Section I

Introducing *SAS Enterprise Guide*

1 *SAS Enterprise Guide* Projects

1.1 A brief history of SAS

The SAS Web site provides a comprehensive history of the software and the company. Here is a synopsis of that information. SAS, an acronym for Statistical Analysis Software, is a set of statistical analysis procedures housed together within a large application. The idea for it was conceived by Anthony J. Barr, a graduate student at North Carolina State University, between 1962 and 1964. Barr collaborated with Jim Goodnight in 1968 to integrate regression and analysis of variance (ANOVA) procedures into the software. The project received a major boost in 1973 from the contribution of John P. Sall. Other participants in the early years of SAS development included Caroll G. Perkins, Jolayne W. Service, and Jane T. Helwig. The SAS Institute was established in Raleigh, NC in 1976 when the first base SAS material was released. The company moved to its present location of Cary, NC in 1980.

SAS began being used on mainframe computers several decades ago. At that time, the only way to instruct the software to perform the statistical analyses was by punching holes on computer cards via a card-reader machine. Later this instruction occurred by typing in this code on an otherwise blank screen. The majority of SAS users still prefer this latter process.

SAS released its first Windows version in 1993. Windows uses a graphical user interface (abbreviated GUI but thought of by most people as a point-and-click interface) to make selections from menus and enter some limited text into dialog screens. These selections are translated "behind the scenes" to SAS code but the code can be viewed by a click of the mouse. *SAS Enterprise Guide* succeeded the Analyst interface and is the third iteration of SAS' GUI. It runs only in the Windows operating environment. Because *SAS Enterprise Guide* writes code and submits it

to SAS as you make selections with the mouse or type text into dialog screens, you also need to be using a computer on which SAS is installed, either a stand-alone personal computer or one that is connected to an organization's network.

This book was written by us based on SAS version 9.1 together with *SAS Enterprise Guide* version 4.0. This configuration is currently available under an organizational license, such as that purchased by a university or government agency. Therefore, certain users may have the software installed on computers owned by the organization, such as computers in a statistics laboratory. This same configuration under the title *SAS Publishing SAS Learning Edition 4.1* was also available from JourneyEd.com at the time we were writing this at a considerably discounted price (compared with the organizational license fees) to students and faculty members to load on their own personal computers with the Windows XP operating system.

1.2 Opening a project

We will assume that the shortcut to *SAS Enterprise Guide* 4.0 is visible on your desktop (if it is not then you can navigate to it in the **Program Files** folder on your internal drive). Open *Enterprise Guide* by double-clicking on its icon. This brings you to the window shown in Figure 1.1.

Everything in *SAS Enterprise Guide* is done within the context of a *project*. A project contains the data set and a history of its use, including the output of any statistical analyses that were performed. This will become quite familiar to you as we work through the chapters of this book; for now, treat this as information that you can read again as necessary. The initial screen for *SAS Enterprise Guide* therefore provides choices of which project or type of project we would like to open. Here are three of the more frequently used options:

- The top portion of our opening screen under **Open a project** lists some of the projects that we have recently opened. If we wished to open one of those, we would simply click on its name.
- If we wished to start a new project, we would select **New Project** in the **New** portion of the screen.
- If we wished to open an existent project whose name is not displayed on the initial screen, we would select **More projects** and then use the menu system to navigate to and open the desired project. Alternatively, we could select **New Project** and then select our project as described in the following section.

