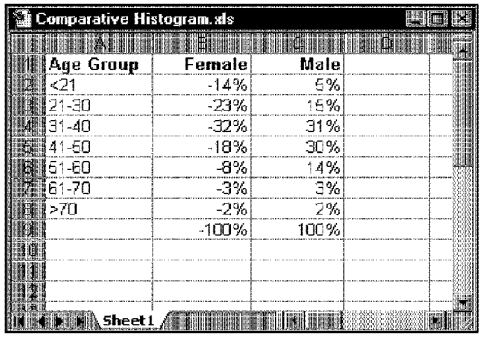
9. Select the first data series and access the Format Data Series dialog box. In the Patterns tab, set Border to None and Area to None. This makes the first data series invisible.
10. Apply other formatting, as desired.

Comparative Histograms

With a bit of creativity, you can create charts that you may have considered impossible with Excel. For example, Figure 16-35 shows data that was used to create the comparative histogram chart shown in Figure 16-36. Such charts often display population data.



Comparative Histogram.xls

	A	B	C
1	Age Group	Female	Male
2	<21	-14%	5%
3	21-30	-23%	15%
4	31-40	-32%	31%
5	41-50	-18%	30%
6	51-60	-8%	14%
7	61-70	-3%	3%
8	>70	-2%	2%
9		-100%	100%
10			
11			
12			

Sheet1

Figure 16-35: Data used in the comparative histogram chart.

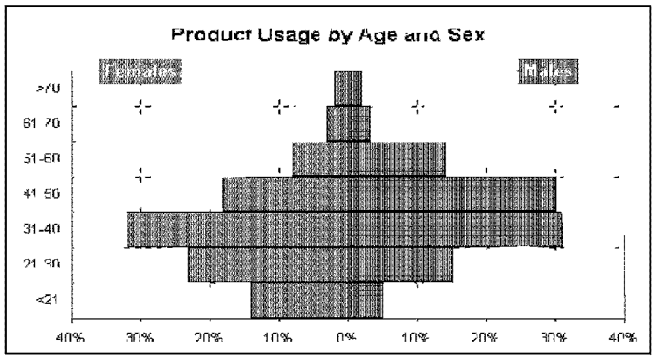


Figure 16-36: Producing this comparative histogram chart requires a few tricks.

A workbook that demonstrates this technique is available on the companion CD-ROM.

Here's how to create the chart:

1. Enter the data as shown in Figure 16-35. Notice that the values for females are entered as negative values.
2. Select A1:C8 and create a 2D bar chart. Use the subtype labeled Clustered Bar.
3. Apply the following custom number format to the horizontal axis: **0%;0%;0%**. This custom format eliminates the negative signs in the percentages.
4. Select the vertical axis and access the Format Axis dialog box. Click the Patterns tab and remove all tick marks. Set the Tick mark labels option to Low. This keeps the axis in the center of the chart but displays the axis labels at the left side.
5. Select either of the data series and then access the Format Data Series dialog box. Click the Options tab and set the Overlap to 100 and the Gap width to 0.
6. Add two text boxes to the chart (**Females** and **Males**), to substitute for the legend.
7. Apply other formatting, as desired.

Charts That Update Automatically

Earlier in this chapter, I discussed several ways to modify the data range used by a chart series. If you have a chart that displays daily sales, for example, you probably need to change the chart's data range each day when you add new data. Although updating a chart's data range isn't difficult, you might be interested in a trick that forces Excel to update the chart's data range whenever you add new data to your worksheet.

A workbook that demonstrates this technique is available on the companion CD-ROM.

To force Excel to update your chart automatically when you add new data, follow these steps:

1. Create the worksheet shown in Figure 16-37.
2. Select Insert• Name• Define to bring up the Define Name dialog box. In the Names in workbook field, enter **Date**. In the Refers to field, enter this formula:

   ```
   =OFFSET(Sheet1!$A$2,0,0,COUNTA(Sheet1! $A:$A)-1)
   ```

3. Click Add. Notice that the `OFFSET` function refers to the first data point (cell A2) and uses the `COUNTA` function to get the number of data points in the

column. Because column A has a heading in row 1, the formula subtracts 1 from the number.

4. Type **Sales** in the Names in workbook field, and in the Refers to field enter:

```
=OFFSET(Sheet1!$B$2,0,0,COUNTA(Sheet1!$B:$B)-1)
```

5. Click Add and then OK to close the dialog box.
6. Activate the chart and select the data series. In this example, the formula in the formula bar will read:

```
=SERIES(Sheet1!$B$1,Sheet1!$A$2:$A$10, Sheet1!$B$2:$B$10,1)
```

7. Replace the range references with the names that you defined in Steps 2 and 4. The formula should read:

```
=SERIES(,Sheet1!Date,Sheet1!Sales,1)
```

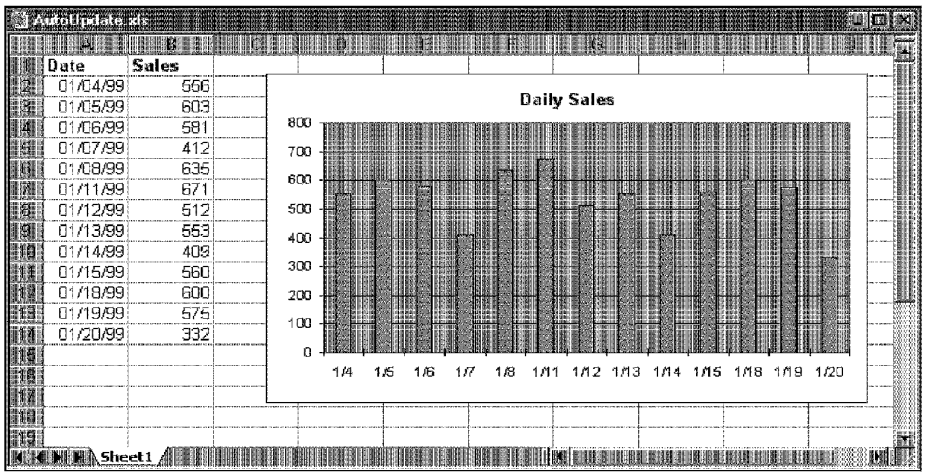


Figure 16-37: This chart is updated automatically whenever you add new data to columns A and B.

After you perform these steps, when you add data to columns A and B, the chart will be updated automatically to show the new data.

To use this technique for your own data, make sure that the first argument for the OFFSET function refers to the first data point, and that the argument for COUNTA refers to the entire column of data. Also, if the columns used for the data contain any other entries, COUNTA will return an incorrect value.

Summary

This chapter picks up where Chapter 13 left off by discussing most of the chart customization options in Excel. This chapter demonstrates how to create combination charts and your own custom chart formats—which let you apply a series of customizations with a single command. This chapter also discusses 3D charts, and concludes with several examples that use chart-making tricks.

CHAPTER 17

Creating Maps with Microsoft Map

In this Chapter

In previous chapters, you saw how you can use a chart to display data in a different — and, usually, more meaningful — way. This chapter explores the topic of mapping and describes how to present geographic information in the form of a map.

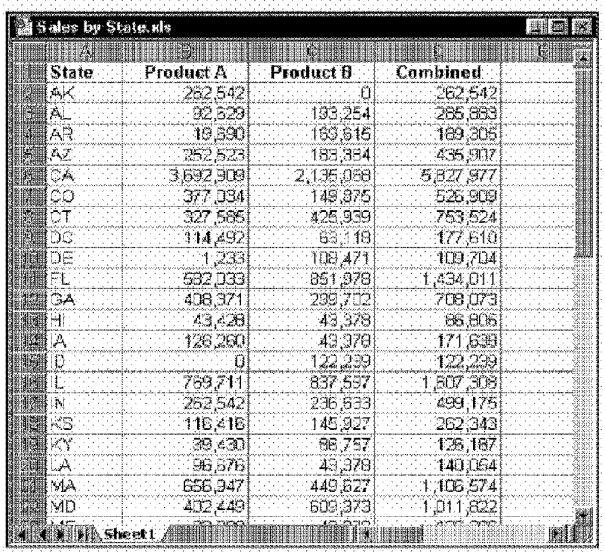
The mapping feature is not actually part of Excel. Rather, this feature uses an OLE server application named *Microsoft Map*, which was developed by MapInfo Corporation. You can use this application to insert maps into other Microsoft Office applications. Because the mapping application is not part of Excel, you'll find that the user interface is quite different from that of Excel. When a map is active, Microsoft Map menus and toolbars replace Excel's menus and toolbars.

Mapping: An Overview

Mapping, like charting, is a tool that visually presents data. People use maps for a variety of purposes, but the common factor in maps is that they work with data that has a basis in geography. If you classify information by state, province, or country, chances are good that you can represent the data on a map. For example, if your company sells its products throughout the United States, showing the annual sales for each state may be useful.

A Mapping Example

Figure 17-1 shows sales data for a company, with the data categorized by state. To understand this information, you would have to spend a lot of time examining the data.



Sales by State.xls

State	Product A	Product B	Combined
AK	262,542	0	262,542
AL	92,629	193,254	285,883
AR	19,390	169,616	189,305
AZ	252,523	183,384	435,907
CA	3,692,909	2,135,068	5,827,977
CO	377,034	149,875	526,909
CT	327,585	425,939	753,524
DC	114,492	63,118	177,610
DE	1,233	108,471	109,704
FL	582,033	851,978	1,434,011
GA	408,371	299,702	708,073
HI	43,428	43,378	86,806
A	128,260	43,378	171,638
D	0	122,259	122,259
L	789,711	837,597	1,807,308
N	262,542	236,633	499,175
KS	116,416	145,927	262,343
KY	39,430	86,757	126,187
LA	96,676	43,378	140,054
MA	656,947	449,627	1,106,574
MD	402,449	609,373	1,011,822

Sheet1

Figure 17-1: Raw data that shows sales by state.

Figure 17-2 shows the same data displayed in a chart. Although an improvement over the raw-data table, this type of presentation doesn't really work, because it has too many data points. In addition, the chart doesn't reveal any information about sales in a particular region.

Figure 17-3 shows the sales data presented as a map (it looks even better in color). This presentation uses different colors to represent various sales ranges. Looking at the map, you can see clearly that this company performs much better in some regions than in others.

The map in Figure 17-3 might be even more revealing if the sales were represented relative to the population of each state; that is, in per capita sales. This population data is available as a sample file on the Office CD (Mapstats.xls).

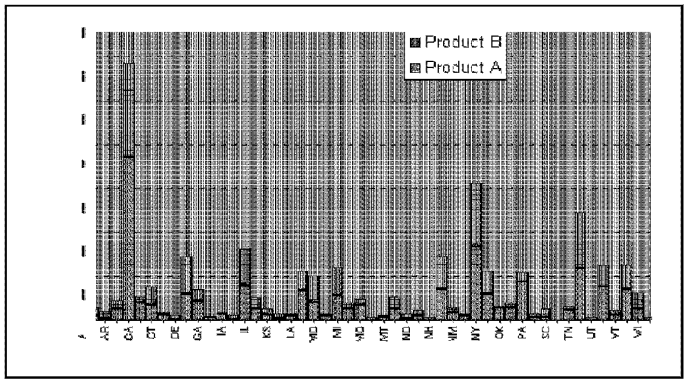


Figure 17-2: The sales data displayed in a chart.

Figure 17-4 shows the contents sheet for this workbook.

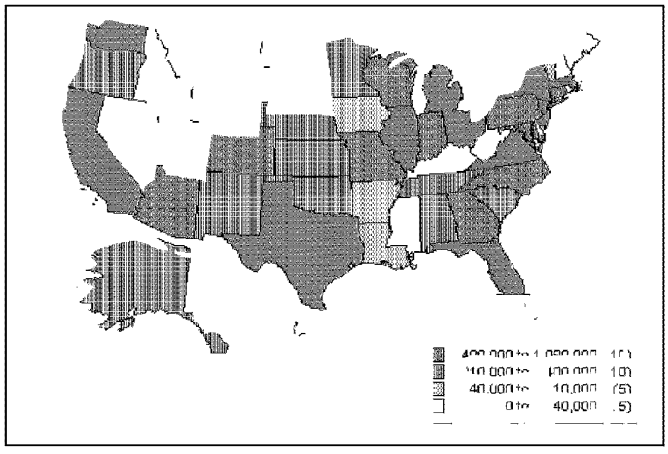


Figure 17-3: The sales data displayed in a map.

Available Maps

The Microsoft Map feature supports a good variety of maps and enables you to create maps in several different formats. A single map can display multiple sets of data, each in a different format. For example, your map can show sales by state and indicate the number of sales offices in each state. In addition, your map can display other accoutrements, such as labels and pin markers.

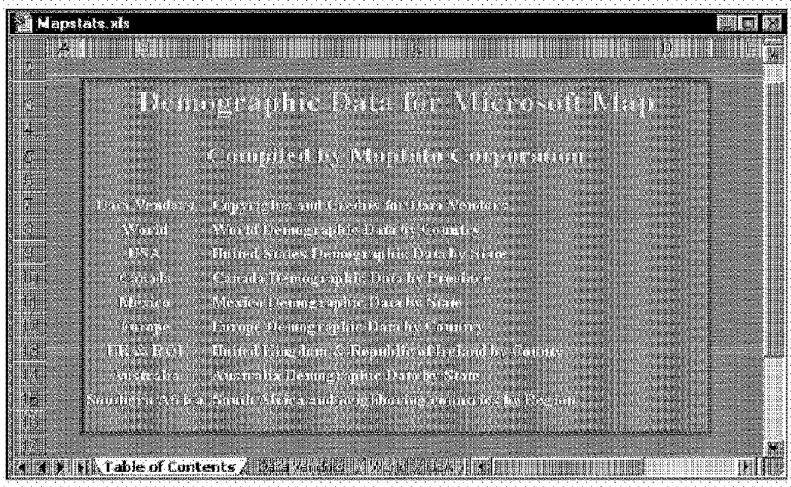


Figure 17-4: The Mapstats workbook contains population statistics that you can use in your maps.

The maps included with Microsoft Map are listed in Table 17-1. As you'll see later in this chapter, a map can be zoomed to display only a portion of it. Therefore, you can use the Europe map to zoom in on a particular region or country.

Table 17-1
Maps Included with Microsoft Map

Map	*Description*
Australia	The continent of Australia, by state
Canada	The country of Canada, by province
Europe	The continent of Europe, by country
Mexico	The country of Mexico, by state
North America	The countries of North America (Canada, U.S., Mexico)
U.K. Standard Regions	The countries of the United Kingdom, by region
U.S. in North America	United States (excluding Alaska and Hawaii insets), by state
U.S. with AK and HI Insets	United States (with Alaska and Hawaii insets), by state
World Countries	The world, by country

If you would like to order additional maps or data from MapInfo, you can contact the company directly or visit its Web site. For information on how to do so, activate a map and click the Help • About command.

Creating a Map

Creating a basic map with Microsoft Map is simple. In almost all cases, however, you'll want to customize the map. This section discusses the basics of mapmaking.

Setting Up Your Data

The Microsoft Map feature works with data stored in a list format (for an example, refer to Figure 17-1). The first column should contain names of map regions (such as states or countries). The columns to the right should contain data for each area. You can have any number of data columns, because you select the columns to use after the map is created.

Creating the Map

To create a map, start by selecting the data. The selection must include one column of area names and at least one column of data. If the columns have descriptive headers, include these in the selection.

Choose Insert • Map (or click the Map button on the Standard toolbar). Click and drag to specify the location and size of the map or just click to create a map of the default size. Unlike charts, maps must be embedded on a worksheet (there are no separate map sheets).

Microsoft Map analyzes the area labels and generates the appropriate map. If two or more maps are possible (or if you've developed any custom map templates), you'll see the Multiple Maps Available dialog box, shown in Figure 17-5. Select the map that you want to use from this list.

Figure 17-5: If multiple maps are available for your data, you can choose from this dialog box which map to use.

You don't have the Insert • Map command?

Excel's mapping feature is performed by an OLE server application. The mapping feature is not an integral part of Excel, and it may not be installed on your system. If you don't see a Map command on the Insert menu, you need to install the mapping feature.

To install the mapping feature, you need to rerun Excel's Setup program (or the Microsoft Office Setup program) and specify the mapping feature. Microsoft Map displays the map using the first column of data. It also displays the Microsoft Map Control dialog box, discussed later in the chapter. When the map is created, it is activated. Whenever a map is activated, Microsoft Map's menus and toolbar replace Excel's menus and toolbars. When you click outside the map, Excel's user interface reappears. You can reactivate a map by double-clicking it.

Setting the Map Format(s)

When a map first appears, the Microsoft Map Control dialog box is visible (see Figure 17-6). Use this dialog box to change the format of the selected map. You can use the Show/Hide Map Control tool to toggle the display of this dialog box.

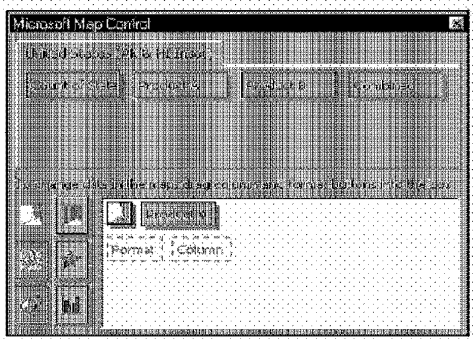


Figure 17-6: The Microsoft Map Control dialog box.

By default, you create maps by using the value-shading map format. You can change the format or display two or more formats on a single map. You use the Microsoft Map Control dialog box by dragging the items in it. The top of the dialog box displays all available data fields (which correspond to the columns that you selected when you created the map). The bottom part contains the map format information. Six format icons on the left determine the map format (described in the sections that follow). You combine a map format icon with one or more data fields by dragging the icon. For example, you can replace the default map format

icon with another one simply by dragging the new icon over the existing one. Some map formats use more than one data field. In such a case, you can drag additional data fields next to the icon.

To change options for a particular map format, either double-click the format icon or use the Map menu and choose the menu command that is appropriate for the format that you want to change. In either case, you get a dialog box that's appropriate for the map format.

The following sections include descriptions (and samples) of each map format supported by Microsoft Map.

Value shading

With this map format, each map region is shaded based on the value of its data. This format is appropriate for data-quantitative information, such as sales, population, and so on. Figure 17-7 shows an example of a map formatted with value shading (this map is zoomed to show only part of the U.S.). In this example, the sales are broken down into four ranges, and each sales range is associated with different shading.

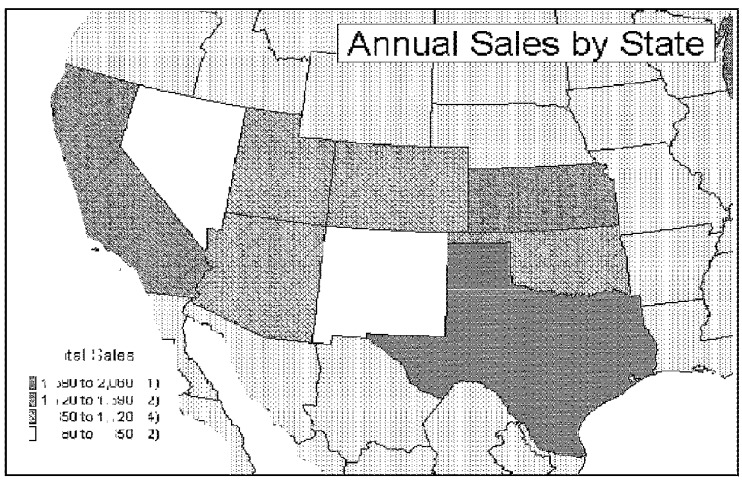


Figure 17-7: This map uses the value-shading format.

You can change the interval ranges in the Format Properties dialog box, shown in Figure 17-8.

When you're viewing a map that uses value shading, choose Map• Value Shading Options, and Microsoft Map opens the value-shading version of the Format Properties dialog box. You can specify the number of value ranges and the method of defining the ranges — an equal number of areas in each range or an equal spread of values in each range. You also can select a color for the shading. The map displays different variations of the single color that you select. You can choose the summary function to use (SUM or AVERAGE). To hide the format from the map, remove the check mark from the Visible check box.

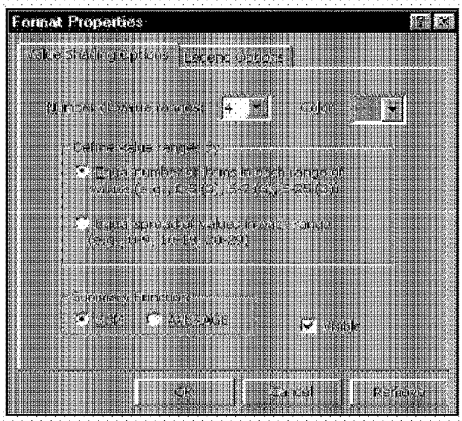


Figure 17-8: The Value Shading Options tab of the Format Properties dialog box.

Category shading

With the category-shading map format, each map region is colored based on a data value. The map legend has one entry (color) for every value of the data range. Therefore, this format is appropriate for data that has a small number of discrete values. For example, you can use the format to identify states that have a sales office, the number of sales reps in a country, and so on. A common use for this format is to identify the states that make up each sales region. Data need not be numeric. For example, the data can consist of text such as Yes and No.

Figure 17-9 shows a map that uses category shading to identify states that met the annual sales goal.

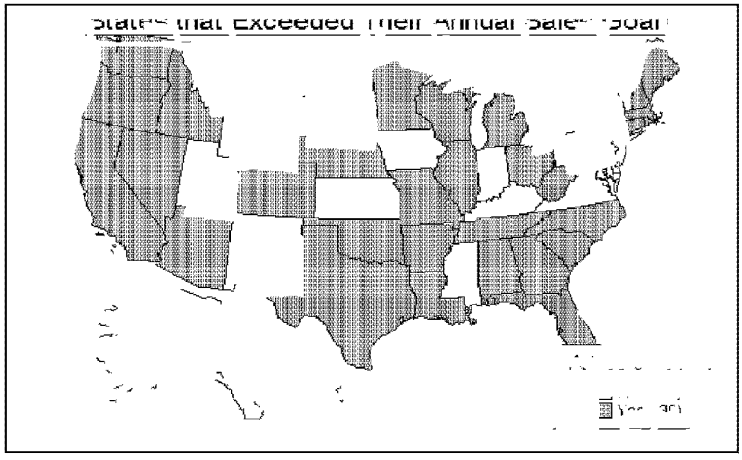


Figure 17-9: This map uses the category-shading format.

To change the colors in the categories, use the Format Properties dialog box. Again, when you're viewing a map that uses category shading, you can open this dialog box by choosing Map • Category Shading Options. Microsoft Map displays the category-shading version of the Format Options dialog box.

Dot density

The dot-density map format displays data as a series of dots. Larger values translate into more dots. The dots are placed randomly within a map region. Figure 17-10 shows an example of a map that uses the dot-density format. This map depicts population in the U.K. and Ireland. Each dot represents 100,000 people.

To change the number of units for each dot or to change the dot size, access the Dot Density Options tab of the Format Properties dialog box, which is shown in Figure 17-11.

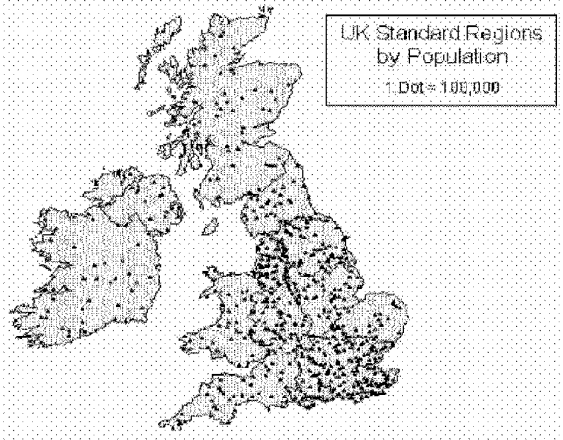


Figure 17-10: A dot-density format map, showing the population of the U.K. and Ireland.

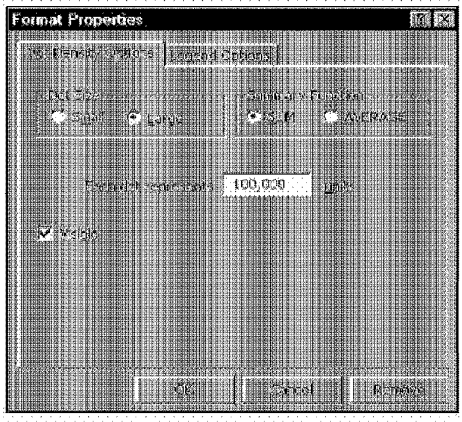


Figure 17-11: The Dot Density Options tab of the Format Properties dialog box.

Graduated symbol

The graduated-symbol map format displays a symbol, the size of which is proportional to the area's data value. Figure 17-12 shows an example of this format. I used a Wingdings font character for the symbol. To change the symbol, use the Graduated Symbol Options dialog box. You can select a font, size, and specific character.

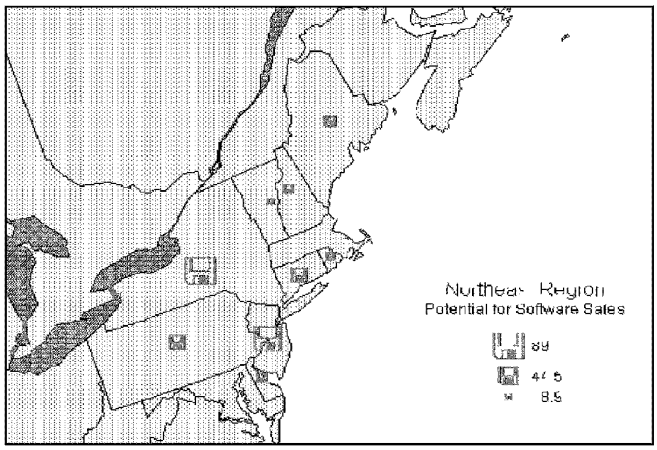


Figure 17-12: A graduated-symbol format map.

Pie chart

The pie-chart map format requires at least two columns of data. Maps with this format display a pie chart within each map region. Figure 17-13 shows an example. This map shows a pie chart that depicts the relative sales of three products for each state.

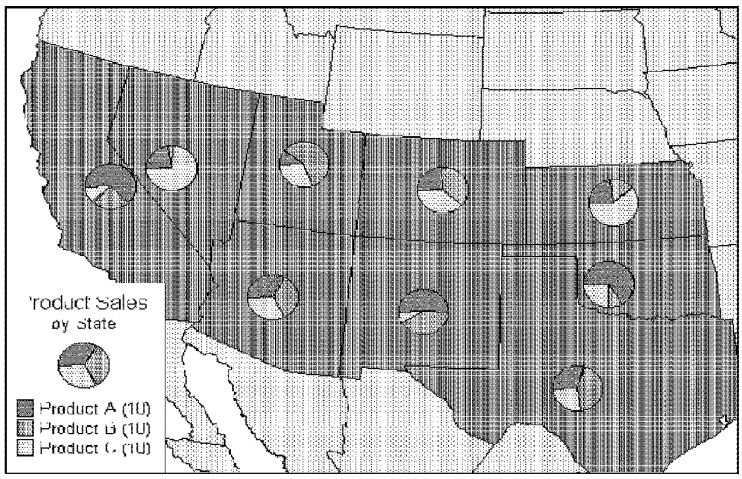


Figure 17-13: A map that uses the pie-chart format.

To change the setting for a pie-chart format map, use the Pie Chart Options tab of the Format Properties dialog box, shown in Figure 17-14. This dialog box enables you to select a color for each pie slice. If you choose the Graduated option, the size of each pie is proportional to the sum or average of the data. If you don't use the Graduated option, you also can set the diameter of the pies.

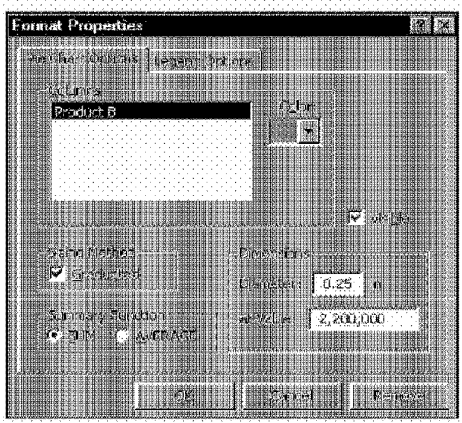


Figure 17-14: The Pie Chart Options tab of the Format Properties dialog box.

Column chart

The column-chart map format is similar to the pie-chart format—except that it displays a column chart instead of a pie chart. Figure 17-15 shows an example.

Combining map formats

As previously mentioned, a single map can include multiple formats for different data. You do this by stacking groups of icons and data fields in the Microsoft Map Control dialog box. For example, you can display sales as value shading and the number of customers as a dot-density map. Each map format has its own legend.

Overlaying multiple map types has no rules, so some experimentation usually is necessary. Unless the map is very simple, however, you're generally better off using only one or two map types per map; otherwise, the map gets so complicated that the original goal (making the data clear) is lost.

Figure 17-16 shows an example of a map that uses two formats. The value-shading format shows sales broken down into four categories. The graduated-symbol format shows the states that have a sales office.

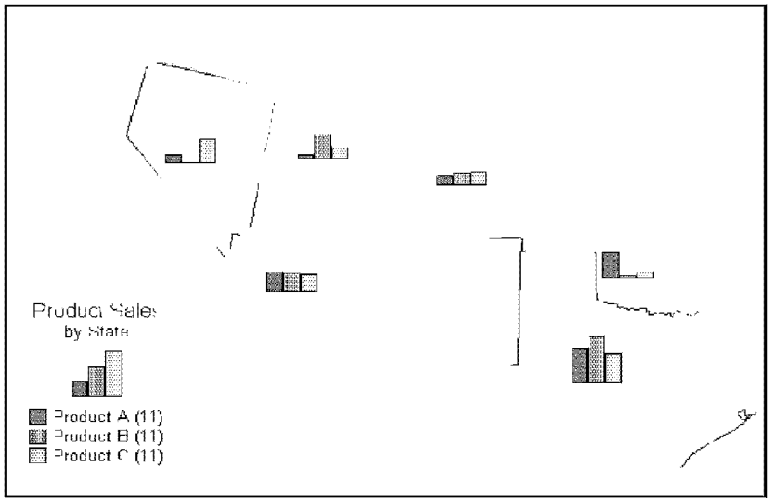


Figure 17-15: A map that uses the column-chart format.

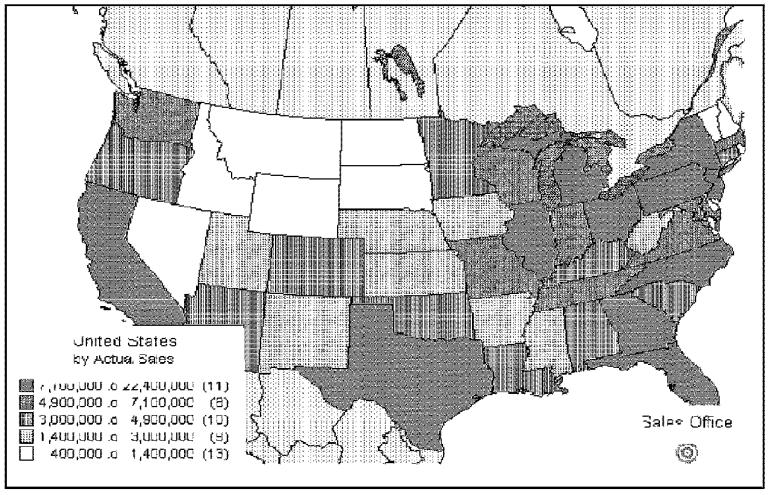


Figure 17-16: An example of a map that uses two map formats: the value-shading format and the graduated-symbol format.

Customizing Maps

After you create a map, you have numerous customization options from which to choose, which are described in the following sections.

Using the Microsoft Map Toolbar

Whenever a map is activated, the Microsoft Map toolbar appears (see the accompanying figure). Note that this isn't one of Excel's toolbars; rather, this is a special toolbar that appears only when a map is activated. This toolbar is handy for manipulating and customizing the map.

The tools on the Microsoft Map toolbar, from left to right, and their corresponding functions are presented in the following table.

Tool	*Purpose*
Select Objects	Changes mouse pointer into an arrow, to select objects in the map
Grabber	Reposition the map within the map window
Center Map	Specify the center of the map
Map Labels	Add geography labels or data values in the map
Add Text	Add free-floating text to the map
Custom Pin Map	Add pins to the map, to indicate specific locations
Display Entire	Displays the entire (unzoomed) map
Redraw Map	Redraws the map
Show/Hide Microsoft Map Control	Toggles the display of the floating Microsoft Map Control dialog box
Zoom Percentage of Map	Changes how much of the map that you view
Help	Provides help for a menu item or toolbar button

Zooming In and Out

Microsoft Map enables you to zoom your map in and out. Zooming in displays less of the map and zooming out displays more of the map (or makes the entire map smaller). Use the Zoom Percentage of Map control on the toolbar (no menu commands exist).

To zoom in, select a zoom percentage greater than 100 percent. To zoom out, select a zoom percentage less than 100 percent. Before you zoom out, you might want to specify the point that will be the center of the map (use the Center Map toolbar button).

Repositioning a Map

You'll find that, after zooming in or out, the map may not be optimally positioned within the map object rectangle. Use the Grabber tool to move the map image within the map object. Just click and drag the map to reposition it.

Adding Labels

Usually, a map doesn't have labels to identify areas. You can't automatically add labels to all areas (for example, all states in the U.S.), but you can add individual labels, one at a time. You also can insert data values that correspond to a particular map region (such as sales for West Virginia).

Use the Label tool to add labels or data values. When you click the Label tool, the dialog box shown in Figure 17-17 appears. The option button labeled Map feature names refers to labels for the various parts of the map (for example, state names in a U.S. map). When you select the Values from option, you can insert data values from a category in the list box. After closing the dialog box, you can drag the mouse pointer over the map. The label or data value appears when the mouse pointer is over a map region. Just click to place the label or data value and then repeat this procedure for each map label or data value that you want to add. Figure 17-18 shows a map that uses labels and data values.

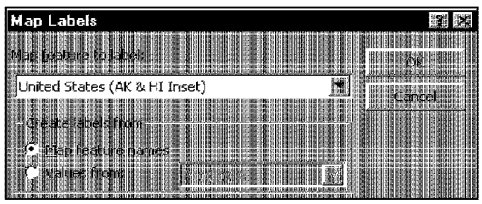


Figure 17-17: The Map Labels dialog box enables you to add labels or data values to your map.

To move a label, click and drag it to a new location. You can change the font, size, or color of a label by double-clicking it. Stretching the label (by dragging a border) also makes the font larger or smaller.

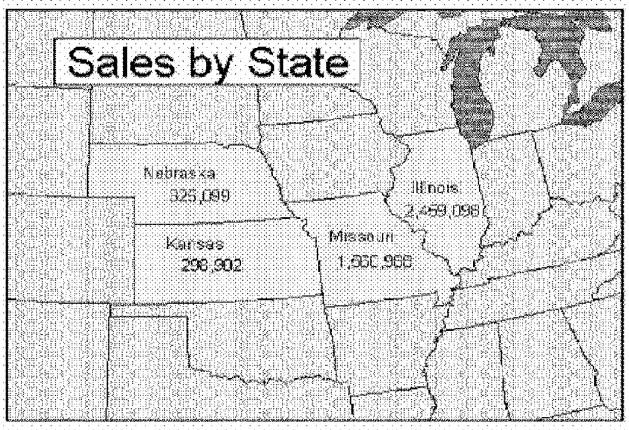


Figure 17-18: This map has labels and data values.

Adding Text

Besides the labels described in the preceding section, you can add free-floating text to your map by using the Text tool. Just click the Text tool, click the area of the map where you want to add text, and enter your text. You can manipulate text the same way that you manipulate labels.

If you don't like the fact that a map title always has a border around it (and the border can't be removed), delete the title and create your own with the Text tool.

Adding Pins to a Map

In some cases, you may want to add one or more identifier icons to your map. This is similar in concept to inserting pins in a wall map to identify various places.

Clicking the Custom Pin Map tool displays a dialog box that asks you to enter a name for a custom pin map (or choose an existing pin map). Enter a descriptive label; you'll be able to bring these same pins into another map (of the same type) later. For example, if you're identifying sales office locations, you can then add the same pins to another map.

When you close the dialog box, the mouse pointer changes to a pushpin. You can place these pins anywhere in your map. When you click the map to place a pin, you also can enter descriptive text. Double-clicking a pin enables you to change the symbol that is used to something other than a pin. Figure 17-19 shows a map with pins added to it.

Figure 17-19: This map has pins to identify specific locations.

Modifying the Legend

You have quite a bit of control over the legend in a map. Note that a map displays a separate legend for each map format that it uses. To modify a legend, double-click it to see the dialog box shown in Figure 17-20.

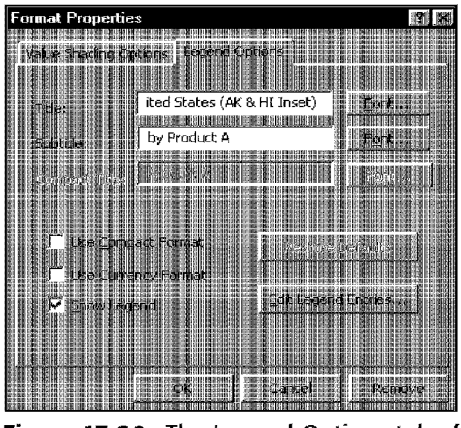


Figure 17-20: The Legend Options tab of the Format Properties dialog box.

You can display a legend in a compact format or in its normal format. A compact format takes up less space, but it doesn't give many details. You also can change the legend's title and subtitle (and enter a different title for a compacted legend). Other buttons enable you to adjust the font (including size and color) and edit the labels that are used in the legend.

To make other changes to the legend—such as changing the number of data ranges that are used—select the appropriate menu item on the Map menu. For example, to change the number of ranges that are used in a value-shading map format, select the Map • Value Shading Options command.

Adding and Removing Features

You can add or remove certain features of a map. When you select Map • Features, you see the Map Features dialog box (see Figure 17-21), which lists all available features for the selected map. To turn on a feature, place a check mark next to it. To turn off a feature, remove the check mark. The features available vary with the map that you're using. If a feature doesn't appear in the list, you can add it by clicking the Add button.

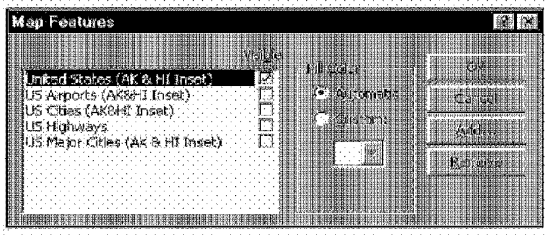


Figure 17-21: The Map Features dialog box.

Figure 17-22 shows a North America map with some features added (major cities, major highways, and world oceans) and some features removed (Canada and Mexico).

Table 17-2 lists the features available for each map. You can, however, add features from different maps—add world oceans to a North America map, for example.

In some cases, you may want your map to display only specific areas. For example, if your company does business in Missouri, Illinois, Kansas, and Nebraska, you can create a map that shows only these four states. Create a map that includes these states and then remove all features from the map by using the Map • Features command. The map then shows only those areas that have data. Figure 17-23 shows an example.

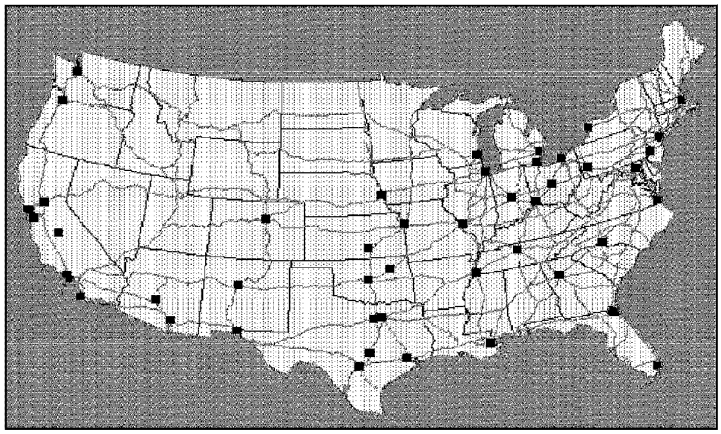


Figure 17-22: This map has features added and removed.

Table 17-2
Available Map Features

Map	*Features*
Australia	Airports, Cities, Highways, Major Cities
Canada	Airports, Cities, Forward Sortation Areas, Highways, Lakes, Major Cities
Europe	Airports, Cities, Highways, Major Cities
Mexico	Cities, Highways, Major Cities
U.K.	2-Digit Post Codes, Airports, Cities, Highways, Major Cities, Standard Regions
U.S. in North America	5-Digit Zip Code Centers, Highways, Major Cities, Great Lakes
U.S. (AK & HI Inset)	Airports, Cities, Major Cities
World	Capitals, Countries, Graticule, Oceans

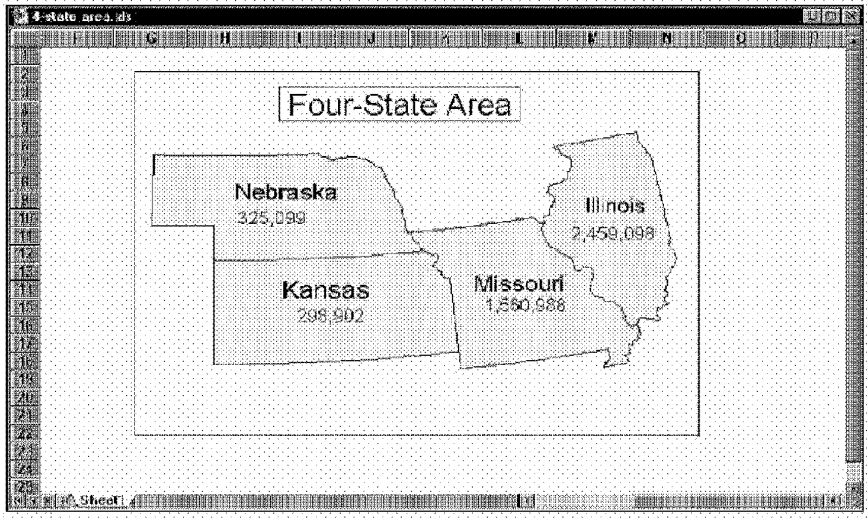


Figure 17-23: This map has all of its features removed, leaving only the states for which data is provided.

Plotting U.S. ZIP Codes

Besides recognizing geographic place names, Microsoft Map recognizes U.S. five-digit ZIP codes. If the data that you select contains more than one type of geographic data (for example, state names and ZIP codes), you need to specify which field to use in the map as the geographic data. Figure 17-24 shows the Specify Geographic Data dialog box, which warns you of the existence of more than one type of data that qualifies as geographic data.

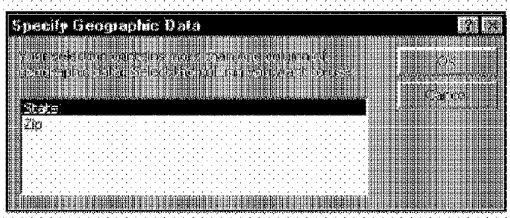


Figure 17-24: The Specify Geographic Data dialog box enables you to select the data to use as geographic data.

If you want to create a map that uses ZIP codes, make sure that your ZIP codes are formatted as values, not as text. Otherwise, Microsoft Map won't recognize them as ZIP codes.

Because ZIP codes are continually being added, Microsoft Map may not recognize all of your ZIP codes. If Microsoft Map encounters an unknown ZIP code, you receive the Resolve Unknown Geographic Data dialog box, shown in Figure 17-25. This dialog box gives you the opportunity to change the ZIP code to another one. Or, you can simply discard that item of data by clicking the Discard button.

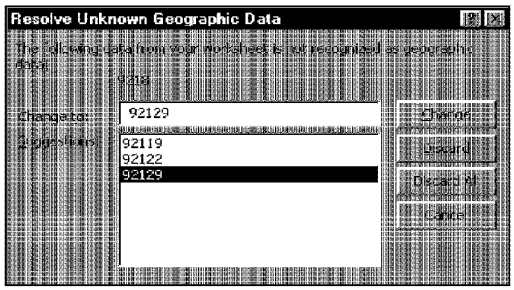


Figure 17-25: Microsoft Map displays the Resolve Unknown Geographic Data dialog box when it doesn't recognize a geographic name.

Figure 17-26 shows a map that depicts customers by their ZIP codes. This is a graduated-symbol map (the default format when ZIP codes are used as data). Note that the symbols appear on the geographic centers of the ZIP codes and don't shade the entire ZIP code areas.

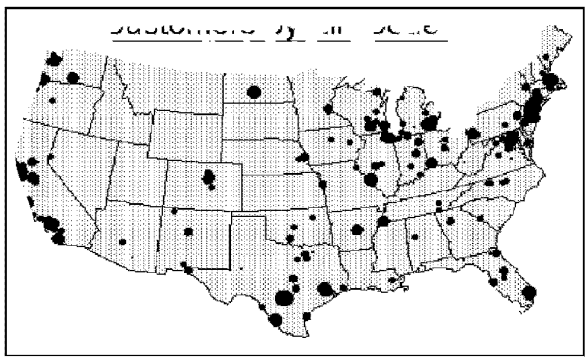


Figure 17-26: This map shows customers by ZIP code centers.

Adding More Data to a Map

After you create a map, you can add additional data to it. Use the Insert • Data command to add data from a worksheet range, or use the Insert • External Data command to add new data from a database file. Make sure that the data includes geographic labels that match the map to which you're adding data.

Map Templates

As you may have figured out by now, getting a map just right can sometimes take a lot of time. Fortunately, you can save a map template, so that you can reuse the settings for another map. To do so, create and customize the map and then choose Map • Save Map Template. You can save a template that includes the following:

- The features that you've added or removed
- A particular view (zoomed in or out)
- Both of the preceding items

Saved templates then appear in the Multiple Maps Available dialog box, which appears when you create a map.

Converting a Map to a Picture

You'll find that working with maps sometimes can be rather sluggish—a great deal of work goes on behind the scenes. When you finish with your map, you can convert it to a static picture that is no longer linked to the data. Click the map once to select it (don't double-click it) and then choose Edit • Copy. Then, select Edit • Paste Special and choose the Picture (Enhanced Metafile) option. This creates an unlinked picture of the map. Then, you can select the original map object and delete it.

