

## Chapter 7

# Summarizing It All

---

### *In This Chapter*

- ▶ Things great and small
  - ▶ Symmetry, peaks, and plateaus
  - ▶ Getting descriptive
  - ▶ Statistics served up on a tray
- 

**M**easures of central tendency and variability are excellent ways of summarizing a set of scores. They aren't the only ways. Central tendency and variability make up a subset of descriptive statistics. Some descriptive statistics are intuitive — like count, maximum, and minimum. Some are not — like skewness and kurtosis.

In this chapter, I discuss descriptive statistics, and I show you Excel's capabilities for calculating them and visualizing them.

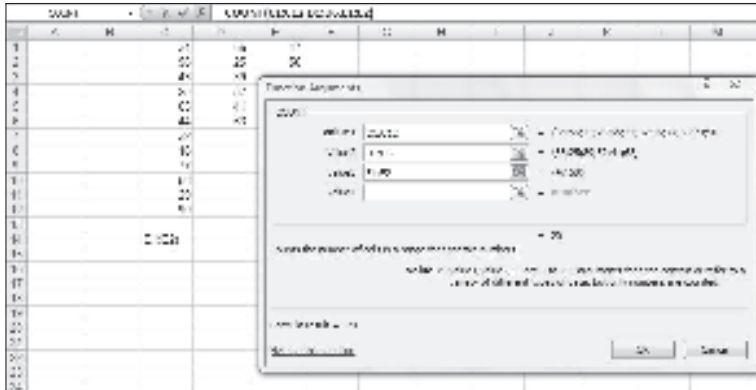
## *Counting Out*

The most fundamental descriptive statistic I can imagine is the number of scores in a set of scores. Excel offers five ways to determine that number. Yes, five ways. Count them.

### ***COUNT, COUNTA, COUNTBLANK, COUNTIF, COUNTIFS***

Given an array of cells, COUNT gives you the amount of those cells that contain numerical data. Figure 7-1 shows that I've entered a group of scores, selected a cell to hold COUNT's result, and opened the Function Arguments dialog box for COUNT. Here are the steps:

**Figure 7-1:**  
The  
Function  
Arguments  
Dialog Box  
for COUNT,  
showing  
multiple  
arguments.



**1. Enter your data into the worksheet and select a cell for the result.**

I entered data into columns C, D, and E to show off COUNT's multi-argument capability. I selected cell C14 to hold the count.

**2. From the Statistical Functions menu, select COUNT and click OK to open the Function Arguments dialog box for COUNT.**

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

In the Number1 box I entered one of the data columns for this example, like C1:C12.

I clicked in the Number2 box and entered another data column. I entered D1:D6.

I clicked in the Number3 box and entered the last column, which in this example is E1:E2.

**4. Click OK to put the result in the selected cell.**

COUNTA works like COUNT, except that its tally includes cells that contain text and logical values in its tally.

COUNTBLANK counts the number of blank cells in an array. In Figure 7-2, I use the numbers from the preceding example, but I extend the array to include cells D7 through D12 and E3 through E12. The array in the Range box is C1:E12. The Argument Functions dialog box for COUNTBLANK shows the number of blank cells (16 for this example).

**Figure 7-2:**  
COUNT  
BLANK  
tallies the  
blank cells  
in a speci-  
fied array.



COUNTIF shows the number of cells whose value meets a specified criterion. Figure 7-3 reuses the data once again, showing the Arguments Function dialog box for COUNTIF. Although the range is C1:E12, COUNTIF doesn't include blank cells.