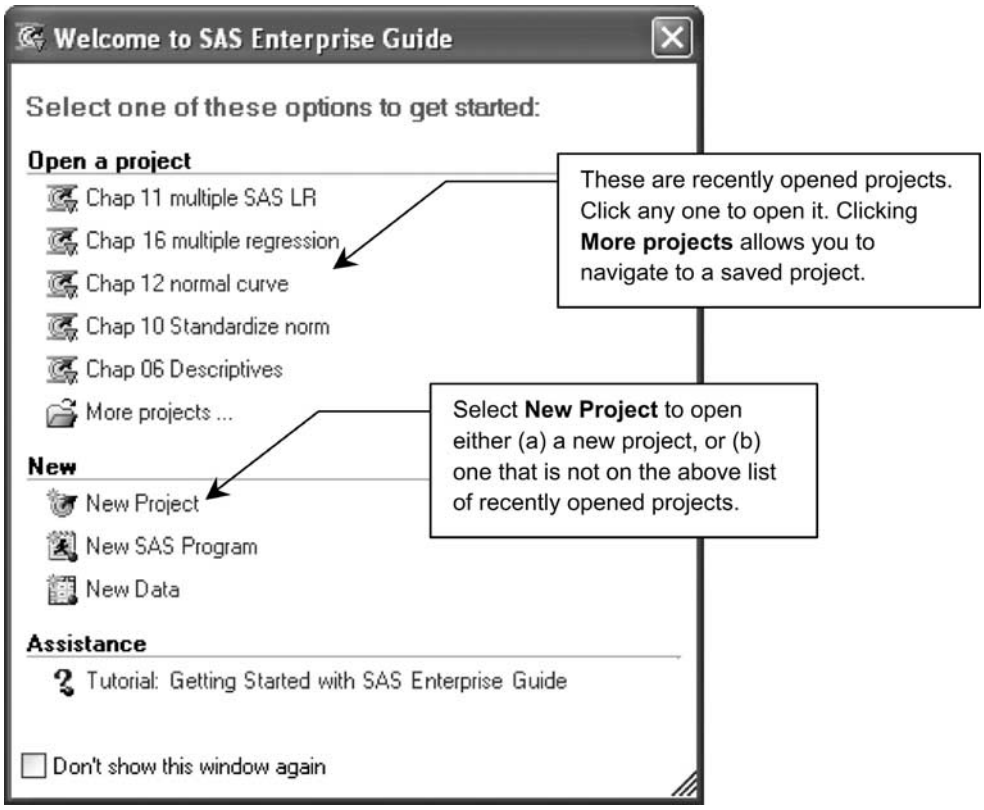

Figure 1.1. The startup screen for *SAS Enterprise Guide*.

1.3 The contents of projects

Selecting **New Project** brings us to the screen shown in Figure 1.2. We are presented with the **Process Flow** screen of the **Project Designer**. It is empty now but at various stages of our work it will contain a data set, the specifications of our analysis, and the results of the data analysis. The screen shows a grid that looks like graph paper – this is the background used by **Process Flow**. Because there is nothing in the project at this time, an empty **Process Flow** window is displayed.

We will open a project in order to show you what a typical project might contain. From the main menu, select **File ➔ Open ➔ Project** (see Figure 1.3). *SAS Enterprise Guide* will require you to indicate where your projects are located (see Figure 1.4); as ours are on the internal drive of our personal computer, we choose **Local Computer**,

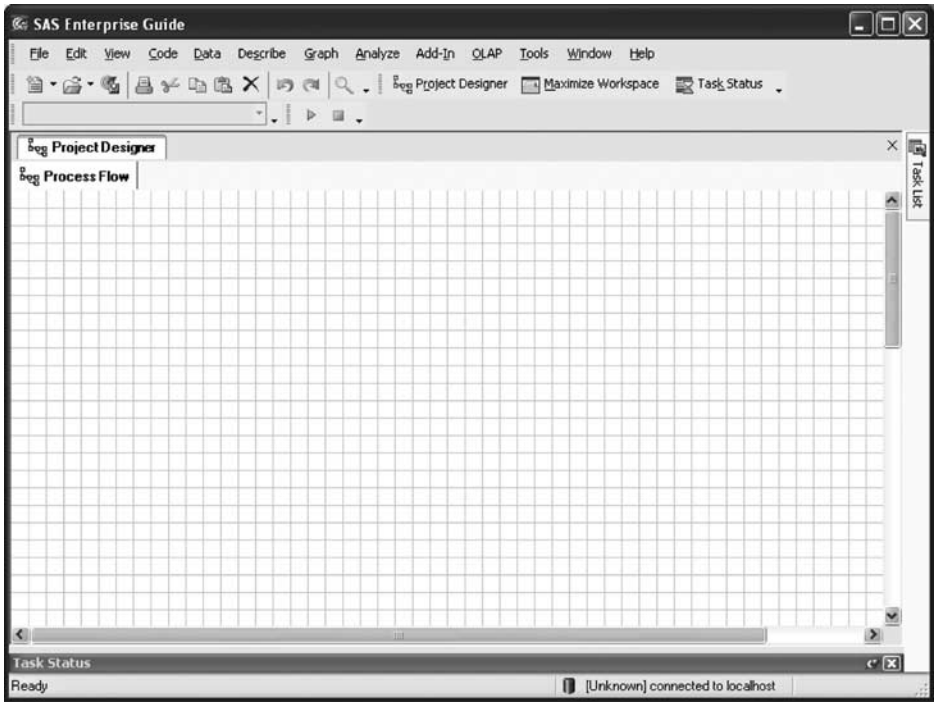

Figure 1.2. The **Project Designer** tab with an empty **Process Flow** window.

