An Introduction to SAS® University Edition

Ron Cody
Contents

List of Programs ......................................................................................................................... xi
About This Book .......................................................................................................................... xvii
About the Author ........................................................................................................................ xxi
Acknowledgments ..................................................................................................................... xxiii

Part 1: Getting Acquainted with the SAS Studio Environment .............................................. 1

Chapter 1: Introduction to the SAS University Edition ......................................................... 3
Introduction: An Overview of SAS and the SAS University Edition ........................................ 3
Obtaining Your Free Copy of SAS ............................................................................................. 4
Installing the SAS University Edition ...................................................................................... 6
Conclusion .................................................................................................................................. 10

Chapter 2: The SAS Studio Interface ...................................................................................... 11
Introduction ................................................................................................................................. 11
Exploring the Built-In Data Sets .............................................................................................. 13
Sorting Your Data ..................................................................................................................... 17
Switching Between Column Names and Column Labels ...................................................... 19
Resizing Tables .......................................................................................................................... 20
Creating Filters .......................................................................................................................... 20
Conclusion .................................................................................................................................. 23

Chapter 3: Importing Your Own Data ..................................................................................... 25
Introduction ................................................................................................................................. 25
Exploring the Utilities Tab ......................................................................................................... 25
Importing Data from an Excel Workbook .............................................................................. 27
Listing the SAS Data Set .......................................................................................................... 32
Importing an Excel Workbook with Invalid SAS Variable Names ....................................... 33
Importing an Excel Workbook That Does Not Have Variable Names .................................. 34
Chapter 14: Subsetting and Combining SAS Data Sets ................................. 193

Introduction ........................................................................................................... 193
Subsetting (Filtering) Data in a SAS Data Set ......................................................... 194
Describing a WHERE= Data Set Option .................................................................. 197
Describing a Subsetting IF Statement .................................................................... 198
A More Efficient Way to Subset Data When Reading Raw Data ............................. 199
Creating Several Data Subsets in One DATA Step .................................................. 199
Combining SAS Data Sets (Combining Rows) ......................................................... 200
Adding a Few Observations to a Large Data Set (PROC APPEND) ......................... 202
Interleaving Data Sets ............................................................................................ 203
Merging Two Data Sets (Adding Columns) .............................................................. 203
Controlling Which Observations Are Included in a Merge (IN= Data Set Option) .... 205
Performing a One-to-Many or Many-to-One Merge ............................................... 208
Merging Two Data Sets with Different BY Variable Names .................................... 210
Merging Two Data Sets with One Character and One Numeric BY Variable ......... 211
Updating a Master File from a Transaction File (UPDATE Statement) .................. 214
Conclusion ............................................................................................................... 216
Problems .................................................................................................................. 217

Chapter 15: Describing SAS Functions ................................................................. 219

Introduction .......................................................................................................... 220
Describing Some Useful Numeric Functions ............................................................. 220
  Function Name: MISSING ..................................................................................... 221
  Function Name: N ................................................................................................. 221
  Function Name: NMISS ....................................................................................... 222
  Function Name: SUM ........................................................................................... 222
  Function Name: MEAN ........................................................................................ 223
  Function Name: MIN ............................................................................................ 223
  Function Name: MAX .......................................................................................... 224
  Function Name: SMALLEST ................................................................................ 224
  Function Name: LARGEST .................................................................................. 225
Programming Example Using the N, NMISS, MAX, LARGEST, and MEAN Functions 226
  Function Name: INPUT ....................................................................................... 227
  CALL Routine: CALL SORTN ............................................................................. 227
  Function Name: LAG ........................................................................................... 229
  Function Name: DIF ............................................................................................ 231
Describing Some Useful Character Functions .............................................................. 231
  Function Names: LENGTHN and LENGTHC ............................................................. 231
  Function Names: TRIMN and STRIP ................................................................. 233
  Function Names: UPCASE, LOWCASE, and PROPCASE (Functions That Change Case) ................................................................. 234
  Function Name: PUT ...................................................................................... 235
  Function Name: SUBSTRN (Newer Version of the SUBSTR Function) ............... 236
  Function Names: FIND and FINDC ................................................................. 237
  Function Names: CAT, CATS, and CATX ......................................................... 238
  Function Names: COUNT and COUNTC .......................................................... 240
  Function Name: COMPRESS ......................................................................... 241
  Function Name: SCAN .................................................................................. 244
  CALL Routine: CALL MISSING ........................................................................ 245
  Function Names: NOTDIGIT, NOTALPHA, and NOTALNUM ......................... 246
  Function Names: ANYDIGIT, ANYALPHA, and ANYALNUM ......................... 246
  Function Name: TRANWRD ............................................................................ 247
Conclusion ............................................................................................................. 247
Problems ................................................................................................................. 248

Chapter 16: Working with Multiple Observations per Subject .................. 251
  Introduction ..................................................................................................... 251
  Useful Tools for Working with Longitudinal Data ........................................... 251
  Describing First. and Last. Variables ............................................................. 252
  Computing Visit-to-Visit Differences ............................................................ 253
  Computing Differences Between the First and Last Visits ......................... 255
  Counting the Number of Visits for Each Patient ........................................... 256
Conclusion ............................................................................................................. 258
Problems ................................................................................................................. 258

