Contents

Preface to the second edition x
Preface to the first edition xiii

1 Getting started 1

1.1 What is statistical programming? 1
1.2 Outline of this book 2
1.3 The R package 3
1.4 Why use a command line? 3
1.5 Font conventions 4
1.6 Installation of R and RStudio 4
1.7 Getting started in RStudio 5
1.8 Going further 6

2 Introduction to the R language 7

2.1 First steps 7
2.2 Basic features of R 11
2.3 Vectors in R 13
2.4 Data storage in R 22
2.5 Packages, libraries, and repositories 27
2.6 Getting help 28
2.7 Logical vectors and relational operators 34
2.8 Data frames and lists 37
2.9 Data input and output 43

3 Programming statistical graphics 49

3.1 High level plots 50
3.2 Choosing a high level graphic 62
3.3 Low level graphics functions 63
3.4 Other graphics systems 70
## Table of Contents

4 Programming with R 76
   4.1 Flow control 76
   4.2 Managing complexity through functions 91
   4.3 The `replicate()` function 97
   4.4 Miscellaneous programming tips 97
   4.5 Some general programming guidelines 100
   4.6 Debugging and maintenance 107
   4.7 Efficient programming 113

5 Simulation 120
   5.1 Monte Carlo simulation 120
   5.2 Generation of pseudorandom numbers 121
   5.3 Simulation of other random variables 126
   5.4 Multivariate random number generation 142
   5.5 Markov chain simulation 143
   5.6 Monte Carlo integration 147
   5.7 Advanced simulation methods 149

6 Computational linear algebra 158
   6.1 Vectors and matrices in R 159
   6.2 Matrix multiplication and inversion 166
   6.3 Eigenvalues and eigenvectors 171
   6.4 Other matrix decompositions 172
   6.5 Other matrix operations 178

7 Numerical optimization 182
   7.1 The golden section search method 182
   7.2 Newton–Raphson 185
   7.3 The Nelder–Mead simplex method 188
   7.4 Built-in functions 191
   7.5 Linear programming 192

Appendix 209
   Review of random variables and distributions 209

Index 212
# Expanded contents

## Preface to the second edition  
**page xi**

## Preface to the first edition  
xiii

## 1 Getting started  
1
  1.1 What is statistical programming?  
  2
  1.2 Outline of this book  
  3
  1.3 The R package  
  3
  1.4 Why use a command line?  
  3
  1.5 Font conventions  
  4
  1.6 Installation of R and RStudio  
  4
  1.7 Getting started in RStudio  
  5
  1.8 Going further  
  6