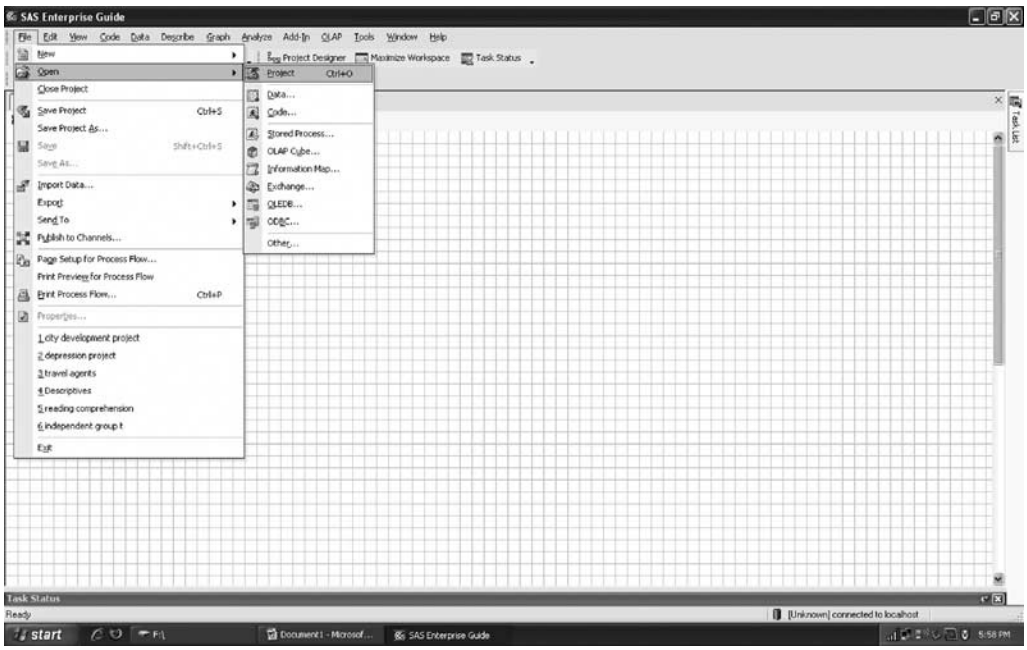

Figure 1.3. Opening a project.

Figure 1.4. Select the system on which your projects are located.

