Chapter 1

The SAS Environment

What you will learn in this chapter

- ✓ overview of SAS environment
- ✓ parts of a program
- ✓ streaming data program

Overview of the SAS environment

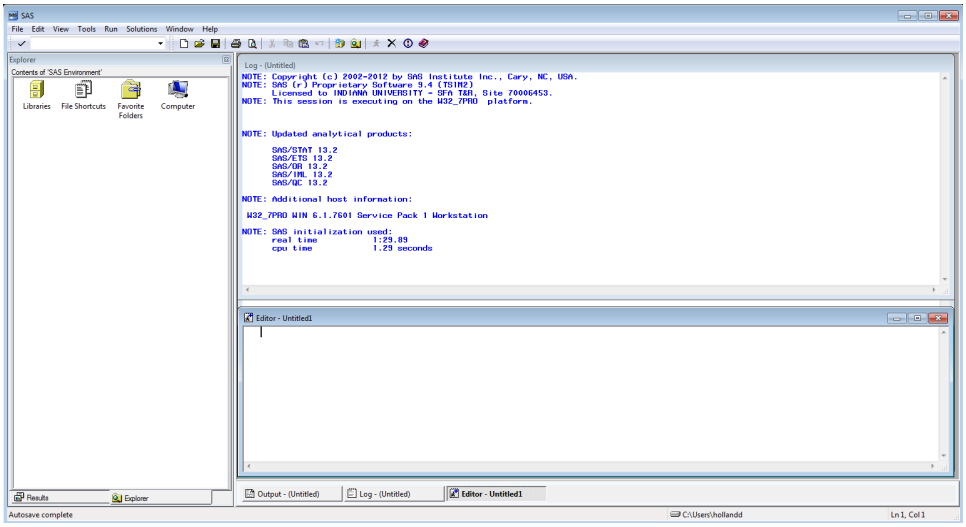
Understanding statistics requires some familiarity with the software program **Statistical Analysis System (SAS)** developed by the SAS Institute, Inc. Many people learn statistics using a pencil and paper approach. In this textbook, you will rely on software to complete the majority of calculations. Relying on software to complete calculations does not relieve one from needing to understand statistical techniques and the assumptions underlying each method. We go more into techniques and assumptions in later chapters. Most pressing is the need to familiarize you with the essential characteristics of the SAS environment. You should make a note, however, that **SAS 9.4** allows users to control many elements of the environment beyond material covered in this chapter or book. In order to help you understand the general appearance of the SAS software upon opening the program, we have to explain various types of information, such as, the windows that you will see, what the colors of the text and lines mean, how you can use the color coded text to keep organized, and other very detailed techniques that should facilitate your organization when using SAS software. At this point you should only read along. We do not suggest opening SAS 9.4 at this time.

Three (3) Windows: Editor, Log, Output

Most people are familiar with some type of word processing program. People routinely open word processing programs by going to the start button, finding the word processing program,

and then clicking it to open the program. Once the word processing program is open there is an unused document waiting for use. This initial document is a template. It has some items formatted for you. For example, it has margins set, a particular font, color of font, the line and paragraph spacing and many other aspects of the document already set-up. SAS software is similar. When you open SAS 9.4, the windows that open have some preset controls or characteristics. Unlike most word processing programs that open only one window (or document), SAS opens several, see Figure 1.1.

