

Richard M. Heiberger • Burt Holland

Statistical Analysis and Data Display

An Intermediate Course with Examples in R

Second Edition

 Springer

Contents

Preface	vii
1 Audience	vii
2 Motivation	viii
3 Structure	viii
4 Computation	ix
4.1 R	ix
4.2 The HH Package in R	x
4.3 S-Plus, now called S+	x
4.4 SAS	xi
5 Chapters in the Second Edition	xi
5.1 Revised Chapters	xi
5.2 Revised Appendices	xii
6 Exercises	xii
1 Introduction and Motivation	1
1.1 Statistics in Context	3
1.2 Examples of Uses of Statistics	4
1.2.1 Investigation of Salary Discrimination	4
1.2.2 Measuring Body Fat	5
1.2.3 Minimizing Film Thickness	5
1.2.4 Surveys	5
1.2.5 Bringing Pharmaceutical Products to Market	6
1.3 The Rest of the Book	6
1.3.1 Fundamentals	6
1.3.2 Linear Models	7
1.3.3 Other Techniques	8
1.3.4 New Graphical Display Techniques	9
1.3.5 Appendices on Software	9
1.3.6 Appendices on Mathematics and Probability	10
1.3.7 Appendices on Statistical Analysis and Writing	10

2	Data and Statistics	13
2.1	Types of Data	13
2.2	Data Display and Calculation	14
2.2.1	Presentation	15
2.2.2	Rounding	15
2.3	Importing Data	16
2.3.1	Datasets for This Book	16
2.3.2	Other Data sources	17
2.4	Analysis with Missing Values	17
2.5	Data Rearrangement	18
2.6	Tables and Graphs	18
2.7	R Code Files for <i>Statistical Analysis and Data Display</i> (HH)	19
2.A	Appendix: Missing Values in R	21
3	Statistics Concepts	29
3.1	A Brief Introduction to Probability	29
3.2	Random Variables and Probability Distributions	30
3.2.1	Discrete Versus Continuous Probability Distributions	31
3.2.2	Displaying Probability Distributions—Discrete Distributions	33
3.2.3	Displaying Probability Distributions—Continuous Distributions	35
3.3	Concepts That Are Used When Discussing Distributions	36
3.3.1	Expectation and Variance of Random Variables	36
3.3.2	Median of Random Variables	37
3.3.3	Symmetric and Skewed Distributions	38
3.3.4	Displays of Univariate Data	39
3.3.5	Multivariate Distributions—Covariance and Correlation	44
3.4	Three Probability Distributions	47
3.4.1	The Binomial Distribution	48
3.4.2	The Normal Distribution	49
3.4.3	The (Student's) t Distribution	50
3.5	Sampling Distributions	54
3.6	Estimation	56
3.6.1	Statistical Models	57
3.6.2	Point and Interval Estimators	58
3.6.3	Criteria for Point Estimators	58
3.6.4	Confidence Interval Estimation	60
3.6.5	Example—Confidence Interval on the Mean μ of a Population Having Known Standard Deviation	61
3.6.6	Example—One-Sided Confidence Intervals	61
3.7	Hypothesis Testing	62
3.8	Examples of Statistical Tests	68
3.9	Power and Operating Characteristic (O.C.) (Beta) Curves	69