**Figure 7-3:**  
COUNTIF  
tallies the  
amount of  
cells whose  
data meet  
a specified  
criterion.



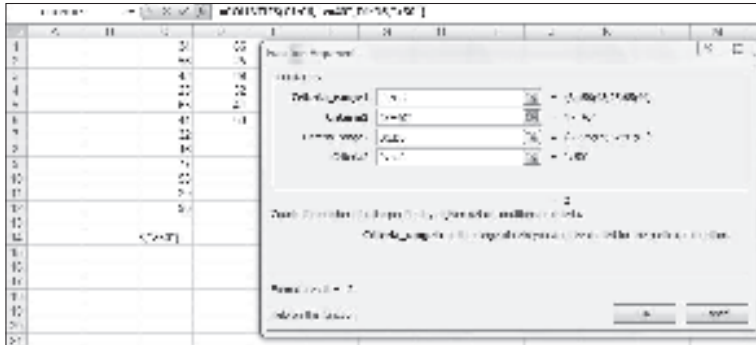
The criterion I used,  $\geq 89$ , tells COUNTIF to count only the cells whose values are greater than or equal to 89. For this example, that count is 1.



This probably won't make much difference as you use this function, but a little quirk of Excel shows up here. If you put double quotes around the criterion, the result appears in the dialog box before you click OK. If you don't, it doesn't. If you click OK without quoting, Excel supplies the quotes and the result appears in the selected cell, and Excel applies the quotes.

COUNTIFS is new in Excel 2007. This function can use multiple criteria to determine the count. If the criteria come from two arrays, they must have the same number of cells. This is because COUNTIFS counts pairs of cells. It includes a pair of cells in the count if one of the cells meets a criterion *and* the other meets a criterion. Take a look at Figure 7-4.

**Figure 7-4:**  
Working  
with  
COUNTIFS.



In this example, COUNTIFS operates in C1:C6 and D1:D6. The criterion for the cells in column C is  $\geq 40$ . The criterion for the cells in column D is  $> 50$ . This means that COUNTIFS counts cell-pairs whose C cell holds a value greater than or equal to 40 and whose D cell holds a value greater than 50. Only two cell-pairs meet these conditions, as the dialog box shows.

You can use a cell range more than once in COUNTIFS. For example

```
=COUNTIFS(C1:C12, ">30", C1:C12, "<60")
```

gives the number of cells in which the value is between 30 and 60 (not including 30 and 60).

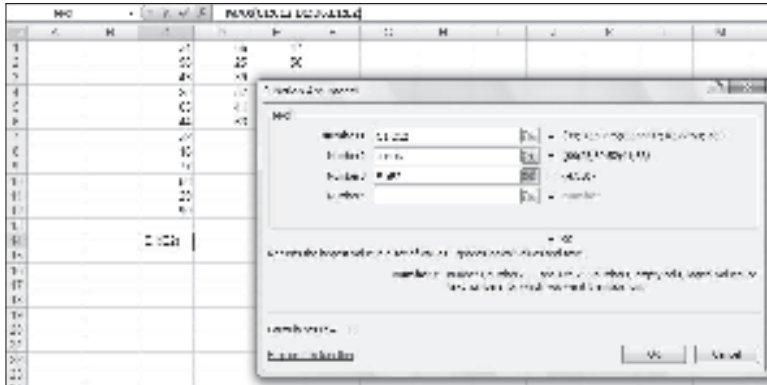
## The Long and Short of It

Two more descriptive statistics that probably require no introduction are the maximum and the minimum. These, of course, are the largest value and the smallest value in a group of scores.

### MAX, MAXA, MIN, and MINA

Excel has worksheet functions that determine a group's largest and smallest values. I show you what MAX is all about. The others work in a similar fashion.

Figure 7-5 reuses the scores from the preceding examples. I selected a cell to hold their maximum value, and opened the Function Arguments dialog box for MAX. Here are the steps:



**Figure 7-5:**  
The  
Function  
Arguments  
dialog box  
for MAX.

**1. Type your data into the worksheet and select a cell to hold the result.**

I entered data into columns C, D, and E to show off MAX's multi-argument capability. For this example, I selected cell C14.

**2. From the Statistical Functions menu select MAX to open the Function Arguments dialog box for MAX.**

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

In the Number1 box, I entered one of the data columns, C1:C12.

Clicking the Number2 box creates and opens the Number3 box. In the Number2 box, I entered another array, D1:D6.

