Table of Contents

Preface xiii

1. Getting Started and Getting Help 1
1.1 Downloading and Installing R 2
1.2 Starting R 4
1.3 Entering Commands 7
1.4 Exiting from R 8
1.5 Interrupting R 9
1.6 Viewing the Supplied Documentation 10
1.7 Getting Help on a Function 11
1.8 Searching the Supplied Documentation 13
1.9 Getting Help on a Package 14
1.10 Searching the Web for Help 16
1.11 Finding Relevant Functions and Packages 18
1.12 Searching the Mailing Lists 19
1.13 Submitting Questions to the Mailing Lists 20

2. Some Basics 23
2.1 Printing Something 23
2.2 Setting Variables 25
2.3 Listing Variables 26
2.4 Deleting Variables 27
2.5 Creating a Vector 28
2.6 Computing Basic Statistics 30
2.7 Creating Sequences 32
2.8 Comparing Vectors 34
2.9 Selecting Vector Elements 35
2.10 Performing Vector Arithmetic 38
2.11 Getting Operator Precedence Right 40
2.12 Defining a Function 41
2.13 Typing Less and Accomplishing More 43

2.14 Avoiding Some Common Mistakes 46

3. Navigating the Software 51
3.1 Getting and Setting the Working Directory 51
3.2 Saving Your Workspace 52
3.3 Viewing Your Command History 53
3.4 Saving the Result of the Previous Command 53
3.5 Displaying the Search Path 54
3.6 Accessing the Functions in a Package 55
3.7 Accessing Built-in Datasets 57
3.8 Viewing the List of Installed Packages 58
3.9 Installing Packages from CRAN 59
3.10 Setting a Default CRAN Mirror 61
3.11 Suppressing the Startup Message 62
3.12 Running a Script 62
3.13 Running a Batch Script 63
3.14 Getting and Setting Environment Variables 66
3.15 Locating the R Home Directory 67
3.16 Customizing R 68

4. Input and Output 71
4.1 Entering Data from the Keyboard 72
4.2 Printing Fewer Digits (or More Digits) 73
4.3 Redirecting Output to a File 74
4.4 Listing Files 75
4.5 Dealing with "Cannot Open File" in Windows 76
4.6 Reading Fixed-Width Records 77
4.7 Reading Tabular Data Files 78
4.8 Reading from CSV Files 80
4.9 Writing to CSV Files 82
4.10 Reading Tabular or CSV Data from the Web 83
4.11 Reading Data from HTML Tables 84
4.12 Reading Files with a Complex Structure 86
4.13 Reading from MySQL Databases 89
4.14 Saving and Transporting Objects 92

5. Data Structures 95
5.1 Appending Data to a Vector 101
5.2 Inserting Data into a Vector 103
5.3 Understanding the Recycling Rule 103
5.4 Creating a Factor (Categorical Variable) 105
5.5 Combining Multiple Vectors into One Vector and a Factor 107
5.6 Creating a List 108

5.7 Selecting List Elements by Position 109
5.8 Selecting List Elements by Name 111
5.9 Building a Name/Value Association List 112
5.10 Removing an Element from a List 114
5.11 Flatten a List into a Vector 115
5.12 Removing NULL Elements from a List 116
5.13 Removing List Elements Using a Condition 117
5.14 Initializing a Matrix 118
5.15 Performing Matrix Operations 119
5.16 Giving Descriptive Names to the Rows and Columns of a Matrix 120
5.17 Selecting One Row or Column from a Matrix 121
5.18 Initializing a Data Frame from Column Data 122
5.19 Initializing a Data Frame from Row Data 123
5.20 Appending Rows to a Data Frame 125
5.21 Preallocating a Data Frame 126
5.22 Selecting Data Frame Columns by Position 127
5.23 Selecting Data Frame Columns by Name 131
5.24 Selecting Rows and Columns More Easily 132
5.25 Changing the Names of Data Frame Columns 133
5.26 Editing a Data Frame 135
5.27 Removing NAs from a Data Frame 136
5.28 Excluding Columns by Name 137
5.29 Combining Two Data Frames 138
5.30 Merging Data Frames by Common Column 140
5.31 Accessing Data Frame Contents More Easily 141
5.32 Converting One Atomic Value into Another 143
5.33 Converting One Structured Data Type into Another 144

