

---

# Contents

**About This Book xvii**

**Acknowledgments xxv**

**About the Author xxvii**

## **Part 1 Data Preparation 1**

### **Chapter 1 Moving, Copying, Importing, and Exporting Data 3**

#### **1.1 LIBNAME Statement Engines 4**

1.1.1 Using Data Access Engines to Read and Write Data 5

1.1.2 Using the Engine to View the Data 6

1.1.3 Options Associated with the Engine 6

1.1.4 Replacing EXCEL Sheets 7

1.1.5 Recovering the Names of EXCEL Sheets 8

#### **1.2 PROC IMPORT and EXPORT 9**

1.2.1 Using the Wizard to Build Sample Code 9

1.2.2 Control through the Use of Options 9

1.2.3 PROC IMPORT Data Source Statements 10

1.2.4 Importing and Exporting CSV Files 12

1.2.5 Preventing the Export of Blank Sheets 15

1.2.6 Working with Named Ranges 16

#### **1.3 DATA Step INPUT Statement 17**

1.3.1 Format Modifiers for Errors 18

1.3.2 Format Modifiers for the INPUT Statement 18

1.3.3 Controlling Delimited Input 20

1.3.4 Reading Variable-Length Records 24

#### **1.4 Writing Delimited Files 28**

1.4.1 Using the DATA Step with the DLM= Option 28

1.4.2 PROC EXPORT 29

1.4.3 Using the %DS2CSV Macro 30

1.4.4 Using ODS and the CSV Destination 31

1.4.5 Inserting the Separator Manually 31

#### **1.5 SQL Pass-Through 32**

1.5.1 Adding a Pass-Through to Your SQL Step 32

1.5.2 Pass-Through Efficiencies 33

#### **1.6 Reading and Writing to XML 33**

1.6.1 Using ODS 34

1.6.2 Using the XML Engine 34

## **Chapter 2 Working with Your Data 37**

- 2.1 Data Set Options 38**
  - 2.1.1 REPLACE and REPEMPTY 40
  - 2.1.2 Password Protection 41
  - 2.1.3 KEEP, DROP, and RENAME Options 42
  - 2.1.4 Observation Control Using FIRSTOBS and OBS Data Set Options 43
- 2.2 Evaluating Expressions 45**
  - 2.2.1 Operator Hierarchy 45
  - 2.2.2 Using the Colon as a Comparison Modifier 46
  - 2.2.3 Logical and Comparison Operators in Assignment Statements 47
  - 2.2.4 Compound Inequalities 49
  - 2.2.5 The MIN and MAX Operators 50
  - 2.2.6 Numeric Expressions and Boolean Transformations 51
- 2.3 Data Validation and Exception Reporting 52**
  - 2.3.1 Date Validation 52
  - 2.3.2 Writing to an Error Data Set 55
  - 2.3.3 Controlling Exception Reporting with Macros 58
- 2.4 Normalizing - Transposing the Data 60**
  - 2.4.1 Using PROC TRANSPOSE 61
  - 2.4.2 Transposing in the DATA Step 63
- 2.5 Filling Sparse Data 65**
  - 2.5.1 Known Template of Rows 65
  - 2.5.2 Double Transpose 67
  - 2.5.3 Using COMPLETYPES with PROC MEANS or PROC SUMMARY 70
  - 2.5.4 Using CLASSDATA 70
  - 2.5.5 Using Preloaded Formats 72
  - 2.5.6 Using the SPARSE Option with PROC FREQ 73
- 2.6 Some General Concepts 73**
  - 2.6.1 Shorthand Variable Naming 73
  - 2.6.2 Understanding the ORDER= Option 77
  - 2.6.3 Quotes within Quotes within Quotes 79
  - 2.6.4 Setting the Length of Numeric Variables 81
- 2.7 WHERE Specifics 82**
  - 2.7.1 Operators Just for the WHERE 83
  - 2.7.2 Interaction with the BY Statement 86
- 2.8 Appending Data Sets 88**
  - 2.8.1 Appending Data Sets Using the DATA Step and SQL UNION 88
  - 2.8.2 Using the DATASETS Procedure's APPEND Statement 90

