Chapter **5**

Opening Data Files

You don't need to put your data into the computer more than once. If you have your data in another program, you can import it from there into SPSS — because every program worth using has some form of output that can serve as input to SPSS. This chapter discusses ways to transfer data into and out of SPSS.

REMEMBER

Sometimes you won't get variable names when you import data from other software, but you can fix that problem in variable view.

Getting Acquainted with the SPSS File Format

SPSS uses its own format, the .sav extension, for storing data and writing files. The file format contains special codes and usually can't be used to export your data to another application. Instead, the file format is used only for saving SPSS data that you want to read back into SPSS at a later time.

The tremendous advantage of using SPSS data format is that it contains not only the cases and variables but also the metadata. For every other type of file that you bring into the software, you'll need to go back to the variable view to set up the metadata, as explained in Chapter 4.

Several example files in this format are copied to your computer as part of the normal SPSS installation. These files are in the same directory as your SPSS installation. To load one of these files, choose File ➪ Open ➪ Data and select the file to be loaded. When you do so, the variable names and data are loaded into SPSS.

If you have data in SPSS, you can save it as a .sav file by choosing File ➪ Save As and providing a name for the file. Or if you've loaded the information from a file, or previously saved a copy of the information to a file, simply choose File ➪ Save to overwrite the previous file with a fresh copy of both variable definitions and data.

WARNING

It's easy to be fooled by the way the SPSS documentation uses the word *file.* If you've defined data and variables in your program, SPSS documentation often refers to it all as a file, even though it may never have been written to disk. SPSS also refers to the material written to disk as a file, so watch the context.

When you write your file to disk, if you don't add the .sav extension to the filename; SPSS adds it for you. When you choose File ➪ Open ➪ Data to display the list of files, you may not see the filename extension (depending on how your Windows system is configured), but it's there.

TIP

When you drag a data file to the SPSS Data Editor window, SPSS chooses the appropriate format and begins reading the data file.

Reading Simple Data from a Text File

This section contains an example of a procedure you can follow to read data from a simple text file into SPSS. The example reads this text file and inserts it into the cells of SPSS. Along the way, SPSS keeps you informed about what's going on so there won't be any big surprises at the end.
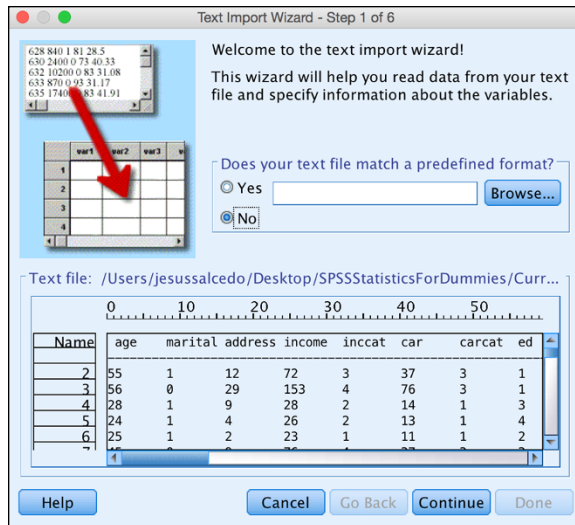
Follow these steps:

1. **Choose File ➪ Import Data ➪ Text Data.**

The file is in the SPSS installation directory.

2. **Select the demo.txt file, and then click Open.**

The Text Import Wizard appears, as shown in Figure 5-1, allowing you to load and format your data.

FIGURE 5-1:
Make sure your data looks reasonable.

3. **Examine the input data.**

   The screen lets you peek at the contents of the input file so you can verify that you've chosen the right file. Also, if your file uses a predefined format (it doesn't in this example), you can select it here and skip some of the later steps. If your data doesn't show up nicely separated into values the way you want, you may be able to correct it in a later step. Don't panic just yet.

4. **Click Continue.**

5. **Specify that the data is delimited and the names are included, as shown in Figure 5-2.**

   As you can see in this example, SPSS takes a guess, but you can also specify how your data is organized. It can be divided using spaces (as in this example), commas, tabs, semicolons, or some combination. Or your data may not be divided — it may be that all the data items are jammed together and each has a fixed width. If your text file includes the names of the variables, you need to tell SPSS.