I clicked in the Number3 box and entered the last array, E1:E2.

**4. Click OK to put the result in the selected cell.**

MAX ignores any text or logical values it encounters along the way. MAXA takes text and logical values into account when it finds the maximum. If MAXA encounters the logical value TRUE, it converts that value to 1. MAXA converts FALSE, or any text other than "TRUE", to 0.

MIN and MINA work the same way as MAX and MAXA, except that they find the minimum rather than the maximum. Take care when you use MINA, because the conversions of logical values and text to 0 and 1 influence the result. With the numbers in the preceding example, the minimum is 22. If you enter FALSE or other text into a cell in any of the arrays, MINA gives 0 as the minimum. If you enter TRUE, MINA gives 1 as the minimum.

## Getting Esoteric

In this section, I discuss some little-used statistics that are related to the mean and the variance. For most people, the mean and the variance are enough to describe a set of data. These other statistics, *skewness* and *kurtosis*, go just a bit further. You might use them someday if you have a huge set of data and you want to provide some in-depth description.

Think of the mean as *locating* a group of scores by showing you where their center is. This is the starting point for the other statistics. With respect to the mean

- ✓ The variance tells you how *spread out* the scores are.
- ✓ Skewness indicates how *symmetrically* the scores are distributed.
- ✓ Kurtosis shows you whether or not your scores are distributed with a *peak* in the neighborhood of the mean.



Skewness and kurtosis are related to the mean and variance in fairly involved mathematical ways. The variance involves the sum of squared deviations of scores around the mean. Skewness depends on cubing the deviations around the mean before you add them all up. Kurtosis takes it all to a higher power — the fourth power, to be exact. I get more specific in the subsections that follow.

### SKEW

Figure 7-6 shows three histograms. The first is symmetric, the other two are not. The symmetry and the asymmetry are reflected in the skewness statistic.

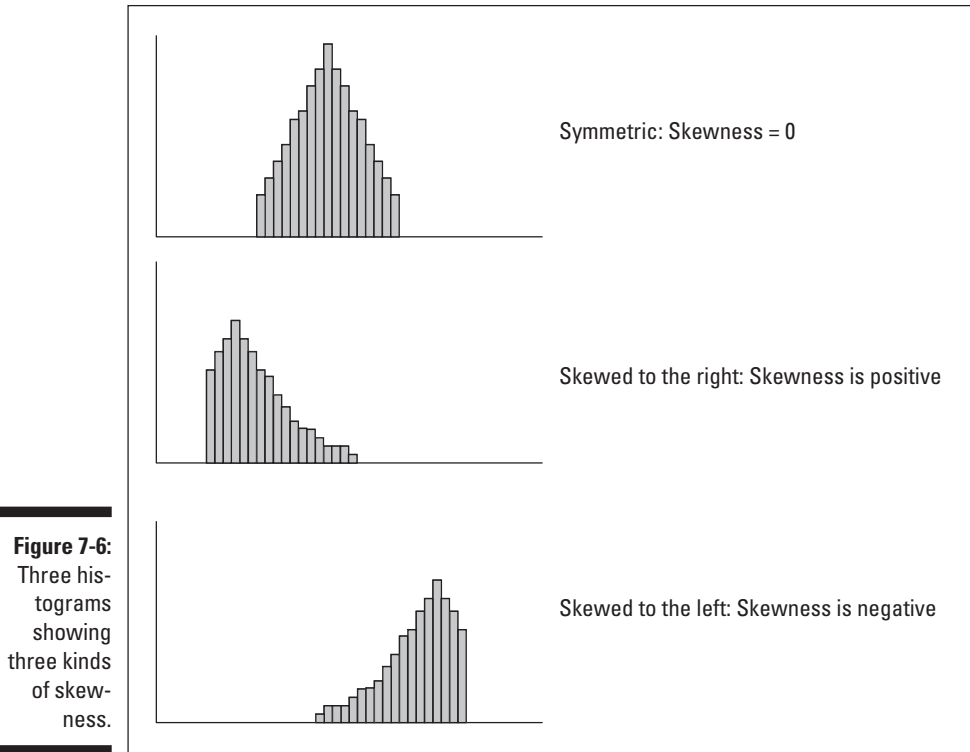
For the symmetric histogram, the skewness is 0. For the second histogram — the one that tails off to the right — the value of the skewness statistic is positive. It's also said to be skewed to the right. For the third histogram (which tails off to the left), the value of the skewness statistic is negative. It's also said to be skewed to the left.