FIGURE 1.1 SAS 9.4 Windows



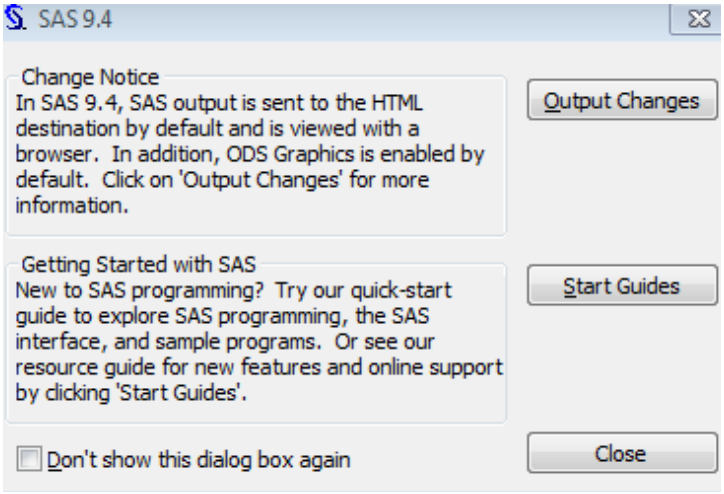
Source: Created with SAS software. Copyright 2009–2011, SAS Institute Inc. Cary, NC, USA. All Rights Reserved. Reproduced with permission of SAS Institute Inc. Cary, NC.

Having several windows open at once takes some getting used to. It is common to lose track of which window you are in. It is common to have some stress getting from one window to another. Later, when we use the software, you should prepare to experiment with transitioning from one window to the other until you become proficient at moving between the windows. If you take the time to do this, you will undoubtedly save yourself some headaches when you are learning more challenging material later. Now you know that SAS 9.4 comes with templates that open when you open the SAS program. You also know that the windows that open are templates and have some pre-set features. The windows that open are the editor, log, and output windows. Note that other windows may also open.

You may see another window open as the top window on your screen. A likely window that you may have to close is shown in Figure

1.2 SAS 9.4 Windows Top Dialogue Box. If you see Figure 1.2 when you open SAS 9.4, just select "close".

FIGURE 1.2 SAS 9.4 Windows Top Dialogue Box



You may save or print the content of each window in much the same way that you print or save a word processing document. You must make sure that you click inside of the window that you want to print or save. Then, go to the file option and make your selection to print or save. We will first explain the editor window.

Editor Window

The **editor window** is where you will place programming language as we move through the book. Everyone has heard of computer programming. In laymen's terms, programming is the use of a language unique to specific software that knows how to translate the language into functional, usable actions. Now, we are concerned that you are worried whether or not the programming required with this book is beyond your capabilities.

Annotated Activity 1.1

If only the output window is visible, how do you make the editor window appear?

_____

_____

_____

_____

Data set

A data set is the file that contains a collection of specific information about more than one element in a population.

No prior programming experience is needed to complete all tasks in this book. Clearly, having some experience would give one an edge and enable one to move more quickly through the book. Nonetheless, no prior programming experience is required.

Usually when SAS 9.4 opens you will see at least three windows. You need to look for the one labeled "Editor" or "Editor-Untitled1" or something similar. The name of the window varies depending upon what version of SAS software that you use, and depending upon how the last user set up the controls on the software. Even still, you should be able to discern the editor window from other windows based upon the window names. On the Window menu, you will see all windows that you have open, such as Editor Windows, Logs, and Results. Another way that you may identify the editor window is by looking at the bottom of the screen for tabs with names on them. The options usually include editor, log, and output. Clicking on the editor tab will open the editor window. Note that usually the default setting is for at least three windows to open and remain open on the screen. The top window is usually the log, the bottom window is usually the editor window, and the window on the left is the explorer/results window.

The editor window is where one places the SAS code. SAS code tells the software where to get the data set, and what to do with the data set and the items contained within the data set. When typing SAS code, you must pay close attention to what you type. Including even one extra period or space in the wrong place will cause your program to get errors, i.e., not work. All work done in the editor window must be saved in a file, as when writing a paper or letter in a word processing document. You have to be prepared to name the file, select a location to save the file, and be able to recall where the file was saved. It is important that you recall that the tab that used to state "Editor" will now indicate the name for your program file and will not state "Editor". This SAS feature is handy when you later have more than one program file open at a time. If you have not saved your program file (your editor window) and you open a second program file it will be named "Editor: untitled2", by default, and so on.

Log Window

Keeping a log of your activities while in SAS is crucial to successful statistical analysis. You may prefer to think of the log as a journal or diary that maintains a record of what you did while working in SAS. The first time you open SAS you will see the log window in the top half of the computer SAS window. At first you might wonder why there is information recorded on a log when you did not yet "do" anything in SAS. SAS begins recording all actions that SAS takes, including opening the editor window, the **log window**, and the **output window**; SAS initialization. SAS records how long it takes to process all requests.

