

## Chapter 1

# Introducing SPSS

“There are three kinds of lies: lies, damn lies, and statistics.” That statement is often attributed to Mark Twain, but that’s not quite right. Mark Twain did say it, but he attributed it to someone else. He indirectly attributed it to Disraeli, but his attribution was vague, and the original statement, if it exists, can’t be located. Speaking statistically, the odds are in favor of us never knowing who said it first.

## *Garbage In, Garbage Out*

Statistical analysis is like a sewer. What you get out of it largely depends on what you put into it.

Over 82 percent of all statistics are made up on the spot to try to prove a point.

You can conclude just about anything if you’re not careful with your data and with your calculations. SPSS watches the performance of the calculations for you, but the raw data, and which calculations should be performed, is up to you.

Let me show you a simple example of using raw data to produce an obviously wrong conclusion. Suppose you want to demonstrate, by sampling, that every odd number is prime. (A prime number can be evenly divided only by 1 and itself.) The first thing to do is gather a collection of data points, as shown in Table 1-1.

<i>Number</i>	<i>Prime?</i>	<i>Comment</i>
1	Yes	It fits the definition exactly
3	Yes	It is certainly both odd and prime
5	Yes	It fits the pattern of primes

*(continued)*

<i>Number</i>	<i>Prime?</i>	<i>Comment</i>
7	Yes	So far, so good
9	No	Must be a bad data point, so throw it out
11	Yes	Now we're back on track
13	Yes	Looking good

Lots of things are already wrong with the data in Table 1-1. For one, the sample is too small. For another, the sampling cannot be considered random. All too often it happens that data points don't fit a preconceived conclusion, so they are omitted. The result of the data in this table can be used as proof of a fact that is dead wrong.

This book is not about the accuracy, correctness, or completeness of the input data. Your data is up to you. This book shows you how to take the numbers you already have, put them into SPSS, crunch them, and display the results so it all makes sense. Gathering valid data and figuring out which crunch to use is up to you.

## *From Whence SPSS?*

SPSS is probably older than you are. In 2007 it becomes 38 years old, and the average age of an American is 35.3.

At Stanford University in the late 1960s, Norman H. Nie, C. Hadlai (Tex) Hull, and Dale H. Bent developed the original software system named Statistical Package for the Social Sciences (SPSS). They needed to analyze a large volume of social science data, so they wrote software to do it. The software package caught on with other folks at universities and, with the open source tradition of the day, the software spread through universities around the country.

The three men produced a manual in the 1970s and the software's popularity took off. A version of it existed for each of the different kinds of mainframe computers of the time. Its popularity spread from universities into other areas of government, and it began to leak out into private enterprise.

In the 1980s, a version of the software was moved to the personal computer, and here we are today.

Maybe it has been continuously successful because the software does such a good job of making predictions, and the SPSS people could always figure out what they should do next.

## The Four Ways to Talk to SPSS

More than one way exists for you to command SPSS to do your bidding. And you don't have to choose one and stick with it — you can perform tasks using whichever of the four interfaces you prefer. You can use any of the four approaches to perform any of the SPSS functions, but which one is best for you depends, to an extent, on the task to be performed and which interface you prefer:

- ✓ **GUI (graphic user interface):** SPSS has a windowing interface and commands can be issued by the mouse through menu selections that cause dialog boxes to appear. This is a fill-in-the-blanks approach to statistical analysis that guides you through the process of making choices and selecting values. The advantage of the GUI approach is that, at each step, SPSS will make sure that you enter everything necessary before proceeding to the next step. This is the preferred interface for those just starting out — and if you don't do much with SPSS, this may be the only interface you ever use.
- ✓ **Syntax:** This is the internal language used to command actions from SPSS. It was known as the command syntax of SPSS, hence its name. It is often referred to as the command language. You can write Syntax commands to directly command SPSS to do anything it is capable of doing. In fact, when you use menu and dialog box selections to command SPSS, you are actually generating Syntax commands internally that do your bidding. That is, the GUI is nothing more than the front end of a Syntax command-writing utility. Writing (and saving) command language programs is a good way to store processes that you expect to repeat. You can even grab a copy of the Syntax commands generated by the GUI and save them to be repeated later.
- ✓ **Python:** This is a general-purpose language that has a collection of SPSS modules written for it, making it possible to write programs that work inside SPSS. It can be run with the Syntax language to command SPSS to perform statistical functions. One advantage of using Python is the fact that it is a modern language and gives you the power and convenience that come with languages today, including the ability to construct a more readable program. In addition, because it's a general-purpose language, you can read and write data from other applications and from other files.
- ✓ **Scripts:** The items that SPSS calls scripts are actually programs written in BASIC. This language is simple and many people are familiar with it. Also, a BASIC program can be written as an *autoscript* — a script that executes automatically when SPSS produces certain output.

