Learn to program SAS® by example!

Learning SAS® by Example, A Programmer's Guide, Second Edition, teaches SAS programming from very basic concepts to more advanced topics. Because most programmers prefer examples rather than reference-type syntax, this book uses short examples to explain each topic. The second edition has brought this classic book on SAS programming up to the latest SAS version, with new chapters that cover topics such as PROC SGPLOT and Perl regular expressions. This book belongs on the shelf (or e-book reader) of anyone who programs in SAS, from those with little programming experience who want to learn SAS to intermediate and even advanced SAS programmers who want to learn new techniques or identify new ways to accomplish existing tasks.

In an instructive and conversational tone, author Ron Cody clearly explains each programming technique and then illustrates it with one or more real-life examples, followed by a detailed description of how the program works. The text is divided into four major sections: Getting Started, DATA Step Processing, Presenting and Summarizing Your Data, and Advanced Topics. Subjects addressed include

- Reading data from external sources
- Learning details of DATA step programming
- Subsetting and combining SAS data sets
- Understanding SAS functions and working with arrays
- Creating reports with PROC REPORT and PROC TABULATE
- Getting started with the SAS macro language
- Leveraging PROC SQL
- Generating high-quality graphics
- Using advanced features of user-defined formats and informats
- Restructuring SAS data sets
- Working with multiple observations per subject
- Getting started with Perl regular expressions

You can test your knowledge and hone your skills by solving the problems at the end of each chapter.

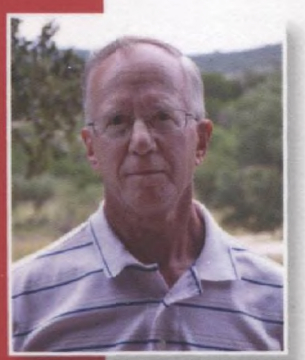

Photo Credit: Jan Cody

RON CODY, EdD, is a retired professor from the Rutgers Robert Wood Johnson Medical School who now works as a national instructor for SAS and as an author of books on SAS and statistics. A SAS user since 1977, Ron's extensive knowledge and innovative style have made him a popular presenter at local, regional, and national SAS conferences. He has authored or co-authored numerous books, as well as countless articles in medical and scientific journals.

FREE DATA on the Web!
support.sas.com/authors

"Ron Cody's newest book can serve not only as a textbook for a SAS programming course, but also as a reference for intermediate-level programmers. The book begins with basic data input methods and advances through merging data sets and more complex tasks.
I would recommend *Learning SAS by Example: A Programmer's Guide* to anyone wishing to learn or refresh his/her SAS programming skills."

KAROL H. KATZ, M.S.
Programmer Analyst
Yale University School of Medicine

SAS® Press
ISBN 978-1-63526-659-7

sas

sas.com/books

Contents

List of Programs xi
Preface xix
About This Book xxi
About the Author xxiii
Acknowledgments xxv

Part 1: Getting Started 1

Chapter 1: What Is SAS? 3
1.1 Introduction 3
1.2 Getting Data into SAS 4
1.3 A Sample SAS Program 4
1.4 SAS Names 6
1.5 SAS Data Sets and SAS Data Types 7
1.6 The SAS Windowing Environment, SAS Enterprise Guide, and the SAS University Edition 7
1.7 Problems 8

Chapter 2: Writing Your First SAS Program 9
2.1 A Simple Program to Read Raw Data and Produce a Report 9
2.2 Enhancing the Program 16
2.3 More on Comment Statements 18
2.4 How SAS Works (a Look inside the "Black Box") 18
2.5 Problems 21