6. Data Transformations 147
6.1 Splitting a Vector into Groups 148
6.2 Applying a Function to Each List Element 149
6.3 Applying a Function to Every Row 151
6.4 Applying a Function to Every Column 152
6.5 Applying a Function to Groups of Data 154
6.6 Applying a Function to Groups of Rows 156
6.7 Applying a Function to Parallel Vectors or Lists 158

7. Strings and Dates 161
7.1 Getting the Length of a String 163
7.2 Concatenating Strings 163
7.3 Extracting Substrings 164
7.4 Splitting a String According to a Delimiter 165
7.5 Replacing Substrings 166

7.6 Seeing the Special Characters in a String 167
7.7 Generating All Pairwise Combinations of Strings 168
7.8 Getting the Current Date 169
7.9 Converting a String into a Date 170
7.10 Converting a Date into a String 171
7.11 Converting Year, Month, and Day into a Date 172
7.12 Getting the Julian Date 173
7.13 Extracting the Parts of a Date 174
7.14 Creating a Sequence of Dates 175

8. Probability 177
8.1 Counting the Number of Combinations 179
8.2 Generating Combinations 180
8.3 Generating Random Numbers 180
8.4 Generating Reproducible Random Numbers 182
8.5 Generating a Random Sample 183
8.6 Generating Random Sequences 184
8.7 Randomly Permuting a Vector 185
8.8 Calculating Probabilities for Discrete Distributions 186
8.9 Calculating Probabilities for Continuous Distributions 188
8.10 Converting Probabilities to Quantiles 189
8.11 Plotting a Density Function 190

9. General Statistics 195
9.1 Summarizing Your Data 197
9.2 Calculating Relative Frequencies 199
9.3 Tabulating Factors and Creating Contingency Tables 200
9.4 Testing Categorical Variables for Independence 201
9.5 Calculating Quantiles (and Quartiles) of a Dataset 201
9.6 Inverting a Quantile 202
9.7 Converting Data to Z-Scores 203
9.8 Testing the Mean of a Sample (*t* Test) 203
9.9 Forming a Confidence Interval for a Mean 205
9.10 Forming a Confidence Interval for a Median 206
9.11 Testing a Sample Proportion 207
9.12 Forming a Confidence Interval for a Proportion 208
9.13 Testing for Normality 209
9.14 Testing for Runs 210
9.15 Comparing the Means of Two Samples 212
9.16 Comparing the Locations of Two Samples Nonparametrically 213
9.17 Testing a Correlation for Significance 215
9.18 Testing Groups for Equal Proportions 216
9.19 Performing Pairwise Comparisons Between Group Means 218

9.20 Testing Two Samples for the Same Distribution 219

10. Graphics 221
10.1 Creating a Scatter Plot 223
10.2 Adding a Title and Labels 225
10.3 Adding a Grid 226
10.4 Creating a Scatter Plot of Multiple Groups 227
10.5 Adding a Legend 229
10.6 Plotting the Regression Line of a Scatter Plot 231
10.7 Plotting All Variables Against All Other Variables 233
10.8 Creating One Scatter Plot for Each Factor Level 233
10.9 Creating a Bar Chart 236
10.10 Adding Confidence Intervals to a Bar Chart 237
10.11 Coloring a Bar Chart 239
10.12 Plotting a Line from x and y Points 241
10.13 Changing the Type, Width, or Color of a Line 242
10.14 Plotting Multiple Datasets 243
10.15 Adding Vertical or Horizontal Lines 245
10.16 Creating a Box Plot 246
10.17 Creating One Box Plot for Each Factor Level 247
10.18 Creating a Histogram 248
10.19 Adding a Density Estimate to a Histogram 250
10.20 Creating a Discrete Histogram 252
10.21 Creating a Normal Quantile-Quantile (Q-Q) Plot 252
10.22 Creating Other Quantile-Quantile Plots 254
10.23 Plotting a Variable in Multiple Colors 256
10.24 Graphing a Function 258
10.25 Pausing Between Plots 259
10.26 Displaying Several Figures on One Page 260
10.27 Opening Additional Graphics Windows 262
10.28 Writing Your Plot to a File 263
10.29 Changing Graphical Parameters 264