## *The Things You Can and Cannot Do with SPSS*

The full-blown SPSS package comes in many parts. The Base system is the center around which the rest of SPSS revolves. You have a Base system. You may also have one or more add-ons. With only one exception, everything described in this book is included in the Base system, so you will be able to do anything you read about. The one exception is the Python programming language, which requires some additional software. But the software is a free download and also comes on the SPSS distribution CD. Chapter 20 describes other modules you can add to your Base system.

SPSS works with numbers. Only. If you cannot express your information as a number, you can't run it through SPSS. You will see names and descriptions seemingly being processed by SPSS, but that's because each name has been assigned a number. That's why survey questions are written like, "How much do you enjoy eating rhubarb? Select your answer: Very much, sort of, don't care, not really, I hate the stuff." A number is assigned to each of the possible answers, and these numbers are fed through the statistical process. SPSS uses the numbers, not the words, so be careful about keeping all your words and numbers straight.

You must keep accurate records describing your data, how you got the data, and what it means. SPSS can do all the calculations for you, but only you can decipher what it means. In *Hitchhiker's Guide to the Galaxy*, a computer the size of a planet crunched on a problem for generations and finally came out with the answer, 42. But the people tending the machine had no idea what the answer meant because they didn't remember the question. They hadn't kept track of their input. You must keep careful track of your data or you may later discover, for example, that what you have interpreted to be a simple increase is actually an increase in your rate of decrease. Oops.

SPSS lets you enter the data and tag it to help keep it organized, but you already have the data written down someplace and fully annotated. Don't you?

## *How SPSS Works*

The developers of SPSS have made every effort to make the software easy to use. This prevents you from making mistakes or even forgetting something. That's not to say it's not possible to do something wrong, but the SPSS software works hard to keep you from running into the ditch. To foul things up, you almost have to work at figuring out a way of doing something wrong.

You always begin by defining a set of *variables*, then you enter data for the variables to create a number of *cases*. For example, if you are doing an analysis of automobiles, each car in your study would be a case. The variables that define the cases could be things such as the year of manufacture, horsepower, and cubic inches of displacement. Each car in the study is defined as a single case, and each case is defined as a set of values assigned to the collection of variables. Every case has a value for each variable. (Well, you can have a missing value, but that's a special situation described later.)

Variables have types. That is, each variable is defined as containing a specific kind of number. For example, a *scale* variable is a numeric measurement, such as weight or miles per gallon. A *categorical* variable contains values that define a category; for example, a variable named `gender` could be a categorical variable defined to contain only values 1 for female and 2 for male. Things that make sense for one type of variable don't necessarily make sense for another. For example, it makes sense to calculate the average miles per gallon, but not the average gender.

After your data is entered into SPSS — your cases are all defined by values stored in the variables — you can run an analysis. You have already finished the hard part. Running an analysis on the data is much easier than entering the data. To run an analysis, you select the one you want to run from the menu, select appropriate variables, and click the OK button. SPSS reads through all your cases, performs the analysis, and presents you with the output.

You can instruct SPSS to draw graphs and charts the same way you instruct it to do an analysis. You select the desired graph from the menu, assign variables to it, and click OK.

When preparing SPSS to run an analysis or draw a graph, the OK button is unavailable until you have made all the choices necessary to produce output. Not only does SPSS require that you select a sufficient number of variables to produce output, it also requires that you choose the right kinds of variables. If a categorical variable is required for a certain slot, SPSS will not allow you to choose any other kind. Whether the output makes sense is up to you and your data, but SPSS makes certain that the choices you make can be used to produce some kind of result.

All output from SPSS goes to the same place — a dialog box named SPSS Viewer. It opens to display the results of whatever you've done. After you have output, if you perform some action that produces more output, the new output is displayed in the same dialog box. And almost anything you do produces output.

## All the Strange Words

Statistics seems to have been born in the land of strange words. Lots of them. If you come across a term that you don't understand, such as *dichotomy*, *variable*, or *kurtosis*, you can look it up in the glossary at the back of the book.

It's not only new words that can trip you up. You will find common words used in a special way. For example, a *break variable* has a special purpose when organizing tabular data.



