

Chapter 2

Understanding Excel's Statistical Capabilities

In This Chapter

- ▶ Working with worksheet functions
 - ▶ Creating a shortcut to statistical functions
 - ▶ Getting an array of results
 - ▶ Naming arrays
 - ▶ Tooling around with analysis
 - ▶ Using Excel's Quick Statistics feature
-

In this chapter, I introduce you to Excel's statistical functions and data analysis tools. If you've used Excel, and I'm assuming you have, you're aware of Excel's extensive functionality, of which statistical capabilities are a subset. Into each worksheet cell you can enter a piece of data, instruct Excel to carry out calculations on data that reside in a set of cells, or use one of Excel's worksheet functions to work on data. Each worksheet function is a built-in formula that saves you the trouble of having to direct Excel to perform a sequence of calculations. As newbies and veterans know, formulas are the business end of Excel. The data analysis tools go beyond the formulas. Each tool provides a set of informative results.

Getting Started

Many of Excel's statistical features are built into its worksheet functions. In previous versions, you accessed the worksheet functions by using the Excel Insert Function button, labeled with the symbol *fx*. Clicking this button opens the Insert Function dialog box, which presents a list of Excel's functions and a capability for searching for Excel functions. Although Excel 2007 provides

easier ways to access the worksheet functions, this latest version preserves this button and offers additional ways to open the Insert Function dialog box. I discuss all of this in more detail in a moment.

Figure 2-1 shows the location of the Insert Function button and the Formula Bar. They're on the right of the Name Box. All three are just below the Ribbon. Inside the Ribbon, in the Formulas tab, is the Function Library.

The Formula Bar is like a clone of a cell you select: Information entered into the Formula Bar goes into the selected cell, and information entered in the selected cell appears in the Formula Bar.

Figure 2-1 shows Excel with the Formulas tab open. This shows you another location for the Insert Function button. Labeled *fx*, it's in the extreme left of the Ribbon, in the Function Library area. As I mention earlier in this section, when you click the Insert Function button, you open the Insert Function dialog box. (See Figure 2-2.)

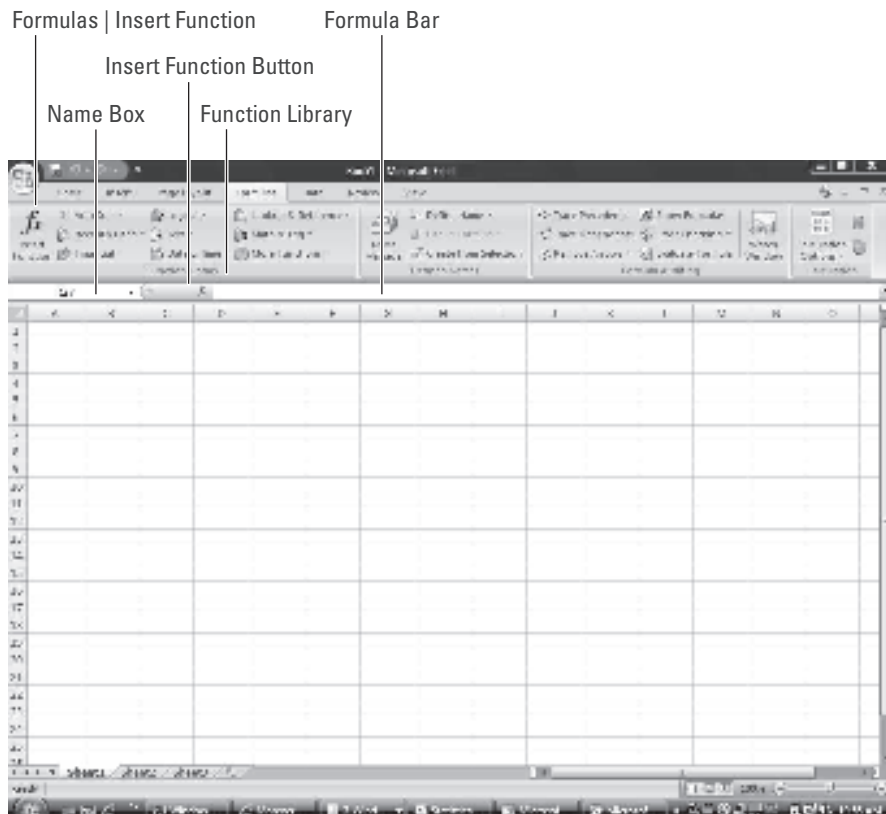
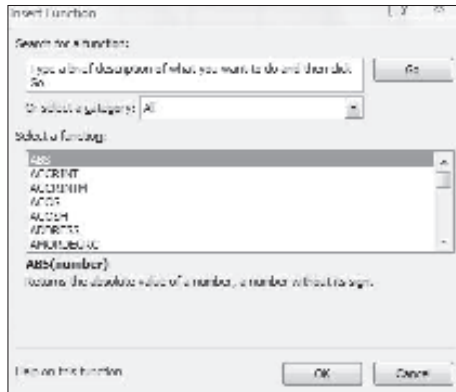


Figure 2-1:
The
Function
Library, the
Name Box,
the Formula
Bar, and
the Insert
Function
button.

Figure 2-2:
The Insert
Function
dialog box.



This dialog box enables you to search for a function that fits your needs, or to scroll through a list of Excel functions.

So in addition to clicking the Insert Function button next to the Formula bar, you can open the Insert Function dialog box by selecting



Formulas | Insert Function

To open the Insert Function dialog box, you can also press Shift+F3.

Because of the way earlier versions of Excel were organized, the Insert Function dialog box was extremely useful. In Excel 2007, however, it's mostly helpful if you're not sure which function to use or where to find it.

The Function Library presents the categories of formulas you can use and makes it convenient for you to access them. Clicking a category button in this area opens a menu of the functions in that category.

Most of the time, I work with Statistical Functions that are easily accessible through the Statistical Functions menu. Sometimes I work with Math functions in the Math & Trig Functions menu. (You see a couple of these later in the chapter.) In Chapter 5, I work with a couple of Logic functions.



