
Contents

Preface	ix
Acknowledgments	xvi
1. Introducing R and S-PLUS	1
1.1 S Basics	1
1.1.1 Interacting with the Interpreter	2
1.1.2 S Functions	4
1.1.3 Vectors and Variables	7
1.1.4 User-Defined Functions	12
1.1.5 Cleaning Up	14
1.1.6 Command Editing and Output Management	15
1.1.7 When Things Go Wrong	16
1.2 An Extended Illustration: Duncan's Occupational-Prestige Regression	18
1.2.1 Reading the Data	19
1.2.2 Examining the Data	21
1.2.3 Regression Analysis	26
1.2.4 Regression Diagnostics	28
1.3 S Functions for Basic Statistics	34
2. Reading and Manipulating Data	37
2.1 Data Input	38
2.1.1 Keyboard Input	38
2.1.2 File Input to a Data Frame	42
2.1.3 Importing Data	46
2.1.4 Accessing Data in S Libraries	47
2.1.5 Getting Data <i>Out</i> of S	48
2.2 Working with Data Frames	49
2.2.1 The Search Path	49
2.2.2 Missing Data	54
2.2.3 Numeric Variables and Factors	60
2.2.4 Modifying Data	62

2.3	Matrices, Arrays, and Lists	68
2.3.1	Matrices	69
2.3.2	Arrays	70
2.3.3	Lists	70
2.3.4	Indexing	71
2.4	Data Attributes, Modes, and Classes*	78
2.4.1	Data in S4	82
3.	Exploring and Transforming Data	85
3.1	Examining Distributions	86
3.1.1	Histograms and Stem-and-Leaf Displays	86
3.1.2	Density Estimates	88
3.1.3	Quantile-Comparison Plots	90
3.1.4	Boxplots	91
3.2	Examining Relationships	93
3.2.1	Scatterplots	93
3.2.2	Bivariate Density Estimates	98
3.2.3	Parallel Boxplots	100
3.3	Examining Multivariate Data	102
3.3.1	Scatterplot Matrices	102
3.3.2	Conditioning Plots	105
3.4	Transforming Data	106
3.4.1	Transformations for Normality and Symmetry	109
3.4.2	Transformations to Equalize Spread	113
3.4.3	Transformations to Linearity	115
4.	Fitting Linear Models	119
4.1	Linear Least Squares Regression	119
4.1.1	Simple Regression	119
4.1.2	Multiple Regression	123
4.1.3	Standardized Regression Coefficients	124
4.2	Dummy-Variable Regression	126
4.2.1	Factors	126
4.2.2	Contrasts	127
4.2.3	Ordered Factors	130
4.2.4	Fitting Additive Dummy-Regression Models	130
4.2.5	Dummy Regression with Interactions	133
4.3	Analysis of Variance Models	136
4.4	User-Specified Contrasts*	142
4.5	General Linear Hypotheses*	145

4.6	Data and Confidence Ellipses	147
4.7	More on <code>lm</code> and Model Formulas	149
4.7.1	formula	149
4.7.2	data	150
4.7.3	subset	151
4.7.4	weights	151
4.7.5	na.action	151
4.7.6	method, model, x, y, qr*	152
4.7.7	singular.ok*	152
4.7.8	contrasts	152
4.7.9	offset (R only)	153
5.	Fitting Generalized Linear Models	155
5.1	The Structure of Generalized Linear Models	155
5.2	Models for Categorical Responses	158
5.2.1	Dichotomous Data	158
5.2.2	Polytomous Data	167
5.3	Poisson Generalized Linear Models for Count Data	177
5.3.1	Poisson Regression	178
5.3.2	Log-Linear Models for Contingency Tables	181
5.4	Odds and Ends	185
5.4.1	Other Generalized Linear Models	185
5.4.2	Arguments to <code>glm</code>	188
5.5	Fitting Generalized Linear Models by Iterated Weighted Least Squares*	189
6.	Diagnosing Problems in Linear and Generalized Linear Models	191
6.1	Unusual Data	192
6.1.1	Outliers: Studentized Residuals	192
6.1.2	Leverage: Hat Values	194
6.1.3	Influence Measures	195
6.2	Nonnormal Errors	201
6.2.1	Box-Cox Transformation of y	203
6.2.2	Constructed-Variable Plot for the Box-Cox Transformation	204
6.3	Nonconstant Error Variance	206
6.3.1	Score Tests for Nonconstant Error Variance	208
6.3.2	Other Approaches to Nonconstant Error Variance	208
6.4	Nonlinearity	210
6.4.1	Component + Residual and CERES Plots	210

6.4.2	Box-Tidwell Transformations of the Predictors	214
6.4.3	Constructed-Variable Plots for Box-Tidwell Transformations	215
6.5	Collinearity and Variable Selection	216
6.5.1	Variance-Inflation Factors	216
6.5.2	Variable Selection	220
6.6	Diagnostics for Generalized Linear Models	225
6.6.1	Outlier, Leverage, and Influence Diagnostics	225
6.6.2	Nonlinearity Diagnostics	230
7.	Drawing Graphs	235
7.1	A General Approach to S Graphics	236
7.1.1	Defining a Coordinate System: plot	236
7.1.2	Adding Graphical Elements: axis, points, lines, text, and so on	239
7.1.3	Specifying Colors	246
7.2	Putting it Together: Effect Displays	247
7.3	Graphics Devices	255
8.	Writing Programs	257
8.1	Defining Functions	258
8.2	Working with Matrices*	261
8.3	Program Control: Conditionals, Loops, and Recursion	268
8.3.1	Conditionals	268
8.3.2	Loops (Iteration)	269
8.3.3	Recursion	272
8.3.4	An Extended Illustration: Binary Logistic Regression*	273
8.4	apply and its Relatives	278
8.5	Object-Oriented Programming in S*	283
8.5.1	S Version 3	283
8.5.2	S Version 4	288
8.6	Writing S Programs	292
	References	295
	Index of Data Sets	299
	Index of Functions, Operators, Control Structures, and Other Symbols	300
	Author Index	305
	Subject Index	306
	About the Author	312