

PART

1

**FUNDAMENTALS
OF DATA
MANAGEMENT**

CHAPTER

1

INTRODUCTION TO SPSS

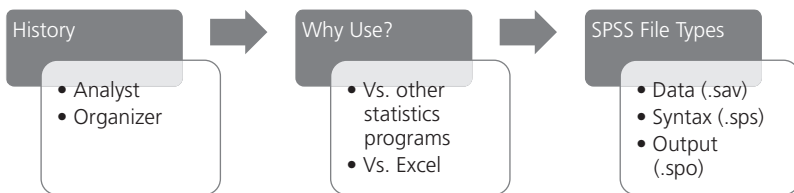


FIGURE 1.1. *History – Why Use? – SPSS File Types.*

Contents

Chapter Learning Objectives

What SPSS Is Used For

SPSS Versus Excel and Others

How the Program Works

Important File Types

The Other File Types

Summary

Key Terms

Discussion Questions

CHAPTER LEARNING OBJECTIVES

1. History of SPSS
2. SPSS advantages (over other programs)
3. Types of files used by SPSS

SPSS is a software program that was created to perform statistical analyses. SPSS is also the name of the company that owns and sells the computer program. The name SPSS is an acronym for Statistical Package for the Social Sciences. Because SPSS is used across many disciplines, the old “social sciences” label is without doubt too narrow for the program’s current market.

SPSS has two primary functions: (1) as a data analyst, and (2) as a data organizer and manager. Although we cover both, the primary focus of this book is on the second function, because although the program itself is a powerful data analyst, you will spend most of your time using the program’s data organizing and managing functions. You need to organize and manage your data *so that you can eventually do analyses with it*—real world data rarely comes in a form that’s immediately ready to analyze.

Like most software programs, SPSS is continually being updated. The current version (15.0 as this book is being written) is quite different from the original and the earlier versions, primarily because those did not have a Microsoft Windows–based interface. The early versions were mainframe and DOS-based programs, which required users to *program* (enter code language) in a less visually appealing and noninteractive computer environment.

In addition to those platform changes, the field of statistics changes over time (improvements and slight alterations are generally made) and new statistics are constantly being discovered. To stay current with statistical advancements and platform limitations and changes (such as

changes in Microsoft Windows), the SPSS program is frequently updated. The information in this book is valid for any version of SPSS. It is written from a PC perspective, but the information is applicable to a Mac audience.

WHAT SPSS IS USED FOR

SPSS has the capability to perform a wide variety of statistical analyses, both *descriptive* and *inferential*. It is used by researchers (biologists, psychologists, sociologists, economists) and public and private sector workers (accountants, human resources professionals, actuaries). These individuals use SPSS to test hypotheses in experimental and field settings, summarize information, and create graphs and figures.

The primary power of SPSS lies in its wide variety of statistical options and its ability to perform these analyses quickly. There are other software programs available that do a nice job of, for example, creating graphs and figures, but SPSS creates graphs and figures *in addition to* its primary purpose of processing data using appropriate statistical formulae.

SPSS Versus Excel and Others

SPSS was designed specifically for the purpose of performing statistical analyses. There are other software programs that perform statistics as their primary purpose (such as SAS, Minitab, and BMDP). There are advantages and disadvantages to each software package; the choice of SPSS over the other statistics-oriented programs was likely made by your class instructor or supervisor. If the names of the other titles look a bit more appealing to you than “SPSS,” you should drop your statistics class and sell this book now, because once you choose one software package, you will generally stick with it.

In addition to statistical processing competitors, SPSS is somewhat crowded upon by spreadsheet programs that have statistical potential. These programs (such as Microsoft Excel) are generally fine—and usually more flexible—for descriptive statistical analyses (statistics that *describe* a given set of numbers: means, standard deviations, correlations), but they are usually less useful if your goal is to perform inferential statistical analyses, such as t-test, ANOVA, or regression—statistics that make an *inference* about a larger group of numbers, given the associations in your small set of numbers (this will be further discussed in Ch. 10–12). If you are doing inferential statistics, or are very concerned

about minimizing error in your processes, you will want to use SPSS instead of Excel or other spreadsheet programs. Spreadsheets are great for organizing data and doing minor computations and manipulations (as well as summarizing information graphically), but if you are doing hypothesis-testing statistics, SPSS is the way to go.

Over the years SPSS has actually slowly grown to look pretty similar (in interface appearance) to Excel. As will be described in Chapter Two, this wasn't always the case. You can think of SPSS as a combination of Word and Excel (to cite the most commonly used word processor and spreadsheet applications). SPSS is conceptually a combination of these two applications, as data is stored in SPSS's Excel-like spreadsheet component and analyses are performed by writing commands in SPSS's Word-similar word processing component.