3.10	Efficiency	71
3.11	Sampling	74
3.11.1	Simple Random Sampling	75
3.11.2	Stratified Random Sampling	76
3.11.3	Cluster Random Sampling	77
3.11.4	Systematic Random Sampling	78
3.11.5	Standard Errors of Sample Means	78
3.11.6	Sources of Bias in Samples	79
3.12	Exercises	80
4	Graphs	85
4.1	What Is a Graph?	86
4.2	Example—Ecological Correlation	87
4.3	Scatterplots	88
4.4	Scatterplot Matrix	89
4.5	Array of Scatterplots	92
4.6	Example—Life Expectancy	93
4.6.1	Study Objectives	93
4.6.2	Data Description	94
4.6.3	Initial Graphs	94
4.7	Scatterplot Matrices—Continued	95
4.8	Data Transformations	100
4.9	Life Expectancy Example—Continued	104
4.10	Color Vision	108
4.11	Exercises	108
4.A	Appendix: R Graphics	111
4.A.1	Cartesian Products	111
4.A.2	Trellis Paradigm	112
4.A.3	Implementation of Trellis Graphics	112
4.A.4	Coordinating Sets of Related Graphs	113
4.A.5	Cartesian Product of Model Parameters	113
4.A.6	Examples of Cartesian Products	114
4.A.7	latticeExtra —Extra Graphical Utilities Based on Lattice	115
4.B	Appendix: Graphs Used in This Book	116
4.B.1	Structured Sets of Graphs	116
4.B.2	Combining Panels	116
4.B.3	Regression Diagnostics	117
4.B.4	Graphs Requiring Multiple Calls to <code>xyp1ot</code>	117
4.B.5	Asymmetric Roles for the Row and Column Sets	119
4.B.6	Rotated Plots	119
4.B.7	Squared Residual Plots	120
4.B.8	Adverse Events Dotplot	120
4.B.9	Microplots	120
4.B.10	Alternate Presentations	120

5 Introductory Inference 123

5.1 Normal (z) Intervals and Tests 123

5.1.1 Test of a Hypothesis Concerning the Mean of a Population Having Known Standard Deviation 124

5.1.2 Confidence Intervals for Unknown Population Proportion p 126

5.1.3 Tests on an Unknown Population Proportion p 127

5.1.4 Example—One-Sided Hypothesis Test Concerning a Population Proportion 127

5.2 t -Intervals and Tests for the Mean of a Population Having Unknown Standard Deviation 129

5.2.1 Example—Inference on a Population Mean μ 130

5.3 Confidence Interval on the Variance or Standard Deviation of a Normal Population 131

5.4 Comparisons of Two Populations Based on Independent Samples 133

5.4.1 Confidence Intervals on the Difference Between Two Population Proportions 133

5.4.2 Confidence Interval on the Difference Between Two Means 134

5.4.3 Tests Comparing Two Population Means When the Samples Are Independent 135

5.4.4 Comparing the Variances of Two Normal Populations ... 138

5.5 Paired Data 139

5.5.1 Example— t -test on Matched Pairs of Means 140

5.6 Sample Size Determination 142

5.6.1 Sample Size for Estimation 143

5.6.2 Sample Size for Hypothesis Testing 144

5.7 Goodness of Fit 148

5.7.1 Chi-Square Goodness-of-Fit Test 149

5.7.2 Example—Test of Goodness-of-Fit to a Discrete Uniform Distribution 150

5.7.3 Example—Test of Goodness-of-Fit to a Binomial Distribution 151

5.8 Normal Probability Plots and Quantile Plots 152

5.8.1 Normal Probability Plots 155

5.8.2 Example—Comparing t -Distributions 156

5.9 Kolmogorov–Smirnov Goodness-of-Fit Tests 158

5.9.1 Example—Kolmogorov–Smirnov Goodness-of-Fit Test .. 158

5.10 Maximum Likelihood 161

5.10.1 Maximum Likelihood Estimation 161

5.10.2 Likelihood Ratio Tests 162

5.11 Exercises 163

- 6 One-Way Analysis of Variance** 167
 - 6.1 Example—Catalyst Data 167
 - 6.2 Fixed Effects 169
 - 6.3 Multiple Comparisons—Tukey Procedure for Comparing
All Pairs of Means 172
 - 6.4 Random Effects 173
 - 6.5 Expected Mean Squares (EMS) 175
 - 6.6 Example—Catalyst Data—Continued 176
 - 6.7 Example—Batch Data 177
 - 6.8 Example—Turkey Data 178
 - 6.8.1 Study Objectives 178
 - 6.8.2 Data Description 179
 - 6.8.3 Analysis 181
 - 6.8.4 Interpretation 181
 - 6.8.5 Specification of Analysis 183
 - 6.9 Contrasts 184
 - 6.9.1 Mathematics of Contrasts 185
 - 6.9.2 Scaling 187
 - 6.10 Tests of Homogeneity of Variance 188
 - 6.11 Exercises 189
 - 6.A Appendix: Computation for the Analysis of Variance 193
 - 6.B Object Oriented Programming 196

