3 Kruskal-Wallis Test 447
Exercise 19.2 449

20
SIMPLE LINEAR REGRESSION 451

20.1 An Example of a Linear Relationship 451
20.2 General Concepts 453
20.2.1 Definition of the Model 453
20.2.2 Estimating the Intercept and Slope Parameters 454
20.2.3 Fitting Linear Models with lm 455
20.2.4 Illustrating Residuals 456
20.3 Statistical Inference 458
20.3.1 Summarizing the Fitted Model 458
20.3.2 Regression Coefficient Significance Tests 459
20.3.3 Coefficient of Determination 460
20.3.4 Other summary Output 460
20.4 Prediction 461
20.4.1 Confidence Interval or Prediction Interval? 461
20.4.2 Interpreting Intervals 462
20.4.3 Plotting Intervals 464
20.4.4 Interpolation vs. Extrapolation 466
Exercise 20.1 467
20.5 Understanding Categorical Predictors 468
20.5.1 Binary Variables: k = 2 468
20.5.2 Multilevel Variables: k > 2 472
20.5.3 Changing the Reference Level 477
20.5.4 Treating Categorical Variables as Numeric 478
20.5.5 Equivalence with One-Way ANOVA 481
Exercise 20.2 482

21
MULTIPLE LINEAR REGRESSION 485

21.1 Terminology 486
21.2 Theory 486
21.2.1 Extending the Simple Model to a Multiple Model 487

21.2.2 Estimating in Matrix Form ... 487
21.2.3 A Basic Example ... 488
21.3 Implementing in R and Interpreting ... 490
21.3.1 Additional Predictors ... 490
21.3.2 Interpreting Marginal Effects ... 493
21.3.3 Visualizing the Multiple Linear Model ... 494
21.3.4 Finding Confidence Intervals ... 495
21.3.5 Omnibus F-Test ... 496
21.3.6 Predicting from a Multiple Linear Model ... 498
Exercise 21.1 ... 499
21.4 Transforming Numeric Variables ... 501
21.4.1 Polynomial ... 502
21.4.2 Logarithmic ... 508
21.4.3 Other Transformations ... 512
Exercise 21.2 ... 512
21.5 Interactive Terms ... 514
21.5.1 Concept and Motivation ... 514
21.5.2 One Categorical, One Continuous ... 515
21.5.3 Two Categorical ... 519
21.5.4 Two Continuous ... 521
21.5.5 Higher-Order Interactions ... 523
Exercise 21.3 ... 525

22
LINEAR MODEL SELECTION AND DIAGNOSTICS 527

22.1 Goodness-of-Fit vs. Complexity ... 527
22.1.1 Principle of Parsimony ... 528
22.1.2 General Guidelines ... 528
22.2 Model Selection Algorithms ... 529
22.2.1 Nested Comparisons: The Partial F-Test ... 529
22.2.2 Forward Selection ... 533
22.2.3 Backward Selection ... 537
22.2.4 Stepwise AIC Selection ... 541
Exercise 22.1 ... 546
22.2.5 Other Selection Algorithms ... 548
22.3 Residual Diagnostics ... 548
22.3.1 Inspecting and Interpreting Residuals ... 549
22.3.2 Assessing Normality ... 554
22.3.3 Illustrating Outliers, Leverage, and Influence ... 555
22.3.4 Calculating Leverage ... 558
22.3.5 Cook's Distance ... 559
22.3.6 Graphically Combining Residuals, Leverage, and Cook's Distance ... 563
Exercise 22.2 ... 567
22.4 Collinearity ... 569
22.4.1 Potential Warning Signs ... 569
22.4.2 Correlated Predictors: A Quick Example ... 569

PART V
ADVANCED GRAPHICS

23
ADVANCED PLOT CUSTOMIZATION 575

23.1 Handling the Graphics Device 576
23.1.1 Manually Opening a New Device 576
23.1.2 Switching Between Devices 577
23.1.3 Closing a Device 578
23.1.4 Multiple Plots in One Device 578
23.2 Plotting Regions and Margins 582
23.2.1 Default Spacing 582
23.2.2 Custom Spacing 583
23.2.3 Clipping 584
23.3 Point-and-Click Coordinate Interaction 586
23.3.1 Retrieving Coordinates Silently 586
23.3.2 Visualizing Selected Coordinates 587
23.3.3 Ad Hoc Annotation 588
Exercise 23.1 589
23.4 Customizing Traditional R Plots 591
23.4.1 Graphical Parameters for Style and Suppression 592
23.4.2 Customizing Boxes 593
23.4.3 Customizing Axes 594
23.5 Specialized Text and Label Notation 597
23.5.1 Font 597
23.5.2 Greek Symbols 598
23.5.3 Mathematical Expressions 599
23.6 A Fully Annotated Scatterplot 601
Exercise 23.2 606