6. **Click Continue.**

7. **Specify how SPSS should interpret the text.**

   For this example, the correct settings are shown in Figure 5-3. You can tell SPSS something about the file and which data you want to read.

   Perhaps some lines at the top of the file should be ignored — this happens when you're reading data from text intended for printing and header information is at the top. By telling SPSS about it, those first lines can be skipped.
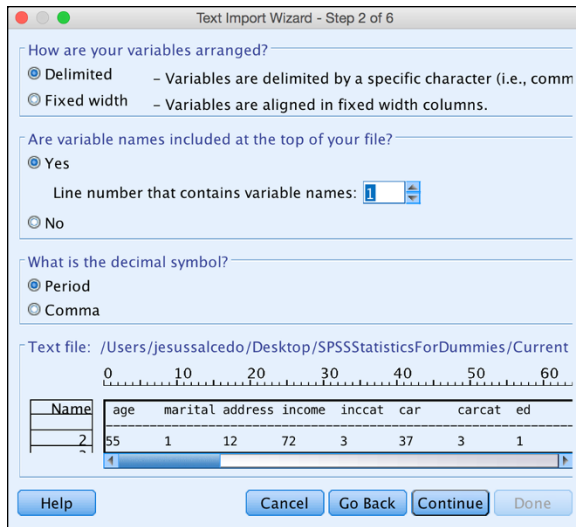
FIGURE 5-2:
Specify whether the fields are delimited and whether the variable names are included.
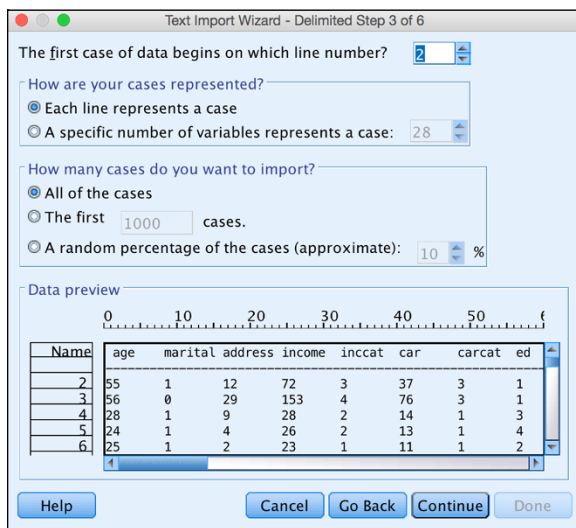
FIGURE 5-3:
Specify where the data appears in the file.

Also, you can have one line of text represent one case (one row of data in SPSS), or you can have SPSS count the variables to determine where each row starts.

And you don't have to read the entire file — you can select a maximum number of lines to read starting at the beginning of the file, or you can select a percentage of the total and have lines of text randomly selected throughout the file. Specifying a limited selection can be useful if you have a large file and would like to test parts of it.
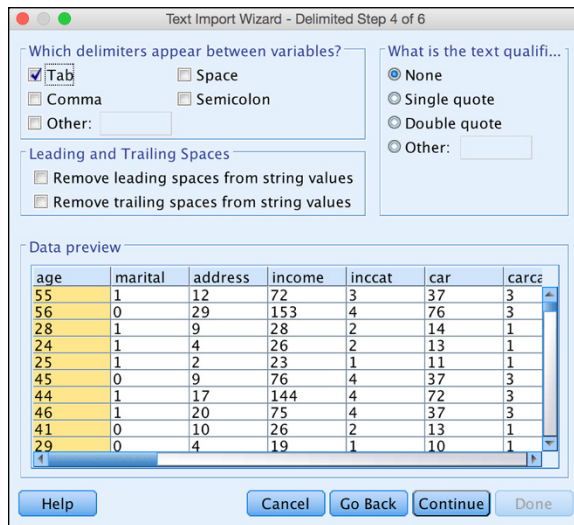
8. **Click Continue.**

9. **Specify tab as the delimiter, as shown in Figure 5-4.**

SPSS knows how to use commas, spaces, tabs, and semicolons as delimiting characters. You can even use some other character as a delimiter by selecting Other and then typing the character into the blank. You can also specify whether your text is formatted with quotes (which is common) and whether you use single or double quotes. Strings must be surrounded in quotes if they contain any of the characters being used as delimiters.