Where do zero, positive, and negative skew come from? They come from this formula:

$$skewness = \frac{\sum (X - \bar{X})^3}{(N - 1)s^3}$$

In the formula,  $\bar{X}$  is the mean of the scores,  $N$  is the number of scores, and  $s$  is the standard deviation.



**Figure 7-6:** Three histograms showing three kinds of skewness.

I include this formula for completeness. If you're ever concerned with skewness, you probably won't use this formula anyway because Excel's SKEW function does the work for you.

To use SKEW:

**1. Type your numbers into a worksheet and select a cell for the result.**

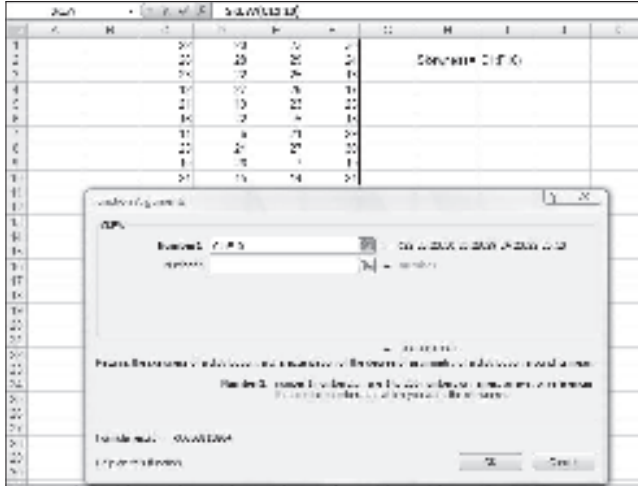
For this example, I've entered scores into the first ten rows of columns C, D, E, and F. (See Figure 7-7.) I selected cell I2 for the result.

**2. From the Statistical Functions menu, select SKEW to open the Function Arguments dialog box for SKEW.**

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

In the Number1 box, enter the array of cells that holds the data. For this example, the array is C1:F10. With the data array entered, the Function Arguments dialog box shows the skewness, which for this example is negative.

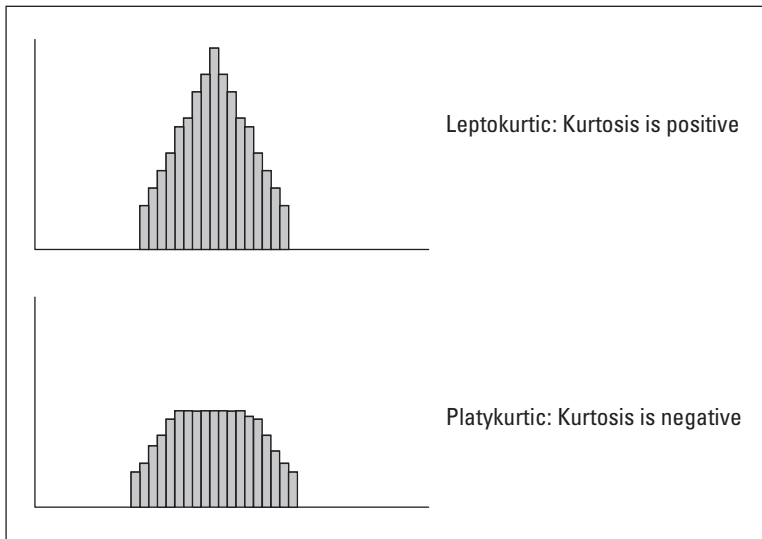
**4. Click OK to put the result into the selected cell.**



**Figure 7-7:**  
Using the  
SKEW  
function to  
calculate  
skewness.

## KURT

Figure 7-8 shows two histograms. The first has a peak at its center, the second is flat. The first is said to be *leptokurtic*. Its kurtosis is positive. The second is *platykurtic*. Its kurtosis is negative.