- 7 Multiple Comparisons** 199
 - 7.1 Multiple Comparison Procedures 200
 - 7.1.1 Bonferroni Method 200
 - 7.1.2 Tukey Procedure for All Pairwise Comparisons 201
 - 7.1.3 The Dunnett Procedure for Comparing One Mean
with All Others 201
 - 7.1.4 Simultaneously Comparing All Possible Contrasts
Scheffé and Extended Tukey 206
 - 7.2 The Mean–Mean Multiple Comparisons Display (MMC Plot) 212
 - 7.2.1 Difficulties with Standard Displays 212
 - 7.2.2 Hsu and Peruggia’s Mean–Mean Scatterplot 217
 - 7.2.3 Extensions of the Mean–Mean Display to Arbitrary
Contrasts 222
 - 7.2.4 Display of an Orthogonal Basis Set of Contrasts 224
 - 7.2.5 Hsu and Peruggia’s Pulmonary Example 228
 - 7.3 Exercises 232

- 8 Linear Regression by Least Squares** 235
 - 8.1 Introduction 235
 - 8.2 Example—Body Fat Data 236
 - 8.2.1 Study Objectives 236
 - 8.2.2 Data Description 236

8.2.3	Data Input	237
8.2.4	One- X Analysis	238
8.3	Simple Linear Regression	238
8.3.1	Algebra	238
8.3.2	Normal Distribution Theory	240
8.3.3	Calculations	240
8.3.4	Residual Mean Square in Regression Printout	247
8.3.5	New Observations	247
8.4	Diagnostics	254
8.5	ECDF of Centered Fitted Values and Residuals	256
8.6	Graphics	259
8.7	Exercises	260
9	Multiple Regression—More Than One Predictor	263
9.1	Regression with Two Predictors—Least-Squares Geometry	263
9.2	Multiple Regression—Two- X Analysis	264
9.3	Multiple Regression—Algebra	266
9.3.1	The Hat Matrix and Leverage	269
9.3.2	Geometry of Multiple Regression	270
9.4	Programming	271
9.4.1	Model Specification	271
9.4.2	Printout Idiosyncrasies	272
9.5	Example—Albuquerque Home Price Data	272
9.5.1	Study Objectives	272
9.5.2	Data Description	273
9.5.3	Data Input	273
9.6	Partial F -Tests	274
9.7	Polynomial Models	277
9.8	Models Without a Constant Term	281
9.9	Prediction	285
9.10	Example—Longley Data	287
9.10.1	Study Objectives	287
9.10.2	Data Description	287
9.10.3	Discussion	288
9.11	Collinearity	290
9.12	Variable Selection	292
9.12.1	Manual Use of the Stepwise Philosophy	293
9.12.2	Automated Stepwise Regression	297
9.12.3	Automated Stepwise Modeling of the Longley Data	299
9.13	Residual Plots	301
9.13.1	Partial Residuals	301
9.13.2	Partial Residual Plots	303
9.13.3	Partial Correlation	303
9.13.4	Added Variable Plots	304
9.13.5	Interpretation of Residual Plots	304

- 9.14 Example—U.S. Air Pollution Data 306
- 9.15 Exercises 310
- 9.A Appendix: Computation for Regression Analysis 314

