

1

Statistics or Sadistics?

It's Up to You

Difficulty Scale ☺ ☺ ☺ ☺ ☺ (really easy)

What you'll learn about in this chapter

- What statistics is all about
- Why you should take statistics
- How to succeed in this course

WHY STATISTICS?

You've heard it all before, right? "Statistics is difficult," "The math involved is impossible," "I don't know how to use a computer," "What do I need this stuff for?" "What do I do next?" and the famous cry of the introductory statistics student, "I don't get it!" Well, relax. Students who study introductory statistics find themselves, at one time or another, thinking about at least one of the above, if not actually sharing it with another student, their spouse, a colleague, or a friend.

And all kidding aside, there are some statistics courses that can easily be described as *sadistics*. That's because the books are repetitiously boring and the authors have no imagination.

That's not the case for you. The fact that you or your instructor has selected *Statistics for People Who (Think They) Hate Statistics, Excel 2010 Edition* shows that you're ready to take the right approach—one that is unintimidating, informative, and applied (and even a little fun) and that tries to teach you what you need to know about using statistics as the valuable tool that it is.

If you're using this book in a class, it also means that your instructor is clearly on your side—he or she knows that statistics can be intimidating but has taken steps to see that it is not intimidating for you. As a matter of fact, we'll bet there's a good chance (as hard as it may be to believe) that you'll be enjoying this class in just a few short weeks.

And Why Excel?

Simple. It's the most popular, most powerful spreadsheet tool available today, and it can be an exceedingly important and valuable tool in learning how to use statistics. In fact, many stats courses taught at the introductory level use Excel as their primary computational tool and ignore other computer programs, such as SPSS and MiniTab. Although we are not going to teach you how to use Excel (see Appendix A for a refresher on some basic tasks), we will show you how to use it to make your statistics learning experience a better one.

But like any program that takes numbers and consolidates and analyzes them, Excel is not a magic bullet or a tool to solve all your problems. It has its limitations. Unless you are an expert programmer and you can program Excel to do just about anything other statistics programs can (and the language you would use is called Visual Basic Applications or VBA), Excel may not look as pretty as other programs dedicated to statistical analysis or offer as many of the same options. But at the level of introductory statistics, it is a very powerful tool that can do an awful lot of very neat things.

A bit of terminology about Excel before we move on. The first ever Excel-like computer application was called Visicalc (thank you, Dan Bricklin and Bob Frankston) and was known as a spreadsheet. Okay—the Excel application is known as a spreadsheet program as well, but each individual sheet is known as a **worksheet**. And worksheets, when combined, constitute what is known as a **workbook**. Fun, huh?

A FIVE-MINUTE HISTORY OF STATISTICS

Before you read any further, it would be useful to have some historical perspective about this topic called statistics. After all, almost every undergraduate in the social, behavioral, and biological sciences and every graduate student in education, nursing, psychology,

social welfare and social services, and anthropology (you get the picture) is required to take this course. Wouldn't it be nice to have some idea from whence the topic it covers came? Of course it would.

Way, way back, as soon as humans realized that counting was a good idea (as in "How many of these do you need to trade for one of those?"), collecting information became a useful skill.

If counting counted, then one would know how many times the sun would rise in one season, how much food was needed to last the winter, and what amount of resources belonged to whom.

That was just the beginning. Once numbers became part of language, it seemed like the next step was to attach these numbers to outcomes. That started in earnest during the 17th century, when the first set of data pertaining to populations was collected. From that point on, scientists (mostly mathematicians, but then physical and biological scientists) needed to develop specific tools to answer specific questions. For example, Francis Galton (a half-cousin of Charles Darwin, by the way), who lived from 1822 to 1911, was very interested in the nature of human intelligence. To explore one of his primary questions regarding the similarity of intelligence among family members, he used a specific statistical tool called the correlation coefficient (first developed by mathematicians), and then he popularized its use in the behavioral and social sciences.

You'll learn all about this tool in Chapter 5. In fact, most of the basic statistical procedures that you will learn about were first developed and used in the fields of agriculture, astronomy, and even politics. Their application to human behavior came much later.

The past 100 years have seen great strides in the invention of new ways to use old ideas. The simplest test for examining the differences between the averages of two groups was first advanced during the early 20th century. Techniques that build on this idea were offered decades later and have been greatly refined. And the introduction of personal computers and such programs as Excel has opened up the use of sophisticated techniques to anyone who wants to explore these fascinating topics.

The introduction of these powerful personal computers has been both good and bad. It's good because most statistical analyses no longer require access to a huge and expensive mainframe computer. Instead, a simple personal computer costing less than \$500 can do 95% of what 95% of the people need. On the other hand, less than adequately educated students (such as your fellow students who passed on taking this course!) will take any old data they have and think that by running them through some sophisticated analysis, they will have reliable, trustworthy, and meaningful

outcomes—not true. What your professor would say is, “Garbage in, garbage out”; if you don’t start with reliable and trustworthy data, what you’ll have after your data are analyzed are unreliable and untrustworthy results.

Today, statisticians in all different areas, from criminal justice to geophysics to psychology, find themselves using basically the same techniques to answer different questions. There are, of course, important differences in how data are collected, but for the most part, the analyses (the plural of *analysis*) that are done following the collection of data (the plural of *datum*) tend to be very similar, even if called something different. The moral here? This class will provide you with the tools to understand how statistics are used in almost any discipline. Pretty neat, and all for just three or four credits.

If you want to learn more about the history of statistics and see a historical time line, great places to start are Saint Anselm’s College at www.anselm.edu/homepage/jpitocch/biostatshist.html and the University of California–Los Angeles at www.stat.ucla.edu/history/.

Okay. Five minutes is up, and you know as much as you need to know about the history of statistics. Let’s move on to what it is (and isn’t).

STATISTICS: WHAT IT IS (AND ISN'T)

Statistics for People Who (Think They) Hate Statistics, Excel 2010 Edition is a book about basic statistics and how to apply them to a variety of different situations, including the analysis and understanding of information.

In the most general sense, **statistics** describes a set of tools and techniques that are used for describing, organizing, and interpreting information or data. Those data might be the scores on a test taken by students participating in a special math curriculum, the speed with which problems are solved, the number of patient complaints when using one type of drug rather than another, the number of errors in each inning of a World Series game, or the average price of a dinner in an upscale restaurant in Santa Fe, New Mexico.

In all of these examples, and the million more we could think of, data are collected, organized, summarized, and then interpreted. In this book, you’ll learn about collecting, organizing, and summarizing data as part of descriptive statistics. And then you’ll learn about interpreting data when you learn about the usefulness of inferential statistics.

What Are Descriptive Statistics?

Descriptive statistics are used to organize and describe the characteristics of a collection of data. The collection is sometimes called a **data set** or just **data**.

For example, the following list shows you the names of 22 college students, their major areas of study, and their ages. If you needed to describe what the most popular college major is, you could use a descriptive statistic that summarizes their most frequent choice (called the mode). In this case, the most common major is psychology. And if you wanted to know the average age, you could easily compute another descriptive statistic that identifies this variable (that one's called the mean). Both of these simple descriptive statistics are used to describe data. They do a fine job allowing us to represent the characteristics of a large collection of data such as the 22 cases in our example.