TIP

You can specify that a data item is missing in your text file. Simply use two delimiters in a row, without intervening data.
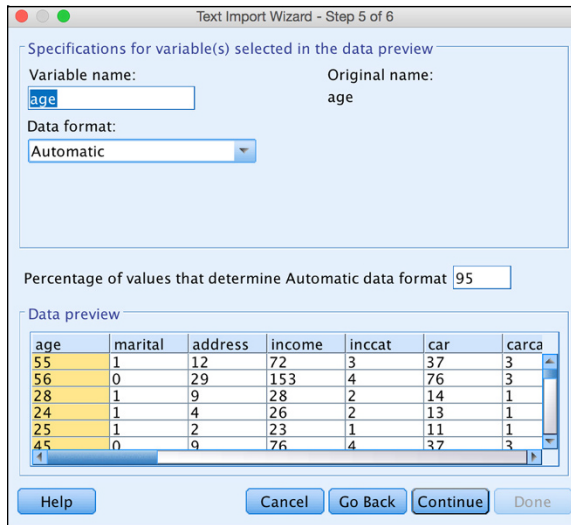

FIGURE 5-4:
Specify the delimiters that go between data items and which quotes to use for strings.

10. **Click Continue.**

11. **Change the variable name and data format (if needed), as shown in Figure 5-5.**

You can change the variable names and specify their types. Here you see that SPSS has made a guess for the type of each variable. If you did not have variables names, SPSS would assign the variables the names V1, V2, V3, and so on. To change a name, select it in the column heading at the bottom of the window, and then type the new name in the Variable Name field at the top. You can select the format from the Data Format drop-down list, if you need to change it.

FIGURE 5-5:
Name your variables and select their data types.

> **REMEMBER**
>
> This step is optional. If you need to refine your data types, you can do so later in the Variable View tab of the Data Editor window. The point here is to get the data into SPSS.

12. **Click Continue.**

13. **In the Would You Like to Save This File Format for Future Use? section, click No, as shown in Figure 5-6.**

Saving the file format for future use is something you would do if you were loading more files of this same format into SPSS — it reduces the number of questions to answer and the amount of formatting to do next time.

> **TECHNICAL STUFF**
>
> In the Would You Like to Paste the Syntax? section, you have the chance to grab a copy of the Syntax language instructions that do all these data manipulations. But unless you know about the Syntax language (as described in Chapters 26 and 27), it's best to pretend that this option doesn't exist for now.

14. **Click the Done button.**

Depending on the type of data conversions and the amount of formatting, SPSS may take a bit of time to finish. Be patient. The Data View tab of the Data Editor window will eventually display your data.
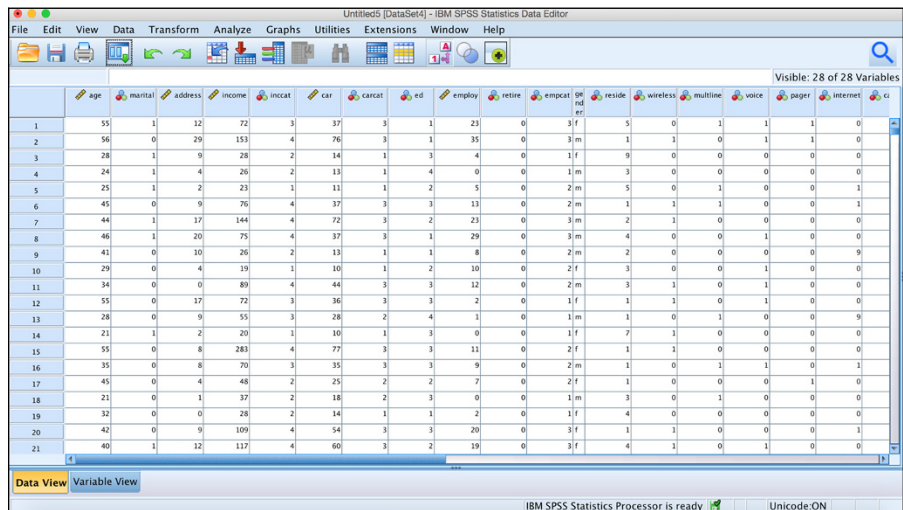
15. **Look at the data. Correct your data types and formats, if necessary. Then save it all to a file by choosing File ⇨ Save As.**

You're instructed to enter a file name. You can just call it **Demo**. The new file will have the .sav extension, which indicates that it's a standard SPSS file.