11. Linear Regression and ANOVA 267
11.1 Performing Simple Linear Regression 269
11.2 Performing Multiple Linear Regression 270
11.3 Getting Regression Statistics 272
11.4 Understanding the Regression Summary 275
11.5 Performing Linear Regression Without an Intercept 278
11.6 Performing Linear Regression with Interaction Terms 279
11.7 Selecting the Best Regression Variables 281
11.8 Regressing on a Subset of Your Data 284
11.9 Using an Expression Inside a Regression Formula 285

11.10 Regressing on a Polynomial 286
11.11 Regressing on Transformed Data 287
11.12 Finding the Best Power Transformation (Box–Cox Procedure) 289
11.13 Forming Confidence Intervals for Regression Coefficients 292
11.14 Plotting Regression Residuals 293
11.15 Diagnosing a Linear Regression 293
11.16 Identifying Influential Observations 296
11.17 Testing Residuals for Autocorrelation (Durbin–Watson Test) 298
11.18 Predicting New Values 300
11.19 Forming Prediction Intervals 301
11.20 Performing One-Way ANOVA 302
11.21 Creating an Interaction Plot 303
11.22 Finding Differences Between Means of Groups 304
11.23 Performing Robust ANOVA (Kruskal–Wallis Test) 308
11.24 Comparing Models by Using ANOVA 309

12. Useful Tricks 313
12.1 Peeking at Your Data 313
12.2 Widen Your Output 314
12.3 Printing the Result of an Assignment 315
12.4 Summing Rows and Columns 315
12.5 Printing Data in Columns 316
12.6 Binning Your Data 317
12.7 Finding the Position of a Particular Value 318
12.8 Selecting Every nth Element of a Vector 319
12.9 Finding Pairwise Minimums or Maximums 320
12.10 Generating All Combinations of Several Factors 321
12.11 Flatten a Data Frame 322
12.12 Sorting a Data Frame 323
12.13 Sorting by Two Columns 324
12.14 Stripping Attributes from a Variable 325
12.15 Revealing the Structure of an Object 326
12.16 Timing Your Code 329
12.17 Suppressing Warnings and Error Messages 329
12.18 Taking Function Arguments from a List 331
12.19 Defining Your Own Binary Operators 332

13. Beyond Basic Numerics and Statistics 335
13.1 Minimizing or Maximizing a Single-Parameter Function 335
13.2 Minimizing or Maximizing a Multiparameter Function 336
13.3 Calculating Eigenvalues and Eigenvectors 338
13.4 Performing Principal Component Analysis 338
13.5 Performing Simple Orthogonal Regression 340

13.6 Finding Clusters in Your Data 342
13.7 Predicting a Binary-Valued Variable (Logistic Regression) 345
13.8 Bootstrapping a Statistic 346
13.9 Factor Analysis 349

14. Time Series Analysis 355
14.1 Representing Time Series Data 356
14.2 Plotting Time Series Data 359
14.3 Extracting the Oldest or Newest Observations 361
14.4 Subsetting a Time Series 363
14.5 Merging Several Time Series 364
14.6 Filling or Padding a Time Series 366
14.7 Lagging a Time Series 368
14.8 Computing Successive Differences 369
14.9 Performing Calculations on Time Series 370
14.10 Computing a Moving Average 372
14.11 Applying a Function by Calendar Period 373
14.12 Applying a Rolling Function 375
14.13 Plotting the Autocorrelation Function 376
14.14 Testing a Time Series for Autocorrelation 377
14.15 Plotting the Partial Autocorrelation Function 378
14.16 Finding Lagged Correlations Between Two Time Series 379
14.17 Detrending a Time Series 382
14.18 Fitting an ARIMA Model 383
14.19 Removing Insignificant ARIMA Coefficients 386
14.20 Running Diagnostics on an ARIMA Model 387
14.21 Making Forecasts from an ARIMA Model 389
14.22 Testing for Mean Reversion 391
14.23 Smoothing a Time Series 393

Index 397