Name	Major	Age	Name	Major	Age
Richard	Education	19	Elizabeth	English	21
Sara	Psychology	18	Bill	Psychology	22
Andrea	Education	19	Hadley	Psychology	23
Steven	Psychology	21	Buffy	Education	21
Jordan	Education	20	Chip	Education	19
Pam	Education	24	Homer	Psychology	18
Michael	Psychology	21	Margaret	English	22
Liz	Psychology	19	Courtney	Psychology	24
Nicole	Chemistry	19	Leonard	Psychology	21
Mike	Nursing	20	Jeffrey	Chemistry	18
Kent	History	18	Emily	Spanish	19

So watch how simple this is. To find the most frequently selected major, just find the one that occurs most often. And to find the average age, just add up all the age values and divide by 22. You're right—the most often occurring major is psychology (9 times), and the average age is 20.3 (actually 20.27). Look, Ma! No hands—you're a statistician.

What Are Inferential Statistics?

Inferential statistics are often (but not always) the next step after you have collected and summarized data. Inferential statistics are

used to make inferences based on a smaller group of data (such as our group of 22 students) about a possibly larger one (such as all the undergraduate students in the College of Arts and Sciences).

This smaller group of data is often called a **sample**, which is a portion, or a subset, of a **population**. For example, all the fifth graders in Newark (you author's fair city of origin), New Jersey, would be a population (the population is all the occurrences with certain characteristics—being in fifth grade and attending school in Newark), whereas a selection of 150 of these students would be a sample.

Let's look at another example. Your marketing agency asks you (a newly hired researcher) to determine which of several names is most appealing for a new brand of potato chip. Will it be Chipsters? FunChips? Crunchies? As a statistics pro (we know we're moving a bit ahead of ourselves, but keep the faith), you need to find a small group of potato chip eaters who are representative of all potato chip fans and ask these people to tell you which one of the three names they like the most. Then, if you did things right, you can easily extrapolate the findings to the huge group of potato chip eaters.

Or let's say you're interested in the best treatment for a particular type of disease. Perhaps you'll try a new drug as one alternative, a placebo (a substance that is known not to have any effect) as another alternative, and nothing as the third alternative to see what happens. Well, you find out that a larger number of patients get better when no action is taken and nature just takes its course! The drug does not have any effect. Then, with that information, you extrapolate to the larger group of patients who suffer from the disease, given the results of your experiment.

In Other Words . . .

Statistics is a tool that helps us understand the world around us. It does so by organizing information we've collected and then letting us make certain statements about how characteristics of those data are applicable to new settings. Descriptive and inferential statistics work hand in hand, and which statistic you use and when depends on the question you want answered.

And today, a knowledge of statistics is more important than ever because it provides us with the tools to make decisions that are based on empirical (observed) evidence and not our own biases or beliefs. Want to know if early intervention programs work? Then test whether they work and provide that evidence to the court

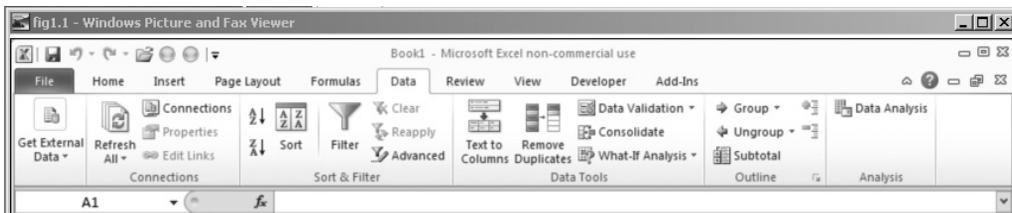
where a ruling will be made on the viability of a new school bond issue that could pay for those programs.

TOOLING AROUND WITH THE ANALYSIS TOOLPAK

An awful lot of what you need to know about using Excel can be found in Appendix A. However, certain Excel procedures are available only if you have the Analysis ToolPak installed (and we use those tools in several chapters throughout the book).

The Analysis ToolPak is a spectacular Excel add-in, a special set of tools that may not have been installed when Excel was originally installed. How do you know if it is installed on the computer you are using? If the Analysis ToolPak doesn't appear on your Data Tab as Data Analysis (as you see in Figure 1.1), you need to install it. Either ask your instructor to have this done on the network level where Excel is installed, or install it on your own machine by doing the following.

Figure 1.1 The Analysis ToolPak on the Data Tab



1. Click the File tab and then click Options.
2. Click Add-Ins and then in the Add-Ins box, select Analysis ToolPak.
3. Click Go.
4. In the Add-Ins box, click the Analysis ToolPak check box and then click OK.

The Analysis ToolPak is an Excel add-in and will be available on the Data tab. An add-in is a program that adds custom commands and features to Microsoft Office. Microsoft Office is Excel's mother and father.

You are now done and ready to make your Excel activities even that much more productive and fun. You can learn how to use the Analysis ToolPak in Little Chapter 1b.

WHAT AM I DOING IN A STATISTICS CLASS?

You probably find yourself using this book for many reasons. You might be enrolled in an introductory statistics class. Or you might be reviewing for your comprehensive exams. Or you might even be reading this on summer vacation (horrors!) in preparation for a more advanced class.

In any case, you are a statistics student, whether you have to take a final exam at the end of a formal course or you're just in it of your own accord. But there are plenty of good reasons to be studying this material—some fun, some serious, and some both.

Here's the list of some of the things that my students hear at the beginning of our introductory statistics course:

1. Statistics 101 or Statistics 1 or whatever it's called at your school looks great listed on your transcript. Kidding aside, this may be a required course for you to complete your major. But even if it is not, having these skills is definitely a big plus when it comes time to apply for a job or for further schooling. And with more advanced courses, your résumé will be even more impressive.
2. If this is not a required course, taking basic statistics sets you apart from those who do not. It shows that you are willing to undertake a course that is above average with regard to difficulty and commitment. And, as the political and economic world becomes more "accountable," more emphasis is being placed on analytic skills. Who knows, this course may be your ticket to a job!
3. Basic statistics is an intellectual challenge of a kind that you might not be used to. There's a good deal of thinking that's required, a bit of math, and some integration of ideas and application. The bottom line is that all this activity adds up to what can be an invigorating intellectual experience because you learn about a whole new area or discipline.
4. There's no question that having some background in statistics makes you a better student in the social or behavioral sciences, because you will have a better understanding not only of what you read in journals but also of what your professors and colleagues may be discussing and doing in and out of class. You will be amazed the first time you say to yourself, "Wow, I actually understand what they're talking about." And it will happen over and over again, because you will have the basic tools necessary to understand exactly how scientists reach the conclusions they do.

5. If you plan to pursue a graduate degree in education, anthropology, economics, nursing, sociology, or any one of many other social, behavioral, and biological pursuits, this course will give you the foundation you need to move further.
6. Finally, you can brag that you completed a course that everyone thinks is the equivalent of building and running a nuclear reactor.

TEN WAYS TO USE THIS BOOK (AND LEARN STATISTICS AT THE SAME TIME!)

Yep. Just what the world needs—another statistics book. But this one is different. It is directed at the student, is not condescending, is informative, and is as basic as possible in its presentation. It makes no presumptions about what you should know before you start and proceeds in slow, small steps, which lets you pace yourself.

However, there has always been a general aura surrounding the study of statistics that it's a difficult subject to master. And we don't say otherwise, because parts of it are challenging. On the other hand, millions and millions of students have mastered this topic, and you can, too. Here are 10 hints to close this introductory chapter before we move on to our first topic.