The log may initially seem redundant with the editor window, as all information contained in the editor window will be recorded in the log. However, you should use the log to check your programming, rather than the editor. The log highlights errors and gives an indication

Annotated Activity 1.2

How is the log helpful with writing programs?

of where the programming error is located. The log uses color codes to highlight various messages to the programmer. Warnings may be in one color but red generally indicates an error, not a warning. Both errors and warnings must be managed by you and cannot be ignored. Note that there are times when warnings do not interfere with the statistical analysis, whereas other times they do. So, the programmer must make this determination.

Output Windows

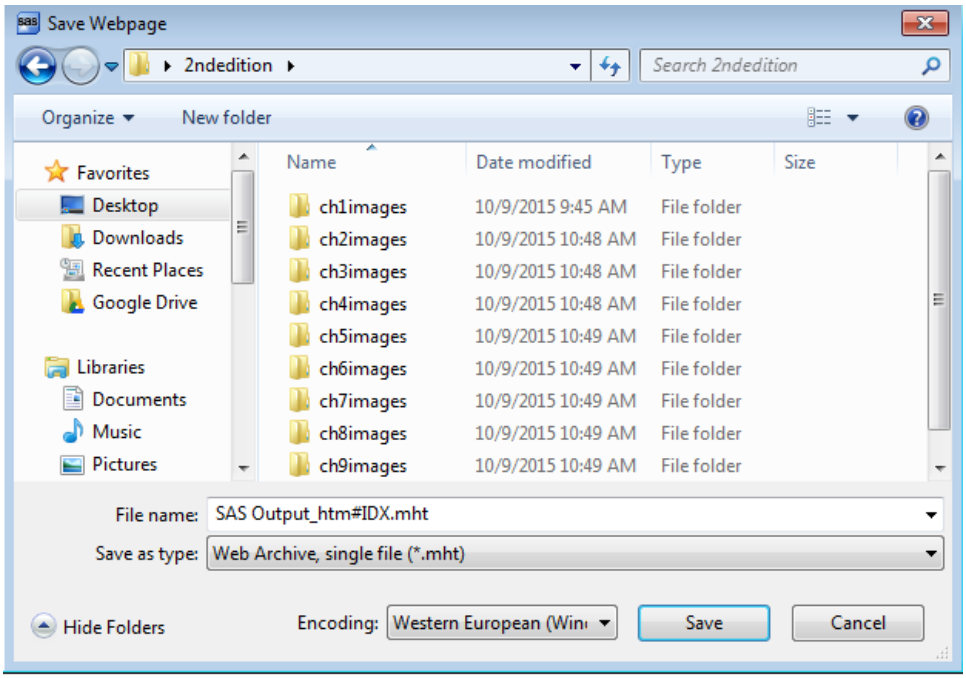
SAS output window is where you will find the results for your analyses. It has all of the results from the statistical analyses that were requested of SAS in the editor window. As you acquire more programming skills, you can program SAS to display the results of the statistical analysis in very particular ways. However, throughout this book, we generally rely on the default output display settings. There are two common Output Windows in SAS 9.4. One is named the "Output" tab that you see when you first open the program. The second place your results will appear are in the "Results Viewer – sashtml" window that does not appear on your screen until you run a SAS program that generates results. Note that there are other output windows that we will review later. The default process for SAS 9.2 and all prior versions was that basically all output was displayed in the "Output" window with SAS monospace font where graphs were not integrated with tables.

Since SAS 9.3, the output is automatically saved in an HTML file and the ODS Graphics are automatically enabled. The results are no longer displayed in the Output window, rather those results are now found in the Results window. The output is displayed in a new style, HTML Blue, designed to include tables with graphics. To access the output, you must click on the appropriately labeled folders in the Results window, drilling down until you see all of your desired results.