**2.9 Finding and Eliminating Duplicates 90**

- 2.9.1 Using PROC SORT 91
- 2.9.2 Using FIRST. and LAST. BY-Group Processing 92
- 2.9.3 Using PROC SQL 93
- 2.9.4 Using PROC FREQ 93
- 2.9.5 Using the Data Component Hash Object 94

**2.10 Working with Missing Values 97**

- 2.10.1 Special Missing Values 97
- 2.10.2 MISSING System Option 98
- 2.10.3 Using the CMISS, NMISS, and MISSING Functions 99
- 2.10.4 Using the CALL MISSING Routine 100
- 2.10.5 When Classification Variables are Missing 100
- 2.10.6 Missing Values and Macro Variables 101
- 2.10.7 Imputing Missing Values 101

**Chapter 3 Just In the DATA Step 103****3.1 Working across Observations 105**

- 3.1.1 BY-Group Processing—Using FIRST. and LAST. Processing 105
- 3.1.2 Transposing to ARRAYS 107
- 3.1.3 Using the LAG Function 108
- 3.1.4 Look-Ahead Using a MERGE Statement 110
- 3.1.5 Look-Ahead Using a Double SET Statement 111
- 3.1.6 Look-Back Using a Double SET Statement 111
- 3.1.7 Building a FIFO Stack 113
- 3.1.8 A Bit on the SUM Statement 114

**3.2 Calculating a Person's Age 114**

- 3.2.1 Simple Formula 115
- 3.2.2 Using Functions 116
- 3.2.3 The Way Society Measures Age 117

**3.3 Using DATA Step Component Objects 117**

- 3.3.1 Declaring (Instantiating) the Object 119
- 3.3.2 Using Methods with an Object 119
- 3.3.3 Simple Sort Using the HASH Object 120
- 3.3.4 Stepping through a Hash Table 121
- 3.3.5 Breaking Up a Data Set into Multiple Data Sets 126
- 3.3.6 Hash Tables That Reference Hash Tables 128
- 3.3.7 Using a Hash Table to Update a Master Data Set 130

**3.4 Doing More with the INTNX and INTCK Functions 132**

- 3.4.1 Interval Multipliers 132
- 3.4.2 Shift Operators 133
- 3.4.3 Alignment Options 134
- 3.4.4 Automatic Dates 136

- 3.5 Variable Conversions 138**
  - 3.5.1 Using the PUT and INPUT Functions 138
  - 3.5.2 Decimal, Hexadecimal, and Binary Number Conversions 143
- 3.6 DATA Step Functions 143**
  - 3.6.1 The ANY and NOT Families of Functions 144
  - 3.6.2 Comparison Functions 145
  - 3.6.3 Concatenation Functions 147
  - 3.6.4 Finding Maximum and Minimum Values 147
  - 3.6.5 Variable Information Functions 148
  - 3.6.6 New Alternatives and Functions That Do More 154
  - 3.6.7 Functions That Put the Squeeze on Values 163
- 3.7 Joins and Merges 165**
  - 3.7.1 BY Variable Attribute Consistency 166
  - 3.7.2 Variables in Common That Are Not in the BY List 169
  - 3.7.3 Repeating BY Variables 170
  - 3.7.4 Merging without a Clear Key (Fuzzy Merge) 171
- 3.8 More on the SET Statement 172**
  - 3.8.1 Using the NOBS= and POINT= Options 172
  - 3.8.2 Using the INDSNAME= Option 174
  - 3.8.3 A Comment on the END= Option 175
  - 3.8.4 DATA Steps with Two SET Statements 175
- 3.9 Doing More with DO Loops 176**
  - 3.9.1 Using the DOW Loop 176
  - 3.9.2 Compound Loop Specifications 178
  - 3.9.3 Special Forms of Loop Specifications 178
- 3.10 More on Arrays 180**
  - 3.10.1 Array Syntax 180
  - 3.10.2 Temporary Arrays 181
  - 3.10.3 Functions Used with Arrays 182
  - 3.10.4 Implicit Arrays 183