1. **You're not dumb.** That's true. If you were, you would not have gotten this far in school. So, treat statistics as you would any other new course. Attend the lectures, study the material, do the exercises in the book and from class, and you'll do fine. Rocket scientists know statistics, but you don't have to be a rocket scientist to succeed in statistics.
2. **How do you know statistics is hard?** Is statistics difficult? Yes and no. If you listen to friends who have taken the course and didn't work hard and didn't do well, they'll surely volunteer to tell you how hard it was and how much of a disaster it made of their entire semester, if not their lives. And let's not forget—we always tend to hear from complainers. So, we'd suggest that you start this course with the attitude that you'll wait and see how it is and judge the experience for yourself. Better yet, talk to several people who have had the class and get a good idea of what they think. Just don't base your expectations on one spoilsport's experience.
3. **Don't skip lessons—work through the chapters in sequence.** *Statistics for People Who (Think They) Hate Statistics, Excel 2010 Edition* is written so that each chapter provides a foundation for the next one in the book. When you are all done with the course, you will (I hope) continue to use this book

as a reference. So if you need a particular value from a table, you might consult Appendix B. Or if you need to remember how to compute the standard deviation, you might turn to Chapter 3. But for now, read each chapter in the sequence that it appears. It's okay to skip around and see what's offered down the road. Just don't study later chapters before you master earlier ones.

4. **Form a study group.** This is a big hint and one of the most basic ways to ensure some success in this course. Early in the semester, arrange to study with friends. If you don't have any who are in the same class as you, then make some new ones or offer to study with someone who looks as happy to be there as you are. Studying with others allows you to help them if you know the material better, or to benefit from those who know some material better than you. Set a specific time each week to get together for an hour and go over the exercises at the end of the chapter or ask questions of one another. Take as much time as you need. Studying with others is an invaluable way to help you understand and master the material in this course.
5. **Ask your teacher questions, and then ask a friend.** If you do not understand what you are being taught in class, ask your professor to clarify it. Have no doubt—if you don't understand the material, then you can be sure that others do not as well. More often than not, instructors welcome questions. And especially because you've read the material before class, your questions should be well informed and help everyone in class to better understand the material.
6. **Do the exercises at the end of a chapter.** The exercises are based on the material and the examples in the chapter they follow. They are there to help you apply the concepts that were taught in the chapter and build your confidence at the same time. How do the exercises do that? An explanation for how each exercise is solved is provided in Appendix D. If you can answer these end-of-chapter exercises, then you are well on your way to mastering the content of the chapter.
7. **Practice, practice, practice.** Yes, it's a very old joke:
Q. How do you get to Carnegie Hall?
A. Practice, practice, practice.
Well, it's no different with basic statistics. You have to use what you learn and use it frequently to master the different ideas and techniques. This means doing the exercises in the back of Chapters 1 through 17 as well as taking advantage of any other opportunities you have to understand what you have learned.
8. **Look for applications to make it more real.** In your other classes, you probably have occasion to read journal articles, talk about the results of research, and generally discuss the importance of the scientific method in your own

area of study. These are all opportunities to see how your study of statistics can help you better understand the topics under class discussion as well as the area of beginning statistics. The more you apply these new ideas, the more full your understanding will be.

9. **Browse.** Read over the assigned chapter first, then go back and read it with more intention. Take a nice leisurely tour of *Statistics for People Who (Think They) Hate Statistics* to see what's contained in the various chapters. Don't rush yourself. It's always good to know what topics lie ahead as well as to familiarize yourself with the content that will be covered in your current statistics class.
10. **Have fun.** This might seem like a strange thing to say, but it all boils down to you mastering this topic rather than letting the course and its demands master you. Set up a study schedule and follow it, ask questions in class, and consider this intellectual exercise to be one of growth. Mastering new material is always exciting and satisfying—it's part of the human spirit. You can experience the same satisfaction here—just keep your eye on the ball and make the necessary commitment to stay current with the assignments and work hard.

And a short note for Mac users. Over the years, the Excel people at Microsoft have become increasingly kind to users of the Macintosh version. The latest versions of Excel for a Windows operating system and a Macintosh operating system are almost identical. One big difference (and it really isn't very big) is the keystrokes that one uses to accomplish particular tasks. So, for example, instead of using the Ctrl+C key combination in Windows to copy highlighted text windows, the Mac uses the Apple or the Command key (the cool little key on the lower left of the keyboard with the four little squiggles) in combination with the C key to accomplish the same. This Apple key is also referred to (believe it or not) as the splat, the cloverleaf, the butterfly, the beanie, and the flower key. Using Excel in one operating system or the other (or both) requires a very similar set of tasks, and you should have no problem making the adjustment. And, both the Windows and the Mac version of Excel can read each other's files, so you are safe exchanging files between one operating system and the other. All that said, if they really want to impress their friends, Mac users can go to System Preferences and reconfigure the keyboard to ensure that Windows and Mac keystrokes are exactly the same!

The biggest difference between the two versions is that the ToolPak is just not available for the Mac version. We suspect that it someday will be, but in the 2011 version it is not. What to do? One solution is to

download StatPlus:mac LE from AnalystSoft at www.analystsoft.com/en/products/statplasmacle/. This software is free and works within Excel in much the same way Windows version of the ToolPak works. You can also upgrade to the Pro version, which will do even more.

ABOUT THOSE ICONS

An icon is a symbol. Throughout *Statistics for People . . .*, you'll see a variety of icons. Here's what each one is and what each represents:



This icon represents information that goes beyond the regular text. We might find it necessary to elaborate on a particular point, and we can do that more easily outside of the flow of the usual material.



Here, we select some more technical ideas and tips to discuss and to inform you about what's beyond the scope of this course. You might find these interesting and useful.



Throughout *Statistics for People . . .*, you'll find a small-steps icon like the one you see here. This indicates that a set of steps is coming up that will direct you through a particular process. These steps have been tested and approved by whatever federal agency approves these things.



That finger with the bow is a cute icon, but its primary purpose is to help reinforce important points about the topic that you just read about. Try to emphasize these points in your studying, because they are usually central to the topic.



Many of the chapters in *Statistics for People . . .* provide detailed information about one or more particular statistical procedures and the computation that accompanies them. The computer icon is used to identify the “Using the Computer to . . .” section of the chapter.



The more Excel icon identifies additional information on the Excel feature that has just been mentioned or worked with.

Appendix A, *Excel-erate Your Learning: All You Need to Know About Excel*, contains a collection of 50 basic and important tasks that anyone who uses Excel should know.

Appendix B contains important tables you will learn about and need throughout the book.

And, in working through the exercises in this book, you will use the data sets in Appendix C. You'll find references to data sets named things such as "Chapter 2 Data Set 1," and each of these sets is shown in Appendix C. You can either enter the data manually or download them from the publisher's site at www.sagepub.com/salkindexcel3e or get them directly from the author at www.onlinefilefolder.com with the following info:

User name: ancillaries

Password: files

KEY TO DIFFICULTY ICONS

To help you along a bit, we placed a difficulty index at the beginning of each chapter. This adds some fun to the start of each chapter, but it's also a useful tip to let you know what's coming and how difficult chapters are in relation to one another.

☹ (very hard)

☹ ☹ (hard)

☹ ☹ ☹ (not too hard, but not easy either)

☺ ☺ ☺ ☺ (easy)

☺ ☺ ☺ ☺ ☺ (very easy)

KEY TO "HOW MUCH EXCEL" ICONS

To help you along a bit more, we placed a "How Much Excel" index at the beginning of each chapter. This adds even more fun (groan) to the start of each chapter, but it also lets you know how much Excel material is contained in the chapter.

