

# Contents

<b>Preface</b>	<b>v</b>
<b>Prologue</b>	<b>1</b>
<b>1 An Introduction to S-PLUS</b>	<b>5</b>
1.1 Introduction . . . . .	5
1.2 Running S-PLUS . . . . .	5
1.3 Elementary commands . . . . .	7
1.4 Vectors . . . . .	9
1.5 Matrices . . . . .	13
1.6 Logical expressions and subsetting matrices and vectors . .	14
1.7 List objects . . . . .	16
1.8 Data frames . . . . .	18
1.9 Reading in data . . . . .	19
1.10 Some commonly used S-PLUS functions for data analysis .	21
1.11 Data manipulation . . . . .	22
1.12 Modeling in S-PLUS . . . . .	25
1.13 User functions . . . . .	26
1.14 Graphics using the command line approach . . . . .	29
1.15 The S-PLUS GUI . . . . .	34
1.16 Exercises . . . . .	34
<b>2 Describing Data</b>	<b>37</b>
2.1 Introduction . . . . .	37

2.2	Graphing and summarizing continuous data . . . . .	37
2.3	Graphing and summarizing categorical data . . . . .	44
2.4	Dotplots for classified continuous data . . . . .	50
2.5	Summary . . . . .	51
2.6	Using S-PLUS . . . . .	55
2.7	Exercises . . . . .	60
<b>3</b>	<b>Basic Inference</b>	<b>63</b>
3.1	Introduction . . . . .	63
3.2	Simple inference for continuous data . . . . .	63
3.2.1	The independent samples $t$ -test . . . . .	63
3.2.2	Paired $t$ -tests . . . . .	70
3.3	Distribution free tests . . . . .	72
3.3.1	A distribution-free rank sum test—Wilcoxon/ Mann-Whitney test . . . . .	72
3.3.2	A distribution-free signed rank test—Wilcoxon test . . . . .	74
3.4	Simple inference for categorical data . . . . .	74
3.4.1	Test for contingency table data . . . . .	74
3.4.2	Fisher's exact test . . . . .	77
3.4.3	The Mantel-Haenszel test . . . . .	78
3.4.4	McNemar's test for correlated proportions . . . . .	78
3.5	An introduction to the bootstrap . . . . .	81
3.6	Summary . . . . .	82
3.7	Using S-PLUS . . . . .	83
3.8	Exercises . . . . .	86
<b>4</b>	<b>Scatterplots, Simple Regression and Smoothing</b>	<b>89</b>
4.1	Introduction . . . . .	89
4.2	The scatterplot . . . . .	89
4.2.1	Jittering a scatterplot . . . . .	93
4.2.2	Including information about marginal distributions on a scatterplot . . . . .	93
4.2.3	Convex hull trimming of bivariate data . . . . .	95
4.2.4	The aspect ratio of a scatterplot . . . . .	96
4.3	Estimating bivariate densities . . . . .	102
4.4	Simple regression and locally weighted regression . . . . .	105
4.5	Summary . . . . .	107
4.6	Using S-PLUS . . . . .	110
4.7	Exercises . . . . .	118
<b>5</b>	<b>Analysis of Variance and Covariance</b>	<b>121</b>
5.1	Introduction . . . . .	121
5.2	A simple one-way example . . . . .	121
5.3	Multiple comparison procedures . . . . .	125
5.4	A factorial experiment . . . . .	127

5.5	Unbalanced designs . . . . .	130
5.6	Analysis of covariance . . . . .	132
5.7	Summary . . . . .	136
5.8	Using S-PLUS . . . . .	138
5.9	Exercises . . . . .	141
<b>6</b>	<b>The Analysis of Longitudinal Data</b>	<b>143</b>
6.1	Introduction . . . . .	143
6.2	Graphical displays of longitudinal data . . . . .	143
6.3	Summary measure analysis of longitudinal data . . . . .	149
6.4	Summary . . . . .	152
6.5	Using S-PLUS . . . . .	153
6.6	Exercises . . . . .	156
<b>7</b>	<b>More Graphics</b>	<b>157</b>
7.1	Introduction . . . . .	157
7.2	The bubbleplot . . . . .	158
7.3	The scatterplot matrix . . . . .	161
7.4	Conditioning plots . . . . .	162
7.5	Trellis graphics . . . . .	168
7.6	Summary . . . . .	169
7.7	Using S-PLUS . . . . .	173
7.8	Exercises . . . . .	176
<b>8</b>	<b>Multiple Linear Regression</b>	<b>179</b>
8.1	Introduction . . . . .	179
8.2	The multiple linear regression model . . . . .	179
	8.2.1 Anesthesia example . . . . .	180
	8.2.2 Cystic fibrosis example . . . . .	180
8.3	Identifying a good model . . . . .	188
	8.3.1 All possible subsets regression . . . . .	188
	8.3.2 Stepwise methods . . . . .	190
8.4	Checking model assumptions: residuals and other regression diagnostics . . . . .	194
8.5	The linear model . . . . .	195
8.6	Summary . . . . .	199
8.7	Using S-PLUS . . . . .	200
8.8	Exercises . . . . .	203
<b>9</b>	<b>Generalized Linear Models I: Logistic Regression</b>	<b>205</b>
9.1	Introduction . . . . .	205
9.2	Logistic regression for grouped data . . . . .	206
9.3	Logistic regression for individual level data . . . . .	213
9.4	Case-control studies . . . . .	215
9.5	Summary . . . . .	216

