
Contents

	Preface to the second edition	<i>page xi</i>
	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

4	Programming with R	76
4.1	Flow control	76
4.2	Managing complexity through functions	91
4.3	The <code>replicate()</code> 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	Review of random variables and distributions	209
	Index	212

Expanded contents

	Preface to the second edition	<i>page xi</i>
	Preface to the first edition	xiii
I	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
2.1.1	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

2.3.9	Matrices and arrays	19
2.4	Data storage in R	22
2.4.1	Approximate storage of numbers	22
2.4.2	Exact storage of numbers	24
2.4.3	Dates and times	25
2.4.4	Missing values and other special values	25
2.5	Packages, libraries, and repositories	27
2.6	Getting help	28
2.6.1	Built-in help pages	28
2.6.2	Built-in examples	29
2.6.3	Finding help when you don't know the function name	30
2.6.4	Some built-in graphics functions	31
2.6.5	Some elementary built-in functions	33
2.7	Logical vectors and relational operators	34
2.7.1	Boolean algebra	34
2.7.2	Logical operations in R	34
2.7.3	Relational operators	36
2.8	Data frames and lists	37
2.8.1	Extracting data frame elements and subsets	39
2.8.2	Taking random samples from populations	40
2.8.3	Constructing data frames	40
2.8.4	Data frames can have non-numeric columns	40
2.8.5	Lists	41
2.9	Data input and output	43
2.9.1	Changing directories	43
2.9.2	<code>dump()</code> and <code>source()</code>	43
2.9.3	Redirecting R output	44
2.9.4	Saving and retrieving image files	45
2.9.5	The <code>read.table</code> function	45

3 | Programming statistical graphics 49

3.1	High level plots	50
3.1.1	Bar charts and dot charts	50
3.1.2	Pie charts	53
3.1.3	Histograms	54
3.1.4	Box plots	55
3.1.5	Scatterplots	57
3.1.6	Plotting data from data frames	57
3.1.7	QQ plots	60
3.2	Choosing a high level graphic	62
3.3	Low level graphics functions	63
3.3.1	The plotting region and margins	63
3.3.2	Adding to plots	64
3.3.3	Adjusting axis tick labels	66
3.3.4	Setting graphical parameters	68
3.4	Other graphics systems	70
3.4.1	The <code>ggplot2</code> package	70
3.4.2	The <code>lattice</code> package	72

3.4.3	The <code>grid</code> package	73
3.4.4	Interactive graphics	74
4	Programming with R	76
4.1	Flow control	76
4.1.1	The <code>for()</code> loop	76
4.1.2	The <code>if()</code> statement	82
4.1.3	The <code>while()</code> loop	86
4.1.4	Newton's method for root finding	87
4.1.5	The <code>repeat</code> loop, and the <code>break</code> and <code>next</code> statements	89
4.2	Managing complexity through functions	91
4.2.1	What are functions?	91
4.2.2	Scope of variables	94
4.2.3	Returning multiple objects	95
4.2.4	Using S3 classes to control printing	95
4.3	The <code>replicate()</code> function	97
4.4	Miscellaneous programming tips	97
4.4.1	Always edit code in the editor, not in the console	97
4.4.2	Documentation using <code>#</code>	98
4.4.3	Neatness counts!	98
4.5	Some general programming guidelines	100
4.5.1	Top-down design	103
4.6	Debugging and maintenance	107
4.6.1	Recognizing that a bug exists	108
4.6.2	Make the bug reproducible	108
4.6.3	Identify the cause of the bug	109
4.6.4	Fixing errors and testing	111
4.6.5	Look for similar errors elsewhere	111
4.6.6	Debugging in RStudio	111
4.6.7	The <code>browser()</code> , <code>debug()</code> , and <code>debugonce()</code> functions	112
4.7	Efficient programming	113
4.7.1	Learn your tools	114
4.7.2	Use efficient algorithms	114
4.7.3	Measure the time your program takes	116
4.7.4	Be willing to use different tools	117
4.7.5	Optimize with care	117
5	Simulation	120
5.1	Monte Carlo simulation	120
5.2	Generation of pseudorandom numbers	121
5.3	Simulation of other random variables	126
5.3.1	Bernoulli random variables	126
5.3.2	Binomial random variables	128
5.3.3	Poisson random variables	132
5.3.4	Exponential random numbers	136
5.3.5	Normal random variables	138
5.3.6	All built-in distributions	140

x	EXPANDED CONTENTS	
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
5.7.1	Rejection sampling	150
5.7.2	Importance sampling	152
6	Computational linear algebra	158
6.1	Vectors and matrices in R	159
6.1.1	Constructing matrix objects	159
6.1.2	Accessing matrix elements; row and column names	161
6.1.3	Matrix properties	163
6.1.4	Triangular matrices	164
6.1.5	Matrix arithmetic	165
6.2	Matrix multiplication and inversion	166
6.2.1	Matrix inversion	167
6.2.2	The <i>LU</i> decomposition	168
6.2.3	Matrix inversion in R	169
6.2.4	Solving linear systems	170
6.3	Eigenvalues and eigenvectors	171
6.4	Other matrix decompositions	172
6.4.1	The singular value decomposition of a matrix	172
6.4.2	The Choleski decomposition of a positive definite matrix	173
6.4.3	The QR decomposition of a matrix	174
6.5	Other matrix operations	178
6.5.1	Kronecker products	179
6.5.2	<code>apply()</code>	179
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
7.5.1	Solving linear programming problems in R	195
7.5.2	Maximization and other kinds of constraints	195
7.5.3	Special situations	196
7.5.4	Unrestricted variables	199
7.5.5	Integer programming	200
7.5.6	Alternatives to <code>lp()</code>	201
7.5.7	Quadratic programming	202
Appendix	Review of random variables and distributions	209
	Index	212