## **Chapter 4 Sorting the Data 185**

- 4.1 PROC SORT Options 186**
  - 4.1.1 The NODUPREC Option 186
  - 4.1.2 The DUPOUT= Option 187
  - 4.1.3 The TAGSORT Option 188
  - 4.1.4 Using the SORTSEQ Option 188
  - 4.1.5 The FORCE Option 190
  - 4.1.6 The EQUALS or NOEQUALS Options 190
- 4.2 Using Data Set Options with PROC SORT 190**
- 4.3 Taking Advantage of Known or Knowable Sort Order 191**

4.4 Metadata Sort Information 193

4.5 Using Threads 194

## **Chapter 5 Working with Data Sets 197**

5.1 Automating the COMPARE Process 198

5.2 Reordering Variables on the PDV 200

5.3 Building and Maintaining Indexes 202

5.3.1 Introduction to Indexing 203

5.3.2 Creating Simple Indexes 204

5.3.3 Creating Composite Indexes 206

5.3.4 Using the IDXWHERE and IDXNAME Options 206

5.3.5 Index Caveats and Considerations 207

5.4 Protecting Passwords 208

5.4.1 Using PROC PWENCODE 208

5.4.2 Protecting Database Passwords 209

5.5 Deleting Data Sets 211

5.6 Renaming Data Sets 211

5.6.1 Using the RENAME Function 212

5.6.2 Using PROC DATASETS 212

## **Chapter 6 Table Lookup Techniques 213**

6.1 A Series of IF Statements—The Logical Lookup 215

6.2 IF -THEN/ELSE Lookup Statements 215

6.3 DATA Step Merges and SQL Joins 216

6.4 Merge Using Double SET Statements 218

6.5 Using Formats 219

6.6 Using Indexes 221

6.6.1 Using the BY Statement 222

6.6.2 Using the KEY= Option 222

6.7 Key Indexing (Direct Addressing)—Using Arrays to Form a Simple Hash 223

6.7.1 Building a List of Unique Values 223

6.7.2 Performing a Key Index Lookup 224

6.7.3 Using a Non-Numeric Index 226

6.8 Using the HASH Object 227

## **Part 2 Data Summary, Analysis, and Reporting 231**

### **Chapter 7 MEANS and SUMMARY Procedures 233**

- 7.1 Using Multiple CLASS Statements and CLASS Statement Options 234**
  - 7.1.1 MISSING and DESCENDING Options 236
  - 7.1.2 GROUPINTERNAL Option 237
  - 7.1.3 Order= Option 238
- 7.2 Letting SAS Name the Output Variables 238**
- 7.3 Statistic Specification on the OUTPUT Statement 240**
- 7.4 Identifying the Extremes 241**
  - 7.4.1 Using the MAXID and MINID Options 241
  - 7.4.2 Using the IDGROUP Option 243
  - 7.4.3 Using Percentiles to Create Subsets 245
- 7.5 Understanding the \_TYPE\_ Variable 246**
- 7.6 Using the CHARTYPE Option 248**
- 7.7 Controlling Summary Subsets Using the WAYS Statement 249**
- 7.8 Controlling Summary Subsets Using the TYPES Statement 250**
- 7.9 Controlling Subsets Using the CLASSDATA= and EXCLUSIVE Options 251**
- 7.10 Using the COMPLETETYPES Option 253**
- 7.11 Identifying Summary Subsets Using the LEVELS and WAYS Options 254**
- 7.12 CLASS Statement vs. BY Statement 255**

### **Chapter 8 Other Reporting and Analysis Procedures 257**

- 8.1 Expanding PROC TABULATE 258**
  - 8.1.1 What You Need to Know to Get Started 258
  - 8.1.2 Calculating Percentages Using PROC TABULATE 262
  - 8.1.3 Using the STYLE= Option with PROC TABULATE 265
  - 8.1.4 Controlling Table Content with the CLASSDATA Option 267
  - 8.1.5 Ordering Classification Level Headings 269
- 8.2 Expanding PROC UNIVARIATE 270**
  - 8.2.1 Generating Presentation-Quality Plots 270
  - 8.2.2 Using the CLASS Statement 273
  - 8.2.3 Probability and Quantile Plots 275
  - 8.2.4 Using the OUTPUT Statement to Calculate Percentages 276
- 8.3 Doing More with PROC FREQ 277**
  - 8.3.1 OUTPUT Statement in PROC FREQ 277
  - 8.3.2 Using the NLEVELS Option 279