How much Excel?

 (just a mention)

  (some)

   (lots)

    (lots and lots)

     (a ton)

GLOSSARY

Bolded terms in the text are included in the glossary at the back of the book.

Summary

That couldn't have been that bad, right? We want to encourage you to continue reading and not worry about what's difficult or time-consuming or too complex for you to understand and apply. Just take one chapter at a time, as you did this one.

Time to Practice

Because there's no substitute for the real thing, Chapters 1 through 17 each end with a set of exercises that will help you review the material that was covered in the chapter. The answers to these exercises can be found near the end of the book in Appendix D.

For example, here is the first set of exercises (but don't look for any answers for these because these are kind of "on your own" answers—each answer is highly tied to your own experiences and interest).

1. Interview someone who uses statistics in his or her everyday work. It might be your adviser, an instructor, a researcher who lives on your block, a market analyst for a company, a city planner, or . . . Ask the person what his or her first statistics course was like. Find out what the person liked and didn't like. See if this individual has any suggestions to help you succeed. And most important, ask the person about how he or she uses these new-to-you tools at work.
2. We hope that you are part of a study group, but, if that is not possible, that you have a telephone, email, instant messaging, or webcam study buddy (or even more than one). Talk to your group or a fellow student in your class about similar likes, dislikes, fears, etc., about the statistics course. What do you have in common? Not in common? Discuss with your fellow student strategies to overcome your fears.
3. Search through your local newspaper and find the results of a survey or interview about any topic. Summarize the results and do the best job you can describing how the researchers who

were involved, or the authors of the survey, came to the conclusions they did. Their methods and reasoning may or may not be apparent. Once you have some idea of what they did, try to speculate as to what other ways the same information might be collected, organized, and summarized.

4. Go to the library and copy a journal article in your own discipline. Then, go through the article and highlight the section (usually the “Results” section) where statistical procedures were used to organize and analyze the data. You don’t know much about the specifics of this yet, but how many of these different procedures (such as *t*-test, mean, and calculation of the standard deviation) can you identify? Can you take the next step and tell your instructor how the results relate to the research question or the primary topic of the research study?
5. Find five websites that contain data on any topic and write a brief description of what type of information is offered and how it is organized. For example, if you go to the mother of all data sites, the US Census (www.census.gov), you’ll find links to Access Tools, which takes you to a page just loaded with links to real live data. Try to find data and information that fit in your own discipline.
6. And the big extra-credit assignment is to find someone who actually uses Excel for daily data analysis needs. Ask why he or she uses Excel rather than a more singularly focused program such as SPSS or MiniTab. You may very well find these good folks in everything from political science to nursing, so search widely!
7. Finally, as your last in this first set of exercises, come up with five of the most interesting questions you can about your own area of study or interest. Do your best to come up with questions for which you would want real, existing information or data to answer. Be a scientist!

1A

All You Need to Know About Formulas and Functions

Difficulty Scale ☺ ☺ (a little tough, but invaluable to stick with)

How much Excel? 📊 📊 📊 📊 (lots and lots)

What you'll learn about in this chapter

- The difference between formulas and functions
- How to create and use a formula
- The important Excel functions
- How to select and use a function

There may be nothing more valuable in your Excel magic tool box than formulas and functions. They both allow you to bypass (very) tedious calculations and get right to the heart of the matter. Both formulas and functions are shortcuts—and both work in different ways and do different things. Let's start with formulas.

WHAT'S A FORMULA?

You probably already know the answer to that question. A **formula** is a set of mathematical operators that performs a particular mathematical task. For example, here's a simple formula:

$$2 + 2 =$$

The operator “+” tells you to add certain values (a 2 and another 2) together to produce the outcome (4). This is a simple one.

Here’s one that’s a bit more advanced and one with which you will become more familiar in Chapter 16 of *Statistics for People . . .*:

$$Y' = bX + a \quad (1a.1)$$

This is the formula that is used to predict the value of Y' from our knowledge of the values of b , X , and a . We’ll worry about what all those mean later, but this is a formula that contains a bunch of symbols and mathematical operators and helps us compute numbers we need to make decisions.

Excel is a formula engine just ready for you to use these tools to make your learning of statistics easier.

Creating a Formula



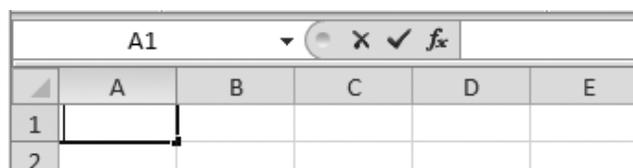
A formula is created through these steps.

1. Click on the **cell** in which you want the results of the formula to appear.
2. Enter an equal sign, which looks like this: =. All formulas begin with an equal sign, no matter what else they contain.
3. Enter the formula. No spaces in formulas please—Excel does not like them.
4. Press the Enter key, and voilà! The results of the formula will appear in the selected cell.

For example, let’s enter the formula that was shown earlier— $2 + 2$ —and see how these steps work.

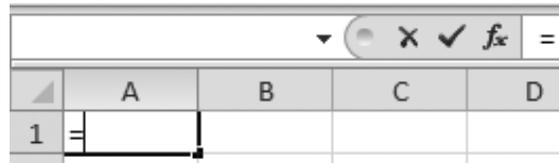
1. As you can see in Figure 1a.1, we selected Cell A1.

Figure 1a.1 Selecting a Cell into Which a Formula Will Be Entered



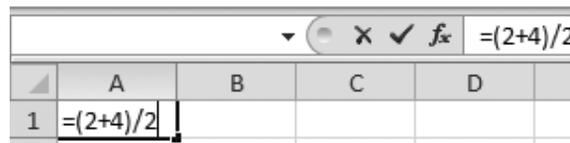
- The equal sign is entered, as shown in Figure 1a.2. And, as you can see, the formula bar at the top of the column becomes active. Everything we enter in Cell A1 will appear in the formula bar.

Figure 1a.2 Entering the Equal Sign to Indicate the Beginning of a Formula



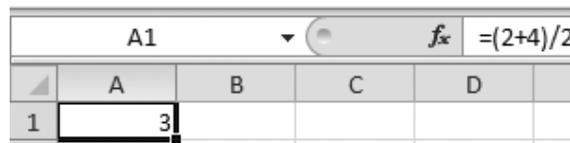
- Enter the rest of the formula, which in this case is $(2+4)/2$, as you see in Figure 1a.3.

Figure 1a.3 Entering the Formula in Cell A1



- Press Enter, and the value of the formula is returned to the cell, as you see in Figure 1a.4.

Figure 1a.4 The Value of a Formula Being Returned to the Cell



A few notes:

- A formula always begins with an equal sign, which tells Excel that what follows is the formula.
- The formula itself always appears in the formula bar.
- The results of the formula (and not the formula itself) are returned to the selected cell.

This is the simplest example of how to use a formula. Formulas can become as complex as you need them to be.

**More Excel**

Want to see the formula behind the scenes in a worksheet? Just use the Ctrl+` key combination to toggle between formulas and the results of those formulas. The ` key is to the left of the number 1 key near the top of the keyboard.

Operator, Operator—Get Me a Formula!