Chapter 17: Describing Arrays ................................................................. 261
  Introduction ..................................................................................................... 261
  What Is an Array? .......................................................................................... 261
  Describing a Character Array ........................................................................ 264
  Performing an Operation on Every Numeric Variable in a Data Set ............... 265
  Performing an Operation on Every Character Variable in a Data Set ............ 265
  Converting a Data Set with One Observation per Subject into a Data Set with Multiple Observations per Subject................................. 266
Converting a Data Set with Multiple Observations per Subject into a Data Set with One Observation per Subject ................................................................. 268
Conclusion .......................................................................................................................... 269
Problems................................................................................................................................ 270

Chapter 18: Displaying Your Data ....................................................................................... 273
Introduction .......................................................................................................................... 273
Producing a Simple Report Using PROC PRINT ................................................................. 273
Using Labels Instead of Variable Names as Column Headings ............................................ 277
Including a BY Variable in a Listing .................................................................................... 279
Including the Number of Observations in a Listing .............................................................. 280
Conclusion .......................................................................................................................... 280
Problems................................................................................................................................ 281

Chapter 19: Summarizing Data with SAS Procedures ......................................................... 283
Introduction .......................................................................................................................... 283
Using PROC MEANS (with the Default Options) ................................................................. 283
Using PROC MEANS Options to Customize the Summary Report ....................................... 286
Computing Statistics for Each Value of a BY Variable ......................................................... 288
Using a CLASS Statement Instead of a BY Statement ......................................................... 289
Including Multiple CLASS Variables with PROC MEANS .................................................. 290
Statistics Broken Down Every Way ....................................................................................... 291
Using PROC MEANS to Create a Summary Data Set ............................................................ 293
Letting PROC MEANS Name the Variables in the Output Data Set ..................................... 294
Creating a Summary Data Set with CLASS Variables ......................................................... 295
Using a Formatted CLASS Variable .................................................................................... 297
Demonstrating PROC UNIVARIATE ................................................................................. 298
Conclusion .......................................................................................................................... 302
Problems................................................................................................................................ 303

Chapter 20: Computing Frequencies ..................................................................................... 305
Introduction .......................................................................................................................... 305
Creating a Data Set to Demonstrate Features of PROC FREQ ............................................ 305
Using PROC FREQ to Generate One-Way Frequency Tables ........................................... 307
Creating Two-Way Frequency Tables .................................................................................... 309
Creating Three-Way Frequency Tables .................................................................................. 311
Using Formats to Create Groups for Numeric Variables ...................................................... 313
Chapter 4: Creating Reports

Introduction ..................................................................................................................37
Using the List Data Task to Create a Simple Listing .................................................38
Filtering Data ..............................................................................................................44
Sorting Data ................................................................................................................46
Outputting PDF and RTF Files ....................................................................................52
Joining Tables (Using the Query Window) .................................................................53
Conclusion ..................................................................................................................63

Introduction

In the last chapter, you saw how to use the Import Data facility on the Utilities tab to import data. In this chapter, you will see how to use several of the most useful tasks as well as the Query tool on the Utilities tab.

In this book, as well as in SAS Studio, you will see the terms SAS data set and table used interchangeably as well as these other equivalent terms: variables are also called columns and observations are called rows.
Using the List Data Task to Create a Simple Listing

SAS Studio has dozens of built-in tasks. If you expand the Tasks tab, you will see the following:

![Figure 4.1: Data Tasks](image)

You can use the List Data task to create a listing of your data set. To demonstrate this, we are going to use the built-in Sashelp data set called Fish. This data set contains information on several species of fish, including weight, length, and width. To create a listing of this data set, expand the list of Data tasks and double-click on List Data. This brings up the screen shown in Figure 4.2:
You can click on the icon at the top-right part of this screen to select the library and data set you wish to list. Because you want a listing of Sashelp.Fish, select this data set.
An Introduction to SAS University Edition

Figure 4.3: Selecting the Fish Data Set in the Sashelp Library

The next step is to click on the plus sign (+) to select which variables you want to include in your listing (see Figure 4.4):

Figure 4.4: Adding Columns (Variables)

When you see the list of variables, you can select them in the usual way (see the instructions below):

To select variables from a list, use one of these two methods: 1) Hold the Ctrl key down and select the variables you want; or 2) click on one variable, hold the Shift key down, and click on another variable—all the variables from the first to the last will be selected.
In this example, you are selecting **Species**, **Weight**, **Height**, and **Width**.

**Figure 4.5: Selecting Variables to List**

![Selecting Variables to List](image)

Click **OK** when you are finished. You can create the listing now or click on the **OPTIONS** tab to customize the listing.

**Figure 4.6: Using the OPTIONS Tab to Customize the Listing**

![Using the OPTIONS Tab to Customize the Listing](image)

Here is the list of options available for the **List Data** task:
You can check or uncheck the **Display row numbers** box to include the Obs column in the listing or omit it. You have a choice of using column labels or column names in the listing. If you are a programmer, you will probably want to see column names as headings—if you are creating the listing for a report, you will probably want to see column labels.