Finding the specific information you need from the results varies by the type of analysis you requested. Initially, finding the needed information to answer research questions is no simple matter. The output can seem rather overwhelming. Be patient with the learning process. There is an adjustment period with each new statistical analysis used. SAS Institute, Inc. offers many online help resources (See, http://support.sas.com/documentation/94/index.html).

As stated previously, SAS output is created and displayed in an Output window. If you wish to save your results to a file, you can click on the Output window, click on the File button at the top left, then click on 'Save As' to display the following window:

FIGURE 1.3 SAS 9.3 Save Results Output As



Source: Created with SAS software. Copyright 2009–2011, SAS Institute Inc. Cary, NC, USA. All Rights Reserved. Reproduced with permission of SAS Institute Inc. Cary, NC.

Simply enter the desired file name in the blank space at the bottom of the window, and click the Save button. Notice that the file is saved under the file name and has a .mht extension.

Color Coding Text

SAS Editor uses colors to specify certain elements of the program language. For example, red might indicate an error in coding. Navy blue represents elements of the program called commands. Keywords that you must define are lighter blue. Green represents comments. Recall that advanced users can change the color scheme and what each color represents. So, you must pay attention to the patterns on the computer used.

Annotated Activity 1.3

What information is contained in the log?

Box 1.1 Student Reaction to the SAS Environment

"*Compared to SPSS*" by Sara Kennedy

If you learned statistics on SPSS first, like I did, or if you have never had any programming experience SAS can be pretty overwhelming at first. Everything is in code, instead of in a matrix as in SPSS. But telling the program explicitly what to do, helped me to learn the differences between statistical tests and which test is appropriate for a specific research question. SAS also allows the user to input comments, and to view exactly how variables have been changed (i.e.: renamed or recoded), which is very helpful when working with large data sets. Manipulating the data can be scary, but SAS logs are user friendly and give the statistician opportunity to determine exactly what may have gone wrong with coding or renaming variables. Plus, isn't it fun to say 'proc freq'?

Window Options

Review the **options** available for each window by clicking inside the window and then notice that the windows options across the top of the screen change. We will highlight only two here. However, you should take time to familiarize yourself with the options for each window. If you click in the Editor window, you will see the options "file" and "run" across the top of the screen. "File" works much the same as it does in a word processing program. One can open, close, and save files from this option. To the right of this option is "run". Selecting "run" tells SAS to process all of the programming that is present in the editor window. Clicking this option after the programming is written generates the results, i.e., output. Once you learn programming and statistics, you will use the RUN command a lot.

Viewing Line Numbers

Writing programs may require hard work. Programs can become very large and unwieldy. In order to maintain organization and for easy reference, you often will want to be able to refer to a specific line of a program. SAS default settings do not include viewing line numbers, so you must set this up. Be prepared to do the following to view line numbers in your program: Click inside of the Editor window then select *Tools/Options/Enhanced Editor/* select the box for "show line numbers".

Commands

A command is the aspect of a SAS program that tells SAS exactly how to manage, manipulate or use an element or more than one element in the data file.

TABLE 1.1 Shortcuts

Ctrl-c	Copy text
Ctrl-v	Paste text
Ctrl-x	Cut text
End	Go to end of line
F4	Places last code in editor window
F5	Open editor window
F6	Opens log window
F7	Opens output window
F8	Submit program
Home	Go to beginning of line
Pg Dn	Go down one screen
Pg Up	Go up one screen
Ctrl-Break	Abort program

Adding Comments

Any time you complete a written project, such as a term paper, essay, or a program, you have a way to keep the material organized. With an essay you would have an introduction, the body of the paper, and then a conclusion. You know that the introduction comes first and so on. Because writing a program is similar in that it has to be organized in a particular way, it is advisable to insert your own comments to remind yourself where you are in the program, what you are trying to accomplish, and what your uncertainties are about your programming. You could also make comments that remind you what research question you are trying to answer, or what homework assignment or number you are completing. You can insert unlimited comments throughout your programs.