- 8.4 Using PROC REPORT to Better Advantage 280**
  - 8.4.1 PROC REPORT vs. PROC TABULATE 280
  - 8.4.2 Naming Report Items (Variables) in the Compute Block 280
  - 8.4.3 Understanding Compute Block Execution 281
  - 8.4.4 Using a Dummy Column to Consolidate Compute Blocks 283
  - 8.4.5 Consolidating Columns 284
  - 8.4.6 Using the STYLE= Option with LINES 285
  - 8.4.7 Setting Style Attributes with the CALL DEFINE Routine 287
  - 8.4.8 Dates within Dates 288
  - 8.4.9 Aligning Decimal Points 289
  - 8.4.10 Conditionally Executing the LINE Statement 290
- 8.5 Using PROC PRINT 291**
  - 8.5.1 Using the ID and BY Statements Together 291
  - 8.5.2 Using the STYLE= Option with PROC PRINT 292
  - 8.5.3 Using PROC PRINT to Generate a Table of Contents 295

## **Chapter 9 SAS/GRAPH Elements You Should Know—Even if You Don’t Use SAS/GRAPH 297**

- 9.1 Using Title Options with ODS 298**
- 9.2 Setting and Clearing Graphics Options and Settings 300**
- 9.3 Using SAS/GRAPH Statements with Procedures That Are Not SAS/GRAPH Procedures 303**
  - 9.3.1 Changing Plot Symbols with the SYMBOL Statement 303
  - 9.3.2 Controlling Axes and Legends 306
- 9.4 Using ANNOTATE to Augment Graphs 309**

## **Chapter 10 Presentation Graphics—More than Just SAS/GRAPH 313**

- 10.1 Generating Box Plots 314**
  - 10.1.1 Using PROC BOXPLOT 314
  - 10.1.2 Using PROC GPLOT and the SYMBOL Statement 315
  - 10.1.3 Using PROC SHEWHART 316
- 10.2 SAS/GRAPH Specialty Techniques and Procedures 317**
  - 10.2.1 Building Your Own Graphics Font 317
  - 10.2.2 Splitting a Text Line Using JUSTIFY= 319
  - 10.2.3 Using Windows Fonts 319
  - 10.2.4 Using PROC GKPI 320
- 10.3 PROC FREQ Graphics 323**

## **Chapter 11 Output Delivery System 325**

- 11.1 Using the OUTPUT Destination 326**
  - 11.1.1 Determining Object Names 326
  - 11.1.2 Creating a Data Set 327
  - 11.1.3 Using the MATCH\_ALL Option 330
  - 11.1.4 Using the PERSIST= Option 330
  - 11.1.5 Using MATCH\_ALL= with the PERSIST= Option 331
- 11.2 Writing Reports to Excel 332**
  - 11.2.1 EXCELXP Tagset Documentation and Options 333
  - 11.2.2 Generating Multisheet Workbooks 334
  - 11.2.3 Checking Out the Styles 335
- 11.3 Inline Formatting Using Escape Character Sequences 337**
  - 11.3.1 Page X of Y 338
  - 11.3.2 Superscripts, Subscripts, and a Dagger 340
  - 11.3.3 Changing Attributes 341
  - 11.3.4 Using Sequence Codes to Control Indentations, Spacing, and Line Breaks 342
  - 11.3.5 Issuing Raw RTF Specific Commands 344
- 11.4 Creating Hyperlinks 345**
  - 11.4.1 Using Style Overrides to Create Links 345
  - 11.4.2 Using the LINK= TITLE Statement Option 347
  - 11.4.3 Linking Graphics Elements 348
  - 11.4.4 Creating Internal Links 350
- 11.5 Traffic Lighting 352**
  - 11.5.1 User-Defined Format 352
  - 11.5.2 PROC TABULATE 353
  - 11.5.3 PROC REPORT 354
  - 11.5.4 Traffic Lighting with PROC PRINT 355
- 11.6 The ODS LAYOUT Statement 356**
- 11.7 A Few Other Useful ODS Tidbits 358**
  - 11.7.1 Using the ASIS Style Attribute 358
  - 11.7.2 ODS RESULTS Statement 358