The option displayed at the bottom of Figure 4.7 gives you the choice of listing all the rows of the table or the first $n$ rows. In this example, you want to see the first seven rows of the Fish data set. This is shown in the next figure:
Figure 4.8: Requesting the First Seven Rows to Be Displayed

Clicking on the **Run** icon generates the following listing:

**Figure 4.9: Listing of the First Seven Rows of the Fish Data Set**

<table>
<thead>
<tr>
<th>Obs</th>
<th>Species</th>
<th>Weight</th>
<th>Height</th>
<th>Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bream</td>
<td>242</td>
<td>11.5200</td>
<td>4.0200</td>
</tr>
<tr>
<td>2</td>
<td>Bream</td>
<td>290</td>
<td>12.4800</td>
<td>4.3056</td>
</tr>
<tr>
<td>3</td>
<td>Bream</td>
<td>340</td>
<td>12.3778</td>
<td>4.8891</td>
</tr>
<tr>
<td>4</td>
<td>Bream</td>
<td>383</td>
<td>12.7300</td>
<td>4.4555</td>
</tr>
<tr>
<td>5</td>
<td>Bream</td>
<td>430</td>
<td>12.4440</td>
<td>5.1340</td>
</tr>
<tr>
<td>6</td>
<td>Bream</td>
<td>450</td>
<td>13.8024</td>
<td>4.9274</td>
</tr>
<tr>
<td>7</td>
<td>Bream</td>
<td>500</td>
<td>14.1795</td>
<td>5.2785</td>
</tr>
</tbody>
</table>
Filtering Data

Another useful task is to filter the table—that is, you select rows that meet predefined criteria. To do this, double-click on Filter Data in the task list.

Figure 4.10: Selecting Filter Data from the Task List

This brings up the following:

Figure 4.11: Selections for Filtering Data
You select the data set as before. Next, you select a variable and a condition for your filter. In this example, you are selecting **Weight** as your variable and **Greater than** as your condition. You can now enter a value for the filter. In this example, you want to see rows in the table where the variable Weight is greater than 1,100.

Figure 4.12: Selecting Rows Where the Weight Is Greater Than 1,100

You can also expand the **OUTPUT DATA SET** option to override a default data set name. In most cases, you will want to supply your own data set name. Here you are naming the data set **Big_Fish**. You can also check or uncheck the **Show Output Data** box. Selecting it (as in this example) generates a listing of the filtered data set:
Figure 4.13: Listing of the Filtered Data Set (Big_Fish)

<table>
<thead>
<tr>
<th>Obs</th>
<th>Species</th>
<th>Weight</th>
<th>Length1</th>
<th>Length2</th>
<th>Length3</th>
<th>Height1</th>
<th>Width1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pike</td>
<td>1250</td>
<td>52</td>
<td>56.0</td>
<td>59.7</td>
<td>10.6863</td>
<td>3.3849</td>
</tr>
<tr>
<td>2</td>
<td>Pike</td>
<td>1600</td>
<td>58</td>
<td>60.0</td>
<td>64.0</td>
<td>9.0000</td>
<td>6.1440</td>
</tr>
<tr>
<td>3</td>
<td>Pike</td>
<td>1550</td>
<td>56</td>
<td>60.0</td>
<td>64.0</td>
<td>9.0000</td>
<td>6.1440</td>
</tr>
<tr>
<td>4</td>
<td>Pike</td>
<td>1650</td>
<td>56</td>
<td>63.4</td>
<td>66.0</td>
<td>10.8120</td>
<td>7.4800</td>
</tr>
</tbody>
</table>

Sorting Data

To sort data, select **Sort Data** from the list:

Figure 4.14: The Sort Data Task

Just as in the previous tasks, you can now choose a data set and options.
You are starting with the Sashelp.Fish data set and requesting a sort based on the variable Weight. You can also choose columns to drop:
You are dropping the three Length variables.

Before you execute the sort, there are several options you should consider. The default sort order is ascending (from smallest to largest). In this case, you want to see the heavier fish at the top of the list, so you choose **Descending** as the sort order.
Expand the **RESULTS** option to either sort in place or create a new data set with the sorted data.

**CAUTION:** Sorting in place replaces the original data set with the sorted data. If you drop columns, they will no longer be in the sorted data set.

In this example, you want to create a new data set called Sorted_Fish:

**Figure 4.18: Naming the Output Data Set**

![Screenshot of naming the output data set](image)

Click on the **Run** icon to see the following screen:
Figure 4.19: Result of Executing the Sort

To see a nicer listing of this data set, go back to the List Data selection and proceed as you did in the first section of this chapter. Once you have opened the List Data task, choose a table to display and any options that you want.
Next, choose which variables to include in the listing.

Finally, select any options that you want. In this example, you want to see the first eight rows of the table.
Figure 4.22: Option to List the First Eight Rows of the Table

Here is the listing:

Figure 4.23: First Eight Rows of the Sorted_Fish Data Set

Outputting PDF and RTF Files

If you want either PDF or RTF (rich text format) output, first click on the SAS Studio Options icon.

Figure 4.24: Click on More Application Options

Click on Preferences, then Results, and then check or uncheck PDF and/or RTF. (Note: These preferences will remain in effect unless you change them later.)
If you have checked one of these file types, after you produce a listing, the two icons **Download results as a PDF file** or **Download results as an RTF file** will be displayed right above your listing (see Figure 4.25). Clicking on either of these icons will output the appropriate file type to a location of your choice.