24
GOING FURTHER WITH THE GRAMMAR OF GRAPHICS 609

24.1 ggplot or qplot? 609
24.2 Smoothing and Shading 611
24.2.1 Adding LOESS Trends 611
24.2.2 Constructing Smooth Density Estimates 614
24.3 Multiple Plots and Variable-Mapped Facets 616
24.3.1 Independent Plots 616
24.3.2 Facets Mapped to a Categorical Variable 619
Exercise 24.1 622
24.4 Interactive Tools in ggvis 623
Exercise 24.2 628

25
DEFINING COLORS AND PLOTTING IN HIGHER DIMENSIONS 631

25.1 Representing and Using Color 631
25.1.1 Red-Green-Blue Hexadecimal Color Codes 632
25.1.2 Built-in Palettes 635
25.1.3 Custom Palettes 636
25.1.4 Using Color Palettes to Index a Continuum 637
25.1.5 Including a Color Legend 641
25.1.6 Opacity 643
25.1.7 RGB Alternatives and Further Functionality 645
Exercise 25.1 646
25.2 3D Scatterplots 649
25.2.1 Basic Syntax 649
25.2.2 Visual Enhancements 650
Exercise 25.2 652
25.3 Preparing a Surface for Plotting 653
25.3.1 Constructing an Evaluation Grid 654
25.3.2 Constructing the z-Matrix 655
25.3.3 Conceptualizing the z-Matrix 656
25.4 Contour Plots 657
25.4.1 Drawing Contour Lines 657
25.4.2 Color-Filled Contours 663
Exercise 25.3 666
25.5 Pixel Images 668
25.5.1 One Grid Point = One Pixel 668
25.5.2 Surface Truncation and Empty Pixels 671
Exercise 25.4 676
25.6 Perspective Plots 679
25.6.1 Basic Plots and Angle Adjustment 679
25.6.2 Coloring Facets 682
25.6.3 Rotating with Loops 686
Exercise 25.5 687

26
INTERACTIVE 3D PLOTS 691

26.1 Point Clouds 691
26.1.1 Basic 3D Cloud 692
26.1.2 Visual Enhancements and Legends 693
26.1.3 Adding Further 3D Components 694
Exercise 26.1 697
26.2 Bivariate Surfaces 699
26.2.1 Basic Perspective Surface 700
26.2.2 Additional Components 701
26.2.3 Coloring by z Value 703
26.2.4 Dealing with the Aspect Ratio 704
Exercise 26.2 706

26.3 Trivariate Surfaces 709
26.3.1 Evaluation Coordinates in 3D 709
26.3.2 Isosurfaces 710
26.3.3 Example: Nonparametric Trivariate Density 715
26.4 Handling Parametric Equations 720
26.4.1 Simple Loci 720
26.4.2 Mathematical Abstractions 725
Exercise 26.3 732

A
INSTALLING R AND CONTRIBUTED PACKAGES 737

A.1 Downloading and Installing R 737
A.2 Using Packages 739
A.2.1 Base Packages 739
A.2.2 Recommended Packages 740
A.2.3 Contributed Packages 740
A.3 Updating R and Installed Packages 746
A.4 Using Other Mirrors and Repositories 747
A.4.1 Switching CRAN Mirror 747
A.4.2 Other Package Repositories 748
A.5 Citing and Writing Packages 748
A.5.1 Citing R and Contributed Packages 748
A.5.2 Writing Your Own Packages 749

B
WORKING WITH RSTUDIO 751

B.1 Basic Layout and Usage 752
B.1.1 Editor Features and Appearance Options 752
B.1.2 Customizing Panes 753
B.2 Auxiliary Tools 754
B.2.1 Projects 754
B.2.2 Package Installer and Updater 755
B.2.3 Support for Debugging 756
B.2.4 Markup, Document, and Graphics Tools 756

REFERENCE LIST 761

INDEX 767