

# Contents

<b>Preface</b>	<b>xi</b>
<b>1 Getting Started With R</b>	<b>1</b>
1.1 R Basics	1
1.1.1 Interacting With the Interpreter	2
1.1.2 R Functions	4
1.1.3 Vectors and Variables	7
1.1.4 Nonnumeric Vectors	10
1.1.5 Indexing Vectors	12
1.1.6 User-Defined Functions	14
1.1.7 Command Editing and Output Management	16
1.1.8 When Things Go Wrong	18
1.1.9 Getting Help and Information	19
1.1.10 Cleaning Up	21
1.1.11 Ending the R Session	22
1.2 An Extended Illustration	22
1.2.1 Reading the Data	23
1.2.2 Examining the Data	24
1.2.3 Regression Analysis	29
1.2.4 Regression Diagnostics	31
1.3 R Functions for Basic Statistics	37
1.4 Generic Functions and Their Methods*	37
1.5 The R Commander Graphical User Interface	41
<b>2 Reading and Manipulating Data</b>	<b>43</b>
2.1 Data Input	44
2.1.1 Keyboard Input	44
2.1.2 File Input to a Data Frame	47
2.1.3 Importing Data	53
2.1.4 Accessing Data in R Packages	55
2.1.5 Getting Data <i>Out</i> of R	56
2.2 Working With Data Frames	56
2.2.1 The Search Path	57
2.2.2 Avoiding <code>attach</code>	61
2.2.3 Missing Data	62
2.2.4 Numeric Variables and Factors	67
2.2.5 Modifying and Transforming Data	68
2.3 Matrices, Arrays, and Lists	74
2.3.1 Matrices	75
2.3.2 Arrays	76

2.3.3	Lists	77
2.3.4	Indexing	77
2.4	Manipulating Character Data	86
2.5	Handling Large Data Sets in R*	92
2.5.1	How Large is “Large”?	93
2.5.2	Speeding Up <code>read.table</code>	94
2.5.3	Accessing Data in Databases	95
2.6	More on the Representation of Data in R*	96
2.6.1	Length, Class, Mode, and Attributes	96
2.6.2	Pitfalls of Floating-Point Arithmetic	102
2.7	Complementary Reading and References	105
<b>3</b>	<b>Exploring and Transforming Data</b>	<b>107</b>
3.1	Examining Distributions	107
3.1.1	Histograms	107
3.1.2	Density Estimation	110
3.1.3	Quantile-Comparison Plots	111
3.1.4	Boxplots	114
3.2	Examining Relationships	115
3.2.1	Scatterplots	115
3.2.2	Parallel Boxplots	121
3.2.3	More on <code>plot</code>	123
3.3	Examining Multivariate Data	124
3.3.1	Three-Dimensional Plots	124
3.3.2	Scatterplot Matrices	125
3.4	Transforming Data	126
3.4.1	Logarithms	127
3.4.2	Power Transformations	131
3.4.3	Transforming Restricted-Range Variables	133
3.4.4	Other Transformations	134
3.4.5	Transformations to Equalize Spread	135
3.4.6	Transformations Toward Linearity	138
3.4.7	Families of Transformations*	140
3.5	Point Labeling and Identification	145
3.5.1	The <code>identify</code> Function	145
3.5.2	Automatic Labeling of Points	147
3.6	Complementary Reading and References	148
<b>4</b>	<b>Fitting Linear Models</b>	<b>149</b>
4.1	Introduction	149
4.2	Linear Least-Squares Regression	151
4.2.1	Simple Regression	151
4.2.2	Multiple Regression	155
4.2.3	Models with Factors and Interactions	157
4.3	Working With Coefficients	171
4.3.1	Confidence Intervals	172

4.3.2	Estimated Coefficients	172
4.3.3	Effect Displays for Linear Models	172
4.3.4	Coefficient Standard Errors*	181
4.3.5	Standardized Regression Coefficients	183
4.3.6	Sandwich Estimates of Coefficient Standard Errors*	184
4.3.7	Bootstrapping Regression Models	187
4.3.8	Confidence Ellipses and Data Ellipses*	188
4.4	Testing Hypotheses About Regression Coefficients	190
4.4.1	Wald Tests	190
4.4.2	Likelihood Ratio Tests and the Analysis of Variance	191
4.4.3	Sequential Analysis of Variance	192
4.4.4	The Anova Function	193
4.4.5	Testing General Linear Hypotheses*	197
4.4.6	Nonlinear Functions of Coefficients*	200
4.4.7	Testing Regression Models Using Simulation*	202
4.5	Model Selection	207
4.6	More on Factors	213
4.6.1	Coding Factors	213
4.6.2	Other Factor Codings	214
4.6.3	Ordered Factors and Orthogonal Polynomials	216
4.6.4	User-Specified Contrasts*	219
4.7	Overparametrized Models*	220
4.7.1	No Intercept	220
4.7.2	Singularity	221
4.8	The Arguments of the <code>lm</code> Function	223
4.8.1	<code>formula</code>	223
4.8.2	<code>data</code>	225
4.8.3	<code>subset</code>	226
4.8.4	<code>weights</code>	226
4.8.5	<code>na.action</code>	226
4.8.6	<code>method, model, x, y, qr*</code>	227
4.8.7	<code>singular.ok*</code>	227
4.8.8	<code>contrasts</code>	227
4.8.9	<code>offset</code>	227
4.9	Using <code>lm</code> Objects	227
4.10	Complementary Reading and References	228
<b>5</b>	<b>Fitting Generalized Linear Models</b>	<b>229</b>
5.1	The Structure of GLMs	230
5.2	The <code>glm</code> Function in R	233
5.3	GLMs for Binary-Response Data	233