The final selection of each category menu (like the Statistical Functions menu) is called Insert Function. Selecting this option is still another way to open the Insert Function dialog box.

The Name Box is something like a running record of what you do in the worksheet. Select a cell, and the cell's address appears in the Name Box. Click the Insert Function button and the name of the function you selected most recently appears in the Name Box.

In addition to the statistical functions, Excel provides a number of data analysis tools you access through the Data tab's Analysis area.

Setting Up for Statistics

In this section, I show you how to use the worksheet functions and the analysis tools.

Worksheet functions in Excel 2007

Because the Ribbon exposes so many of Excel's capabilities, it's not necessary to bury them in menus any more. As I point out in the preceding section, the Function Library area of the Formulas tab shows all the categories of worksheet functions.

The steps in using a worksheet function are:

- 1. Type your data into a data array and select a cell for the result.**
- 2. Select the appropriate formula category and choose your function from its pop-up menu.**
Doing this opens the Function Arguments dialog box.
- 3. In the Function Arguments dialog box, type the appropriate values for the function's arguments.**

Argument is a term from mathematics. It has nothing to do with debates, fights, or confrontations. In mathematics, an argument is a value on which a function does its work.

- 4. Click OK to put the result into the selected cell.**

Yes, that's all there is to it.

To give you an example, I explore a function that typifies how Excel's worksheet functions work. This function, SUM, adds up the numbers in cells you specify and returns the sum in still another cell that you specify. Although adding numbers together is an integral part of statistical number crunching, SUM is not in the Statistical category. It is, however, a typical worksheet function and it shows a familiar operation.

Here, step by step, is how to use SUM.

1. Enter your numbers into an array of cells and select a cell for the result.

In this example, I've entered 45, 33, 18, 37, 32, 46, 39 into cells C2 through C8, and selected C9 to hold the sum.

2. Select the appropriate formula category and choose your function from its pop-up menu.

This opens the Function Arguments dialog box.

I selected Formulas | Math & Trig

and scrolled down to find and choose SUM.

3. In the Function Arguments dialog box, enter the appropriate values for the arguments.

Excel guesses that you want to sum the numbers in cells C2 through C8 and identifies that array in the Number1 box. Excel doesn't keep you in suspense: The Function Arguments dialog box shows the result of applying the function. In this example, the sum of the numbers in the array is 250. (See Figure 2-3.)

4. Click OK to put the sum into the selected cell.

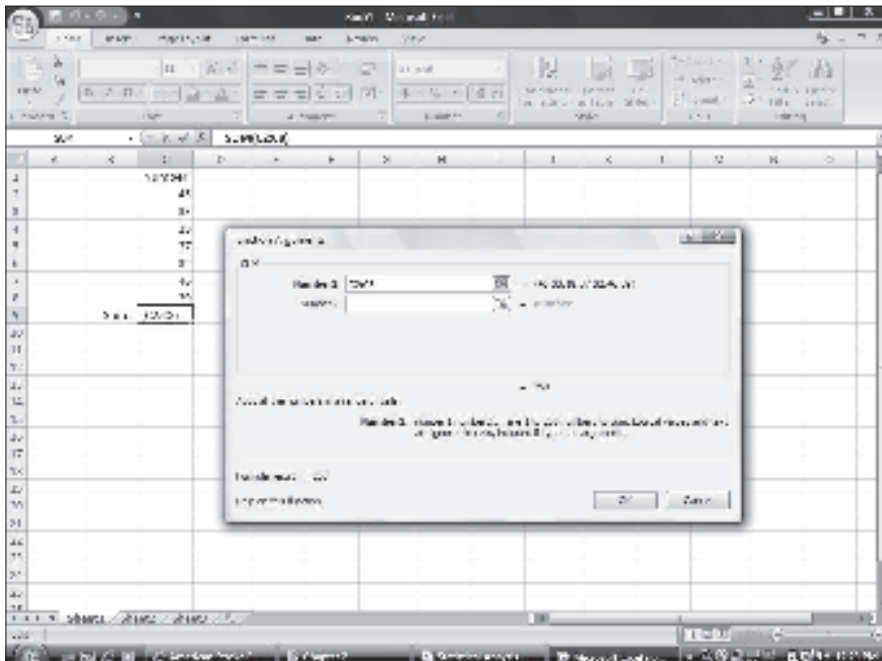


Figure 2-3:
Using SUM.

Note a couple of points. First, as Figure 2-3 shows, the Formula Bar holds

```
=SUM ( C2 : C8 )
```

This formula indicates that the value in the selected cell equals the sum of the numbers in cells C2 through C8.



After you get familiar with a worksheet function and its arguments, you can bypass the menu and type the function directly into the cell or into the formula bar, beginning with “=”. When you do, Excel opens a helpful menu as you type the formula. (See Figure 2-4.) The menu shows possible formulas beginning with the letter(s) you type, and you can select one by double-clicking it.

Figure 2-4:

As you type a formula, Excel opens a helpful menu.



Another noteworthy point is the set of boxes in the Function Arguments dialog box in Figure 2-3. In the figure you see just two boxes, Number1 and Number2. The data array appears in Number1. So what’s Number2 for?

The Number2 box allows you to include an additional argument in the sum. And it doesn’t end there. Click in the Number2 box and the Number3 box appears. Click in the Number3 box, and the Number4 box appears . . . and on and on. The limit is 255 boxes, with each box corresponding to an argument. A value can be another array of cells anywhere in the worksheet, a number, an arithmetic expression that evaluates to a number, a cell ID, or a name that you have attached to a range of cells. (Regarding that last one: Read the upcoming section “What’s in a name? An array of possibilities.”) As you type in values, the SUM dialog box shows the updated sum. Clicking OK puts the updated sum into the selected cell.