**Figure 4.25: Downloading a PDF File**

![Figure 4.25: Downloading a PDF File](image)

Here is a listing of the PDF file:

**Figure 4.26: Listing of the PDF File**

<table>
<thead>
<tr>
<th>Species</th>
<th>Weight</th>
<th>Height</th>
<th>Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bream</td>
<td>242</td>
<td>11.5200</td>
<td>4.0200</td>
</tr>
<tr>
<td>Bream</td>
<td>200</td>
<td>12.4500</td>
<td>4.3056</td>
</tr>
<tr>
<td>Bream</td>
<td>340</td>
<td>12.3778</td>
<td>4.6981</td>
</tr>
<tr>
<td>Bream</td>
<td>363</td>
<td>12.7300</td>
<td>4.4555</td>
</tr>
<tr>
<td>Bream</td>
<td>430</td>
<td>12.4440</td>
<td>5.1340</td>
</tr>
</tbody>
</table>

**Joining Tables (Using the Query Window)**

The last topic in this chapter describes how to use the **Query** utility to join two tables. Two data sets, ID_Name and Grades, were created to explain how the joining process works. Here is a listing of these two data sets:
Figure 4.27: Listing of Data Sets ID_Name and Grades

There are several features of these two data sets that are important for you to notice. First, the ID variable in the ID_Name data set is in order—in the data set Grades, it is not. As you will see, this does not cause a problem—the Query tool automatically sort the data sets. Also, ID 003 is in data set ID_Name and not in data set Grades; ID 006 is in Grades but not in ID_Name. The goal is to join these two tables based on the ID column.

If you want to play along with this demonstration, you can run the program shown next to create these two tables (you may need to refer to the programming chapters to see how to do this):

Figure 4.28: Program to Create Data Sets ID_Name and Grades

```sas
data ID_Name;
  informat ID $3. Name $12.;
  input ID Name;
  datalines;
  001 Ron
  002 Jan
  003 Peter
  004 Paul
  005 Mary;
  data Grades;
  informat ID $3.;
  input ID Grade1-Grade3;
  datalines;
  005 78 80 82
  002 100 90 95
  001 99 95 98
  030 65 67 69
  004 85 86 84;
```

The first step in joining these two tables is to select **Query** from the **Utilities** tab.
This brings up the following screen:

The next step is to open the Libraries tab and find the Work library.
Left-click on each file and drag it into the Query window. (If you drag the second file on top of the first file, SAS Studio automatically assumes that you want to perform a join operation.)
In the pull-down menu on the **TABLES** tab, select **Join** (unless you dragged the second file on top of the first, in which case the Query tool assumes that you want a join).
This brings up the following screen:

**Figure 4.34: Getting Ready to Join the Two Tables**

Select **ID_Name** for the **Left table** and **Grades** for the **Right table**. For **Join type**, select **Inner join**.
Click **Save**. You now see the two tables with a Venn diagram that represents an inner join. If you are familiar with SQL, you already know the four types of joins. For those readers who are not, here is the explanation:

Because some IDs are only in one table, you have some decisions to make about how you want to handle the join. The most common join, selected in this example, is an **inner join**. This type of join includes only those rows where there is a matching ID in both files. An **outer join** includes all rows from both tables (with missing values in the rows from the table that does not contain an ID). Finally, the other two joins are a **left join** and a **right join**. In a left join, all IDs from the left table are included—in a right join, all IDs from the right table are included.
The next step is to name the columns that you want to use to join the tables. In this example, because ID is in both tables, the Query tool automatically selects ID for the join variables. You are free to select any variable from each file to construct the join, even if the variable names are not the same in the two files.

Next, select which columns you want in the joined table. To do this, click on the COLUMNS tab. Select the columns in the usual way, and drag them to the Add columns area.
The figure below shows the final list of columns in the joined table.

The last step is to click on the **PROPERTIES** tab and indicate if you want a table or a report. If you choose a table, you can name the location (the Work library in this example) and the table name.
Clicking on the Run icon finishes the join. A snapshot view of the resulting table is produced.

You can use the List Data task to create a listing of the resulting table. It is shown in Figure 4.42:
Figure 4.42: Listing of Combined Data Set

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>ID2</th>
<th>Grade1</th>
<th>Grade2</th>
<th>Grade3</th>
</tr>
</thead>
<tbody>
<tr>
<td>005</td>
<td>Mary</td>
<td>005</td>
<td>78</td>
<td>80</td>
<td>82</td>
</tr>
<tr>
<td>002</td>
<td>Jan</td>
<td>002</td>
<td>100</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>001</td>
<td>Ron</td>
<td>001</td>
<td>99</td>
<td>95</td>
<td>98</td>
</tr>
<tr>
<td>004</td>
<td>Paul</td>
<td>004</td>
<td>95</td>
<td>86</td>
<td>84</td>
</tr>
</tbody>
</table>

Because this was an inner join, only those IDs that were in both tables are listed in the final table.