If you convert a map to a picture, no way exists to link data back to the picture. If any of your data changes or you want to make modifications to the map, you have to re-create the map.

Learning More

The Microsoft Map feature is relatively complex, and it definitely takes time to master. The best way to master this feature is simply to create some maps and perform customizations. As previously mentioned, the user interface is different

from Excel's, so you'll have to try some new techniques. Generally, you can find your way around maps by doing the following:

- Double-clicking objects
- Right-clicking objects
- Exploring the menus (they change somewhat, depending on the type of map)
- Using the Microsoft Map toolbar

Summary

This chapter covers Excel's new Microsoft Map feature — which is actually an OLE server application developed by MapInfo Corporation. Some data is more appropriate for a map than for a chart, and this chapter demonstrates the difference. This chapter also describes the basics of creating and customizing maps and provides an example of each map format.

CHAPTER 18

Creating and Using Worksheet Outlines

In This Chapter

Introducing Worksheet Outlines

Creating an Outline

Using Outlines

If you use a word processor, you may be familiar with the concept of an outline. Most word processors have an outline mode that lets you view only the headings and subheadings in your document. You can easily expand a heading to show the detail (that is, the text) below it. To write this book, I used the outline feature in my word processor extensively.

Excel also is capable of using outlines, and understanding this feature can make working with certain types of worksheets much easier for you.

Introducing Worksheet Outlines

You can use outlines to create summary reports in which you don't want to show all the details. You'll find that some worksheets are more suitable for outlines than others. If your worksheet uses hierarchical data with subtotals, it's probably a good candidate for an outline.

An Example

The best way to understand how worksheet outlining works is to look at an example. Figure 18-1 shows a simple budget model without an outline. Subtotals are used to calculate subtotals by region and by quarter.

Two-Way Outline Example.xls

	A	B	C	D	E	F	G	H	I
1	State	Jan	Feb	Mar	Q1 Total	Apr	May	Jun	Q2 Total
2	California	1118	1960	1252	4330	1271	1557	1679	4507
3	Washington	1247	1238	1028	3513	1345	1784	1574	4703
4	Oregon	1460	1954	1726	5140	1461	1764	1144	4369
5	Nevada	1345	1375	1075	3795	1736	1555	1372	4663
6	**West Total**	**5170**	**6527**	**5081**	**16778**	**5813**	**6660**	**5769**	**18242**
7	New York	1429	1316	1993	4738	1832	1740	1191	4763
8	New Jersey	1735	1406	1224	4365	1706	1320	1290	4316
9	Massachusetts	1099	1233	1110	3442	1637	1512	1006	4155
10	Florida	1705	1792	1225	4722	1946	1327	1367	4630
11	**East Total**	**5968**	**5747**	**5552**	**17267**	**7121**	**5899**	**4844**	**17864**
12	Kentucky	1109	1078	1155	3342	1993	1082	1551	4626
13	Oklahoma	1309	1045	1641	3995	1924	1499	1941	5364
14	Missouri	1511	1744	1414	4669	1243	1493	1820	4556
15	Illinois	1539	1493	1211	4243	1165	1013	1445	3623

Sheet1

Figure 18-1: A typical budget model with subtotals.

Figure 18-2 shows the same worksheet after the outline was created. Notice that Excel adds a new border to the left of the screen. This border contains controls that enable you to determine which level to view. This particular outline has three levels: States, Regions (each region consists of states), and Grand Total (the sum of each region's subtotal). In Figure 18.2, the outline is fully expanded so that you can see all the data.

Two-Way Outline Example.xls

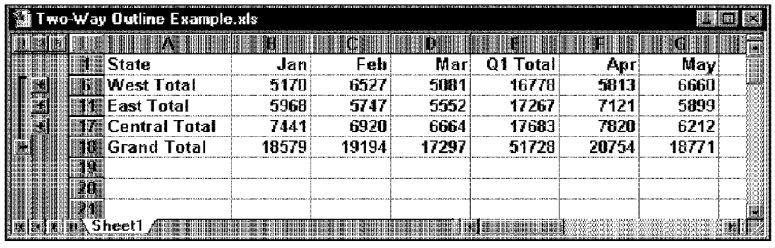
	A	B	C	D	E	F	G
1	State	Jan	Feb	Mar	Q1 Total	Apr	May
2	California	1118	1960	1252	4330	1271	1557
3	Washington	1247	1238	1028	3513	1345	1784
4	Oregon	1460	1954	1726	5140	1461	1764
5	Nevada	1345	1375	1075	3795	1736	1555
6	**West Total**	**5170**	**6527**	**5081**	**16778**	**5813**	**6660**
7	New York	1429	1316	1993	4738	1832	1740
8	New Jersey	1735	1406	1224	4365	1706	1320
9	Massachusetts	1099	1233	1110	3442	1637	1512
10	Florida	1705	1792	1225	4722	1946	1327
11	**East Total**	**5968**	**5747**	**5552**	**17267**	**7121**	**5899**
12	Kentucky	1109	1078	1155	3342	1993	1082
13	Oklahoma	1309	1045	1641	3995	1924	1499
14	Missouri	1511	1744	1414	4669	1243	1493
15	Illinois	1539	1493	1211	4243	1165	1013
16	Kansas	1973	1560	1243	4776	1495	1125
17	**Central Total**	**7441**	**6920**	**6664**	**17683**	**7820**	**6212**
18	**Grand Total**	**18579**	**19194**	**17297**	**51728**	**20754**	**18771**
19							

Sheet1

Figure 18-2: The budget model after creating an outline.

Figure 18-3 depicts the outline displayed at the second level. Now, the outline shows only the totals for the regions (the detail rows are hidden). You can partially

expand the outline to show the detail for a particular region. Collapsing the outline to level 1 shows only the headers and the Grand Total row.



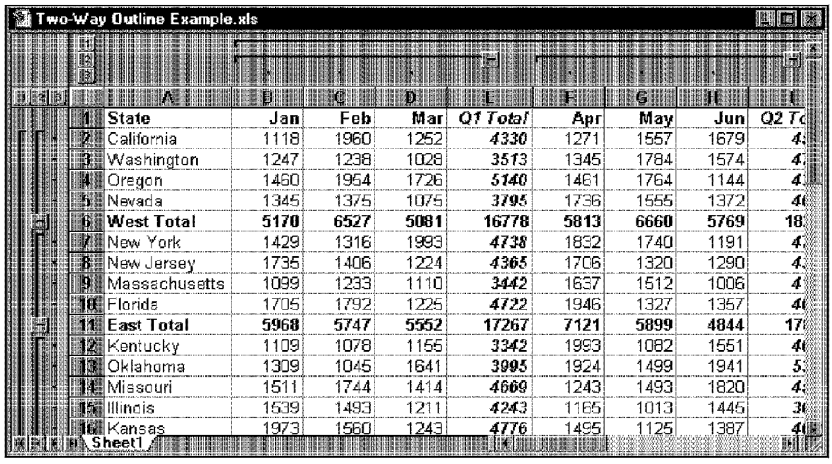
Two-Way Outline Example.xls

	A	B	C	D	E	F	G
1	State	Jan	Feb	Mar	Q1 Total	Apr	May
6	West Total	5170	6527	5081	16778	5813	6660
11	East Total	5968	5747	5552	17267	7121	5899
17	Central Total	7441	6920	6664	17683	7820	6212
18	Grand Total	18579	19194	17297	51728	20754	18771
19							
20							
21							

Sheet1

Figure 18-3: The budget model after collapsing the outline to the second level.

Excel can create outlines in both directions. In the preceding examples, the outline was a row (vertical) outline. Figure 18-4 shows the same model after a column (horizontal) outline was added. Now, Excel displays another border at the top.



Two-Way Outline Example.xls

	A	B	C	D	E	F	G	H	I
1	State	Jan	Feb	Mar	*Q1 Total*	Apr	May	Jun	*Q2 Tc*
2	California	1118	1960	1252	***4330***	1271	1557	1679	***4***
3	Washington	1247	1238	1028	***3513***	1345	1784	1574	***4***
4	Oregon	1460	1954	1726	***5140***	1461	1764	1144	***4***
5	Nevada	1345	1375	1075	***3795***	1736	1555	1372	***4***
6	**West Total**	**5170**	**6527**	**5081**	**16778**	**5813**	**6660**	**5769**	**18**
7	New York	1429	1316	1993	***4738***	1832	1740	1191	***4***
8	New Jersey	1735	1406	1224	***4365***	1706	1320	1290	***4***
9	Massachusetts	1099	1233	1110	***3442***	1637	1512	1006	***4***
10	Florida	1705	1792	1225	***4722***	1946	1327	1367	***4***
11	**East Total**	**5968**	**5747**	**5552**	**17267**	**7121**	**5899**	**4844**	**17**
12	Kentucky	1109	1078	1155	***3342***	1993	1082	1551	***4***
13	Oklahoma	1309	1045	1641	***3995***	1924	1499	1941	***5***
14	Missouri	1511	1744	1414	***4669***	1243	1493	1820	***4***
15	Illinois	1539	1493	1211	***4243***	1165	1013	1445	***3***
16	Kansas	1973	1560	1243	***4776***	1495	1125	1387	***4***

Sheet1

Figure 18-4: The budget model after adding a column outline.

If you create both a row and a column outline in a worksheet, you can work with each outline independent of the other. For example, you can show the row outline at the second level and the column outline at the first level. Figure 18-5 shows the model with both outlines collapsed at the second level. The result is a nice summary table that gives regional totals by quarter.

You'll find the workbook used in the preceding examples on this book's CD-ROM.

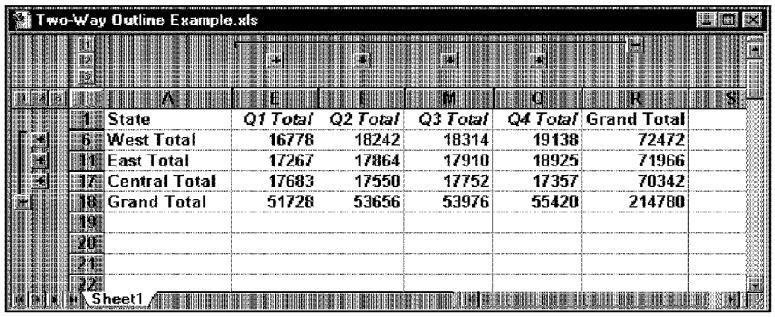


Figure 18-5: The budget model with both outlines collapsed at the second level.

More About Outlines

The following are points to keep in mind about worksheet outlines:

- A single worksheet can have only one outline (row, column, or both). If you need to create more than one outline, move the data to a new worksheet.
- You can either create an outline manually or have Excel do it for you automatically. If you choose the latter option, you may need to do some preparation to get the worksheet in the proper format.
- You can create an outline for either all data on a worksheet or just a selected data range.
- You can remove an outline with a single command.
- You can hide the outline symbols (to free screen space) but retain the outline.
- You can have up to eight nested levels in an outline.

Worksheet outlines can be quite useful. But if your main objective is to summarize a large amount of data, you might be better off using a pivot table. A pivot table is much more flexible and doesn't require that you create the subtotal formulas; it does the summarizing for you automatically.

Pivot tables are discussed in Chapter 25.

Creating an Outline

In this section, you learn the two ways to create an outline: automatically and manually. But, before getting into the details of those two methods, the all-important first step is examined: getting your data ready for outlining.

Preparing the Data

Before you create an outline, you need to ensure the following:

- The data is appropriate for an outline
- The formulas are set up properly

Determining appropriate data

What type of data is appropriate for an outline? Generally, the data should be arranged in a hierarchy, such as a budget that consists of an arrangement similar to the following:

Company

Division

Department

Budget Category

Budget Item

In this case, each budget item (for example, airfare and hotel expenses) is part of a budget category (for example, travel expenses). Each department has its own budget, and the departments are rolled up into divisions. The divisions make up the company. This type of arrangement is well-suited for a row outline — although most of your outlines probably won't have this many levels.

Once created, you can view the information at any level of detail that you want. When you need to create reports for different levels of management, try using an outline. Upper management may want to see only the Division totals. Division managers may want to see totals by department, and each department manager needs to see the full details for his or her department.

As demonstrated at the beginning of the chapter, you can include time-based information that is rolled up into larger units (such as months and quarters) in a column outline. Column outlines work just like row outlines, however, and the levels need not be time-based.

Setting up the formulas

Before you create an outline, you need to make sure that all the summary formulas are entered correctly and consistently. *Consistently* means that the formulas are in the same relative location. Generally, formulas that compute summary formulas (such as subtotals) are entered below the data to which they refer. In some cases, however, the summary formulas are entered above the referenced cells. Excel can handle either method, but you must be consistent throughout the range that you outline. If the summary formulas aren't consistent, automatic outlining won't produce the results that you want.

If your summary formulas aren't consistent (that is, some are above and some are below the data), you still can create an outline, but you must do it manually.

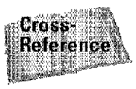
Creating an Outline Automatically

Excel can create an outline for you automatically in a few seconds, whereas it might take you ten minutes or more to do the same thing manually.

To have Excel create an outline, move the cell pointer anywhere within the range of data that you're outlining. Then, choose Data • Group and Outline • Auto Outline. Excel analyzes the formulas in the range and creates the outline. Depending on the formulas that you have, Excel creates a row outline, a column outline, or both.

If the worksheet already has an outline, Excel asks whether you want to modify the existing outline. Click Yes to force Excel to remove the old outline and create a new one.

Excel automatically creates an outline when you use the Data • Subtotals command, which inserts subtotal formulas automatically if you set up your data as a list.



The Data • Subtotals command is discussed in Chapter 23 (see the section "Creating Subtotals").

Creating an Outline Manually

Usually, letting Excel create the outline is the best approach. It's much faster and less error-prone. If the outline that Excel creates isn't what you have in mind, however, you can create an outline manually.

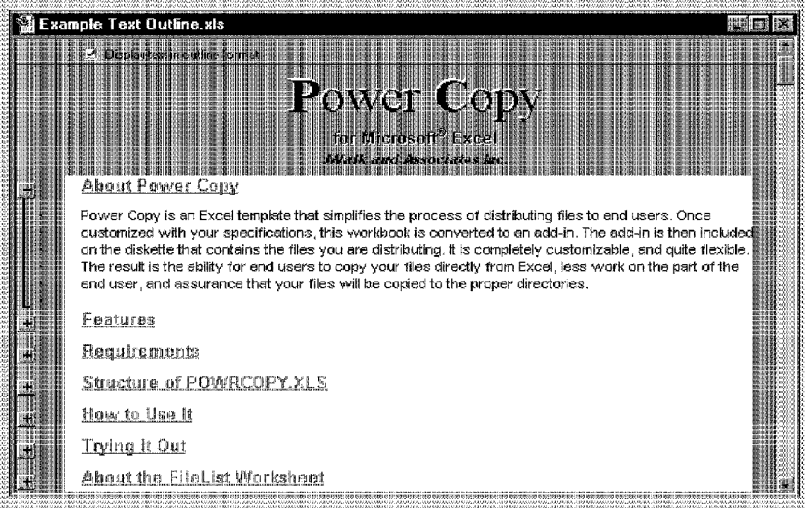
When Excel creates a row outline, the summary rows all must be above the data or below the data (they can't be mixed). Similarly, for a column outline, the summary columns all must be to the right of the data or to the left of the data. If your worksheet doesn't meet these requirements, you have two choices:

- Rearrange the worksheet so that it does meet the requirements
- Create the outline manually

Using an Outline for Text

If you need to present lots of textual information in a workbook—as in user instructions, for example—consider arranging the information in the form of an outline. The accompanying figure shows an example that I developed for one of my shareware products. The user manual is contained on a worksheet, and I created an outline to make locating a specific section easier. I also used a simple macro, attached to a check box, to make it easy for users to expand and collapse the outline.

The workbook shown in the figure is available on this book's CD-ROM.



You also need to create an outline manually if the range doesn't contain any formulas. You may have imported a file and want to use an outline to display it better. Because Excel uses the formulas to determine how to create the outline, it is not able to make an outline without formulas.

Creating an outline manually consists of creating groups of rows (for row outlines) or groups of columns (for column outlines). To create a group of rows, click the row numbers for all the rows that you want to include in the group—but do not select the row that has the summary formulas. Then, choose Data • Group and Outline • Group. Excel displays outline symbols for the group. Repeat this for each group that you want to create. When you collapse the outline, Excel hides rows in the group. But the summary row, which is not in the group, remains in view.

If you select a range of cells (rather than entire rows or columns) before you create a group, Excel displays a dialog box asking what you want to group. It then groups entire rows or columns based on the range that you select.

You also can select groups of groups to create multilevel outlines. When you create multilevel outlines, always start with the innermost groupings and then work your way out. If you realize that you grouped the wrong rows, you can ungroup the group by selecting Data • Group and Outline • Ungroup.

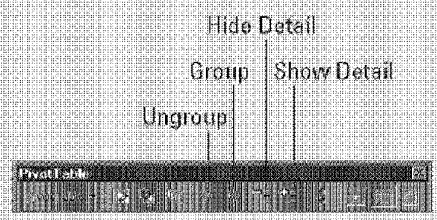
Excel has toolbar buttons that speed up the process of grouping and ungrouping (see the sidebar "Outlining Tools"). You also can use the following keyboard shortcuts:

- **Alt+Shift+right arrow:** Groups selected rows or columns
- **Alt+Shift+left arrow:** Ungroups selected rows or columns

Creating outlines manually can be confusing at first, but if you stick with it, you'll become a pro in no time.

Outlining Tools

Excel doesn't have a toolbar devoted exclusively to outlining, but it *does* have one that comes close. The Pivot Table toolbar (see accompanying figure) includes four tools that are handy for working with outlines.



The relevant Pivot Table toolbar buttons are as follows:

Button Name	*What It Does*
Ungroup	Ungroups selected rows or columns
Group	Groups selected rows or columns
Show Detail	Shows details of selected summary cell
Hide Detail	Hides details of selected summary cell

Using Outlines

This section discusses the basic operations that you can perform with a worksheet outline.

Displaying Levels

To display various outline levels, click the appropriate outline symbol. These symbols consist of buttons with numbers on them (1, 2, and so on) and buttons with either a plus sign (+) or a minus sign (–).

Clicking the 1 button collapses the outline so that it displays no detail, just the highest summary level of information. Clicking the 2 button expands the outline to show one level, and so on. The number of numbered buttons depends on the number of outline levels. Choosing a level number displays the detail for that level, plus any lower levels. To display all levels — the most detail — click the highest-level number.

You can expand a particular section by clicking its + button, or you can collapse a particular section by clicking its – button. In short, you have complete control over the details that Excel exposes or hides in an outline.

If you prefer, you can use the Hide Detail and Show Detail commands on the Data • Group and Outline menu, to hide and show details, respectively. Or, you can use one of the buttons on the Pivot Table toolbar to hide or show information.

If you constantly adjust the outline to show different reports, consider using the Custom Views feature to save a particular view and give it a name. Then, you can quickly switch among the named views. Use the View • Custom Views command for this.

Applying Styles to an Outline

When you create an outline, you can have Excel automatically apply named styles to the summary rows and columns.

Chapter 11 discusses named styles.

Excel uses styles with names in the following formats (where *n* corresponds to the outline level):

- RowLevel_*n*
- ColLevel_*n*

For example, the named style that is applied to the first row level is RowLevel_1. These styles consist only of formats for the font. Using font variations makes distinguishing various parts of the outline a bit easier. You can, of course, modify the styles in any way that you want. For example, you can use the Format • Style command to change the font size or color for the RowLevel_1 style. After you do so, all the RowLevel_1 cells take on the new formatting. Figure 18-6 shows an outline with the automatic outline styles assigned.

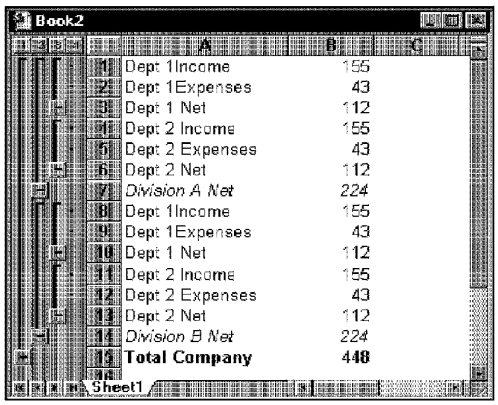


Figure 18-6: This outline has automatic styles.

You can have Excel automatically apply the styles when it creates an outline, or you can apply them after the fact. You control this in the Settings dialog box, shown in Figure 18-7. This dialog box appears when you select Data • Group and Outline • Settings.

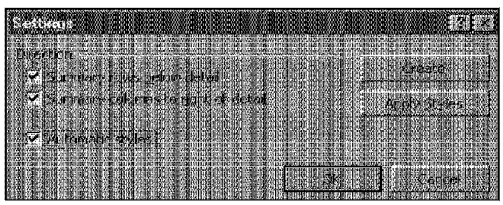


Figure 18-7: The Settings dialog box.

If the *Automatic styles* check box contains a check when you create the outline, Excel automatically applies the styles. To apply styles to an existing outline, select the outline, choose Data • Group and Outline • Settings, and then click the Apply Styles button. Notice that you also can create an outline by using this dialog box.

You may prefer to use Excel's Format• AutoFormat command to format an outline. Several of the AutoFormats use different formatting for summary cells.

Adding Data to an Outline

You may need to add additional rows or columns to an outline. In some cases, you may be able to insert new rows or columns without disturbing the outline, and the new rows or columns become part of the outline. In other cases, you'll find that the new row or column is not part of the outline. If you create the outline automatically, just select Data• Group and Outline• Auto Outline again. Excel makes you verify that you want to modify the existing outline. If you create the outline manually, you need to make the adjustments manually, as well.

Removing an Outline

If you no longer need an outline, you can remove it by selecting Data• Group and Outline• Clear Outline. Excel fully expands the outline by displaying all hidden rows and columns, and the outline symbols disappear. The outline styles remain in effect, however.

You can't "undo" removing an outline, so make sure that you *really* want to remove the outline before you select this command.

Hiding the Outline Symbols

The outline symbols Excel displays when an outline is present take up quite a bit of space (the exact amount depends on the number levels). If you want to see as much as possible onscreen, you can temporarily hide these symbols, without removing the outline. The following are the two ways to do this:

- Open the Options dialog box, select the View tab, and remove the check from the Outline Symbols check box
- Press Ctrl+8

When you hide the outline symbols, the outline still is in effect, and the worksheet displays the data at the current outline level. That is, some rows or columns may be hidden.

To redisplay the outline symbols, either place a check mark in the Outline Symbols check box in the Options dialog box or press Ctrl+8.

The Custom Views feature, which saves named views of your outline, also saves the status of the outline symbols as part of the view, enabling you to name some views with the outline symbols and other views without them.

Creating Charts from Outlines

A worksheet outline also is a handy way to create summary charts. If you have a large table of data, creating a chart usually produces a confusing mess. But, if you create an outline first, then you can collapse the outline and select the summary data for your chart. Figure 18-8 shows an example of a chart created from a collapsed outline. When you expand an outline from which you created a chart, the chart shows the additional data.

If your chart shows all the data in the outline, even when it's collapsed, remove the check from the Plot Visible Cells Only check box in the Chart tab in the Options dialog box.

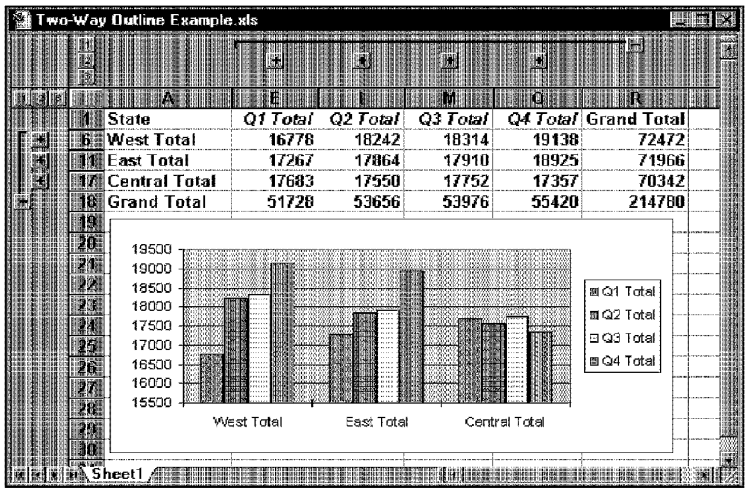


Figure 18-8: This chart was created from the summary cells in an outline.

Summary

This chapter discusses the advantages of creating an outline from worksheet data. It teaches you how to create row outlines and column outlines, either automatically or manually. It also discusses how to use an outline after it is created.

CHAPTER 19

Linking and Consolidating Worksheets

In This Chapter

Linking Workbooks

Consolidating Worksheets

This chapter discusses two procedures that are common in the world of spreadsheets: linking and consolidation. *Linking* is the process of using references to cells in external workbooks to get data into your worksheet. *Consolidation* combines or summarizes information from two or more worksheets (which can be in multiple workbooks).

Linking Workbooks

When you link worksheets, you connect them together in such a way that one depends on the other. The workbook that contains the link formulas (or external reference formulas) is called the *dependent* workbook. The workbook that contains the information used in the external reference formula is called the *source* workbook. Note, importantly, that you don't need to open the source workbook when you link it to the dependent workbook.

You also can create links to data in other applications, such as a database program or a word processor. This is a completely different procedure and is the topic of Chapter 29.

Why Link Workbooks?

When you consider linking workbooks, you might ask yourself the following question: If Workbook A needs to access data in another workbook (Workbook B), why not just enter the data into Workbook A in the first place? In some cases, you can. But the real value of linking becomes apparent when you continually update the source workbook. Creating a link in Workbook A to Workbook B means that, in Workbook A,

you always have access to the most recent information in Workbook B, because Workbook A is updated whenever Workbook B changes.

Linking workbooks also can be helpful if you need to consolidate different files. For example, each regional sales manager might store data in a separate workbook. You can create a summary workbook that first uses link formulas to retrieve specific data from each manager's workbook and then calculates totals across all regions.

Linking also is useful as a way to break up a large model into smaller files. You can create smaller workbook modules that are linked together with a few key external references. Often, this approach makes your model easier to deal with and uses less memory.

Linking has its downside, however. As you'll see later, external reference formulas are somewhat fragile, and accidentally severing the links that you create is relatively easy. You can prevent this from happening if you understand how linking works. Later in the chapter, some of the problems that may arise are discussed, as well as how to avoid them (see "Potential Problems with External Reference Formulas").

Creating External Reference Formulas

The following are the ways that you can create an external reference formula:

- **Type the cell references manually.** These references may be lengthy, because they include workbook and sheet names (and, possibly, even drive and path information). The advantage of manually typing the cell references is that the source workbook doesn't have to be open.
- **Point to the cell references.** If the source workbook is open, you can use the standard pointing techniques to create formulas that use external references.
- **With the source workbook open, select Edit• Paste Special with the Paste Link button.**
- **Use Excel's Data• Consolidate command.** This method is discussed later in the chapter (see "Consolidating Worksheets by Using Data• Consolidate").

Understanding the link formula syntax

This section discusses the concept of external references. The general syntax for an external reference formula is as follows:

```
=[WorkbookName]SheetName!CellAddress
```

Precede the cell address by the workbook name (in brackets), the worksheet name, and an exclamation point. Here's an example of a formula that uses cell A1 in the Sheet1 worksheet of a workbook named Budget:

```
=[Budget.xls]Sheet1!A1
```

If the workbook name or the sheet name in the reference includes one or more spaces, you must enclose the text in single quotation marks. For example, here's a formula that refers to cell A1 on Sheet1 in a workbook named Annual Budget:

```
='[Annual Budget]Sheet1'!A1
```

When a formula refers to cells in a different workbook, that other workbook doesn't need to be open. If the workbook is closed and not in the current folder, you must add the complete path to the reference; for example:

```
='C:\MSOffice\Excel\Budget Files\[Annual Budget]Sheet1'!A1
```

Creating a link formula by pointing

As previously mentioned, you can directly enter external reference formulas, but doing so can cause errors, because you must have every bit of information exactly correct. Instead, have Excel build the formula for you, as follows:

1. Open the source workbook.
2. Select the cell in the dependent workbook that will hold the formula.
3. Enter the formula. When you get to the part that requires the external reference, activate the source workbook and select the cell or range.
4. Finish the formula and press Enter.

You'll see that when you point to the cell or range, Excel automatically takes care of the details and creates a syntactically correct external reference. When you point to a cell reference by using the procedure outlined in the preceding steps, the cell reference is always an absolute reference (such as $A$1). If you plan to copy the formula to create additional link formulas, you can change the absolute reference to a relative reference by removing the dollar signs.

As long as the source workbook remains open, the external reference doesn't include the path to the workbook. If you close the source workbook, however, the external reference formulas change to include the full path. If you use the File • Save As command to save the source workbook with a different name, Excel changes the external references to use the new filename.

Pasting links

The Paste Special command provides another way to create external reference formulas:

1. Open the source workbook.
2. Select the cell or range that you want to link and then copy it to the Clipboard.
3. Activate the dependent workbook and select the cell in which you want the link formula to appear. If you're pasting a range, just select the upper-left cell.
4. Choose Edit • Paste Special and then click the Paste Link button.

Working with External Reference Formulas

You need to understand that a single workbook can contain links that refer to any number of different source workbooks. This section discusses what you need to know about working with links.

Creating links to unsaved workbooks

Excel enables you to create link formulas to unsaved workbooks, and even to nonexistent workbooks. Assume that you have two workbooks open and you haven't saved either of them (they have the names Book1 and Book2). If you create a link formula to Book1 in Book2 and then save Book2, Excel displays the dialog box shown in Figure 19-1. Generally, you should avoid this situation. Simply save the source workbook first.

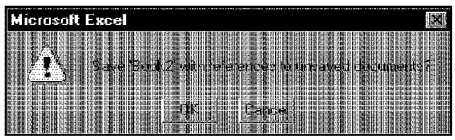


Figure 19-1: This message indicates that the workbook you're saving contains references to a workbook that you haven't yet saved.

You also can create links to documents that don't exist. You might want to do this if you'll be using a source workbook from a colleague, but the file hasn't arrived. When you enter an external reference formula that refers to a nonexistent workbook, Excel displays its File Not Found dialog box, shown in Figure 19-2. If you click Cancel, the formula retains the workbook name that you entered, but it returns an error. When the source workbook becomes available, the error goes away and the formula displays its proper value.

Opening a workbook with external reference formulas

When you open a workbook that contains one or more external reference formulas, Excel retrieves the current values from the source workbooks and calculates the formulas.

If Excel can't locate a source workbook that's referred to in a link formula, it displays its File Not Found dialog box and prompts you to supply a workbook to use for the source workbook.

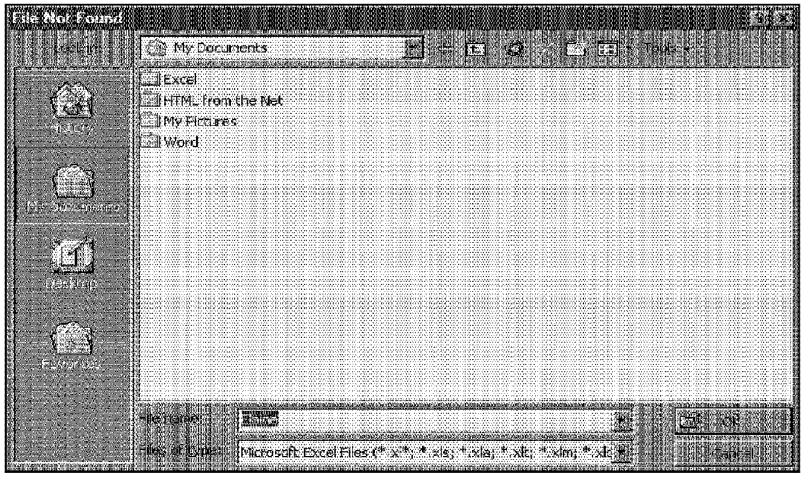


Figure 19-2: When you enter a formula that refers to a nonexistent workbook, Excel displays this dialog box to help you locate the file.

Examining links

If your workbook uses several workbook links, you might want to see a list of source workbooks. To do so, choose the Edit • Links command. Excel responds with the Links dialog box, shown in Figure 19-3. This dialog box lists all source workbooks, plus other types of links to other documents.

These other types of links are explained in Chapter 29.

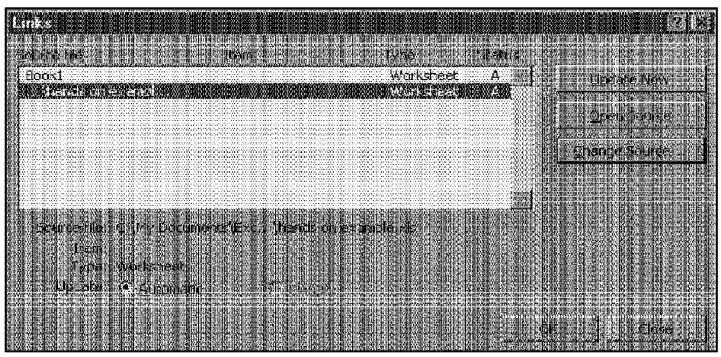


Figure 19-3: The Links dialog box lists all link sources.

Updating links

If you want to ensure that your link formulas have the latest values from their source workbooks, you can force an update. This step might be necessary if you just learned that someone made changes to the source workbook and saved the latest version to your network server.

To update linked formulas with their current value, open the Links dialog box (choose Edit • Links), choose the appropriate source workbook, and then click the Update Now button. Excel updates the link formulas with the latest version of the source workbook.

Excel always sets worksheet links to the Automatic update option in the Links dialog box, and you can't change them to Manual. This means that Excel updates the links only when you open the workbook. Excel doesn't automatically update links when the source file changes.

Changing the link source

A time may come when you need to change the source workbook for your external references. For example, you might have a worksheet that has links to a workbook named Preliminary Budget, but you later receive a finalized version named Final Budget.

You *could* change all the cell links manually, or you could simply change the link source. Do this in the Links dialog box. Select the source workbook that you want to change and click the Change Source button. Excel displays a dialog box that enables you to select a new source file. After you select the file, all external reference formulas are updated.

Severing links

If you have external references in a workbook and then decide that you no longer need the links, you can convert the external reference formulas to values, thereby severing the links. To do so, follow these steps:

1. Select the range that contains the external reference formulas and copy it to the Clipboard.
2. Choose the Edit • Paste Special command. Excel displays the Paste Special dialog box.
3. Select the Values option and click OK.
4. Press Esc to cancel cut-copy mode.

All formulas in the selected range are converted to their current values.

Potential Problems with External Reference Formulas

Using external reference formulas can be quite useful, but the links may be unintentionally severed. In almost every case, you'll be able to reestablish lost links. If you open the workbook and Excel can't locate the file, you're presented with a dialog box that enables you to specify the workbook and re-create the links. You also can change the source file by using the Change Source button in the Links dialog box. The following sections discuss some pointers that you must remember when you use external reference formulas.

Renaming or moving a source workbook

If you rename the source document or move it to a different folder, Excel won't be able to update the links. You need to use the Links dialog box and specify the new source document.

Using the File • Save As command

If both the source workbook and the destination workbook are open, Excel doesn't display the full path in the external reference formulas. If you use the File • Save As command to give the source workbook a new name, Excel modifies the external references to use the new workbook name. In some cases, this may be what you want. But in other cases, it may not. Bottom line? Be careful when you use the File • Save As command with a workbook that is linked to another workbook.

Modifying a source workbook

If you open a workbook that is a source workbook for another workbook, be extremely careful if you don't open the destination workbook at the same time. For example, if you add a new row to the source workbook, the cells all move down one row. When you open the destination workbook, it continues to use the old cell references — which are now invalid.You can avoid this problem in the following ways:

- **Open the destination workbook when you modify the source workbook.** If you do so, Excel adjusts the external references in the destination workbook when you make changes to the source workbook.
- **Use names rather than cell references in your link formula.** This is the safest approach.

Intermediary links

Excel doesn't place many limitations on the complexity of your network of external references. For example, Workbook A can contain external references that refer to Workbook B, which can contain an external reference that refers to Workbook C. In this case, a value in Workbook A can ultimately depend on a value in Workbook C. Workbook B is an *intermediary link*.

I don't recommend these types of links, but if you must use them, be aware that Excel doesn't update external reference formulas if the workbook isn't open. In the preceding example, assume that Workbooks A and C are open. If you change a value in Workbook C, Workbook A won't reflect the change, because you didn't open Workbook B (the intermediary link).

Consolidating Worksheets

The term *consolidation*, in the context of worksheets, refers to several operations that involve multiple worksheets or multiple workbook files. In some cases, consolidation involves creating link formulas. Here are two common examples of consolidation:

- The budget for each department in your company is stored in a separate worksheet in a single workbook. You need to consolidate the data and create a company-wide budget.
- Each department head submits his or her budget to you in a separate workbook. Your job is to consolidate these files into a company-wide budget.

These tasks can be very difficult or quite easy; the tasks are easy if the information is laid out exactly the same in each worksheet (as you'll see shortly).

If the worksheets aren't laid out identically, they may be similar enough. In the second example, some budget files submitted to you may be missing categories that aren't used by a particular department. In this case, you can use a handy feature in Excel that matches data by using row and column titles. This feature is discussed later in the chapter (see "Consolidating Worksheets by Using Data • Consolidate).

If the worksheets bear little or no resemblance to each other, your best bet may be to edit the sheets so that they correspond to one another. In some cases, simply reentering the information in a standard format may be more efficient.

You can use any of the following techniques to consolidate information from multiple workbooks:

- Use external reference formulas
- Copy the data and use the Paste Special command
- Use Excel's Data • Consolidate command
- Use a pivot table (discussed in Chapter 25)

Using Links to Recover Data from Corrupted Files

Sooner or later (with luck, later), it's bound to happen. You attempt to open an Excel workbook, and you get an error telling you that Excel can't access the file. Most of the time, this indicates that the file (somehow) got corrupted. If you're lucky, you have a recent backup. If you're *very* lucky, you haven't made any changes to the file since you backed it up. But assume that you fell a bit behind on your backup procedures, and the dead file is the only version you have.

Although I don't know of any method to fully recover a corrupt file, I'll share with you a method that sometimes enables you to recover at least some of the data from worksheets in the file (values, not formulas). Your actual success depends on how badly the file is corrupted.

This technique involves creating an external reference formula that refers to the corrupt file. You need to know the names of the worksheets that you want to recover. For example, assume that you have a workbook named Summary Data that you can't open. Further, assume that this workbook is stored on the C drive in a folder named Sheets. This workbook has one sheet, named Sheet1. Here's how to attempt to recover the data from this worksheet:

1. Open a new workbook.
2. In cell A1, enter the following external reference formula:

   ```
   ='C:\Sheets\[Summary Data]Sheet1'!A1
   ```

   If you're lucky, this formula returns the value in cell A1 of Sheet1 in the corrupt file.
3. Copy down this formula and to the right to recover as many values as you can.
4. Convert the external reference formulas in the new workbook to values and then save the workbook.

If the corrupt file has additional worksheets, repeat these steps for any other worksheets in the workbook (you need to know the exact sheet names).

Consolidating Worksheets by Using Formulas

Consolidating with formulas simply involves creating formulas that use references to other worksheets or other workbooks. The primary advantages to using this method of consolidation are the following:

- Dynamic updating—if the values in the source worksheets change, the formulas are updated automatically.
- The source workbooks don't need to be open when you create the consolidation formulas.

If you are consolidating the worksheets in the same workbook—and if all the worksheets are laid out identically—the consolidation task is quite simple. You can just use standard formulas to create the consolidations. For example, to compute the total for cell A1 in worksheets named Sheet2 through Sheet10, enter the following formula:

```
=SUM(Sheet2:Sheet10!A1)
```

You can enter this formula manually or use the multisheet selection technique discussed in Chapter 8 (see "Selecting Multisheet Ranges"). You can then copy this formula to create summary formulas for other cells. Figure 19-4 shows this technique at work.

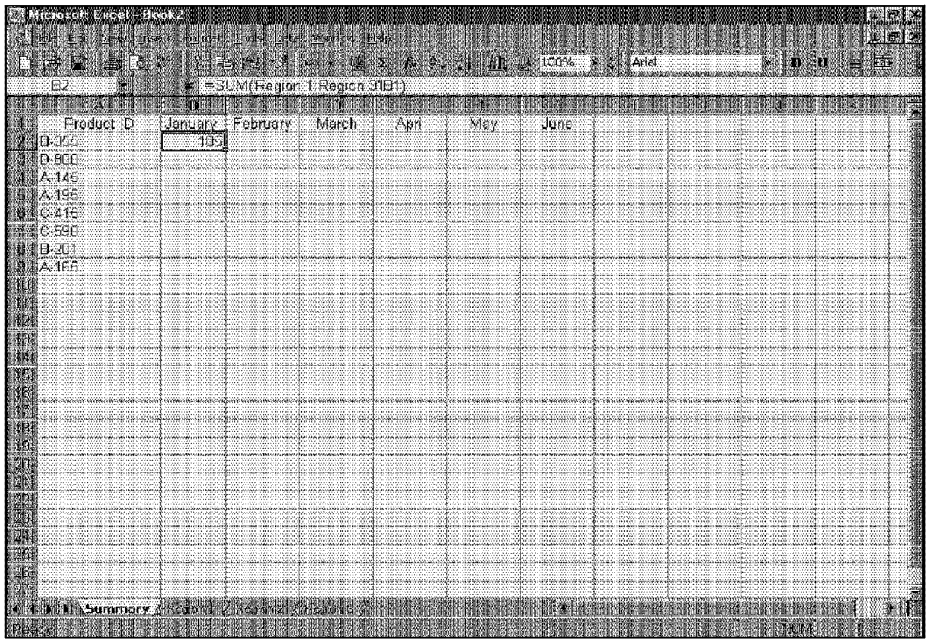


Figure 19-4: Consolidating multiple worksheets by using formulas.

If the consolidation involves other workbooks, you can use external reference formulas to perform your consolidation. For example, if you want to add the values in cell A1 from Sheet1 in two workbooks (named Region1 and Region2), you can use the following formula:

```
=[Region1.xls]Sheet1!A1+[Region2.xls]Sheet1!A1
```

You can include any number of external references in this formula, up to the 1,024-character limit for a formula. However, if you use many external references, such a formula can be quite lengthy and confusing, if you need to edit it.

Remember that Excel expands the references to include the full path—which can increase the length of the formula. Therefore, this expansion may cause the formula to exceed the limit, thus creating an invalid formula.

If the worksheets that you're consolidating aren't laid out the same, you can still use formulas—but you have to ensure that each formula refers to the correct cell.

Consolidating Worksheets by Using Paste Special

Another method of consolidating information is to use the Edit • Paste Special command. This method is applicable only when all the worksheets that you're consolidating are open. The disadvantage—a major disadvantage—is that the consolidation isn't dynamic. In other words, it doesn't generate a formula. So, if any data that was consolidated changes, the consolidation is no longer accurate.

This technique takes advantage of the fact that the Paste Special command can perform a mathematical operation when it pastes data from the Clipboard. Figure 19-5 shows the Paste Special dialog box.

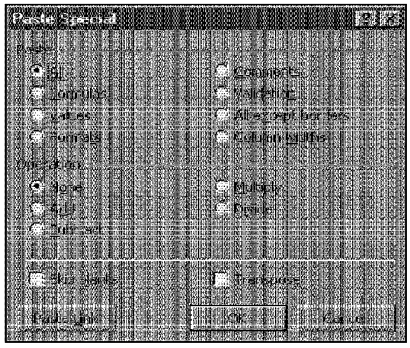


Figure 19-5: The Paste Special dialog box.

Here's how to use this method:

1. Copy the data from the first source range.
2. Activate the destination workbook and select the cell in which you want to place the consolidation formula.
3. Select Edit • Paste Special, click the Add option, and then click OK.

Repeat these steps for each source range that you want to consolidate. As you can see, this can be quite error-prone and isn't really a good method of consolidating data.

Consolidating Worksheets by Using Data • Consolidate

For the ultimate in data consolidation, use Excel's Data • Consolidate command. This method is quite flexible, and in some cases, it even works if the source worksheets aren't laid out identically. This technique can create consolidations that are static (no link formulas) or dynamic (with link formulas). The Data • Consolidate command supports the following methods of consolidation:

- **By position:** This method is accurate only if the worksheets are laid out identically.
- **By category:** Excel uses row and column labels to match data in the source worksheets. Use this option if the data is laid out differently in the source worksheets or if some source worksheets are missing rows or columns.

Figure 19-6 shows the Consolidate dialog box, which appears when you select Data • Consolidate. The following list is a description of the controls in this dialog box:

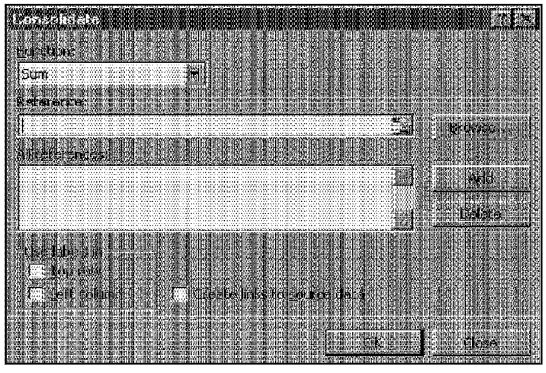


Figure 19-6: The Consolidate dialog box enables you to specify ranges to consolidate.

- **Function list box:** Specify the type of consolidation. Usually, you use Sum, but you also can select from ten other options: Count, Avg, Max, Min, Product, Count Nums, StdDev (standard deviation), StdDevp (population standard deviation), Var (variance), or Varp (population variance).
- **Reference text box:** Specify a range from a source file that you want to consolidate. You can enter the range reference manually or use any standard pointing technique (if the workbook is open). After you enter the range in this box, click the Add button to add it to the All References list. If you consolidate by position, don't include labels in the range. If you consolidate by category, *do* include labels in the range.