You have just seen that even the simplest formulas consist of operators. In this case, the only operators are a plus (+) sign and a division (/) sign, which direct Excel to add the two values you see in Figure 1a.3, divide by 2, and return the sum to Cell A1.

Addition and division are just two types of operation you can perform. The most important operations and the symbols you use to accomplish them are shown in the following table.

Operator	Symbol	Example	What It Does
Addition	+ (plus)	=2+5	Adds 2 and 5.
Subtraction	- (minus)	=5-3	Subtracts 3 from 5.
Division	/ (slash)	=10/5	Divides 10 by 5.
Multiplication	* (asterisk)	=2*5	Multiplies 2 times 5.
Power of	^ (caret)	=4^2	Takes 4 to the power of 2, or squares 4.

Beware the Parentheses

When you create a formula that goes beyond a very simple one, it is critical for you to consider the order in which operations are carried out and the use of parentheses.

Let's say that we want to find the average score on a weekly test given each Friday for a month and the scores range from 0 to 100.

Here are Willy's scores:

Week 1	78
Week 2	56
Week 3	85
Week 4	92 (Willy finally got it!)

We need to create a formula that will add all of the scores together and divide the sum by 4. We'll name each score w_1 , w_2 , w_3 , and w_4 . Here's one way we might try it:

$$w_1 + w_2 + w_3 + w_4/4$$

Oops! What this does is it adds w_1 , w_2 , and w_3 together and then adds that sum to the value of w_4 divided by 4. This is not what we want!

Rather, take a look at this formula:

$$(w_1 + w_2 + w_3 + w_4)/4$$

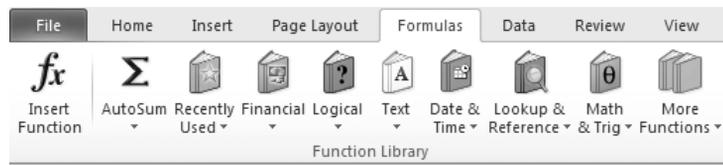
This is more like it. Here, the four values were summed and then divided by 4. This one works. The lesson? Watch your parentheses when you're working with anything beyond the most simple equation.

WHAT'S A FUNCTION?

You know that a formula is a simple set of symbols (such as numbers and operators) that performs some calculation and results in an outcome in the cell where the formula lives.

A **function** is nothing other than a predefined formula. The good people who created Excel developed a whole bunch of these that can do many different things, but for our purposes throughout *Statistics for People . . .*, we deal only with those that are relevant to the chapters of this book. These functions fall under the general Formulas tab on the ribbon, as you see in Figure 1a.5.

Figure 1a.5 The Formulas Tab on the Excel Ribbon



For example, there are groups of financial functions, logical functions, text functions, and others. But we're going to focus only on the functions that fall in the category of statistical functions (which is only visible on your screen when you click on the More Functions drop-down box, which you see in Figure 1a.5, and then click Statistical).

And, out of all of these statistical functions, we deal only with the ones that are relevant to the material covered in this book, such as **AVERAGE** (guess what that does) and **T.TEST** (guess, but you

probably don't yet know). Some are too advanced for us to bother with, such as FISHER and GAMALIN. We'll leave those for the next course or for you to explore on your own.

Using a Function

Unlike a formula, a function is not created by you. You just tell it which values (in which cells) you want to work with. And, every formula contains two different elements—the name of the function and the argument of the function.

Here's a very simple function that averages a given set of numbers. In this example, this function averages the numbers in Cells A1 through A3:

`=AVERAGE(A1:A3)`

The name of the function is AVERAGE, and the argument is A1:A3—the cells on which you want the function to perform its magic. And as you can see, functions (like formulas) always, always, always begin with an equal sign.

Here's another function that produces the sum of a set of cells:

`=SUM(A1:A3)`

Simple, right? And, you may be thinking, "Well, why not just use a formula in this case?" and you could. But what if you needed the sum of a set of 3,267 values like this?

`=SUM(A1:A3267)`

Or what if you needed a fancy-schmancy calculation that includes functions that are very complex? Those are both different stories, and you really didn't want to type in `=(A1+A2+A3+A4 . . .)` until you got to A3267, right? We thought not.

So, let's get to the way that we use a function, and as an example, we'll use the AVERAGE function.

To use this (or any other function), you follow three steps.

1. Enter the function in the cell where you want the results to appear.
2. Enter the range of cells on which you want the function to operate.
3. Press the Enter key, and voilà! There you have it.

However, there are several ways to accomplish these three steps, and let's deal with those now.

Inserting a Function (When You Know the Function's Name and How It Works)

Here's the old-fashioned way.

1. Enter the function in the cell where you want the results to appear.

For example, in Figure 1a.6, you can see a data set of 10 values. We are going to find the average of those values using the AVERAGE function. And, to make things a bit clearer, we entered a text label in the cell to the left of where we want the sum to appear.

Figure 1a.6 Creating a Data Set and the Location of the AVERAGE Function

	A	B
1		Value
2		3
3		4
4		2
5		3
6		4
7		5
8		4
9		3
10		2
11		3
12	Average	

2. Type `=AVERAGE(B2:B11)` in Cell B12.
3. Press the Enter key, and presto: As you see in Figure 1a.7, the sum shows up in Cell B12, and in the formula bar, you can see the structure of the function.

Notice that the results of the function (3.3) are returned to the same cell (B12) where the function was entered. Pretty cool.

And not very difficult. And very convenient. Remember that you can do this with any function. But how do you know what the structure of the function is? That's where the next step comes in.

Figure 1a.7 The Completed AVERAGE Function

		B12		fx		=AVERAGE(B2:B11)	
	A	B	C	D	E		
1		Value					
2		3					
3		4					
4		2					
5		3					
6		4					
7		5					
8		4					
9		3					
10		2					
11		3					
12	Average	3.3					



More Excel

Okay—so how do you know what function to use? Well, certainly one way is through exploring different functions and finding out what they do (which you will do throughout *Statistics for People . . .*). Another is by using Excel Help (press F1 at any time and enter the terms on which you want help). And another way is to look at Table 1a.1 at the end of this little chapter, which gives you a heads-up on which functions we'll be mentioning (some in great detail and others just in passing) throughout the book and what they do.

Inserting a Function Using the Insert Function (fx) Command

Let's use the same example, the AVERAGE function, and assume you haven't used it before but know this is the one you want to use.

We're using the same data as shown in Figure 1a.6. First, erase the results of the function in Cell B12 by selecting the cell and hitting the space bar once and then Enter.

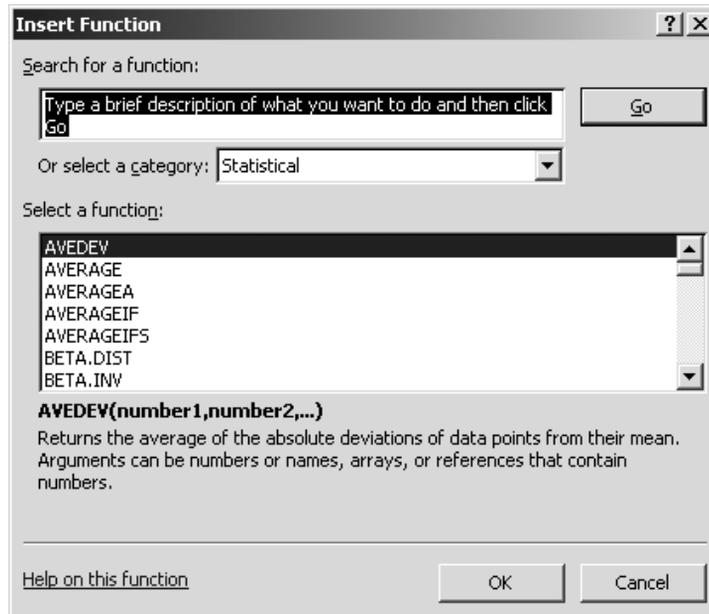
1. Select Cell B12.
2. Click the Formulas tab and the Insert Function command (fx). When you do this, you will see the Insert Function dialog box as shown in Figure 1a.8.



