# Contents

Acknowledgments vii  
About This Book ix  

## Chapter 1  
**Introducing Data Relationships, Techniques for Data Manipulation, and Access Methods**  
1  
Overview 1  
Determining Data Relationships 1  
Understanding the Methods for Combining SAS Data Sets 3  
Understanding Access Methods: Sequential versus Direct 7  
Understanding the Tools for Combining SAS Data Sets 8  
Understanding the Tools for Processing Information in Groups 10  
Choosing between the DATA Step and PROC SQL 10  
Choosing between MODIFY and UPDATE 11  

## Chapter 2  
**Combining Data Sets Vertically: Concatenating, Interleaving, and Appending Data Sets**  
13  
Example 2.1  Concatenating Data Sets 14  
Example 2.2  Interleaving Observations from Two or More Data Sets Based on a Common Variable 17  
Example 2.3  Appending One Data Set to the End of Another Data Set 20  
Example 2.4  Selecting Unique Rows When Concatenating Tables 23  
Example 2.5  Selecting Rows in Common When Concatenating Tables 26  
Example 2.6  Selecting Observations Unique to Each Data Set When Concatenating Data Sets 28  

## Chapter 3  
**Combining Data Sets Horizontally: Match-Merging Data Sets by Value**  
31  
Example 3.1  Merging Data Sets by a Common Variable 33  
Example 3.2  Merging Observations from Multiple Data Sets by a Common Variable 36  
Example 3.3  Combining Observations When Variable Values Do Not Match Exactly 40  
Example 3.4  Combining Observations by the Formatted Value of a Variable 42  
Example 3.5  Combining Multiple Tables When the Matching Column Has Different Attributes 45  
Example 3.6  Combining Rows When There Is No Common Column 49  
Example 3.7  Matching Observations Randomly 52  
Example 3.8  Combining Multiple Data Sets without a Variable Common to All the Data Sets 55  
Example 3.9  Generating Every Combination of Rows (Cartesian Product) between Tables 58  
Example 3.10  Generating Every Combination of Rows between Tables Based on a Common Column 61
Example 3.11 Generating Every Combination of Observations between Data Sets Based on a Common Variable When an Index Is Available

Example 3.12 Combining and Collapsing Observations Based on a Common Variable

Example 3.13 Combining and Collapsing Observations Based on a Common Variable When the Transaction Data Set Is Indexed

Chapter 4 Using Lookup Tables to Match Data

Example 4.1 Performing a Simple Table Lookup

Example 4.2 Performing a Table Lookup in a Small Lookup Data Set

Example 4.3 Performing a Table Lookup in a Large, Non-Indexed Lookup Data Set

Example 4.4 Performing Multiple Lookups for Each Observation in a Data Set

Example 4.5 Performing a Table Lookup When the Lookup Data Set Is Indexed

Example 4.6 Performing a “Chained” Lookup

Chapter 5 Combining Summary and Detail Data

Example 5.1 Adding Values to All Observations in a Data Set

Example 5.2 Adding Values from the Last Observation in a Data Set to All Observations in a Data Set

Example 5.3 Computing Summary Data and Combining Them with Detail Data

Example 5.4 Subsetting a Table Based on the Calculated Average of a Group

Example 5.5 Calculating Totals across a BY Group to Produce Cumulative and Grand Totals

Example 5.6 Calculating Percentages and Statistics That One Row Contributes to a BY Group

Chapter 6 Updating Data Sets by Match-Merging by Value

Example 6.1 Updating a Data Set and Controlling Whether Common Variables Are Overwritten with Missing Values

Example 6.2 Updating a Data Set and Allowing Some Values to Be Updated with Missing Values

Example 6.3 Merging Data Sets and Conditionally Overwriting Common Variables

Example 6.4 Adding Observations and Variables to the Master Data Set When Duplicate Matching Variable Values Exist in the Transaction Data Set

Example 6.5 Saving Observations from Only the Master Data Set When the Transaction Data Set Contains Duplicates
## Chapter 7  Modifying Data Sets in Place

| Example 7.1 | Modifying All Observations in a Data Set in Place | 164 |
| Example 7.2 | Modifying a Non-Indexed Data Set in Place by Matching by a Common Variable | 167 |
| Example 7.3 | Modifying an Indexed Master Data Set in Place | 172 |
| Example 7.4 | Modifying an Indexed Master Data Set in Place When Both the Master and Transaction Data Sets Contain Duplicate Key Values | 176 |

## Chapter 8  Manipulating Data from a Single Source

| Example 8.1 | Performing a Simple Subset | 182 |
| Example 8.2 | Separating Unique Observations from Duplicate Observations Based on BY Values | 185 |
| Example 8.3 | Separating Completely Duplicate Observations from Unique Observations | 187 |
| Example 8.4 | Separating the First Observation in a BY Group from the Other Observations in the BY Group | 190 |
| Example 8.5 | Accessing a Specific Number of Observations from the Beginning and End of a Data Set | 193 |
| Example 8.6 | Creating a Customized Sort Order without Adding a New Column to the Table | 196 |
| Example 8.7 | Adding New Observations to the End of a Data Set | 199 |
| Example 8.8 | Adding Observations to a Data Set Based on the Value of a Variable | 201 |
| Example 8.9 | Adding Observations to a SAS Data Set So the Values of a Variable Are Consecutive throughout the BY Group | 203 |
| Example 8.10 | Adding Rows to a Table So That All Possible Values of Specific Columns Are Present in Each BY Group | 208 |
| Example 8.11 | Expanding Single Observations into Multiple Observations | 211 |
| Example 8.12 | Collapsing Observations within a BY Group into a Single Observation | 215 |
| Example 8.13 | Obtaining the Previous Value of a Variable within a BY Group | 220 |
| Example 8.14 | Comparing the Value of a Variable to Its Value in the Next Observation | 225 |
| Example 8.15 | Applying the Same Operation to a Group of Variables | 227 |
| Example 8.16 | Obtaining Hierarchical Data from a Table and Matching Them to the Rows in the Same Table | 229 |
| Example 8.17 | Combining Generation Data Sets | 232 |

## Chapter 9  Manipulating Data with Utilities and Functions

| Example 9.1 | Converting Variable Types from Character to Numeric and Vice Versa | 238 |
| Example 9.2 | Determining the Type of a Variable’s Content | 240 |
| Example 9.3 | Determining Whether a Variable Is Character or Numeric | 245 |
| Example 9.4 | Specifying a Numeric or Character Format at Run Time | 251 |
| Example 9.5 | Creating Columns That Contain the Attributes of Other Columns | 254 |
Example 9.6  Sorting Variable Values within an Observation  258
Example 9.7  Shifting Nonmissing Values Left in an Observation  262
Example 9.8  Generating Random Numbers within a Range of Values  265
Example 9.9  Selecting Observations at Random from a Data Set without Replacement  268
Example 9.10  Selecting Equal-Sized Samples from Different Groups  273
Example 9.11  Creating SAS Datetime Values and Computing the Difference between Two Datetime Values  278
Example 9.12  Creating a SAS Time Value from a Character Value  280
Example 9.13  Calculating a Person’s Age  282
Example 9.14  Incrementing a Date by an Interval  285
Example 9.15  Determining the Number of U.S. Business Days between Two Dates  290
Example 9.16  Counting the Occurrences of a String  292
Example 9.17  Extracting a Character String without Breaking the Text in the Middle of a Word  294
Example 9.18  Cleaning Character Data Values  296
Example 9.19  Validating and Standardizing Complex Character Data  300

Index  307