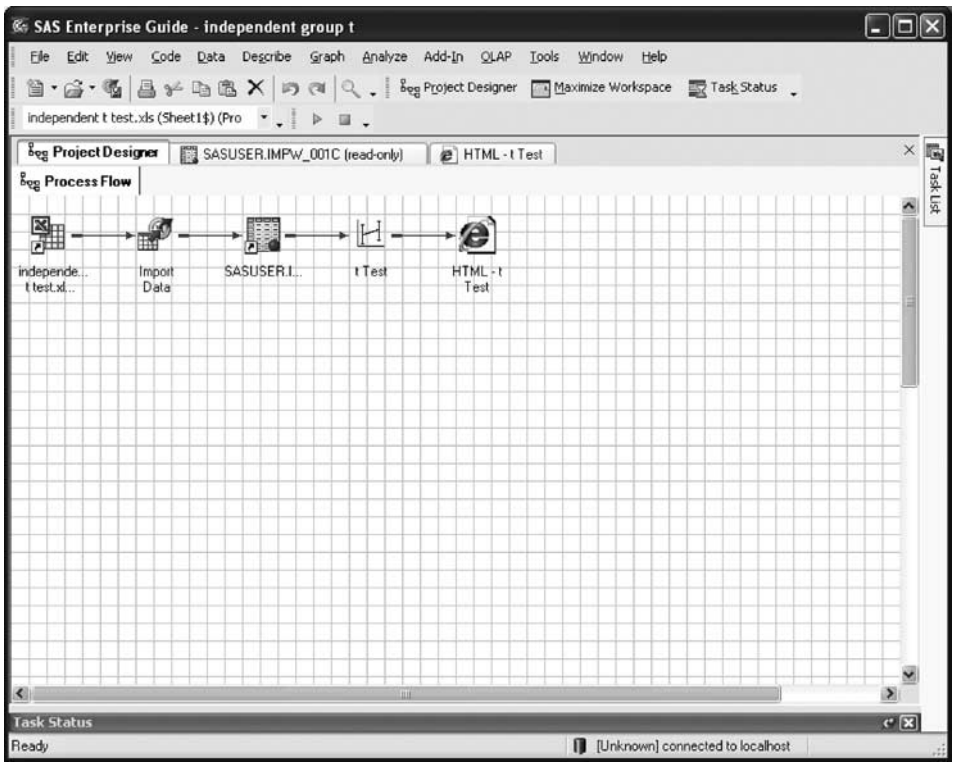

Figure 1.5. The **Process Flow** window for a project named **t test**.

navigate to the folder on our desktop containing our projects, and select **independent group *t***.

We have opened a project whose **Process Flow** screen is displayed in Figure 1.5. It is named **independent group t**, as can be seen in the Windows title bar at the top

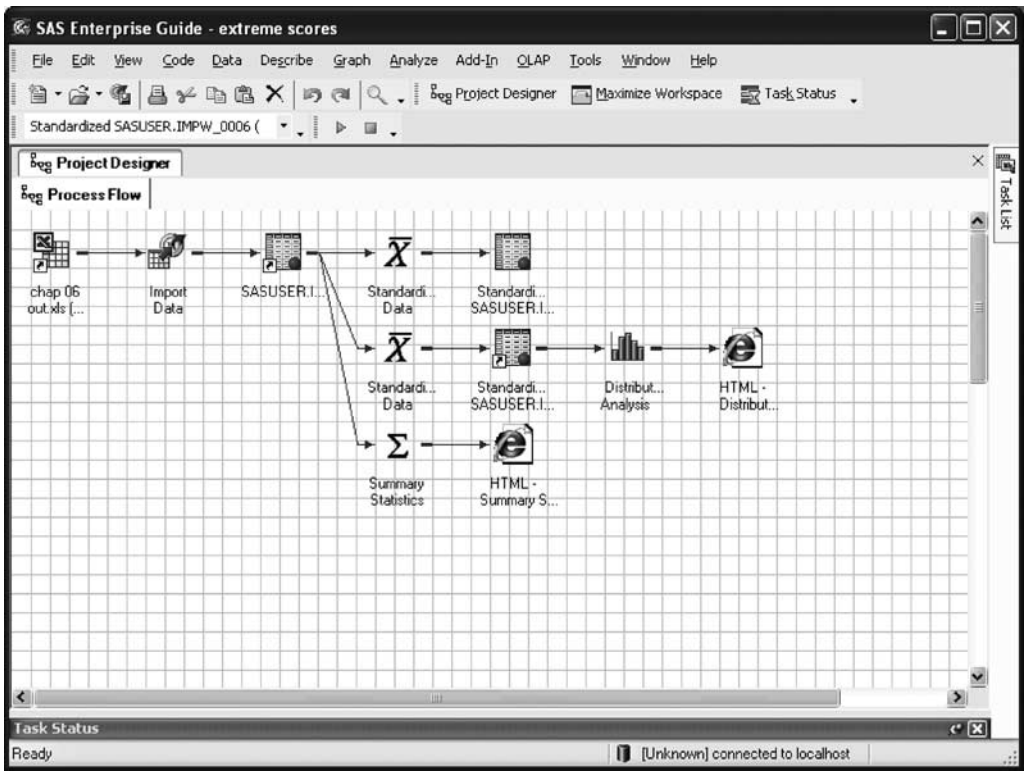

Figure 1.6. The **Process Flow** window for **Standardize**.

of the screen. **Process Flow** is a pictorial representation of the history of the project. Reading the icons from left to right unfolds the following story:

- The first icon represents an Excel file. At the time we began this project, the data were imported from an Excel file named **independent t test**.
- The second icon shows that the data in the Excel file were imported into *SAS Enterprise Guide*.
- The third icon stands for the *SAS Enterprise Guide* data set. The name **SASUSER** is read as "SAS user."
- The fourth icon represents the statistical analysis procedure **t test**.
- The fifth icon represents the output file. Results of a statistical procedure are placed in output files, which can have different formats. This output file is in HTML format, and this is how we display output in this book. We will talk more about this and other output formats in Chapter 4.