- **All references list box:** Contains the list of references that you have added with the Add button.
- **Use labels in check boxes:** Use to instruct Excel to perform the consolidation by examining the labels in the top row, the left column, or both positions. Use these options when you consolidate by category.
- **Create links to source data check box:** When you select this option, Excel creates an outline that consists of external references to the destination cells in the destination worksheet. Additionally, Excel includes summary formulas in the outline. If you don't select this option, the consolidation doesn't use formulas.
- **Browse button:** Displays a dialog box that enables you to select a workbook to open. It inserts the filename in the Reference box, but you have to supply the range reference.
- **Add button:** Adds the reference in the Reference box to the All References list.
- **Delete button:** Deletes the selected reference from the All References list.

An example

The simple example in this section demonstrates the power of the Data • Consolidate command. Figure 19-7 shows three single-sheet workbooks that will be consolidated. These worksheets report product sales for three months. Notice, however, that they don't all report on the same products. In addition, the products aren't even listed in the same order. In other words, these worksheets aren't laid out identically—which makes creating consolidation formulas difficult.

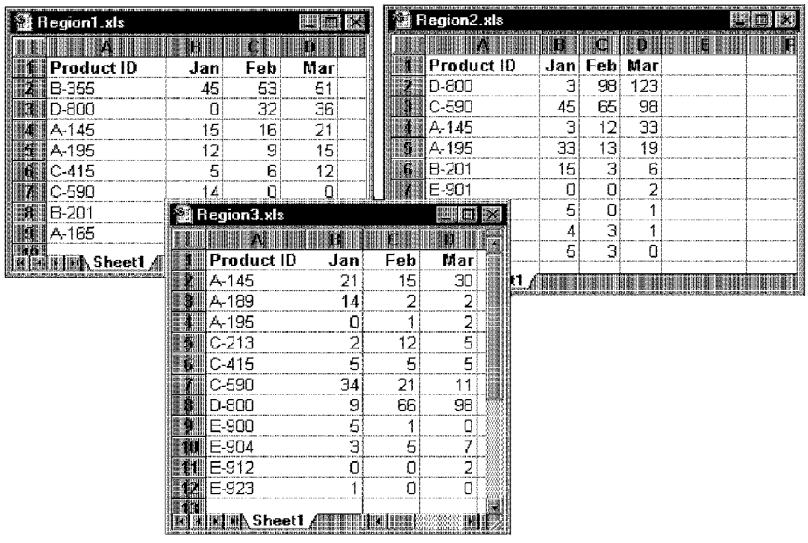


Figure 19-7: Three worksheets to be consolidated.

To consolidate this information, start with a new workbook. The source workbooks can be open or not — it doesn't matter. Follow these steps to consolidate the workbooks:

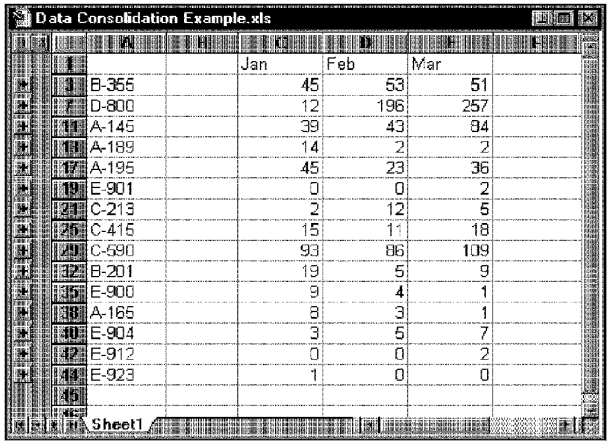
1. Select Data • Consolidate. Excel displays its Consolidate dialog box.
2. Select the type of consolidation summary that you want to use. Use Sum for this example.
3. Enter the reference for the first worksheet to consolidate. If the workbook is open, you can point to the reference. If it's not open, click the Browse button to locate the file on disk. The reference must include a range. Use **A1:D100**. This range is larger than the actual range to consolidate, but using this range ensures that the consolidation still works if new rows are added to the source file. When the reference in the Reference box is correct, click Add to add it to the All References list.
4. Enter the reference for the second worksheet. You can simply edit the existing reference by changing Region1 to **Region2** and then clicking Add. This reference is added to the All References list.
5. Enter the reference for the third worksheet. Again, you can simply edit the existing reference by changing Region2 to **Region3** and then clicking Add. This final reference is added to the All References list.
6. Because the worksheets aren't laid out the same, select the Left column and Top row check boxes to force Excel to match the data by using the labels.
7. Select the Create links to source data check box to make Excel create an outline with external references.
8. Click OK to begin the consolidation.

In seconds, Excel creates the consolidation, beginning at the active cell. Figure 19-8 shows the result. Notice that Excel created an outline, which is collapsed to show only the subtotals for each product. If you expand the outline, you can see the details. Examine it further, and you'll discover that each detail cell is an external reference formula that uses the appropriate cell in the source file. Therefore, the destination range is updated automatically if any data is changed.

More about consolidation

Excel is very flexible regarding the sources that you can consolidate. You can consolidate data from the following:

- Workbooks that are open
- Workbooks that are closed (you have to enter the reference manually — but you can use the Browse button to get the filename part of the reference)
- The same workbook in which you're creating the consolidation



Data Consolidation Example.xls

	A	B	C	D	E	F
1			Jan	Feb	Mar	
3	B-355		45	53	51	
7	D-800		12	196	257	
11	A-145		39	43	84	
13	A-189		14	2	2	
17	A-195		45	23	36	
19	E-901		0	0	2	
21	C-213		2	12	5	
25	C-415		15	11	18	
29	C-590		93	86	109	
32	B-201		19	5	9	
35	E-900		9	4	1	
38	A-165		8	3	1	
40	E-904		3	5	7	
42	E-912		0	0	2	
44	E-923		1	0	0	
45						

Sheet1

Figure 19-8: The result of the consolidation.

And, of course, you can mix and match any of the preceding choices in a single consolidation.

Excel remembers the references that you entered in the Consolidate dialog box and saves them with the workbook. Therefore, if you want to refresh a consolidation later, you won't have to reenter the references.

If you perform the consolidation by matching labels, be aware that the matches must be exact. For example, *Jan* does not match *January*. The matching isn't case-sensitive, however, so *April* does match *APRIL*. In addition, the labels can be in any order, and they need not be in the same order in all the source ranges.

If you don't choose the Create links to source data check box, Excel doesn't create formulas, which generates a static consolidation. If the data on any of the source worksheets changes, the consolidation doesn't update automatically. To update the summary information, you need to select the destination range and repeat the Data • Consolidate command.

If you name the destination range **Consolidate_Area**, you don't need to select it before you update the consolidation. Consolidate_Area is a name that has special meaning to Excel.

If you choose the Create links to source data check box, Excel creates an outline. This is a standard worksheet outline, and you can manipulate it by using the techniques described in Chapter 18.

Summary

This chapter discusses two important spreadsheet procedures: linking and consolidation. *Linking* is the process of referring in one worksheet to cells in external workbooks. *Consolidation* is the process of combining or summarizing information from two or more worksheets (which can be in multiple workbooks). This chapter covers various methods of linking and consolidation, and lists potential pitfalls.

CHAPTER 20

Creating and Using Array Formulas

In This Chapter

This chapter introduces a concept that may be new to you: *array formulas*. Understanding this special type of formula may open a whole new world of analytical capability. Working with arrays (rather than with individual cells) requires a different type of mind-set. Some people never quite get the hang of arrays, and others take to this concept quickly. If you're in the former group, don't despair. Using array formulas can be considered an optional skill.

Introducing Arrays

This chapter discusses two concepts:

- **Array:** A collection of cells or values that is operated on as a group. An array can be stored in cells or can be a named constant that consists of multiple elements.
- **Array formula:** A formula that uses one or more arrays either directly or as arguments for a function. An array formula can occupy one or more cells.

If you've ever done any computer programming, you've probably been exposed to arrays. An *array* is a collection of items. Excel's arrays can be one-dimensional or two-dimensional. These dimensions correspond to rows and columns. For example, a *one-dimensional array* can be a cell range that occupies cells in one row (a horizontal array) or one column (a vertical array). A *two-dimensional array* occupies cells in one or more rows and columns.

You can perform operations on arrays by using *array formulas*. For example, if you construct an array formula to multiply a five-item vertical array by another five-column vertical array, the result is another five-column vertical array that consists of each element in the first array multiplied by each corresponding element in the second array. Because Excel can fit only one value in a cell, the results of an operation such as this one occupy five cells — and the same array formula is in each of the five cells.

Figure 20-1 illustrates this example. Each cell in the range C1:C5 holds the same formula: {=A1:A5*B1:B5}. The result occupies five cells and contains each element of the first array multiplied by each corresponding element in the second array. The brackets around the formula designate it as an array formula (more about this later in "Entering an Array Formula").

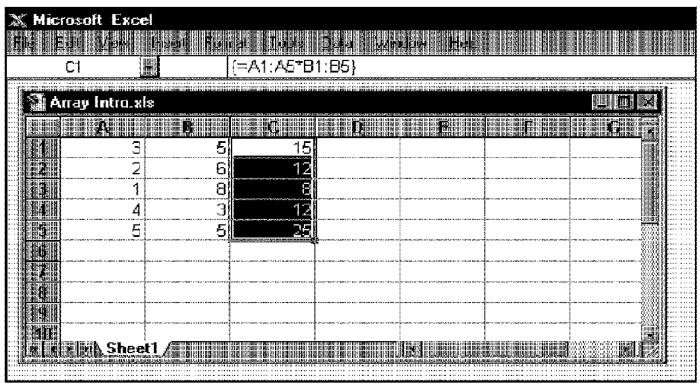


Figure 20-1: A single array formula entered in the range C1:C5 produces results in five cells.

As you will see, arrays have their pros and cons. At the very least, this feature provides an alternative way of doing some operations and is the only way to perform others.

Advantages of Array Formulas

The following are some of the advantages that may be obtained by using array formulas (as opposed to single-cell formulas):

- Much more efficient to work with
- Eliminate the need for intermediary formulas
- Enable you to do things that would otherwise be difficult or impossible
- Use less memory

Disadvantages of Array Formulas

This list shows a few disadvantages of array formulas:

- Some large arrays can slow your spreadsheet recalculation time to a crawl.
- Arrays can make your worksheets more difficult for others to understand.
- You must remember to enter an array formula with a special key sequence (Ctrl+Shift+Enter). Otherwise, the result isn't what you expect.
- Array formulas cannot be exported to other spreadsheet formats (such as Lotus 1-2-3).

Understanding Arrays

This section presents several examples to help clarify the concept of arrays. As always, you can get more from this chapter if you follow along on your own computer.

Array Formulas Versus Standard Formulas

You often can use a single array formula to substitute for a range of copied formulas. Figure 20-2 shows two examples; the upper worksheet uses standard single-result formulas. The formulas use the SQRT function to calculate the square roots of the values in column A. The formula =SQRT(A3) was entered into cell B3 and copied to the three cells below it. This example uses four different formulas to calculate the results in column B.

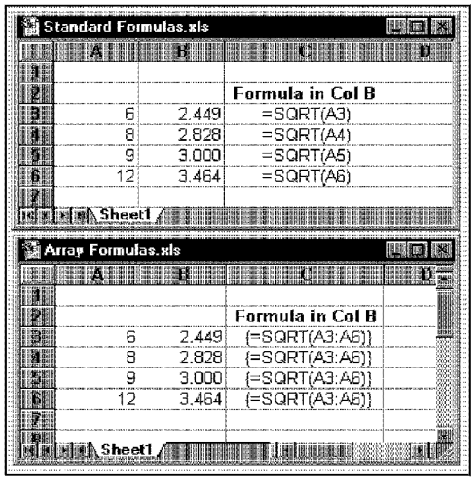


Figure 20-2: These workbooks accomplish the same result, but one uses standard formulas and the other uses an array formula.

The lower workbook uses a single array formula, which is inserted into all four cells. Use the following steps to enter this array formula:

1. Select the range B3:B6.
2. Enter **SQRT(A3:A6)**.
3. Press Ctrl+Shift+Enter to designate the formula as an array formula.

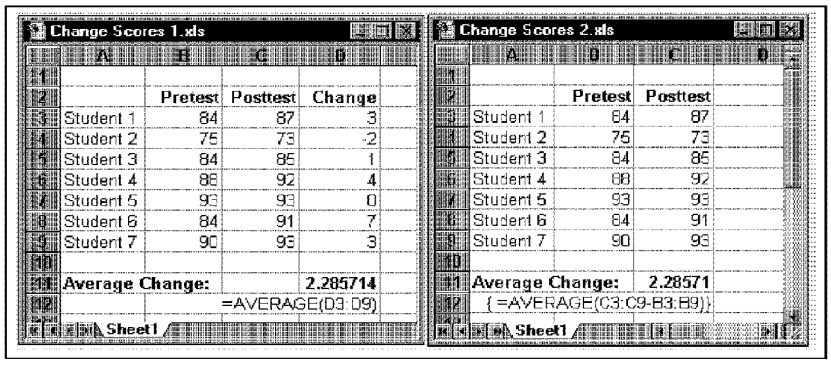
Excel enters the array formula into the three selected cells. It also adds brackets around the formula to indicate that it's an array formula. The key point here is that this example uses only one formula, but the results appear in four different cells, because the formula is operating on a four-cell array.

To further demonstrate that this is, in fact, one formula, try to edit one of the cells in B3:B6. You find that Excel doesn't let you make any changes. To modify an array formula that uses more than one cell, you must select the entire array before you edit the formula.

Virtually no advantage is gained by using an array formula in the preceding example (except perhaps to save the time that it takes to copy the formula). The real value of array formulas becomes apparent as you work through this chapter.

An Array Formula in One Cell

Figure 20-3 shows another example. The worksheet on the left uses standard formulas to calculate the average change from the pretest to the posttest. The worksheet on the right also calculates the average changes, but it uses an array formula. This array formula resides in only one cell, because the result is a single value. This is an example of how an array formula can eliminate the need for intermediary formulas. As you can see, you don't need to include an additional column to calculate the change in scores.



Change Scores 1.xls

	A	B	C	D
1				
2		Pretest	Posttest	Change
3	Student 1	84	87	3
4	Student 2	75	73	-2
5	Student 3	84	85	1
6	Student 4	88	92	4
7	Student 5	93	93	0
8	Student 6	84	91	7
9	Student 7	90	93	3
10				
11	Average Change:			2.285714
12			=AVERAGE(D3:D9)	

Change Scores 2.xls

	A	B	C	D
1				
2		Pretest	Posttest	
3	Student 1	84	87	
4	Student 2	75	73	
5	Student 3	84	85	
6	Student 4	88	92	
7	Student 5	93	93	
8	Student 6	84	91	
9	Student 7	90	93	
10				
11	Average Change:		2.28571	
12		{=AVERAGE(C3:C9-B3:B9)}		

Figure 20-3: Using an array formula to eliminate intermediary formulas.

The formula in cell C11 is as follows:

```
{=AVERAGE(C3:C9-B3:B9)}
```

This array formula operates on two arrays, which are stored in cells. It subtracts each element of B3:B9 from the corresponding element in C3:C9 and produces (in memory) a new seven-element array that holds the result. The AVERAGE function computes the average of the elements in the new array, and the result is displayed in the cell.

Looping with Arrays

Excel's array feature enables you to perform individual operations on each cell in a range — in much the same way as a program language's looping feature enables you to work with elements of an array. For example, assume that you have a range of cells (named Data) that contains positive and negative values. You need to compute the average of just the positive values in the range. Figure 20-4 shows an example of this.

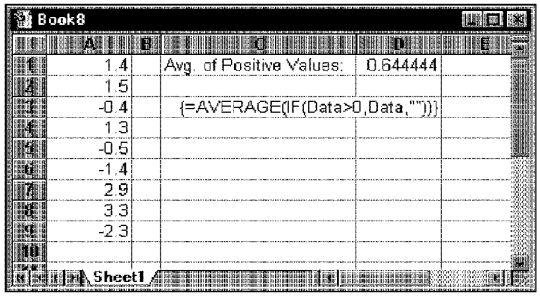


Figure 20-4: You can use an array formula to calculate the average of only the positive values in this range.

One approach is to sort the data and then use the AVERAGE function to calculate the average on only the positive values. A more efficient approach uses the following array formula:

```
{=AVERAGE(IF(Data>0,Data,""))}
```

The IF function in this formula checks each element in the input range to see whether it's greater than zero. If so, the IF function returns the value from the input range; otherwise, it returns an empty string. The result is an array that's identical to the input array, except that all nonpositive values are replaced with a null string (the third argument of the IF functions). The AVERAGE function then computes the average of this new array, and the result is displayed in the cell.

The preceding problem can also be solved with the following nonarray formula:

```
=SUMIF(Data,">0",Data)/COUNTIF(Data,">0")
```

Many similar operations can't be performed with a standard formula, however. For example, to calculate the median of the positive values in a range, an array formula is the only solution.

Some more useful examples that use arrays are presented later in this chapter, but for now, some rules are provided for how to work with arrays and array formulas.

Working with Arrays

This section deals with the mechanics of selecting arrays and entering and editing array formulas. These procedures are a little different from working with ordinary ranges and formulas.

Entering an Array Formula

When you enter an array formula into a cell or range, you must follow a special procedure, so that Excel knows that you want an array formula rather than a normal formula. You enter a normal formula into a cell by pressing Enter. You enter an array formula into one or more cells by pressing Ctrl+Shift+Enter.

You can easily identify array formulas, because they are enclosed in brackets in the formula bar. For example, {=SQRT(A1:A12)} is an array formula.

Don't enter the brackets when you create an array formula; Excel inserts them for you. If the result of an array formula consists of more than one value, you must select all the cells before you enter the formula. If you fail to do this, only the first result shows.

Editing an Array Formula

If an array formula occupies multiple cells, you must edit the entire range as though it is a single cell. The key point to remember is that you can't change just one element of an array formula. If you attempt to do so, Excel displays the messages shown in Figure 20-5.

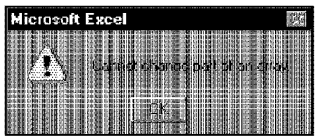


Figure 20-5: Excel's warning message reminds you that you can't edit just one cell of a multicell array.

The following rules apply to multicell array formulas. (If you try to do any of these things, Excel lets you know about it.):

- You can't change the contents of any cell that makes up an array formula.
- You can't move cells that make up part of an array formula. You can, however, move an entire array formula.
- You can't delete cells that form part of an array formula, but you can delete an entire array.
- You can't insert new cells into an array range; this rule includes inserting rows or columns that would add new cells to an array range.

To edit an array formula, select all the cells in the array range and activate the formula bar as usual (click it or press F2). Excel removes the brackets from the formula while you're editing it. Edit the formula and then press Ctrl+Shift+Enter to enter the changes. All the cells in the array now reflect your editing changes.

Selecting an Array Range

You can select an array range manually by using the normal selection procedures. Or, you can use either of the following methods:

- Move to any cell in the array range. Select Edit • Go To (or press F5), click the Special button, and then choose the Current Array option. Click OK to close the dialog box.
- Move to any cell in the array range and press Ctrl+/ to select the entire array.

Formatting Arrays

Although you can't change any part of an array formula without changing all parts, you're free to apply formatting to the entire array or to only parts of it.

Using Array Constants

So far, the examples in this chapter have used cell ranges to hold arrays. You can also use constant values as an array. These constants can be entered directly into a formula or defined by using the Define Name dialog box. Array constants can be used in array formulas in place of a reference to a range of cells. To use an array constant in an array formula, type the set of values directly into the formula and enclose it in brackets. If you defined a name for the array constant, you can use the name instead.

Array constants can be either one-dimensional or two-dimensional. One-dimensional arrays can be either vertical or horizontal. The elements in a one-dimensional horizontal array are separated by commas. The following example is a one-dimensional horizontal array:

```
{1,2,3,4,5}
```

Because this array constant has five values, it requires five cells (in a row). To enter this array into a range, select a range that consists of one row and five columns. Then, enter **={1,2,3,4,5}** and press Ctrl+Shift+Enter.

When you use array constants, you must enter the brackets. Excel doesn't provide them for you. The following example is another horizontal array; it has seven elements:

```
{"Sun","Mon","Tue","Wed","Thu","Fri","Sat"}
```

Figure 20-6 demonstrates how to create a named array constant by using the Define Name dialog box.

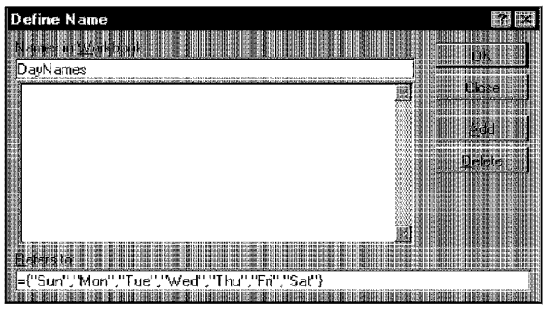


Figure 20-6: Creating an array constant in the Define Name dialog box.

The elements in a one-dimensional vertical array are separated by semicolons. The following is a six-element vertical array:

```
{10;20;30;40;50;60}
```

The following is another example of a vertical array; this one has four elements:

```
{"Widgets";"Sprockets";"Do-Dads";"Thing-A-Majigs"}
```

Two-dimensional arrays also separate the elements in a single row with commas and separate the rows with semicolons. The next example is a 3 × 4 array (three rows, each of which occupies four columns):

```
{1,2,3,4;5,6,7,8;9,10,11,12}
```

Figure 20-7 shows how this array appears in a worksheet. First, the array constant was created and named MyArray. Then, A1:D3 was selected and =MyArray was entered. The array formula was entered into the range by pressing Ctrl+Shift+Enter.

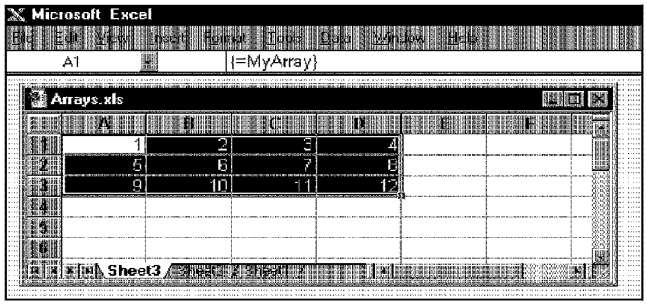


Figure 20-7: An array constant used in a formula.

You can't list cell references, names, or formulas in an array formula in the same way as you list constants. For example, {2*3,3*3,4*3} isn't valid, because it lists formulas. {A1,B1,C1} isn't valid, either, because it lists cell references. Instead, you should use a range reference, such as {A1:C1}.

You must remember an array's dimensions when you're performing operations on it. Consider the following array formula:

```
={2,3,4}*{10,11}
```

This formula multiplies a 1 × 3 array by a 1 × 2 array. Excel returns an array with three values: 20, 33, and #N/A. Because the second array wasn't large enough, Excel generated #N/A as the third element of the result.

Examples of Using Array Formulas

Perhaps the best way to learn about array formulas is by following examples and adapting them to your own needs. This section presents useful examples that give you a good idea of how you can use array formulas.

All the examples presented in this section can be found in a workbook on this book's companion CD-ROM.

Using an Array Constant

Figure 20-8 shows a practical example of an array constant.

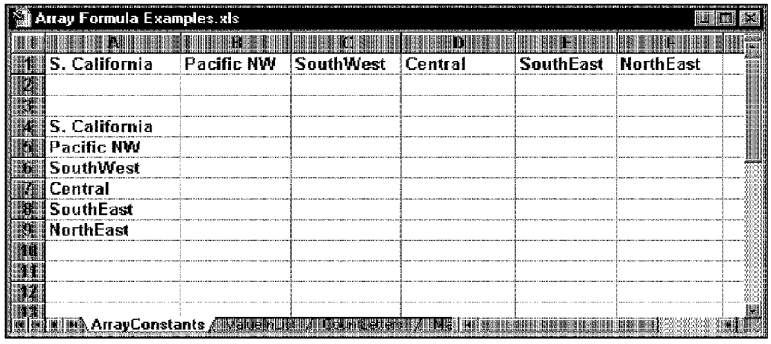


Figure 20-8: Using an array constant to enter the names of sales regions.

The following steps demonstrate how to create this example:

1. Define the following constant, named SalesRegions:

```
={"S. California";"Pacific NW";"SouthWest";"Central";"
SouthEast";" NorthEast"}
```

   Because the elements are separated by semicolons, this is a vertical array.

2. Select A4:A9 and enter **=SalesRegions**.
3. Press Ctrl+Shift+Enter.

The worksheet in Figure 20-8 also shows the sales regions displayed horizontally. To do this, select A1:F1 and then enter the following formula (by pressing Ctrl+Shift +Enter):

```
{=TRANSPOSE(SalesRegions)}
```

The TRANSPOSE function converts a horizontal array to a vertical array (and vice versa).

The method just described is one of several ways to enter a stored list quickly into a range of cells. Perhaps a better approach is to create a custom list in the Custom Lists panel of the Options dialog box.

Identifying a Value in a Range

To determine whether a particular value is contained in a range, choose Edit • Find. But, you also can make this determination by using an array formula. Figure 20-9 shows a worksheet with a list of names (named Names). An array formula in cell E4 checks the name that is entered into cell B1 (named TestValue). If the name exists, it displays the text Name is in the list. Otherwise, it displays Name not found.

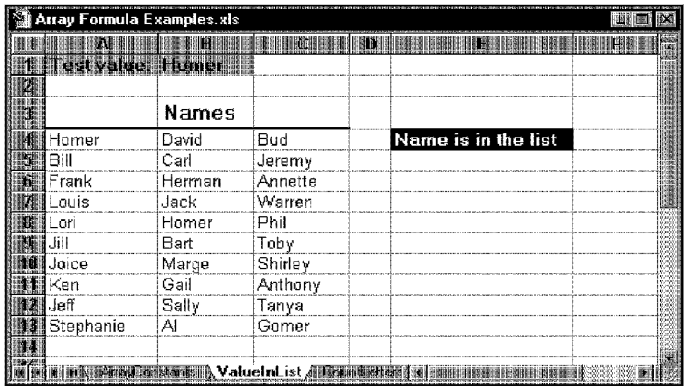


Figure 20-9: Determining whether a range contains a particular value.

The formula in cell E4 is as follows:

```
{=IF(OR(TestValue=Names),"Name is in the list","Name not
found")}
```

This formula compares TestValue to each cell in the range Names. It builds a new array that consists of logical TRUE or FALSE values. The OR function returns TRUE if any one of the values in the new array is TRUE. The IF function determines which message to display based on the result.

Counting Characters in a Range

This example demonstrates how to use nested functions in an array formula to loop through each element in the input range. Figure 20-10 shows a worksheet with text entered in a range named WordList.

The array formula in cell B1 is as follows:

```
{=SUM(LEN(WordList))}
```

This formula is quite straightforward. It creates an array that consists of the length of each word in the WordList range. Then, it uses the SUM formula to add the values in this new array. You can accomplish this without an array formula by using an additional column of formulas and then summing the results.

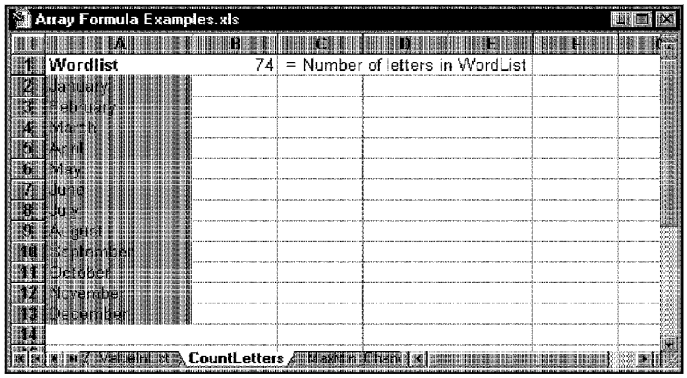


Figure 20-10: This array formula counts the number of characters in a range of text.

Computing Maximum and Minimum Changes

Figure 20-11 shows another example of how an array formula can eliminate the need for intermediary formulas. This worksheet shows two test scores for a group of students. Array formulas compare the two tests and calculate the largest decrease and the largest increase.

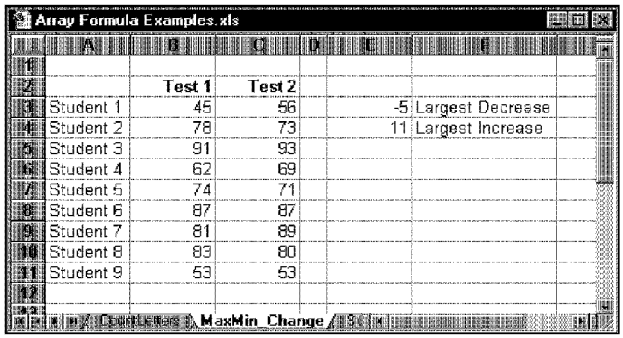


Figure 20-11: Array formulas determine the largest decrease and the largest increase in test scores.

The formulas are as follows:

```
E3:      {=MIN(C3:C11-B3:B11)}
E4:      {=MAX(C3:C11-B3:B11)}
```

Looping Through Characters in a Cell

The following array formula calculates the sum of the digits in an integer, which is stored in a cell named Number:

```
{=SUM(VALUE(MID(Number,ROW($A$1:OFFSET($A$1,LEN(Number)-
1,0)),1)))}
```

This is a rather complex formula that makes use of an interesting trick; thus, it is explained next one part at a time, so that you can see how it works. (Figure 20-12 shows an example.)

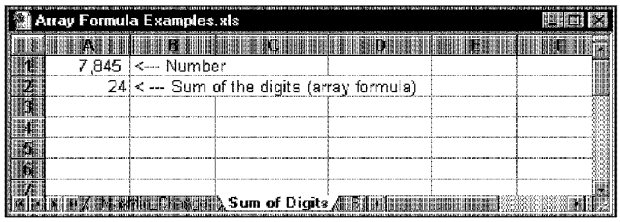


Figure 20-12: An array formula calculates the sum of the digits in a value.

You may be confused by the ROW function (this is the trick). This function is used to generate an array of consecutive integers, beginning with 1 and ending with the number of digits in the absolute value of Number.

If Number is 489, then LEN(Number) is 3. The ROW function can then be simplified as follows:

```
{=ROW($A$1:OFFSET($A$1,3-1,0))}
```

This formula generates an array with three elements: {1,2,3}, which is used as the second argument for the MID function (the third argument is 1). The MID part of the formula, simplified a bit and expressed as values, is the following:

```
{=MID(489,{1,2,3},1)}
```

This formula generates an array with three elements: {4,8,9}. By simplifying again and adding the SUM function, the formula becomes as follows:

```
{=SUM({4,8,9})}
```

This produces the result of 21.

The following is another version of this formula that also works with negative numbers. The ABS function is added to calculate the absolute value of the result:

```
{=SUM(VALUE(MID(ABS(Number),ROW($A$1:OFFSET($A$1,LEN(ABS(Number
))-1,0)),1)))}
```

Summing Every *n*th Value in a Range

The next example can be quite useful. Suppose that you have a range of values and you want to compute the sum of every third value in the list — the first, the fourth, the seventh, and so on. You can't accomplish this with a standard formula. The following array formula does the job, however. It assumes that a cell named Nth determines which values to sum, and that the range to sum is named Data.

```
{=IF(nth=0,0,SUM(IF(MOD(ROW($A$1:OFFSET($A$1,COUNT(Data)-
1,0)),nth)=0,Data,0)))}
```

The formula uses the MOD function to determine which values to sum. The first argument for the MOD function is as follows:

```
ROW($A$1:OFFSET($A$1,COUNT(Data)-1,0))
```

This expression generates an array that begins with 1 and ends with the number of cells in the Data range. If the MOD function returns 0, the value is included in the array to sum.

Notice that a special case exists for when Nth is 0 (that is, sum every cell in the range), because the MOD function returns an error when its second argument is 0.

This formula has a limitation: It works only when Data consists of a single column of values, because it uses the ROW function to determine the element in the array.

Figure 20-13 shows an example that uses the preceding array formula, plus a series of intermediary formulas to calculate the result without using an array formula.

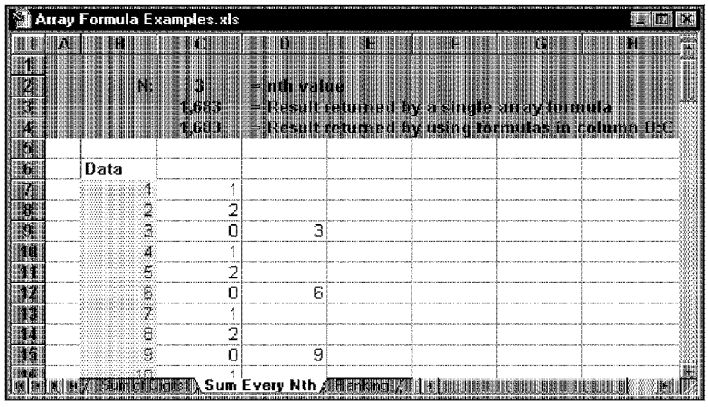


Figure 20-13: You can use an array formula to sum every *n*th element in a range—or use a series of intermediary formulas (a less-efficient approach).

An Alternate Method of Ranking

Often, computing rank orders for a range of data is helpful. If you have a worksheet with the annual sales figures for 20 salespeople, for example, you may want to know how each person ranks, from highest to lowest.

If you do this sort of work, you've probably discovered Excel's RANK function. You also may have noticed, however, that the ranks produced by this function don't handle ties the way that you may like. For example, if two values are tied for third place, they both receive a rank of 3. Many people prefer to assign each an average (or midpoint) of the ranks—that is, a rank of 3.5 for both values tied for third place.

Figure 20-14 shows a worksheet that uses two methods to rank a column of values (named Sales). The first method (column C) uses Excel's RANK function. Column D uses array formulas to compute the ranks.

The following is the array formula in cell D2:

```
{=IF((SUM(IF(Sales=B2,1)))=1,(SUM(IF(Sales>=B2,1,0))),(SUM(IF(S
ales>=B2,1)))-((SUM(IF(Sales=B2,1)))-1)*0.5)}
```

This formula was entered into cell D2 and then copied to the cells below it.

The formula is rather complex, but breaking it down into parts should help you understand how it works.

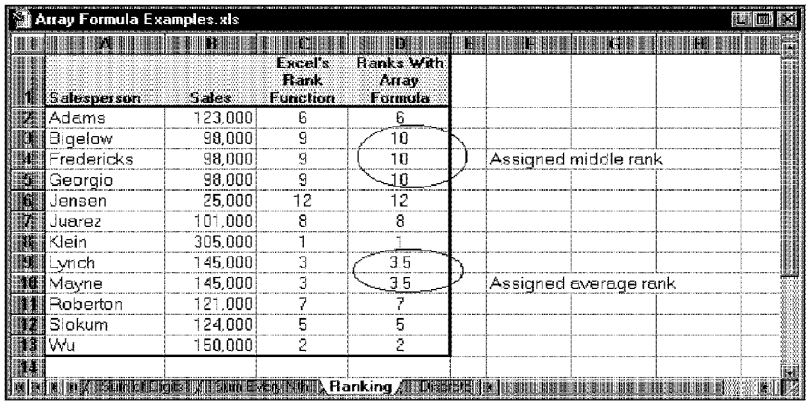


Figure 20-14: Ranking data with Excel's RANK function and with array formulas.

Frequency Distributions

Before Excel 5, the only way to calculate frequency distributions was to use array formulas. Beginning with Excel 5, however, the COUNTIF function provides a more direct way to generate frequency distributions.

Figure 20-15 shows a worksheet with a series of scores in column A that range from 1 to 4. Column D contains array formulas to calculate the frequency of each score.

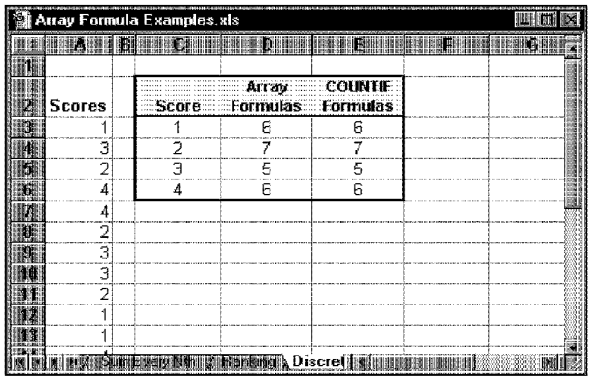


Figure 20-15: Calculating discrete frequency distributions by using array formulas and COUNTIF functions.

The formula in D6 is as follows:

```
{=SUM(IF(Scores=C3,1))}
```

The corresponding formulas in column E use the COUNTIF function. The following is the formula in cell E6:

```
=COUNTIF(Scores,C3)
```

Both of these methods count specific values. But what if the scores are noninteger values, as in Figure 20-16? Both types of formulas require modification to handle noninteger data.

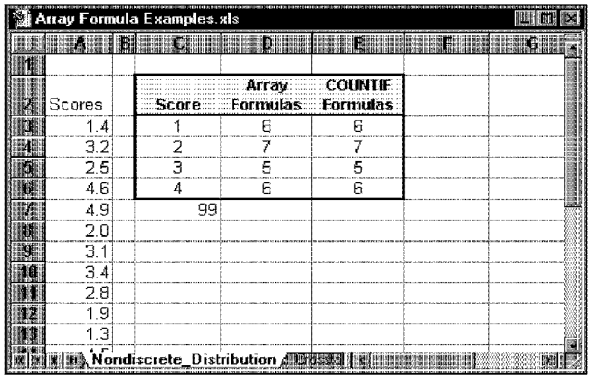


Figure 20-16: Calculating nondiscrete frequency distributions by using array formulas and COUNTIF functions.

The array formula can be modified as follows:

```
=SUM(IF(Scores>=C3,1))-SUM(IF(Scores>=C4,1))
```

The following is the revised COUNTIF formula:

```
=COUNTIF(Scores,">="&C3)-COUNTIF(Scores,">="&C4)
```

The array formula requires you to add an additional value in column C, so that the last array formula doesn't refer to an empty cell (I added a value of 99).

You also can compute distributions by using the Histogram tool in the Analysis ToolPak (see Chapter 28). An advantage to using arrays or COUNTIF functions, however, is that these procedures are dynamic and display the correct values if you change the input data.

Dynamic Crosstabs

The preceding section demonstrates that using COUNTIF is better than using array formulas to calculate frequency distributions. This section demonstrates how to extend these distributions into another dimension and create crosstabs. In this case, an array formula is the only method that can get the job done. This technique enables you to create a dynamic crosstab table that is updated automatically whenever the data is changed. Even a pivot table can't do that!

The worksheet in Figure 20-17 shows a simple expense account listing. Each item consists of the date, the expense category, and the amount spent. Each column of data is a named range, indicated in the first row.

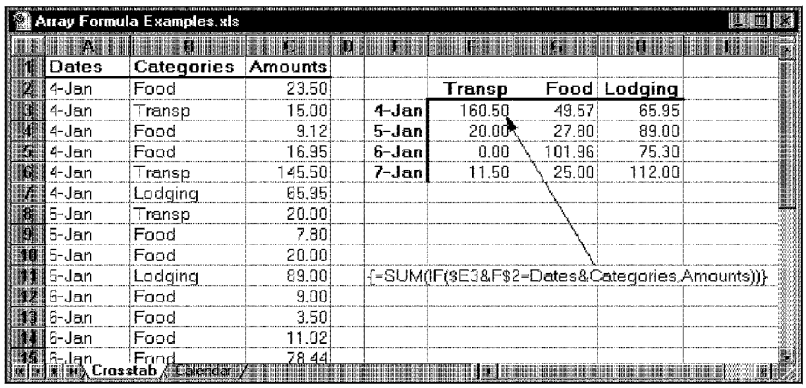


Figure 20-17: You can use array formulas to summarize data such as this in a dynamic crosstab table.

Array formulas were used to summarize this information into a handy table that shows the total expenses, by category, for each day. Cell F3 contains the following array formula, which was copied to the remaining 11 cells in the table:

```
{=SUM(IF($E3&F$2=DATES&CATEGORIES,AMOUNTS))}
```

These array formulas display the totals for each day, by category.

This formula operates similarly to the more simple one demonstrated in the preceding section. This formula has a few new twists, however. Rather than count the number of entries, the formula adds the appropriate value in the Amounts range. It does so, however, only if the row and column names in the summary table match the corresponding entries in the DATES and CATEGORIES ranges. It does the comparison by concatenating (using the & operator) the row and column names and comparing the resulting string to the concatenation of the corresponding DATES and CATEGORIES values. If the two match, the =SUM function kicks in and adds the corresponding value in the AMOUNTS range.

This technique can be customized, of course, to hold any number of different categories and any number of dates. You can eliminate the dates, in fact, and substitute people's names, departments, regions, and so on.

You also can cross-tabulate data by creating a pivot table. But, unlike a pivot table, using the procedure described here is completely dynamic (a pivot table must be updated if the data changes).

Pivot tables are discussed in Chapter 25.

Returning the Last Value in a Column

Suppose that you have a worksheet that you update frequently and need to determine the most recently entered value in a column. The following array formula returns the contents of the last nonempty cell in the first 500 rows of column A:

```
=INDIRECT(ADDRESS(MAX((ROW(1:500)*(A1:A500<>""))),COLUMN(A:A)))
```

You can modify this formula to work with a different column, and with a different number of rows in the column. To use a different column, change the column references from A to whichever column you need. To check more than 500 rows, change the two references to row 500.

Returning the Last Value in a Row

The following array formula is similar to the previous formula, but it returns the last nonempty cell in a row (in this case, row 1):

```
=INDIRECT(ADDRESS(1,(MAX((TRANSPOSE(ROW(1:256))*(1:1<>""))))))
```

To use this formula for a different row, change the first argument for the ADDRESS function to the new ADDRESS function and the 1:1 reference to correspond to the row.

A Single-Formula Calendar

The final array formula example is perhaps the most impressive. Figure 20-18 shows a monthly calendar that is calculated using a single array formula entered in B6:H11. This workbook includes a few additional bells and whistles. For example, you can choose the month and year to display by using dialog box controls that are inserted directly on the worksheet. When you change the month or year, the calendar is updated immediately.

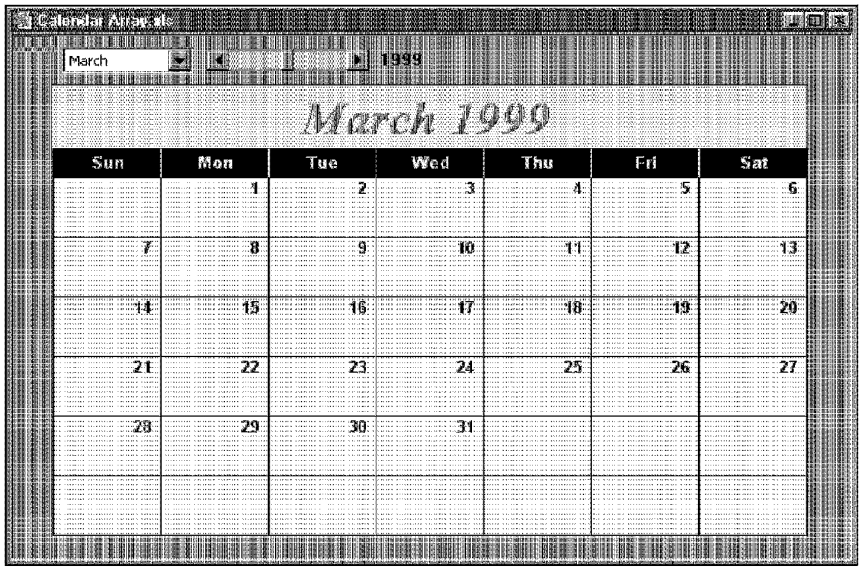


Figure 20-18: This calendar is calculated with a single array formula.

The array formula is as follows:

```
{=IF(MONTH(StartDate)<>MONTH(StartDate-StartDOW+Week*7+Weekday-
1),"",StartDate-StartDOW+Week*7+Weekday-1)}
```

This formula uses a few cell references (StartDate and StartDOW) and two named array constants, which are defined as follows:

```
Week:    ={0;1;2;3;4;5}
Weekday: ={1,2,3,4,5,6,7}
```

I leave it up to you to figure out how this works. Suffice it to say that it took more than a few minutes to develop.

The companion CD-ROM contains a workbook that uses the single-formula calendar.

Tips for Array Formulas

If you've followed along in this chapter, you probably understand the advantages of using array formulas. As you gain more experience with arrays, you undoubtedly will discover some disadvantages.

The primary problem with array formulas is that they slow your worksheet's recalculations, especially if you use large arrays. On a faster system, this may not be a problem. But if you have a slower system and speed is of the essence, you should probably avoid using large arrays.

Array formulas are one of the least understood features of Excel. Consequently, if you plan to share a worksheet with someone who may need to make modifications, you should probably avoid using array formulas. Encountering an array formula when you don't know what it is can be confusing.

You may also discover that you can easily forget to enter an array formula by pressing Ctrl+Shift+Enter. If you edit an existing array, you still must use these keys to complete the edits. Except for logical errors, this is probably the most common problem that users have with array formulas. If you press Enter by mistake after editing an array formula, just double-click the cell to get back into Edit mode and then press Ctrl+Shift+Enter.

Summary

This chapter introduces the concept of *array formulas*, a special type of formula that operates on a group of cells. You can write an array formula by entering a single formula that performs an operation on multiple inputs and produces multiple results—with each result displayed in a separate cell. This chapter also presents several practical examples of array formulas.

CHAPTER 21

Using Excel in a Workgroup

In This Chapter

If you use Excel on a standalone computer—a PC that's not connected to a network—you can skip this chapter, because it applies only to users who run Excel on a network.

Using Excel on a Network

A computer network consists of a group of PCs that are linked. A common type of network uses a *client-server model*, in which one or more PCs on the network act as dedicated *servers*, because they store files centrally and supply information, while user PCs are called *clients* (they use data in the centrally stored files on the server). Other networks are *peer-to-peer networks* that don't have a central server. Users on a network can perform the following tasks:

- Access files on other systems
- Share files with other users
- Share resources such as printers and fax modems
- Communicate with each other electronically

In many offices, networks now perform functions that formerly required a mainframe system and *dumb* terminals. Networks are usually less expensive, easier to expand, more manageable, and more flexible in terms of software availability than a mainframe system.

This chapter discusses the Excel features that are designed for network users.