## **Part 3 Techniques, Tools, and Interfaces 361**

### **Chapter 12 Taking Advantage of Formats 363**

- 12.1 Using Preloaded Formats to Modify Report Contents 364**
  - 12.1.1 Using Preloaded Formats with PROC REPORT 365
  - 12.1.2 Using Preloaded Formats with PROC TABULATE 367
  - 12.1.3 Using Preloaded Formats with the MEANS and SUMMARY Procedures 369

<b>12.2</b>	<b>Doing More with Picture Formats</b>	<b>370</b>
12.2.1	Date Directives and the DATATYPE Option	371
12.2.2	Working with Fractional Values	373
12.2.3	Using the MULT and PREFIX Options	374
12.2.4	Display Granularity Based on Value Ranges – Limiting Significant Digits	376
<b>12.3</b>	<b>Multilabel (MLF) Formats</b>	<b>377</b>
12.3.1	A Simple MLF	377
12.3.2	Calculating Rolling Averages	378
<b>12.4</b>	<b>Controlling Order Using the NOTSORTED Option</b>	<b>381</b>
<b>12.5</b>	<b>Extending the Use of Format Translations</b>	<b>382</b>
12.5.1	Filtering Missing Values	382
12.5.2	Mapping Overlapping Ranges	383
12.5.3	Handling Text within Numeric Values	383
12.5.4	Using Perl Regular Expressions within Format Definitions	384
12.5.5	Passing Values to a Function as a Format Label	384
<b>12.6</b>	<b>ANYDATE Informats</b>	<b>388</b>
12.6.1	Reading in Mixed Dates	389
12.6.2	Converting Mixed DATETIME Values	389
<b>12.7</b>	<b>Building Formats from Data Sets</b>	<b>390</b>
<b>12.8</b>	<b>Using the PVALUE Format</b>	<b>392</b>
<b>12.9</b>	<b>Format Libraries</b>	<b>393</b>
12.9.1	Saving Formats Permanently	393
12.9.2	Searching for Formats	394
12.9.3	Concatenating Format Catalogs and Libraries	394
<b>Chapter 13</b>	<b>Interfacing with the Macro Language</b>	<b>397</b>
<b>13.1</b>	<b>Avoiding Macro Variable Collisions—Make Your Macro Variables %Local</b>	<b>398</b>
<b>13.2</b>	<b>Using the SYMPUTX Routine</b>	<b>400</b>
13.2.1	Compared to CALL SYMPUT	401
13.2.2	Using SYMPUTX to Save Values of Options	402
13.2.3	Using SYMPUTX to Build a List of Macro Variables	402
<b>13.3</b>	<b>Generalized Programs—Variations on a Theme</b>	<b>403</b>
13.3.1	Steps to the Generalization of a Program	403
13.3.2	Levels of Generalization and Levels of Macro Language Understanding	405
<b>13.4</b>	<b>Utilizing Macro Libraries</b>	<b>406</b>
13.4.1	Establishing an Autocall Library	406
13.4.2	Tracing Autocall Macro Locations	408
13.4.3	Using Stored Compiled Macro Libraries	408
13.4.4	Macro Library Search Order	409