- 10 Multiple Regression—Dummy Variables, Contrasts, and Analysis of Covariance 315**
 - 10.1 Dummy (Indicator) Variables 315
 - 10.2 Example—Height and Weight 316
 - 10.2.1 Study Objectives 316
 - 10.2.2 Data Description 317
 - 10.2.3 Data Problems 317
 - 10.2.4 Three Variants on the Analysis 320
 - 10.3 Equivalence of Linear Independent X -Variables (such as Contrasts) for Regression 322
 - 10.4 Polynomial Contrasts and Orthogonal Polynomials 325
 - 10.4.1 Specification and Interpretation of Interaction Terms 330
 - 10.5 Analysis Using a Concomitant Variable (Analysis of Covariance—ANCOVA) 330
 - 10.6 Example—Hot Dog Data 332
 - 10.6.1 Study Objectives 332
 - 10.6.2 Data Description 332
 - 10.6.3 One-Way ANOVA 332
 - 10.6.4 Concomitant Explanatory Variable—ANCOVA 333
 - 10.6.5 Tests of Equality of Regression Lines 340
 - 10.7 `ancovaplot` Function 341
 - 10.8 Exercises 342

- 11 Multiple Regression—Regression Diagnostics 345**
 - 11.1 Example—Rent Data 345
 - 11.1.1 Study Objectives 345
 - 11.1.2 Data Description 345
 - 11.1.3 Rent Levels 346
 - 11.1.4 Alfalfa Rent Relative to Other Rent 350
 - 11.2 Checks on Model Assumptions 356
 - 11.2.1 Scatterplot Matrix 356
 - 11.2.2 Residual Plots 356
 - 11.3 Case Statistics 362
 - 11.3.1 Leverage 363
 - 11.3.2 Deleted Standard Deviation 364
 - 11.3.3 Standardized and Studentized Deleted Residuals 365
 - 11.3.4 Cook’s Distance 366
 - 11.3.5 DFFITS 367
 - 11.3.6 DFBETAS 369
 - 11.3.7 Residuals vs Leverage 371
 - 11.3.8 Calculation of Regression Diagnostics 372
 - 11.4 Exercises 373

12	Two-Way Analysis of Variance	377
12.1	Example—Display Panel Data	377
12.1.1	Study Objectives	377
12.1.2	Data Description	378
12.1.3	Analysis Goals	378
12.2	Statistical Model	382
12.3	Main Effects and Interactions	383
12.4	Two-Way Interaction Plot	385
12.5	Sums of Squares in the Two-Way ANOVA Table	386
12.6	Treatment and Blocking Factors	387
12.7	Fixed and Random Effects	388
12.8	Randomized Complete Block Designs	388
12.9	Example—The Blood Plasma Data	389
12.9.1	Study Objectives	389
12.9.2	Data Description	390
12.9.3	Analysis	390
12.10	Random Effects Models and Mixed Models	393
12.11	Example—Display Panel Data—Continued	394
12.12	Studentized Range Distribution	396
12.13	Introduction to Nesting	397
12.13.1	Example—Workstation Data	397
12.13.2	Data Description	397
12.13.3	Analysis Goals	398
12.14	Example—The <i>Rhizobium</i> Data	400
12.14.1	Study Objectives	400
12.14.2	Data Description	400
12.14.3	First <i>Rhizobium</i> Experiment: Alfalfa Plants	401
12.14.4	Second <i>Rhizobium</i> Experiment: Clover Plants	401
12.14.5	Initial Plots	401
12.14.6	Alfalfa Analysis	403
12.14.7	Clover Analysis	406
12.15	Models Without Interaction	417
12.16	Example—Animal Feed Data	418
12.16.1	Study Objectives	418
12.16.2	Analysis	418
12.17	Exercises	421
12.A	Appendix: Computation for the Analysis of Variance	425
13	Design of Experiments—Factorial Designs	427
13.1	A Three-Way ANOVA—Muscle Data	427
13.2	Latin Square Designs	435
13.2.1	Example—Latin Square	437
13.3	Simple Effects for Interaction Analyses	441
13.3.1	Example—The <i>filmcoat</i> Data	442
13.3.2	Study Objectives	442