File Reservations

Networks provide users with the ability to share information stored on other computer systems. Most networks have one

or more file servers attached. A file server stores files that members of a workgroup share. A network's file server may contain, for example, files that store customer lists, price lists, and form letters. Keeping these files on a file server has two major advantages:

- It eliminates the need to have multiple copies of the files stored locally on user PCs.
- It ensures that the file is always up to date; for example, if everyone makes changes to the same shared copy of a customer list, there's little likelihood that the portions of the list will be correct while other portions will be obsolete.

Some software applications are *multiuser applications*. Most database software applications, for example, enable multiple users to work simultaneously on the same database files. One user may be updating customer records in the database, while another is extracting records. But what if a user is updating a customer record and another user wants to make a change to that same record? Multiuser database software contains record-locking safeguards that ensure only one user at a time can modify a particular record.

Excel is *not* a multiuser application. When you open an Excel file, the entire file is loaded into memory. If the file is accessible to other users, you wouldn't want someone else to open a file that you've opened. If Excel allowed you to open and change a file that someone else on a network has already opened, the following scenario could happen.

Assume that your company keeps its sales information in an Excel file that is stored on a network server. Elaine wants to add this week's data to the file, so she loads it from the server and begins adding new information. A few minutes later, Albert loads the file to correct some errors that he noticed last week. Elaine finishes her work and saves the file. A while later, Albert finishes his corrections and saves the file. Albert's file overwrites the copy that Elaine saved, and her additions are gone.

This scenario *can't happen*, because Excel uses a concept known as *file reservation*. When Elaine opens the sales file, she has the reservation for the file. When Albert tries to open the file, Excel informs him that Elaine is using the file. If he insists on opening it, Excel opens the file as *read-only*. In other words, Albert can open the file, but he can't save it under the same name. Figure 21-1 shows the message that Albert receives if he tries to open a file that is in use by someone else.

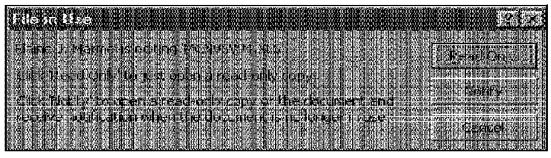


Figure 21-1: The File in Use dialog box appears if you try to open a file that someone else is using.

Albert has these three choices:

- **Select Cancel, wait a while, and try again.** He may call Elaine and ask her when she expects to be finished.
- **Select Read Only.** This lets him open the file to read it, but doesn't let him save changes to the same filename.
- **Select Notify, which opens the file as read-only.** Excel pops up a message when Elaine is finished using the file.

Figure 21-2 shows the message that Albert receives when the file is available.

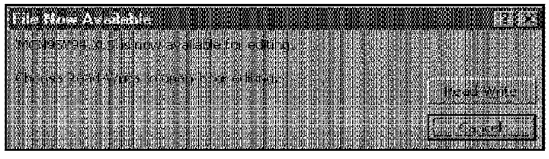


Figure 21-2: The File Now Available dialog box pops up with a new message when the file is available for editing.

Shared Workbooks

Although Excel isn't a multiuser application, it does support a feature known as *shared workbooks,* which enables multiple users to work on the same workbook simultaneously. Excel keeps track of the changes and provides appropriate prompts to handle conflict.

Appropriate Workbooks for Sharing

Although you can designate any workbook as a shared list, only certain workbooks contain information that is appropriate for sharing. The following are examples of workbooks that work well as shared lists:

- **Project tracking:** You may have a workbook that contains status information for projects. If multiple people are involved in the project, they can make changes and updates to the parts that are relevant.
- **Customer lists:** With customer lists, changes usually occur infrequently, but records are added and deleted.
- **Consolidations:** You may create a budget workbook in which each department manager is responsible for his or her department's budget. Usually, each department's budget appears on a separate sheet, with one sheet serving as the consolidation sheet.

Limitations of Shared Workbooks

If you plan to designate a workbook as shared, be aware that you cannot perform any of the following actions while sharing the workbook:

- Delete worksheets or chart sheets.
- Insert or delete a block of cells. However, you can insert or delete entire rows and columns.
- Merge cells.
- Define or apply conditional formats.
- Set up or change data-validation restrictions and messages.
- Insert or change charts, pictures, drawings, objects, or hyperlinks.
- Assign or modify a password to protect individual worksheets or the entire workbook.
- Create or modify pivot tables, scenarios, outlines, or data tables.
- Insert automatic subtotals.
- Make changes to dialog boxes or menus.
- Write, change, view, record, or assign macros. However, you can record a macro in a shared workbook that you store in another, unshared workbook.

Designating a Workbook As a Shared Workbook

To designate a workbook as a shared workbook, select Tools • Share Workbook. Excel displays the dialog box that is shown in Figure 21-3. This dialog box has two tabs: Editing and Advanced. In the Editing tab, select the check box to allow changes by multiple users and then click OK. Excel then prompts you to save the workbook.

When you open a shared workbook, the window's title bar displays [Shared]. If you no longer want other users to be able to use the workbook, remove the check mark from the Share Workbook dialog box and save the workbook.

Whenever you're working with a shared workbook, you can find out whether any other users are working on the workbook. Choose Tools • Share Workbook, and the Share Workbook dialog box lists the names of the other users who have the file open, as well as the time that each user opened the workbook.

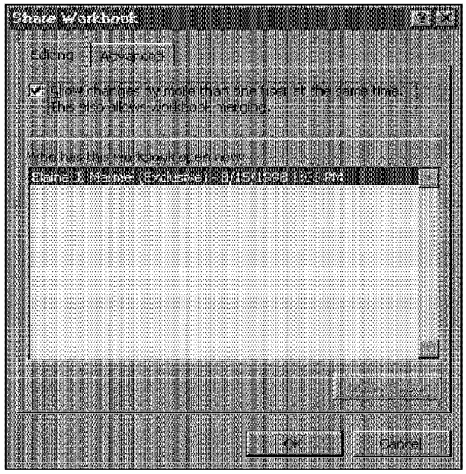


Figure 21-3: The Share Workbook dialog box lets you specify a workbook as a shared workbook.

Advanced Settings

Excel allows you to set options for shared workbooks. Select Tools • Share Workbook and click the Advanced tab to access these options (see Figure 21-4).

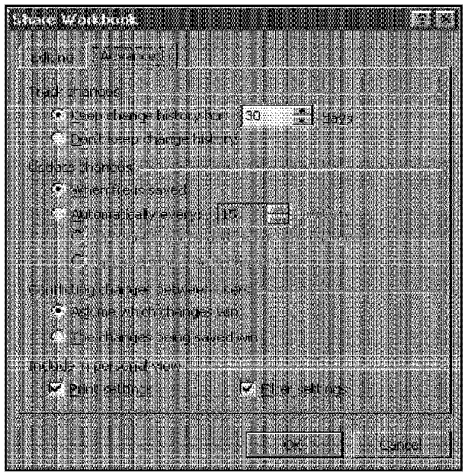


Figure 21-4: The Advanced tab of the Share Workbook dialog box.

Tracking changes

Excel can keep track of the workbook's changes — something known as *change history*. When you designate a workbook as a shared workbook, Excel automatically turns on the change history option, enabling you to view information about previous (and perhaps conflicting) changes to the workbook. You can turn off change history by selecting the option labeled Don't keep change history. You can also specify the number of days for which Excel tracks change history.

Updating changes

While you're working on a shared workbook, you can use the standard File • Save command to update the workbook with your changes. The Update changes settings determine what happens when you save a shared workbook:

- **When file is saved:** You receive updates from other users when you save your copy of the shared workbook.
- **Automatically every:** Lets you specify a time period for receiving updates from other users of the workbook. You can also specify whether Excel should save your changes automatically, too, or just show you the changes made by other users.

Conflicting changes between users

As you may expect, multiple users working on the same file can result in some conflicts. For example, assume that you're working on a shared customer database workbook, and another user also has the workbook open. If you and the other user both make a change to the same cell, a conflict occurs. You can specify the manner in which Excel resolves the conflicts by selecting one of two options in the Advanced tab of the Share Workbook dialog box:

- **Ask me which changes win:** If you select this option, Excel displays a dialog box to let you determine how to settle the conflict.
- **The changes being saved win:** If you select this option, your changes always take precedence.

Include in personal view

The final section of the Advanced tab of the Share Workbook dialog box enables you to specify settings that are specific to your view of the shared workbook. You can choose to use your own print settings and your own data-filtering settings. If you don't place checks in these check boxes, you can't save your own print and filter settings.

Mailing and Routing Workbooks

Excel provides a few additional workgroup features. To use these features, your system must have one of the following items installed:

- Office 2000
- Microsoft Exchange
- A mail system that is compatible with MAPI (Messaging Application Programming Interface)
- Lotus cc:Mail
- A mail system that is compatible with VIM (Vendor Independent Messaging)

The procedures vary, depending on the mail system that you have installed; for this reason, discussions in the following sections are general in nature. For specific questions, consult your network administrator.

Mailing a Workbook As an E-mail Attachment

Electronic mail, or *e-mail*, is commonplace in most offices, and is an extremely efficient means of communication. Unlike a telephone, e-mail doesn't rely on the recipient of the message being available when you want to send the message.

In addition to sending messages by e-mail, you can send complete files—including Excel workbooks. Like a growing number of software applications, Excel is *mail-enabled,* which means that you don't have to leave Excel to send a worksheet to someone by e-mail.

To send a copy of your workbook to someone on your network, select File • Send To • Mail Recipient (as Attachment). Excel creates an e-mail message with a copy of the workbook attached, using your default e-mail program; in Figure 21-5, Excel opened Outlook Express to send the workbook. You send this e-mail message the same way that you send any message—from your e-mail program. You also can send the message to multiple recipients, the same way that you send any e-mail message to multiple recipients.

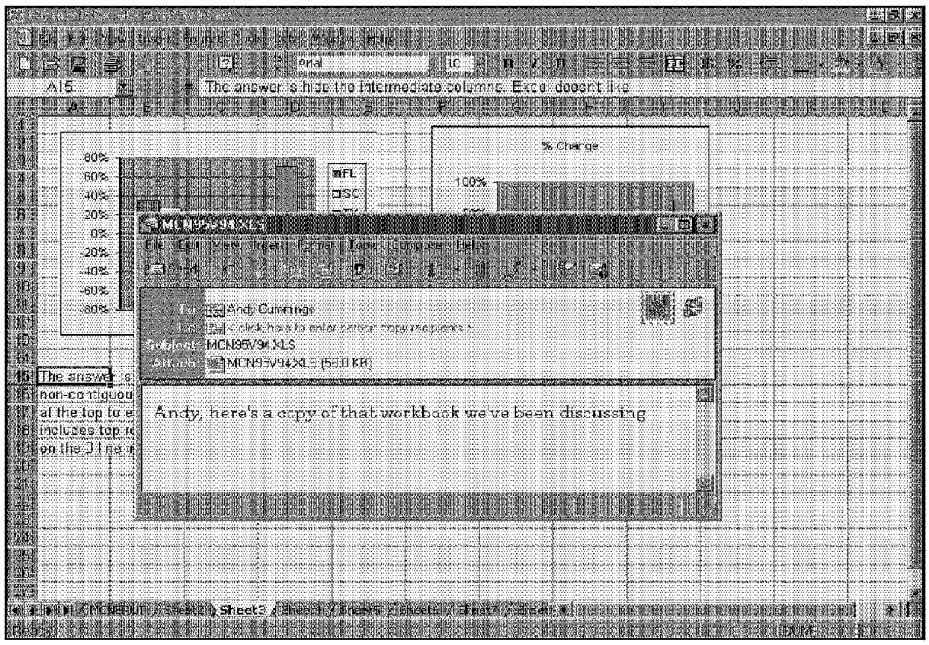


Figure 21-5: Sending a workbook as an attachment to an e-mail message.

When you send any file by using an e-mail program, you send a *copy* of the file. If the recipient makes changes to the notebook, the changes do not appear in your copy of the workbook.

Routing a Workbook to Others

If you choose File • Send To • Routing Recipient, Excel enables you to attach a routing slip to a workbook, similar to the one you see in Figure 21-6. Routing a workbook is most useful when you want the first person in the group to review (and possibly edit) the workbook and then send it to the next person on the list. For example, if you're responsible for your department's budget, you may need input from Alice—and her input may depend on Andy's input. You can set up the workbook and then route it to the others so that they can make their respective additions. When you set up the routing slip, you can tell Excel to return the workbook to you when the routing is finished.

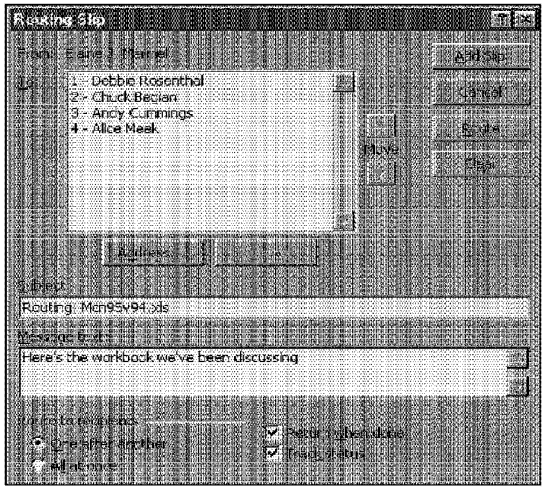


Figure 21-6: Routing a workbook.

When you route a workbook, you have the following two options:

- **Sequential routing:** Enables you to route the workbook sequentially to workgroup members. When the first recipient is finished, the workbook goes to the second recipient. When the second recipient is finished, the workbook goes to the third, and so on. When all recipients have received the workbook, it can be returned to you. Choose One after another at the bottom of the Routing Slip dialog box for this type of routing.
- **Simultaneous routing:** Enables you to route the workbook to all recipients at once. You receive a copy of the workbook from each recipient (not just one copy). This type of routing is useful if you want to solicit comments from a group of coworkers, and you want the responses back quickly (you don't want to wait until a single worksheet makes the circuit). Choose All at once at the bottom of the Routing Slip dialog box for this type of routing.

Click Route to route the workbook immediately. If you don't want to route immediately, click Add Slip. Later, when you're ready to route, choose File • Send To • Next Routing Recipient. Either choice places the workbook in the outgoing mail folder of your e-mail program. To actually route the workbook, open your e-mail program to send the message.

Whether you route or attach a workbook to an e-mail message, Excel uses your e-mail program. Since you can send a workbook to a number of people, either as an e-mail attachment or by using a routing slip, the distinction between the two methods lies in the distinction between sequential and simultaneous routing. If you choose simultaneous routing and you *don't* place a check in the Return when done check box, routing and attaching are identical, because you can't guarantee a reply to e-mail.

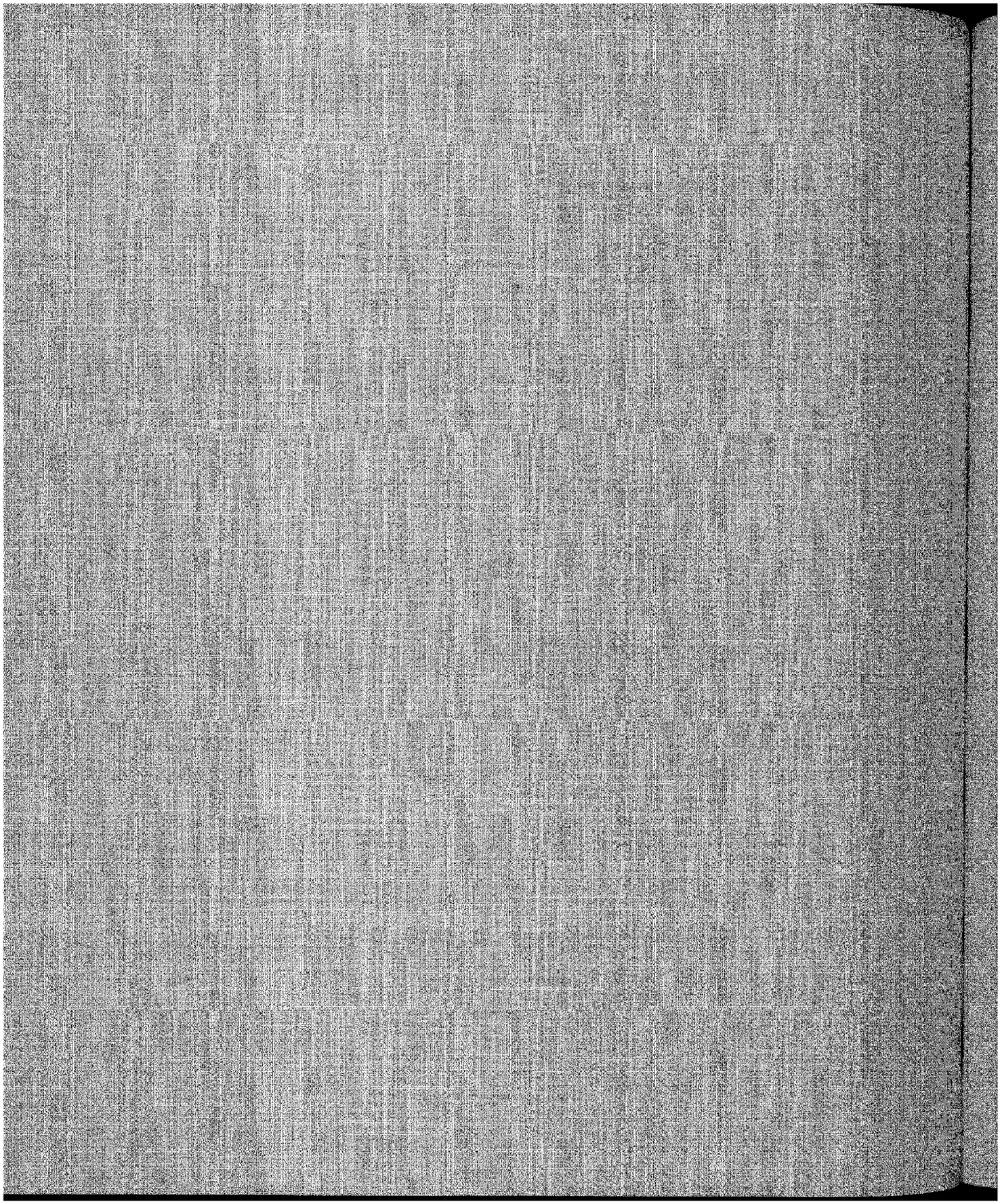
Summary

This chapter presents a basic overview of using Excel in a network environment. It explains how the concept of a file reservation prevents two users from modifying a workbook simultaneously. Excel's shared workbook feature, however, lets multiple users work on a single workbook at the same time. The chapter concludes with a discussion of mailing and routing workbooks.

Analyzing Data

PART IV

In This Part



CHAPTER 22

Importing Data from Other Sources

In This Chapter

When you get right down to it, Excel can be described as a tool that manipulates data—the numbers and text that you use in a worksheet. But before you can manipulate data, it must be present in a worksheet. This chapter describes a variety of data-importing techniques.

An Overview of Importing

The following are the six basic ways to import data into Excel:

- Enter the data manually by typing values and text into cells
- Generate data by using formulas or macros
- Use Query (or a pivot table) to import data from an external database
- Import data from an HTML document on the Internet or a corporate intranet
- Copy data from another application by using the Windows Clipboard
- Import data from another (non-Excel) file

This chapter deals primarily with the last two methods: Clipboard copying and foreign-file importing.

Chapter 29 is somewhat related to this topic. It deals with linking to and from other applications and embedding objects. Querying external databases is covered in Chapter 24, and pivot tables are covered in Chapter 25. Chapter 30 discusses how Excel works with the Internet.

A Few Words About Data

Data is a broad concept that means different things to different people. Data is basically raw information that can come in any number of forms. For example, data can be numbers, text, or a combination. Most of what you do in Excel involves manipulating data in one way or another.

As computers become more commonplace, data is increasingly available in machine-readable formats (otherwise known as *files*). Not too long ago, major data suppliers provided printed reports to their clients. Now, data suppliers commonly offer a choice of formats: paper or disk.

Data that is stored in files can be in a wide variety of formats. Common file formats for distributing data include Lotus 1-2-3 files (WKS and WK1), dBASE files (DBF), and text files (which come in several varieties). Excel's file format is rather complex, and the format tends to change with every new version of Excel. Consequently, the Excel file format is not widely used for the general distribution of data.

The file format for Excel 2000 files is the same as the file format for Excel 97. However, if you use an Excel 2000 file in Excel 97, you will have access only to Excel 97 features.

As an Excel user, you need to understand the types of data that you can access either directly or indirectly.

File Formats Supported by Excel

Rarely does a computer user work with only one application or interact only with people using the same applications he or she uses. Suppose that you're developing a spreadsheet model that uses data from last year's budget, which is stored in your company's mainframe. You can request a printout of the data, of course, and manually enter it into Excel. If the amount of data isn't too large, this route may be the most efficient. But what if you have hundreds of entries to make? Your mainframe probably can't generate an Excel workbook file, but an excellent chance exists that it can send the report to a text file, which you can then import into an Excel worksheet. Potentially, you can save several hours of work and virtually eliminate data-entry errors.

As you know, Excel's native file format is an XLS file. In addition, Microsoft included the capability to read other file formats directly. For example, you can open a file that was created in several other spreadsheet products, such as Lotus 1-2-3 and Quattro Pro. Table 22-1 lists all the file formats that Excel can read (excluding its own file types).

To open any of these files, choose File • Open and select the file type from the drop-down list labeled Files of type (see Figure 22-1)to display only the files of the selected

type in the file list. If the file is a text file, Excel's Text Import Wizard appears, to help you interpret the file. The Text Import Wizard is discussed later in this chapter.

Table 22-1
File Formats Supported by Excel

File Type	*Description*
Text	Space delimited, tab delimited, and comma delimited
Lotus 1-2-3	Spreadsheet files generated by Lotus 1-2-3 for DOS Release 1.*x*, Release 2.*x*, Release 3.*x*, and 1-2-3 for Windows
Quattro Pro/DOS	Files generated by Novell's Quattro Pro for DOS spreadsheet
Microsoft Works 2.0	Files generated by Microsoft Works 2.0
dBASE	Database files in the DBF format
SYLK	Files generated by Microsoft's MultiPlan spreadsheet
Data Interchange Format	Files generated by the VisiCalc spreadsheet
HTML	Files developed for the World Wide Web
Quattro Pro for Windows	Files generated by Corel's Quattro Pro for Windows spreadsheet

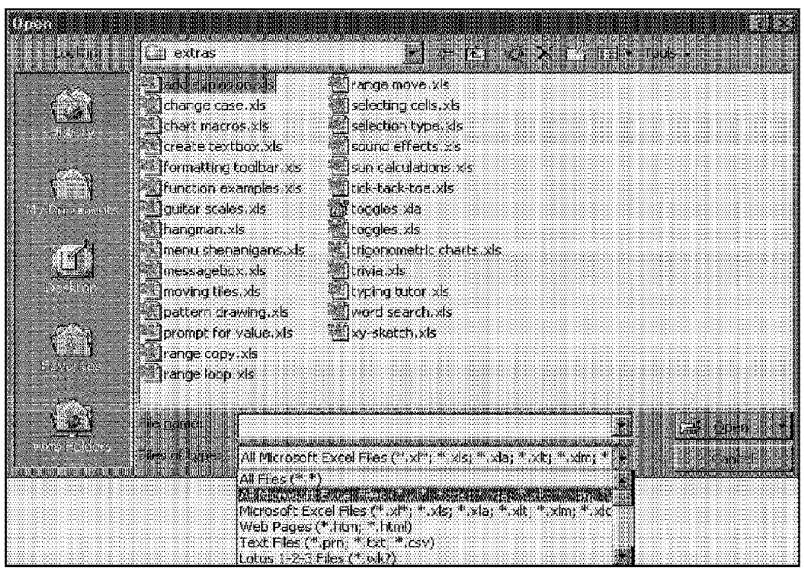


Figure 22-1: Use the Open dialog box to import a foreign file.

You should understand, however, that being able to read a file and translating it perfectly are two different matters. In some cases, you may encounter one or more of the following problems while reading a foreign file into Excel:

- Some formulas aren't translated correctly
- Unsupported functions aren't translated
- Formatting is incorrect
- Column widths are incorrect

When you open a file that wasn't produced by Excel, examine it carefully to ensure that Excel retrieved the data correctly.

The following sections discuss the various types of files that Excel can read. Each section discusses a file type and lists that file type's associated file extensions.

If a colleague sends you a file that Excel can't open, don't give up. Simply ask your colleague to save the spreadsheet in a format that Excel *can* read. For example, many applications can save files in 1-2-3 format, and most applications can export to a text file format.

Lotus 1-2-3 Spreadsheet Files

Lotus spreadsheets come in several flavors:

- **WKS files:** Single-sheet files used by 1-2-3 Release 1.*x* for DOS. Excel can read and write these files. If you export a workbook to a WKS file, Excel saves only the active worksheet, because 1-2-3 Release 1.*x* supports only one worksheet per workbook.
- **WK1 files:** Single-sheet files used by 1-2-3 Release 2.*x* for DOS. The formatting for these files is stored in ALL files (produced by the Allways add-in) or FM1 files (produced by the WYSIWYG add-in). Excel can read and write all of these file types. When you save a file to the WK1 format, you can choose which (if any) type of formatting file to generate. And, like WKS files, if you export a workbook to a WK1 file, Excel saves only the active worksheet.
- **WK3 files:** Multisheet (potentially) files generated by 1-2-3 Release 3.*x* for DOS, 1-2-3 Release 4.*x* for DOS, and 1-2-3 Release 1.*x* for Windows. The formatting for these files is stored in FM3 files (produced by the WYSIWYG add-in). Excel can read and write WK3 files with or without the accompanying FM3 file.
- **WK4 files:** Multisheet (potentially) files generated by 1-2-3 Release 4.*x* for Windows and 1-2-3 Release 5.*x* for Windows. Lotus combined formatting and data into one file, eliminating the separate formatting file. Excel can read and write these files.
- **1-2-3 files:** Multisheet (potentially) files generated by 1-2-3 97 (also known as Release 6) and 1-2-3 Millenium Edition (also known as Release 7). Excel can neither read nor write these files.

If you plan to import or export 1-2-3 files, I urge you to read the online Help for general guidelines and for specific types of information that may not be translated.

Excel evaluates some formulas differently from 1-2-3. To ensure complete compatibility when you work with an imported 1-2-3 file, choose Tools • Options, select the Transition tab, and then check the box labeled Transition Formula Evaluation.

Quattro Pro Spreadsheet Files

Quattro Pro files exist in several versions:

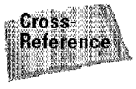
- **WQ1 files:** Single-sheet files generated by Quattro Pro for DOS Versions 1, 2, 3, and 4. Excel can read and write these files. If you export a workbook to a WQ1 file, Excel saves only the active worksheet.
- **WQ2 files:** Multisheet (potentially) files generated by Quattro Pro for DOS Version 5. Excel can neither read nor write this file format.
- **WB1 files:** Multisheet (potentially) files generated by Quattro Pro for Windows Versions 1 and 5 (there are no Versions 2 through 4). Excel can read (but not write) this file format.
- **WB2 files:** Multisheet (potentially) files generated by Quattro Pro for Windows Version 6. Excel can neither read nor write this file format.

Database File Formats

DBF files are single-table database files generated by dBASE and several other database programs. Excel can read and write DBF files up to and including dBASE 4.

If you have Microsoft Access installed on your system, you can take advantage of a feature that converts a worksheet list into an Access database file. To use this feature, you must install the Access Links add-in in Excel (you need your Office 2000 CD-ROM). Use the Data • Convert to MS Access command.

Excel can't read or write any other database file formats directly. If you install the Query add-in, however, you can use Query to access many other database file formats and then copy or link the data into an Excel worksheet.



See Chapter 24 for details on how to use Query to copy or link data from other database file formats into an Excel worksheet.

Text File Formats

Text files simply contain data — no formatting. The following relatively standard text file formats exist, although no standard file extensions exist:

- **Tab-delimited files:** Each line consists of fields that are separated by tabs. Excel can read these files, converting each line to a row and each field to a column. Excel also can write these files, using TXT as the default extension.
- **Comma-separated files:** Each line consists of fields that are separated by commas. Sometimes, text appears in quotation marks. Excel can read these files, converting each line to a row and each field to a column. Excel can also write these files, using CSV as the default extension.
- **Space-delimited files:** Each line consists of fields that are separated by spaces. Excel can read these files, converting each line to a row and each field to a column. Excel also can write these files, using PRN as the default extension.

If you want your exported text file to use a different extension, specify the complete filename and extension in quotation marks. For example, saving a workbook in comma-separated format normally uses the CSV extension. If you want your file to be named output.txt (with a TXT extension), enter **"output.txt"** in the File name box in the Save As dialog box.

When you attempt to load a text file into Excel, the Text Import Wizard kicks in to help you specify how you want Excel to retrieve the file (discussed in detail later in the chapter).

HTML Files

Excel can read and save files in HTML (Hypertext Markup Language) format, a file format that is used on the World Wide Web. And, through the use of XML (Extensible Markup Language), HTML files retain all document properties, including fonts and formatting.

Using Excel, you can edit any Excel document from within a Web browser. While you are viewing a page that was created in an Office application, such as Excel, click the Edit button on the browser's toolbar. Office opens the document in the application that was used to create it. You can then edit the Web page and resave it in any of the file formats that the application supports or in HTML.

Other File Formats

The following are two other types of file formats that you will rarely encounter; I haven't seen a DIF file in ages, and I've never seen a SYLK file.

- **Data Interchange Format (DIF):** Used by VisiCalc. Excel can read and write these files.
- **Symbolic Link (SYLK):** Used by MultiPlan. Excel can read and write these files.

Using the Clipboard to Get Data

Using the Windows Clipboard is a another method of importing data into your worksheet. The process involves selecting data from another application and copying the data to the Clipboard. Then, you reactivate Excel and paste the information to the worksheet. The exact results that you get can vary quite a bit, depending on the type of data that you copied and the Clipboard formats that it supports. Obviously, you must have a copy of the other application installed on your system.

If you copy information from another Office application, you use the Office Clipboard, not the Windows Clipboard. The Office Clipboard supports copying and pasting of all formats used in all Office applications.

About the Clipboard

As you read in Chapter 8, Office 2000 provides Windows 95 or Windows 98 with two clipboards. The original Windows Clipboard remains; whenever you cut or copy information from a Windows program, Windows stores the information on the Windows Clipboard, which is an area of memory. Each time that you cut or copy information, Windows replaces the information previously stored on the Clipboard with the new information that you cut or copied. The Windows Clipboard can store data in a variety of formats. Because Windows manages it, information on the Windows Clipboard can be pasted to other Windows applications, regardless of where it originated. Normally, you can't see information stored on the Windows Clipboard (nor would you want to).

To view the Windows Clipboard contents, you can run the Clipboard Viewer program, which comes with Windows. The Clipboard Viewer may or may not be installed on your system (it is not installed by default). You can use the Clipboard Viewer to view only the last piece of information that you copied to the Office Clipboard.

When you copy or cut data to the Clipboard, the source application places one or more formats on the Clipboard along with the data. Different applications support different Clipboard formats. When you paste Clipboard data into another application, the destination application determines which format it can handle and typically selects the format that either provides the most information or is appropriate for where you are pasting it. If you view cells copied from Excel in the Clipboard Viewer, by default, you'll see a row/column reference rather than the actual information that you copied (see Figure 22-2). In some cases, you can use the Display command in the Clipboard Viewer application to view the Clipboard data in a different format. For example, you can display a range of cells from Excel as a picture, bitmap, text, OEM text, or a DIB bitmap. Figures 22-3, 22-4, 22-5, and 22-6, show examples of the same Excel range as it appears in the Clipboard Viewer when using different Display formats.

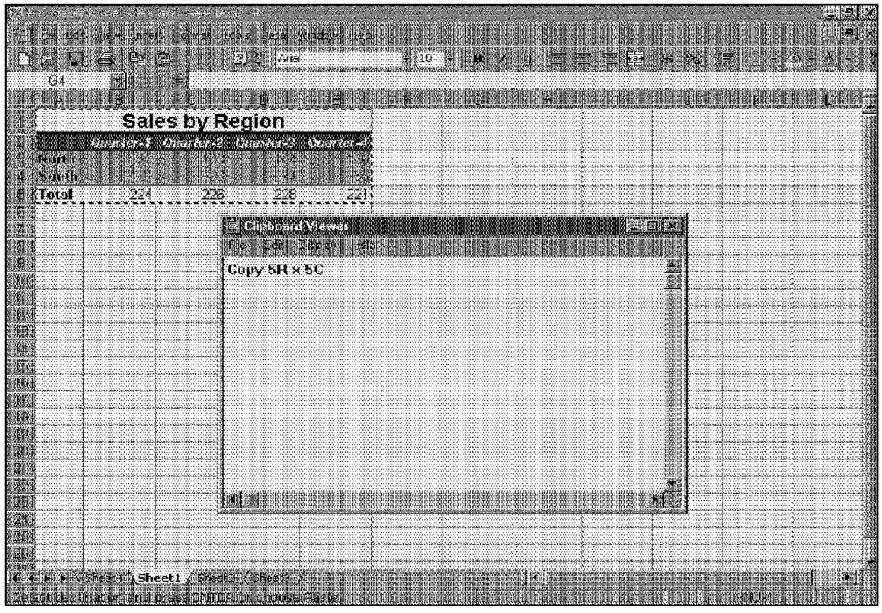


Figure 22-2: The Windows Clipboard Viewer application displaying Excel 2000 data in default format, Display Text.

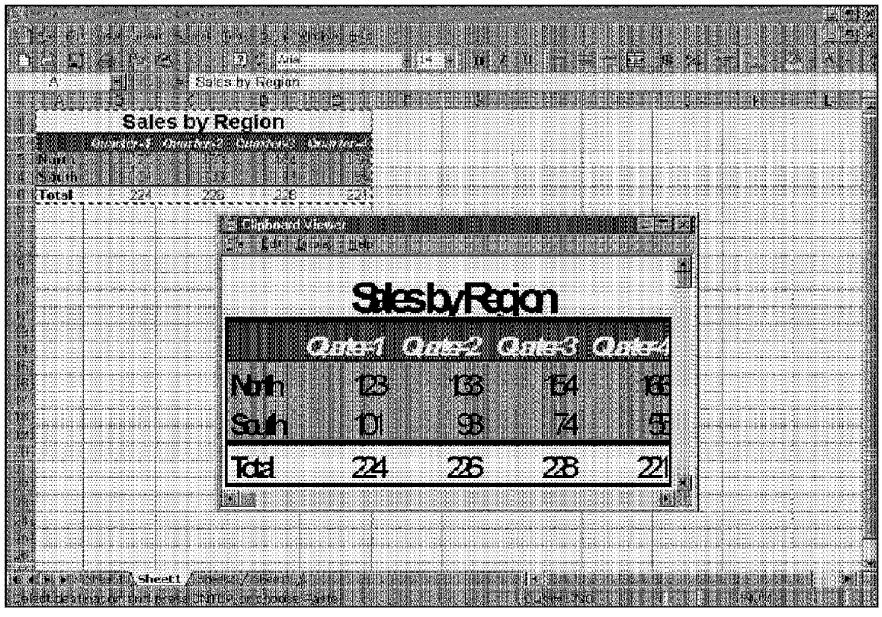


Figure 22-3: The same data in Picture format.

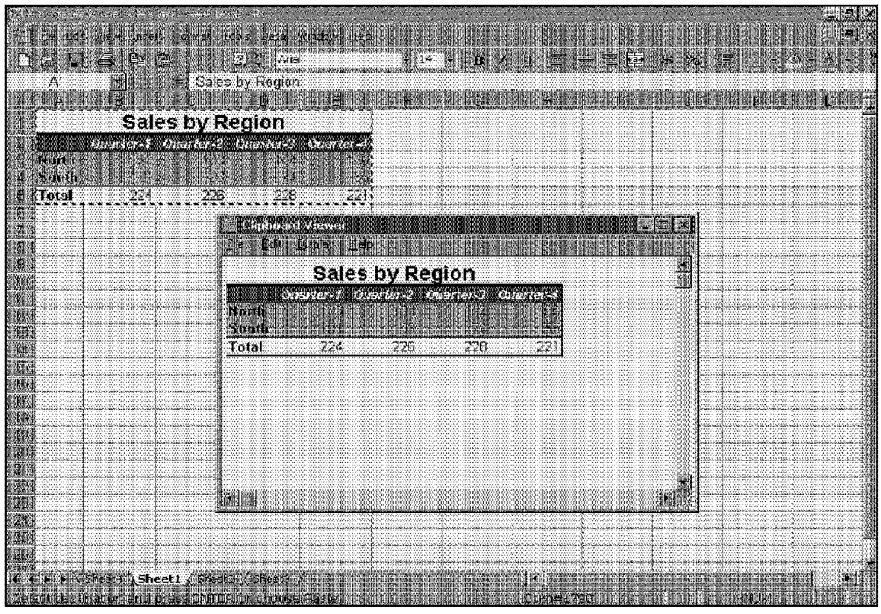


Figure 22-4: Both Bitmap and DIB Bitmap format closely resemble the formatted appearance of the data in Excel.

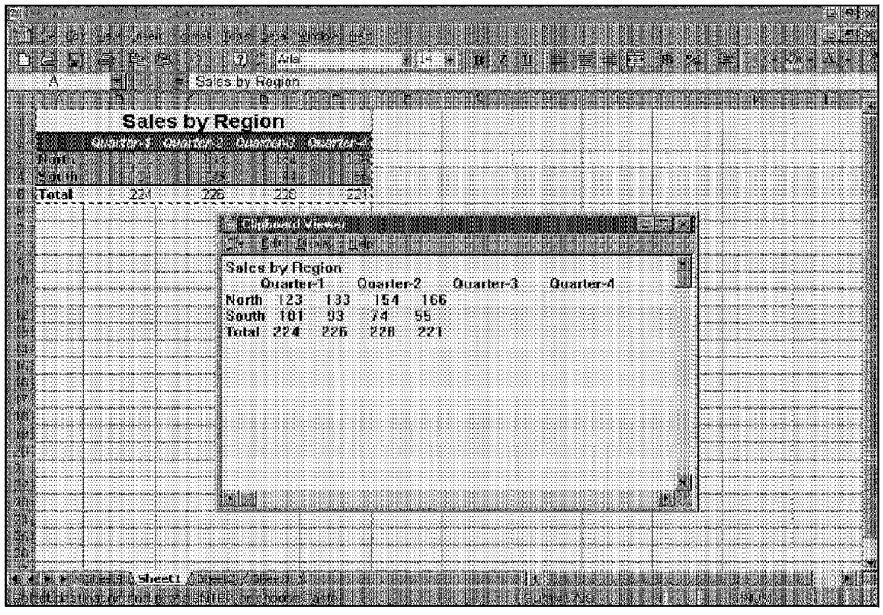


Figure 22-5: The Text format shows the text similar to the way that it appears in the Notepad.

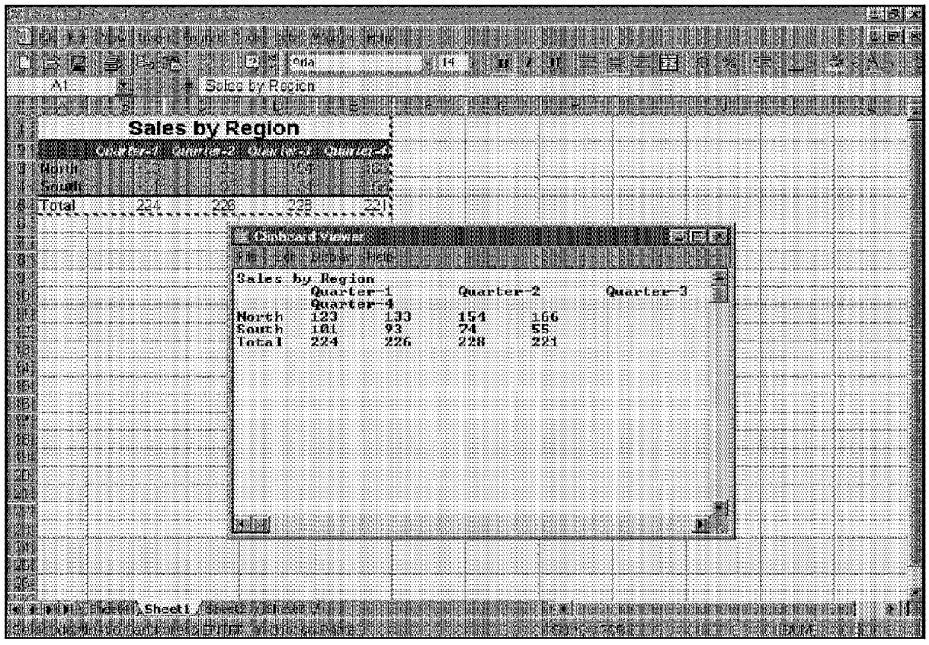


Figure 22-6: The OEM text format.

Importantly, the format that you select in the Clipboard Viewer *doesn't* affect how Excel copies the data. In some cases, however, you can use Excel's Edit • Paste Special command to select alternate methods of pasting the data.

Copying Data from Another Windows Application

Copying data from one Windows application to another is quite straightforward. The *source application* contains the data that you're copying, and the *destination application* receives the data that you're copying. Use the following steps to copy data from one application into another:

1. Activate the source document window that contains the information you want to copy.
2. Select the information that you want to copy by using the mouse or the keyboard. If Excel is the source application, this information can be a cell, range, chart, or drawing object.
3. Select Edit • Copy (or any available shortcut). A copy of the information is sent to the Windows Clipboard. If you're copying from an Office application, a copy of the information is also sent to the Office Clipboard.
4. Activate the destination application. If it isn't open, you can start it without affecting the contents of the Clipboard.

5. Move to the position to which you want to paste in the destination application.
6. Select Edit • Paste from the menu in the destination application. If the Clipboard contents aren't appropriate for pasting, the Paste command is grayed (not available).

In Step 3, you also can select Edit • Cut from the source application menu. This step erases the selection from the source application after it's placed on the Clipboard.

Many Windows applications use a common keyboard convention for the Clipboard commands. Generally, this technique is a bit faster than using the menus, because these keys are adjacent to each other. The shortcut keys and their equivalents are the following:

Ctrl+C Edit • Copy

Ctrl+X Edit • Cut

Ctrl+V Edit • Paste

You need to understand that Windows applications vary in how they respond to data that you paste from the Clipboard. If the Edit • Paste command isn't available (it is grayed on the menu) in the destination application, the application can't accept the information from the Clipboard. If you copy a table from Word for Windows to Excel, the data translates into cells perfectly—complete with formatting. Copying data from other applications may not work as well; for example, you may lose the formatting, or you may end up with all the data in a single column rather than in separate columns. As discussed later in this chapter, you can use the Convert Text to Columns Wizard to convert this data into columns.

If you plan to do a great deal of copying and pasting between two applications, experiment until you understand how the two applications can handle each other's data.

Copying Data from a Non-Windows Application

You also can use the Windows Clipboard with non-Windows applications running in a DOS window. As you may know, you can run non-Windows programs from Windows either in a window or in full-screen mode (the application takes over the complete screen).

When you're running a non-Windows application in Windows, you can press Alt+Print Screen to copy the entire screen to the Clipboard. The screen contents can then be pasted into a Windows application (including Excel). To copy only part of the screen, you must run the application in a window: press Alt+Enter to toggle between full-screen mode and windowed mode. You can then click the Control menu, choose Edit • Mark, and select text from the window. This window may or may not have a toolbar displayed. If it does not, follow these steps:

1. Right-click the title bar and select the Toolbar option.

2. Click the Mark tool and select the text to copy.
3. Click the Copy tool to copy the selected text to the Clipboard.
4. Activate Excel.
5. Select Edit • Paste to copy the Clipboard data into your worksheet.

Figure 22-7 shows Quattro Pro running in a DOS window. Some text is selected.

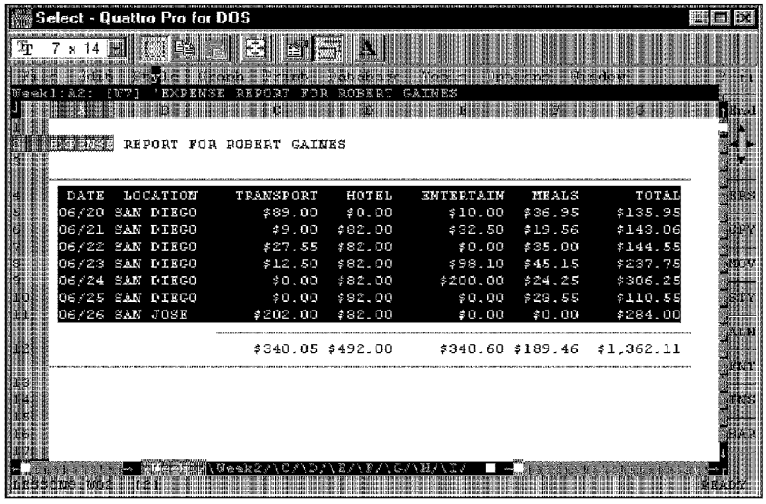


Figure 22-7: Copying data from Quattro Pro for DOS.

If you use this technique and copy to Excel, the information is pasted as text in a single column. In other words, even if you copy information from neatly formatted columns, it's all pasted into a single column in Excel. But don't fret—you can use Excel's Convert Text to Columns Wizard to convert this data into columns.

You're limited to copying one screen of information at a time—you can't scroll the DOS application while you're selecting text.

Importing Text Files

Text files (sometimes referred to as ASCII files) are usually considered to be the lowest-common-denominator file type. Such files contain only data, with no formatting. Consequently, most applications can read and write text files. So, if all else fails, you can probably use a text file to transfer data between two applications that don't support a common file format. Because text files are so commonly used, this entire section is devoted to discussing them and explaining how to use Excel's Text Import Wizard.