You won’t find this multiargument capability on every worksheet function. Some are designed to work with just one argument. For the ones that do work with multiple arguments, however, you can incorporate data that resides all over the worksheet. Figure 2-5 shows a worksheet with a Function Arguments dialog box that includes data from two arrays of cells, two arithmetic expressions, and one cell. Notice the format of the function in the Formula Bar (a comma separates successive arguments).

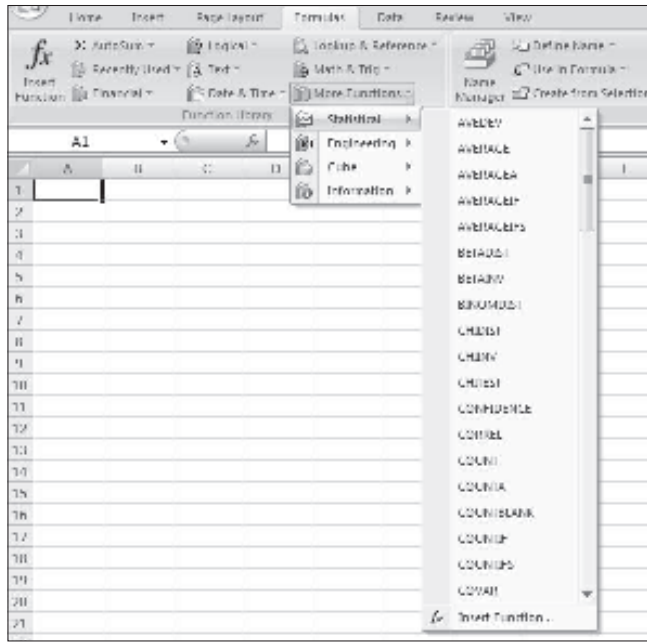


Figure 2-6:
Accessing
Excel's
Statistical
Functions.

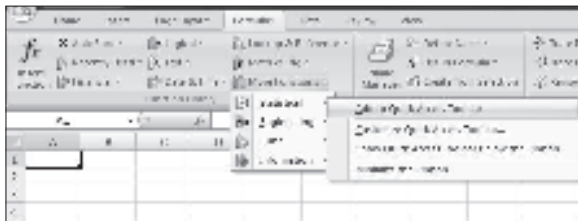
Although Excel has buried the statistical functions several layers deep, you can use a handy Excel 2007 technique to make them as accessible as any of the other categories: You add them to the Quick Access Toolbar in the upper-left corner. (Every Office 2007 application has one.)

To do this, select

Formulas | More Functions

and right-click on Statistical. On the pop-up menu, pick the first option; Add to Quick Access Toolbar. (See Figure 2-7.) Doing this adds a button to the Quick Access Toolbar. Clicking the new button's down arrow opens the pop-up menu of statistical functions. (See Figure 2-8.)

Figure 2-7:
Adding the
Statistical
functions to
the Quick
Access
Toolbar.



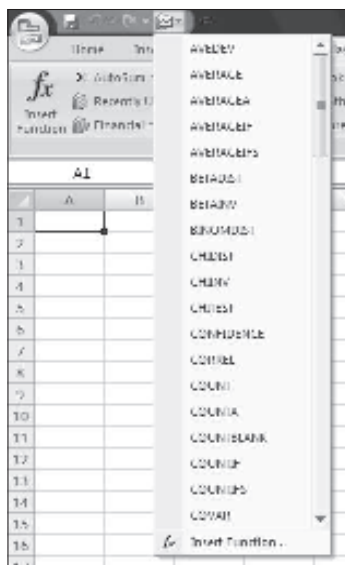


Figure 2-8:
The
Statistical
Functions
menu.

From now on, when I deal with a statistical function, I assume that you've created this shortcut, so you can quickly open the menu of statistical functions. The next section provides an example.

Array functions

Most of Excel's built-in functions are formulas that calculate a single value (like a sum) and put that value into a worksheet cell. Excel has another type of function. It's called an *array function* because it calculates multiple values and puts those values into an array of cells, rather than into a single cell.

FREQUENCY is a good example of an array function (and it's an Excel statistical function, too). Its job is to summarize a group of scores by showing how the scores fall into a set of intervals that you specify. For example, given these scores

77, 45, 44, 61, 52, 53, 68, 55

and these intervals

50, 60, 70, 80

FREQUENCY shows how many are less than or equal to 50 (2 in this example), how many are greater than 50 and less than or equal to 60 (that would be 3), and so on. The number of scores in each interval is called a *frequency*. A table of the intervals and the frequencies is called a *frequency distribution*.

Here's an example of how to use FREQUENCY:

1. Enter the scores into an array of cells.

Figure 2-9 shows a group of scores in cells B2 through B16.

2. Enter the intervals into an array.

I've put the intervals in C2 through C9.

3. Select an array for the frequencies.

I've put Frequency as the label at the top of column D, so I select D2 through D10 for the resulting frequencies. Why the extra cell? FREQUENCY returns a vertical array that has one more cell than the frequencies array.

4. From the Statistical Functions menu, select FREQUENCY to open the Function Arguments dialog box.

I used the shortcut I installed on the Quick Access Toolbar to open this menu and select FREQUENCY.