**Figure 7-8:**  
Two his-  
tograms  
showing  
two kinds  
of kurtosis.



Negative? Wait a second. How can that be? I mentioned earlier that kurtosis involves the sum of fourth powers of deviations from the mean. Because four is an even number, even the fourth power of a negative deviation is positive. If you're adding all positive numbers, how can kurtosis ever be negative?

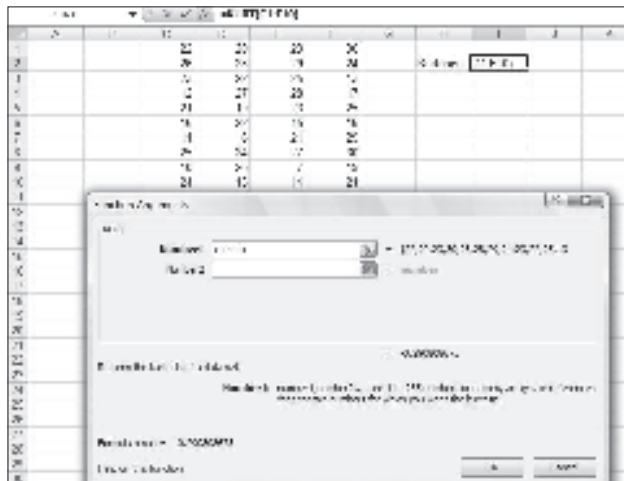
Here's how. The formula for kurtosis is

$$kurtosis = \frac{\sum (X - \bar{X})^4}{(N - 1)s^4} - 3$$

where  $\bar{X}$  is the mean of the scores,  $N$  is the number of scores, and  $s$  is the standard deviation.

Uh . . . why 3? The 3 comes into the picture because that's the kurtosis of something special called the *standard normal distribution*. (I discuss the normal distribution at length in Chapter 8.) Technically, statisticians refer to this formula as *kurtosis excess* — meaning that it shows the kurtosis in a set of scores that's in excess of the standard normal distribution's kurtosis. If you're about to ask the question “Why is the kurtosis of the standard normal distribution equal to 3?” don't ask.

This is another formula you'll probably never use because Excel's KURT function takes care of business. Figure 7-9 shows the scores from the preceding example, a selected cell, and the Function Arguments dialog box for KURT.



**Figure 7-9:**  
Using KURT  
to calculate  
kurtosis.

To use KURT:

**1. Enter your numbers into a worksheet and select a cell for the result.**

For this example, I entered scores into the first ten rows of columns C, D, E, and F. I selected cell I2 for the result.

**2. From the Statistical Functions menu, select KURT to open the Function Arguments dialog box for KURT.**

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

In the Number1 box, I entered the array of cells that holds the data. Here, the array is C1:F10. With the data array entered, the Function Arguments dialog box shows the kurtosis, which for this example is negative.

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

## *Tuning In the Frequency*

Although the calculations for skewness and kurtosis are all well and good, it's helpful to see how the scores are distributed. To do this, you create a *frequency distribution*, a table that divides the possible scores into intervals and shows the number (the frequency) of scores that fall into each interval.

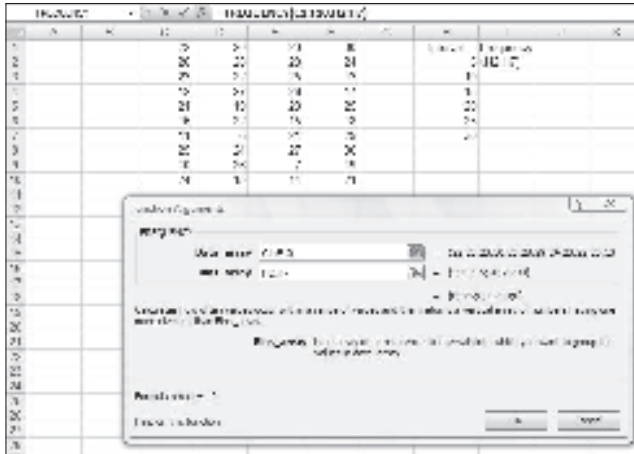
Excel gives you two ways to create a frequency distribution. One is a worksheet function, the other is a data analysis tool.