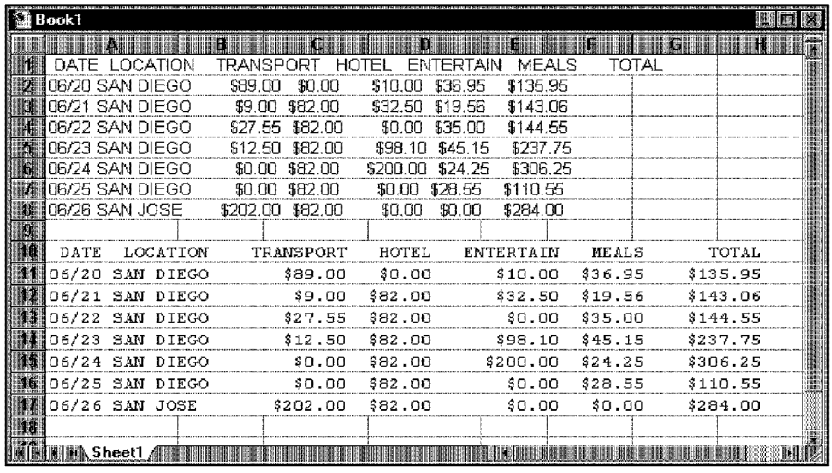
About Text Files

You may find it helpful to think of some text files in terms of a database table. Each line in the text file corresponds to a database record, and each record consists of a number of fields. In Excel, each line (or record) is imported to a separate row, and each field goes into a separate column. Text files come in two types: delimited and nondelimited.

Text files consist of plain text and end-of-line markers. *Delimited* text files use a special character to separate the fields on each line—typically a comma, a space, or a tab (but occasionally, you'll see other delimiters used). In addition, text is usually (but not always) enclosed in quotation marks.

Nondelimited files don't contain a special field-separator character. Often, however, each field is a fixed length, enabling you easily to break each line of text into separate columns. When you view a nondelimited file, the data often appears to be in columns.

If you use a proportional font, such as Arial or Times Roman, the fields of text file may appear to not line up, although they actually do. In proportional font sets, each character uses a different amount of horizontal space. For best results, use a nonproportional font, such as Courier New, when working with text files. Excel uses Courier New in its Text Import Wizard dialog box. Figure 22-8 shows the same text displayed in Arial and Courier New fonts.



Book1

	A	B	C	D	E	F	G	H
1	DATE	LOCATION	TRANSPORT	HOTEL	ENTERTAIN	MEALS	TOTAL	
2	06/20	SAN DIEGO	$89.00	$0.00	$10.00	$36.95	$135.95	
3	06/21	SAN DIEGO	$9.00	$82.00	$32.50	$19.56	$143.06	
4	06/22	SAN DIEGO	$27.55	$82.00	$0.00	$35.00	$144.55	
5	06/23	SAN DIEGO	$12.50	$82.00	$98.10	$45.15	$237.75	
6	06/24	SAN DIEGO	$0.00	$82.00	$200.00	$24.25	$306.25	
7	06/25	SAN DIEGO	$0.00	$82.00	$0.00	$28.55	$110.55	
8	06/26	SAN JOSE	$202.00	$82.00	$0.00	$0.00	$284.00	
9								
10	DATE	LOCATION	TRANSPORT	HOTEL	ENTERTAIN	MEALS	TOTAL	
11	06/20	SAN DIEGO	$89.00	$0.00	$10.00	$36.95	$135.95	
12	06/21	SAN DIEGO	$9.00	$82.00	$32.50	$19.56	$143.06	
13	06/22	SAN DIEGO	$27.55	$82.00	$0.00	$35.00	$144.55	
14	06/23	SAN DIEGO	$12.50	$82.00	$98.10	$45.15	$237.75	
15	06/24	SAN DIEGO	$0.00	$82.00	$200.00	$24.25	$306.25	
16	06/25	SAN DIEGO	$0.00	$82.00	$0.00	$28.55	$110.55	
17	06/26	SAN JOSE	$202.00	$82.00	$0.00	$0.00	$284.00	
18								

Sheet1

Figure 22-8: Using a proportional font may obscure columns in a text file.

Excel is quite versatile when importing text files. If each line of the text file is identically laid out, importing is usually problem-free. But if the line contains mixed information, you may need to do some additional work to make the data usable.

For example, you create text files in some programs by sending a printed report to a disk file rather than to the printer. These reports often have extra information, such as page headers and footers, titles, summary lines, and so on.

Using the Text Import Wizard

Prior versions of Excel treated importing text files differently from other types of database information. In Excel 2000, if you use the technique described in this section, you'll create a Text File Query, which you can refresh in the same way that you refresh Database and Web queries. This new feature will make easier the lives of those who need to regularly import text files, because they won't need to "set up" the import each time. When you want to update the Excel file that you create by importing a text file, choose Data • Refresh Data. Highlight the text file that you originally imported and click the Import button. Excel automatically updates the Excel version of the file with any new data that may appear in the text file.

For more information on Database queries, see Chapter 24. For more information on Web queries, see Chapter 30.

To import a text file into Excel, choose Data • Get External Data • Import Text File. In the Import Text File dialog box, navigate to the folder containing the file that you want to import. The dialog box then displays text files that have an extension of TXT. If the text file that you're importing has a different extension, select the All Files option. Or, you can enter the filename directly into the File name box, if you know the file's name.

Excel displays its Text Import Wizard, a series of interactive dialog boxes in which you specify the information that Excel requires to break the lines of the text file into columns. You can truly appreciate this time-saving feature if, in a previous life, you struggled with the old data-parsing commands that are found in other spreadsheet programs and older versions of Excel.

Text Import Wizard: Step 1 of 3

Figure 22-9 shows the first of three Text Import Wizard dialog boxes. In the Original data type section, verify the type of data file (Excel almost always guesses correctly). You also can indicate the row that Excel should use to start importing. For example, if the file has a title in the first row, you may want to skip the first line.

Notice that you can preview the file at the bottom of the dialog box, using the scrollbars to view more of the file. If the characters in the file don't look right, you may need to change the File Origin; this determines which character set to use (in many cases, it doesn't make any difference). After you finish with this step, click the Next button to move to Step 2.

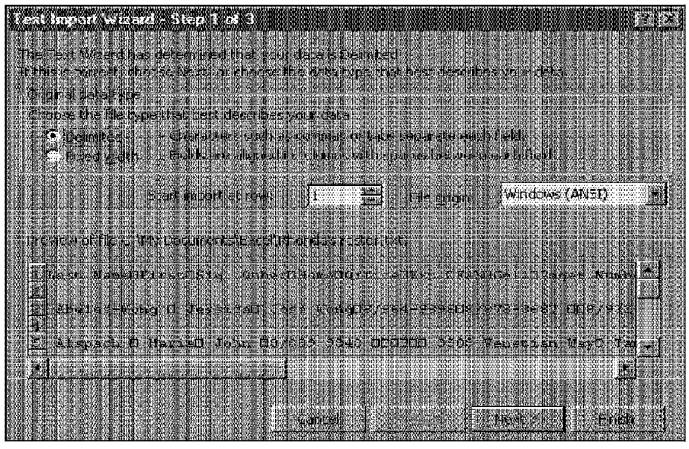


Figure 22-9: Step 1 of the Text Import Wizard.

Text Import Wizard: Step 2 of 3

The dialog box that you see for Step 2 of the Text Import Wizard varies, depending on your choice in the Original data type section in Step 1. If you selected Delimited, you see the dialog box shown in Figure 22-10. You can specify the type of delimiter, the text qualifier, and whether to treat consecutive delimiters as a single delimiter; choosing to treat consecutive delimiters as a single delimiter tells Excel to skip empty columns. The Data preview section displays vertical lines to indicate how Excel will break up the fields. The Data preview section changes as you make choices in the dialog box.

If you selected Fixed width, you see the dialog box shown in Figure 22-11. At this point, Excel attempts to identify the column breaks and displays vertical break lines to represent how it will break fields apart into columns. If Excel guesses wrong, you can move the lines, insert new ones, or delete lines that Excel proposes. You'll see instructions in the dialog box.

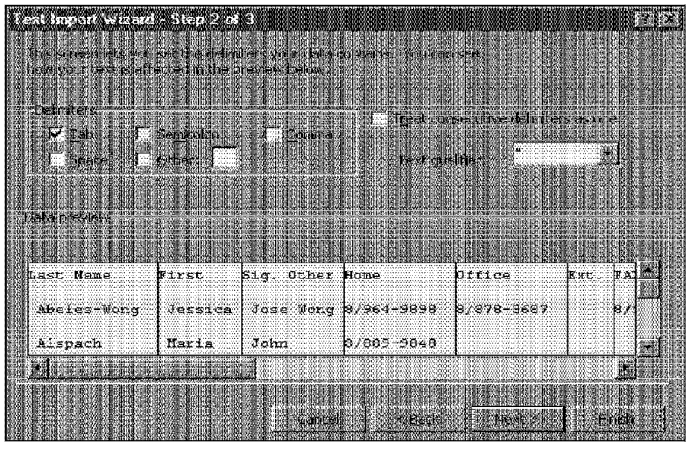


Figure 22-10: Step 2 of the Text Import Wizard (for delimited files).

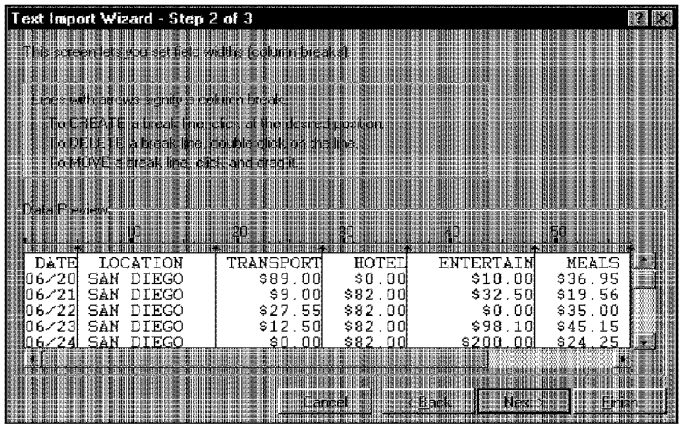


Figure 22-11: Step 2 of the Text Import Wizard (for fixed-width files).

If you're importing a print image file that includes page headers, you can ignore them when you specify the column indicators. Rather, base the columns on the data. When the file is imported, you can then delete the rows that contain the page headers.

When you're satisfied with how the column breaks look, click Next to move to the final step. Or, you can click Back to return to Step 1 and change the file type.

Text Import Wizard: Step 3 of 3

Figure 22-12 shows the last of the three Text Import Wizard dialog boxes. In this dialog box, you can select individual columns and specify the formatting to apply (General, Text, or Data). You also can specify columns to skip—they aren't imported. If you click the Advanced button, you'll see the dialog box shown in Figure 22-13, in which you can specify characters to use as decimal and thousands separators. When you're satisfied with the results, click Finish. Excel prompts you for the starting cell location for the imported data; when you click OK, Excel imports the data and displays the External Data toolbar, which helps you to work with the imported text file (see Figure 22-14). For example, if you click the Data Range Properties tool, you see the External Data Range Properties dialog box, shown in Figure 22-15, which you can use to change how Excel treats the imported file.

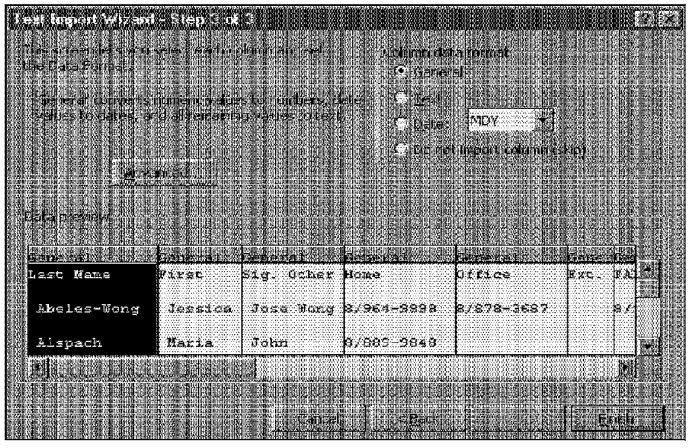


Figure 22-12: Step 3 of the Text Import Wizard.

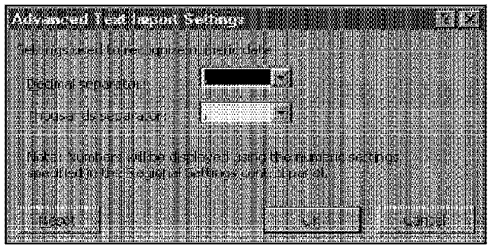


Figure 22-13: The Advanced Text Import Settings dialog box.

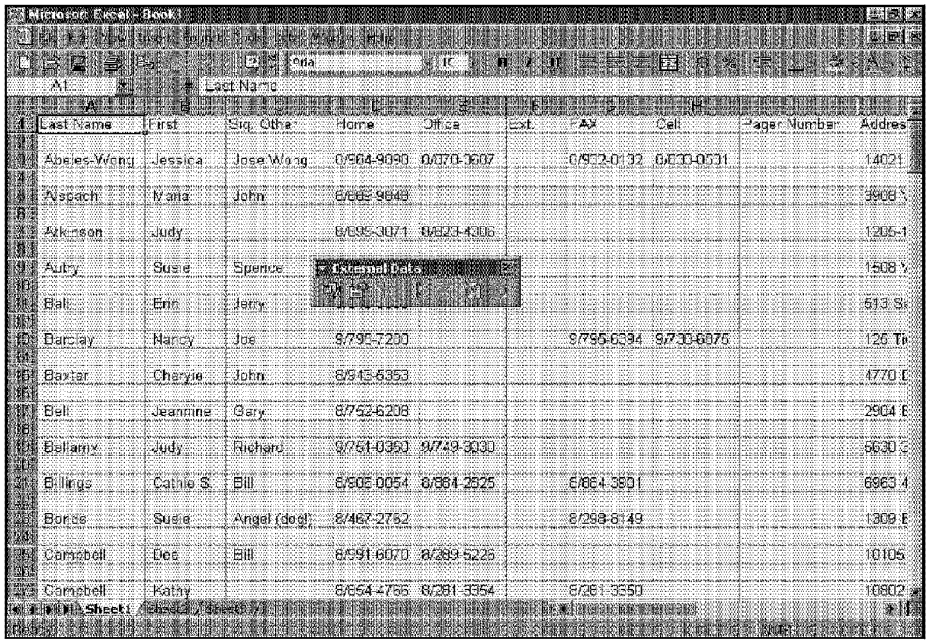


Figure 22-14: Imported text file.

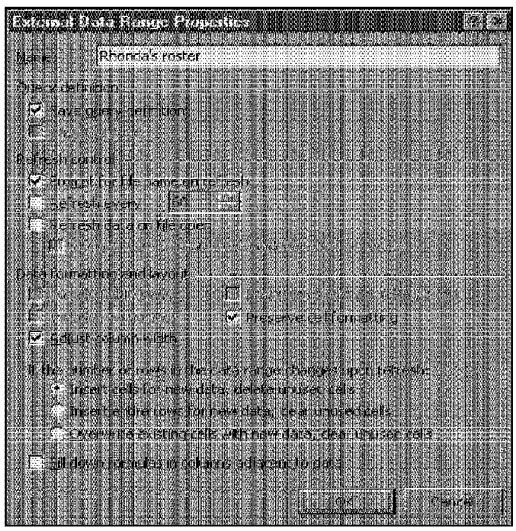


Figure 22-15: Use this dialog box to change the way that Excel treats the imported file.

If the results aren't what you expect, close the workbook and try again (text importing often involves trial and error). Don't forget that you can scroll the Data Preview window to make sure that all the data is converted properly. With some files, however, importing all the data properly is impossible. In such cases, you may want to import the file as a single column of text and then break lines into columns selectively. The procedure for doing this is discussed in the next section.

Using the Text to Columns Wizard

Excel can parse text that is stored in a column. Start by selecting the text (in a single column). Then, choose Data • Text to Columns, and Excel displays the first of three Text to Columns Wizard dialog boxes. These dialog boxes are identical to those used for the Text Import Wizard, except that the title bar text is different.

Unfortunately, you can't use the Data • Text to Columns command on a multiple selection; this would be quite handy for parsing imported files with several different layouts. Even worse, you can't use the Edit • Repeat command to repeat the Text to Columns command.

Summary

This chapter identifies the various sources for getting data into Excel: entering data manually, generating data from formulas or macros, using Query or pivot tables, copying data using the Clipboard, and importing foreign files (including text files) into Excel. The chapter focuses on Clipboard operations and file importing.

CHAPTER 23

Working with Lists

In This Chapter

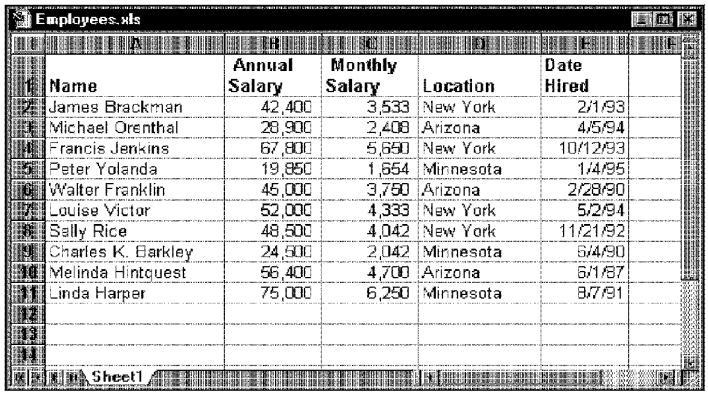
Research conducted by Microsoft indicates that Excel is frequently used to manage lists, or *worksheet databases*. This chapter covers list management and demonstrates useful techniques that involve lists.

What Is a List?

A list is essentially an organized collection of information. More specifically, a list consists of a row of headers (descriptive text), followed by additional rows of data, which can be values or text. You may recognize this as a database table — which is exactly what it is. Beginning with Excel 5, Microsoft uses the term *list* to refer to a database stored in a worksheet and the term *database* to refer to a table of information stored in an external file. To avoid confusion, I adhere to Microsoft's terminology.

I cover external database files in Chapter 24.

Figure 23-1 shows an example of a list in a worksheet. This particular list has its headers in row 1 and has 10 rows of data. The list occupies four columns. Notice that the data consists of several different types: text, values, and dates. Column C contains a formula that calculates the monthly salary from the value in column B.



Employees.xls

	A	B	C	D	E	F
1	Name	Annual Salary	Monthly Salary	Location	Date Hired	
2	James Brackman	42,400	3,533	New York	2/1/93	
3	Michael Orenthal	28,900	2,408	Arizona	4/5/94	
4	Francis Jenkins	67,800	5,650	New York	10/12/93	
5	Peter Yolanda	19,850	1,654	Minnesota	1/4/95	
6	Walter Franklin	45,000	3,750	Arizona	2/28/90	
7	Louise Victor	52,000	4,333	New York	5/2/94	
8	Sally Rice	48,500	4,042	New York	11/21/92	
9	Charles K. Barkley	24,500	2,042	Minnesota	6/4/90	
10	Melinda Hintquest	56,400	4,700	Arizona	6/1/87	
11	Linda Harper	75,000	6,250	Minnesota	8/7/91	
12						
13						
14						

Sheet1

Figure 23-1: An example of a list.

People often refer to the columns in a list as *fields* and to the rows as *records*. Using this terminology, the list shown in the figure has five fields (Name, Annual Salary, Monthly Salary, Location, and Date Hired) and ten records.

The size of the lists that you develop in Excel is limited by the size of a single worksheet. In other words, a list can have no more than 256 fields and can consist of no more than 65,535 records (one row contains the field names). A list of this size would require a great deal of memory and even then may not be possible. At the other extreme, a list can consist of a single cell—not very useful, but it's still considered a list.

In versions of Excel prior to Excel 97, a list was limited to 16,383 records.

What Can You Do with a List?

Excel provides several tools to help you manage and manipulate lists. Consequently, people use lists for a wide variety of purposes. For some users, a list is simply a method to keep track of information (for example, customer lists); others use lists to store data that ultimately will appear in a report. Common list operations include:

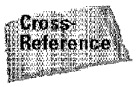
- Entering data into the list
- Filtering the list to display only the rows that meet certain criteria
- Sorting the list
- Inserting formulas to calculate subtotals
- Creating formulas to calculate results on the list filtered by certain criteria
- Creating a summary table of the data in the list (this is done using a pivot table; see Chapter 25).

With the exception of the last item, these operations are covered in this chapter.

Designing a List

Although Excel is quite accommodating when it comes to the information that is stored in a list, planning the organization of your list information will pay off. The following are some guidelines to keep in mind when creating lists:

- Insert descriptive labels (one for each column) in the first row of the list, called the *header row*. If you use lengthy labels, consider using the Wrap Text format so that you don't have to widen the columns.



See Chapter 11 for information on the Wrap Text format.

- Make sure each column contains the same type of information. For example, don't mix dates and text in a single column.
- You can use formulas that perform calculations on other fields in the same record. If you use formulas that refer to cells outside the list, make these absolute references; otherwise, you get unexpected results when you sort the list.
- Don't leave any empty rows within the list. For list operations, Excel determines the list boundaries automatically, and an empty row signals the end of the list.
- For best results, try to keep the list on a worksheet by itself. If you must place other information on the same worksheet as the list, place the information above or below the list. In other words, don't use the cells to the left or the right of a list.
- Select Window • Freeze Panes to make sure that you can see the headings when you scroll the list.
- You can preformat entire columns to ensure that the data has the same format. For example, if a column contains dates, format the entire column with the desired date format.

Many people find working in spreadsheets most appealing because changing the layout is relatively easy. Lists behave no differently than any other kind of data in Excel; changing a list's layout is also easy. For example, you may create a list and then decide that it needs another column (field). No problem. Just insert a new column, give it a field name, and Excel expands your list. If you've ever used a database management program, you can appreciate the simplicity of this layout change.

Entering Data into a List

You can enter data into a list in three ways:

- Manually, using all standard data entry techniques
- By importing it or copying it from another file
- By using a dialog box

There's really nothing special about entering data into a list. You just navigate through the worksheet and enter the data into the appropriate cells.

Excel has two features that assist with repetitive data entry:

- **AutoComplete.** When you begin to type in a cell, Excel scans up and down the column for entries that match what you're typing. If it finds a match, Excel fills in the rest of the text automatically. Press Enter to make the entry. You can turn this feature on or off in the Edit tab of the Options dialog box.
- **Pick Lists.** You can right-click on a cell and select Pick from list from the shortcut menu. Excel displays a list box that shows all entries in the column (see Figure 23-2). Click on the one that you want to enter into the cell (no typing is required).

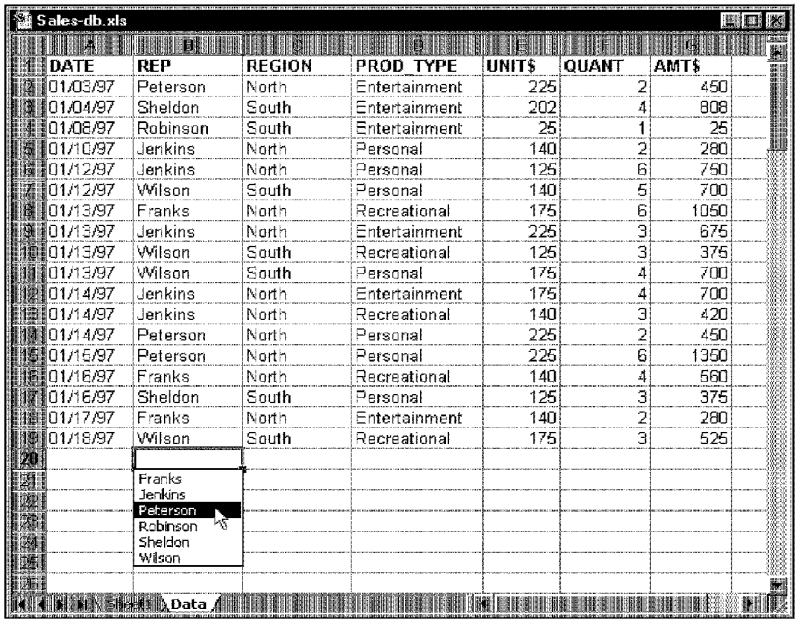


Figure 23-2: Choosing the Pick from list command on the shortcut menu gives you a list of all items in the current column.

If you prefer to use a dialog box for your data entry, Excel accommodates you. To display a data entry dialog box, move the cell pointer anywhere within the list and choose Data • Form. Excel determines the boundaries of your list and displays a dialog box showing each field in the list. Figure 23-3 depicts an example of such a dialog box. Fields that have a formula don't have an edit box.

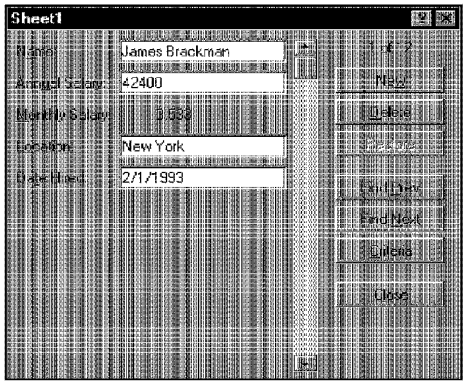


Figure 23-3: The Data • Form command gives you a handy data entry dialog box.

If the number of fields exceeds the limit of your display, the dialog box contains two columns of field names. If your list consists of more than 32 fields, however, the Data • Form command doesn't work. You must forego this method of data entry and enter the information directly into the cells.

Entering Data with the Data Form Dialog Box

When the data form dialog box appears, Excel displays the first record in the list. Notice the indicator in the upper-right corner of the dialog box that tells you the number of the selected record and the total number of records in the list.

To enter a new record, click on the New button to clear the fields. Then you can enter the new information into the appropriate fields. Use Tab or Shift+Tab to move among the fields. When you click on New (or Close), Excel appends the data that you entered to the bottom of the list. You also can press Enter, which is equivalent to clicking on the New button. If the list contains any formulas, Excel enters them for you automatically into the new record.

If you named the range of your list Database, Excel automatically extends the range definition to include the new row(s) that you add to the list using the data form dialog box. Note that this works only if you name the list Database; any other name doesn't work.

Other Uses for the Data Form Dialog Box

You can use the data form dialog box for more than just data entry. You can edit existing data in the list, view data one record at a time, delete records, and display records that meet certain criteria.

The dialog box contains a number of additional buttons, which are described as follows:

- **Delete:** Deletes the displayed record.
- **Restore:** Restores any information that you edited. You must click on this button before you click on the New button.
- **Find Prev:** Displays the previous record in the list. If you entered a criterion, this button displays the previous record that matches the criterion.
- **Find Next:** Displays the next record in the list. If you entered a criterion, this button displays the next record that matches the criterion.
- **Criteria:** Clears the fields and lets you enter a criterion upon which to search for records. For example, to locate records that have a salary greater than $50,000, enter **>50000** into the Salary field. Then you can use the Find Next and Find Prev buttons to display the qualifying records.
- **Close:** Closes the dialog box (and enters the data that you were entering, if any).

Using Microsoft Access Forms for Data Entry

If you have Microsoft Access installed on your system, you can use its form creation tools to develop a data entry form for an Excel worksheet. This feature uses the Access Links add-in, which must be loaded. When the add-in is loaded, you have a new command: Data • Access Form.

Choosing this command starts Access (if it's not already running) and begins its Form Wizard. Use the Form Wizard to create the data entry form. You can then use this form to add data to your Excel worksheet. Access's Form Wizard places a button on your worksheet that contains the text View Access Form. Click on this button to use the form. Figure 23-4 shows an Access form being used to enter data into an Excel worksheet.

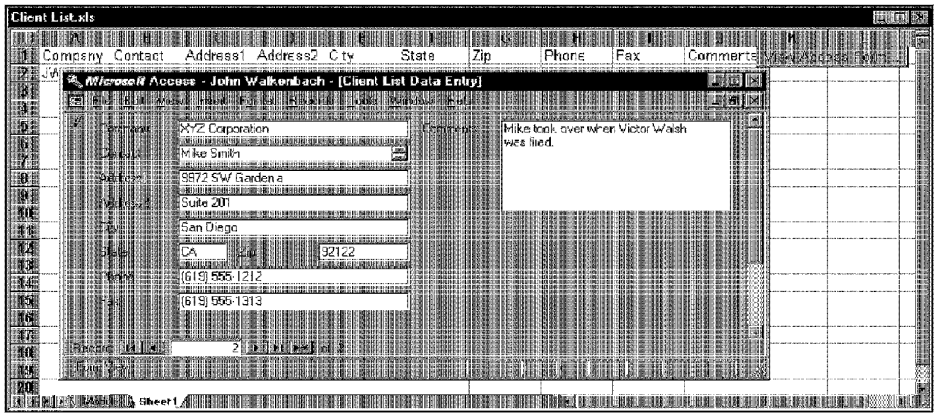


Figure 23-4: This form, developed in Microsoft Access, is being used to enter data into an Excel worksheet.

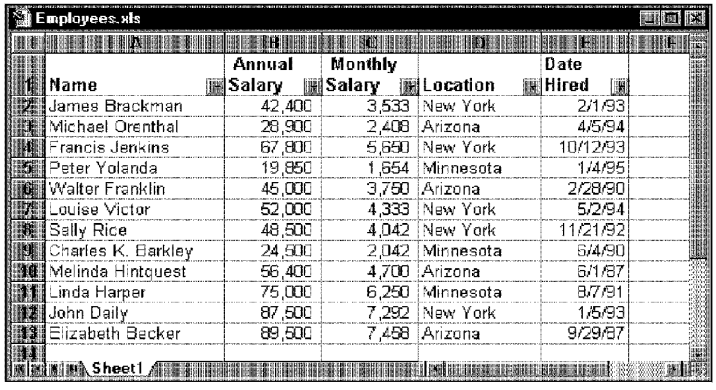
Filtering a List

Filtering a list is the process of hiding all rows in the list except those that meet some criteria that you specify. For example, if you have a list of customers, you can filter the list to show only those who live in New Jersey. Filtering is a common (and very useful) technique. Excel provides two ways to filter a list:

- AutoFilter, for simple filtering criteria
- Advance Filter, for more complex filtering

AutoFiltering

To use Excel's AutoFilter feature to filter a list, place the cell pointer anywhere within the list and then choose Data • Filter • AutoFilter. Excel analyzes your list and adds drop-down arrows to the field names in the header row, as shown in Figure 23-5.



Employees.xls

	A	B	C	D	E	F
1	Name	Annual Salary	Monthly Salary	Location	Date Hired	
2	James Brackman	42,400	3,533	New York	2/1/93	
3	Michael Orenthal	28,900	2,408	Arizona	4/5/94	
4	Francis Jenkins	67,800	5,650	New York	10/12/93	
5	Peter Yolanda	19,850	1,654	Minnesota	1/4/95	
6	Walter Franklin	45,000	3,750	Arizona	2/28/90	
7	Louise Victor	52,000	4,333	New York	5/2/94	
8	Sally Rice	48,500	4,042	New York	11/21/92	
9	Charles K. Barkley	24,500	2,042	Minnesota	6/4/90	
10	Melinda Hintquest	56,400	4,700	Arizona	6/1/87	
11	Linda Harper	75,000	6,250	Minnesota	8/7/91	
12	John Daily	87,500	7,292	New York	1/5/93	
13	Elizabeth Becker	89,500	7,458	Arizona	9/29/87	
14						

Sheet1

Figure 23-5: When you choose the Data • Filter • AutoFilter command, Excel adds drop-down arrows to the field names in the header row.

When you click on the arrow in one of these drop-down lists, the list expands to show the unique items in that column. Select an item, and Excel hides all rows except those that include the selected item. In other words, Excel filters the list by the item that you selected.

After you filter the list, the status bar displays a message that tells you how many rows qualified. In addition, the drop-down arrow changes color to remind you that you filtered the list by a value in that column.

AutoFiltering has a limit. Only the first 999 unique items in the column appear in the drop-down list. If your list exceeds this limit, you can use advanced filtering, which is described later.

Besides showing every item in the column, the drop-down list includes five other items:

- **All:** Displays all items in the column. Use this to remove filtering for a column.
- **Top 10:** Filters to display the "top 10" items in the list; this is discussed later.
- **Custom:** Lets you filter the list by multiple items; this is discussed later.
- **Blanks:** Filters the list by showing rows that contain blanks in this column.
- **NonBlanks:** Filters the list by showing rows that contain non-blanks in this column.

To display the entire list again, click on the arrow and choose All—the first item in the drop-down list. Or, you can select Data • Filter • Show All.

To move out of Autofilter mode and remove the drop-down arrows from the field names, choose Data • Filter • AutoFilter again to remove the check mark from the AutoFilter menu item and restore the list to its normal state.

If you have any formulas that refer to data in a filtered list, be aware that the formulas don't adjust to use only the visible cells. For example, if a cell contains a formula that sums values in column C, the formula continues to show the sum for *all* the values in column C—not just those in the visible rows. To solve this problem, use database functions, which I describe later in this chapter.

Multicolumn AutoFiltering

Sometimes you may need to filter a list by values in more than one column. Figure 23-6 shows a list comprised of several fields.

sales.xls

	A	B	C	D	E
1	Month	State	Product	Price	From Ad
2	Jan	CA	Printer	208	Yes
3	Jan	CA	Printer	203	No
4	Jan	IL	Printer	468	No
5	Jan	IL	Printer	226	No
6	Jan	NY	Printer	484	Yes
7	Jan	NY	Printer	373	Yes
8	Jan	CA	Modem	249	Yes
9	Jan	CA	Modem	329	No
10	Jan	IL	Modem	760	Yes
11	Jan	IL	Modem	959	No
12	Jan	NY	Modem	419	No
13	Jan	NY	Modem	555	No
14	Jan	CA	HardDrive	287	Yes
15	Jan	CA	HardDrive	758	No
16	Jan	IL	HardDrive	651	Yes
17	Jan	IL	HardDrive	233	No
18	Jan	NY	HardDrive	332	Yes
19	Jan	NY	HardDrive	652	Yes
20	Jan	CA	Mouse	748	No
21	Jan	CA	Mouse	811	No

Product Sales

Figure 23-6: The list before filtering by multiple columns.

Assume that you want to see the records that show modems sold in February. In other words, you want to filter out all records except those in which the Month field is *Feb* and the Product field is *Modem*.

First, get into Autofilter mode. Then click on the drop-down arrow in the Month field and select *Feb* to filter the list to show only records with *Feb* in the Month field. Then click on the drop-down arrow in the Product field and select *Modem*, filtering the filtered list to show only records that contain *Modem* in the Product column — resulting in a list filtered by values in two columns. Figure 23-7 shows the result.

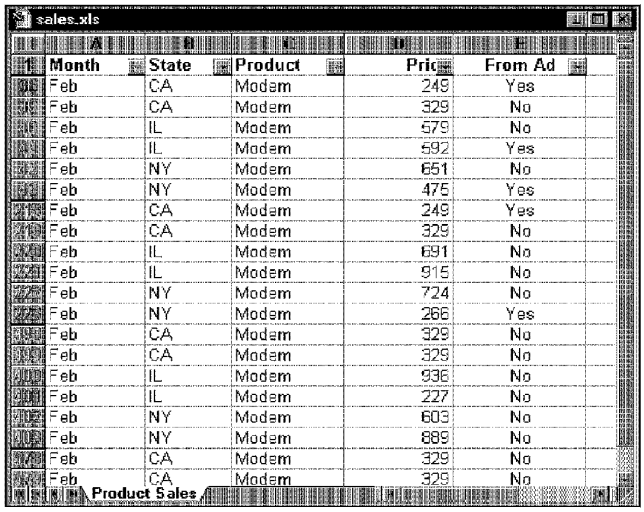


Figure 23-7: The same list filtered by values in two columns.

You can filter a list by any number of columns. Excel applies a different color to the drop-down arrows in the columns that have a filter applied.

Custom AutoFiltering

Usually, AutoFiltering involves selecting a single value for one or more columns. If you choose the Custom option in a drop-down list, you gain a bit more flexibility in filtering the list; Excel displays a dialog box like the one shown in Figure 23-8.

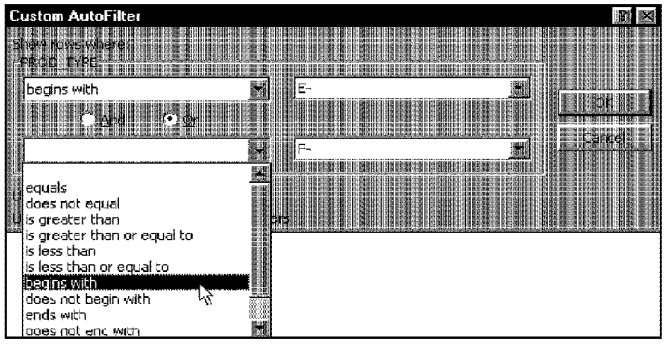


Figure 23-8: The Custom AutoFilter dialog box gives you more filtering options.

The Custom AutoFilter dialog box lets you filter in several ways:

- **Values above or below a specified value.** For example, sales amounts greater than 10,000.
- **Values within a range.** For example, sales amounts greater than 10,000 AND sales amounts less than 50,000.
- **Two discrete values.** For example, state equal to *New York* OR state equal to *New Jersey*.
- **Approximate matches.** You can use the * and ? wildcards to filter in a number of other ways. For example, to display only those customers whose last name begins with *B*, use **B***.

Custom AutoFiltering can be useful, but it definitely has limitations. For example, if you want to filter the list to show only three values in a field (such as New York or New Jersey or Connecticut), you can't do it by AutoFiltering. Such filtering tasks require the advanced filtering feature, which I discuss later in this chapter.

Top 10 AutoFiltering

Sometimes you may want to use a filter on numerical fields to show only the highest or lowest values in the list. For example, if you have a list of employees, you may want to identify the 12 employees with the longest tenure. You could use the custom AutoFilter option, but then you must supply a cutoff date (which you may not know). The solution is to use Top 10 AutoFiltering.

Top 10 AutoFiltering is a generic term; it doesn't limit you to the top *10* items. In fact, it doesn't even limit you to the *top* items. When you choose the Top 10 option from a drop-down list, you see dialog box that is shown in Figure 23-9.

Figure 23-9: The Top 10 AutoFilter gives you more AutoFilter options.

You can choose either Top or Bottom and specify any number. Suppose, for example, that you want to see the 12 employees with the longest tenure. Choose Bottom and 12 to filter the list and show the 12 rows with the smallest values in the Date Hired field. You also can choose Percent or Value in this dialog box. For example, you can filter the list to show the Bottom 5 percent of the records.

Charting filtered list data

You can create some interesting multipurpose charts that use data in a filtered list. The technique is useful because only the visible data appears in the chart. When you change the AutoFilter criteria, the chart updates itself to show only the visible cells.

For this technique to work, select the chart and make sure that the Plot Visible Cells Only option is enabled on the Chart tab of the Options dialog box.

Figure 23-10 shows an example of a chart created with an unfiltered list. It shows sales data for three months for each of four sales regions.

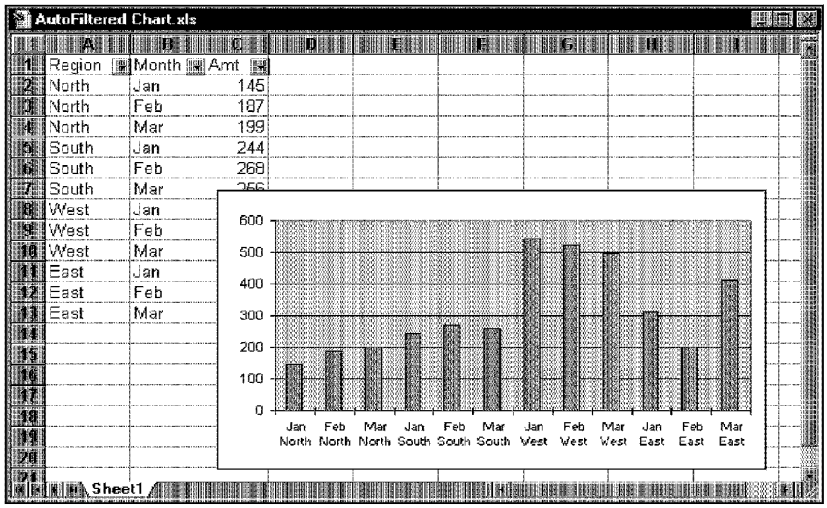


Figure 23-10: This chart was created from an unfiltered list.

Figure 23-11 shows the same chart, but the list was filtered to show only the North sales region. You can apply other filters, and the chart updates automatically. This technique lets a single chart show several different views of the data.

Advanced Filtering

In many cases, AutoFiltering does the job. But if you run up against its limitations, you need to use advanced filtering. Advanced filtering is much more flexible than AutoFiltering, but it takes a bit of up-front work to use it. Advanced filtering provides you with the following capabilities:

- You can specify more complex filtering criteria.
- You can specify computed filtering criteria.
- You can extract a copy of the rows that meet the criteria to another location.

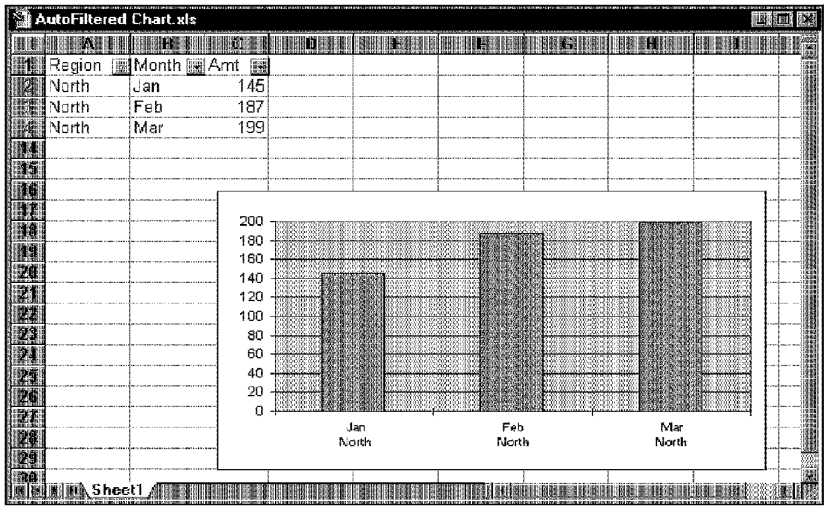


Figure 23-11: The chart from the previous figure, after filtering the list.

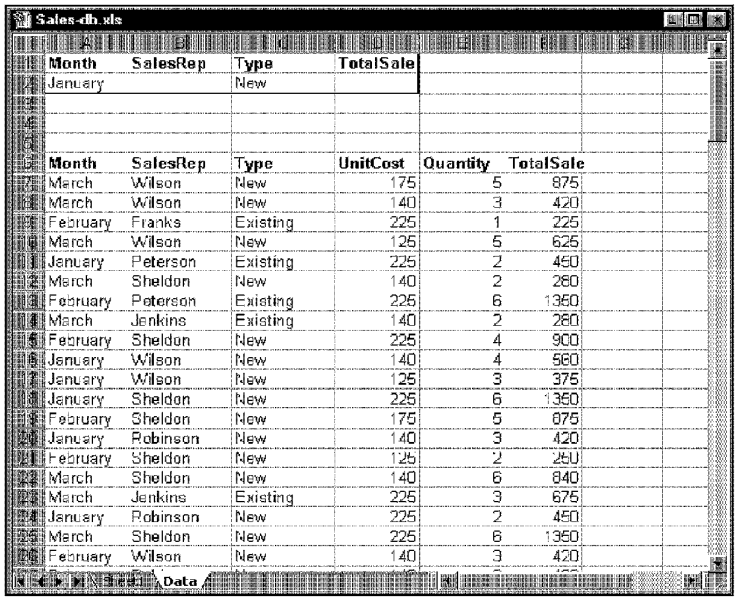
Setting up a criteria range

Before you can use the advanced filtering feature, you must set up a *criteria range*, a designated range on a worksheet that conforms to certain requirements. The criteria range holds the information that Excel uses to filter the list. It must conform to the following specifications:

- It must consist of at least two rows, and the first row must contain some or all field names from the list.
- The other rows of the criteria range must consist of your filtering criteria.

Although you can put the criteria range anywhere in the worksheet, it's a good idea not to put it in rows where you placed the list. Because Excel hides some of these rows when filtering the list, you may find that your criteria range is no longer visible after filtering. Therefore, you should generally place the criteria range above or below the list.

Figure 23-12 shows a criteria range, located in A1:D2, above the list that it uses. Only some field names appear in the criteria range. You don't need to include, in the criteria range, field names for fields that you don't use in the selection criteria.



Sales-db.xls

	A	B	C	D	E	F
1	Month	SalesRep	Type	TotalSale		
2	January		New			
3						
4						
5						
6	Month	SalesRep	Type	UnitCost	Quantity	TotalSale
7	March	Wilson	New	175	5	875
8	March	Wilson	New	140	3	420
9	February	Franks	Existing	225	1	225
10	March	Wilson	New	125	5	625
11	January	Peterson	Existing	225	2	450
12	March	Sheldon	New	140	2	280
13	February	Peterson	Existing	225	6	1350
14	March	Jenkins	Existing	140	2	280
15	February	Sheldon	New	225	4	900
16	January	Wilson	New	140	4	560
17	January	Wilson	New	125	3	375
18	January	Sheldon	New	225	6	1350
19	February	Sheldon	New	175	5	875
20	January	Robinson	New	140	3	420
21	February	Sheldon	New	125	2	250
22	March	Sheldon	New	140	6	840
23	March	Jenkins	Existing	225	3	675
24	January	Robinson	New	225	2	450
25	March	Sheldon	New	225	6	1350
26	February	Wilson	New	140	3	420

Sheet1 | Data

Figure 23-12: A criteria range for a list.

In this example, the criteria range has only one row of criteria. The fields in each row of the criteria range (except for the header row) are joined with an AND operator. Therefore, the filtered list shows rows in which the Month column equals *January* AND the Type column equals *New*. In other words, the list displays only sales to new customers made in January.

To perform the filtering, choose Data • Filter • Advanced filter. Excel displays the dialog box that is shown in Figure 23-13. Specify the list range and the criteria range, and make sure that you select the option labeled Filter the List in-place. Click on OK, and Excel filters the list by the criteria that you specified.

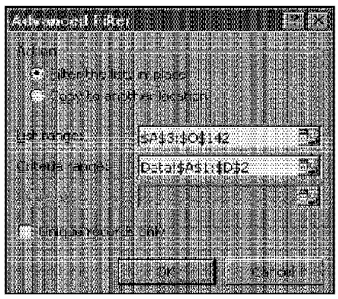


Figure 23-13: The Advanced Filter dialog box.