- 13.5 Metadata-Driven Programs 409**
  - 13.5.1 Processing across Data Sets 409
  - 13.5.2 Controlling Data Validations 410
- 13.6 ~~Hard Coding~~—Just Don't Do It 415**
- 13.7 Writing Macro Functions 417**
- 13.8 Macro Information Sources 420**
  - 13.8.1 Using SASHELP and Dictionary tables 420
  - 13.8.2 Retrieving System Options and Settings 422
  - 13.8.3 Accessing the Metadata of a SAS Data Set 424
- 13.9 Macro Security and Protection 426**
  - 13.9.1 Hiding Macro Code 426
  - 13.9.2 Executing a Specific Macro Version 427
- 13.10 Using the Macro Language IN Operator 430**
  - 13.10.1 What Can Go Wrong 430
  - 13.10.2 Using the MINOPERATOR Option 431
  - 13.10.3 Using the MINDELIMITER= Option 432
  - 13.10.4 Compilation vs. Execution for these Options 432
- 13.11 Making Use of the MFILE System Option 433**
- 13.12 A Bit on Macro Quoting 434**

## **Chapter 14 Operating System Interface and Environmental Control 437**

- 14.1 System Options 438**
  - 14.1.1 Initialization Options 438
  - 14.1.2 Data Processing Options 441
  - 14.1.3 Saving SAS System Options 444
- 14.2 Using an AUTOEXEC Program 446**
- 14.3 Using the Configuration File 446**
  - 14.3.1 Changing the SASAUTOS Location 447
  - 14.3.2 Controlling DM Initialization 449
- 14.4 In the Display Manager 449**
  - 14.4.1 Showing Column Names in ViewTable 450
  - 14.4.2 Using the DM Statement 451
  - 14.4.3 Enhanced Editor Options and Shortcuts 452
  - 14.4.4 Macro Abbreviations for the Enhanced Editor 456
  - 14.4.5 Adding Tools to the Application Tool Bar 461
  - 14.4.6 Adding Tools to Pull-Down and Pop-up Menus 463
  - 14.4.7 Adding Tools to the KEYS List 466
- 14.5 Using SAS to Write and Send E-mails 467**

- 14.6 Recovering Physical Location Information 468**
  - 14.6.1 Using the PATHNAME Function 468
  - 14.6.2 SASHELP VIEWS and DICTIONARY Tables 468
  - 14.6.3 Determining the Executing Program Name and Path 469
  - 14.6.4 Retrieving the UNC (Universal Naming Convention) Path 470

## **Chapter 15 Miscellaneous Topics 473**

- 15.1 A Few Miscellaneous Tips 474**
  - 15.1.1 Customizing Your NOTES, WARNINGS, and ERRORS 474
  - 15.1.2 Enhancing Titles and Footnotes with the #BYVAL and #BYVAR Options 475
  - 15.1.3 Executing OS Commands 477
- 15.2 Creating User-defined Functions Using PROC FCMP 479**
  - 15.2.1 Building Your Own Functions 479
  - 15.2.2 Storing and Accessing Your Functions 481
  - 15.2.3 Interaction with the Macro Language 482
  - 15.2.4 Viewing Function Definitions 483
  - 15.2.5 Removing Functions 484
- 15.3 Reading RTF as Data 485**
  - 15.3.1 RTF Diagram Completion 486
  - 15.3.2 Template Preparation 486
  - 15.3.3 RTF as Data 487

## **Appendix A Topical Index 489**

## **Appendix B Usage Index 491**

- Global Statements and Options 492**
  - Statements, Global 492
  - Macro Language 493
  - GOPTIONS, Graphics 493
  - Options, System 493
  - Options, Data Set 495
- Procedures: Steps, Statements, and Options 495**
  - Procedures 495
- DATA Step: Statements and Options 500**
  - Statements, DATA Step 500
  - Format Modifiers 501
  - Functions 501
  - Hash Object 504

**Output Delivery System, ODS 504**

ODS Destinations and Tagsets 504

ODS Attributes 505

ODS Options 505

ODS Statements 506

**SAS Display Manager 506**

Display Manager Commands 506

**References 507**

**User Publications 507**

**Generally Good Reading—Lots More to Learn 518**

SAS Documentation 518

SAS Usage Notes 518

Discussion Forums 518

Newsletters, Corporate and Private Sites 519

User Communities 519

Publications 519

Learning SAS 520

**Index 521**