



From *Learning SAS® by Example*. Full book available for purchase [here](#).

Contents

List of Programs xv

Preface xxix

Acknowledgments xxxi

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 7
- 1.5 SAS Data Sets and SAS Data Types 8
- 1.6 The SAS Display Manager and SAS Enterprise Guide 9
- 1.7 Problems 9

Chapter 2 Writing Your First SAS Program 11

- 2.1 A Simple Program to Read Raw Data and Produce a Report 11
- 2.2 Enhancing the Program 18
- 2.3 More on Comment Statements 20
- 2.4 How SAS Works (a Look Inside the “Black Box”) 22
- 2.5 Problems 25

Part 2 DATA Step Processing 27

Chapter 3 Reading Raw Data from External Files 29

- 3.1 Introduction 30
- 3.2 Reading Data Values Separated by Blanks 30
- 3.3 Specifying Missing Values with List Input 32
- 3.4 Reading Data Values Separated by Commas (CSV Files) 33
- 3.5 Using an Alternative Method to Specify an External File 34

- 3.6 Reading Data Values Separated by Delimiters Other Than Blanks or Commas 34
- 3.7 Placing Data Lines Directly in Your Program (the DATALINES Statement) 36
- 3.8 Specifying INFILE Options with the DATALINES Statement 37
- 3.9 Reading Raw Data from Fixed Columns—Method 1: Column Input 37
- 3.10 Reading Raw Data from Fixed Columns—Method 2: Formatted Input 39
- 3.11 Using a FORMAT Statement in a DATA Step versus in a Procedure 43
- 3.12 Using Informats with List Input 43
- 3.13 Supplying an INFORMAT Statement with List Input 45
- 3.14 Using List Input with Embedded Delimiters 46
- 3.15 Problems 47

Chapter 4 Creating Permanent SAS Data Sets 53

- 4.1 Introduction 54
- 4.2 SAS Libraries—The LIBNAME Statement 54
- 4.3 Why Create Permanent SAS Data Sets? 55
- 4.4 Examining the Descriptor Portion of a SAS Data Set Using PROC CONTENTS 56
- 4.5 Listing All the SAS Data Sets in a SAS Library Using PROC CONTENTS 59
- 4.6 Viewing the Descriptor Portion of a SAS Data Set Using the SAS Explorer 60
- 4.7 Viewing the Data Portion of a SAS Data Set Using PROC PRINT 63
- 4.8 Viewing the Data Portion of a SAS Data Set Using the SAS VIEWTABLE Window 64
- 4.9 Using a SAS Data Set as Input to a DATA Step 65
- 4.10 DATA _NULL_: A Data Set That Isn't 67
- 4.11 Problems 68

Chapter 5 Creating Formats and Labels 71

- 5.1 Adding Labels to Your Variables 71
- 5.2 Using Formats to Enhance Your Output 73
- 5.3 Regrouping Values Using Formats 76
- 5.4 More on Format Ranges 78
- 5.5 Storing Your Formats in a Format Library 79
- 5.6 Permanent Data Set Attributes 80
- 5.7 Accessing a Permanent SAS Data Set with User-Defined Formats 82
- 5.8 Displaying Your Format Definitions 83
- 5.9 Problems 84

Chapter 6 Reading and Writing Data from an Excel Spreadsheet 87

- 6.1 Introduction 87
- 6.2 Using the Import Wizard to Convert a Spreadsheet to a SAS Data Set 88
- 6.3 Creating an Excel Spreadsheet from a SAS Data Set 93
- 6.4 Using an Engine to Read an Excel Spreadsheet 95
- 6.5 Using the SAS Output Delivery System to Convert a SAS Data Set to an Excel Spreadsheet 96
- 6.6 Problems 98

Chapter 7 Performing Conditional Processing 101

- 7.1 Introduction 102
- 7.2 The IF and ELSE IF Statements 102
- 7.3 The Subsetting IF Statement 105
- 7.4 The IN Operator 107
- 7.5 Using a SELECT Statement for Logical Tests 108
- 7.6 Using Boolean Logic (AND, OR, and NOT Operators) 109
- 7.7 A Caution When Using Multiple OR Operators 111
- 7.8 The WHERE Statement 112
- 7.9 Some Useful WHERE Operators 113
- 7.10 Problems 114

Chapter 8 Performing Iterative Processing: Looping 117

- 8.1 Introduction 117
- 8.2 DO Groups 118
- 8.3 The Sum Statement 120
- 8.4 The Iterative DO Loop 125
- 8.5 Other Forms of an Iterative DO Loop 129
- 8.6 DO WHILE and DO UNTIL Statements 131
- 8.7 A Caution When Using DO UNTIL Statements 134
- 8.8 LEAVE and CONTINUE Statements 135
- 8.9 Problems 137