Conclusion

Only a few of the more popular data tasks were described in this chapter. Once you get the knack of running a task, you should feel confident in trying out some of the other data tasks in the list. The decision to use a task or write a SAS program is a personal choice. For those with programming experience, writing a program may be the way to go—for those folks who are new to SAS and just want to get things done, using the tasks is a great way to go. Or you can do both! Use a task or a utility to get the basic program written for you, and then take that and edit it to do more.
Index

A
age
  computing from date of birth 190
  grouping using conditional processing 153–156
AND operator 158–159, 196
ANYALNUM function 246–247
ANYALPHA function 246–247
ANYDIGIT function 246–247
APPEND procedure 202
arguments 156
ARRAY statement 263, 264, 267–268
arrays
  about 261–263
  character 264
  converting data sets 266–268, 268–269
  performing operations on character variables in data sets 265–266
  performing operations on numeric variables in data sets 265
  problems 270–271
ascending sort 18
ASSIGNMENT statement 173
@ sign 115, 199
Autocomplete feature (SAS Studio) 93, 94
AUTONAME option 294–295
B
backslash (\) 126
Bar Chart task, DATA tab 81–83
bar charts
  adding formats to 188–189
  creating with response variables 81–83
  “black box” 106
Boolean logic, using in conditional processing 158–159
brackets 263
built-in data sets (SAS Studio) 13–17
built-in formats 144–145
BY statement 203, 205, 253, 257, 289–290
BY variable
  computing statistics for values of 288–289
  including in listing 279
C
CALL MISSING function 245
CALL routines 220
CALL SORTN function 227–229
Carpenter, Art
  Carpenter's Complete Guide to the PROC REPORT Procedure 281
  Carpenter's Complete Guide to the PROC REPORT Procedure (Carpenter) 281
case-sensitivity, in variable names 95
CAT function 238–240
CATS function 238–240
CATX function 238–240
Change button 30
classification variables 70–72
character arrays 264
class variables
  creating summary data set with 295–297
  formatted 297–298
  including with MEANS procedure 290–291
classification variables 70–72
CLM function 286
CODE tab 100–101
Cody, Ronald P.
  Learning SAS by Example 119
  SAS Functions by Example, Second Edition 190
  SAS Statistics by Example 303
column input 112–114
column pointer 115
columns
  adding 40
  switching between names and labels 19–20
COLUMNS tab 60–61
comma-separated values files, reading data in 110–111
compile stage 106
COMPRESS function 241–244
concatenation 200–201
conditional processing
  about 153
  checking data errors using 156–157
  grouping age using 153–156
  IN operator 157–158
  OR operators 159–161
  problems 162–163
  using Boolean logic 158–159
constants, date 184–185
CONTENTS procedure 123–124, 212
CONTINUE statement 176–177
contributor variables 205–206
COUNT function 240–241
COUNTC function 240–241
CSV files, importing data from 35–36
curly braces ({} 263
CV function 286
D
data
  See also displaying data
  See also graphing data
  See also importing data
  See also reading data
  See also summarizing data
  checking errors using conditional processing 156–157
  filtering 44–46, 194–197
  sorting 17–19, 46–52
DATA= procedure 124
data sets
  See also permanent data sets
  about 123–125, 193–194
APPEND procedure 202
  combining 200–201
  controlling observations included in 205–207
  converting 266–269
  creating permanent 133–134
  creating shared folders 128–132
  creating subsets in one DATA step 199–200
  creating to demonstrate features of FREQ procedure 305–307
  IF statement 198
  interleaving 203
  listing 32–33
  many-to-one merge 208–209
  merging 203–214
  merging with different BY variable names 210–211
  merging with one character and one numeric
    BY variable 211–214
  names of 95
  one-to-many merge 208–209
  performing operations on character variables in 265–266
  performing operations on numeric variables in 265
  problems 136, 217–218
  reading from permanent 134–135
  renaming 30
  shared folders 126–128
  subsetting (filtering) data in 194–197, 199
  temporary versus permanent 125–126
  UPDATE statement 214–216
  WHERE= option 197–198
DATA statement 128, 200
DATA steps
  about 91
  creating data subsets in 199–200
  FORMAT statement in 146
  how it works 97–101
DATA tab
  Bar Chart task 81–83
  Statistics task 66
Data Table icon 67
DATALINES statement 138
data-set-name 125–126

dates
  about 181
  adding formats to bar charts 188–189
  constants 184–185
  creating from month, day, and year values 184
  extracting day of the week, day of the month, and year from 186–188
  problems 191–192
  reading from text data 181–184
  YRDIF function 190

day, creating dates from values for 184
DAY function 186–188

day of the month, extracting from 186–188

day of the week, extracting from 186–188
delimited data, reading 102–103
delimiters
  reading data separated by other 111–112
  reading values separated by 107–110
descending sort 18, 19, 48
DIF function 231, 254
displaying data
  about 273
  including BY variables in listing 279
  including number of observations in listing 280
  PRINT procedure 273–277
  problems 281
  using labels as column headings 277–278
DLM= option, INFILE statement 111–112
DO group 166
DO loops
  See also looping
    about 166–169, 263
    with character values 170–172
    graphing equations with 169–170
DO UNTIL statement 174–175
DO WHILE statement 172–174
dollar sign ($) 102, 264
downloading SAS University Edition 6
DROP statement 213

E
ELSE IF statement 154
END statement 167
equations, graphing with DO loops 169–170
The Essential Guide to SAS Dates and Times, Second Edition (Morgan) 190
Excel files, reading 116–119
Excel workbooks
  importing data from 27–32
  importing data from with invalid SAS variable names 33–34
  importing data from with no variable names 34–35
execute stage 106
expression 172
external files, reading data from
  See reading data