- 13.3.3 Data Description 442
- 13.3.4 Data Analysis 443
- 13.4 Nested Factorial Experiment 448
 - 13.4.1 Example—Gunload Data 448
 - 13.4.2 Example—Turkey Data (Continued) 451
- 13.5 Specification of Model Formulas 456
 - 13.5.1 Crossing of Two Factors 462
 - 13.5.2 Example—Dummy Variables for Crossed Factors
Nested Within Another Factor—Turkey Data
(Continued Again) 464
- 13.6 Sequential and Conditional Tests 464
 - 13.6.1 SAS Terminology for Conditional Sums of Squares 466
 - 13.6.2 Example—Application to Clover Data 468
 - 13.6.3 Example—Application to Body Fat Data 470
- 13.7 Exercises 472
- 13.A Appendix: Orientation for Boxplots 478
- 14 Design of Experiments—Complex Designs 479**
 - 14.1 Confounding 479
 - 14.2 Split Plot Designs 481
 - 14.3 Example—Yates Oat Data 482
 - 14.3.1 Alternate Specification 487
 - 14.3.2 Polynomial Effects for Nitrogen 489
 - 14.4 Introduction to Fractional Factorial Designs 492
 - 14.4.1 Example— 2^{8-2} Design 493
 - 14.4.2 Example— 2^{5-1} Design 495
 - 14.5 Introduction to Crossover Designs 497
 - 14.5.1 Example—Two Latin Squares 498
 - 14.6 ANCOVA with Blocks: Example—Apple Tree Data 501
 - 14.6.1 Study Objectives 502
 - 14.6.2 Data Description 502
 - 14.6.3 Data Analysis 502
 - 14.6.4 Model 1: `yield ~ block + pre * treat` 504
 - 14.6.5 Model 2: `yield.block ~ pre.block * treat` 506
 - 14.6.6 Model 3: `yield.block ~ pre.block` 508
 - 14.6.7 Model 4: `yield.block ~ treat` 509
 - 14.6.8 Model 5: `yield.block ~ pre.block + treat` 509
 - 14.6.9 Model 6: `yield.block.pre ~ treat` 511
 - 14.6.10 Multiple Comparisons 513
 - 14.7 Example—testscore 516
 - 14.7.1 Study Objectives 516
 - 14.7.2 Data Description 516
 - 14.7.3 Analysis—Plots 517
 - 14.7.4 Analysis—ANOVA 517
 - 14.7.5 Summary of ANOVA 520

14.8	The Tukey One Degree of Freedom for Nonadditivity	524
14.8.1	Example—Crash Data—Study Objectives	524
14.8.2	Data Description	525
14.8.3	Data Analysis	525
14.8.4	Theory	534
14.9	Exercises	535
15	Bivariate Statistics—Discrete Data	539
15.1	Two-Dimensional Contingency Tables—Chi-Square Analysis	539
15.1.1	Example—Drunkness Data	539
15.1.2	Chi-Square Analysis	542
15.2	Two-Dimensional Contingency Tables—Fisher’s Exact Test	545
15.2.1	Example—Do Juvenile Delinquents Eschew Wearing Eyeglasses?	545
15.3	Simpson’s Paradox	548
15.4	Relative Risk and Odds Ratios	552
15.4.1	Glasses (Again)	552
15.4.2	Large Sample Approximations	553
15.4.3	Example—Treating Cardiac Arrest with Therapeutic Hypothermia	555
15.5	Retrospective and Prospective Studies	558
15.6	Mantel–Haenszel Test	559
15.7	Example—Salk Polio Vaccine	563
15.8	Example—Adverse Experiences	565
15.9	Ordered Categorical Scales, Including Rating Scales	567
15.9.1	Display of Professional Challenges Dataset	568
15.9.2	Single-Panel Displays	570
15.9.3	Multiple-Panel Displays	572
15.10	Exercises	573
16	Nonparametrics	577
16.1	Introduction	577
16.2	Sign Test for the Location of a Single Population	578
16.3	Comparing the Locations of Paired Populations	581
16.3.1	Sign Test	581
16.3.2	Wilcoxon Signed-Ranks Test	582
16.4	Mann–Whitney Test for Two Independent Samples	586
16.5	Kruskal–Wallis Test for Comparing the Locations of at Least Three Populations	590
16.6	Exercises	591
17	Logistic Regression	593
17.1	Example—The Space Shuttle Challenger Disaster	595
17.1.1	Study Objectives	595
17.1.2	Data Description	595