## ***FREQUENCY***

I show you the FREQUENCY worksheet function in Chapter 2 when I introduce array functions. Here, I give you another look. In the upcoming example, I reuse the data from the skewness and kurtosis discussions so you can see what the distribution of those scores looks like.

Figure 7-10 shows the data once again, along with a selected array, labeled Frequency. I've also added the label Intervals to a column, and in that column I put the interval boundaries. Each number in that column is the upper bound of an interval. The figure also shows the Function Arguments dialog box for FREQUENCY.

**Figure 7-10:**  
Finding the  
frequencies  
in an array  
of cells.



This is an array function, so the steps are a bit different from the functions I showed you so far in this chapter.

**1. Enter the scores into an array of cells.**

The array, as in the preceding examples is C1:F10.

**2. Enter the intervals into an array.**

I entered 5, 10, 15, 20, 25, and 30 into H2:H7.

**3. Select an array for the resulting frequencies.**

I put Frequency as the label at the top of column I, so I selected I2 through I7 to hold the resulting frequencies.

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

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

In the Data\_array box I entered the cells that hold the scores. In this example, that's C1:F10.

FREQUENCY refers to intervals as “bins,” and holds the intervals in the Bins\_array box. For this example, H2:H7 goes into the Bins\_array box.

After I identified both arrays, the Function Arguments dialog box shows the frequencies inside a pair of curly brackets. Look closely at Figure 7-10 and you see that Excel adds a frequency of zero to the end of the set of frequencies.

**6. Press Ctrl+Shift+Enter to close the Function Arguments dialog box.**

Use this keystroke combination because FREQUENCY is an array function.

When you close the Function Arguments dialog box, the frequencies go into the appropriate cells, as Figure 7-11 shows.

	A	B	C	D	E	F	G	H	I
1			22	20	23	20		interval	frequency
2			26	28	29	24		5	0
3			23	22	25	13		10	0
4			19	27	28	17		15	5
5			21	19	23	25		20	5
6			16	22	15	18		25	15
7			11	6	21	29		30	10
8			25	24	27	30			
9			10	26	7	19			
10			24	15	14	21			
11									

**Figure 7-11:**  
FREQUENCY's  
frequencies.



If I had assigned the name *Data* to C1:F10 and the name *Interval* to H2:H7, and used those names in the Function Arguments dialog box, the resulting formula would have been

```
=FREQUENCY (Data, Interval)
```

which might be easier to understand than

```
=FREQUENCY (C1 : F10, H2 : H7)
```

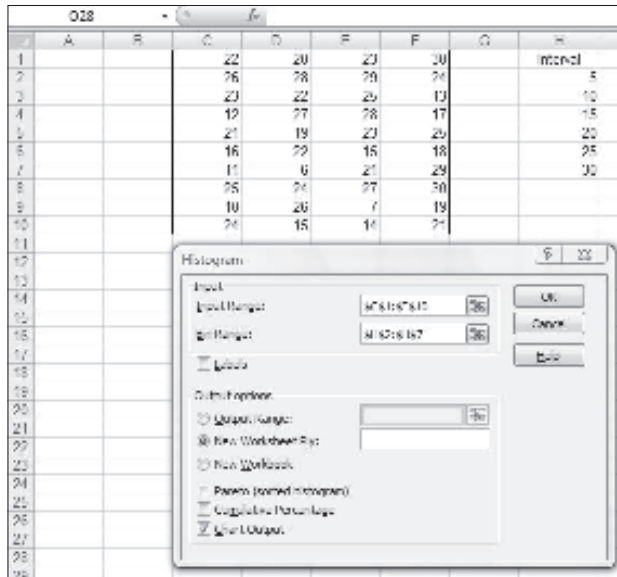
(Don't remember how to assign a name to a range of cells? Take another look at Chapter 2.)