Part 2: DATA Step Processing 23

Chapter 3: Reading Raw Data from External Files 25
3.1 Introduction 25
3.2 Reading Data Values Separated by Blanks 26
3.3 Specifying Missing Values with List Input 27
3.4 Reading Data Values Separated by Commas (CSV Files) 27
3.5 Using an alternative Method to Specify an External File 28
3.6 Reading Data Values Separated by Delimiters Other Than Blanks or Commas 29
3.7 Placing Data Lines Directly in Your Program (the DATALINES Statement) 30
3.8 Specifying INFILE Options with the DATALINES Statement 30
3.9 Reading Raw Data from Fixed Columns—Method 1: Column Input 31
3.10 Reading Raw Data from Fixed Columns—Method 2: Formatted Input 32
3.11 Using a FORMAT Statement in a DATA Step versus in a Procedure 36
3.12 Using Informats with List Input 36
3.13 Supplying an INFORMAT Statement with List Input 37
3.14 Using List Input with Embedded Delimiters 37
3.15 Problems 38

Chapter 4: Creating Permanent SAS Data Sets 43
4.1 Introduction 43
4.2 SAS Libraries—The LIBNAME Statement 43
4.3 Why Create Permanent SAS Data Sets? 45
4.4 Examining the Descriptor Portion of a SAS Data Set Using PROC CONTENTS 45
4.5 Listing All the SAS Data Sets in a SAS Library Using PROC CONTENTS 48
4.6 Viewing the Descriptor Portion of a SAS Data Set Using a Point-and-Click Approach 48
4.7 Viewing the Data Portion of a SAS Data Set Using PROC PRINT 51
4.8 Using a SAS Data Set as Input to a DATA Step 52
4.9 DATA _NULL_: A Data Set That Isn't 53
4.10 Problems 55

Chapter 5: Creating Labels and Formats 57
5.1 Adding Labels to Your Variables 57
5.2 Using Formats to Enhance Your Output 58
5.3 Regrouping Values Using Formats 61
5.4 More on Format Ranges 63
5.5 Storing Your Formats in a Format Library 64
5.6 Permanent Data Set Attributes 64
5.7 Accessing a Permanent SAS Data Set with User-Defined Formats 66
5.8 Displaying Your Format Definitions 67
5.9 Problems 68

Chapter 6: Reading and Writing Data from an Excel Spreadsheet 71
6.1 Introduction 71
6.2 Using the Import Wizard to Convert a Spreadsheet to a SAS Data Set 71
6.3 Creating an Excel Spreadsheet from a SAS Data Set 76
6.4 Using an Engine to Read an Excel Spreadsheet 78
6.5 Using the SAS Output Delivery System to Convert a SAS Data Set to an Excel Spreadsheet 79
6.6 A Quick Look at the Import Utility in SAS Studio 80
6.7 Problems 84

Chapter 7: Performing Conditional Processing 87
7.1 Introduction 87
7.2 The IF and ELSE IF Statements 87
7.3 The Subsetting IF Statement 90
7.4 The IN Operator 91
7.5 Using a SELECT Statement for Logical Tests 92
7.6 Using Boolean Logic (AND, OR, and NOT Operators) 93
7.7 A Caution When Using Multiple OR Operators 94
7.8 The WHERE Statement 95
7.9 Some Useful WHERE Operators 96
7.10 Problems 96

Chapter 8: Performing Iterative Processing: Looping 99
8.1 Introduction 99
8.2 DO Groups 99
8.3 The Sum Statement 101

8.4 The Iterative DO Loop 106
8.5 Other Forms of an Iterative DO Loop 110
8.6 DO WHILE and DO UNTIL Statements 112
8.7 A Caution When Using DO UNTIL Statements 115
8.8 LEAVE and CONTINUE Statements 116
8.9 Problems 117

Chapter 9: Working with Dates 121

9.1 Introduction 121
9.2 How SAS Stores Dates 121
9.3 Reading Date Values from Text Data 122
9.4 Computing the Number of Years between Two Dates 124
9.5 Demonstrating a Date Constant 125
9.6 Computing the Current Date 126
9.7 Extracting the Day of the Week, Day of the Month, Month, and Year from a SAS Date 126
9.8 Creating a SAS Date from Month, Day, and Year Values 127
9.9 Substituting the 15th of the Month when the Day Value Is Missing 128
9.10 Using Date Interval Functions 129
9.11 Problems 134