17.1.3	Graphical Display	596
17.1.4	Numerical Display	599
17.2	Estimation	603
17.3	Example—Budworm Data	605
17.4	Example—Lymph Nodes	609
17.4.1	Data	610
17.4.2	Data Analysis	610
17.4.3	Additional Techniques	612
17.4.4	Diagnostics	619
17.5	Numerical Printout	619
17.6	Graphics	620
17.6.1	Conditioned Scatterplots	620
17.6.2	Common Scaling in Comparable Plots	621
17.6.3	Functions of Predicted Values	622
17.7	Model Specification	622
17.7.1	Fitting Models When the Response Is Dichotomous	622
17.7.2	Fitting Models When the Response Is a Sample Proportion	623
17.8	LogXact	624
17.9	Exercises	625
18	Time Series Analysis	631
18.1	Introduction	631
18.2	The ARIMA Approach to Time Series Modeling	632
18.2.1	AutoRegression (AR)	633
18.2.2	Moving Average (MA)	634
18.2.3	Differencing	635
18.2.4	Autoregressive Integrated Moving Average (ARIMA)	635
18.3	Autocorrelation	636
18.3.1	Autocorrelation Function (ACF)	636
18.3.2	Partial Autocorrelation Function (PACF)	637
18.4	Analysis Steps	637
18.5	Some Algebraic Development, Including Forecasting	640
18.5.1	The General ARIMA Model	640
18.5.2	Special Case—The AR(1) Model	641
18.5.3	Special Case—The MA(1) Model	642
18.6	Graphical Displays for Time Series Analysis	642
18.7	Models with Seasonal Components	648
18.7.1	Multiplicative Seasonal ARIMA Models	648
18.7.2	Example—co ₂ ARIMA(0, 1, 1) × (0, 1, 1) ₁₂ Model	649
18.7.3	Determining the Seasonal AR and MA Parameters	649
18.8	Example of a Seasonal Model—The Monthly co ₂ Data	650
18.8.1	Identification of the Model	650
18.8.2	Parameter Estimation and Diagnostic Checking	652
18.8.3	Forecasting	661

18.9	Exercises	661
18.A	Appendix: Construction of Time Series Graphs	692
18.A.1	Characteristics of This Presentation of the Time Series Plot	694
18.A.2	Characteristics of This Presentation of the Sample ACF and PACF Plots	695
18.A.3	Construction of Graphical Displays	695
18.A.4	Functions in the HH package for R	696
A	R	699
A.1	Installing R —Initial Installation	699
A.1.1	Packages Needed for This Book—Macintosh and Linux	700
A.1.2	Packages and Other Software Needed for This Book—Windows	700
A.1.3	Installation Problems—Any Operating System	703
A.1.4	XLConnect: All Operating Systems	703
A.2	Installing R —Updating	704
A.3	Using R	704
A.3.1	Starting the R Console	704
A.3.2	Making the Functions in the HH Package Available to the Current Session	705
A.3.3	Access HH Datasets	705
A.3.4	Learning the R Language	705
A.3.5	Duplicating All HH Examples	706
A.3.6	Learning the Functions in R	706
A.3.7	Learning the lattice Functions in R	707
A.3.8	Graphs in an Interactive Session	707
A.4	S/R Language Style	708
A.5	Getting Help While Learning and Using R	711
A.6	R Inexplicable Error Messages—Some Debugging Hints	712
B	HH	715
B.1	Contents of the HH Package	715
B.2	R Scripts for all Figures and Tables in the Book	716
B.2.1	Macintosh	716
B.2.2	Linux	716
B.2.3	Windows	717
B.3	Functions in the HH Package	717
B.4	HH and S+	717
C	Rcmdr: R Commander	719