## Data analysis tool: Histogram

Here's another way to create a frequency distribution — with the Histogram data analysis tool. To show you that the two methods are equivalent, I use the data from the FREQUENCY example. Figure 7-12 shows the data along with the Histogram dialog box.

The steps are:

1. Enter the scores into an array, and enter intervals into another array.
2. Click on **Data | Data Analysis** to open the Data Analysis dialog box.
3. From the Data Analysis dialog box, select **Histogram** to open the Histogram dialog box.



**Figure 7-12:**  
The  
Histogram  
analysis  
tool.

**4. In the Histogram dialog box, enter the appropriate values.**

The data are in cells C1 through F10, so C1:F10 goes into the Input Range box. The easiest way to enter this array is to click on C1, press and hold the Shift key, and then click F10. Excel puts the absolute reference format (\$C\$1:\$F\$10) into the Input Range box.

In the Bin Range box, I enter the array that holds the intervals. In this example, that's H2 through H7. I click on H2, press and hold the Shift key, and then click H7. The absolute reference format (\$H\$2:\$H\$7) appears in the Bin Range box.

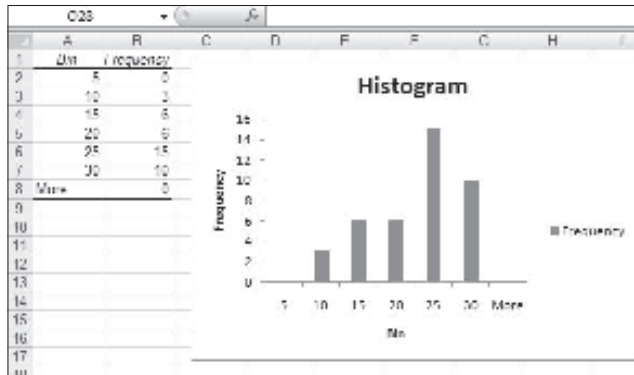
**5. Click the New Worksheet Ply radio button to create a new tabbed page and to put the results on the new page.**

**6. Click the Chart Output checkbox to create a histogram and visualize the results.**

**7. Click OK to close the dialog box.**

Figure 7-13 shows Histogram's output. The table matches up with what FREQUENCY produces. Notice that Histogram adds "More" to the Bin column. The size of the histogram is somewhat smaller when it first appears. I used the mouse to stretch the histogram and give it the appearance you see in the figure. The histogram shows that the distribution does tail off to the left (consistent with the negative skewness statistic) and seems to not have a distinctive peak (consistent with the negative kurtosis statistic).

**Figure 7-13:**  
The Histogram tool's output (after I stretched the chart).



By the way, the other checkbox options on the Histogram dialog box are Pareto chart and Cumulative percentage. The Pareto chart sorts the intervals in order from highest frequency to lowest before creating the graph. Cumulative percentage shows the percentage of scores in an interval combined with the percentages in all the preceding intervals. Checking this box also puts a cumulative percentage line in the histogram.

## *Can You Give Me a Description?*

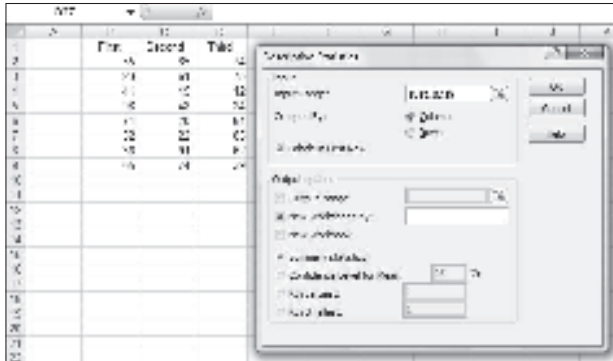
If you're dealing with individual descriptive statistics, the worksheet functions I've discussed get the job done nicely. If you want an overall report that presents just about all the descriptive statistical information in one place, use the Data Analysis tool I describe in the next section.