## 2 Introduction to the R language  
7
  2.1 First steps  
  7
    2.1.1 R can be used as a calculator  
    7
    2.1.2 Named storage  
    9
    2.1.3 Quitting R  
   10
  2.2 Basic features of R  
  11
    2.2.1 Functions  
    11
    2.2.2 R is case-sensitive  
    12
    2.2.3 Listing the objects in the workspace  
    13
  2.3 Vectors in R  
  13
    2.3.1 Numeric vectors  
    13
    2.3.2 Extracting elements from vectors  
    14
    2.3.3 Vector arithmetic  
    15
    2.3.4 Simple patterned vectors  
    16
    2.3.5 Vectors with random patterns  
    17
    2.3.6 Character vectors  
    17
    2.3.7 Factors  
    18
    2.3.8 More on extracting elements from vectors  
    19
## EXPANDED CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.9</td>
<td>Matrices and arrays</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>Data storage in R</td>
<td></td>
</tr>
<tr>
<td>2.4.1</td>
<td>Approximate storage of numbers</td>
<td>22</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Exact storage of numbers</td>
<td>24</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Dates and times</td>
<td>25</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Missing values and other special values</td>
<td>25</td>
</tr>
<tr>
<td>2.5</td>
<td>Packages, libraries, and repositories</td>
<td>27</td>
</tr>
<tr>
<td>2.6</td>
<td>Getting help</td>
<td></td>
</tr>
<tr>
<td>2.6.1</td>
<td>Built-in help pages</td>
<td>28</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Built-in examples</td>
<td>29</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Finding help when you don’t know the function name</td>
<td>30</td>
</tr>
<tr>
<td>2.6.4</td>
<td>Some built-in graphics functions</td>
<td>31</td>
</tr>
<tr>
<td>2.6.5</td>
<td>Some elementary built-in functions</td>
<td>33</td>
</tr>
<tr>
<td>2.7</td>
<td>Logical vectors and relational operators</td>
<td>34</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Boolean algebra</td>
<td>34</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Logical operations in R</td>
<td>34</td>
</tr>
<tr>
<td>2.7.3</td>
<td>Relational operators</td>
<td>36</td>
</tr>
<tr>
<td>2.8</td>
<td>Data frames and lists</td>
<td></td>
</tr>
<tr>
<td>2.8.1</td>
<td>Extracting data frame elements and subsets</td>
<td>39</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Taking random samples from populations</td>
<td>40</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Constructing data frames</td>
<td>40</td>
</tr>
<tr>
<td>2.8.4</td>
<td>Data frames can have non-numeric columns</td>
<td>40</td>
</tr>
<tr>
<td>2.8.5</td>
<td>Lists</td>
<td>41</td>
</tr>
<tr>
<td>2.9</td>
<td>Data input and output</td>
<td></td>
</tr>
<tr>
<td>2.9.1</td>
<td>Changing directories</td>
<td>43</td>
</tr>
<tr>
<td>2.9.2</td>
<td><code>dump()</code> and <code>source()</code></td>
<td>43</td>
</tr>
<tr>
<td>2.9.3</td>
<td>Redirecting R output</td>
<td>44</td>
</tr>
<tr>
<td>2.9.4</td>
<td>Saving and retrieving image files</td>
<td>45</td>
</tr>
<tr>
<td>2.9.5</td>
<td>The <code>read.table()</code> function</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>Programming statistical graphics</td>
<td>49</td>
</tr>
<tr>
<td>3.1</td>
<td>High level plots</td>
<td></td>
</tr>
<tr>
<td>3.1.1</td>
<td>Bar charts and dot charts</td>
<td>50</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Pie charts</td>
<td>53</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Histograms</td>
<td>54</td>
</tr>
<tr>
<td>3.1.4</td>
<td>Box plots</td>
<td>55</td>
</tr>
<tr>
<td>3.1.5</td>
<td>Scatterplots</td>
<td>57</td>
</tr>
<tr>
<td>3.1.6</td>
<td>Plotting data from data frames</td>
<td>57</td>
</tr>
<tr>
<td>3.1.7</td>
<td>QQ plots</td>
<td>60</td>
</tr>
<tr>
<td>3.2</td>
<td>Choosing a high level graphic</td>
<td>62</td>
</tr>
<tr>
<td>3.3</td>
<td>Low level graphics functions</td>
<td>63</td>
</tr>
<tr>
<td>3.3.1</td>
<td>The plotting region and margins</td>
<td>63</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Adding to plots</td>
<td>64</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Adjusting axis tick labels</td>
<td>66</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Setting graphical parameters</td>
<td>68</td>
</tr>
<tr>
<td>3.4</td>
<td>Other graphics systems</td>
<td></td>
</tr>
<tr>
<td>3.4.1</td>
<td>The <code>ggplot2</code> package</td>
<td>70</td>
</tr>
<tr>
<td>3.4.2</td>
<td>The <code>lattice</code> package</td>
<td>72</td>
</tr>
</tbody>
</table>
## 4 Programming with R

### 4.1 Flow control
- 4.1.1 The *for()* loop
- 4.1.2 The *if()* statement
- 4.1.3 The *while()* loop
- 4.1.4 Newton's method for root finding
- 4.1.5 The *repeat* loop, and the *break* and *next* statements

### 4.2 Managing complexity through functions
- 4.2.1 What are functions?
- 4.2.2 Scope of variables
- 4.2.3 Returning multiple objects
- 4.2.4 Using S3 classes to control printing

### 4.3 The *replicate()* function

### 4.4 Miscellaneous programming tips
- 4.4.1 Always edit code in the editor, not in the console
- 4.4.2 Documentation using 
- 4.4.3 Neatness counts!

### 4.5 Some general programming guidelines
- 4.5.1 Top-down design

### 4.6 Debugging and maintenance
- 4.6.1 Recognizing that a bug exists
- 4.6.2 Make the bug reproducible
- 4.6.3 Identify the cause of the bug
- 4.6.4 Fixing errors and testing
- 4.6.5 Look for similar errors elsewhere
- 4.6.6 Debugging in RStudio
- 4.6.7 The *browser()*,* debug()*,* and *debugonce()* functions

### 4.7 Efficient programming
- 4.7.1 Learn your tools
- 4.7.2 Use efficient algorithms
- 4.7.3 Measure the time your program takes
- 4.7.4 Be willing to use different tools
- 4.7.5 Optimize with care

## 5 Simulation

### 5.1 Monte Carlo simulation

### 5.2 Generation of pseudorandom numbers

### 5.3 Simulation of other random variables
- 5.3.1 Bernoulli random variables
- 5.3.2 Binomial random variables
- 5.3.3 Poisson random variables
- 5.3.4 Exponential random numbers
- 5.3.5 Normal random variables
- 5.3.6 All built-in distributions