Multiple criteria

If you use more than one row in the criteria range, the criteria in each row are joined with an OR operator. A criteria range can have any number of rows, each of which is joined to the others with an OR operator. Figure 23-14 shows a criteria range (A1:D3) with two rows of criteria.

Sales-db.xls

	Month	SalesRep	Type	TotalSale
2	January		New	
3	February			>1000
4				
5				

	Month	SalesRep	Type	UnitCost	Quantity	TotalSale
6	March	Wilson	New	175	5	875
7	March	Wilson	New	140	3	420
8	February	Franks	Existing	225	1	225
9	March	Wilson	New	125	5	625
10	January	Peterson	Existing	225	2	450
11	March	Sheldon	New	140	2	280
12	February	Peterson	Existing	225	6	1350
13	March	Jenkins	Existing	140	2	280
14	February	Sheldon	New	225	4	900
15	January	Wilson	New	140	4	560
16	January	Wilson	New	125	3	375
17	January	Sheldon	New	225	6	1350
18	February	Sheldon	New	175	5	875
19	January	Robinson	New	140	3	420
20	February	Sheldon	New	125	2	250
21	March	Sheldon	New	140	6	840
22	March	Jenkins	Existing	225	3	675
23	January	Robinson	New	225	2	450
24	March	Sheldon	New	225	6	1350
25	February	Wilson	New	140	3	420

Sheet1 Data

Figure 23-14: This criteria range has two sets of criteria.

In this example, the filtered list shows rows in either of the following:

- *The Month field is January AND the Type field is New.*
- The Month field is *February* AND the Total Sale field is greater than 1000.

You cannot filter this way with AutoFiltering.

Types of criteria

The entries that you make in the criteria range can be either of the following:

- Text or value criteria. The filtering involves comparisons to a value or string, using operators such as equal (=), greater than (>), not equal to (<>), and so on.
- **Computed criteria.** The filtering involves a computation of some sort.

Text or value criteria

Table 23-1 lists the comparison operators that you can use with text or value criteria.

Table 23-1
Comparison Operators

Operator	*Comparison Type*
=	Equal to
>	Greater than
>=	Greater than or equal to
<	Less than
<=	Less than or equal to
<>	Not equal to

Table 23-2 shows examples of criteria that use strings.

Table 23-2
Examples of String Criteria

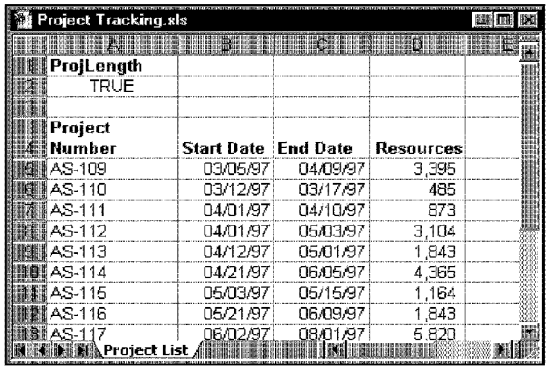
Criteria	*Effect*
>K	Text that begins with *L* through *Z*
<>C	All text, except text that begins with C
="January"	Text that matches January
Sm*	Text that begins with Sm
s*s	Text that begins with s and ends with s
s?s	Three-letter text that begins with *s* and ends with *s*

The text comparisons are not case sensitive. For example, si* matches *Simpson* as well as *sick*.

Computed criteria

Using computed criteria can make filtering even more powerful. Computed criteria filter the list based one or more calculations. Figure 23-15 shows a simple list that consists of project numbers, start dates, end dates, and resources. Above the list, in range A1:A2, is the criteria range. Notice, however, that this criteria range does

not use a field header from the list — it uses a new field header. A computed criteria essentially computes a new field for the list. Therefore, you must supply new field names in the first row of the criteria range.



Project Tracking.xls

	A	B	C	D	E
1	ProjLength				
2	TRUE				
3					
4	Project Number	Start Date	End Date	Resources	
5	AS-109	03/05/97	04/09/97	3,395	
6	AS-110	03/12/97	03/17/97	485	
7	AS-111	04/01/97	04/10/97	873	
8	AS-112	04/01/97	05/03/97	3,104	
9	AS-113	04/12/97	05/01/97	1,843	
10	AS-114	04/21/97	06/05/97	4,365	
11	AS-115	05/03/97	05/15/97	1,164	
12	AS-116	05/21/97	06/09/97	1,843	
13	AS-117	06/02/97	08/01/97	5,820	

Project List

Figure 23-15: This list is to be filtered using computed criteria.

Cell A2 contains the following formula:

```
=C5-B5+1>=30
```

This formula returns a logical value of either *True* or *False*. The result of the formula refers to cells in the first row of data in the list; it does *not* refer to the header row. When you filter the list by this criterion, the list shows only rows in which the project length (End Date–Start Date+1) is greater than or equal to 30 days. In other words, Excel bases the comparison on a computation.

You could accomplish the same effect, without using a computed criterion, by adding a new column to the list that contains a formula to calculate the project length. Using a computed criterion, however, eliminates the need to add a new column.

To filter the list to show only the projects that use above average resources, you could use the following computed criteria formula:

```
=D5>AVERAGE(D:D)
```

This filters the list to show only the rows in which the value of the Resources field is greater than the average of the Resources field.

Keep in mind the following items when using computed criteria:

- Don't use a field name in the criteria range that appears in the list. Create a new field name or just leave the cell blank.

- You can use any number of computed criteria and mix and match them with noncomputed criteria.
- Don't pay attention to the values returned by formulas in the criteria range. These refer to the first row of the list.
- If your computed formula refers to a value outside the list, use an absolute reference rather than a relative reference. For example, use $C$1 rather than C1.
- Create your computed criteria formulas using the first row of data in the list (not the field names). Make these references relative, not absolute. For example, use C5 rather than $C$5.

Other advanced filtering operations

The Advanced Filter dialog box gives you two other options:

- Copy to Another Location
- Unique Records Only

Both of these advanced filtering options are discussed below.

Copying qualifying rows

If you choose the Copy to Another Location option in the Advanced Filter dialog box, Excel copies the qualifying rows to another location in the worksheet or a different worksheet. You specify the location for the copied rows in the Copy to edit box. Note that the list itself is not filtered when you use this option.

Displaying only unique rows

Choosing the option labeled Unique records only hides all duplicate rows that meet the criteria that you specify. If you don't specify a criteria range, this option hides all duplicate rows in the list.

Using Database Functions with Lists

It's important to understand that Excel's worksheet functions don't ignore hidden cells. Therefore, if you have a SUM formula that calculates the total of the values in a column of a list, the formula returns the same value when you filter the list.

To create formulas that return results based on filtering criteria, you need to use Excel's database worksheet functions. For example, you can create a formula that calculates the sum of values in a list that meets certain criteria. Set up a criteria range as described previously. Then enter a formula such as the following:

```
=DSUM(ListRange,FieldName,Criteria)
```

In this case, ListRange refers to the list, FieldName refers to the field name cell of the column that you are summing, and Criteria refers to the criteria range.

Excel's database functions are listed in Table 23-3.

Table 23-3
Excel's Database Worksheet Functions

Function	*Description*
DAVERAGE	Returns the average of selected database entries
DCOUNT	Counts the cells containing numbers from a specified database and criteria
DCOUNTA	Counts nonblank cells from a specified database and criteria
DGET	Extracts from a database a single record that matches the specified criteria
DMAX	Returns the maximum value from selected database entries
DMIN	Returns the minimum value from selected database entries
DPRODUCT	Multiplies the values in a particular field of records that match the criteria in a database
DSTDEV	Estimates the standard deviation based on a sample of selected database entries
DSTDEVP	Calculates the standard deviation based on the entire population of selected database entries
DSUM	Adds the numbers in the field column of records in the database that match the criteria
DVAR	Estimates variance based on a sample from selected database entries
DVARP	Calculates variance based on the entire population of selected database entries

Refer to Chapter 10 for general information about using worksheet functions.

Sorting a List

In some cases, the order of the rows in your list doesn't matter. But in other cases, you want the rows to appear in a specific order. For example, in a price list, you may want the rows to appear in alphabetical order by product name. This makes the products easier to locate in the list. Or, if you have a list of accounts receivable information, you may want to sort the list so that the higher amounts appear at the top of the list (in descending order).

Rearranging the order of the rows in a list is called *sorting*. Excel is quite flexible when it comes to sorting lists, and you can often accomplish this task with the click of a mouse button.

Simple Sorting

To quickly sort a list in ascending order, move the cell pointer into the column that you want to sort. Then click on the Sort Ascending button the Standard toolbar. The Sort Descending button works the same way, but it sorts the list in descending order. In both cases, Excel determines the extent of your list and sorts all the rows in the list.

When you sort a filtered list, Excel sorts only the visible rows. When you remove the filtering from the list, the list is no longer sorted.

Be careful if you sort a list that contains formulas. If the formulas refer to cells in the list that are in the same row, you don't have any problems. But if the formulas refer to cells in other rows in the list or to cells outside the list, the formulas will not be correct after you sort the list. If formulas in your list refer to cells outside the list, make sure that the formulas use an absolute cell reference.

More Complex Sorting

Sometimes, you may want to sort by two or more columns. This is relevant to break ties. A tie occurs when rows with duplicate data remain unsorted. Figure 23-16 shows an example of an unsorted list. If you sort this list by Month, Excel places the rows for each month together. But you may also want to show the Sales Reps in ascending order within each month. In this case, you would need to sort by two columns (Month and Sales Rep). Figure 23-17 shows the list after sorting by these two columns.

You can use the Sort Ascending and Sort Descending buttons to do this—but you need to do two sorts. First, sort by the Sales Reps column, and then sort by the Month column. As I explain in the next section, Excel provides a way to accomplish multicolumn sorting with a single command.

Sales-db.xls

	A	B	C	D	E	F
7	Month	Sales Rep	Type	Unit Cost	Quantity	Total Sale
8	May	Sheldon	Existing	125	1	125
9	January	Sheldon	Existing	175	1	175
10	January	Sheldon	New	140	6	840
11	January	Jenkins	New	225	1	225
12	February	Robinson	New	225	1	225
13	March	Wilson	Existing	125	4	500
14	April	Robinson	Existing	125	2	250
15	February	Sheldon	Existing	175	1	175
16	March	Robinson	Existing	125	1	125
17	May	Jenkins	New	225	3	675
18	April	Jenkins	New	225	2	450
19	February	Wilson	Existing	125	5	625
20	February	Jenkins	New	225	2	450
21	January	Franks	New	225	4	900
22	May	Wilson	New	225	1	225
23	January	Sheldon	New	225	1	225
24	March	Jenkins	New	225	2	450
25	March	Jenkins	Existing	125	5	625
26	April	Peterson	New	140	2	280
27	February	Franks	Existing	175	2	350
28	May	Robinson	New	140	3	420
29	April	Peterson	Existing	175	6	1050
30	February	Robinson	New	225	3	675

SALES-DB

Figure 23-16: This list is unsorted.

Sales-db.xls

	A	B	C	D	E	F
7	Month	Sales Rep	Type	Unit Cost	Quantity	Total Sale
8	January	Franks	New	225	4	900
9	January	Franks	Existing	175	1	175
10	January	Franks	Existing	175	5	875
11	January	Franks	New	225	1	225
12	January	Franks	Existing	175	1	175
13	January	Franks	Existing	125	3	375
14	January	Jenkins	New	225	1	225
15	January	Jenkins	Existing	125	1	125
16	January	Jenkins	New	140	3	420
17	January	Jenkins	Existing	175	2	350
18	January	Jenkins	New	140	1	140
19	January	Peterson	Existing	125	1	125
20	January	Peterson	Existing	125	3	375
21	January	Peterson	New	140	1	140
22	January	Peterson	New	225	1	225
23	January	Robinson	New	140	2	280
24	January	Robinson	Existing	125	5	625
25	January	Robinson	Existing	175	4	700
26	January	Sheldon	Existing	175	1	175
27	January	Sheldon	New	140	6	840
28	January	Sheldon	New	225	1	225
29	January	Sheldon	Existing	125	2	250
30	January	Sheldon	Existing	175	5	875

SALES-DB

Figure 23-17: The list after sorting on two fields.

Excel's Sorting Rules

Because cells can contain different types of information, you may be curious about how Excel sorts this information. For an ascending sort, the information appears in the following order:

1. Values: Excel sorts numbers from smallest negative to largest positive, and treats dates and times as values. In all cases, Excel sorts using the actual values in cells (not their formatted appearance).
2. Text: In alphabetical order, as follows: 0 1 2 3 4 5 6 7 8 9 (space) ! " # $ % & ' ( ) * + , - . / : ; < = > ? @ [\ ] ^ _ ` { | } ~ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z.

   By default, sorting is not case sensitive. You can change this behavior, however, in the Sort Options dialog box (described in this chapter).
3. Logical values: False comes before True.
4. Error values: Error values (such as #VALUE! and #NA) appear in their original order; Excel does not sort them by error type.
5. Blank cells: Blanks cells always appear last.

Sorting in descending order reverses this sequence — except that blank cells *still* appear last.

The Sort dialog box

If you want to sort by more than one field, choose Data• Sort. Excel displays the dialog box that is shown in Figure 23-18. Simply select the first sort field from the drop-down list labeled Sort By, and specify Ascending or Descending order. Then, do the same for the second sort field. If you want to sort by a third field, specify the field in the third section. If the Header Row option is set, the first row (field names) is not affected by the sort. Click on OK, and the list's rows rearrange in a flash.

If the sorting didn't occur as you expected, select Edit• Undo (or press Ctrl+Z) to undo the sorting.

What if you need to sort your list by more than three fields? It can be done, but it takes an additional step. For example, assume that you want to sort your list by five fields: Field1, Field2, Field3, Field4, and Field5. Start by sorting by Field3, Field4, and Field5. Then re-sort the list by Field1 and Field2. In other words, sort the three "least important" fields first; they remain in sequence when you do the second sort.

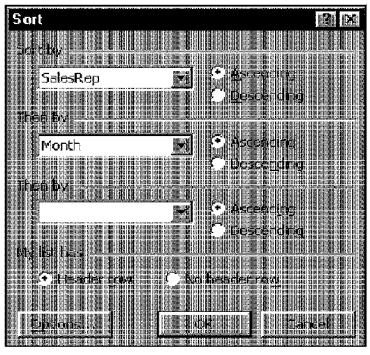


Figure 23-18: The Sort dialog box lets you sort by up to three columns.

Often, you want to keep the records in their original order but perform a temporary sort just to see how it looks. The solution is to add an additional column to the list with sequential numbers in it (don't use formulas to generate these numbers, but you can use the Fill command). Then, after you sort, you can return to the original order by resorting on the field that contains the sequential numbers. You can also use Excel's undo feature to return the list to its original order. If you use an additional column, you can perform other operations while the list is temporarily sorted (and these operations won't be undone when you undo the sort operation).

Sort options

When you click on the Options button in the Sort dialog box, Excel displays the Sort Options dialog, shown in Figure 23-19.

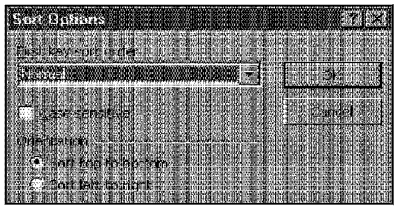


Figure 23-19: The Sort Options dialog gives you some additional sorting options.

These options are described as follows:

- **First key sort order:** Lets you specify a custom sort order for the sort (see the next section).

How Excel Identifies a Header Row

When you use the Data• Sort command, there's no need to select the list before you choose the command. That's because Excel examines the active cell position and then establishes the list's boundaries for you. In addition, Excel attempts to determine whether the list contains a header row. If the list has a header row, Excel excludes this row from the sort.

How does this happen? I'm not sure exactly, but the following seems to be Excel's "thought" process:

1. Select the current region. (You can do this manually: press F5, click on the Special button, select the Current Region option, and click on OK.)
2. Examine the first row of the selection.
3. Determine whether the first row contains any blanks. If so, this list has no header row.
4. Determine whether the first row contains text. If so, check the other cells. If they also contain text, this list has no header row.
5. Determine whether the first row contains uppercase text while the list itself contains lowercase or proper case text. If so, this list has a header row.
6. Determine whether the cells in the first row are formatted differently from the other cells in the list. If so, this list has a header row.

Knowing this information can help you eliminate incorrect sorting. For example, if you want to sort a range that doesn't have a header row, you need to make sure that Excel doesn't sort the data as if it had a header row. For best results, use the Sort Ascending and Sort Descending toolbar buttons only when the data that you're sorting has headers. If your data contains no headers, select Data• Sort and make sure that the No Header Row option is selected.

- **Case sensitive:** Makes the sorting case sensitive so that uppercase letters appear before lowercase letters in an ascending sort. Normally, sorting ignores the case of letters.
- **Orientation:** Enables you to sort by columns rather than by rows (the default).

Using a Custom Sort Order

Excel typically sorts either numerically or alphabetically, depending on the data being sorted. In some cases, however, you may want to sort your data in other ways. For example, if your data consists of month names, you usually want it to appear in month order rather than alphabetically. You can use the Sort Options dialog box to perform such a sort. Select the appropriate list from the drop-down list labeled First key sort order. Excel, by default, has four "custom lists," and you can define your own. Excel's custom lists are as follows:

- **Abbreviated days:** Sun, Mon, Tue, Wed, Thu, Fri, Sat
- **Days:** Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday
- **Abbreviated months:** Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
- **Months:** January, February, March, April, May, June, July, August, September, October, November, December

Note that the abbreviated days and months do not have periods after them. If you use periods for these abbreviations, Excel doesn't recognize them (and doesn't sort them correctly).

You may want to create a custom list. For example, your company may have several stores, and you want the stores to be listed in a particular order (not alphabetically). If you create a custom list, sorting puts the items in the order that you specify in the list. You must use the Data • Sort command to sort by a custom list (click on the Options button to specify the custom list).

To create a custom list, use the Custom Lists tab of the Options dialog box, as shown in Figure 23-20. Select the NEW LIST option, and make your entries (in order) in the List Entries box. Or, you can import your custom list from a range of cells by selecting the range and then clicking the Import button.

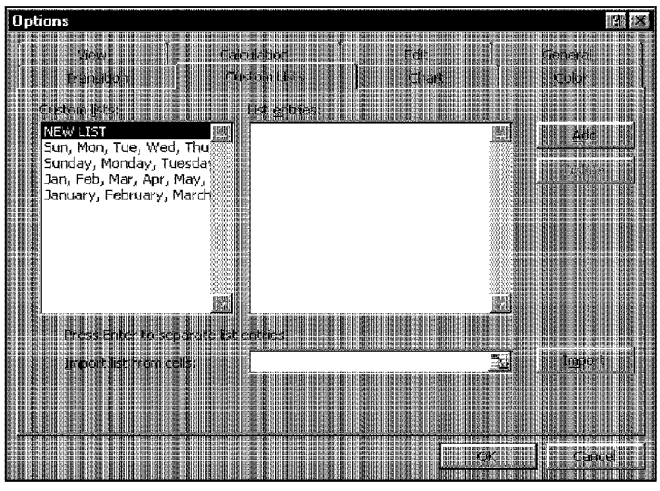


Figure 23-20: Excel lets you create custom sorting lists.

Custom lists also work with the AutoFill handle in cells. If you enter the first item of a custom list and then drag the cell's AutoFill handle, Excel fills in the remaining list items automatically.

Sorting Nonlists

You can, of course, sort any range in a worksheet — it doesn't have to be a list. You need to be aware of a few things, however. The Sort Ascending and Sort Descending toolbar buttons may assume (erroneously) that the top row is a header row and not include these cells in the sort (see the sidebar, "How Excel identifies a header row," earlier in this chapter).

Therefore, to avoid potential errors when sorting non-lists, don't use these toolbar buttons. Rather, select the entire range, and select Data • Sort (making sure that you choose the No Header Row option).

Creating Subtotals

The final topic of this chapter is automatic subtotals — a handy feature that can save you a great deal of time. To use this feature, your list must be sorted, because the subtotals are inserted whenever the value in a specified field changes. Figure 23-21 shows an example of a list, sorted by the Month field, which is appropriate for subtotals.

Sales-db.xls

	A	B	C	D	E	F	G
1	Month	Sales Rep	Type	Unit Cost	Quantity	Total Sale	
2	January	Franks	New	225	4	900	
3	January	Franks	Existing	175	5	875	
4	January	Franks	New	225	1	225	
5	January	Franks	Existing	175	1	175	
6	January	Jenkins	New	225	1	225	
7	January	Jenkins	Existing	125	1	125	
8	February	Franks	New	225	4	900	
9	February	Jenkins	New	225	2	450	
10	February	Jenkins	New	225	3	675	
11	February	Jenkins	New	225	3	675	
12	February	Jenkins	New	225	3	675	
13	February	Jenkins	Existing	175	1	175	
14	February	Peterson	New	225	1	225	
15	February	Peterson	New	225	2	450	
16	March	Peterson	Existing	125	2	250	
17	March	Peterson	New	225	2	450	
18	March	Robinson	Existing	125	1	125	
19	March	Robinson	Existing	125	5	625	
20	March	Robinson	New	225	4	900	
21	April	Franks	New	175	4	700	
22	April	Franks	New	175	3	525	
23	April	Jenkins	New	225	2	450	
24	April	Jenkins	New	140	3	420	

SALES DB

Figure 23-21: This list is a good candidate for subtotals, which are inserted at each change of the month.

To insert subtotal formulas into a list automatically, move the cell pointer anywhere in the list and choose Data • Subtotals. You see the dialog box shown in Figure 23-22.

Figure 23-22: The Subtotal dialog box automatically inserts subtotal formulas into a sorted list.

This dialog box offers the following choices:

- **At Each Change in:** This drop-down list displays all fields in your list. You must have sorted the list by the field that you choose.
- **Use Function: Choose from 11 functions:** You should normally use Sum (the default).
- **Add Subtotal to:** This list box shows all the fields in your list. Place a check mark next to the field or fields that you want to subtotal.
- **Replace Current Subtotals:** If this box is checked, Excel removes any existing subtotal formulas and replaces them with the new subtotals.
- **Page Break Between Groups:** If this box is checked, Excel inserts a manual page break after each subtotal.
- **Summary Below Data:** If this box is checked, Excel places the subtotals below the data (the default). Otherwise, the subtotal formulas appear above the totals.
- **Remove All:** This button removes all subtotal formulas in the list.

When you click on OK, Excel analyzes the list and inserts formulas as specified—and creates an outline for you. The formulas all use the SUBTOTAL worksheet function.

When you add subtotals to a filtered list, the subtotals may no longer be accurate when the filter is removed.

Figure 23-23 shows a worksheet after adding subtotals.

Sales-db.xls

	A	B	C	D	E	F
1	Month	Sales Rep	Type	Unit Cost	Quantity	Total Sale
2	January	Franks	New	225	4	900
3	January	Franks	Existing	175	5	875
4	January	Franks	New	225	1	225
5	January	Franks	Existing	175	1	175
6	January	Jerkins	New	225	1	225
7	January	Jerkins	Existing	125	1	125
8	**January Total**					2525
9	February	Franks	New	225	4	900
10	February	Jerkins	New	225	2	450
11	February	Jerkins	New	225	3	675
12	February	Jerkins	New	225	3	675
13	February	Jerkins	New	225	3	675
14	February	Jerkins	Existing	175	1	175
15	February	Peterson	New	225	1	225
16	February	Peterson	New	225	2	450
17	**February Total**					4225
18	March	Peterson	Existing	125	2	250
19	March	Peterson	New	225	2	450
20	March	Robinson	Existing	125	1	125
21	March	Robinson	Existing	125	5	625
22	March	Robinson	New	225	4	900
23	**March Total**					2350
24	April	Franks	New	175	4	700

SALES-DB

Figure 23-23: Excel added the subtotal formulas automatically—and even created an outline.

Summary

In this chapter, I discuss lists. A list is simply a database table that is stored on a worksheet. The first row of the list (the header row) contains field names, and subsequent rows contain data (records). I offer some pointers on data entry and discuss two ways to filter a list to show only rows that meet certain criteria. AutoFiltering is adequate for many tasks, but if your filtering needs are more complex, you need to use advanced filtering. I end the chapter with a discussion of sorting and Excel's automatic subtotal feature.

CHAPTER 24

Using External Database Files

In This Chapter

The preceding chapter described how to work with lists that are stored in a worksheet. Many users find that worksheet lists are sufficient for their data tracking. Others, however, choose to take advantage of Excel's capability to access data that is stored in external database files. That's the topic of this chapter.

Why Use External Database Files?

Accessing external database files from Excel is useful when you have the following situations:

- You need to work with a very large database.
- You share the database with others; that is, other users have access to the database and may need to work with the data at the same time.
- You want to work with only a subset of the data—data that meets certain criteria that you specify.
- The database is in a format that Excel can't read.

If you need to work with external databases, you may prefer Excel to other database programs. The advantage? After you bring the data into Excel, you can manipulate and format it by using familiar tools.

As you may know, Excel can read some database files directly—specifically, those produced by various versions of dBASE (with a DBF extension). If the database has fewer than 65,535 records and no more than 255 fields, you can load the entire file into a worksheet, memory permitting. Even if you have enough memory to load such a large file, however, Excel's performance would likely be poor.

In many cases, you may not be interested in all the records or fields in the file. Instead, you may want to bring in just the

data that meets certain criteria. In other words, you want to *query* the database and load into your worksheet a subset of the external database that meets the criteria. Excel makes this type of operation relatively easy.

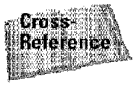
To perform queries using external databases, Microsoft Query must be installed on your system. If Query is not installed, you will be prompted to install it when you select the Data • Get External Data • Create New Query command. You must rerun the Excel (or Microsoft Office) setup program and install Query.

In previous versions of Excel, using Microsoft Query required that you load an add-in. That is no longer necessary, starting with Excel 97, although the add-in is still included for compatibility purposes.

To work with an external database file from Excel, use the Query application that is included with Excel. The general procedure is as follows:

1. Activate a worksheet.
2. Choose Data • Get External Data • New Database Query. This starts Query.
3. Specify whether you want to use Query directly or use Query Wizard.
4. Specify the database that you want to use and then create a *query*—a list of criteria that specifies which records you want.
5. Specify how you want the data returned that passes your query—either to a worksheet or as a pivot table.

You can choose to save the query in a file so that you can reuse it later. This means that modifying the query or *refreshing* it (updating it with any changed values) is a simple matter. This is particularly useful when the data resides in a shared database that is continually being updated.



The next chapter discusses pivot tables. You can create a pivot table using data in an external file, and use Query to retrieve data.

Using Query: An Example

The best way to become familiar with Query is to walk through an example.

The Database File

The file that is used in this example is named Budget.dbf.

If you want to try this example, you can find it on this book's CD-ROM.

This database file is a dBASE IV database with a single table that consists of 15,840 records. This file contains the following fields:

- **Sort:** A numeric field that holds record sequence numbers.
- **Division:** A text field that specifies the company division (Asia, Europe, N. America, Pacific Rim, or S. America).
- **Department:** A text field that specifies the department within the division. Each division is organized into the following departments: Accounting, Advertising, Data Processing, Human Resources, Operations, Public Relations, R&D, Sales, Security, Shipping, and Training.
- **Category:** A text field that specifies the budget category. The four categories are Compensation, Equipment, Facility, and Supplies & Services.
- **Item:** A text field that specifies the budget item. Each budget category has different budget items. For example, the Compensation category includes the following items: Benefits, Bonuses, Commissions, Conferences, Entertainment, Payroll Taxes, Salaries, and Training.
- **Month:** A text field that specifies the month (abbreviated as Jan, Feb, and so on).
- **Budget:** A numeric field that stores the budgeted amount.
- **Actual:** A numeric field that stores the actual amount spent.
- **Variance:** A numeric field that stores the difference between the Budget and Actual fields.

The Task

The objective of this exercise is to develop a report that shows the first quarter (January through March) actual compensation expenditures of the training department in the North American division. In other words, the query will extract records for which the following applies:

- The Division is N. America
- The Department is Training
- The Category is Compensation
- The Month is Jan, Feb, or Mar

Using Query to Get the Data

You *could* import the entire dBASE file into a worksheet and then choose Data • Filter • AutoFilter to filter the data as required. This approach would work, because the file has fewer than 65,535 records, which isn't always the case. Using Query, however, you import only the data that's required.

Some Database Terminology

People who spend their days working with databases seem to have their own special language. The following terms can help you hold your own among a group of database experts:

External database: A collection of data that is stored in one or more files (not Excel files). Each file of a database holds a single table, and tables are comprised of records and fields.

Field: In a database table, an element of a record that corresponds to a column.

ODBC: An acronym for Open DataBase Connectivity, a standard developed by Microsoft that uses drivers to access database files in different formats. Microsoft Query comes with drivers for Access, dBASE, FoxPro, Paradox, SQL Server, Excel workbooks, and ASCII text files. ODBC drivers for other databases are available from Microsoft and third-party providers.

Query: To search a database for records that meet specific criteria. This term is also used as a noun; you can write a query, for example.

Record: In a database table, a single element that corresponds to a row.

Refresh: To rerun a query to get the latest data. This is applicable when the database contains information that is subject to change, as in a multiuser environment.

Relational database: A database that is stored in more than one table or file. At least one common field (sometimes called the *key field*) connects the tables.

Result set: The data that is returned by a query, usually a subset of the original database. Query returns the result set to your Excel workbook or to a pivot table.

SQL: An acronym for Structured Query Language (usually pronounced *sequel*). Query uses SQL to query data that is stored in ODBC databases.

Table: A record- and field-oriented collection of data. A database consists of one or more tables.

Starting query

Begin with an empty worksheet. Select Data • Get External Data • New Database Query; this action launches and activates Microsoft Query, a separate application. Excel continues to run, and you can switch back and forth between Query and Excel, as needed.

Selecting a data source

When Query starts, it displays the Choose Data Source dialog box, shown in Figure 24-1. This dialog box contains three tabs:

- Databases: Lists the data sources that are known to Query — this tab may be empty, depending on which data sources are defined on your system.
- Queries: Contains a list of stored queries. Again, this may or may not be empty.
- OLAP Cubes: Lists OLAP databases (see sidebar) that are available for Query.

Figure 24-1: The Choose Data Source dialog box.

OLAP Databases

OLAP is an acronym for *online analytical processing*; OLAP presents a new way to organize large databases to suit the way that you analyze and manage information. In an OLAP database, data is organized by level of detail. In a business database, for example, you might want to track sales around the world for the products of a particular company. In an OLAP organization of this information, you would need to consider where and when each product was sold, as well as which product was sold. Each of these aspects of the OLAP database is called a *dimension*, and each dimension is comprised of several fields that can be organized hierarchically, by level of detail. You might call the "where" dimension the Location dimension, and it might contain, for example, fields for country, region, and city. The Time dimension, containing information about when the product was sold, might contain fields for month, date, day, and year.

Dimensions in an OLAP database combine to provide information about the intersecting points; because you can combine several dimensions, OLAP databases are called *cubes*.

You can connect Excel to an OLAP data source created with either the Microsoft DSS (Decision Support Services) Analysis server or other third-party OLAP products that provide data source drivers that are compatible with OLD-DB for OLAP. You connect to an OLAP cube the same way that you connect to other external data sources. Excel can display data that you retrieve from an OLAP cube as either a PivotTable or a PivotChart. You cannot display OLAP data as an external data range of the type discussed in this chapter.

If you've previously worked with a particular database, its name appears in the list of databases. Otherwise, you need to identify the source.

In the Databases tab, select the <New Database Source> option and click OK. This displays the Create New Data Source dialog box, shown in Figure 24-2.

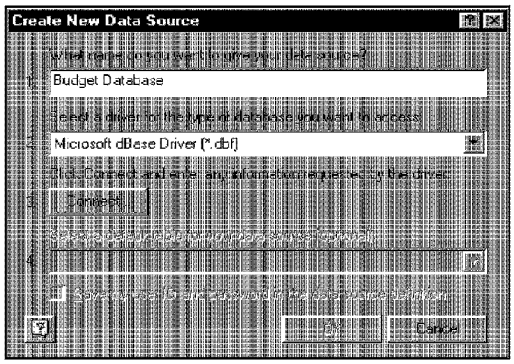


Figure 24-2: The Create New Data Source dialog box.

The Create New Data Source dialog box has the following four numbered parts:

1. Enter a descriptive name for the data source. For this example, the name is Budget Database.
2. Select a driver for the data source by selecting from the list of installed drivers. Because the database file in this example is a dBASE file, select the driver named Microsoft dBASE Driver.
3. The Connect button displays the ODBC Setup dialog box that asks for information specific to the driver that you select in Step 2. In this dialog box, you select the directory where the database is located.
4. Select the default data table that you want to use (this step is optional). If the database requires a password, you can also specify that the password be saved with the Data Source definition.

After you supply all the information in the Create New Data Source dialog box, click OK, and Excel redisplays the Choose Data Source dialog box—which now includes the data source that you created.

You have to go through these steps only once for each data source. The next time that you access Query, the Budget Database (and any other database sources that you define) appears in the Choose Data Source dialog box.

Using the ODBC Manager

Occasionally, you may need to edit data sources – for example, if you move your database files to a new location. You can do this by using the ODBC Manager utility. This program is available in the Windows Control Panel (it's called *32-bit ODBC*). This utility also lets you add new data sources and remove those that you no longer need.

Use Query Wizard?

The Choose Data Source dialog box has a check box at the bottom that lets you specify whether to use Query Wizard to create your query. Query Wizard walks you through the steps that are used to create your query, and if you use Query Wizard, you don't have to deal directly with Query. I highly recommend using Query Wizard — and the examples in this chapter use this tool.

In the Choose Data Sources dialog box, make sure that you check the Query Wizard check box at the bottom of the dialog box and then click OK to start Query Wizard.

Query Wizard: Choosing the columns

In the first step of Query Wizard (see Figure 24-3), select the database columns that you want to appear in your query.

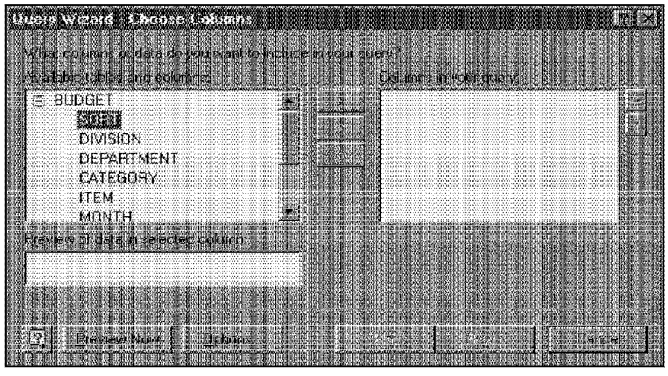


Figure 24-3: In the first step of Query Wizard, you select the columns to use in your query.

The columns that you select determine the fields from the database that Query returns to Excel. Recall that the query for this example involves selecting records based on the following fields: Division, Department, Month, Category, and Actual. You also want to add the Item field. The left tab of the dialog box shows all the available columns. To add a column to the right tab, select the column and click the > button (or, you can double-click the column name).

After you finish adding the columns, the Query Wizard dialog box looks like Figure 24-4.

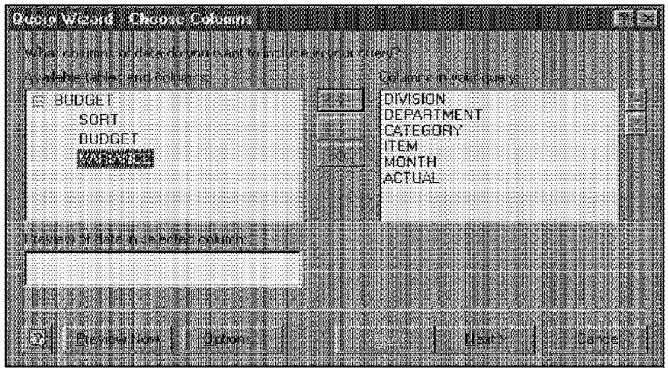


Figure 24-4: Six columns have been added to the query.

If you want to see the data for a particular column, select the column and click the Preview Now button. If you accidentally add a column that you don't need, select it in the right tab and click the < button to remove it. After you select all the columns for the query, click the Next button.

Query Wizard: Filtering data

In the second Query Wizard dialog box, you specify your record selection criteria—how you want to filter the data. This step is optional. If you want to retrieve all the data, just click the Next button to proceed. Figure 24-5 shows the Filter Data dialog box of Query Wizard.

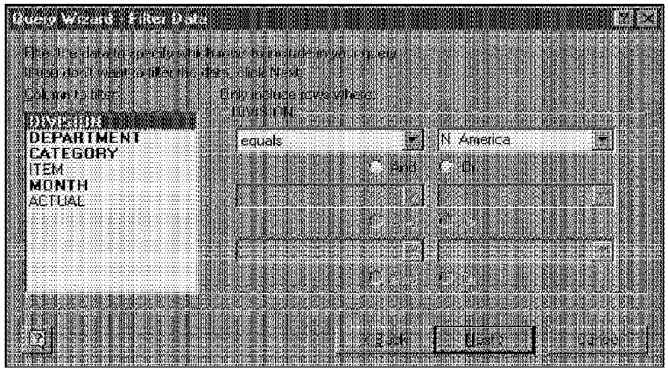


Figure 24-5: In the second step of Query Wizard, you specify how you want to filter the data.

For the example, you don't need all records. Recall that you're interested only in the records in which one of the following applies:

- The Division is N. America.
- The Department is Training.
- The Category is Compensation.
- The Month is Jan, Feb, or Mar.

You enter the criteria by column. In this case, you need to specify four criteria (one for each of four columns):

1. In the Column to filter column, select DIVISION. In the right tab, select equals from the first drop-down list, and N. America from the second drop-down list.
2. In the Column to filter column, select DEPARTMENT. In the right tab, select equals from the first drop-down list, and Training from the second drop-down list.
3. In the Column to filter column, select CATEGORY. In the right tab, select equals from the first drop-down list, and Compensation from the second drop-down list.
4. In the Column to filter column, select MONTH. In the right tab, select equals from the first drop-down list, and Jan from the second drop-down list. Because this column is filtered by multiple values, click the Or option and then select equals and Feb from the drop-down lists in the second row. Finally, select equals and Mar from the drop-down lists in the second row.

To review the criteria that you've entered, select the column from the Column to filter list. Query Wizard displays the criteria that you entered for the selected column. After you enter all the criteria, click Next.

Query Wizard: Sort order

The third step of the Query Wizard enables you to specify how you want the records to be sorted (see Figure 24-6). This step is optional, and you can click Next to move to the next step if you don't want the data sorted or prefer to sort it after it's returned to your worksheet.

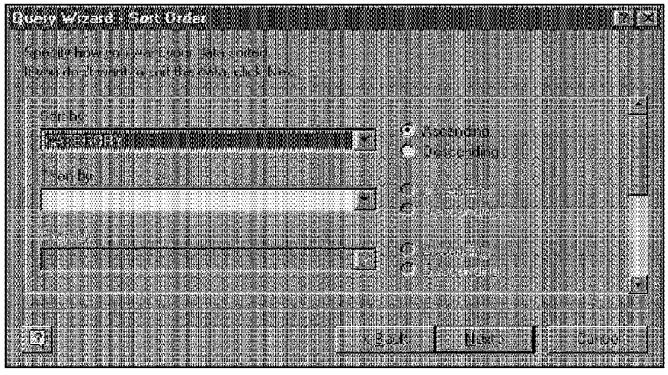


Figure 24-6: In the third step of Query Wizard, you specify the sort order.

For this example, sort by CATEGORY in Ascending order. You can specify as many sort fields as you like. Click Next to move to the next step.

Query Wizard: Finish

The final step of Query Wizard, shown in Figure 24-7, lets you do the following things:

- Give the query a name
- Save it to a file, so that it can be reused
- Specify what to do with the data

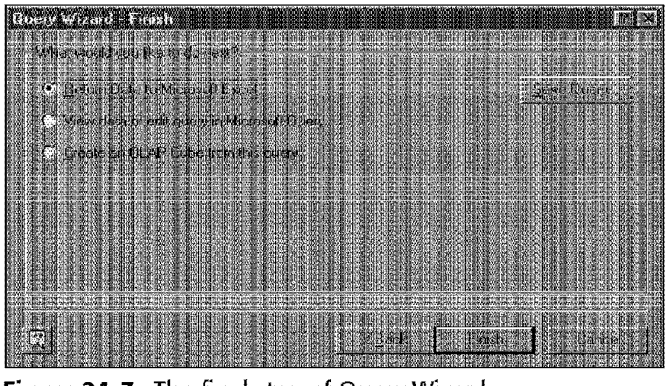


Figure 24-7: The final step of Query Wizard.

Normally, you want to return the data to Excel. If you know how to use the Microsoft Query application, you can return the data to Query and examine it, or even modify the selection criteria. Or, you can create an OLAP cube to use in a PivotTable or PivotChart report.

If you plan to reuse this query, you should save it to a file. Click the Save Query button, and you are prompted for a filename. After you make your choices, click Finish.

Query Operators

The following table lists and describes the operators that are available when you create a query. These operators give you complete control over which rows are returned.

Operator	*What It Does*
equals	Field is identical to value
does not equal	Field is not equal to value
is greater than	Field is greater than value
is greater than or equal to	Field is greater than or equal to value
is less than	Field is less than value
is less than or equal to	Field is less than or equal to value
is one of	Field is in a list of values, separated by commas
is not one of	Field is not in a list of values, separated by commas
is between	Field is between two values, separated by commas
is not between	Field is not between two values, separated by commas
begins with	Field begins with the value
does not begin with	Field does not begin with value
ends with	Field ends with value
does not end with	Field does not end with value
contains	Field contains value
does not contain	Field does not contain value
like	Field is like value (using * and ? wildcard characters)
not like	Field is not like value (using * and ? wildcard characters)
is Null	Field is empty
is not Null	Field is not empty

Specifying a Location for the Data

Figure 24-8 shows the Returning External Data to Microsoft Excel dialog box, which appears when you click the Finish button in the Query Wizard dialog box.

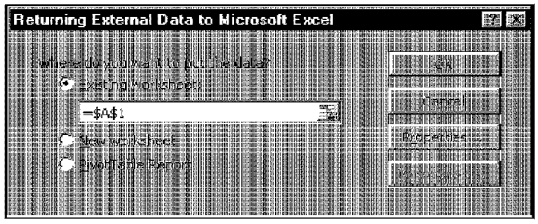


Figure 24-8: Specifying what to do with the data.

You can select from the following choices:

- **Existing worksheet:** You can specify the upper-left cell.
- **New worksheet:** Excel can insert a new worksheet and insert the data beginning in cell A1.
- **Pivot Table Report:** Excel can display its Pivot Table Wizard, so that you can specify the layout for a pivot table (see Chapter 25).

Figure 24-9 shows the data that is returned to a worksheet.

Book1

	DIVISION	DEPARTMENT	CATEGORY	ITEM	MONTH	ACTUAL
2	N. America	Training	Compensation	Payroll Taxes	Feb	3642
3	N. America	Training	Compensation	Benefits	Jan	3283
4	N. America	Training	Compensation	Bonuses	Jan	3331
5	N. America	Training	Compensation	Commissions	Jan	3143
6	N. America	Training	Compensation	Payroll Taxes	Jan	3516
7	N. America	Training	Compensation	Training	Jan	4058
8	N. America	Training	Compensation	Conferences	Jan	4281
9	N. America	Training	Compensation	Entertainment	Jan	3344
10	N. America	Training	Compensation	Salaries	Feb	3972
11	N. America	Training	Compensation	Benefits	Feb	3995
12	N. America	Training	Compensation	Salaries	Jan	4313
13	N. America	Training	Compensation	Commissions	Feb	3288
14	N. America	Training	Compensation	Entertainment	Mar	3205
15	N. America	Training	Compensation	Training	Feb	3757
16	N. America	Training	Compensation	Conferences	Feb	4055
17	N. America	Training	Compensation	Entertainment	Feb	3724
18	N. America	Training	Compensation	Salaries	Mar	3748
19	N. America	Training	Compensation	Benefits	Mar	3608
20	N. America	Training	Compensation	Bonuses	Mar	3609
21	N. America	Training	Compensation	Commissions	Mar	3271
22	N. America	Training	Compensation	Payroll Taxes	Mar	3547
23	N. America	Training	Compensation	Training	Mar	3678
24	N. America	Training	Compensation	Conferences	Mar	4146
25	N. America	Training	Compensation	Bonuses	Feb	2611

Sheet1

Figure 24-9: The results of the query.

Working with an External Data Range

Excel stores the data that Query returns in either a worksheet or a pivot table. When Excel stores data in a worksheet, it stores the data in a specially named range, known as an *external data range;* Excel creates the name for this range automatically.

This section describes what you can do with the data that Excel receives from Query and stores in a worksheet.

Adjusting External Data Range Properties

You can adjust various properties of the external data range by using the External Data Range Properties dialog box (see Figure 24-10).

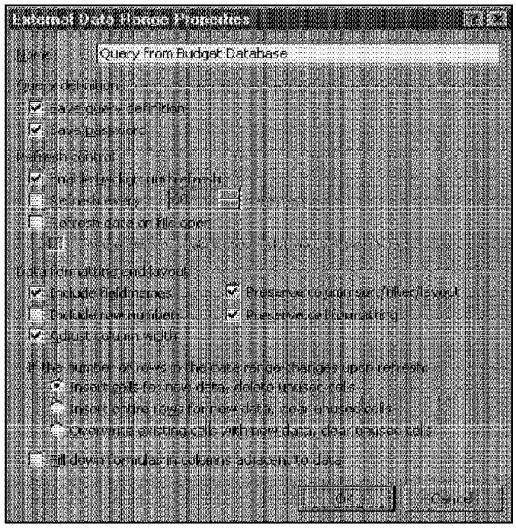


Figure 24-10: The External Data Range Properties dialog box enables you to specify various options for an external data range.

To display this dialog box, the cell pointer must be within the external data range. You can open this dialog box by using any of three methods:

- Right-click and select Data Range Properties from the shortcut menu.
- Select Data • Get External Data • Data Range Properties.
- Click the Data Range properties tool on the External Data toolbar (this toolbar appears automatically when you perform a query).