Chapter 9 Working with Dates 141

- 9.1 Introduction 142
- 9.2 How SAS Stores Dates 142
- 9.3 Reading Date Values from Raw Data 143
- 9.4 Computing the Number of Years between Two Dates 146
- 9.5 Demonstrating a Date Constant 147
- 9.6 Computing the Current Date 148
- 9.7 Extracting the Day of the Week, Day of the Month, Month, and Year from a SAS Date 149
- 9.8 Creating a SAS Date from Month, Day, and Year Values 150
- 9.9 Substituting the 15th of the Month when the Day Value Is Missing 151
- 9.10 Using Date Interval Functions 152
- 9.11 Problems 157

Chapter 10 Subsetting and Combining SAS Data Sets 161

- 10.1 Introduction 162
- 10.2 Subsetting a SAS Data Set 162
- 10.3 Creating More Than One Subset Data Set in One DATA Step 163
- 10.4 Adding Observations to a SAS Data Set 164
- 10.5 Interleaving Data Sets 167
- 10.6 Combining Detail and Summary Data 168

- 10.7 Merging Two Data Sets 170
- 10.8 Omitting the BY Statement in a Merge 172
- 10.9 Controlling Observations in a Merged Data Set 173
- 10.10 More Uses for IN= Variables 175
- 10.11 When Does a DATA Step End? 176
- 10.12 Merging Two Data Sets with Different BY Variable Names 177
- 10.13 Merging Two Data Sets with Different BY Variable Data Types 179
- 10.14 One-to-One, One-to-Many, and Many-to-Many Merges 181
- 10.15 Updating a Master File from a Transaction File 183
- 10.16 Problems 185

Chapter 11 Working with Numeric Functions 189

- 11.1 Introduction 190
- 11.2 Functions That Round and Truncate Numeric Values 190
- 11.3 Functions That Work with Missing Values 192
- 11.4 Setting Character and Numeric Values to Missing 193
- 11.5 Descriptive Statistics Functions 194
- 11.6 Computing Sums within an Observation 196
- 11.7 Mathematical Functions 197
- 11.8 Computing Some Useful Constants 198
- 11.9 Generating Random Numbers 199
- 11.10 Special Functions 201
- 11.11 Functions That Return Values from Previous Observations 204
- 11.12 Problems 207

Chapter 12 Working with Character Functions 211

- 12.1 Introduction 212
- 12.2 Determining the Length of a Character Value 212
- 12.3 Changing the Case of Characters 213
- 12.4 Removing Characters from Strings 214
- 12.5 Joining Two or More Strings Together 215
- 12.6 Removing Leading or Trailing Blanks 217

- 12.7 Using the COMPRESS Function to Remove Characters from a String 218
- 12.8 Searching for Characters 220
- 12.9 Searching for Individual Characters 223
- 12.10 Searching for Words in a String 223
- 12.11 Searching for Character Classes 225
- 12.12 Using the NOT Functions for Data Cleaning 226
- 12.13 Describing a Real Blockbuster Data Cleaning Function 227
- 12.14 Extracting Part of a String 228
- 12.15 Dividing Strings into Words 230
- 12.16 Comparing Strings 232
- 12.17 Performing a Fuzzy Match 234
- 12.18 Substituting Characters or Words 235
- 12.19 Problems 238

Chapter 13 Working with Arrays 243

- 13.1 Introduction 244
- 13.2 Setting Values of 999 to a SAS Missing Value for Several Numeric Variables 244
- 13.3 Setting Values of NA and ? to a Missing Character Value 247
- 13.4 Converting All Character Values to Lowercase 248
- 13.5 Using an Array to Create New Variables 249
- 13.6 Changing the Array Bounds 250
- 13.7 Temporary Arrays 251
- 13.8 Loading the Initial Values of a Temporary Array from a Raw Data File 253
- 13.9 Using a Multidimensional Array for Table Lookup 254
- 13.10 Problems 257

Part 3 Presenting and Summarizing Your Data 259

Chapter 14 Displaying Your Data 261

- 14.1 Introduction 262
- 14.2 The Basics 262
- 14.3 Changing the Appearance of Your Listing 263
- 14.4 Changing the Appearance of Values 265
- 14.5 Controlling the Observations That Appear in Your Listing 266
- 14.6 Adding Additional Titles and Footnotes to Your Listing 268
- 14.7 Changing the Order of Your Listing 270
- 14.8 Sorting by More Than One Variable 272
- 14.9 Labeling Your Column Headings 273
- 14.10 Adding Subtotals and Totals to Your Listing 274
- 14.11 Making Your Listing Easier to Read 277
- 14.12 Adding the Number of Observations to Your Listing 279
- 14.13 Double-Spacing Your Listing 280
- 14.14 Listing the First n Observations of Your Data Set 281
- 14.15 Problems 283