HOW THE PROGRAM WORKS

When you access the program, by selecting the SPSS icon from your desktop or Start menu, one of two things will happen. First, a Help dialog box may appear that asks "What would you like to do?" There is an option at the bottom of this box: "Don't show this dialog in the future." Go ahead and select that option (you don't want to see this box every time you open SPSS; that would get somewhat annoying). Second, an empty data file may open. This is a good thing—it's what you want to occur.

Important File Types

There are three very important types of files used by SPSS: the aforementioned data file type (identified by an .sav file extension), a syntax file type (identified by an .sps file extension), and an output file type (identified by an .spo file extension).

Data Files. *Data (.sav)* files are where you store your numbers and data. Typically you have one *original* data file for each major project that you are working on. Figure 1.2 shows what an empty .sav file (containing no data) looks like.

Notice that the empty data file has (1) a grid of cells (seventy-two cells and part of another twelve are visible in the screen capture), (2) commands and toolbar icons (including the File, Edit, and View commands familiar in Microsoft applications, but also Transform, Analyze, Graphs, and more), and (3) spreadsheet tabs (here labeled Data View and Variable View, which are unique to SPSS and not a shared feature of

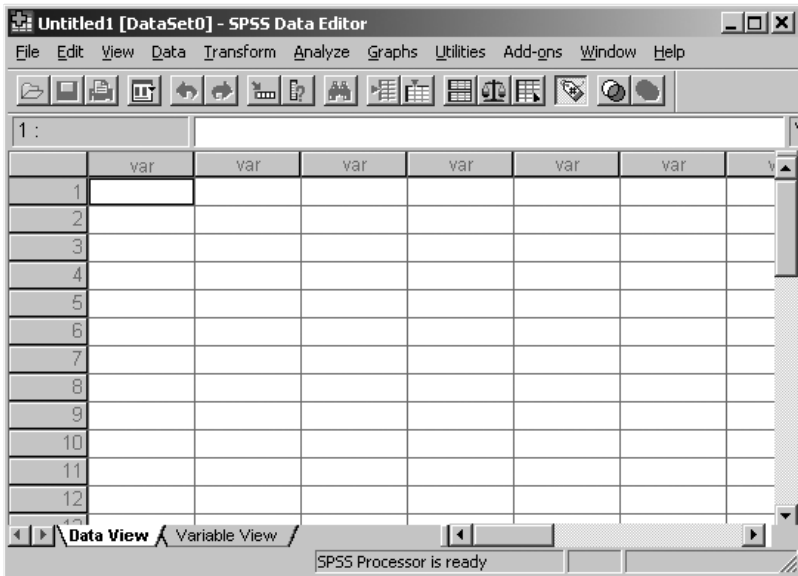


FIGURE 1.2. *The SPSS Data Editor.*

Excel). [More will be said about data files later; your take-away message here should be “Ah, that’s where I put my data; numbers go in there somehow.”

Syntax Files. *Syntax* (.*sps*) files are where you write your syntax diary (of where your data came from, and what analyses you’d like to do, and so on). Conceptualizing your .*sps* files as a diary constitutes one of the most important elements of this book. Chapter 2 will highlight some lessons for this importance. Figure 1.3 shows an empty syntax file (with no diary entries or statistical commands).

Just as you put your numbers in your data file, you identify what analyses you’d like to do in your syntax file. Notice that: (1) the commands are similar to the commands of the .*sav* file (with the exception of a Run command included in the .*sps* file), (2) there is a blank white space instead of a grid (you actually type entries, just as you would in a word processor), and (3) there are no tabs at the bottom of the window (that is, no Data View or Variable View).

These two file types (data [.*sav*] and syntax [.*sps*]) are the ones that you as an SPSS user work with. You must enter information into both the .*sav* file and the .*sps* file for each project that you do.