The following list describes the options in the External Data Range Properties dialog box:

- **Name:** The name of the external data range. You can change this name or use the default name that Excel creates. Excel substitutes, in the range name, the underscore character for any spaces that you see in the Name box of the External Data Range Properties box.
- **Query definition:** If you check Save query definition, Excel stores the query definition with the external data range, enabling you to refresh the data or edit the query, if necessary. If the database requires a password, you can also store the password so that you don't need to enter it when you refresh the query.
- **Refresh control:** Determines how and when Excel refreshes the data.
- **Data formatting and layout:** Determines the appearance of the external data range.

The External Data Range Properties dialog box has quite a few options. For specific details, click the Help icon in the title bar and then click an option in the dialog box.

You can manipulate data returned from a query just like any other worksheet range. For example, you can sort the data, format it, or create formulas that use the data.

In prior versions of Excel, if you intend to refresh the query, you need to keep the external data range intact. That is, you can't insert new rows or columns in the external data range, because refreshing the query causes the external range to be rewritten. Similarly, you lose any formatting that you applied to the external data range when you refresh the query.

In Excel 2000, refreshing a query *does not* overwrite the external data range. You are free to format the external data range or insert rows and columns. You also can include formulas in those rows and columns that refer to other parts of the external data range. Your work *will not* be destroyed when you refresh the query.

Refreshing a Query

After performing a query, you can save the file and then retrieve it later. The file contains the data that you originally retrieved from the external database. The external database may have changed, however, in the interim.

If you checked the Save query definition option in the External Data Range Properties dialog box, then Excel saves the query definition with the workbook. Simply move the cell pointer anywhere within the external data table in the worksheet and then use one of the following methods to refresh the query:

- Right-click and select Refresh Data from the shortcut menu
- Select Data • Refresh Data
- Click the Refresh Data tool on the External Data toolbar

Excel launches Query and uses your original query to bring in the current data from the external database.

If you find that refreshing the query causes undesirable results, use Excel's Undo feature to "unrefresh" the data.

Making Multiple Queries

A single workbook can hold as many external data ranges as you need. Excel gives each query a unique name, and you can work with each query independently. Excel automatically keeps track of the query that produces each external data range.

Copying or Moving a Query

After performing a query, you may want to copy or move the external data range, which you can do by using the normal copy, cut, and paste techniques. However, make sure that you copy or cut the entire external data range—otherwise, the underlying query is not copied, and the copied data cannot be refreshed.

Deleting a Query

If you decide that you no longer need the data that is returned by a query, you can delete it by selecting the entire external data range and choosing Edit • Delete.

If you simply press Delete, the contents of the cells are erased, but the underlying query remains. Excel displays a dialog box asking whether you want to delete the query. If you choose No, you can refresh the query, and the deleted cells appear again, including any formatting that you applied to them.

When you refresh, Query returns only data that is retrieved from the external database. If you delete rows or columns that you inserted into the external data range, Query does not redisplay those rows and columns when you refresh.

Changing Your Query

If you bring the query results into your worksheet and discover that you don't have what you want, you can modify the query. Move the cell pointer anywhere within the external data table in the worksheet and then use one of the following methods to refresh the query:

- Right-click and select Edit Query from the shortcut menu
- Select Data • Get External Data • Edit Query
- Click the Edit Query tool on the External Data toolbar

Excel then launches (or activates) Query, and you can change the original query. After you finish, choose File • Return Data to Microsoft Excel. Excel reactivates, executes the modified query, and updates the external data range.

Using Microsoft Query Without Query Wizard

Previous sections in this chapter describe how to use Query Wizard to create a database query. Query Wizard is essentially a "front end" for Microsoft Query. In some cases, you may want to use Query itself rather than Query Wizard.

When you select Data • Get External Data • Create New Query, the Choose Data Source dialog box gives you the option of whether to use Query Wizard. If you choose not to use Query Wizard, you work directly with Microsoft Query.

Creating a Query

Before you can create a query, you must display the Criteria pane. In Query, open the View menu and confirm that a check appears next to the Criteria command. If you don't see a check, choose View • Criteria to display the Criteria pane in the middle of the window. (See Figure 24-11.)

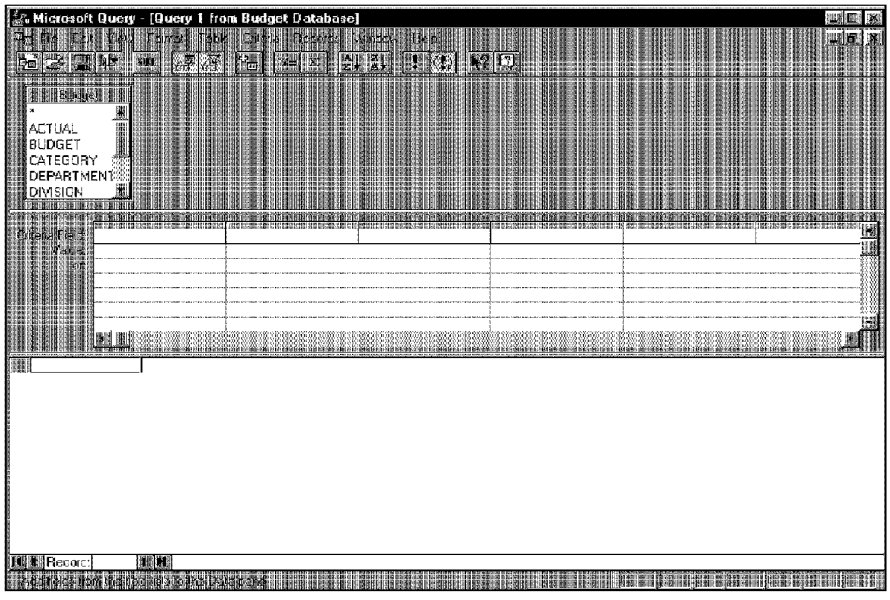


Figure 24-11: Microsoft Query, displaying the Criteria pane.

The Query window has three panes, which are split vertically:

- **Tables pane:** The top pane, which holds the data tables for the database. Each data table window has a list of the fields in the table.
- **Criteria pane:** The middle pane, which holds the criteria that determine the rows that the query returns.
- **Data pane:** The bottom pane, which holds the data that passes the criteria.

Creating a query consists of the following steps:

1. Drag fields from the Tables pane to the Data pane. You can drag as many fields as you want. These fields are the columns that the query will return. You can also double-click a field instead of dragging it.
2. Enter criteria in the Criteria pane. When you activate this pane, the first row (labeled Criteria Field) displays a drop-down list that contains all the field names. Select a field and enter the criteria below it. Query updates the Data pane automatically, treating each row like an OR operator.
3. Choose File • Return Data to Microsoft Excel to execute the query and place the data in a worksheet or pivot table.

Figure 24-12 shows how the query for the example presented earlier in this chapter appears in Query.

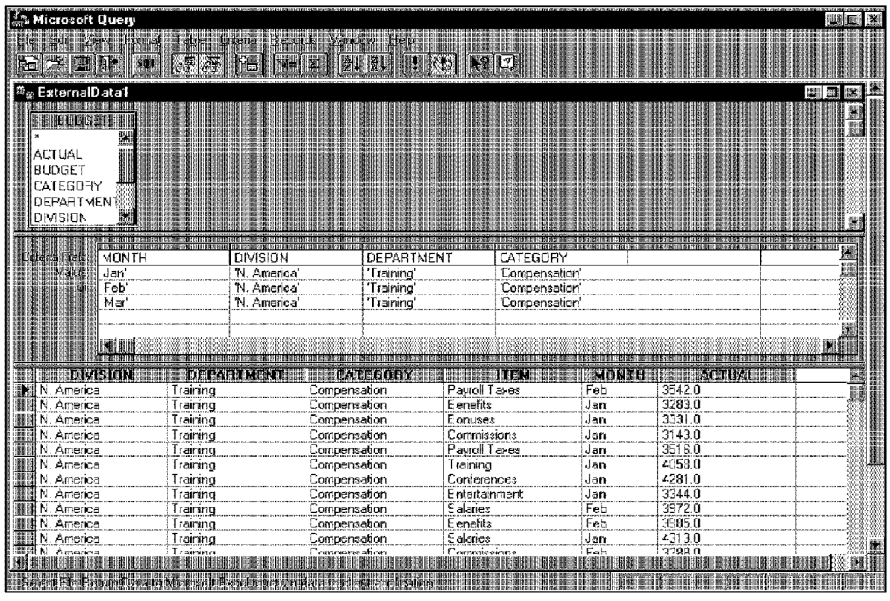


Figure 24-12: The center pane contains a query definition.

Running Microsoft Query by Itself

Normally, you run Query from Excel. But because Query is a standalone application, you also can run it directly. The executable file is named msqry32.exe and its location can vary (use the Windows Find File feature to locate this program on your system).

If you run Query by itself, you can't return the data to Excel automatically. You can, however, use the Clipboard to copy data from the data pane to any application that you want (including Excel).

Using Multiple Database Tables

The example in this chapter uses only one database table. Some databases, however, use multiple tables. These databases are known as *relational databases*, because a common field links the tables. Query lets you use any number of tables in your queries. To see an example of a relational database, load the sample database (called Northwind Traders) that's provided with Microsoft Query. This particular database has six tables.

Adding and Editing Records in External Database Tables

To add, delete, and edit data when you are using Query, make sure that a check appears next to the Records • Allow Editing command. Of course, you can't edit a database file that's set up as read-only. In any case, you need to be careful with this feature, because your changes are saved to disk as soon as you move the cell pointer out of the record that you're editing (you do not need to choose File • Save).

Formatting Data

If you don't like the data's appearance in the data pane, you can change the font used, by selecting Format • Font. Be aware that selective formatting isn't allowed (unlike in Excel); changing the font affects all the data in the data pane.

Sorting Data

If you need to view the data in the data pane in a different order, choose Records • Sort (or click the Sort Ascending or Sort Descending toolbar icon).

Learning More

This chapter isn't intended to cover every aspect of Microsoft Query. Rather, it discusses the basic features that are used most often. In fact, if you use Query Wizard, you may never need to interact with Query itself. But, if you do need to use Query, you can experiment and consult the online Help to learn more. As with anything related to Excel, the best way to master Query is to use it — preferably with data that's meaningful to you.

Summary

This chapter introduces Microsoft Query — a standalone application that can be executed by Excel. Use Query to retrieve data from external database files. You can specify the criteria, and Query returns the data to your Excel worksheet.

CHAPTER 25

Analyzing Data with Pivot Tables

In This Chapter

Excel provides many data analysis tools, but the pivot table feature may be the most useful overall. Pivot tables are valuable for summarizing information that is contained in a database, which can be stored in a worksheet or in an external file.

This chapter demonstrates this innovative feature and suggests how you can use it to view your data in ways that you may not have imagined.

What Is a Pivot Table?

A *pivot table* provides a dynamic summary of data that is contained in a database or list. A pivot table enables you to create frequency distributions and cross-tabulations of several different data dimensions. In addition, you can display subtotals and any level of detail that you want. But, as explained later in this chapter, a pivot table *isn't* appropriate for all databases.

The best way to understand the concept of a pivot table is to see one. Start with Figure 25-1, which shows the data that is being used to create the pivot table in this chapter.

Banking.xls

	A	B	C	D	E	F
1	Date	Amount	AcctType	OpenedBy	Branch	Customer
2	09/01/97	340	Checking	New Accts	Central	Existing
3	09/01/97	15,759	CD	Teller	Westside	Existing
4	09/01/97	15,276	CD	New Accts	North County	Existing
5	09/01/97	12,000	CD	New Accts	Westside	Existing
6	09/01/97	5,000	CD	New Accts	North County	Existing
7	09/01/97	7,000	Savings	New Accts	North County	New
8	09/01/97	90,000	CD	New Accts	Central	Existing
9	09/01/97	124	Checking	Teller	Central	Existing
10	09/01/97	400	Checking	Teller	Central	Existing
11	09/01/97	100	Checking	New Accts	Central	Existing
12	09/01/97	14,644	CD	New Accts	Westside	New
13	09/01/97	5,000	Savings	New Accts	Westside	Existing
14	09/01/97	4,623	Savings	New Accts	North County	Existing
15	09/01/97	5,879	Checking	New Accts	Central	Existing
16	09/01/97	3,171	Checking	New Accts	Westside	Existing
17	09/01/97	4,000	Savings	New Accts	Central	Existing
18	09/01/97	5,000	Checking	New Accts	Central	Existing
19	09/01/97	16,000	CD	New Accts	Central	New
20	09/01/97	50,000	Savings	New Accts	Central	Existing
21	09/01/97	13,636	CD	New Accts	North County	Existing

September

Figure 25-1: This database is used to create a pivot table.

This database consists of daily new-account information for a three-branch bank. The database contains 350 records and tracks:

- The date that each account was opened
- The opening amount
- The account type (CD, checking, savings, or IRA)
- Who opened the account (a teller or a new-account representative)
- The branch at which it was opened
- Whether a new customer or an existing customer opened the account

You can find this workbook on this book's CD-ROM; it is used in many examples throughout the chapter.

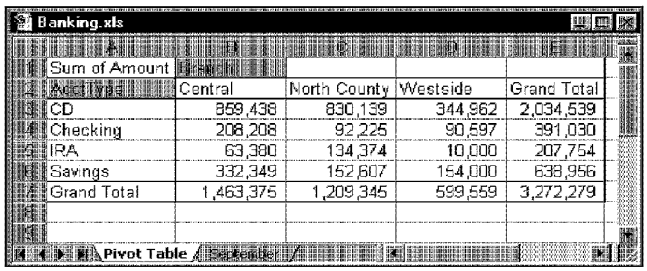
The bank database contains a lot of information, but it's not all that revealing, because, in its present form, the information is difficult to understand. If the data were summarized, it would be more useful. Summarizing a database is essentially the process of answering questions about the data. Here are a few questions that may be of interest to the bank's management:

- What is the total deposit amount for each branch, broken down by account type?
- How many accounts were opened at each branch, broken down by account type?
- What's the dollar distribution of the different account types?

- What types of accounts do tellers most often open?
- How is the Central branch doing compared to the other two branches?
- Which branch opens the most accounts for new customers?

You can use a pivot table to answer questions like these. It takes only a few seconds and doesn't require a single formula.

Figure 25-2 shows a pivot table created from the database that is displayed in Figure 25-1. This pivot table shows the amount of new deposits, broken down by branch and account type. This summary is one of hundreds that you can produce from this data.



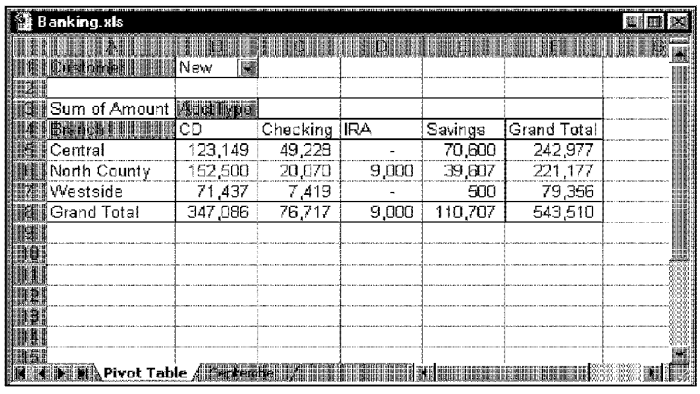
Banking.xls

Sum of Amount	AcctType			
Branch	Central	North County	Westside	Grand Total
CD	859,438	830,139	344,962	2,034,539
Checking	208,208	92,225	90,697	391,030
IRA	63,380	134,374	10,000	207,754
Savings	332,349	152,607	154,000	638,956
Grand Total	1,463,375	1,209,345	599,659	3,272,279

Pivot Table

Figure 25-2: A simple pivot table.

Figure 25-3 shows another pivot table that is generated from the bank data. This pivot table uses a page field for the Customer item. In this case, the pivot table displays the data only for new customers. Notice that the orientation of the table is changed. (Branches appear in rows and AcctType appears in columns.)



Banking.xls

Customer	New				
Sum of Amount	AcctType				
Branch	CD	Checking	IRA	Savings	Grand Total
Central	123,149	49,228	-	70,600	242,977
North County	152,500	20,070	9,000	39,607	221,177
Westside	71,437	7,419	-	500	79,356
Grand Total	347,086	76,717	9,000	110,707	543,510

Pivot Table

Figure 25-3: A pivot table that uses a page field.

Data Appropriate for a Pivot Table

Before getting into the details of pivot tables, you need to understand the type of data that's relevant to this feature. The data that you're summarizing must be in the form of a database (although an exception to this does exist, which is discussed later in the chapter). You can store the database in either a worksheet (such a database is sometimes known as a table) or an external database file. Although Excel can convert any database to a pivot table, not all databases benefit.

Generally speaking, fields in a database table can be one of two types:

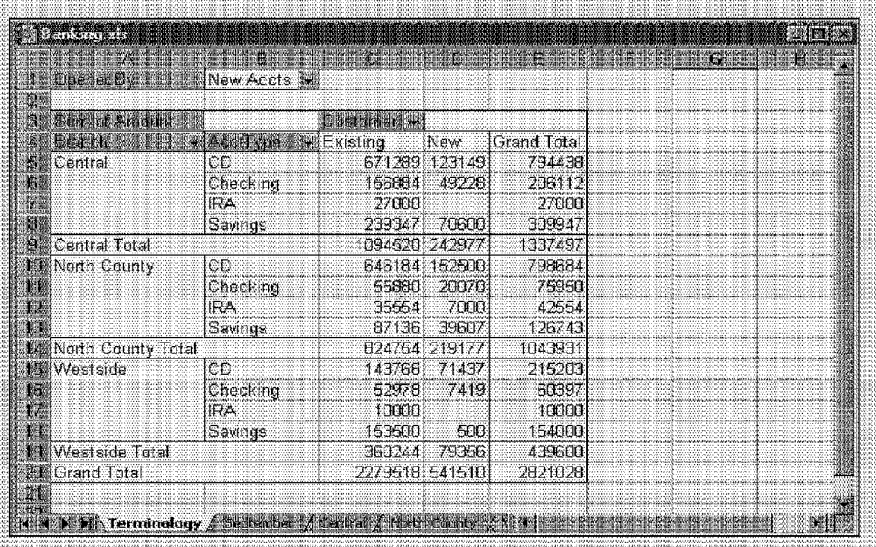
- **Data:** Contains a value. In Figure 25-1, the Amount field is a data field.
- **Category:** Describes the data. In Figure 25-1, the Date, AcctType, OpenedBy, Branch, and Customer fields are category fields, because they describe the data in the Amount field.

Pivot Table Terminology

If you're new to Excel, the concept of a pivot table may be a bit baffling. As far as I know, Microsoft invented the name *pivot table*. Understanding the terminology associated with pivot tables is important. Refer to the accompanying figure to get your bearings.

- **Column field:** A field that has a column orientation in the pivot table. Each item in the field occupies a column. In the figure, Customer is a column field, and it has two items (Existing and New). Column fields can be nested.
- **Data area:** The cells in a pivot table that contain the summary data. Excel offers several ways to summarize the data (sum, average, count, and so on). In the figure, the Data area includes C5:E20.
- **Grand totals:** A row or column that displays totals for all cells in a row or column in a pivot table. You can specify that grand totals be calculated for rows, columns, or both (or neither). The pivot table in the figure has grand totals for rows and columns.
- **Group:** A collection of items that are treated as a single item. You can group items manually or automatically (group dates into months, for example).
- **Item:** An element in a field that appears as a row or column header in a pivot table. In the figure, Existing and New are items for the Customer field. The Branch field has three items: Central, North County, and Westside. AcctType has four items: CD, Checking, IRA, and Savings.
- **Page field:** A field that has a page orientation in the pivot table—similar to a slice of a three-dimensional cube. Only one item in a page field can be displayed at one time. In the figure, OpenedBy is a page field that's displaying the NewAccts item; the pivot table shows data only for NewAccts.
- **Refresh:** To recalculate the pivot table after changes to the source data have been made.

- **Row field:** A field that has a row orientation in the pivot table. Each item in the field occupies a row. Row fields can be nested. In the figure, Branch and AcctType are both row fields.
- **Source data:** The data used to create a pivot table. It can reside in a worksheet or an external database.
- **Subtotals:** A row or column that displays subtotals for detail cells in a row or column in a pivot table.



Branch	AcctType	Existing	New	Grand Total
Central	CD	671289	123149	794438
	Checking	156884	49228	206112
	IRA	27000		27000
	Savings	239347	70600	309947
Central Total		1094520	242977	1337497
North County	CD	643184	152500	795684
	Checking	55880	20070	75950
	IRA	35554	7000	42554
	Savings	87136	39607	126743
North County Total		824754	219177	1043931
Westside	CD	143766	71437	215203
	Checking	52978	7419	60397
	IRA	10000		10000
	Savings	153500	500	154000
Westside Total		360244	79356	439600
Grand Total		2279518	541510	2821028

A single database table can have any number of data fields and category fields. When you create a pivot table, you usually want to summarize one or more of the data fields. The values in the category fields, on the other hand, appear in the pivot table as rows, columns, or pages.

Exceptions exist, however, and you may find that Excel's pivot table feature is useful even for databases that don't contain actual numerical data fields. The database in Figure 25-4, for example, doesn't contain numerical data fields, but you can create a useful pivot table that counts fields rather than sums them.

You can summarize information in pivot tables by using methods other than summing. For example, the pivot table that you see in Figure 25-5 cross-tabulates the Month Born field by the Sex field, and the intersecting cells show the count for each combination of city and sex.

Name List.xls

	A	B	C
1	Employee	Month Born	Sex
2	Miller	September	Female
3	Santos	February	Female
4	Alios	June	Male
5	Chan	December	Female
6	Henderson	March	Male
7	Klinger	July	Female
8	Rosarita	June	Male
9	Fuller	February	Male
10	Wilson	January	Female
11	Quigley	July	Male
12	Ross-Jacobs	April	Male
13	Ocarina	August	Female
14	Yulanderpol	November	Female
15	Franklin	June	Female

Sheet1

Figure 25-4: This database doesn't have any numerical fields, but you can use it to generate a pivot table.

Name List.xls

	A	B	C	D	E
1	Count of Employee	Sex			
2	Month Born	Female	Male	Grand Total	
3	January	2	2	4	
4	February	2	2	4	
5	March	0	5	5	
6	April	1	2	3	
7	May	2	0	2	
8	June	3	3	6	
9	July	3	2	5	
10	August	3	2	5	
11	September	4	0	4	
12	October	2	2	4	
13	November	3	1	4	
14	December	2	2	4	
15	Grand Total	27	23	50	

Sheet2 Sheet1

Figure 25-5: This pivot table summarizes non-numeric fields by displaying a count rather than a sum.

Creating a Pivot Table

This section walks you through the steps to create a pivot table by using the PivotTable and PivotChart Wizard. You access the PivotTable and PivotChart Wizard by choosing Data • PivotTable and PivotChart Report.

This section uses the banking account workbook, which is available on this book's CD-ROM.

Identifying Where the Data Is Located

When you choose Data • PivotTable and PivotChart Report, the first of several dialog boxes appears (see Figure 25-6).

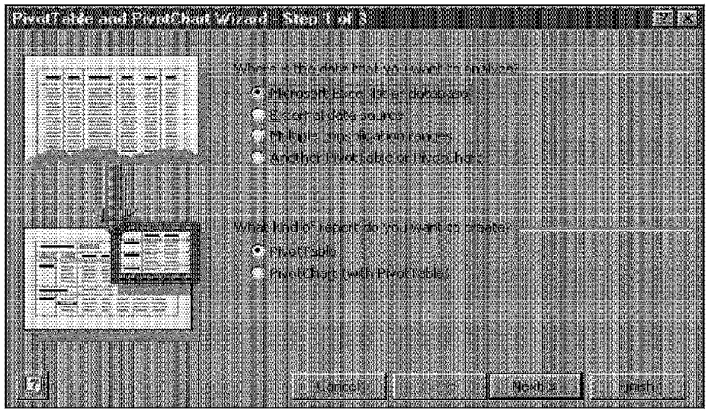


Figure 25-6: The first of three PivotTable and PivotChart Wizard dialog boxes.

In this step, you identify the data source. The possible data sources are described in the following sections.

You see different dialog boxes while you work through the Wizard, depending on the location of the data that you want to analyze. The following sections present the Wizard dialog boxes for data located in an Excel list or database, in the context of describing the various possible data sources.

Excel list or database

Usually, the data that you analyze is stored in a worksheet database—which is also known as a *list*. Databases stored in a worksheet are limited to 65,535 records and 256 fields. Working with a database of this size isn't efficient, however (and memory may not even allow it). The first row in the database should be field names. No other rules exist. The data can consist of values, text, or formulas.

External data source

If you use the data in an external database for a pivot table, the data is retrieved by using Query (a separate application). You can use dBASE files, SQL server data, or other data that your system is set up to access. You are prompted for the data source in Step 2 of the PivotTable and PivotChart Wizard.

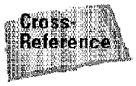
Pivot Tables and OLAP Cubes

In Chapter 24, the sidebar, "OLAP Databases" explains that OLAP (*online analytical processing*) presents a new way to organize large databases, to suit the way that you analyze and manage information. In an OLAP database, data is organized by level of detail, and the various aspects of data contained in an OLAP database are called *dimensions*. Because you combine dimensions to obtain information, OLAP databases are called *cubes*.

Generally, creating pivot tables in Excel from OLAP databases is faster than creating pivot tables from other types of external databases, because the OLAP server, not Excel, computes summarized values. Excel, therefore, receives less data from an OLAP cube when you create a pivot table or pivot chart.

Excel contains OLAP Cube Wizard, which helps you to organize data from external relational databases into OLAP cubes. Excel also contains Offline Cube Wizard, which enables you to create cube files that you can query even when you're not connected to your network. Distribute cube files over a network or on the Web to provide access to part, but not all, of a database.

You can create cube files only if you use an OLAP provider that supports creating cube files, such as Microsoft DSS Analysis server.



Chapter 24 discusses external database access, including Query. If you plan to create a pivot table by using data in an external database, you should consult Chapter 24 before proceeding.

Multiple consolidation ranges

You also can create a pivot table from multiple tables. This procedure is equivalent to consolidating the information in the tables. When you create a pivot table to consolidate information in tables, you have the added advantage of using all of the pivot table tools while you work with the consolidated data. (An example of this is presented later in the chapter.)

Chapter 19 discusses other consolidation techniques.

Another pivot table

Excel enables you to create a pivot table from an existing pivot table. Actually, this is a bit of a misnomer. The pivot table that you create is based on the *data* that the first pivot table uses (not the pivot table itself). If the active workbook has no pivot tables, this option is grayed, meaning you can't choose it.

If you need to create more than one pivot table from the same set of data, the procedure is more efficient (in terms of memory usage) if you create the first pivot table and then use that pivot table as the source for subsequent pivot tables.

Specifying the Data

To move on to the next step of the Wizard, click the Next button. Step 2 of the PivotTable and PivotChart Wizard prompts you for the data. Remember, the dialog box varies, depending on your choice in the first dialog box; Figure 25-7 shows the dialog box that appears when you select an Excel list or database in Step 1.

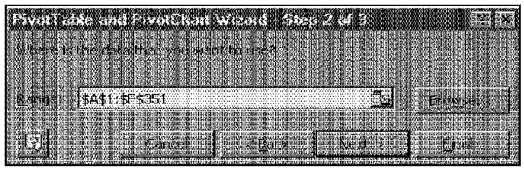


Figure 25-7: In Step 2, you specify the data range.

If you place the cell pointer anywhere within the worksheet database when you select Data • PivotTable Report, Excel identifies the database range automatically in Step 2 of the PivotTable and PivotChart Wizard.

You can use the Browse button to open a different worksheet and select a range. To move on to Step 3, click the Next button.

Completing the Pivot Table

The following sections outline how to complete the pivot table. The first step is to determine the pivot table's location. The dialog box for the final step of the PivotTable and PivotChart Wizard is shown in Figure 25-8. In this step, you specify the location for the pivot table.

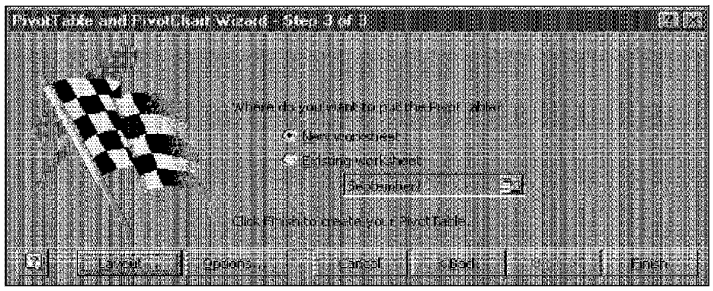


Figure 25-8: In Step 3, you specify the pivot table's location.

If you select the New worksheet option, Excel inserts a new worksheet for the pivot table. If you select the Existing worksheet option, the pivot table appears on the current worksheet (you can specify the starting cell location).

Pivot table options

You can click the Options button to select some options that determine how the table appears. Refer to the sidebar "Pivot Table Options," later in this chapter. Click OK to redisplay the PivotTable and PivotChart Wizard – Step 3 of 3 dialog box.

Setting up the layout of the pivot table

You can set up the layout of the pivot table in two different ways: either by using the PivotTable and PivotChart Wizard or by using the PivotTable toolbar directly on the worksheet.

Using a dialog box to lay out a pivot table

Click the Layout button of the last Wizard dialog box to see the dialog box shown in Figure 25-9. The fields in the database appear as buttons along the right side of the dialog box. Simply drag the buttons to the appropriate area of the pivot table diagram.

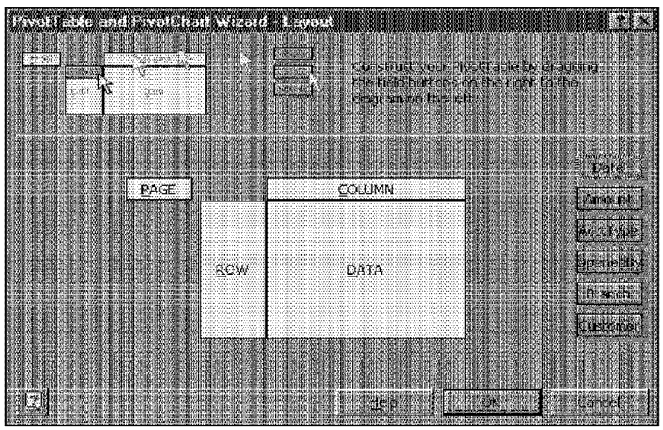


Figure 25-9: Specify the table layout.

The pivot table diagram has four areas

- **Page:** Values in the field appear as page items in the pivot table.
- **Row:** Values in the field appear as row items in the pivot table.
- **Data:** The field is summarized in the pivot table.
- **Column:** Values in the field appear as column items in the pivot table.

You can drag as many field buttons as you want to any of these locations, and you don't have to use all the fields. Any fields that you don't use simply don't appear in the pivot table.

When you drag a field button to the Data area, the PivotTable and PivotChart Wizard applies the Sum function if the field contains numeric values, and the Count function if the field contains non-numeric values.

While you're setting up the pivot table, you can double-click a field button to customize it. You can specify, for example, that a particular field be summarized as a Count or other function. You also can specify which items in a field to hide or omit. Be aware, however, that you can customize fields at any time after the pivot table is created; this is demonstrated later in this chapter.

If you drag a field button to an incorrect location, just drag it off the table diagram to get rid of it.

Figure 25-10 shows how the dialog box looks after some field buttons were dragged to the pivot table diagram. This pivot table displays the sum of the Amount field, broken down by AcctType (as rows) and Customer (as columns). In addition, the Branch field appears as a page field. Click OK to redisplay the PivotTable and PivotChart Wizard – Step 3 of 3 dialog box.

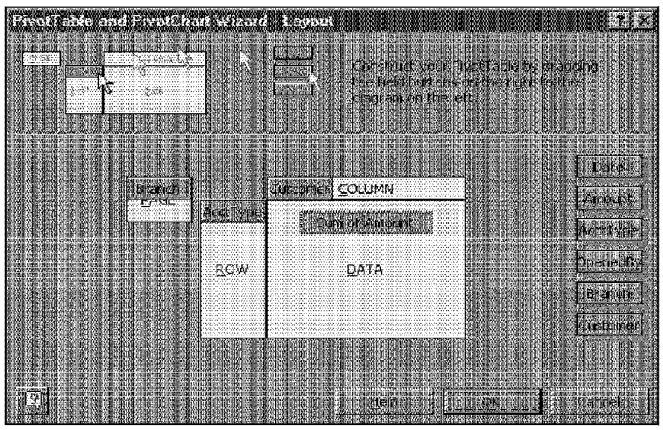


Figure 25-10: The table layout after dragging field buttons to the pivot table diagram.

Laying out a pivot table by using the PivotTable toolbar

Starting in Excel 2000, you can lay out a pivot table directly in a worksheet by using the PivotTable toolbar. The technique is very similar to the one just described, because you still drag and drop fields.

EXCEL 2000 The PivotTable toolbar is new to Excel 2000.

Complete the first two steps of the PivotTable and PivotChart Wizard. If you want, set options for the pivot table by using the Options button that appears in the third dialog box of the Wizard. Don't bother with the Layout button, however. Select a location for the pivot table and choose Finish. Excel displays a pivot table template similar to the one you see in Figure 25-11. The template provides you with hints about where to drop various types of fields.

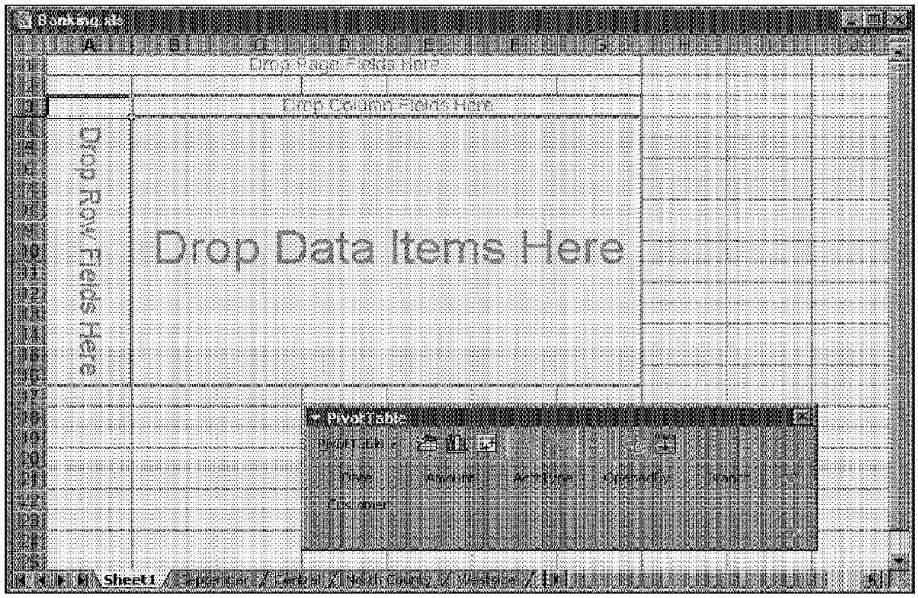


Figure 25-11: Use the PivotTable toolbar to drag and drop fields onto the pivot table template that Excel displays.

Drag and drop fields from the PivotTable toolbar onto the template. As you point at buttons on the toolbar, you'll see tool tips that instruct you to drag the field to the template. Excel continues to update the pivot table as you drag and drop fields; for this reason, you'll find this method easiest to use if you drag and drop data items last.

If you make a mistake, simply drag the field off the template and drop it anyplace on the worksheet — Excel removes it from the pivot table template. All fields remain on the PivotTable toolbar, even if you use them.

The finished product

When you click the Finish button in this last Wizard dialog box, Excel creates the PivotTable. Figure 25-12 shows the result of this example.

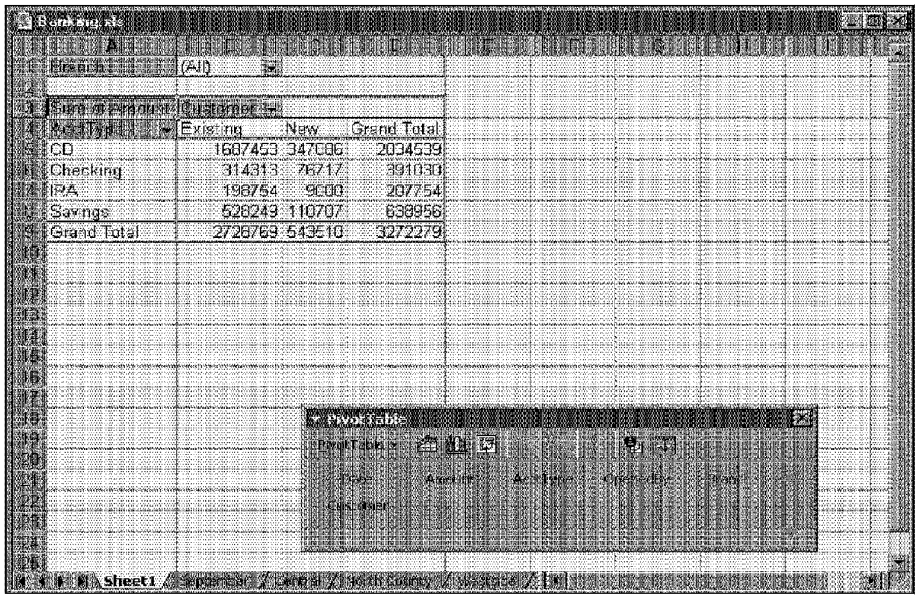
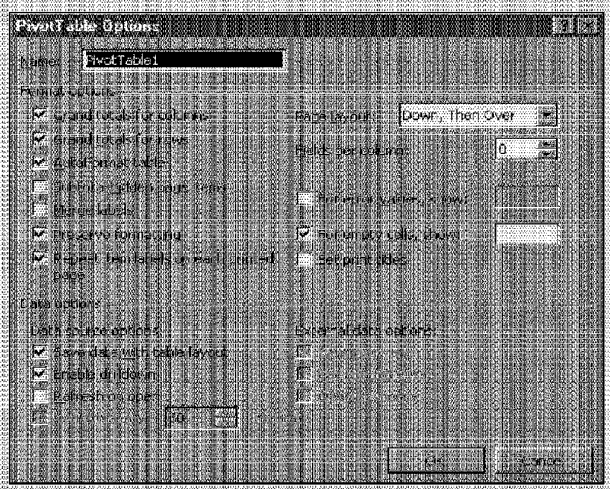


Figure 25-12: The pivot table that is created by the PivotTable and PivotChart Wizard.

Notice that the page field is displayed as a drop-down box. You can choose which item in the page field to display by choosing it from the list. You also can choose an item called All, which displays all the data.

Pivot Table Options

Excel provides plenty of options that determine how your pivot table looks and works. To access these options, click the Options button in the final step of the PivotTable and PivotChart Wizard. You can also access this dialog box after you create the pivot table. Right-click any cell in the pivot table and then select Options from the shortcut menu. The accompanying figure shows the PivotTable Options dialog box.



The PivotTable Options dialog box contains the following choices:

- **Name:** You can provide a name for the pivot table. Excel provides default names in the form of PivotTable1, PivotTable2, and so on.
- **Grand totals for columns:** Check this box if you want Excel to calculate grand totals for items that are displayed in columns.
- **Grand totals for rows:** Check this box if you want Excel to calculate grand totals for items that are displayed in rows.
- **AutoFormat table:** Check this box if you want Excel to apply one of its AutoFormats to the pivot table. Excel uses the AutoFormat even if you rearrange the table layout.
- **Subtotal hidden page items:** Check this box if you want Excel to include hidden items in the page fields in the subtotals.
- **Merge labels:** Check this box if you want Excel to merge the cells for outer row and column labels. Doing so may make the table more readable.
- **Preserve formatting:** Check this box if you want Excel, when it updates the pivot table, to keep any of the formatting that you applied.
- **Repeat item labels on each printed page:** Check this box to set row titles that appear on each page when you print a PivotTable report.

- **Page layout:** You can specify the order in which you want the page fields to appear.
- **Fields per column:** You can specify the number of page fields to show before starting another row of page fields.
- **For error values, show:** You can specify a value to show for pivot table cells that display an error.
- **For empty cells, show:** You can specify a value to show for pivot table cells that are empty.
- **Set print titles:** Check this box to set column titles that appear at the top of each page when you print a PivotTable report.
- **Save data with table layout:** If you check this option, Excel stores an additional copy of the data (called a *pivot table cache*), enabling Excel to recalculate the table more quickly when you change the layout. If memory is an issue, you should keep this option unchecked (updating is then a bit slower).
- **Enable drilldown:** If checked, you can double-click a cell in the pivot table to view details.
- **Refresh on open:** If checked, the pivot table is refreshed whenever you open the workbook.
- **Refresh every *x* minutes:** If you are connected to an external database, you can specify how often you want the pivot table refreshed while the workbook is open.
- **Save password:** If you use an external database that requires a password, you can store the password as part of the query, so that you don't have to reenter it.
- **Background query:** If checked, Excel runs the external database query in the background while you continue your work.
- **Optimize memory:** This option reduces the amount of memory that is used when you refresh an external database query.

Working with Pivot Tables

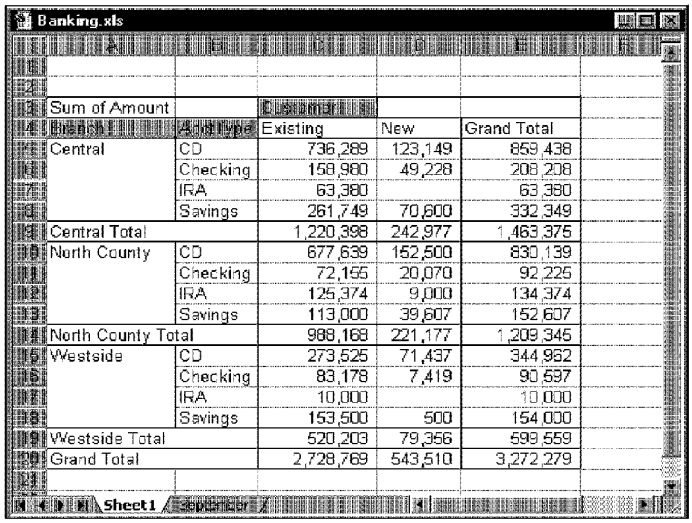
After you create a pivot table, it's not a static object. You can continue to modify and tweak it until it looks exactly how you want it to look. This section discusses modifications that you can make to a pivot table.

You'll find the PivotTable toolbar quite useful when you work with pivot tables. This toolbar appears automatically when you activate a worksheet that contains a pivot table.

Changing the Pivot Table Structure

Notice that a pivot table, when displayed in a worksheet, includes the field buttons. You can drag any of the field buttons to a new position in the pivot table (known as *pivoting*). For example, you can drag a column field to the row position. Excel immediately redisplays the pivot table to reflect your change. You also can change the order of the row fields or the column fields by dragging the buttons. This action affects how Excel nests the fields and can have a dramatic effect on the appearance of the table.

Figure 25-13 shows the pivot table that was created in the preceding example, after making a modification to the table's structure. The page field button (Branch) has been dragged to the row position. The pivot table now shows details for each item in the AcctType field for each branch.



Banking.xls

Sum of Amount		Customer		
Branch	AcctType	Existing	New	Grand Total
Central	CD	736,289	123,149	859,438
	Checking	158,980	49,228	208,208
	IRA	63,380		63,380
	Savings	261,749	70,600	332,349
Central Total		1,220,398	242,977	1,463,375
North County	CD	677,639	152,500	830,139
	Checking	72,155	20,070	92,225
	IRA	125,374	9,000	134,374
	Savings	113,000	39,607	152,607
North County Total		988,168	221,177	1,209,345
Westside	CD	273,525	71,437	344,962
	Checking	83,178	7,419	90,597
	IRA	10,000		10,000
	Savings	153,500	500	154,000
Westside Total		520,203	79,356	599,559
Grand Total		2,728,769	543,510	3,272,279

Sheet1 | September

Figure 25-13: This pivot table has two row fields.

Describing how to change the layout of a pivot table is more difficult than doing it. I suggest that you create a pivot table and experiment by dragging around field buttons to see what happens.

A pivot table is a special type of range, and (with a few exceptions) you can't make any changes to it. For example, you can't insert or delete rows, edit results, or move cells. If you attempt to do so, Excel displays an appropriate error message.

Removing a Field

To remove a field from a pivot table, click the field button and drag it away from the pivot table. The mouse pointer changes to include a button with an X across it. Release the mouse button, and Excel updates the table to exclude the field.

Adding a New Field

To add a new field to the pivot table, select any field in the pivot table. Then, drag the field that you want to add from the PivotTable toolbar onto the pivot table. Excel updates the pivot table with the new field.

You also can add fields from the PivotTable and PivotChart Wizard; choose Data • PivotTable and PivotChart Report to start the Wizard.

Refreshing a Pivot Table

Notice that pivot tables don't contain formulas. Rather, Excel recalculates the pivot table every time that you make a change to it. If the source database is large, some delay may occur while this recalculation takes place, but for small databases, the update is virtually instantaneous.

In some cases, you may change the source data. When this happens, Excel doesn't update the pivot table automatically. Rather, you must refresh it manually. To refresh a pivot table, you can use any of the following methods:

- Choose Data • Refresh Data
- Right-click anywhere in the pivot table and select Refresh Data from the shortcut menu
- Click the Refresh Data tool on the PivotTable toolbar

Customizing a Pivot Table Field

Several options are available for fields within a pivot table. To access these options, simply double-click a field button (or right-click and select Field from the shortcut menu). Excel displays a PivotTable Field dialog box, like the one shown in Figure 25-14.

You can modify any of the following items:

- **Name:** Changes the name that is displayed on the field button. You can also make this change directly by editing the cell that holds the field button.

- **Orientation:** Changes how the field's items are displayed. You can also take the more direct approach of dragging the field button to another location, as described previously.
- **Subtotals:** Lets you change the type of subtotaling that is displayed. Subtotaling is relevant only if you have more than one field displayed as rows or columns. You can make a multiple selection in the list box, which results in more than one line of subtotals. To eliminate subtotals, click the None option.
- **Hide items:** Enables you to hide (not display) one or more items from a field. Click the specific item names that you want to hide.