The glossary is always there, ready to explain the meaning of those strange terms.

## All Those Files

Input data and statistics are stored in files. Different kinds of files. Some files contain numbers and definitions of numbers. Some files contain graphics. Some files contain both.

The examples in this book require the use of files that contain data configured to demonstrate capabilities of SPSS. These files are all in one of two places. Most are in the same directory you use to install SPSS. That is, the action of installing SPSS also installs a number of data files ready to be loaded into SPSS and used for analysis. A few of the files used in the examples can be found in the compressed file `spss.zip` found at this Web site:

```
www.dummies.com/go/spss
```

You can also get the files from the author's Web site:

```
www.belugalake.com/spss
```

After you have downloaded the zip file to your system, you need to decompress (unzip) it into separate files and directories. If you don't have an unzipper, and would like to get one, enter the search word *unzip* into Google. There are free ones and commercial ones. A popular commercial product that runs as a windowing program and uses mouse controls can be found at the following Web site:

```
http://www.winzip.com
```

If you don't want to buy WinZip, you can download a free trial version that will work just fine for this job.

If you want, the Web site is configured so that you can download the files in the form ready to be used. Doing it this way is a bit easier, but only if you don't need all the files. Each file will have to be downloaded individually.

## Where to Get Help When You Need It

You're not alone. Some immediate help comes directly from the SPSS software package, and other help can be found on the Internet. If you find yourself stumped on some point, you can look in several places:



- ✓ **Topics:** Choosing Help⇒Topics from the main window of the SPSS application is your gateway to immediate help. The help is somewhat terse, but it will often be exactly what you need. You will find all the information in one large help document, presented to you one page at a time. Choose Contents to select a heading from an extensive table of contents, choose Index to search for a heading by entering its name, or choose Search to enter a string search inside the body of the help text.

In the help directory, the titles in all uppercase are descriptions of Syntax language commands.

- ✓ **Tutorial:** Choose Help⇒Tutorial to open a dialog box with the outline of a tutorial that guides you through many parts of SPSS. You can start at the beginning and view each lesson in turn, or you can select your subject and view just that.
- ✓ **Case Studies:** Choose Help⇒Case Studies to open a dialog box containing examples in a format similar to that of the Tutorial selection. You can select titles from its outline and view descriptions and examples of specific instances of using SPSS. You will also find descriptions of the different types of calculations. This is a good place to look if some particular analysis type is eluding your comprehension.
- ✓ **Statistics Coach:** Choose Help⇒Statistics Coach if you have a good idea of what you want to do but need some specific information on how to go about doing it.
- ✓ **Command Syntax Reference:** Choose Help⇒Command Syntax Reference to display more than 2000 pages of references to the Syntax language in your PDF viewer. The regular help topics, mentioned previously, provide a brief overview of each topic, but this document is much more detailed.
- ✓ **Python:** Choose Help⇒Programmability to display a 100-page PDF document on programming SPSS using Python.

## *Your Most Valuable Possession*

The most valuable possession you have in dealing with statistics is not your computer. It's not your SPSS software. It's not this book, or any other book you may be using to learn statistics. You can lose any one of those, but any one of them can be replaced.

Your most valuable possession is your data. Sure, you can always go and get more data, but you can't go and get the same data. The world doesn't hold still long enough. Make sure you make backup copies of your data.



Back up your data to memory that does not live in the same building with the computer you are using. You can swap backups with a friend, or if you have access to a remote Web site, you can stuff files in a blind directory.

This message about backing up your data comes to you from someone who has been stung. Twice. And I don't want to talk about it again. Ever.

## *You Can Dive As Deep As You Want to Go*

SPSS makes no effort to keep anything a secret. It is designed to be as easy to use as possible, so you really don't have to know that much to make it work. However, if you want to understand how things are working internally, you can find out if you dig. And you don't have to dig very far. Choosing Help is the first step to finding out anything you want to know about what's going on inside.

Let's say you are working on your numbers and want to use some specific algorithm to do your calculations. SPSS has been at this longer than you have, so the algorithm you want to use is almost certainly built in. If you are not sure exactly what SPSS is doing to calculate some of the numbers, you can go to the Help menu and the PDF documents on the documentation CD to find out how the calculations are being performed. But, before you start looking, make sure you really want to know because the equations and how they are applied are explained in excruciating detail.

The purpose of this book is to give the shallow divers enough information to be able to swim, and show the deeper divers how to begin. I don't explain all the details because there are too many. There's simply not enough room in a book this size to explain SPSS in depth.