F
FILE PRINT statement 185, 207, 253
files, supported 26
filtering data 44–46, 194–197
filters
  about 20
  creating 20–23
  removing 22–23
FIND function 237–238
FINDC function 237–238
First. variable 252–253
first visits, computing difference between last visit and 255–256
FIRSTOBS= option 277
fixed columns, reading data in 112–116
FMTLIB option 148
folders, shared 126–128
FORMAT procedure 139, 147–148
FORMAT statement 142, 146, 184, 297–298, 313
formats
  about 137–144
  adding to bar charts 188–189
  built-in 144–145
  comparing FORMAT statements in procedures and DATA steps 146
Index

formats (continued)
  creating groups for numeric variables using 313
  examples of 145–146
  making permanent 147–149
formatted input 112, 114–116
formatting 139
forward slash (/) 126
frequencies
  about 305
  creating data sets to demonstrate FREQ procedure 305–307
  creating groups for numeric variables using formats 313
  creating three-way frequency tables 311–312
  creating two-way frequency tables 309–311
  generating one-way frequency tables with FREQ procedure 307–309
  problems 314
frequency bar charts, creating 77–81
frequency tables 74
full join 205
functions
  about 220
  ANYALNUM 246–247
  ANYALPHA 246–247
  ANYDIGIT 246–247
  CALL MISSING 245
  CALL SORTN 227–229
  CAT 238–240
  CATS 238–240
  CATX 238–240
  character 231–247
  CLM 286
  COMPRESS 241–244
  COUNT 240–241
  COUNTC 240–241
  CV 286
  DAY 186–188
  DIF 231
  FIND 237–238
  FINDC 237–238
  INPUT 227
  LAG 229–231
  LARGEST 225, 226–227
  LENGTHC 231–233
  LENGTHN 231–233
  LOWCASE 234–235
  MAX 224, 226–227, 286
  MAXDEC=n 286–287
  MDY 184
  MEAN 223, 226–227, 286
  MEDIAN 286
  MIN 223–224, 286
  MISSING 156, 185, 221
  N 221–222, 226–227, 286
  NMISS 222, 226–227, 286
  NOTALNUM 246
  NOTALPHA 246
  NOTDIGIT 246
  numeric 220–231
  problems 248–250
  PROPCASE 234–235
  PUT 235–236
  RAND 284
  ROUND 190
  SCAN 244–245
  SMALLEST 224–225
  STD 286
  STDERR 286
  STRIP 233–234
  SUBSTRN 236–237
  SUM 222–223
  TRANWRD 247
  TRIMN 233–234
  UPCASE 234–235
  WEEKDAY 186–188, 220
  YEAR 186–188
  YRDIF 190
Index 333

G
Graph tasks 78
graphing data
about 77
adding group variables 83–84
creating bar charts with response
variables 81–83
creating frequency bar charts 77–81
creating pie charts 84–86
creating scatter plots 86–88
graphing equations, with DO loops 169–170
group variables 83–84

H
histogram 69
HISTOGRAM statement 299, 302
home screen (SAS Studio) 13

I
ID statement 139, 274–275, 299
IF statement 154–156, 198
IF-THEN-ELSE logic, checking for data
errors using 156–157
Import Data task 26, 27–28
Import window 29–30
importing data
from CSV files 35–36
from Excel workbooks 27–32
Excel workbooks with invalid SAS
variable names 33–34
Excel workbooks with no variable names
34–35
listing SAS data sets 32–33
Utilities tab 25–27
IN operator 157–158, 196
INFILE statement 96, 108, 110, 111–112,
114, 127, 132, 138
informat 112
INFORMAT statement 109
inner join 59, 206
INPUT CSV tab (SAS Studio) 118
INPUT function 227
INPUT statement 96–97, 98, 102, 108–109, 171,
199
INPUT XLSX tab (SAS Studio) 118
installing SAS University Edition 6–9
interleaving data sets 203
invalid SAS names 95, 118–119
iterative processing
See looping