After you complete Step 6 of 6 in the wizard, the data appears in SPSS as shown in Figure 5-7.
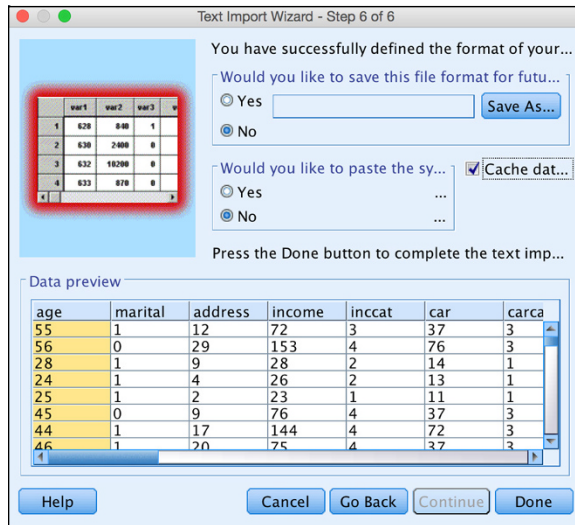
Transferring Data from Another Program

You can get your data into SPSS from a file created by another program, but it isn't always easy. SPSS knows how to read some file formats, but if you're not careful, you'll find your data stored in an odd file format. Deciphering some file formats can be as confusing as Klingon trigonometry. SPSS can read only from file formats it knows.

SPSS recognizes the file formats of several applications. Here's a complete list:

- **IBM SPSS Statistics (.sav):** IBM SPSS Statistics data, and also the format used by the DOS program SPSS/PC+

- **dBase (.dbf):** An interactive database system

- **Text (.txt, .csv, .dat, and .tab):** A standard text file that contains unformatted text

- **CSV (.csv):** A plain text file that contains a list of data

- **Microsoft Excel (.xls, .xlsx, .xlsm):** A spreadsheet for performing calculations on numbers in a grid

- **Portable (.por):** A portable format read and written by other versions of SPSS, including other operating systems

- **Lotus (.w):** A spreadsheet for performing calculations with numbers in a grid

- **SAS (.sas7bdat, .sd7, .sd2, .ssd01, ssd04, and .xpt):** Statistical analysis software

- **Stata (.dta):** Statistical analysis and graphics software

- **Sylk (.slk):** A symbolic link file format for transporting data from one application to another

- **Systat (.syd and .sys):** Software that produces statistical and graphical results

- **Database:** Any database with an ODBC connection

- **Cognos TM1:** A budgeting and forecasting program

- **Cognos Business Intelligence:** A web-based reporting tool

Although SPSS knows how to read any of these formats, you may still need to make a decision from time to time about how SPSS should import your dataset. But you have some advantages:

- You know exactly what you want — the form of data appearing in SPSS is simple, and what you see is what you get.

- SPSS has some reasonable defaults and makes some good guesses along the way.

- You can always fiddle with things after you've loaded the data.

REMEMBER

You're only reading from the data file, so you can't hurt it. Besides, you have everything safely backed up, don't you? If the process gets hopelessly balled up, you can always call it quits and start over. That's the way we do it — we think of it as a learning process.

Reading an Excel file

SPSS knows how to read Excel files directly. If you want to read data from an Excel file, we suggest you read the steps in "Reading Simple Data from a Text File," earlier in this chapter, because the two processes are similar. If you understand the decisions you have to make in reading a text file, reading from an Excel file will be duck soup.
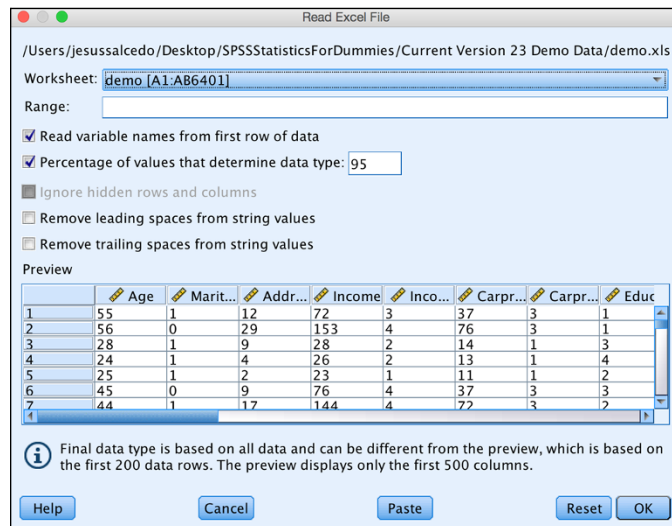
Do the following to read this data into SPSS:

1. **Choose File ⇨ Import Data ⇨ Excel.**

2. **Select the demo.xls file and then click Open.**

   The file is in the SPSS installation directory.

3. **Select the data to include.**

   An Excel file can contain more than one worksheet, and you can choose the one you want from the drop-down list, as shown in Figure 5-8.

FIGURE 5-8: Select which data in the spreadsheet to include.

Also, if you've elected to read only part of the data, enter the Excel cell numbers of the upper-left and lower-right corners here. You can specify if the names of variables appear in the first row and what percentage of data to use to determine the variable type. Finally you can remove any leading or trailing spaces.

TIP

It's a good idea to inspect the data preview to make sure the variables and data are read properly.

4. **Click OK.**

Your data appears in the SPSS window.

5. **Switch to the Variable View tab to examine the variable definitions and make any changes.**

SPSS makes a bunch of assumptions about your data, and some of those assumptions are probably wrong.

6. **Save the file using your chosen SPSS name, and you're off and running.**

Reading from an unknown program type

SPSS can import most file types, but there are exceptions. However, almost any file format in current use can be exported as a text file and then imported into SPSS.

REMEMBER

SPSS has been around for more than a half century, and organizations, such as the IRS, have hundreds of millions of rows of SPSS data. Very large files or older data-sets are almost always stored as text files. So if you're going to be an SPSS power user, you'll need to get comfortable with text files.

If your data is in an application that can't directly create a file of a type that SPSS can read, getting the data into SPSS may be easier than you think. If you can get the information out of your application and into a text file, it's fairly easy to have SPSS read the text file.

TECHNICAL STUFF

Years ago, it was not uncommon that an unusual proprietary format could not be exported to a text format but could be easily imported into SPSS. Although this situation is now rare, expert users of SPSS syntax can import even the most arcane data formats.

The data file you output to SPSS doesn't have to include the variable names, just the values that go into the variables. You can format the data in the file by using spaces, tabs, commas, or semicolons to separate data items. Such dividers are known as *delimiters.* Another method of formatting data avoids delimiters. In that method, you don't have to separate the individual data items, but you must make

each data item a specific length, because you have to tell SPSS exactly how long each item is.

The most intuitive format is to have one *case* (one row of data) per line of text. That way, the data items in your text file are in the same positions they'll be in when they're read into SPSS. Alternatively, you can have all your data formatted as one long stream, but you'll have to tell SPSS how many items go into each case.

REMEMBER Always save this kind of raw data as simple text; the file you store it in should have the .txt extension so SPSS can recognize it for what it is.

An even more roundabout way to bring in data from another application is by selecting, copying, and pasting the data you want, but we don't recommend doing this method. The places you're copying from and pasting to are usually larger than the screen, so highlighting and selecting can be tricky. You must be ready to choose Edit ⇨ Undo when necessary.

A better method is to write the data to a file in a format understood by SPSS, and then read that file into SPSS. SPSS knows how to read some file formats directly. Using such a file as an intermediary means you'll have an extra backup copy of your data, and that's never a bad idea.

TIP When you drag a data file to the SPSS Data Editor window, SPSS chooses the appropriate format and begins reading the data file.

Saving Data

Writing data from SPSS is easier than reading data into SPSS. All you do is choose File ⇨ Save As, select your file type, and then enter a file name.

You have lots of file types to choose from. You can write your data not only in 3 plain-text formats, but also in 3 Excel spreadsheet formats, 3 Lotus formats, 3 dBase formats, 9 SAS formats, 1 SYLK format, 1 Portable format, and 18 Stata formats.

TIP If you'll be exporting data from SPSS into another application, find out what kinds of files the other application can read, and then use SPSS to write in one of those formats.