More Excel

A very nice shortcut to the function command (fx) can be found on the Formula bar just to the left of where you see any information that is entered into a cell. Just click that, and you get the Insert Function dialog box.

Figure 1a.8 The Insert Function Dialog Box

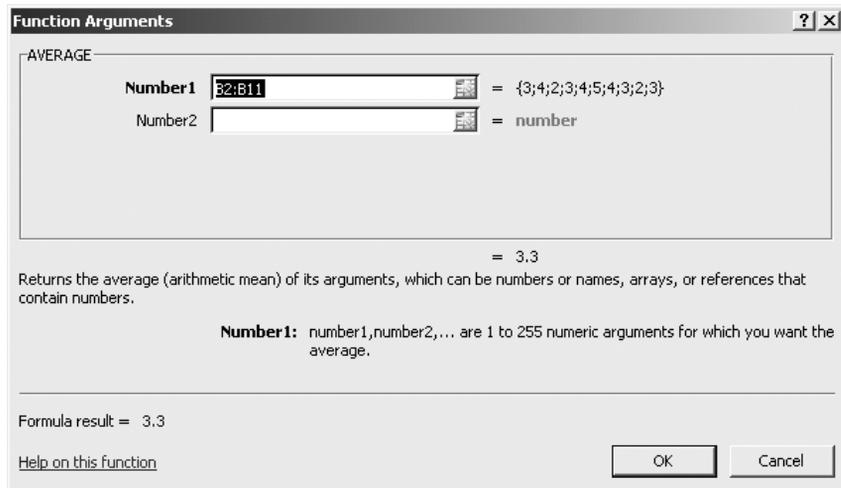


3. Now you can do one of two things:
 - a. Type a brief description for what you want to do, such as AVERAGE, and click the Go button; or
 - b. Find the function you want in the list of functions and double-click on it.

We did option (b), and when we did, the Function arguments dialog box appeared, as shown in Figure 1a.9. Notice that Excel automatically assumed that we wanted to average all of the values above the current cell, and it completed the cell range in the Number 1 box.



Don't get too excited. A function's argument is not really an argument like a disagreement. An argument in mathematical terms is a set of premises, and that's exactly what you need to provide within the parentheses of any function—a set of premises that the function is to carry out.

Figure 1a.9 The Function Arguments Dialog Box

Let's take a look at the different elements in this dialog box.

- There's the name of the function, *AVERAGE*.
 - Then there are text boxes where you enter the range of cells (the argument) on which you want the function to perform its duty. Note that Excel is pretty smart and automatically enters the range of cells it thinks you want to average. Notice also that the actual numbers you want to average are listed to the right of the text box.
 - In the middle of the dialog box (on the left) is a description of what the function does, and to the right of that is the value the function will return to the cell in which it is located (in this case it's 3.3).
 - The syntax (or directions) of how to put the function together is given near the bottom.
 - The formula result is shown again at the bottom left.
 - Finally, there is a link to a place to get help if you need it.
4. Click OK, and you will see the same result as you saw in Figure 1a.7. We entered the function using the Insert Function command instead of by typing its name, but we get the same result.

Inserting a Function Using Formulas → More Functions → Statistical

That really says it all. Just follow these three mouse clicks, and you will see a list of all the statistical functions that are available.

Selecting any of them (such as AVERAGE) provides you with the same dialog box you see in Figure 1a.9.



More Excel

Most functions can do a lot more than first appears. For example, with the most simple of functions such as SUM, you can enter the following variations and get the following results. Excel functions are so useful because they are so flexible.

If you enter the following formula ...	Excel does this ...
=SUM(3,4)	Adds the values to get 7.
=SUM(A2:A4)	Adds the values located in Cells A2 through A4.
=SUM(A2:A4,6)	Adds the values located in Cells A2 through A4 and also adds the value of 6 to that sum.
=SUM(A6:A8,4)	Adds the values located in Cells A6 to A8 and adds the value of 4 to that sum.



More Excel

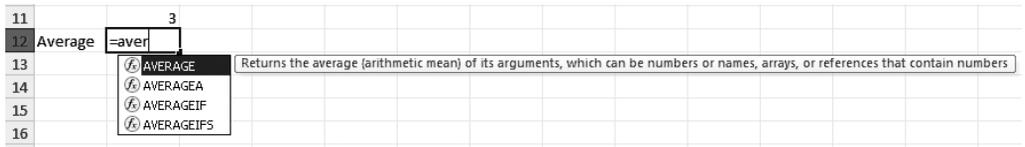
Now you know two ways to insert a function in a worksheet—by typing its name or selecting it through the Insert Function dialog box. And once the specific Function Arguments dialog box (like the one you see in Figure 1a.9) is open, you can just enter the cell addresses in the appropriate text box. However, you can also just click in the cell address box and then drag the mouse over the cell addresses you want to include in that box. This is good. But there's another nifty way to go about this. You can click on the Collapse button (which looks like this , which will shrink the entire dialog box and allow you to select the cells you want using the mouse directly on the worksheet. Then click the Expand button, and the dialog box returns to its normal size with the cell addresses included.



More Excel

If you insert a function by typing it directly into a cell, the 2010 version of Excel provides a list of similarly spelled functions as well as a tip about how to use your function, as you see in Figure 1a.10. Here we typed in =aver, and as we typed, Excel provided a list of various average functions plus a tip as to what the AVERAGE function does.

Figure 1a.10



Using Functions in Formulas

Time to get a bit fancy.

Now, formulas and functions are basically the same animal—they carry out instructions. There’s just no reason why you can’t include a function in a formula.

For example, let’s say that you have three job evaluation scores (Eval 1, Eval 2, and Eval 3) as you see in Figure 1a.11. You also have a Fudge Factor (in column E), and that’s a value you can use to increase or decrease an employee’s score at your discretion. For example, for employee KH, you want to increase that score by 3%, so you multiply the average evaluation score (from Eval 1, Eval 2, and Eval 3) by 1.03. Figure 1a.11 shows the formula using the AVERAGE function (which you will learn how to use in Chapter 2).

Figure 1a.11 Using a Function in a Formula

	A	B	C	D	E	F
1	Name	Eval 1	Eval 2	Eval 3	Fudge Factor	Final Score
2	HY	67	76	34	1.10	64.90
3	NM	65	78	32	1.02	59.50
4	GG	45	98	34	1.23	72.57
5	DF	65	98	54	1.06	76.67
6	KH	76	78	58	1.03	72.79
7	RR	32	76	75	1.04	63.44
8	YR	45	77	54	1.05	61.60
9	HH	43	76	54	1.00	57.67
10	JU	34	54	33	1.01	40.74
11	WE	32	47	78	1.04	54.43

As you can see in the formula bar shown in Figure 1a.11, the formula Cell F2 looks like this:

$$=AVERAGE(B2:D2)*E2$$

And it reads like this: The contents of Cells B2 through D2 are averaged, and then that value is multiplied by the contents of Cell E2. We copied the formula from Cell F2 to Cells F3 through Cell F11, and the results are shown in Column F.