J
joining
defined 203
tables 53–63

L
LABEL procedure 277–278
LABEL statement 150
labels
creating variable 150–151
using instead of variable names 277–278
LAG function 229–231
LARGEST function 225, 226–227
Last. variable 252–253
last visits, computing difference between first visit
and 255–256
Learning SAS by Example (Cody) 119
LEAVE statement 176–177
left join 59
LENGTH statement 109, 158, 264
LENGTHC function 231–233
LENGTHN function 231–233
LIBNAME statement 126, 127–128, 147
libraries
creating with SAS Studio 127–128
defined 125
LIBRARIES tab 103, 194–195
library-reference (libref) 125–126, 149
Likert scale 139
Linux, virtualization software for 5
List Data task
about 32–33
creating listings with 38–43
options for 41–42
list input 107
listings
   creating with List Data task  38–43
   SAS data sets  32–33
LOG window 99
logical comparison operators  154
longitudinal data, tools for working with  251–252
looping
   about 165
   CONTINUE statement 176–177
   DO group 166
   DO loops 166–172
   DO loops with character values 170–172
   DO UNTIL statement 174–175
   DO WHILE statement 172–174
   graphing equations with DO loop 169–170
   LEAVE statement 176–177
   problems 178–179
   UNTIL condition 175–176
   WHILE condition 174
LOWCASE function 234–235
M
   many-to-one merge 208–209
   MAX function 224, 226–227, 286
   MAXDEC= n function 286–287
   MDY function 184
   MEAN function 223, 226–227, 286
   MEANS procedure
      about 146
      breaking down statistics 291–292
      CLASS statement 289–290
      computing statistics for values of BY
      variables 288–289
      creating summary data set with 293–294
      creating summary data set with CLASS
      variables 295–297
   including CLASS variables with 290–291
   naming variables in output data set with 294–295
   BY statement 289–290
   using formatted CLASS variables 297–298
   using to customize Summary reports 286–287
   using with default options 283–285
   MEDIAN function 286
   MERGE statement 205, 216
   merging data sets 203–205, 210–214
   metadata 123
   MIDPOINTS= option 302
   MIN function 223–224, 286
   MISSING function 156, 185, 221
   MISSOVER option 108, 110
   month, creating dates from values for 184
   Morgan, Derek
My Folders tab 133–134
My Libraries tab (SAS Studio) 14
N
   N function 221–222, 226–227, 286
   navigation pane (SAS Studio) 13, 25
   NMISS function 222, 226–227, 286
   NOCUM statement 308
   NOOBS option, PRINT procedure 160
   NOPRINT option, MEANS procedure 293–294
   NOT operator 158–159
   NOTALNUM function 246
   NOTALPHA function 246
   NOTDIGIT function 246
   Notifications box 12
   number of visits per patient, counting 256–258
   numeric functions 220–231
   numeric variables
      creating groups for using formats 313
      performing operations on 265
      summarizing 65–70
OBS= option 276–277
observations
  about 251
  computing differences between first and last visits 255–256
  computing visit-to-visit differences 253–255
  controlling included 205–207
  counting number of visits per patient 256–258
  defined 99
  First. variable 252–253
  including number of in listing 280
  Last. variable 252–253
  longitudinal data 251–252
  problems 258–259
one-to-many merge 208–209
One-Way Frequencies task 72–75
one-way frequency tables, generating using FREQ procedure 307–309
open code 98
operations
  performing on character variables in data sets 265–266
  performing on numeric variables in data sets 265
options
  AUTONAME 294–295
  CHISQ TABLES 310
  DLM= 111–112
  FIRSTOBS= 277
  FMTLIB 148
  MIDPOINTS= 302
  MISSOVER 108, 110
  NOOBS 160
  NOPRINT 293–294
  OBS= 276–277
  ORDER= 313
  PLOTS 73
  PRINTALLTYPES 291–292
  RENAME= 210–214

TABLES 140–141
WHERE= 197–198
Options icon (SAS Studio) 52
OPTIONS statement 100
OPTIONS tab
  Graphs tasks 79–80
  One-Way Frequencies task 73
  Statistics task 66–68
OR operators 158–159, 159–161
ORDER= option 313
OS X, virtualization software for 5
outer join 59
OUTPUT statement 167, 177, 200, 268, 293–297
outputting PDF and RTF files 52–53