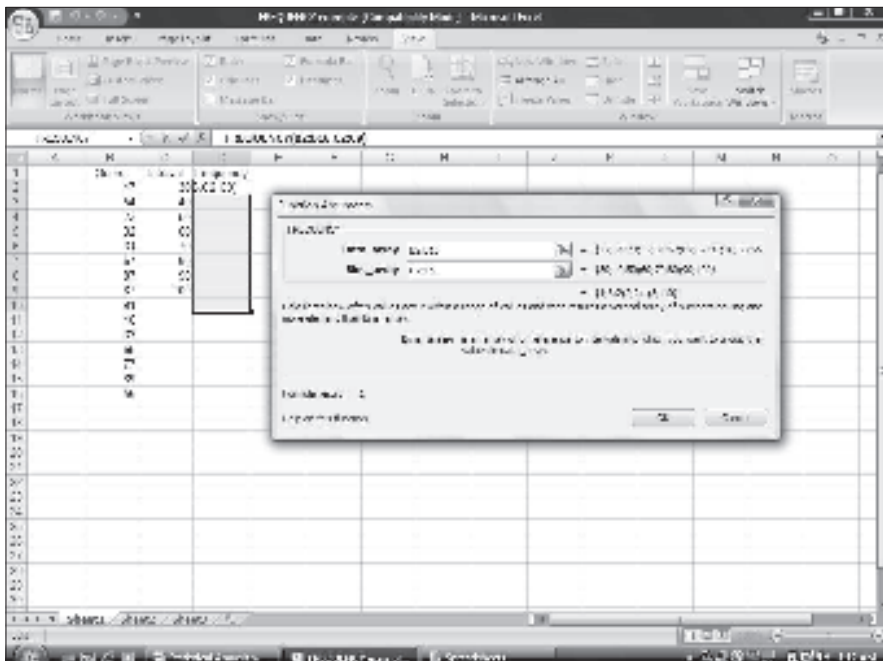


Figure 2-9:
Working
with
FREQUENCY.

5. In the Function Arguments dialog box, enter the appropriate values for the arguments.

I begin with the Data_array box. In this box I entered the cells that hold the scores. In this example, that's B2:B16. I'm assuming you know Excel well enough to know how to do this in several ways.

Next, I identify the intervals array. FREQUENCY refers to intervals as "bins," and holds the intervals in the Bins_array box. For this example, C2:C9 goes into the Bins_array box. After identifying both arrays, the Insert Function dialog box shows the frequencies inside a pair of curly brackets.

6. Press Ctrl+Shift+Enter to Close the Function Arguments dialog box and put the values in the selected array.

This is VERY important. Because the dialog box has an OK button, the tendency is to click OK, thinking that puts the results into the worksheet. That doesn't get the job done when you work with an array function, however. Always use the keystroke combination Ctrl+Shift+Enter to close the Function Arguments dialog box for an array function.

After closing the Function Arguments dialog box, the frequencies go into the appropriate cells, as Figure 2-10 shows.

	Score	Interval	Frequency
2	47	20	3
3	56	40	3
4	23	20	2
5	39	50	2
6	31	70	2
7	57	80	1
8	57	90	3
9	57	100	1
10	51		2
11	46		
12	23		
13	56		
14	77		
15	50		
16	56		

Figure 2-10:
The finished frequencies.

Note the formula in the Formula Bar:

```
{= FREQUENCY (B2 : B16 , C2 : C9 ) }
```

The curly brackets are Excel's way of telling you that this is an array function.



I'm not one to repeat myself, but in this case I'll make an exception. As I said in Step 6, press Ctrl+Shift+Enter whenever you work with an array function. Keep this in mind because the Arguments Function dialog box doesn't provide any reminders. If you click OK after you enter your arguments into an array function, you'll be very frustrated. Trust me.

What's in a name? An array of possibilities

As you get more into Excel's statistical features, you work increasingly with formulas that have multiple arguments. Oftentimes, these arguments refer to arrays of cells, as in the preceding examples.

If you apply meaningful names to these arrays, it helps you keep straight what you're doing. Also, if you come back to a worksheet after being away from it for a while, meaningful array names can help you quickly get back into the swing of things. Another benefit: If you have to explain your worksheet and its formulas to others, meaningful array names are tremendously helpful.

Excel gives you an easy way to attach a name to a group of cells. In Figure 2-11, column C is named Revenue_Millions, indicating "Revenue in millions of dollars." As it stands, that just makes it a bit easier to read the column. If I explicitly tell Excel to treat Revenue_Millions as the name of the array of cells C2 through C13, however, I can use Revenue_Millions whenever I refer to that array of cells.

	A	B	C
1	Year	Region	Revenue_Millions
2	2008	North	20
3	2008	South	22
4	2008	East	19
5	2008	West	25
6	2007	North	28
7	2007	South	28
8	2007	East	21
9	2007	West	27
10	2006	North	37
11	2006	South	29
12	2006	East	25
13	2006	West	31
14			
15			
16			

Figure 2-11:
Defining
names for
arrays of
cells.

Why did I use Revenue_Millions and not Revenue (Millions) or Revenue In Millions or Revenue: Millions? Excel doesn't like blank spaces or symbols in its names. In fact, here are four rules to follow when you supply a name for a range of cells:

- ✓ Begin a name with an alphabetic character — a letter rather than a number or a punctuation mark.
- ✓ As I just mentioned, make sure that the name contains no spaces or symbols. Use an underscore to denote a space between words in the name.

- ✓ Be sure that the name is unique within the worksheet.
- ✓ Be sure that the name doesn't duplicate any cell reference in the worksheet.

Here's how to define a name:

1. Put a descriptive name at the top of a column (or to the left of a row) you want to name.

Figure 2-10 shows this.

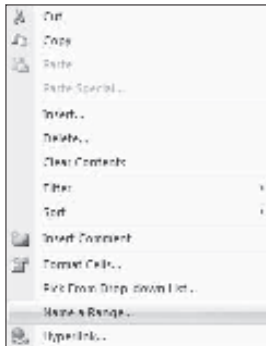
2. Select the range of cells you want to name.

For this example, that's cells C2 through C13. Why not include C1? I explain in a second.

3. Right-click on the selected range.

This opens the menu shown in Figure 2-12.

Figure 2-12:
Right-clicking a selected cell range opens this pop-up menu.



4. From the pop-up menu, select Name a Range.

This selection opens the New Name dialog box (see Figure 2-13). As you can see, Excel knows that Revenue_Millions is the name for the array, and that Revenue_Millions refers to cells C2 through C13. When presented with a selected range of cells to name, Excel looks for a nearby name — just above a column or just to the left of a row. If no name is present, you get to supply one in the New Name dialog box. (The New Name dialog box is also accessible by choosing Formula | Define Name.)

Figure 2-13:
The New
Name dialog
box.