### *Data analysis tool: Descriptive Statistics*

In Chapter 2, I show you the Descriptive Statistics tool to introduce Excel's data analysis tools. Here's a slightly more complex example. Figure 7-14 shows three columns of scores and the Descriptive Statistics dialog box. I've labeled the columns First, Second, and Third so you can see how this tool incorporates labels.

Here are the steps for using this tool:

1. Enter the data into an array.



**Figure 7-14:**  
The  
Descriptive  
Statistics  
tool at work.

2. Select **Data | Data Analysis** to open the **Data Analysis** dialog box.
3. Choose **Descriptive Statistics** to open the **Descriptive Statistics** dialog box.
4. In the **Descriptive Statistics** dialog box, enter the appropriate values.
 

In the **Input Range** box, I enter the data. The easiest way to do this is to move the cursor to the upper-left cell (B1), press the Shift key, and click the lower-right cell (D9). That puts `$B$1:$D$9` 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 headings.
7. Click the **New Worksheet Ply** radio button to create a new tabbed sheet within the current worksheet, and to send the results to the newly created sheet.
8. Click the **Summary Statistics** checkbox, and leave the others unchecked.
9. Click **OK** to close the dialog box.

The new tabbed sheet (*ply*) opens, displaying statistics that summarize the data.

As Figure 7-15 shows, the statistics summarize each column separately. When this page first opens, the columns that show the statistic names are too narrow, so the figure shows what the page looks like after I widened the columns.

	A	B	C	D	E	F
1	First		Second		Third	
2						
3	Mean	55.5	Mean	55.75	Mean	55.875
4	Standard Error	8.343209627	Standard Error	8.49766	Standard Error	9.292062
5	Median	53.5	Median	54.5	Median	58
6	Mode	#N/A	Mode	#N/A	Mode	#N/A
7	Standard Deviation	23.50787072	Standard Deviation	25.85311	Standard Deviation	28.25616
8	Sample Variance	195.0571429	Sample Variance	221.6429	Sample Variance	250.4107
9	Kurtosis	0.28877548	Kurtosis	-1.38777	Kurtosis	-1.30336
10	Skewness	0.567263219	Skewness	0.12026	Skewness	0.65704
11	Range	75	Range	71	Range	73
12	Minimum	29	Minimum	20	Minimum	12
13	Maximum	58	Maximum	51	Maximum	55
14	Sum	444	Sum	446	Sum	426
15	Count	8	Count	8	Count	8
16						

**Figure 7-15:**  
The  
Descriptive  
Statistics  
tool's  
output.

The Descriptive Statistics tool gives values for these statistics: mean, standard error, median, mode, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, sum, and count. Except for standard error and range, I've discussed all of them.

Range is just the difference between the maximum and the minimum. Standard error is more involved, and I defer the explanation until Chapter 9. For now, I'll just say that standard error is the standard deviation divided by the square root of the sample size and leave it at that.

By the way, one of the checkboxes left unchecked in the example's Step 6 provides something called the *Confidence Limit of the Mean*, which I also defer until Chapter 9. The remaining two checkboxes, Kth Largest and Kth Smallest, work like the functions LARGE and SMALL.

## Instant Statistics

Suppose you're working with a cell range full of data. You might like to quickly know the status of the average and perhaps some other descriptive statistics about the data without going to the trouble of using several Statistical functions.

You can customize the Status bar at the bottom of the worksheet to track these values for you and display them whenever you select the cell range. To do this, right-click the status bar to open the Customize Status Bar menu. (See Figure 7-16.) In the area second from the bottom, checking all the items displays the values I mention in the preceding section (along with the count of items in the range — numerical and non-numerical).