We're Taking Names: Naming Ranges

It's certainly easy enough to enter cell addresses such as A1:A3—not much work involved there.

But what if you're dealing with a really large worksheet with hundreds of columns and rows and thousands of cells? Wouldn't it be nice if you could just enter a name that represents a certain range of cells rather than having to remember all those cell addresses? Desire it no more. Excel allows you to name a range, or a collection of cells.

For example, in Figure 1a.11, if you want to average the employees' second evaluations, instead of using the cell addresses C2:C11, why not just give the range of cells a name, such as `eval2` or `EVAL_2` (no spaces, please!)? Then, the average for that set of scores using the `AVERAGE` function would look like this—

`=AVERAGE(EVAL_2)`

rather than like this—

`=AVERAGE(C2:C11)`

And it gets better—you can just paste that name into any formula or function with a few clicks. You don't even have to type anything! Here's how to assign a name to a range of cells:

1. Highlight the range of cells you want to name.
2. Click the Name box at the left end of the formula bar.
3. Type the name that you want to use to refer to your selection, as shown in Figure 1a.12 (we used `Eval_2`). And again, no spaces please—Excel does not like them in name ranges.
4. Press Enter.

Using Ranges

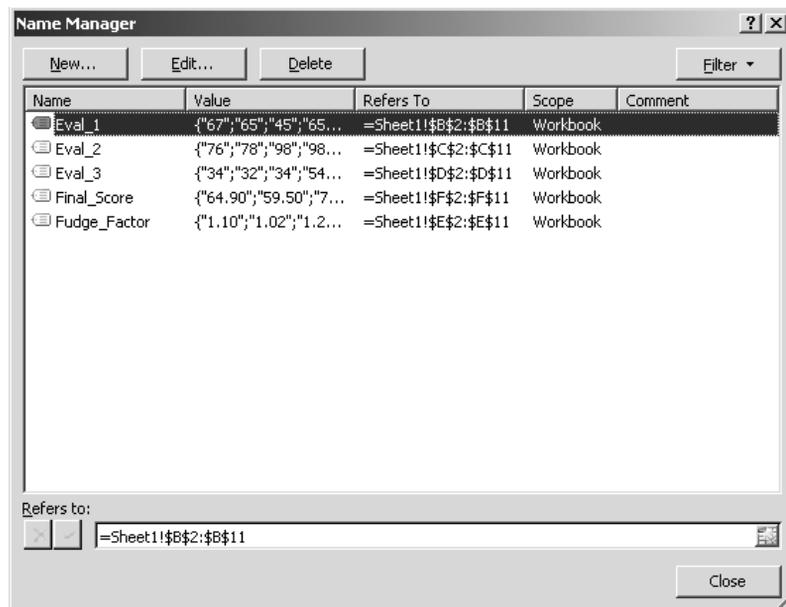
Once a range is defined, you can use the name assigned instead of a cell range. If you remember that the data are named, you can just

Figure 1a.12 Naming a Selection of Cells as a Range

		Eval_2		fx		76	
	A	B	C	D	E	F	
1	Name	Eval 1	Eval 2	Eval 3	Fudge Factor	Final Score	
2	HY	67	76	34	1.10	64.90	
3	NM	65	78	32	1.02	59.50	
4	GG	45	98	34	1.23	72.57	
5	DF	65	98	54	1.06	76.67	
6	KH	76	78	58	1.03	72.79	
7	RR	32	76	75	1.04	63.44	
8	YR	45	77	54	1.05	61.60	
9	HH	43	76	54	1.00	57.67	
10	JU	34	54	33	1.01	40.74	
11	WE	32	47	78	1.04	54.43	

enter that name by using the drop-down menu now available in the Name Manager dialog box (on the Formulas tab), as shown in Figure 1a.13.

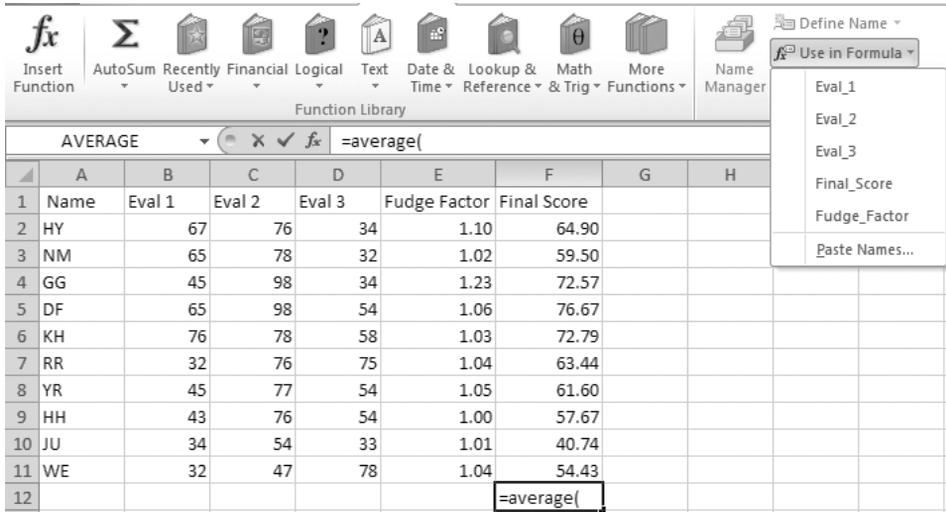
Figure 1a.13 Seeing Range Names



Let's use the ranges that were defined and compute the average of all the Fudge Factor scores.

1. Click on Cell F12 where the average will be placed.
2. Type =AVERAGE(.
3. Click Formulas → Use in Formula, and you will see the Use in Formula drop-down menu, as shown in Figure 1a.14.

Figure 1a.14 The Menu for Selecting a Range of Cells



4. Click on Fudge_Factor.
5. Type) .
6. Double-click on Final Score and take a look at Figure 1a.15, where you can see how the name was used in the function rather than the cell addresses of F2 through F11.

Figure 1a.15 Inserting a Cell Range into a Function

	A	B	C	D	E	F
1	Name	Eval 1	Eval 2	Eval 3	Fudge Factor	Final Score
2	HY	67	76	34	1.10	64.90
3	NM	65	78	32	1.02	59.50
4	GG	45	98	34	1.23	72.57
5	DF	65	98	54	1.06	76.67
6	KH	76	78	58	1.03	72.79
7	RR	32	76	75	1.04	63.44
8	YR	45	77	54	1.05	61.60
9	HH	43	76	54	1.00	57.67
10	JU	34	54	33	1.01	40.74
11	WE	32	47	78	1.04	54.43
12						62.43

Summary

This may be a “little” chapter, but it contains some of the most useful tools that Excel has to offer. In fact, the uses of formulas and functions are really limited only by your imagination. As you use Excel, you will find more and more ways to make these powerful tools work exactly for you.