When you select a range of cells like a column with a name at the top, you can include the cell with the name in it and Excel attaches the name to the range. *I strongly advise against doing this.* Why? If I select C1 through C13, the name Revenue_Millions refers to cells C1 through C13, not C2 through C13. In that case, the first value in the range is text and the others are numbers.

For a formula like SUM (or SUMIF or SUMIFS, which I discuss next), this doesn't make a difference: In those formulas, Excel just ignores values that aren't numbers. If you have to use the whole array in a calculation, however, it makes a *huge* difference: Excel thinks the name is part of the array and tries to use it in the calculation. You'll see this in the next section on creating your own array formulas.

5. Click OK.

Excel attaches the name to the range of cells.

Now I have the convenience of using the name in a formula. Here, selecting a cell (like C14) and entering the SUM formula directly into C14 opens the boxes in Figure 2-14.

	A	B	C	D
1	Year	Region	Revenue Millions	
2	2006	North	20	
3	2006	South	22	
4	2006	East	19	
5	2006	West	25	
6	2007	North	26	
7	2007	South	28	
8	2007	East	21	
9	2007	West	27	
10	2008	North	32	
11	2008	South	29	
12	2008	East	25	
13	2008	West	31	
14			=sum(Revenue_Millions)	
15				
16				
17				

Figure 2-14:
Entering
a formula
directly into
a cell opens
these boxes.

As the figure shows, the boxes open as I type. Pressing the Tab key fills in the formula in a way that Excel understands. I have to supply the close parenthesis (see Figure 2-15) and type Enter to see the result.

Using the named array, then, the formula is

```
=SUM(Revenue_Millions)
```

which is more descriptive than

```
=SUM(C2:C13)
```

A couple of Excel 2007's new formulas show just how convenient this naming capability is. These formulas, SUMIF and SUMIFS, add a set of numbers if specific conditions in one cell range (SUMIF) or in more than one cell range (SUMIFS) are met. SUMIFS is new in Excel 2007.

To take full advantage of naming, I name both column A (Year) and column B (Region) in the same way I named column C.

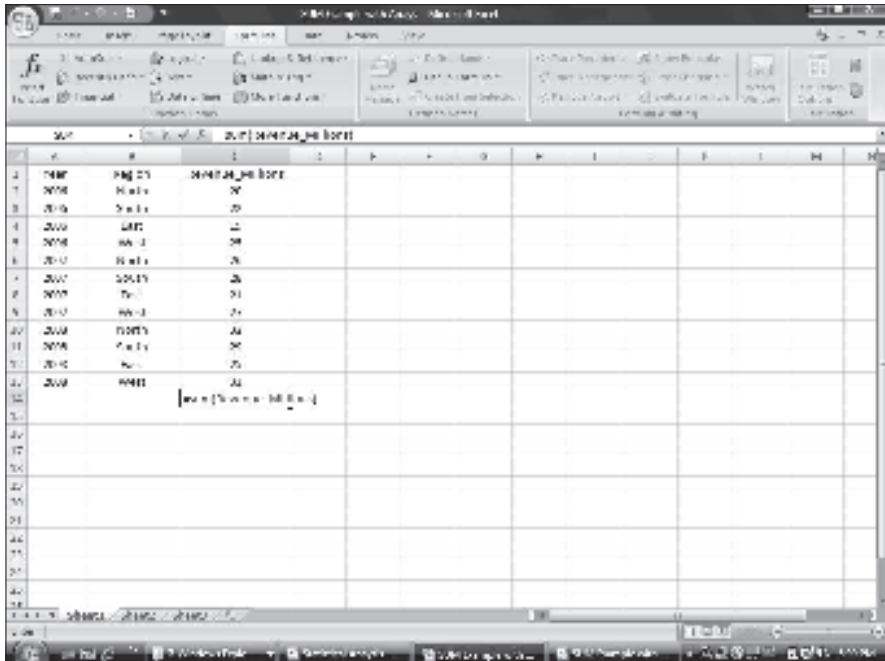


Figure 2-15: Completing the formula.



When you define a name for a cell range like B2:B13 in this example, beware: Excel can be a bit quirky when the cells hold names. Excel might guess that the name in the uppermost cell is the name you want to assign to the cell range. In this case, Excel guesses “North” for the name, rather than “Region.” If that happens, you make the change in the New Name dialog box.

To keep track of the names in a worksheet, selecting

Formula | Name Manager

opens the Name Manager box shown in Figure 2-16. The nearby buttons in the Defined Names area are also useful.

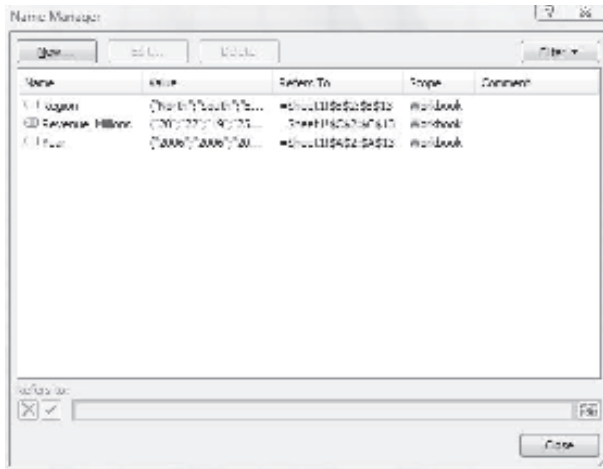


Figure 2-16:
Managing
the Defined
Names in a
worksheet.

Next, I sum the data in column C, but only for the North Region. That is, I only consider a cell in column C if the corresponding cell in column B contains “North.” To do this, I followed these steps:

- 1. Select a cell for the formula result.**

My selection here is C15.