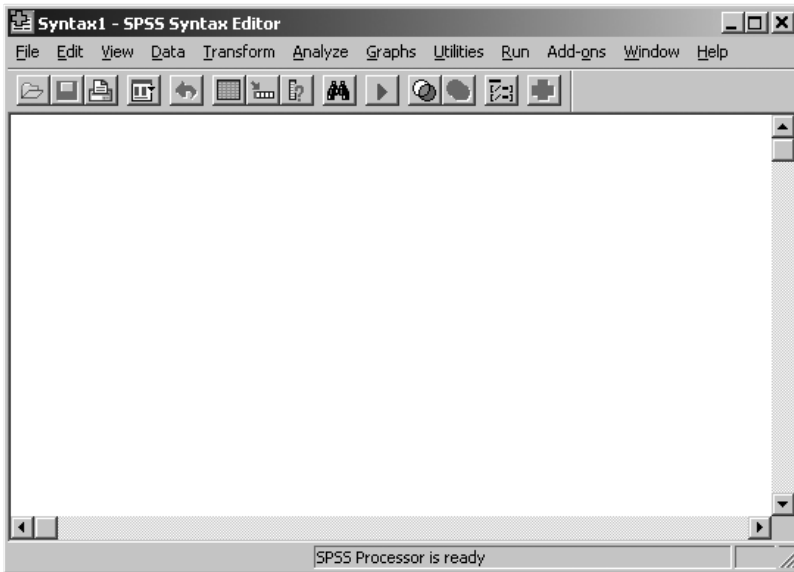


FIGURE 1.3. *The SPSS Syntax Editor.*

Output Files. *Output (.spo)* files are different from data and syntax files. *You* do not create output files—these are created by the SPSS program. If you have a syntax file that is full (with writing in it, instead of a big blank space as in the Figure 1.3 screen capture) and you have a data file that has numbers in it (instead of the blank white cells in Figure 1.2), the result of the interaction of the .sav and the .sps is an output (.spo) file. Figure 1.4 shows a blank output file.

You will rarely see an empty output file like this one, because output files are typically generated, not conjured up as this one has been. When you perform an analysis or a data manipulation—whenever something is done to your data, an output file appears. Output files are also like SPSS security enforcers—any time anything illegal or otherwise noteworthy is done to data, an output file will pop up and report the incident.

Output files have two panes: the narrower left pane and the wider right pane. The right-hand pane presents all of the detailed output you request from your selected analyses. The left pane is merely an outline of the detailed information provided in the right-hand pane. This outline is

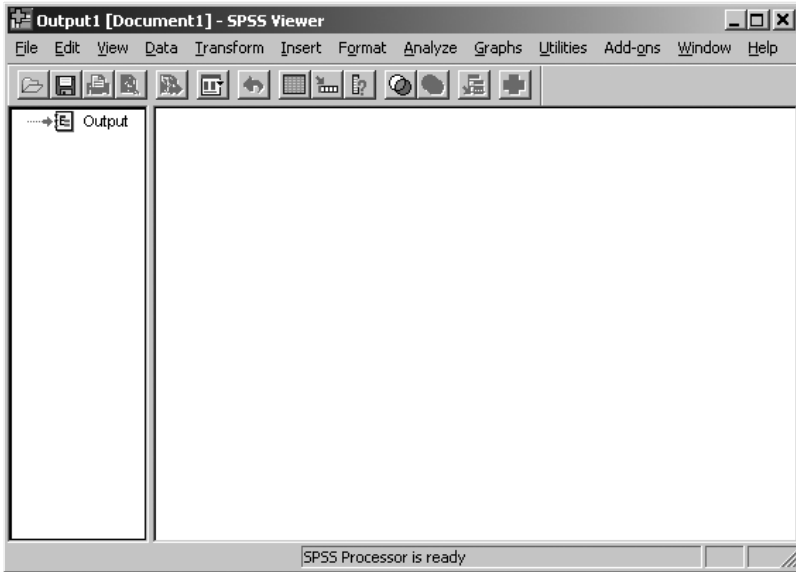


FIGURE 1.4. *The SPSS Output Viewer.*

sometimes useful to help you navigate through the detailed output on the right if you have performed a lot of analyses.

The Other File Types

There are other SPSS file types you will likely encounter, but the big three that you should be concerned with are the data (.sav), syntax (.sps), and output (.spo) files. If after reading this book you decide that you like SPSS enough (or are going to use it enough) to learn more, you can pursue learning about the program's other capabilities.

SUMMARY

SPSS is used not only to perform data analyses but also used to organize and manipulate data. SPSS is considered superior to data spreadsheet programs because with SPSS you can *keep a record* of what you have done. This is done through keeping a *syntax diary*. The program works through the interaction of a data file (.sav) and a syntax file (.sps). When the numbers (data in a .sav file) and instructions (syntax in an .sps file) interact, the result is an output file (.spo).

KEY TERMS

Data Anything informative.

Data (.sav) files Where data is stored in SPSS.

Descriptive statistics Information summarizing a set of numbers.

Excel Computer spreadsheet application.

Inferential statistics Probability-based information relating sample to population characteristics.

Output (.spo) files Where results are reported in SPSS.

SPSS Computer program used to do data manipulations and analyses.

Syntax (.sps) files Where operations are specified in SPSS.

DISCUSSION QUESTIONS

1. Why is SPSS superior to Excel?
2. What are some advantages and disadvantages associated with SPSS's evolution toward an Excel-Word hybrid?