parentheses (()) 263
PDF files, outputting 52–53
permanent data sets
  compared with temporary data sets 125–126
  creating 133–134
  reading from 134–135
pie charts, creating 84–86
plots 68–69
PLOTS option 73
plus sign (+) 40, 67
PRINT procedure 98, 100, 103, 109, 139, 146, 160, 185, 196, 273–278, 279, 280, 293–294
PRINTALLTYPES option 291–292
problems
  arrays 270–271
  conditional processing 162–163
  displaying data 281
  frequencies 314
  looping 178–179
  observations 258–259
  procedures 303
  reading data 120–122
  SAS data sets 136, 217–218
  SAS dates 191–192
  SAS functions 248–250
procedure options 105
procedures
  See also MEANS procedure
  about 283
  APPEND 202
  CONTENTS 123–124, 212
  DATA= 124
  FORMAT 139, 147–148
  FORMAT statement in 146
  how they work 104–106
  LABEL 277–278
  PRINT 98, 100, 103, 109, 139, 146, 160, 185, 196, 273–278, 279, 280, 293–294
  problems 303
  SGPLOT 170
  SORT 205, 291
  TRANSPOSE 267, 268
  UNIVARIATE 298–302
processing
  See conditional processing
  See looping
programming
  about 91
  "black box" 106
  DATA step 97–101
  INPUT statement 102
  procedures 104–106
  reading delimited data 102–103
  SAS Studio programming windows 92
  writing 92–97
PROPCASE function 234–235
PROPERTIES tab 61–62
PUT function 235–236
PUT statement 157, 185, 206, 207, 213, 253
Q
QQPLOT statement 299, 302
Query task (Utilities tab) 54–55
Query window, joining tables using 53–63
QUIT statement 91
R
RAND function 284
reading data
  about 107
  comma-separated values files 110–111
  Excel files 116–119
  in fixed columns 112–116
  problems 120–122
  separated by other delimiters 111–112
  values separated by delimiters 107–110
reading SAS data sets
  See data sets
removing filters 22–23
RENAME= option 210–214
renaming SAS data set 30
reports
  about 37
  creating simple listings with List Data task 38–43
  filtering data 44–46
  joining tables 53–63
  outputting PDF and RTF files 52–53
  producing using PRINT procedure 273–277
  sorting data 46–52
resizing tables 20
response variables, creating bar charts with 81–83
RESULTS window 99–100
RETAIN statement 255–256, 269
right join 59
ROUND function 190
rows, combining 200–201
RTF files, outputting 52–53
RUN icon 31, 98
RUN statement 91, 97, 138
S
SAS data sets
  See data sets
SAS dates
  See dates
SAS functions
  See functions
SAS Functions by Example, Second Edition
(Cody) 190
SAS Log 109–110
SAS procedures
  See procedures
SAS programming
  See programming
SAS Statistics by Example (Cody) 303
SAS Studio
  about 11–13
  Autocomplete feature 93, 94
  built-in data sets 13–17
  creating filters 20–23
  creating libraries with 127–128
  INPUT CSV tab 118
  INPUT XLSX tab 118
  LIBRARIES tab 103, 194–195
  navigation pane 13, 25
  options 93
  preferences 94
  programming windows 92
  resizing tables 20
  Snippets tab 117
  sorting data 17–19
  summarizing data using 65–75
  switching between column names and
  labels 19–20
SAS University Edition
  See also specific topics
  about 3
  downloading 6–9
  installing 6–9
  obtaining a free copy of 4–6
  updates for 9
Sashelp library 14–15
SAVE AS icon 100–101
SCAN function 244–245
scatter plots 86–88
Select File button 28–29
semicolons 91
SET statement 134–135, 157, 196, 200–201,
  202, 203, 253, 257, 268
Settings screen (List Data task) 38
SGPLOT procedure 170
  shared folders 126–132
  slash (/) 105
SMALLEST function 224–225
SORT procedure 205, 291
  sorting data 17–19, 46–52
square brackets ([]) 263
Start Menu 9
statements
  BY 203, 205, 253, 257, 289–290
  ARRAY 263, 264, 267–268
  ASSIGNMENT 173
  CLASS 289–290
  CONTINUE 176–177
  DATA 128, 200
  DATALINES 138
  DO UNTIL 174–175
  DO WHILE 172–174
  DROP 213
  ELSE IF 154
  END 167
  FILE PRINT 185, 207, 253
  FORMAT 142, 146, 184, 297–298, 313
  HISTOGRAM 299, 302
  ID 139, 274–275, 299
  IF 154–156, 198
  INFILE 96, 108, 110, 111–112, 114, 127, 132,
    138
  INFORMAT 109
  INPUT 96–97, 98, 102, 108–109, 171, 199
  LABEL 150
  LEAVE 176–177
  LENGTH 109, 158, 264
  LIBNAME 126, 127–128, 147
  MERGE 205, 216
  NOCUM 308
  OPTIONS 100
  PUT 157, 185, 206, 207, 213, 253
  QQPLOT 299, 302
  QUIT 91
  RETAIN 255–256, 269
  RUN 91, 97, 138
Statements (continued)
  SUM 257
  TABLES 104–105
  TITLE 98, 149, 274
  UPDATE 214–216
  VALUE 140
  VAR 98, 274–276, 285, 287, 299
  WHERE 195–196, 197–198
Statistics task 65–66
STD function 286
STDDERR function 286
STRIP function 233–234
subsetting
  See data set
SUBSTRN function 236–237
SUM function 222–223
SUM statement 257
summarizing data
  adding classification variables 70–72
  character variables 72–75
  numeric variables 65–70
summary data set
  creating with CLASS variables 295–297
  creating with MEANS procedure 293–294
Summary reports, using MEANS procedure to customize 286–287
Summary Statistics task 67
supported files 26
Text data, reading dates from 181–184
three-way frequency tables 311–312
TITLE statement 98, 149, 274
trailing @ 199
TRANSPOSE procedure 267, 268
TRANWRD function 247
TRIMN function 233–234
two-way frequency tables 309–311
U
UNIVARIATE procedure 298–302
UNTIL condition 175–176
UPCASE function 234–235
UPDATE statement 214–216
updates, for SAS University Edition 9
Utilities tab 25–27, 54–55
V
valid SAS names 95
VALUE statements 140
VAR statement 98, 274–276, 285, 287, 299
variable labels 150–151
variable names
  about 95
  importing Excel workbooks with invalid 33–34
  importing Excel workbooks with no 34–35
  using labels instead of 277–278
variables
  character 72–75, 265–266
  choosing which to list 50–52
  classification 70–72
  contributor 205–206
  defined 99
  group 83–84
  naming in output data set with MEANS procedure 294–295
  numeric 65–70, 265
  response 81–83
  selecting 40, 73
Variables pane (SAS Studio) 15–16
View tab (SAS Studio) 19
VirtualBox 11
visit-to-visit differences, computing 253–255
VMWare Player, downloading SAS University Edition using 6–9

W

WEEKDAY function 186–188, 220
WHERE= option 197–198
WHERE statement 195–196, 197–198
WHILE condition 174
Windows, virtualization software for 5
work area, expanding in SAS Studio 16–17
Work library 31
workbooks
   See Excel workbooks
writing SAS data sets
   See data sets

Y

year
   creating dates from values for 184
   extracting from 186–188
YEAR function 186–188
YRDIF function 190
About The Author

Ron Cody, EdD, a retired professor from the Rutgers Robert Wood Johnson Medical School now works as a private consultant and a national instructor for SAS Institute Inc. 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, such as Learning SAS by Example: A Programmer's Guide; SAS Statistics by Example, Applied Statistics and the SAS Programming Language, Fifth Edition; The SAS Workbook; The SAS Workbook Solutions; Cody's Data Cleaning Techniques Using SAS, Second Edition; Longitudinal Data and SAS: A Programmer's Guide; SAS Functions by Example, Second Edition, and Cody's Collection of Popular Programming Tasks and How to Tackle Them, and Test Scoring and Analysis Using SAS, as well as countless articles in medical and scientific journals.

Learn more about this author by visiting his author page at http://support.sas.com/cody. There you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more.
Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

support.sas.com/bookstore for additional books and resources.