

S-Plus is a powerful environment for statistical and graphical analysis of data. It provides the tools to implement many statistical ideas which have been made possible by the widespread availability of workstations having good graphics and computational capabilities. This book is a guide to using S-Plus to perform statistical analyses and provides both an introduction to the use of S-Plus and a course in modern statistical methods.

The aim of the book is to show how to use S-Plus as a powerful and graphical system. Readers are assumed to have a basic grounding in statistics, and so the book is intended for would-be users of S-Plus, and both students and researchers using statistics. Throughout, the emphasis is on presenting practical problems and full analyses of real data sets. Many of the methods discussed are state-of-the-art approaches to topics such as linear and non-linear regression models, robust and smooth regression methods, survival analysis, multivariate analysis, tree-based methods, time series, and spatial statistics. All the data sets and S-Plus functions used are supplied with the book on a diskette.

ISBN 0-387-94350-1
ISBN 3-540-94350-1



Contents

Preface	v
Typographical Conventions	xiii
1 Introduction	1
1.1 A quick overview of S	3
1.2 Getting started	4
1.3 Bailing out	6
1.4 Getting help with functions and features	7
1.5 An introductory session	8
1.6 What next?	16
2 The S Language	17
2.1 A concise description of S objects	17
2.2 Calling conventions for functions	25
2.3 Arithmetical expressions	26
2.4 Reading data	32
2.5 Finding S objects	36
2.6 Character vector operations	38
2.7 Indexing vectors, matrices and arrays	40
2.8 Matrix operations	45
2.9 Functions operating on factors and lists	51
2.10 Input/Output facilities	54
2.11 Customizing your S environment	57
2.12 History and audit trails	59
2.13 Exercises	59
3 Graphical Output	61
3.1 Graphics devices	61
3.2 Basic plotting functions	65
3.3 Enhancing plots	70

3.4	Conditioning plots	74
3.5	Fine control of graphics	76
3.6	Exercises	83
4	Programming in S	85
4.1	Control structures	85
4.2	Writing your own functions	90
4.3	Finding errors	97
4.4	Calling the operating system	103
4.5	Some more advanced features. Recursion and frames	105
4.6	Generic functions and object-oriented programming	110
4.7	Using C and FORTRAN routines	113
4.8	Exercises	119
5	Distributions and Data Summaries	121
5.1	Probability distributions	121
5.2	Generating random data	123
5.3	Data summaries	125
5.4	Classical univariate statistics	129
5.5	Density estimation	134
5.6	Bootstrap and permutation methods	141
5.7	Exercises	146
6	Linear Statistical Models	147
6.1	A linear regression example	147
6.2	Model formulae	153
6.3	Regression diagnostics	157
6.4	Safe prediction	161
6.5	Factorial designs and designed experiments	162
6.6	An unbalanced four-way layout	169
6.7	Multistratum models	177
7	Generalized Linear Models	183
7.1	Functions for generalized linear modelling	187
7.2	Binomial data	189
7.3	Poisson models	196
7.4	A negative binomial family	200

8 Robust Statistics	203
8.1 Univariate samples	204
8.2 Median polish	210
8.3 Robust regression	212
8.4 Resistant regression	217
8.5 Multivariate location and scale	222
9 Non-linear Regression Models	223
9.1 Fitting non-linear regression models	224
9.2 Parametrized data frames	226
9.3 Using function derivative information	226
9.4 Non-linear fitted model objects and method functions	229
9.5 Taking advantage of linear parameters	230
9.6 Examples	231
9.7 Assessing the linear approximation: profiles	237
9.8 General minimization and maximum likelihood estimation	239
10 Modern Regression	247
10.1 Additive models and scatterplot smoothers	247
10.2 Projection-pursuit regression	255
10.3 Response transformation models	258
10.4 Neural networks	261
10.5 Conclusions	265
11 Survival Analysis	267
11.1 Estimators of survivor curves	269
11.2 Parametric models	273
11.3 Cox proportional hazards model	279
11.4 Further examples	285
11.5 Expected survival rates	298
11.6 Superseded functions	299
12 Multivariate Analysis	301
12.1 Graphical methods	301
12.2 Cluster analysis	311
12.3 Discriminant analysis	315
12.4 An example: <i>Leptograpsus variegatus</i> crabs	322

13 Tree-based Methods	329
13.1 Partitioning methods	330
13.2 Cutting trees down to size	342
13.3 Low birth weights revisited	345
14 Time Series	349
14.1 Second-order summaries	352
14.2 ARIMA models	361
14.3 Seasonality	367
14.4 Multiple time series	373
14.5 Nottingham temperature data	376
14.6 Other time-series functions	380
14.7 Backwards compatibility	382
15 Spatial Statistics	383
15.1 Interpolation and kriging	383
15.2 Point process analysis	392
References	397
 Appendices	
A Datasets and Software	407
A.1 Directories	408
A.2 Sources of machine-readable versions	411
A.3 Caveat	412
B Common S-PLUS Functions	413
C S versus S-PLUS	429
D Using S Libraries	431
D.1 Creating a library	433
D.2 Sources of libraries	434
E Command Line Editing	437
F Answers to Selected Exercises	439
Index	447