Multiple analyses can be performed and preserved in projects. Figure 1.6 displays the **Process Flow** screen for another project. The large X-bar symbol represents

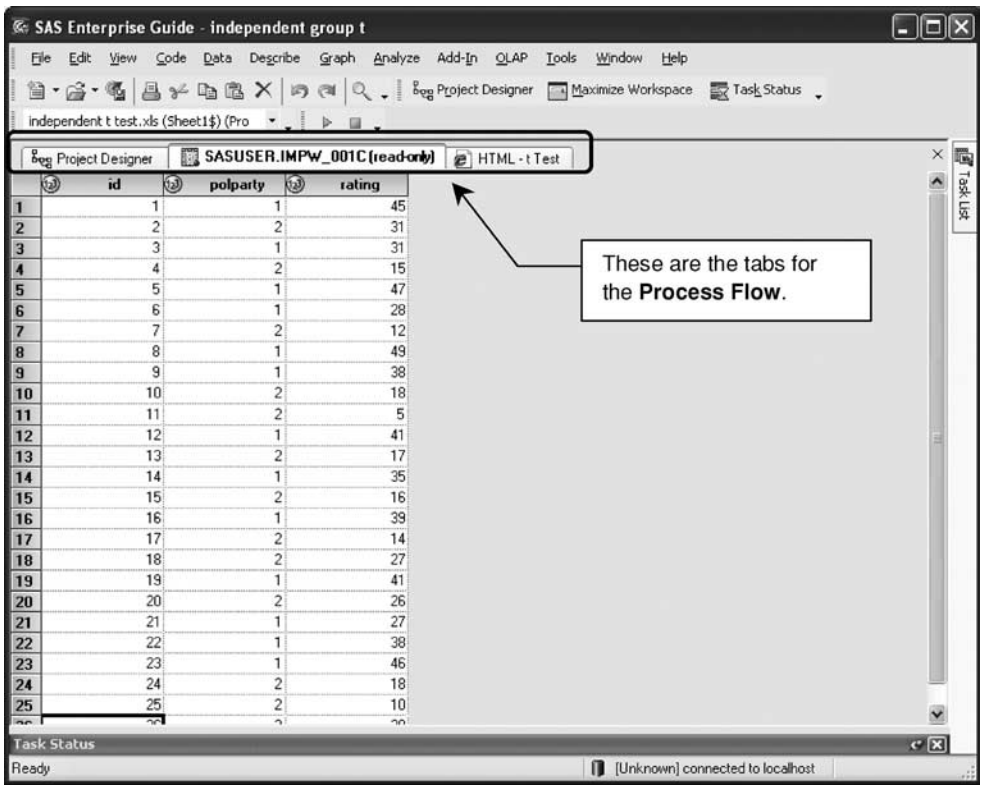

Figure 1.7. A view of the data set with the **Process Flow** tabs just above it.

standardization of a variable. As we can see, two standardizations and one **Summary Statistics** procedure (uppercase Greek sigma) have been performed; to picture this, a set of different arrows emerge from the data set on the first row.

1.4 Navigating tabs in the Process Flow screen

We return to **Process Flow** for the project named **independent group t** as shown before in Figure 1.5. By clicking on the icon for the data set, we can display it. This is shown in Figure 1.7. Each column is a variable.

Our interest for the moment is in the tabs just above the data set. The **Project Designer** tab is the one furthest to the left and is dimmed on the screen, indicating that it is not currently active. The **Project Designer** tab contains the pictorial representation of the project in the form of the **Process Flow** screen.

The active tab is labeled **SASUSER.IMPW_001C (read-only)**. It refers to the displayed data set. Here is what the parts of the label mean:

- As before, the name **SASUSER** is read as "SAS user."
- The expression **IMPW** indicates that the file was imported (**IMP**) from someplace unspecified in the label and that it is in the Working Library of SAS (**W**).
- The number **001C** is just a count of the work we have done during the current session.
- The expression **read-only** reminds us that to protect data sets from unintentionally being changed, they are opened in a protected or read-only mode. When we wish to modify the data set in some way, such as computing a new variable from the existent ones, it will be necessary to actively (and easily) turn off the read-only protection.

The tab furthest to the right is the output file named the **HTML t test**. By clicking on it we would open the output file.