Chapter 10: Subsetting and Combining SAS Data Sets 137

10.1 Introduction 137
10.2 Subsetting a SAS Data Set 137
10.3 Creating More Than One Subset Data Set in One DATA Step 139
10.4 Adding Observations to a SAS Data Set 139
10.5 Interleaving Data Sets 142
10.6 Combining Detail and Summary Data 143
10.7 Merging Two Data Sets 144
10.8 Omitting the BY Statement in a Merge 146
10.9 Controlling Observations in a Merged Data Set 147
10.10 More Uses for IN= Variables 149
10.11 When Does a DATA Step End? 150
10.12 Merging Two Data Sets with Different BY Variable Names 151
10.13 Merging Two Data Sets with Different BY Variable Data Types 152
10.14 One-to-One, One-to-Many, and Many-to-Many Merges 154
10.15 Updating a Master File from a Transaction File 156
10.16 Problems 158

Chapter 11: Working with Numeric Functions 161

11.1 Introduction 161
11.2 Functions That Round and Truncate Numeric Values 161
11.3 Functions That Work with Missing Values 163
11.4 Setting Character and Numeric Values to Missing 163
11.5 Descriptive Statistics Functions 164
11.6 Computing Sums within an Observation 166
11.7 Mathematical Functions 167
11.8 Computing Some Useful Constants 168
11.9 Generating Random Numbers 168
11.10 Special Functions 172
11.11 Functions That Return Values from Previous Observations 174

11.12 Sorting Within an Observations—a Game Changer 177
11.13 Problems 179

Chapter 12: Working with Character Functions 181
12.1 Introduction 181
12.2 Determining the Length of a Character Value 182
12.3 Changing the Case of Characters 182
12.4 Removing Characters from Strings 183
12.5 Joining Two or More Strings Together 184
12.6 Removing Leading or Trailing Blanks 185
12.7 Using the COMPRESS Function to Remove Characters from a String 186
12.8 Searching for Characters 189
12.9 Searching for Individual Characters 190
12.10 Searching for Words in a String 191
12.11 Searching for Character Classes 192
12.12 Using the NOT Functions for Data Cleaning 193
12.13 Extracting Part of a String 194
12.14 Dividing Strings into Words 196
12.15 Performing a Fuzzy Match 197
12.16 Substituting Strings or Words 199
12.17 Problems 201

Chapter 13: Working with Arrays 205
13.1 Introduction 205
13.2 Setting Values of 999 to a SAS Missing Value for Several Numeric Variables 206
13.3 Setting Values of NA and ? to a Missing Character Value 207
13.4 Converting All Character Values to Propercase 208
13.5 Using an Array to Create New Variables 209
13.6 Changing the Array Bounds 210
13.7 Temporary Arrays 210
13.8 Loading the Initial Values of a Temporary Array from a Raw Data File 211
13.9 Using a Multidimensional Array for Table Lookup 212
13.10 Problems 215

Part 3: Presenting and Summarizing Your Data 217

Chapter 14: Displaying Your Data 219
14.1 Introduction 219
14.2 The Basics 219
14.3 Changing the Appearance of Your Listing 220
14.4 Changing the Appearance of Values 222
14.5 Controlling the Observations That Appear in Your Listing 223
14.6 Adding Titles and Footnotes to Your Listing 224
14.7 Changing the Order of Your Listing 226
14.8 Sorting by More Than One Variable 228
14.9 Labeling Your Column Headings 229
14.10 Adding Subtotals and Totals to Your Listing 230
14.11 Making Your Listing Easier to Read 233
14.12 Adding the Number of Observations to Your Listing 234

14.13 Listing the First n Observations of Your Data Set 234
14.14 Problems 235

Chapter 15: Creating Customized Reports 237
15.1 Introduction 237
15.2 Using PROC REPORT 238
15.3 Selecting the Variables to Include in Your Report 239
15.4 Comparing Detail and Summary Reports 240
15.5 Producing a Summary Report 241
15.6 Demonstrating the FLOW Option of PROC REPORT 242
15.7 Using Two Grouping Variables 243
15.8 Changing the Order of Variables in the COLUMN Statement 244
15.9 Changing the Order of Rows in a Report 245
15.10 Applying the ORDER Usage to Two Variables 246
15.11 Creating a Multi-Column Report 247
15.12 Producing Report Breaks 248
15.13 Using a Nonprinting Variable to Order a Report 251
15.14 Computing a New Variable with PROC REPORT 252
15.15 Computing a Character Variable in a COMPUTE Block 253
15.16 Creating an ACROSS Variable with PROC REPORT 254
15.17 Using an ACROSS Usage to Display Statistics 255
15.18 Problems 256

