Contents

About This Book ix
Acknowledgments xiii

Part 1 The Basics—Including Some Nuts and Bolts 1

Chapter 1 Let's Get Started—Preliminaries and a SAS Quick Start 3

1.1 Statistical Computing versus Programming versus Managing Data 4
1.2 Good Programming Practice 4
1.3 SAS Program Structure 9
1.4 What Is a SAS Data Set? 16
1.5 Internally Documenting SAS Programs 18
1.6 Summary 19
1.7 References 19
1.8 Exercises 19

Chapter 2 Reading, Combining, and Managing Data for Later Analysis 23

2.1 Temporary versus Permanent Status of Data Sets 24
2.2 Reading Data into a SAS Data Set 27
2.3 Writing Out a File or Making a Simple Report 39
2.4 Concatenating Data Sets and Adding Observations 46
2.5 Merging Data Sets and Adding Variables 50
2.6 Database Processing with PROC SQL 58
2.7 Summary 70
2.8 References 71
2.9 Exercises 71
2.10 Self-Study Laboratory Explorations 73

Chapter 3 Using SAS Procedures 85

3.1 SAS System Options 86
3.2 Statements That Can Modify the Output of Most Procedures 89
3.3 Defining Your Own Formats for Variable Values 91
3.4 Selecting or Stratifying an Analysis by Values of a Variable ... 93
3.5 Displaying Data Set Properties and Observations 95
3.6 Using PROC PRINT to List the Observations in a Data Set 96
3.7 Basic Graphical Displays 97
3.8 Using Scatter Plots to Display Relationships between Numeric Variables 100
3.9 Summarizing Categorical Variables 106
3.10 Summarizing Numeric Variables 108
3.11 Selecting a Simple Random Sample 115
3.12 Randomly Assigning Treatments to Observations 116
3.13 Summary 118
3.14 References 118
3.15 Exercises 119

Chapter 4 Complex Table Construction and Output Control 121

4.1 Introducing PROC TABULATE 122
4.2 Building from Simple Specifications 124
4.3 Enhancing PROC TABULATE Output 128
4.4 Using the Output Delivery System 132
4.5 Summary 151
4.6 References 151
4.7 Exercises 151

Chapter 5 Basic Models in SAS 155

5.1 Overview of Modeling 156
5.2 Linear Regression Models 156
5.3 ANOVA Models—PROC GLM for a One-Way ANOVA 177
5.4 ANOVA Models—PROC GLM for an ANOVA Model with Two or More Factors 188
5.5 Summary 194
5.6 References 194
5.7 Exercises 194

Chapter 6 Producing Statistical Graphics in SAS 197

6.1 Graphics in SAS 197
6.2 ODS Statistical Graphics 198
6.3 Modifying Graphics Using the ODS Graphics Editor 206
6.4 Graphing with Styles and Templates 209
6.5 Statistical Graphics—Entering the Land of SG Procedures 216
6.6 Case Study: Using the SG Procedures 224
6.7 Summary 233
6.8 References 234
6.9 Exercises 235

Chapter 7 Traditional SAS Graphics 237

7.1 Traditional SAS Graphics 238
7.2 Customizing Graphics 238
7.3 Why You Need to Learn about Annotate Data Sets 244
7.4 Case Study: Comparing Distributions of Responses 250
7.5 Descriptive Displays of Spatial Data 254
7.6 Summary 256
7.7 References 257

7.8 Exercises 257
7.9 Appendix: Complete Ohio County Population Data Set 258

Part 2 Doing More with Programming 261

Chapter 8 Formatting Variables, Recoding Variables, and Writing Programs 263

8.1 Internal Representations and Output Displays 264
8.2 Character, Numeric, Time, and Date Formats 272
8.3 Recoding and Transforming Variables in a DATA Step 277
8.4 Ordering How Tasks Are Done 285
8.5 What Goes and What Stays in a Data Set 287
8.6 Structured Thinking about Writing Programs 289
8.7 Case Study 1: Is the Two-Sample *t*-Test Robust Enough for Heterogeneous Variances? 290
8.8 Case Study 2: Monte Carlo Integration to Estimate an Integral 305
8.9 Case Study 3: Simple Percentile-Based Bootstrap 310
8.10 Throw Out Your Tables of Statistical Distributions 312
8.11 Generating Variables Using Random Number Generators 320
8.12 Summary 322
8.13 References 323
8.14 Exercises 323

Chapter 9 Programming in a DATA Step 325

9.1 Storage Bins for Collections of Values 326
9.2 Case Study 1: Monte Carlo *p*-Value for Test of Spatial Randomness 332
9.3 Remembering Variable Values across Observations 336
9.4 Case Study 2: Randomization Test for the Equality of Two Populations 340

9.5 Summary 346
9.6 References 346
9.7 Exercises 347

Chapter 10 Macro Programming 351

10.1 What Is a Macro and Why Would You Use It? 352
10.2 Motivation for Macros: Numerical Integration to Determine P(0<Z<1.645) 352
10.3 Processing Macros 357
10.4 Macro Variables, Parameters, and Functions 357
10.5 Conditional Execution, Looping, and Macros 360
10.6 Debugging Macro Code and Programs 367
10.7 Saving Macros 374
10.8 Functions and Routines for Macros 375
10.9 Bonus Material: Processing Multiple Data Sets 379
10.10 Summary 385
10.11 References 385
10.12 Exercises 385

Chapter 11 Programming with Matrices and Vectors 389

11.1 Defining a Matrix and Subscripting 390
11.2 Using Diagonal Matrices and Stacking Matrices 396
11.3 Using Elementwise Operations, Repeating, and Multiplying Matrices 399
11.4 Importing a Data Set into SAS/IML and Exporting Matrices from SAS/IML to a Data Set 400
11.5 Case Study 1: Monte Carlo Integration to Estimate π 402
11.6 Case Study 2: Bisection Root Finder 403
11.7 Case Study 3: Randomization Test Using Matrices Imported from PROC PLAN 406
11.8 Case Study 4: SAS/IML Module to Implement Monte Carlo Integration to Estimate π 408

11.9 Introducing SAS/IML Studio 411
11.10 Summary 422
11.11 References 422
11.12 Exercises 423

Index 427