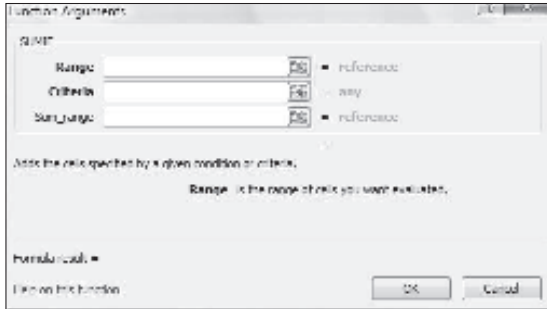
- 2. Select the appropriate formula category and choose your function from its pop-up menu.**

This opens the Function Arguments dialog box.

I selected Formulas | Math & Trig

and scrolled down the menu to find and choose SUMIF. This selection opens the Function Arguments dialog box shown in Figure 2-17.

Figure 2-17:
The
Function
Arguments
dialog box
for SUMIF.

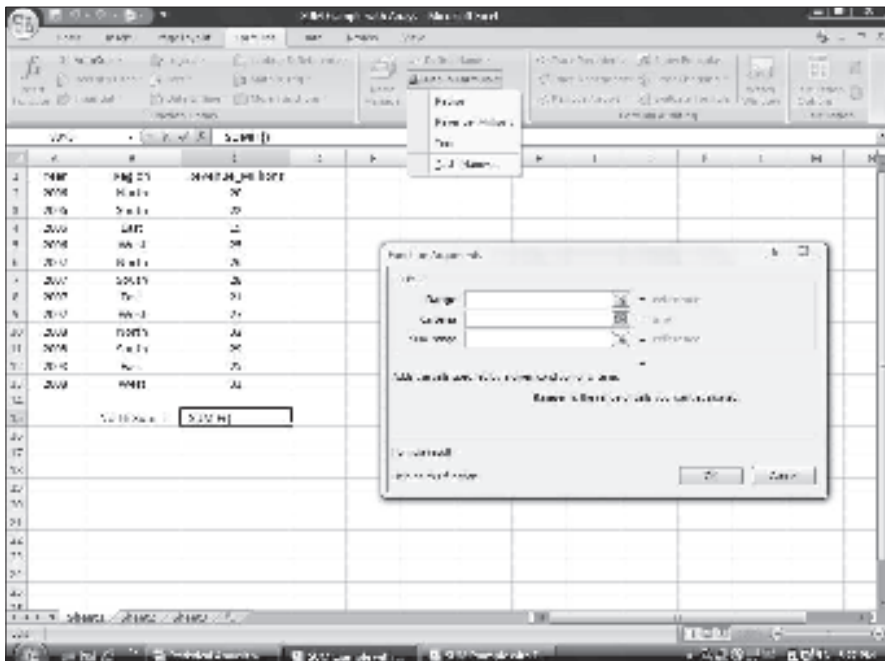


SUMIF has three arguments. The first, Range, is the range of cells to evaluate for the condition to include in the sum (North, South, East, or West in this example). The second, Criteria, is the specific value in the Range (North, for this example). The third, Sum_range, holds the values I sum.

3. In the Function Arguments dialog box, enter the appropriate values for the arguments.

Here's where another Defined Names button comes in handy. In that Ribbon area, click the down arrow next to Use in Formula to open the drop-down list shown in Figure 2-18.

Figure 2-18:
The Use
In Formula
drop-down
list.

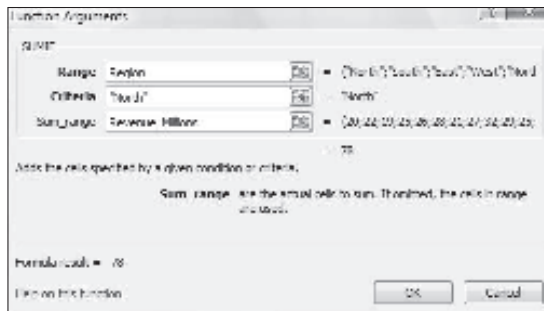


Selecting from this list fills in the Function Arguments dialog box, as shown in Figure 2-19. I had to type North into the Criteria box. Excel adds the double quotes.

4. Click OK.

The result appears in the selected cell. For this example, that's 78.

Figure 2-19:
Completing
the Function
Arguments
dialog box
for SUMIF.



In the formula bar,

```
=SUMIF(Region,"North", Revenue_Millions)
```

appears. I can type it exactly that way into the formula bar, without the dialog box or the drop-down list.

The formula in the formula bar is easier to understand than

```
= SUMIF(B2:B13,"North", C2:C13)
```

isn't it?

Incidentally, the same cell range can be both the Range and the Sum_range. For example, to sum just the cells for which Revenue_Millions is less than 25, that's

```
=SUMIF(Revenue_Millions, "< 25", Revenue_Millions)
```

The second argument (Criteria) is always in double-quotes.

What about SUMIFS? That one is useful if I want to find the sum of revenues for North but only for the years 2006 and 2007. Follow these steps to use SUMIFS to find this sum:

1. Select a cell for the formula result.

The selected cell is C17.

2. Select the appropriate formula category and choose your function from its pop-up menu.

This opens the Function Arguments dialog box.

For this example, the selection is SUMIFS from the

Formulas | Math & Trig

menu, opening the Functions Arguments dialog box shown in Figure 2-20.

3. In the Function Arguments dialog box, enter the appropriate values for the arguments.

Notice that in SUMIFS the Sum_range argument appears first. In SUMIF, it appears last. The appropriate values for the arguments appear in Figure 2-20.

4. The formula in the Formula bar is

```
=SUMIFS(Revenue_Millions,Year,"<2008",Region,"North")
```

5. Click OK.

The answer, 46, appears in the selected cell.

With unnamed arrays, the formula would have been

```
=SUMIFS(C2:C13,A2:A13,"<2008",B2:B13,"North")
```

which seems much harder to comprehend.