The first SAS coding that you learn in this book is how to insert your comments into the program. In the editor window you merely type

```
*place your comments here;
```

To clarify this syntax, first you place an asterisk. You can either have spaces or not, then type your comment and follow it by a semi-colon. *Note that all commands in SAS must end with a semi-colon.*

Parts of a Program

Understanding how the SAS program is organized is the first real step to being able to write a simple SAS program. The two parts of the program are referred to as the data step and the procedure step.

Data Step

The data step is where one can tell SAS what information is considered the data set or where SAS can locate the data set. You can also write SAS code to make changes to the data set or to tell SAS how to make new variables from existing variables in the data set. The data step allows users to make changes to existing variables. Although all programs do not require a data step, we will usually have one throughout the examples in this book.

Procedures Step

The procedure step is where one can tell SAS to complete statistical tests that we will cover later in the textbook. Initially you will use the procedure step to run frequencies, averages, and other descriptive statistics. Within the **procedure step** is where you will request SAS to calculate the majority of your statistical tests. The procedure step begins with the word PROC. Every procedure statement must end with a semicolon. Procedure statements can run across multiple lines. One must place the semi-colon at the very end of the statement, not at the end of a line. One example of what not to do: I once knew a person who used the procedure statement PROC PRINT while using a very large dataset. With very large data sets every command takes more time to complete than in small data sets. So, SAS took some time to run the frequencies. It ran frequencies on every variable in the dataset, something one does not usually do in very large datasets. Rather, one is more likely to select only certain variables to run frequencies. We show you later how to select only certain cases to run any SAS procedure. The other mistake the person made was to use **PROC PRINT** without specifying the variables to print. Most times one would not want to print every frequency of every variable in a dataset, unless the dataset was relatively small, or the person was familiar enough with the dataset to know that there were few continuous variables in the dataset. In this case, the person should have known that the dataset was very large. It had over 30,000 cases and over 1000 variables. Many of the variables were continuous with large ranges, such as household income, and family income. Thus, the person wasted three reams of paper printing the frequency for every variable in the dataset, that is, until the secretary canceled the print job. We do not know how much paper would have been wasted had the secretary not canceled the print request. Thus, look at your output and know what you want to print. Use PROC

Variable

A a variable is one characteristic of an element in a dataset. Examples include age, sex, race, education, income, health status, and self-esteem of the people in a dataset.

Continuous variable

Variables that have the ability to take on any value across the range of the variable itself. Examples include age, precisely calculated, height, and body mass index (BMI).

vs.

Discrete variable

Variables that can have only certain values. Examples include sex, education level, and racial/ethnic group.

Note: Researchers often discretize continuous variables. They convert variables from continuous to discrete. Example: change age from days old to two categories: old and not old.

Descriptive statistics

For researchers to understand overall basic patterns in the dataset, they use descriptive statistics. One might use descriptive statistics to determine the average age of people in the dataset, or to determine how many people answered yes or no to a particular question on a survey. Examples include mean, standard deviation, median, histogram, stem-and-leaf plots, correlation, and frequency distributions.

Annotated Activity 1.4

Write the SAS code to insert the following comment on your output/printout: I can do this!

PRINT only if you are familiar with your dataset and what information the command will print.

Guided Streaming [Inline] Data Program

Because it is good to rehearse what you plan to do before you do it. We provide here a simple way to use a data set in SAS. At this point, you do not need to do the following; you should follow along and think through the process as we review it.

We need to open SAS 9.4. Every bureaucracy has unique ways to access the software. Thus, you will have to seek assistance from your technical support staff to know how to access the SAS program. Also, you may have purchased your own copy of the program. In this case, follow the instructions that came with the software to open SAS.