Chapter 16: Summarizing Your Data 261
16.1 Introduction 261
16.2 PROC MEANS—Starting from the Beginning 261
16.3 Adding a BY Statement to PROC MEANS 263
16.4 Using a CLASS Statement with PROC MEANS 264
16.5 Applying a Format to a CLASS Variable 265
16.6 Deciding between a BY Statement and a CLASS Statement 266
16.7 Creating Summary Data Sets Using PROC MEANS 266
16.8 Outputting Other Descriptive Statistics with PROC MEANS 267
16.9 Asking SAS to Name the Variables in the Output Data Set 268
16.10 Outputting a Summary Data Set: Including a BY Statement 269
16.11 Outputting a Summary Data Set: Using a CLASS Statement 270
16.12 Using Two CLASS Variables with PROC MEANS 271
16.13 Selecting Different Statistics for Each Variable 276
16.14 Printing all Possible Combinations of Your Class Variables 276
16.15 Problems 278

Chapter 17: Counting Frequencies 281
17.1 Introduction 281
17.2 Counting Frequencies 281
17.3 Selecting Variables for PROC FREQ 284
17.4 Using Formats to Label the Output 286
17.5 Using Formats to Group Values 288
17.6 Problems Grouping Values with PROC FREQ 289
17.7 Displaying Missing Values in the Frequency Table 290
17.8 Changing the Order of Values in PROC FREQ 292
17.9 Producing Two-Way Tables 294

17.10 Requesting Multiple Two-Way Tables 295
17.11 Producing Three-Way Tables 296
17.12 Problems 297

Chapter 18: Creating Tabular Reports 299
18.1 Introduction 299
18.2 A Simple PROC TABULATE Table 300
18.3 Describing the Three PROC TABULATE Operators 301
18.4 Using the Keyword ALL 303
18.5 Producing Descriptive Statistics 304
18.6 Combining CLASS and Analysis Variables in a Table 306
18.7 Customizing Your Table 307
18.8 Demonstrating a More Complex Table 309
18.9 Computing Row and Column Percentages 310
18.10 Displaying Percentages in a Two-Dimensional Table 312
18.11 Computing Column Percentages 313
18.12 Computing Percentages on Numeric Variables 314
18.13 Understanding How Missing Values Affect PROC TABULATE Output 315
18.14 Problems 318

Chapter 19: Introducing the Output Delivery System 325
19.1 Introduction 325
19.2 Sending SAS Output to an HTML File 325
19.3 Creating a Table of Contents 327
19.4 Selecting a Different HTML Style 328
19.5 Choosing Other ODS Destinations 329
19.6 Selecting or Excluding Portions of SAS Output 329
19.7 Sending Output to a SAS Data Set 333
19.8 Problems 335

Chapter 20: Creating Charts and Graphs 337
20.1 Introduction 337
20.2 Creating Bar Charts 337
20.3 Displaying Statistics for a Response Variable 340
20.4 Creating Scatter Plots 341
20.5 Adding a Regression Line and Confidence Limits to the Plot 342
20.6 Generating Time Series Plots 343
20.7 Describing Two Methods of Generating Smooth Curves 345
20.8 Generating Histograms 346
20.9 Generating a Simple Box Plot 347
20.10 Producing a Box Plot with a Grouping Variable 348
20.11 Demonstrating Overlays and Transparency 349
20.12 Problems 351

Part 4: Advanced Topics 357

Chapter 21: Using Advanced INPUT Techniques 359
21.1 Introduction 359
21.2 Handling Missing Values at the End of a Line 360
21.3 Reading Short Data Lines 362