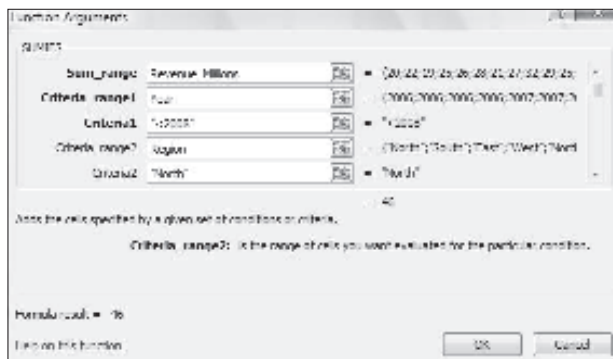


Figure 2-20:
The Completed Function Arguments dialog box for SUMIFS.



A defined name involves absolute referencing. (See Chapter 1.) Therefore, if you try to autofill from a named array, you'll be in for an unpleasant surprise: Rather than autofilling a group of cells, you'll be copying a value over and over again.

Here's what I mean. Suppose you assign the name `Series_1` to `A2:A11` and `Series_2` to `B2:B11`. In `A12`, you calculate `SUM(Series_1)`. Being clever, you figure you'll just drag the result from `A12` to `B12` to calculate `SUM(Series_2)`. What do you find in `B12`? `SUM(Series_1)`, that's what.

Creating your own array formulas

In addition to Excel's built-in array formulas, you can create your own. To help things along, you can incorporate named arrays.

Figure 2-21 shows two named arrays, `X` and `Y` in columns `C` and `D`. `X` refers to `C2` through `C5` (not `C1` through `C5`!) and `Y` refers to `D2` through `D5` (not `D1` through `D5`!) `XY` is the column header for column `F`. Each cell in column `F` will store the product of the corresponding cell in column `C` and the corresponding cell in column `D`.

	A	B	C	D	E	F
1			X	Y		XY
2			12	8		
3			14	9		
4			15	10		
5			16	11		
6						
7						

Figure 2-21:
Two named
arrays.

An easy way to enter the products, of course, is to just set `F2` equal to `C2*E2` and then autofill the remaining applicable cells in column `F`.

Just to illustrate array formulas, though, I follow these steps to work on the data in the worksheet in Figure 2-21.

- 1. Select the array that will hold the answers to the array formula.**

That would be `F2` through `F5`, or `F2:F5` in Excel-speak. Figure 2-21 shows the array selected.

- 2. Into the selected array, type the formula.**

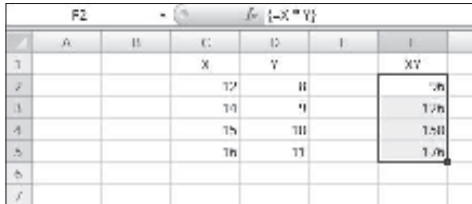
The formula here is `=X * Y`

3. Press Ctrl+Shift+Enter (not Enter).

The answers appear in F2 through F5, as Figure 2-22 shows. Note the formula $\{=X*Y\}$

in the formula bar. As I told you earlier, the curly brackets indicate an array formula.

Figure 2-22:
The results of the array formula $\{=X * Y\}$.



Another thing I mention earlier in this chapter: When you name a range of cells, make sure that the named range does *not* include the cell with the name in it. If it does, an array formula like $\{=X * Y\}$ tries to multiply the letter X by the letter Y to produce the first value, which is impossible and results in the exceptionally ugly #VALUE! error.

Using data analysis tools

Excel has a set of sophisticated tools for data analysis. Table 2-1 lists the tools I cover. (The one I don't cover, Fourier Analysis, is extremely technical.) Some of the terms in the table may be unfamiliar to you, but you'll know them by the time you finish this book.

Table 2-1 Excel's Data Analysis Tools	
<i>Tool</i>	<i>What It Does</i>
Anova: Single Factor	Analysis of variance for two or more samples
Anova: Two Factor with Replication	Analysis of variance with two independent variables, and multiple observations in each combination of the levels of the variables
Anova: Two Factor without Replication	Analysis of variance with two independent variables, and one observation in each combination of the levels of the variables

(continued)

<i>Tool</i>	<i>What It Does</i>
Correlation	With more than two measurements on a sample of individuals, calculates a matrix of correlation coefficients for all possible pairs of the measurements
Covariance	With more than two measurements on a sample of individuals, calculates a matrix of covariances for all possible pairs of the measurements
Descriptive Statistics	Generates a report of central tendency, variability, and other characteristics of values in the selected range of cells
Exponential Smoothing	In a sequence of values, calculates a prediction based on a preceding set of values, and on a prior prediction for those values
F-Test Two Sample for Variances	Performs an F-test to compare two variances
Histogram	Tabulates individual and cumulative frequencies for values in the selected range of cells
Moving Average	In a sequence of values, calculates a prediction which is the average of a specified number of preceding values
Random Number Generation	Provides a specified amount of random numbers generated from one of seven possible distributions
Rank and Percentile	Creates a table that shows the ordinal rank and the percentage rank of each value in a set of values
Regression	Creates a report of the regression statistics based on linear regression through a set of data containing one dependent variable and one or more independent variables
Sampling	Creates a sample from the values in a specified range of cells
t-Test: Two Sample	Three t-test tools test the difference between two means. One assumes equal variances in the two samples. Another assumes unequal variances in the two samples. The third assumes matched samples.
z-Test: Two Sample for Means	Performs a two-sample z-test to compare two means when the variances are known

In order to use these tools, you first have to load them into Excel.

To start, click the Office Button and select Excel Options. Doing this opens the Excel Options dialog box. Then follow these steps:

1. In the Excel Options dialog box, select Add-Ins.

Oddly enough, this opens a list of add-ins.

2. Near the bottom of the list, you see a drop-down list labeled Manage. From this list, select Excel Add-Ins.

3. Click Go.

This opens the Add-Ins dialog box. (See Figure 2-23.)

4. Click the check box next to Analysis Toolpak and then click OK.



Figure 2-23:
The Add-Ins
dialog box.

When Excel finishes loading the Toolpak, you'll find a Data Analysis button in the Analysis area of the Data tab. In general, the steps for using a data analysis tool are:

1. Enter your data into an array.

2. Click Data | Data Analysis to open the Data Analysis dialog box.

3. In the Data Analysis dialog box select the data analysis tool you want to work with.

4. Click OK (or just double-click the selection) to open the dialog box for the selected tool.

5. In the tool's dialog box, enter the appropriate information.

I know this sounds like a cop-out, but each tool is different.