Now that we have SAS open you will see the following windows: the log, the editor, and the results/explorer window. We click inside of the editor window, then type Program 1.1 exactly as it is written below:

Program 1.1 Read Data in Program to Temporary Data Set while Using Comments

```
*** SAS PROGRAMMING *** ;
*** DATE (UPDATE THIS EACH TIME) *** ;
*** YOUR NAME *** ;
*** HOMEWORK # OR NAME (UPDATE THIS EACH TIME) *** ;
*** DATE DUE *** ;
*** PROGRAM 1.1 BEGINS HERE *** ;
TITLE1 'ASSMT1.SAS' ;
TITLE2 'READ DATA FROM WITHIN PROGRAM' ;
TITLE3 'CREATE TEMPORARY SAS DATASET CALLED ONE' ;
TITLE4 'YOUR NAME';
OPTIONS COMPRESS = YES ;
DATA ONE;
INPUT NAME $ AGE HEIGHT WEIGHT GENDER $ ;
LABEL NAME = 'NAME'
AGE = 'AGE IN YEARS'
```

```
   HEIGHT = 'HEIGHT IN INCHES'
   WEIGHT = 'WEIGHT IN POUNDS' ;
datalines ;
   JAKE 62 72 185 M
   GARY 57 65 135 M
   ROSE 47 62 100 F
   NANCY 53 66 115 F
;
PROC PRINT ;
RUN ;
*** PROGRAM ENDS HERE *** ;
```


All SAS programs must end with RUN. If you forget to include "run;" SAS will not run your program properly. You may have to use "ctrl-break" to cancel the program submission to SAS. All green lines are comments because they begin with an asterisk and end with a semi-colon.

Next, we click on the menu item that looks like a little guy running to submit the program for SAS to process. We save our output, we save our log, and we save our program by clicking in each window, then going to "file", and selecting "save as". We name the files in the same way that we do any word processing file. We named the files: "assmt1". All files should have the same name and *can* have the same name because the extensions on the file types are unique. So our program file name will be "assmt1.sas". Our output file name will be "assmt1.lst". Our log file name will be "assmt1.log". SAS data files have a file name extension of "*.sas7bat".

Using the same name for all files will help keep the programs, logs, and output organized by assignment, project, or technique. Perhaps you will find a unique way to organize your SAS files.

After submitting the program, we examine the log and the output windows. If there are errors, we go back to the editor window to correct the errors and run the program again. Resave the program using the original name "assmt1". Print the output, log, and program by clicking in a particular window then go to the "file" option and select "print".

Key Terms

commands
comments
continuous variable
data set
data step
discrete variable
EDITOR
editor window
guided
INLINE
LOG
log window
OPTIONS
OUTPUT
output window
PROC
PROC PRINT
procedures step
descriptive statistics
SAS 9.2
SAS 9.3
Statistical Analysis System (SAS)
streaming
variable

Review

Name and describe the two parts of a SAS program.

_____
_____
_____
_____

Describe how to save and print a program.

Questions

1. When (or why) does the name of the Editor tab change?
   a. It does not change.
   b. Changes in the name of the Editor tab are automatically generated by the software.
   c. The name on the Editor tab changes to the name of the saved file.
   d. None of the above.

2. Where is the SAS code placed?
   a. You place SAS code in the Log.
   b. You place SAS code in the Editor window.
   c. You place SAS code in the Output window.
   d. You can place SAS code in any of the windows.

3. What does the SAS code do?
   a. It tells SAS where to get your data.
   b. SAS code tells SAS where to store your data.
   c. None of the above.
   d. Both a and b.

4. Where does SAS indicate programming errors and problems?
   a. in the Editor window
   b. in the Log window
   c. in the Results window
   d. None of the above.

5. What must all commands in SAS end with?
   a. a period
   b. a colon
   c. a semi-colon
   d. "run"

6. What must all SAS programs end with?
   a. a period
   b. a colon
   c. PROC
   d. "run"

7. Which of the following SAS Windows open when you first open SAS?
   a. Editor Window
   b. Log Window
   c. Results Window
   d. Both a and b

8. Which of the following places comments in programs?
   a. information contained within quotation marks
   b. information contained within parentheses
   c. information contained between an asterisk and a semi-colon
   d. information contained between a hash-tag and a semi-colon

9. Warning for errors in programs are found in
   a. the Log Window in a unique color, usually red color font
   b. the log in the same color as other text
   c. the Editor Window, usually as red color font
   d. the Editor Window in the same color as the other text

10. What are the two parts of a SAS program?
    a. Editor and Log
    b. Syntax and images
    c. Editor and Results
    d. Data Step and Procedure Step