9.6	Using S-PLUS . . . . .	218
9.7	Exercises . . . . .	222
<b>10</b>	<b>Generalised linear models II: Poisson regression</b>	<b>223</b>
10.1	Introduction . . . . .	223
10.2	Poisson regression examples . . . . .	225
10.3	Overdispersion . . . . .	230
10.4	Diagnostics . . . . .	234
10.5	Summary . . . . .	234
10.6	Using S-PLUS . . . . .	237
10.7	Exercises . . . . .	242
<b>11</b>	<b>Linear Mixed Models I</b>	<b>243</b>
11.1	Introduction . . . . .	243
11.2	Variance components . . . . .	243
11.3	Mixed effects ANOVA model . . . . .	248
11.4	Linear regression with a random intercept . . . . .	251
11.5	Random coefficient models . . . . .	256
11.6	Summary . . . . .	260
11.7	Using S-PLUS . . . . .	262
11.8	Exercises . . . . .	268
<b>12</b>	<b>Linear Mixed Models II</b>	<b>269</b>
12.1	Introduction . . . . .	269
12.2	Autocorrelated residuals . . . . .	269
12.3	Heteroscedasticity . . . . .	272
12.4	Multilevel models . . . . .	274
12.5	Assessing model assumptions . . . . .	277
12.6	Summary . . . . .	282
12.7	Using S-PLUS . . . . .	283
12.8	Exercises . . . . .	287
<b>13</b>	<b>Generalized Additive Models</b>	<b>291</b>
13.1	Introduction . . . . .	291
13.2	Additive models . . . . .	292
13.3	Fitting additive models . . . . .	292
13.4	Some examples of the use of additive models . . . . .	294
13.4.1	Heights and resting pulse and oxygen uptake and ex- pired ventilation . . . . .	294
13.4.2	Maximal static expiratory pressure for patients with cystic fibrosis . . . . .	294
13.5	Using a logistic additive model . . . . .	296
13.6	Summary . . . . .	298
13.7	Using S-PLUS . . . . .	301
13.7.1	Using the S language . . . . .	301

13.8 Exercises . . . . .	303
<b>14 Nonlinear models</b>	<b>305</b>
14.1 Introduction . . . . .	305
14.2 Nonlinear models . . . . .	305
14.3 Two examples of nonlinear models . . . . .	306
14.4 Age of onset of schizophrenia . . . . .	314
14.5 Summary . . . . .	317
14.6 Using S-PLUS . . . . .	322
14.6.1 Using the S language . . . . .	322
14.7 Exercises . . . . .	326
<b>15 Regression Trees</b>	<b>327</b>
15.1 Introduction . . . . .	327
15.2 Tree-based models . . . . .	328
15.3 Birthweight of babies . . . . .	330
15.3.1 Some simple examples of regression trees for the birth- weight data . . . . .	332
15.4 More complex trees for the birthweight data . . . . .	335
15.5 Summary . . . . .	337
15.6 Using S-PLUS . . . . .	342
15.6.1 Using the S language . . . . .	342
15.7 Exercises . . . . .	344
<b>16 Survival Analysis I</b>	<b>345</b>
16.1 Introduction . . . . .	345
16.2 The survivor function . . . . .	346
16.3 The log-rank test . . . . .	350
16.4 The hazard function . . . . .	352
16.5 Summary . . . . .	353
16.6 Using S-PLUS . . . . .	356
16.6.1 Using the S language . . . . .	356
16.7 Exercises . . . . .	358
<b>17 Survival Analysis II: Cox's Regression</b>	<b>361</b>
17.1 Introduction . . . . .	361
17.2 Cox's regression . . . . .	361
17.3 Left truncation . . . . .	364
17.4 Stratification . . . . .	368
17.5 Diagnostics and assessing the proportional hazards assumption . . . . .	370
17.6 Summary . . . . .	373
17.7 Using S-PLUS . . . . .	375
17.7.1 Using the S language . . . . .	375
17.8 Exercises . . . . .	378

<b>18 Principal Components and Factor Analysis</b>	<b>381</b>
18.1 Introduction . . . . .	381
18.2 Principal components analysis . . . . .	381
18.2.1 Principal components analysis of head lengths . . . .	382
18.2.2 Principal components analysis of nutrients in foodstuffs	384
18.3 Exploratory factor analysis . . . . .	389
18.4 Principal components analysis and factor analysis compared	393
18.5 Summary . . . . .	396
18.6 Using S-PLUS . . . . .	397
18.6.1 Using the S language . . . . .	397
18.7 Exercises . . . . .	399
<b>19 Cluster Analysis</b>	<b>401</b>
19.1 Introduction . . . . .	401
19.2 Agglomerative hierarchical clustering . . . . .	401
19.2.1 Clustering foodstuffs . . . . .	402
19.3 $k$ -means clustering . . . . .	403
19.4 Classification maximum likelihood . . . . .	409
19.5 Summary . . . . .	412
19.6 Using S-PLUS . . . . .	413
19.6.1 Using the S language . . . . .	413
19.7 Exercises . . . . .	416
<b>20 Discriminant Function Analysis</b>	<b>417</b>
20.1 Introduction . . . . .	417
20.2 The linear discriminant function . . . . .	418
20.3 Discriminant functions when group covariance matrices cannot be assumed to be equal . . . . .	424
20.4 Discriminant analysis with more than two groups . . . . .	425
20.5 Summary . . . . .	431
20.6 Using S-PLUS . . . . .	433
20.6.1 Using the S language . . . . .	433
20.7 Exercises . . . . .	437
<b>A The S-PLUS GUI</b>	<b>439</b>
A.1 Introduction . . . . .	439
A.2 Statistical analysis using dialogue boxes . . . . .	439
A.3 Graphics using the GUI . . . . .	443
<b>B Answers to selected exercises</b>	<b>447</b>
<b>References</b>	<b>465</b>
<b>Index</b>	<b>473</b>