21.4 Reading External Files with Lines Longer Than 32,767 Characters 363
21.5 Detecting the End of the File 364
21.6 Reading a Portion of a Raw Data File 365
21.7 Reading Data from Multiple Files 366
21.8 Reading Data from Multiple Files Using a FILENAME Statement 366
21.9 Reading External Filenames from a Data File 366
21.10 Reading Multiple Lines of Data to Create One Observation 367
21.11 Reading Data Conditionally (the Single Trailing @ Sign) 369
21.12 More Examples of the Single Trailing @ Sign 372
21.13 Creating Multiple Observations from One Line of Input 372
21.14 Using Variable and Informat Lists 373
21.15 Using Relative Column Pointers to Read a Complex Data Structure Efficiently 374
21.16 Problems 375

Chapter 22: Using Advanced Features of User-Defined Formats and Informats 377

22.1 Introduction 377
22.2 Using Formats to Recode Variables 377
22.3 Using Formats with a PUT Function to Create New Variables 379
22.4 Creating User-Defined Informats 380
22.5 Reading Character and Numeric Data in One Step 382
22.6 Using Formats (and Informats) to Perform Table Lookup 384
22.7 Using a SAS Data Set to Create a Format 386
22.8 Updating and Maintaining Your Formats 390
22.9 Using Formats within Formats 392
22.10 Multilabel Formats 394
22.11 Using the INPUTN Function to Perform a More Complicated Table Lookup 398
22.12 Problems 401

Chapter 23: Restructuring SAS Data Sets 403

23.1 Introduction 403
23.2 Converting a Data Set with One Observation per Subject to a Data Set with Several Observations per Subject: Using a DATA Step 403
23.3 Converting a Data Set with Several Observations per Subject to a Data Set with One Observation per Subject: Using a DATA Step 406
23.4 Converting a Data Set with One Observation per Subject to a Data Set with Several Observations per Subject: Using PROC TRANSPOSE 407
23.5 Converting a Data Set with Several Observations per Subject to a Data Set with One Observation per Subject: Using PROC TRANSPOSE 408
23.6 Problems 409

Chapter 24: Working with Multiple Observations per Subject 413

24.1 Introduction 413
24.2 Identifying the First or Last Observation in a Group 413
24.3 Counting the Number of Visits Using PROC FREQ 416
24.4 Computing Differences between Observations 418
24.5 Computing Differences between the First and Last Observation in a BY Group Using the LAG Function 420
24.6 Computing Differences between the First and Last Observation in a BY Group Using a RETAIN Statement 421

24.7 Using a Retained Variable to "Remember" a Previous Value 422
24.8 Problems 423

Chapter 25: Introducing the SAS Macro Language 425
25.1 Introduction 425
25.2 Macro Variables: What Are They? 425
25.3 Some Built-In Macro Variables 426
25.4 Assigning Values to Macro Variables with a %LET Statement 427
25.5 Demonstrating a Simple Macro 427
25.6 Describing Positional and Keyword Macro Parameters 429
25.7 A Word about Tokens 431
25.8 Another Example of Using a Macro Variable as a Prefix 432
25.9 Using a Macro Variable to Transfer a Value between DATA Steps 433
25.10 Problems 435

Chapter 26: Introducing the Structured Query Language 437
26.1 Introduction 437
26.2 Some Basics 437
26.3 Joining Two Tables (Merge) 439
26.4 Left, Right, and Full Joins 444
26.5 Concatenating Data Sets 446
26.6 Using Summary Functions 449
26.7 Demonstrating the ORDER Clause 450
26.8 An Example of Fuzzy Matching 451
26.9 Problems 452

Chapter 27: Introducing Perl Regular Expressions 455
27.1 Introduction 455
27.2 Describing the Syntax of Regular Expressions 455
27.3 Testing That Social Security Numbers Are in Standard Form 456
27.4 Checking for Valid ZIP Codes 457
27.5 Verifying That Phone Numbers Are in a Standard Form 458
27.6 Describing the PRXPARSE Function 460
27.7 Problems 461

Solutions to Odd-Numbered Exercises 463

Index 497