Chapter 15 Creating Customized Reports 287

- 15.1 Introduction 288
- 15.2 Using PROC REPORT 289
- 15.3 Selecting Variables to Include in Your Report 291
- 15.4 Comparing Detail and Summary Reports 291
- 15.5 Producing a Summary Report 293
- 15.6 Demonstrating the FLOW Option of PROC REPORT 294
- 15.7 Using Two Grouping Variables 296
- 15.8 Changing the Order of Variables in the COLUMN Statement 297
- 15.9 Changing the Order of Rows in a Report 299
- 15.10 Applying the ORDER Usage to Two Variables 300
- 15.11 Creating a Multi-Column Report 301
- 15.12 Producing Report Breaks 303
- 15.13 Using a Nonprinting Variable to Order a Report 306
- 15.14 Computing a New Variable with PROC REPORT 307
- 15.15 Computing a Character Variable in a COMPUTE Block 308

- 15.16 Creating an ACROSS Variable with PROC REPORT 310
- 15.17 Modifying the Column Label for an ACROSS Variable 311
- 15.18 Using an ACROSS Usage to Display Statistics 311
- 15.19 Problems 313

Chapter 16 Summarizing Your Data 319

- 16.1 Introduction 320
- 16.2 PROC MEANS—Starting from the Beginning 320
- 16.3 Adding a BY Statement to PROC MEANS 323
- 16.4 Using a CLASS Statement with PROC MEANS 324
- 16.5 Applying a Format to a CLASS Variable 325
- 16.6 Deciding between a BY Statement and a CLASS Statement 327
- 16.7 Creating Summary Data Sets Using PROC MEANS 327
- 16.8 Outputting Other Descriptive Statistics with PROC MEANS 328
- 16.9 Asking SAS to Name the Variables in the Output Data Set 329
- 16.10 Outputting a Summary Data Set: Including a BY Statement 330
- 16.11 Outputting a Summary Data Set: Including a CLASS Statement 331
- 16.12 Using Two CLASS Variables with PROC MEANS 333
- 16.13 Selecting Different Statistics for Each Variable 337
- 16.14 Problems 338

Chapter 17 Counting Frequencies 341

- 17.1 Introduction 342
- 17.2 Counting Frequencies 342
- 17.3 Selecting Variables for PROC FREQ 345
- 17.4 Using Formats to Label the Output 346
- 17.5 Using Formats to Group Values 347
- 17.6 Problems Grouping Values with PROC FREQ 349
- 17.7 Displaying Missing Values in the Frequency Table 351
- 17.8 Changing the Order of Values in PROC FREQ 353
- 17.9 Producing Two-Way Tables 356

- 17.10 Requesting Multiple Two-Way Tables 358
- 17.11 Producing Three-Way Tables 358
- 17.12 Problems 360

Chapter 18 Creating Tabular Reports 363

- 18.1 Introduction 364
- 18.2 A Simple PROC TABULATE Table 364
- 18.3 Describing the Three PROC TABULATE Operators 366
- 18.4 Using the Keyword ALL 369
- 18.5 Producing Descriptive Statistics 370
- 18.6 Combining CLASS and Analysis Variables in a Table 372
- 18.7 Customizing Your Table 374
- 18.8 Demonstrating a More Complex Table 377
- 18.9 Computing Row and Column Percentages 379
- 18.10 Displaying Percentages in a Two-Dimensional Table 381
- 18.11 Computing Column Percentages 382
- 18.12 Computing Percentages on Numeric Variables 384
- 18.13 Understanding How Missing Values Affect PROC TABULATE Output 385
- 18.14 Problems 390

Chapter 19 Introducing the Output Delivery System 397

- 19.1 Introduction 397
- 19.2 Sending SAS Output to an HTML File 398
- 19.3 Creating a Table of Contents 400
- 19.4 Selecting a Different HTML Style 401
- 19.5 Choosing Other ODS Destinations 402
- 19.6 Selecting or Excluding Portions of SAS Output 403
- 19.7 Sending Output to a SAS Data Set 407
- 19.8 Problems 409

Chapter 20 Generating High-Quality Graphics 411

- 20.1 Introduction 412
- 20.2 Some Basic Concepts 412
- 20.3 Producing Simple Bar Charts Using PROC GCHART 413
- 20.4 Creating Pie Charts 415
- 20.5 Creating Bar Charts for a Continuous Variable 416