- D RExcel: Embedding R inside Excel on Windows** 733
 - D.1 Installing RExcel for Windows 734
 - D.1.1 Install R 734
 - D.1.2 Install Two R Packages Needed by RExcel 734
 - D.1.3 Install RExcel and Related Software 735
 - D.1.4 Install Rcmdr to Work with RExcel 735
 - D.1.5 Additional Information on Installing RExcel 735
 - D.2 Using RExcel 736
 - D.2.1 Automatic Recalculation of an R Function 736
 - D.2.2 Transferring Data To/From R and Excel 737
 - D.2.3 Control of a lattice Plot from an Excel/Rcmdr Menu ... 740

- E Shiny: Web-Based Access to R Functions** 743
 - E.1 NTplot 744
 - E.2 bivariateNormal 745
 - E.3 bivariateNormalScatterplot 746
 - E.4 PopulationPyramid 747

- F R Packages** 749
 - F.1 What Is a Package? 749
 - F.2 Installing and Loading R Packages 749
 - F.3 Where Are the Packages on Your Computer? 750
 - F.4 Structure of an R Package 751
 - F.5 Writing and Building Your Own Package 751
 - F.6 Building Your Own Package with Windows 752

- G Computational Precision and Floating-Point Arithmetic** 753
 - G.1 Examples 753
 - G.2 Floating Point Numbers in the IEEE 754 Floating-Point Standard 755
 - G.3 Multiple Precision Floating Point 756
 - G.4 Binary Format 757
 - G.5 Round to Even 758
 - G.6 Base-10, 2-Digit Arithmetic 758
 - G.7 Why Is .9 Not Recognized to Be the Same as (.3 + .6)? 760
 - G.8 Why Is $(\sqrt{2})^2$ Not Recognized to Be the Same as 2? 760
 - G.9 zapsmall to Round Small Values to Zero for Display 760
 - G.10 Apparent Violation of Elementary Factoring 762
 - G.11 Variance Calculations 763
 - G.12 Variance Calculations at the Precision Boundary 763
 - G.13 Can the Answer to the Calculation be Represented? 769
 - G.14 Explicit Loops 770

- H Other Statistical Software** 773

I	Mathematics Preliminaries	775
I.1	Algebra Review	775
I.1.1	Line	776
I.1.2	Parabola	776
I.1.3	Ellipse	777
I.1.4	Simultaneous Equations	777
I.1.5	Exponential and Logarithm Functions	778
I.1.6	Asymptote	780
I.2	Elementary Differential Calculus	780
I.3	An Application of Differential Calculus	781
I.4	Topics in Matrix Algebra	782
I.4.1	Elementary Operations	784
I.4.2	Linear Independence	786
I.4.3	Rank	787
I.4.4	Quadratic Forms	787
I.4.5	Orthogonal Transformations	788
I.4.6	Orthogonal Basis	788
I.4.7	Matrix Factorization— <i>QR</i>	789
I.4.8	Modified Gram–Schmidt (MGS) Algorithm	789
I.4.9	Matrix Factorization—Cholesky	793
I.4.10	Orthogonal Polynomials	793
I.4.11	Projection Matrices	794
I.4.12	Geometry of Matrices	795
I.4.13	Eigenvalues and Eigenvectors	796
I.4.14	Singular Value Decomposition	797
I.4.15	Generalized Inverse	801
I.4.16	Solving Linear Equations	803
I.5	Combinations and Permutations	804
I.5.1	Factorial	804
I.5.2	Permutations	804
I.5.3	Combinations	805
I.6	Exercises	805
J	Probability Distributions	807
J.1	Continuous Central Distributions	808
J.1.1	Beta	808
J.1.2	Cauchy	809
J.1.3	Chi-Square	809
J.1.4	Exponential	810
J.1.5	F	811
J.1.6	Gamma	811
J.1.7	Log Normal	812
J.1.8	Logistic	813
J.1.9	Normal	814
J.1.10	Studentized Range Distribution	815