Note that these tabs mirror the **Process Flow** screen and can be used to navigate between its elements directly. If the number of tabs exceeds the horizontal space allowed on the tab bar, scroll arrows will appear at the far right of the tab bar.

1.5 The main *SAS Enterprise Guide* menu

Figure 1.8 shows a portion of an existent *SAS Enterprise Guide* process flow. At the top of the window the main *SAS Enterprise Guide* menu (**File**, **Edit**, and so on) appears. You will make use of some of these menus much more frequently than others. When you click on one of these menu items, you will open a secondary menu from which you select what you would like to do. Very briefly, these menu items contain the following:

- **File:** Contains a variety of functions including **Open**, **Import Data**, **Print Preview** (the data set name will appear here), and **Exit**.
- **Edit:** Allows you to **Cut**, **Copy**, **Paste**, **Select All**, and so on.
- **View:** Controls **Toolbars**, **Task Status**, and so on.
- **Code:** Allows you to run the analysis that has been set up, stop the processing, and deal with macros.
- **Data:** Allows you to deal with the data set; among other things, you can select options to **sort** (reorder) the cases, **Transpose** the rows and columns, and **standardize** the data.

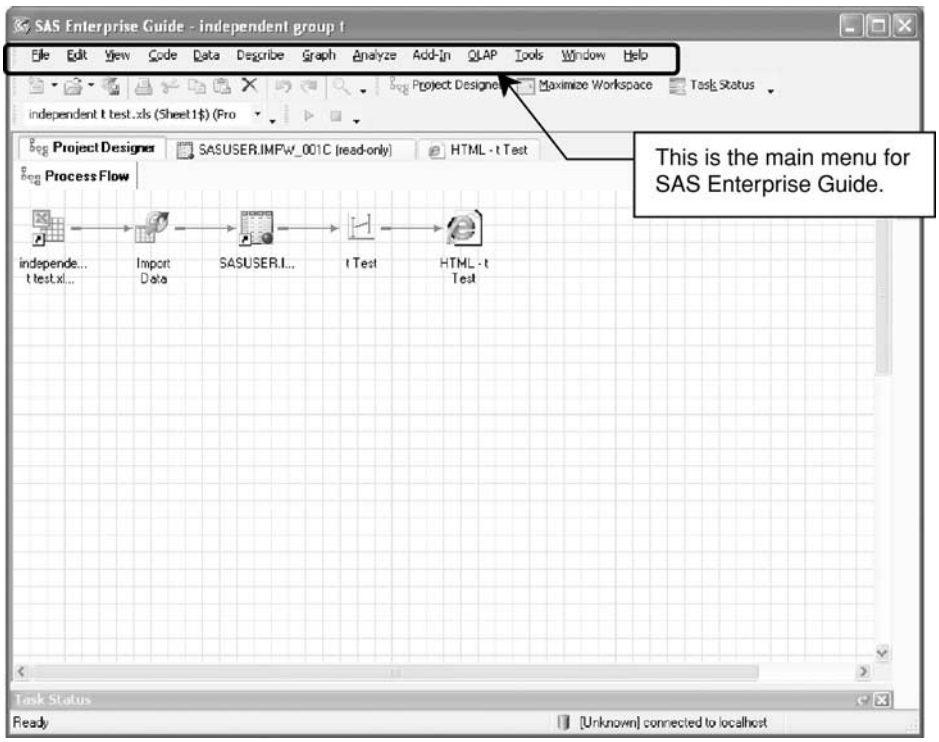

Figure 1.8. The *SAS Enterprise Guide* main menu appears at the top of the screen.

- **Describe:** Allows you to **List Data** (e.g., identify each case by variables that you designate), acquire **Summary Statistics**, and produce a **Frequency** table on a specified variable.
- **Graph:** Contains a variety of preformatted ways to plot your data.
- **Analyze:** Contains the statistical procedures you use to analyze your data.
- **Add-In:** Gets you to the **Add-In Manager**, which allows you to add, remove, and update commonly used procedures, such as **Standardize Data** and **Summary Statistics**.
- **OLAP:** This acronym stands for online analytical processing. According to the SAS Web site, the OLAP Server is a multidimensional data store designed to provide quick access to presummarized data generated from vast amounts of detailed data.
- **Tools:** Allows you to access sample data sets through **SAS Enterprise Guide Explorer**, place your project in a particular library through **Assign Library**, and produce your statistical output in HTML, PDF, RTF, and other formats through **Options**.