Time to Practice

1. Create formulas for the following in an Excel worksheet:
 - a. Add the values of 3 and 5 to one another.
 - b. Subtract the value of 5 from 10 and multiply the outcome by 7.
 - c. Average the values 5, 6, 7, and 8.
 - d. Find the sum of the squared values of 3, 4, and 5.
2. Use the AVERAGE function to add the values of 3, 5, and 7 and divide the total by 3.
3. What would the function look like for summing the values in Cells A1 through A5?
4. What would the function look like for averaging the values in Cells A1 through A5?
5. Create a worksheet with any five values in Cells A1 through A5, name the range of these cells Test_Scores, create a formula in Cell A6 that will sum them and find the average. Can you think of another way to do this?

Table 1a.1 The Functions You’ll Love to Love. Note that many functions that appeared in Excel 2007 version (and earlier) have changed slightly. For example, TTEST is now T.TEST (leaving the general Excel user, of course, clueless as to why the change).

The Function Name	What It Does	The Chapter in Which You’ll Read About It
AVERAGE	Returns the average of its arguments.	2
GEOMEAN	Returns the geometric mean.	2
MEDIAN	Returns the median of the given numbers.	2
MODE.SNGL	Returns the most common value in a data set.	2

(Continued)

Table 1a.1 (Continued)

The Function Name	What It Does	The Chapter in Which You'll Read About It
MODE.MULTI	Returns the most common values in a data set.	2
QUARTILE.INC	Returns the quartile of a data set.	2
STDEV.S	Estimates standard deviation based on a sample.	3
STDEV.P	Calculates standard deviation based on the entire population.	3
VAR.S	Estimates variance based on a sample.	3
VAR.P	Calculates variance based on the entire population.	3
KURT	Returns the kurtosis of a data set.	4
SKEW	Returns the skewness of a distribution.	4
CORREL	Returns the correlation coefficient between two data sets.	5 and 16
PEARSON	Returns the Pearson product moment correlation coefficient.	5 and 16
NORM.S.DIST	Returns the standard normal cumulative distribution.	8
STANDARDIZE	Returns a normalized value.	8
Z.TEST	Returns the probability of a one-tailed z value.	10
T.DIST	Returns the Student's t distribution.	11, 12
T.TEST	Returns the probability associated with a Student's t -test.	11, 12
F.DIST	Returns the F probability distribution.	13
F.TEST	Returns the result of an F -test.	13
FORECAST	Returns a value along a linear trend.	16
FREQUENCY	Returns a frequency distribution as a vertical array.	16
LINEST	Returns the parameters of a linear trend.	16
SLOPE	Returns the slope of the linear regression line.	16
STEYX	Returns the standard error of the predicted y value for each x in the regression.	16
TREND	Returns values along a linear trend.	16
CHISQ.DIST	Returns the one-tailed probability of the chi-square distribution.	17
CHISQ.TEST	Returns the test for independence.	17

IB

All You Need to Know About Using the Amazing Analysis ToolPak

Difficulty Scale ☺ ☺ (a little tough,
but invaluable to stick with)

How much Excel? 📊 📊 📊 📊 📊 (a ton)

What you'll learn about in this chapter

- What the Analysis ToolPak is and what it does

(Almost) everything you need to know about Excel, you can learn in Appendix A. But certain Excel procedures are available only if you have the Analysis ToolPak (which used to be called the Data Analysis ToolPak) installed, and we use those tools in several chapters in this book. Excel refers to this set of tools as the Analysis ToolPak, but you will see it on your screen as Data Analysis Tools or Data Analysis—no worries. The Analysis ToolPak is an Excel add-in—a special set of tools that may not have been installed when Excel was originally installed.

How do you know if it is installed on the computer you are using? If the Data Analysis item doesn't appear on the Data tab (usually at the right-hand side of the Data tab), you need to install the Analysis ToolPak. Either ask your instructor to have this done on the network level where Excel is installed, or install it on your own machine as discussed earlier.

The Analysis ToolPak is easy to use. You just follow the instructions and identify the analysis you want to perform and the data on which you want it performed, and you're done. Throughout *Statistics for People . . .*, we will show you in detailed steps how this is done.

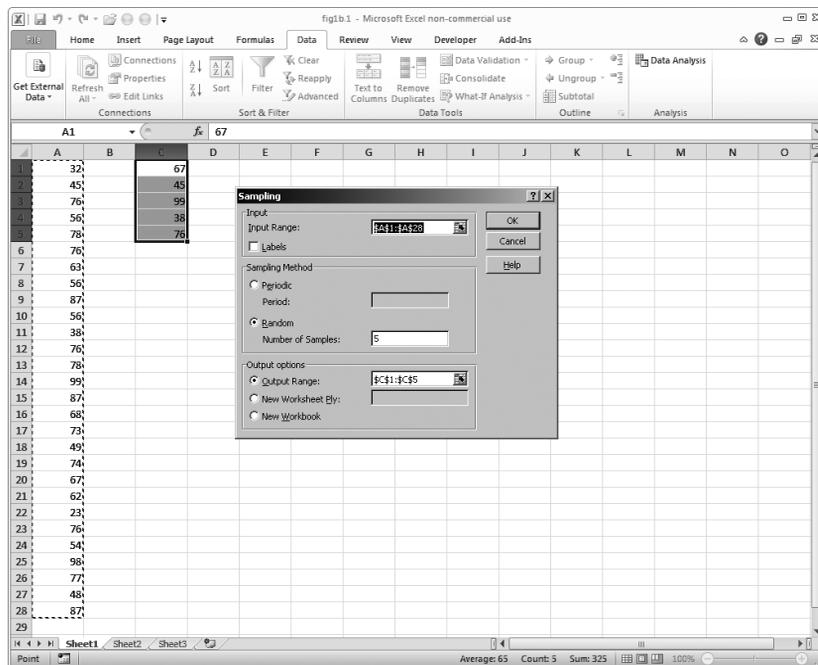
One more note . . . As you already know from Table 1a.1, there are many different functions, and some of these have counterparts in the Analysis ToolPak. For example, you can perform a test for the difference between the means of two groups using a function (T.TEST) or using the Analysis ToolPak (and using the *t*-test . . . tool). That's all for the better. You will often have more than one way to analyze your data—always a nice position to be in.

A LOOK AT THE ANALYSIS TOOLPAK

For now, let's just look at some output where we took a random sample of 5 numbers from a group of 25 numbers. In Figure 1b.1, you see the data in Column A; the sample of 5 numbers in Column C; and the Sampling dialog box that we used from the Analysis ToolPak to tell Excel what to sample, how many to sample, and where to put the results. Much more about this later in the book, but we thought you'd like to see how this cool set of tools works.

For our purposes, we will be working with the following Analysis ToolPak tools (we'll be working with about three fourths of

Figure 1b.1 Using the Sampling Tool From the Analysis ToolPak



them—the others are more advanced than what we need). And in the appropriate chapter (such as Chapter 2), you will learn about that particular tool (such as Descriptive Statistics):

ANOVA
Correlation
Descriptive Statistics
Histogram
Moving Average
Random Number Generation
Rank and Percentile
Regression
Sampling
t-test
Z-test

DON'T HAVE IT?

The Analysis ToolPak is an Excel add-in. An add-in is a program that adds custom commands and features to Microsoft Office. Microsoft Office is Excel's mother and father.

To load the Analysis ToolPak into Excel, follow these steps.

1. Click the File Tab and then click Options.
2. Click Add-Ins and then, in the Manage box, select Analysis ToolPak.
3. Click Go.
4. In the Add-Ins dialog box, select the Analysis ToolPak check box and then click OK.

And you are done and ready to make your Excel activities even that much more productive and fun.