6. Click OK to close the dialog box and see the results.

Here's an example to get you accustomed to using these tools. In this example, I go through the Descriptive Statistics tool. This tool calculates a number of statistics that summarize a set of scores.

1. Enter your data into an array.

Figure 2-24 shows an array of numbers in cells B2 through B9, with a column header in B1.

2. Click Data | Data Analysis to open the Data Analysis dialog box.

3. Click Descriptive Statistics and click OK (or just double-click Descriptive Statistics) to open the Descriptive Statistics dialog box.

4. Identify the data array.

In the Input Range box, enter the cells that hold the data. For this example, that's B1 through B9. The easiest way to do this is to move the cursor to the top cell (B1), press the Shift key, and click the bottom cell (B9). That puts the absolute reference format `B1:B9` into Input Range.

5. Click the Columns radio button to indicate that the data are organized by columns.

6. Check the Labels in First Row checkbox, because the Input Range includes the column heading.

7. Click the New Worksheet Ply radio button, if it isn't already selected.

This tells Excel to create a new tabbed sheet within the current worksheet, and to send the results to the newly created sheet.

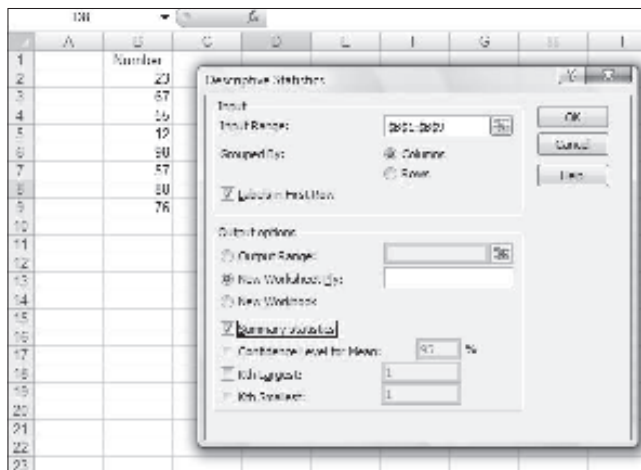


Figure 2-24:
Working
with the
Descriptive
Statistics
Analysis
tool.

8. Click the **Summary Statistics** checkbox and leave the others unchecked. Click **OK**.

The new tabbed sheet (ply) opens, displaying statistics that summarize the data. Figure 2-25 shows the new ply, after I widened Column A.

Figure 2-25:
The output of the Descriptive Statistics Analysis tool.

	number
Mean	58.9
Standard Error	10.55498
Median	62
Mode	49.4
Standard Deviation	29.0202
Sample Variance	841.1729
Kurtosis	0.68084
Skewness	-0.50676
Range	86
Minimum	12
Maximum	98
Sum	476
Count	8

For now, I won't tell you the meaning of each individual statistic in the Summary Statistics display. I leave that for Chapter 7 when I delve more deeply into descriptive statistics.

Accessing Commonly Used Functions

Need quick access to a few commonly used Statistical functions? You can get to AVERAGE, MIN (minimum value in a selected cell range), and MAX (maximum value in a selected range) by clicking the down arrow next to a button on the Home tab. Clicking this down arrow also gets you to the Mathematical functions SUM and COUNT NUMBERS (counts the numerical values in a cell range).

For some reason, this button is in the Editing area. It's labeled Σ . Figure 2-26 shows you exactly where it is and the menu its down arrow opens.

By the way, if you just click the button

Home | Σ

and not the down arrow, you get SUM.

The last selection on that menu is yet another way to open the Insert Function dialog box.

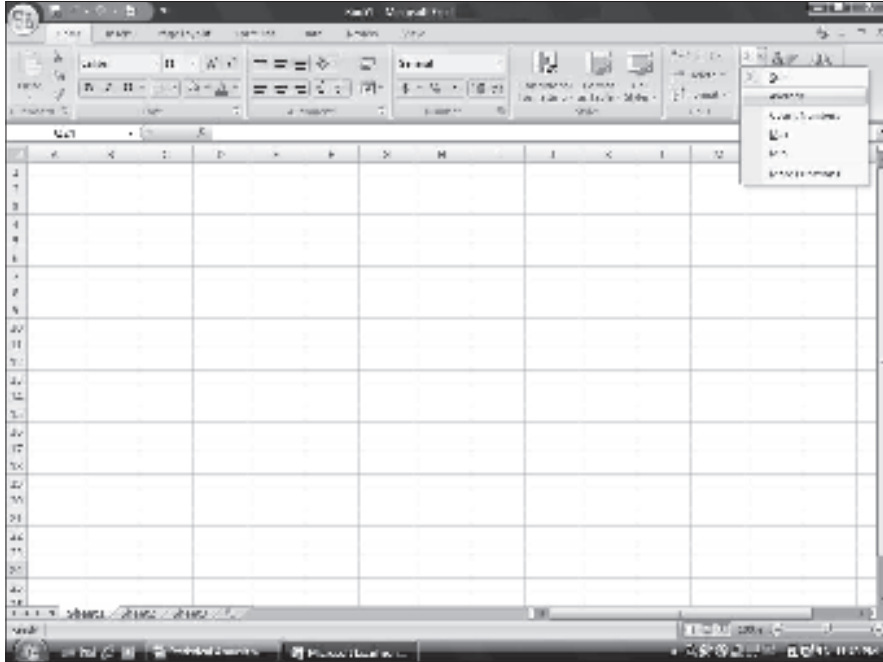


Figure 2-26:
The Home
| Σ button
and the
menu its
down arrow
opens.



One nice thing about using this menu — it eliminates a step: When you select a function, you don't have to select a cell for the result. Just select the cell range and the function inserts the value in a cell immediately after the range.