- 20.6 Creating Charts with Values Representing Categories 418
- 20.7 Creating Bar Charts Representing Sums 420
- 20.8 Creating Bar Charts Representing Means 422
- 20.9 Adding Another Variable to the Chart 423
- 20.10 Producing Scatter Plots 425
- 20.11 Connecting Points 427
- 20.12 Connecting Points with a Smooth Line 430
- 20.13 Problems 431

Part 4 Advanced Topics 435

Chapter 21 Using Advanced INPUT Techniques 437

- 21.1 Introduction 438
- 21.2 Handling Missing Values at the End of a Line 438
- 21.3 Reading Short Data Lines 440
- 21.4 Reading External Files with Lines Longer Than 256 Characters 443
- 21.5 Detecting the End of the File 443
- 21.6 Reading a Portion of a Raw Data File 445
- 21.7 Reading Data from Multiple Files 446
- 21.8 Reading Data from Multiple Files Using a FILENAME Statement 447
- 21.9 Reading External Filenames from a Data File 447
- 21.10 Reading Multiple Lines of Data to Form One Observation 448
- 21.11 Reading Data Conditionally (the Single Trailing @ Sign) 451
- 21.12 More Examples of the Single Trailing @ Sign 453
- 21.13 Creating Multiple Observations from One Line of Input 454
- 21.14 Using Variable and Informat Lists 455
- 21.15 Using Relative Column Pointers to Read a Complex Data Structure Efficiently 456
- 21.16 Problems 458

Chapter 22 Using Advanced Features of User-Defined Formats and Informats 462

- 22.1 Introduction 462
- 22.2 Using Formats to Recode Variables 462
- 22.3 Using Formats with a PUT Function to Create New Variables 463
- 22.4 Creating User-Defined Informats 464
- 22.5 Reading Character and Numeric Data in One Step 467
- 22.6 Using Formats (and Informats) to Perform Table Lookup 470
- 22.7 Using a SAS Data Set to Create a Format 471
- 22.8 Updating and Maintaining Your Formats 477
- 22.9 Using Formats within Formats 479
- 22.10 Using Multilabel Formats 482
- 22.11 Using the INPUTN Function to Perform a More Complicated Table Lookup 485
- 22.12 Problems 490

Chapter 23 Restructuring SAS Data Sets 493

- 23.1 Introduction 494
- 23.2 Converting a Data Set with One Observation per Subject to a Data Set with Several Observations per Subject: Using a DATA Step 494
- 23.3 Converting a Data Set with Several Observations per Subject to a Data Set with One Observation per Subject: Using a DATA Step 496
- 23.4 Converting a Data Set with One Observation per Subject to a Data Set with Several Observations per Subject: Using PROC TRANSPOSE 498
- 23.5 Converting a Data Set with Several Observations per Subject to a Data Set with One Observation per Subject: Using PROC TRANSPOSE 500
- 23.6 Problems 501

Chapter 24 Working with Multiple Observations per Subject 505

- 24.1 Introduction 506
- 24.2 Identifying the First or Last Observation in a Group 506
- 24.3 Counting the Number of Visits Using PROC FREQ 509

24.4	Counting the Number of Visits Using PROC MEANS	511
24.5	Computing Differences between Observations	512
24.6	Computing Differences between the First and Last Observation in a BY Group Using the LAG Function	514
24.7	Computing Differences between the First and Last Observation in a BY Group Using a RETAIN Statement	515
24.8	Using a Retained Variable to “Remember” a Previous Value	517
24.9	Problems	518
Chapter 25	Introducing the SAS Macro Language	521
25.1	Introduction	522
25.2	Macro Variables: What Are They?	522
25.3	Some Built-In Macro Variables	523
25.4	Assigning Values to Macro Variables with a %LET Statement	524
25.5	Demonstrating a Simple Macro	525
25.6	A Word about Tokens	527
25.7	Another Example of Using a Macro Variable as a Prefix	529
25.8	Using a Macro Variable to Transfer a Value between DATA Steps	530
25.9	Problems	532
Chapter 26	Introducing the Structured Query Language	535
26.1	Introduction	536
26.2	Some Basics	536
26.3	Joining Two Tables (Merge)	539
26.4	Left, Right, and Full Joins	543
26.5	Concatenating Data Sets	546
26.6	Using Summary Functions	549
26.7	Demonstrating an ORDER Clause	551
26.8	An Example of Fuzzy Matching	551
26.9	Problems	553
	Solutions to Odd-Numbered Problems	557
	Index	601