- J.1.11 (Student’s) *T* 816
- J.1.12 Uniform 816
- J.1.13 Weibull 817
- J.2 Noncentral Continuous Probability Distributions 817
 - J.2.1 Chi-Square: Noncentral 819
 - J.2.2 *T*: Noncentral 819
 - J.2.3 *F*: Noncentral 820
- J.3 Discrete Distributions 820
 - J.3.1 Discrete Uniform 822
 - J.3.2 Binomial 822
 - J.3.3 Geometric 823
 - J.3.4 Hypergeometric 824
 - J.3.5 Negative Binomial 825
 - J.3.6 Poisson 825
 - J.3.7 Signed Rank 826
 - J.3.8 Wilcoxon 827
- J.4 Multivariate Distributions 828
 - J.4.1 Multinomial 828
 - J.4.2 Multivariate Normal 829
- K Working Style** 831
 - K.1 Text Editor 831
 - K.1.1 Requirements for an Editor 832
 - K.1.2 Choice of Editor 833
 - K.2 Types of interaction with **R** 833
 - K.3 Script File 834
 - K.4 Directory Structure 834
 - K.4.1 Directory Structure of This Book 835
 - K.4.2 Directory Structure for Users of This Book 835
 - K.4.3 Other User Directories 836
- L Writing Style** 837
 - L.1 Typographic Style 837
 - L.2 Graphical Presentation Style 841
 - L.2.1 Resolution 841
 - L.2.2 Aspect Ratio 841
 - L.2.3 Other Features 843
 - L.3 English Writing Style 844
 - L.4 Programming Style and Common Errors 845
 - L.5 Presentation of Results 847
- M Accessing **R** Through a Powerful Editor—With **Emacs** and **ESS** as the Example** 851
 - M.1 Emacs Features 852
 - M.1.1 Text Editing 852

M.1.2	File Comparison	852
M.1.3	Buffers	853
M.1.4	Shell Mode	854
M.1.5	Controlling Other Programs	854
M.2	ESS	854
M.2.1	Syntactic Indentation and Color/Font-Based Source Code Highlighting	856
M.2.2	Partial Code Evaluation	856
M.2.3	Object Name Completion	857
M.2.4	Process Interaction	857
M.2.5	Interacting with Statistical Programs on Remote Computers	858
M.2.6	Transcript Editing and Reuse	858
M.2.7	Help File Editing (R)	858
M.3	Learning Emacs	858
M.3.1	GUI (Graphical User Interface)	859
M.3.2	Keyboard Interface	859
M.4	Nuisances with Windows and Emacs	860
M.5	Requirements	860
N	L^AT_EX	863
N.1	Organization Using L^AT_EX	863
N.2	Setting Equations	864
N.3	Coordination with R	864
N.4	Global Changes: Specification of Fonts	864
O	Word Processors and Spreadsheets	867
O.1	Microsoft Word	867
O.1.1	Editing Requirements	868
O.1.2	SWord	868
O.2	Microsoft Excel	869
O.2.1	Database Management	869
O.2.2	Organizing Calculations	869
O.2.3	Excel as a Statistical Calculator	869
	References	873
	Index of Datasets	887
	Index	889