5.3.1	Fitting the Binary Logistic- Regression Model	235
5.3.2	Parameter Estimates for Logistic Regression	237
5.3.3	Analysis of Deviance for Logistic Regression	238
5.3.4	Fitted Values	239
5.4	Binomial Data	240
5.5	Poisson GLMs for Count Data	246
5.6	Loglinear Models for Contingency Tables	250
5.6.1	Two-Dimensional Tables	250
5.6.2	Three-Dimensional Contingency Tables	253
5.6.3	Sampling Plans for Loglinear Models	255
5.6.4	Response Variables	256
5.6.5	Preparing Data for Fitting a Loglinear Model	258
5.7	Multinomial Response Data	259
5.8	Nested Dichotomies	264
5.9	Proportional-Odds Model	269
5.10	Extensions	272
5.10.1	More on the <code>Anova</code> Function	272
5.10.2	Gamma Models	273
5.10.3	Quasi-Likelihood Estimation	276
5.10.4	Overdispersed Binomial and Poisson Models	277
5.10.5	“Rolling Your Own” GLM*	281
5.11	Arguments to <code>glm</code>	281
5.11.1	<code>weights</code>	281
5.11.2	<code>start</code>	282
5.11.3	<code>offset</code>	282
5.11.4	<code>control</code>	282
5.11.5	<code>model, method, x, y</code>	282
5.12	Fitting GLMs by Iterated Weighted Least Squares*	283
5.13	Complementary Reading and References	284
<b>6</b>	<b>Diagnosing Problems in Linear and Generalized Linear Models</b>	<b>285</b>
6.1	Residuals	286
6.2	Basic Diagnostic Plots	287
6.2.1	Plotting Residuals	288
6.2.2	Marginal Model Plots	290
6.2.3	Added-Variable Plots	292
6.3	Unusual Data	294
6.3.1	Outliers and Studentized Residuals	294
6.3.2	Leverage: Hat-Values	296
6.3.3	Influence Measures	298
6.4	Transformations After Fitting a Regression Model	302
6.4.1	Transforming the Response	303

6.4.2	Predictor Transformations	309
6.5	Nonconstant Error Variance	314
6.5.1	Spread-Level Plots	315
6.5.2	Score Tests for Nonconstant Error Variance	316
6.5.3	Other Approaches to Nonconstant Error Variance	316
6.6	Diagnostics for Generalized Linear Models	317
6.6.1	Residuals and Residual Plots	317
6.6.2	Influence Measures	320
6.6.3	Graphical Methods: Added-Variable and Component-Plus-Residual Plots	321
6.7	Collinearity and Variance Inflation Factors	325
6.8	Complementary Reading and References	328
<b>7</b>	<b>Drawing Graphs</b>	<b>329</b>
7.1	A General Approach to R Graphics	330
7.1.1	Defining a Coordinate System: <code>plot</code>	330
7.1.2	Graphics Parameters: <code>par</code>	332
7.1.3	Adding Graphical Elements: <code>axis</code> , <code>points</code> , <code>lines</code> , <code>text</code> , etcetera	334
7.1.4	Specifying Colors	342
7.2	Putting It Together: Explaining Nearest-Neighbor Kernel Regression	344
7.2.1	Finer Control Over Plot Layout	350
7.3	Lattice and Other Graphics Packages in R	352
7.3.1	The Lattice Package	352
7.3.2	Maps	354
7.3.3	Other Notable Graphics Packages	356
7.4	Graphics Devices	357
7.5	Complementary Reading and References	358
<b>8</b>	<b>Writing Programs</b>	<b>359</b>
8.1	Defining Functions	360
8.2	Working With Matrices*	364
8.2.1	Basic Matrix Arithmetic	365
8.2.2	Matrix Inversion and the Solution of Linear Simultaneous Equations	366
8.2.3	Example: Linear Least-Squares Regression	367
8.2.4	Eigenvalues and Eigenvectors	369
8.2.5	Miscellaneous Matrix Computations	369
8.3	Program Control: Conditionals, Loops, and Recursion	370
8.3.1	Conditionals	370
8.3.2	Iteration (Loops)	373
8.3.3	Recursion	375
8.4	<code>apply</code> and Its Relatives	377

8.4.1	To Loop or Not To Loop?	382
8.5	Illustrative R Programs*	386
8.5.1	Binary Logistic Regression	386
8.5.2	Numbers Into Words	392
8.6	Improving R Programs*	398
8.6.1	Debugging R Code	399
8.6.2	Profiling R Functions	402
8.7	Object-Oriented Programming in R*	406
8.7.1	The S3 Object System	406
8.7.2	The S4 Object System	410
8.8	Writing Statistical-Modeling Functions in R*	413
8.9	Environments and Scope in R*	417
8.9.1	Basic Definitions: Frames, Environments, and Scope	417
8.9.2	Lexical Scoping in R	418
8.10	R Programming Advice	422
8.11	Complementary Reading and References	424
<b>References</b>		<b>425</b>
<b>Author Index</b>		<b>431</b>
<b>Subject Index</b>		<b>433</b>
<b>Command Index</b>		<b>442</b>
<b>Data Set Index</b>		<b>449</b>
<b>Package Index</b>		<b>450</b>