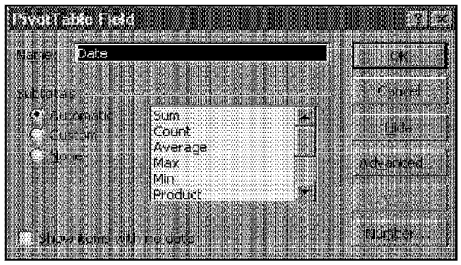


Figure 25-14: Double-clicking a PivotTable field button displays a dialog box like this one.

Excel includes some additional field options that you can specify by clicking the Advanced button in the PivotTable Field dialog box. These options let you specify how the field items are sorted and how many items to show (for example, just the top ten).

Formatting a Pivot Table

When you create a pivot table, Excel, by default, applies an AutoFormat to the table (you can change this by clicking the Options button in Step 3). After Excel creates the pivot table, you can always specify a different AutoFormat. Place the cell pointer in the pivot table and click the Format Report tool on the PivotTable toolbar. Excel displays the AutoFormat dialog box. Select an AutoFormat and click OK.

To change the number format for the pivot table data, use the following procedure:

1. Select any cell in the pivot table's Data area.
2. Right-click and choose PivotTable Field Settings from the shortcut menu. Excel displays its PivotTable Field dialog box.
3. Click the Number button.
4. Select the number format that you want to use.

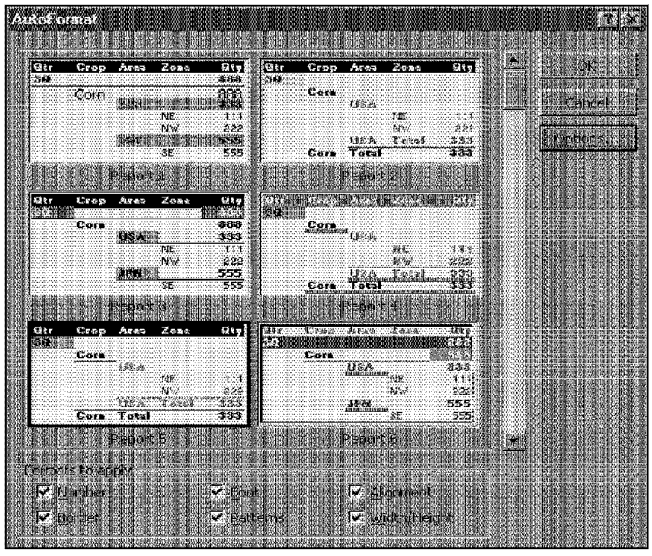


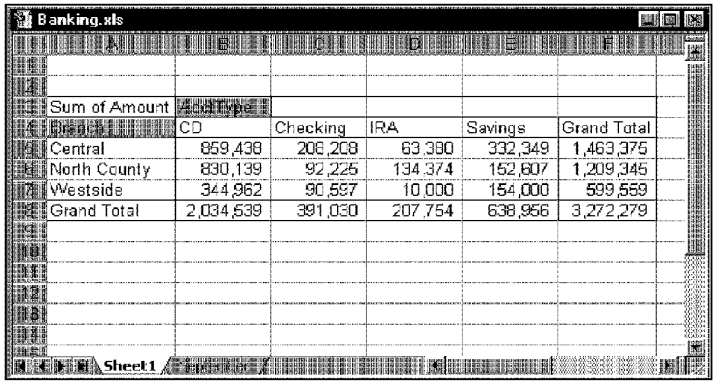
Figure 25-15: Change the formatting of a pivot table.

If you want Excel to preserve all the formatting that you perform on individual cells, make sure that the Preserve formatting option is turned on. You do this in the PivotTable Options dialog box (right-click a cell and select Table Options from the shortcut menu). If this option is not turned on, Excel returns the formats to the default formats when you refresh the pivot table.

Grouping Pivot Table Items

Grouping pivot table items is a handy feature that enables you to group specific items in a field. If one of the fields in your database consists of dates, for example, the pivot table displays a separate row or column for every date. You may find that grouping the dates into months or quarters and then hiding the details is more useful. Fortunately, this is easy to do.

Figure 25-16 shows a pivot table that was created using the bank database. It shows total balances for each AcctType (column field) by the Branch (row field). To create a report that compares the Central branch to the other two branches combined, create a group that consists of the Westside and North County branches.



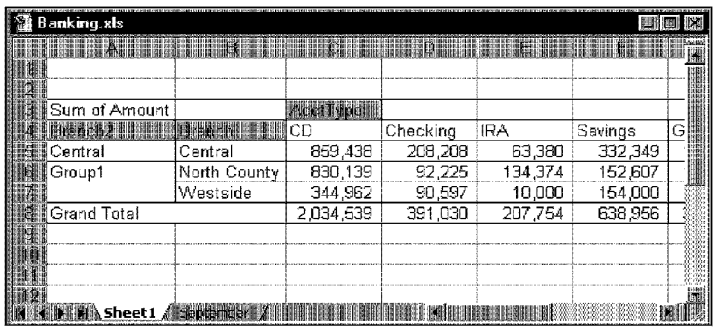
Banking.xls

Sum of Amount	AcctType				
Branch	CD	Checking	IRA	Savings	Grand Total
Central	859,438	208,208	63,380	332,349	1,463,375
North County	830,139	92,225	134,374	152,607	1,209,345
Westside	344,962	90,597	10,000	154,000	599,559
Grand Total	2,034,539	391,030	207,754	638,956	3,272,279

Sheet1

Figure 25-16: This version of the pivot table shows balances for each account type, by branch.

To create the group, select the cells that you want to group — in this case, A6:A7. Then, choose Data • Group and Outline • GroupPivotTable. Excel creates a new field called Branch2, which has two items: Central and Group1 (see Figure 25-17). At this point, you can remove the original Branch field (drag away the field button) and change the names of the field and the items. Figure 25-18 shows the pivot table after making these modifications.



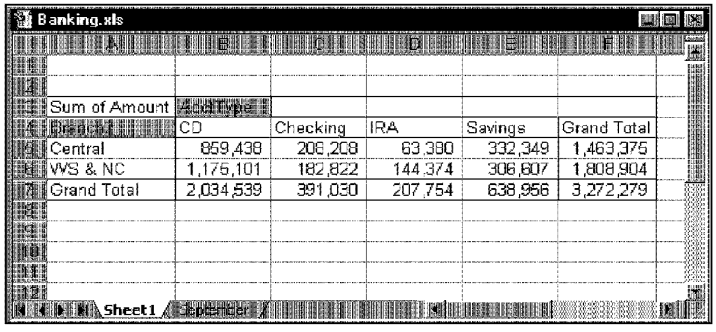
Banking.xls

Sum of Amount		AcctType				
Branch2	Branch	CD	Checking	IRA	Savings	G
Central	Central	859,438	208,208	63,380	332,349	
Group1	North County	830,139	92,225	134,374	152,607	
	Westside	344,962	90,597	10,000	154,000	
Grand Total		2,034,539	391,030	207,754	638,956	

Sheet1

Figure 25-17: The pivot table after grouping the North County and Westside branches.

The new field name can't be an existing field name. If it is, Excel adds the field to the pivot table. In this example, you can't rename Branch2 to Branch.



Banking.xls

Sum of Amount	AcctType				
Branch	CD	Checking	IRA	Savings	Grand Total
Central	859,438	208,208	63,380	332,349	1,463,375
WS & NC	1,175,101	182,822	144,374	306,807	1,808,904
Grand Total	2,034,539	391,030	207,754	638,956	3,272,279

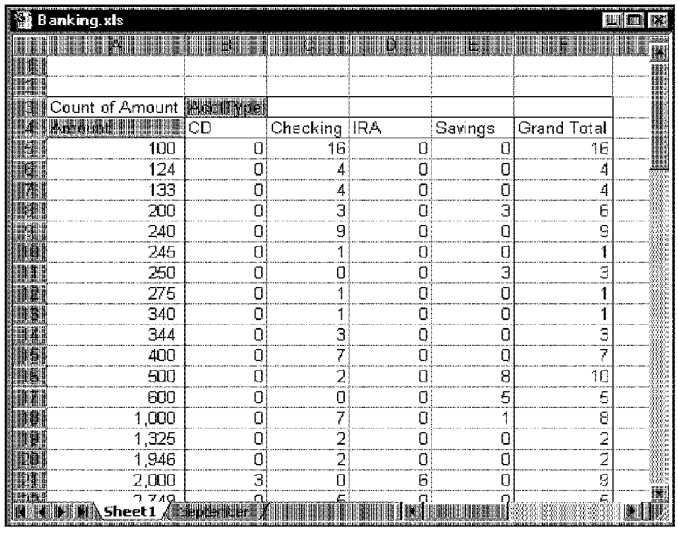
Sheet1 Spender

Figure 25-18: The pivot table after removing the original Branch field and renaming the new field and items.

Tip

If the items that you want to group are not adjacent to each other, you can make a multiple selection by pressing Ctrl and selecting the items that make up the group.

If the field items that you want to group consist of values, dates, or times, you can let Excel do the grouping for you. Figure 25-19 shows part of another pivot table that was generated from the bank database. This time, Amount is used for the row field and AcctType for the column field. The Data area shows the count for each combination. This report isn't useful, because the Amount field contains so many different items. The report can be salvaged, however, by grouping the items into bins.



Banking.xls

Count of Amount	AcctType				
Amount	CD	Checking	IRA	Savings	Grand Total
100	0	16	0	0	16
124	0	4	0	0	4
133	0	4	0	0	4
200	0	3	0	3	6
240	0	9	0	0	9
245	0	1	0	0	1
250	0	0	0	3	3
275	0	1	0	0	1
340	0	1	0	0	1
344	0	3	0	0	3
400	0	7	0	0	7
500	0	2	0	8	10
600	0	0	0	5	5
1,000	0	7	0	1	8
1,325	0	2	0	0	2
1,946	0	2	0	0	2
2,000	3	0	6	0	9
2,740	0	5	0	0	5

Sheet1 Spender

Figure 25-19: This isn't a useful pivot table report, because the Amount field contains too many different items.

To create groups automatically, select any item in the Amount field. Then, choose Data • Group and Outline • Group. Excel displays the Grouping dialog box, shown in Figure 25-20. By default, it shows the smallest and largest values — but you can change these to whatever you want. To create groups of $5,000 increments, enter **0** for the Starting at value, **100000** for the Ending at value, and **5000** for the By value (as shown in Figure 25-20). Click OK, and Excel creates the groups. Figure 25-21 shows the result, which is much more meaningful than the ungrouped data.

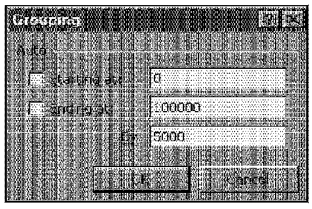


Figure 25-20: The Grouping dialog box instructs Excel to create groups automatically.

Banking.xls

Count of Amount	Account Type				
Amount	CD	Checking	IRA	Savings	Grand Total
0-4999	3	127	6	36	172
5000-9999	4	18	13	31	66
10000-14999	56	2	8	1	67
15000-19999	19	0	0	2	21
20000-24999	0	0	0	1	1
25000-29999	1	0	0	1	2
30000-34999	0	0	0	2	2
35000-39999	2	0	0	0	2
40000-44999	0	0	0	1	1
45000-49999	1	0	0	0	1
50000-54999	4	0	0	1	5
65000-69999	0	0	0	2	2
75000-79999	5	0	0	0	5
90000-94999	3	0	0	0	3
Grand Total	98	147	27	78	360

Sheet1

Figure 25-21: The pivot table after grouping the Amount field items.

Seeing the Details

Each cell in the Data area of a pivot table represents several records in the source database. You may be interested in seeing exactly which fields contribute to a summary value in the pivot table. Using the banking example, you may want to see a list

of the records that constitute the total CD accounts in the Central branch. To do so, double-click the appropriate summary cell in the Data area. Excel creates a new worksheet with the records that were used to create the summary. Figure 25-22 shows an example.

If double-clicking a cell doesn't work, make sure that the Enable drilldown option is turned on in the PivotTable Options dialog box (right-click a pivot table cell and select Options from the shortcut menu).

Displaying a Pivot Table on Different Sheets

If your pivot table displays a field in the Page position, you can see only one slice of the data at a time, by using the drop-down list box. Excel has an option, however, that puts each item from a page field on a separate sheet, creating a three-dimensional block of data. Click the PivotTable button on the PivotTable toolbar and choose Show Pages from the shortcut menu (or right-click the pivot table and select Show Pages from the shortcut menu). Excel displays the Show Pages dialog box, shown in Figure 25-23, which lists the page fields in your PivotTable. Select the fields that you want, and Excel inserts enough new sheets to accommodate each item in that field.

Banking.xls

	A	B	C	D	E	F	G
1	Date	Amount	AcctType	OpenedBy	Branch	Customer	
2	09/29/97	2000	CD	New Accts	Central	New	
3	09/29/97	11000	CD	New Accts	Central	New	
4	09/01/97	90000	CD	New Accts	Central	Existing	
5	09/29/97	14548	CD	New Accts	Central	Existing	
6	09/29/97	15000	CD	New Accts	Central	Existing	
7	09/29/97	17000	CD	Teller	Central	Existing	
8	09/29/97	90000	CD	New Accts	Central	Existing	
9	09/29/97	15208	CD	New Accts	Central	Existing	
10	09/01/97	16000	CD	New Accts	Central	New	
11	09/04/97	13000	CD	New Accts	Central	Existing	
12	09/04/97	13519	CD	New Accts	Central	New	
13	09/26/97	15208	CD	New Accts	Central	Existing	
14	09/26/97	13519	CD	New Accts	Central	Existing	
15	09/26/97	13000	CD	New Accts	Central	Existing	
16	09/25/97	15208	CD	New Accts	Central	Existing	
17	09/04/97	14548	CD	New Accts	Central	Existing	
18	09/22/97	2000	CD	Teller	Central	Existing	
19	09/04/97	11000	CD	New Accts	Central	New	
20	09/04/97	35000	CD	New Accts	Central	Existing	
21	09/22/97	13519	CD	New Accts	Central	Existing	

Sheet2

Figure 25-22: Double-clicking a cell in the Data area of a pivot table generates a new worksheet with the underlying data.

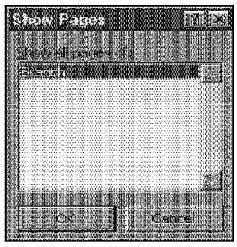


Figure 25-23: The Show Pages dialog box enables you to display each page field item on a separate worksheet.

Inserting a Calculated Field into a Pivot Table

As previously noted, a pivot table is a special type of data range, and you can't insert new rows or columns into a pivot table. This means that you can't insert formulas to perform calculations with the data in a pivot table. However, you can create calculated fields for a pivot table. A *calculated field* consists of a calculation that can involve other fields.

You cannot create a calculated field in a pivot table that is based on an OLAP database.

In the banking example, for instance, assume that management wants to increase deposits by 15 percent and wants to compare the projected deposits to the current deposits. In this situation, you can use a calculated field. Calculated fields must reside in the Data area of the pivot table (you can't use them in the Page, Row, or Column areas).

Use the following procedure to create a calculated field that consists of the Amount field multiplied by 1.15 (that is, a 15 percent increase):

1. Move the cell pointer anywhere within the pivot table.
2. Right-click and choose Formulas • Calculated Field from the shortcut menu. Excel displays the Insert Calculated Field dialog box, shown in Figure 25-24.
3. Enter a descriptive name for the field and specify the formula. The formula can use other fields, but can't use worksheet functions. For this example, the name is Projected, and the formula is the following:

   ```
   =Amount*1.15
   ```

4. Click Add to add this new field.
5. To create additional calculated fields, repeat Steps 3 and 4. Click OK to close the dialog box.

After you create the field, Excel adds it to the Data area of the pivot table. You can treat it just like any other field, with one exception: you can't move it to the Page, Row, or Column area (it must remain in the Data area). Figure 25-25 shows a pivot table with a calculated field (called Projected).

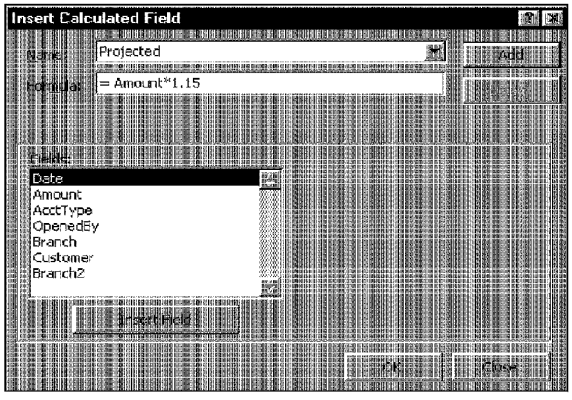


Figure 25-24: The Insert Calculated Field dialog box.

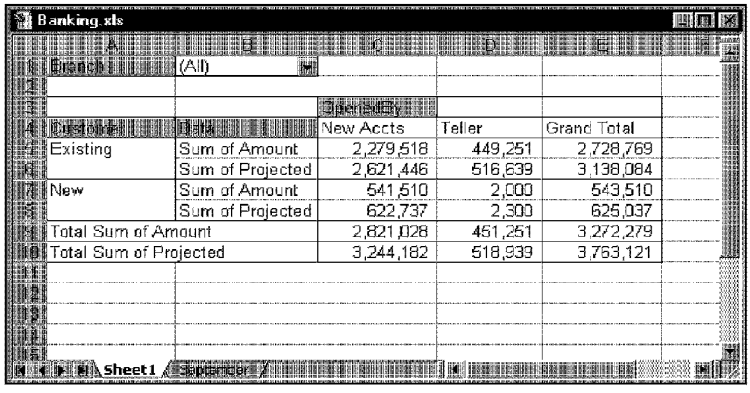


Figure 25-25: This pivot table uses a calculated field.

The formulas that you develop can also use worksheet functions, but the functions cannot refer to cells or named ranges.

Inserting a Calculated Item into a Pivot Table

The previous section explains how to create a calculated field. Excel also enables you to create *calculated items* for a pivot table field. For example, if you have a field named Months, you can create a calculated item (called Q1, for example) that displays the sum of January, February, and March. You can also do this by grouping the items — but using grouping hides the individual months and shows only the total of the group. Creating a calculated item for quarterly totals shows the total and the individual months. Calculated items must reside in the Page, Row, or Column area of a pivot table (you can't use calculated items in the Data area).

You can't create a calculated item in a pivot table based on an OLAP database.

In the banking example, management may want to look at CD accounts combined with savings accounts; you can show this information by creating a calculated item. To create a calculated item, use these steps:

1. Move the cell pointer to a Row, Column, or Page area of the pivot table. The cell pointer cannot be in the Data area.
2. Right-click and choose Formulas • Calculated Item from the shortcut menu. Excel displays the Insert Calculated Item dialog box, as shown in Figure 25-26.
3. Enter a name for the new item and specify the formula. The formula can use items in other fields, but can't use worksheet functions. For this example, the new item is named **CD & Savings**, and the formula is as follows:

   ```
   =CD + Savings
   ```

4. Click Add.
5. Repeat Steps 3 and 4 to create additional items. Click OK to close the dialog box.

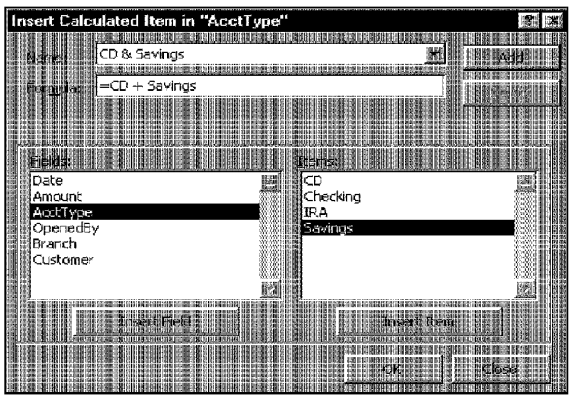


Figure 25-26: The Insert Calculated Item dialog box.

After you create the item, it appears in the pivot table. Figure 25-27 shows the pivot table after adding a calculated item.

If you use a calculated item in your pivot table, you may need to turn off the Grand Total display, to avoid double-counting.

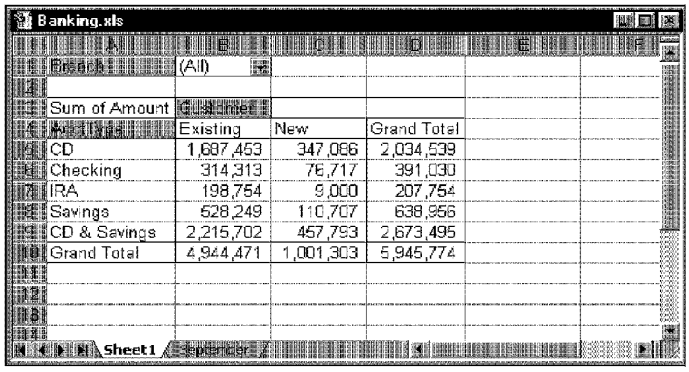


Figure 25-27: This pivot table uses a calculated item.

Pivot Table Examples

This section describes additional examples of pivot tables, to spark your creativity and help you apply some of these techniques to your own data.

The best way to master pivot tables is to work with them, using your own data—not just read about them.

If you want to work with some prefab pivot tables, I've developed a few for you to use, which you can find on this book's CD-ROM.

Using a Pivot Table to Consolidate Sheets

Chapter 19 discusses several ways to consolidate data across different worksheets or workbooks. Excel's pivot table feature gives you yet another consolidation option. Figure 25-28 shows three worksheets, each containing monthly sales data for a store (three different stores) in a music store chain. The goal is to consolidate this information into a single pivot table. In this example, all the source data is in a single workbook, but you can consolidate data from different workbooks.

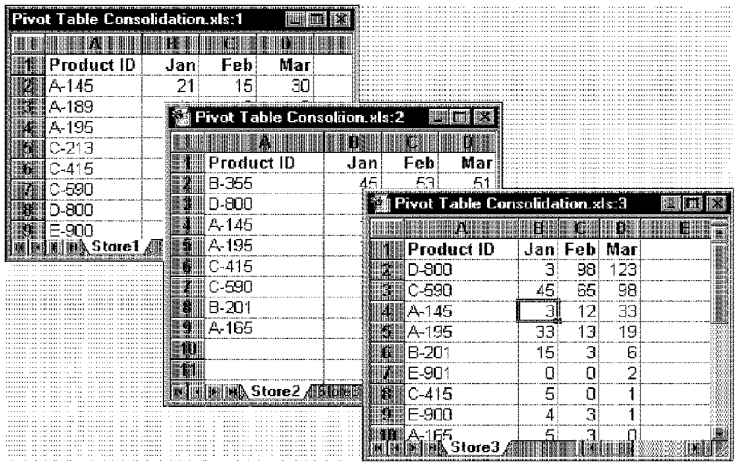


Figure 25-28: You can use a pivot table to consolidate these three worksheets.

You can find this workbook on this book's CD-ROM.

Use the following steps to create this pivot table:

1. Start with a new worksheet named **Summary.**
2. Choose Data • PivotTable and PivotChart Report, to display the PivotTable and PivotChart Wizard.
3. Select the Multiple Consolidation Ranges option and then click Next.
4. In Step 2a of the PivotTable Wizard, select the option labeled Create a single page field for me. Click Next.
5. In Step 2b, specify the ranges to be consolidated. The first range is **Store1!$A$1:$D$12** (you can enter this directly or point to it). Click Add to add this range to the All Ranges list.
6. Repeat this for the other two ranges (see Figure 25-29). Click Next to continue to Step 3.
7. The dialog box in Step 3 of the PivotTable Wizard should look familiar. Click Finish.

Figure 25-30 shows the pivot table. It uses generic names, which you can change to more meaningful names.

In Step 2a of the PivotTable Wizard, you can choose the option labeled I will create the page fields. Doing so enables you to provide an item name for each item in the page field (rather than the generic Item1, Item2, and Item3).

Figure 25-29: Step 2b of the PivotTable Wizard.

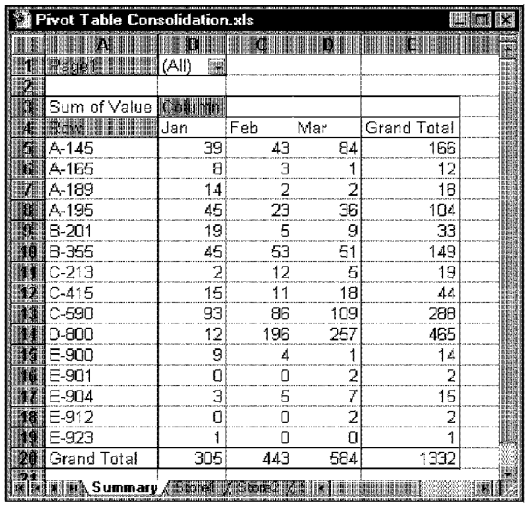


Figure 25-30: This pivot table uses data from three ranges.

Creating Charts from a Pivot Table

A Pivot chart report is a chart that is linked to a pivot table. By using the PivotTable and PivotChart Wizard, you can create simultaneously both a pivot table and a linked chart; you can use the techniques described earlier to drag and drop fields onto the pivot chart or the pivot table. To simultaneously create a pivot table and a pivot chart, choose PivotChart (with PivotTable) in the first dialog box of the PivotTable and PivotChart Wizard. Excel creates a new worksheet and a new chart sheet; both will contain templates for the pivot table and the pivot chart, respectively. Drag fields from the PivotTable toolbar onto either the chart or the table—

simply switch between the sheets in the workbook to choose the sheet with which you want to work.

EXCEL 2000 Pivot chart reports are a new feature of Excel 2000.

Although you can create a pivot chart by using the PivotTable and PivotChart Wizard, you'll find it easier to create the chart from an existing pivot table. While viewing the pivot table, click the Chart Wizard button on the PivotTable toolbar. Excel immediately creates a chart sheet in the workbook based on the pivot table. Figure 25-31 shows the pivot table used as the foundation for the pivot chart shown in Figure 25-32. Excel updates this chart whenever you make changes to the pivot table.

Tip A pivot chart is always created on a separate Chart sheet. To convert the chart to an embedded chart on a worksheet, activate the Chart sheet and select Chart • Location. Select the second option (as object in) and specify a worksheet for the chart.

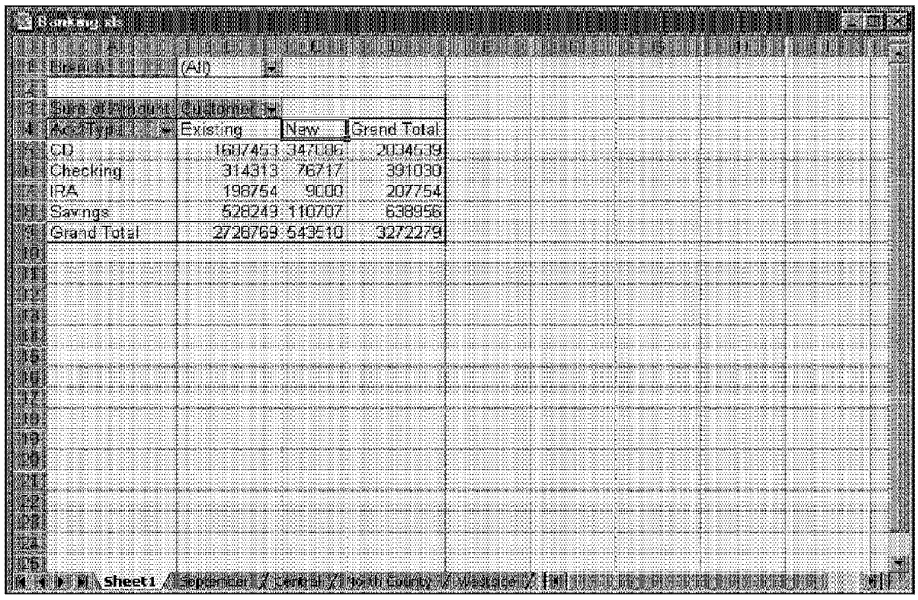


Figure 25-31: The pivot table from which the chart in Figure 25-32 was created.

Analyzing Survey Data

This example demonstrates how to use a pivot table to analyze survey data that was obtained via a questionnaire. Figure 25-33 shows part of the raw data that is typical of data collected from a survey questionnaire. Each record represents the responses for one respondent.

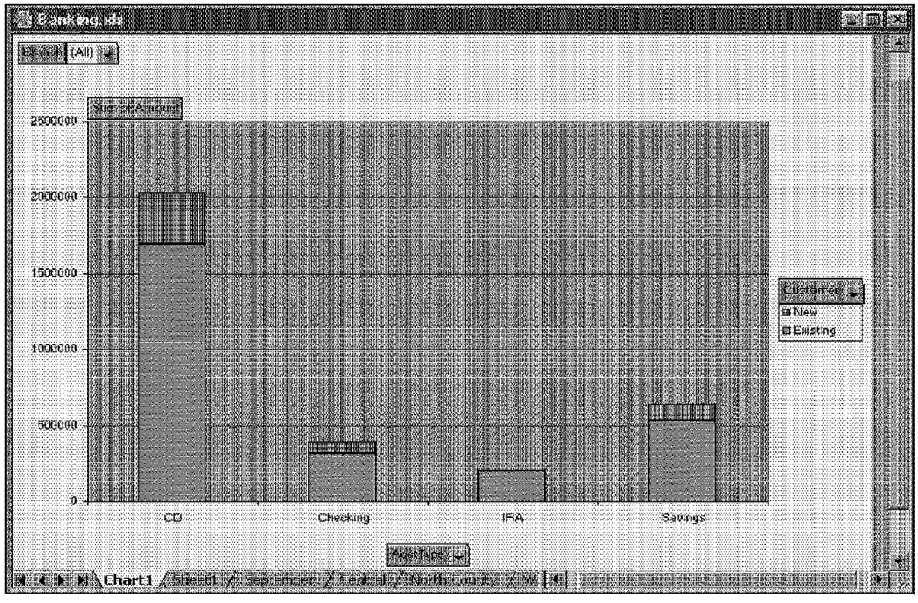


Figure 25-32: The chart changes based on the pivot table.

Pivot Table Survey Analysis.xls

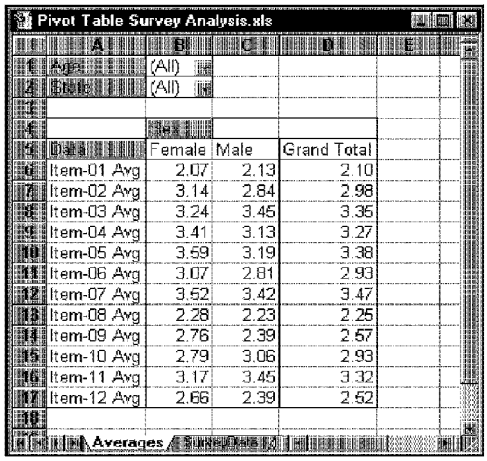
	A	B	C	D	E	F	G	H	I	J	
1	Name	Sex	Age	State	Item01	Item02	Item03	Item04	Item05	Item06	Iter
2	Subject1	Male	40	Illinois	1	4	4	4	1	1	
3	Subject2	Female	31	Illinois	2	5	1	1	4	2	
4	Subject3	Male	56	New York	1	1	4	2	3	3	
5	Subject4	Male	55	Illinois	2	1	3	5	1	2	
6	Subject5	Female	47	New York	2	2	5	5	4	2	
7	Subject6	Female	51	Illinois	2	4	3	3	1	1	
8	Subject7	Female	48	California	2	4	5	4	5	3	
9	Subject8	Male	39	New York	3	2	1	2	3	4	
10	Subject9	Female	37	California	3	4	4	4	5	1	
11	Subject10	Male	38	New York	2	1	5	5	5	1	
12	Subject11	Male	38	California	4	3	3	2	1	2	
13	Subject12	Female	46	California	2	1	4	5	5	5	
14	Subject13	Female	48	Illinois	4	3	4	3	2	5	
15	Subject14	Female	56	New York	2	3	4	2	1	1	

SurveyData

Figure 25-33: Use a pivot table to tabulate this survey data.

You can find the workbook used in this example on this book's CD-ROM.

Figure 25-34 shows a pivot table that was created to calculate averages for each of the 12 survey items, broken down by sex. Additional page fields enable you to examine the results easily by an age group or a particular state. Or, for a more complex pivot table, you can drag one or both of the page fields to a row or column position.



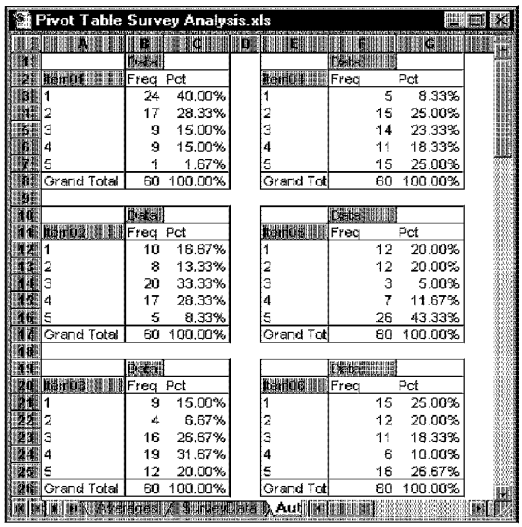
Pivot Table Survey Analysis.xls

	A	B	C	D	E
1	Age	(All)			
2	Edu	(All)			
3					
4		Sex			
5	Data	Female	Male	Grand Total	
6	Item-01 Avg	2.07	2.13	2.10	
7	Item-02 Avg	3.14	2.84	2.98	
8	Item-03 Avg	3.24	3.45	3.35	
9	Item-04 Avg	3.41	3.13	3.27	
10	Item-05 Avg	3.59	3.19	3.38	
11	Item-06 Avg	3.07	2.81	2.93	
12	Item-07 Avg	3.52	3.42	3.47	
13	Item-08 Avg	2.28	2.23	2.25	
14	Item-09 Avg	2.76	2.39	2.57	
15	Item-10 Avg	2.79	3.06	2.93	
16	Item-11 Avg	3.17	3.45	3.32	
17	Item-12 Avg	2.66	2.39	2.52	
18					

Averages | Survey Data 2

Figure 25-34: This pivot table calculates averages for each item.

Figure 25-35 shows another sheet in the workbook. This sheet contains 12 separate pivot tables, one for each survey item. Each pivot table displays the frequency of responses and the percentage of responses. Although you could create each table manually, the workbook includes a macro that creates them all in just a few seconds.



Pivot Table Survey Analysis.xls

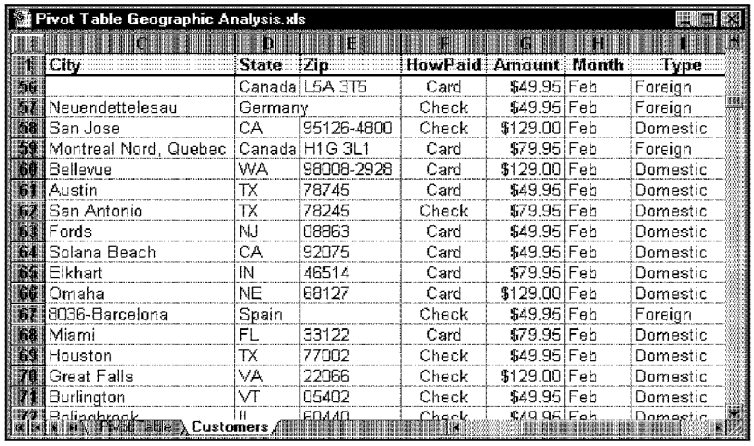
	A	B	C	D	E	F	G
1		Data				Data	
2	Item01	Freq	Pct		Item01	Freq	Pct
3	1	24	40.00%		1	5	8.33%
4	2	17	28.33%		2	15	25.00%
5	3	9	15.00%		3	14	23.33%
6	4	9	15.00%		4	11	18.33%
7	5	1	1.67%		5	15	25.00%
8	Grand Total	60	100.00%		Grand Tot	60	100.00%
9							
10		Data				Data	
11	Item02	Freq	Pct		Item02	Freq	Pct
12	1	10	16.67%		1	12	20.00%
13	2	8	13.33%		2	12	20.00%
14	3	20	33.33%		3	3	5.00%
15	4	17	28.33%		4	7	11.67%
16	5	5	8.33%		5	26	43.33%
17	Grand Total	60	100.00%		Grand Tot	60	100.00%
18							
19		Data				Data	
20	Item03	Freq	Pct		Item03	Freq	Pct
21	1	9	15.00%		1	15	25.00%
22	2	4	6.67%		2	12	20.00%
23	3	16	26.67%		3	11	18.33%
24	4	19	31.67%		4	6	10.00%
25	5	12	20.00%		5	16	26.67%
26	Grand Total	60	100.00%		Grand Tot	60	100.00%

Averages | Survey Data | Au

Figure 25-35: This sheet contains 12 pivot tables, created by a macro.

Customer Geographic Analysis

As a byproduct of creating a pivot table, you end up with a list of unique entries in a field. Figure 25-36 shows part of a database that tracks customers. The field of interest is the State field (which holds the country in the case of non-U.S. orders). The Type field contains a formula that returns either Foreign or Domestic, depending on the length of the entry in the State field. The goal of this example is to create a map that shows sales by state.



Pivot Table Geographic Analysis.xls

	C	D	E	F	G	H	I
1	City	State	Zip	HowPaid	Amount	Month	Type
56		Canada	L5A 3T5	Card	$49.95	Feb	Foreign
57	Neuendettelesau	Germany		Check	$49.95	Feb	Foreign
58	San Jose	CA	95126-4800	Check	$129.00	Feb	Domestic
59	Montreal Nord, Quebec	Canada	H1G 3L1	Card	$79.95	Feb	Foreign
60	Bellevue	WA	98008-2928	Card	$129.00	Feb	Domestic
61	Austin	TX	78745	Card	$49.95	Feb	Domestic
62	San Antonio	TX	78245	Check	$79.95	Feb	Domestic
63	Fords	NJ	08863	Card	$49.95	Feb	Domestic
64	Solana Beach	CA	92075	Card	$49.95	Feb	Domestic
65	Elkhart	IN	46514	Card	$79.95	Feb	Domestic
66	Omaha	NE	68127	Card	$129.00	Feb	Domestic
67	8036-Barcelona	Spain		Check	$49.95	Feb	Foreign
68	Miami	FL	33122	Card	$79.95	Feb	Domestic
69	Houston	TX	77002	Check	$49.95	Feb	Domestic
70	Great Falls	VA	22066	Check	$129.00	Feb	Domestic
71	Burlington	VT	05402	Check	$49.95	Feb	Domestic
72	Bolingbrook	IL	60440	Check	$49.95	Feb	Domestic

PT-Cities | Customers

Figure 25-36: This customer database would make a good map, but the data is not in the proper format.

You can find this workbook on the CD-ROM.

Figure 25-37 shows a pivot table created from the data displayed in Figure 25-36. It displays the data in terms of total amount, plus a count. Three page fields were used to filter the data.

Figure 25-38 shows the map that was created by using Excel's mapping feature (described fully in Chapter 17).

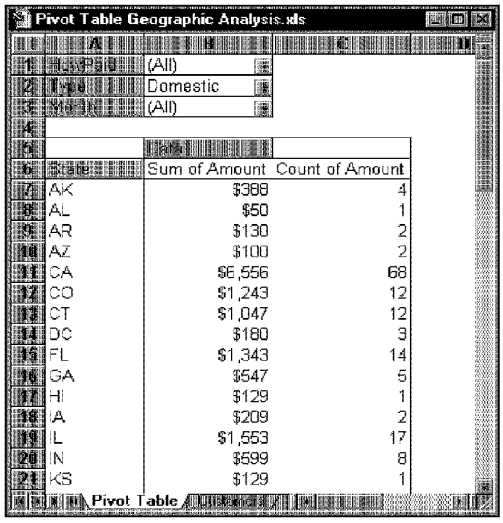


Figure 25-37: This pivot table contains perfect input for an Excel map.

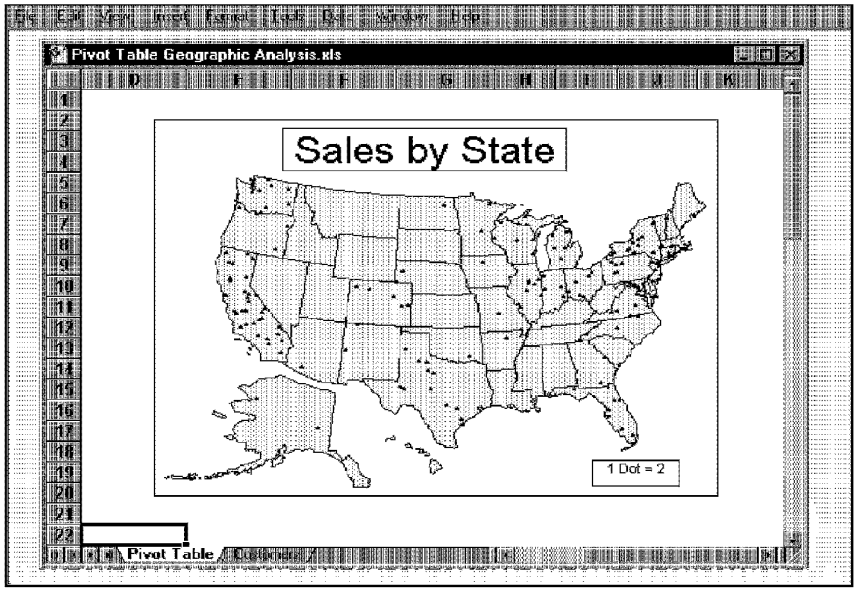


Figure 25-38: This map was created from the data in the pivot table.

Grouping by Month and Years

The final pivot table example (see Figure 25-39) demonstrates some techniques that involve grouping by dates. The worksheet contains daily pricing data for two years. I created a macro to change the grouping to days, weeks, months, quarters, or years. The macro also changes the range that is used in the chart.

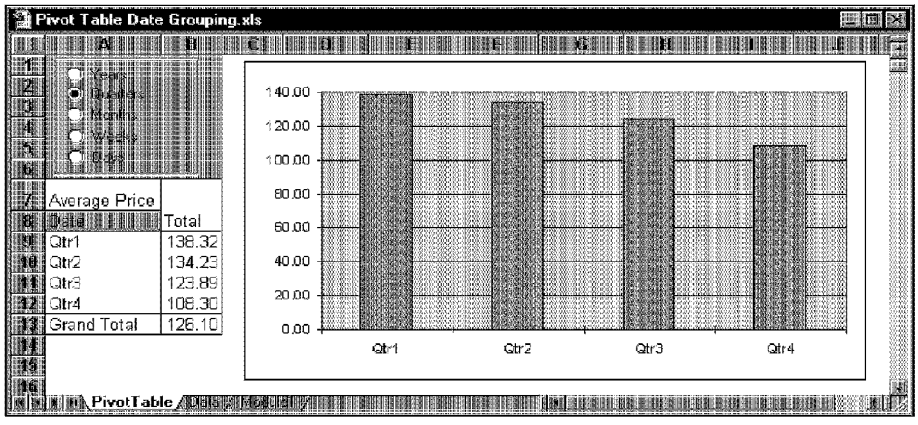


Figure 25-39: Clicking an option button executes a macro that changes the date grouping and updates the chart.

You can find this workbook on the CD-ROM.

Summary

This chapter discusses Excel's pivot table feature, which enables you to summarize data from a database that can be stored in a worksheet or in an external file. The examples in this chapter demonstrate some useful techniques. The best way to master this feature, however, is to use a database with which you're familiar and experiment until you understand how it works.

CHAPTER 26

Performing Spreadsheet What-If Analysis

In This Chapter

One of the most appealing aspects of a spreadsheet program—including Excel—is that you can use formulas to create dynamic models that instantly recalculate when you change values in cells to which the formulas refer. When you change values in cells in a systematic manner and observe the effects on specific formula cells, you're performing a type of *what-if* analysis. What-if analysis is the process of asking questions such as, "What if the interest rate on the loan changes to 8.5 rather than 9.0 percent?" or "What if we raise the prices of our products by 5 percent?"

If you set up your spreadsheet properly, answering such questions is a matter of plugging in new values and observing the results of the recalculation. Excel provides useful tools to assist you in your what-if endeavors.

A What-If Example

Figure 26-1 shows a spreadsheet that calculates information pertaining to a mortgage loan. The worksheet is divided into two sections: the input cells and the result cells. Column D shows the formulas stored in column C.

With this worksheet, you can easily answer the following what-if questions:

- What if I can negotiate a lower purchase price on the property?
- What if the lender requires a 20-percent down payment?
- What if I can get a 40-year mortgage?
- What if the interest rate decreases to 7.5 percent?

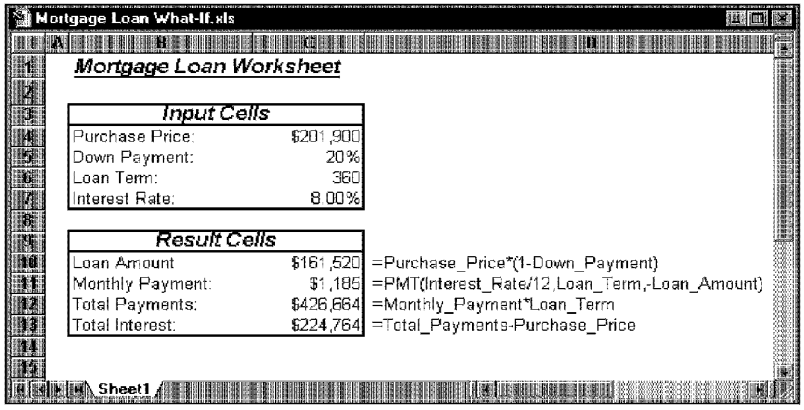


Figure 26-1: This worksheet model uses four input cells to produce the results in the formulas.

You can answer these questions simply by changing the values in the cells in range C4:C7 and observing the effects in the dependent cells (C10:C13). You can, of course, vary any number of input cells simultaneously.

Hard Code Values? No Way!

The mortgage calculation example, simple as it is, demonstrates an important point about spreadsheet design: You should always set up your worksheet so that you have maximum flexibility to make changes. Perhaps the most fundamental rule of spreadsheet design is the following:

Do not hard code (store) values in a formula. Rather, store the values in separate cells, and use cell references in the formula.

The term *hard code* refers to the use of actual values, or *constants*, in a formula. In the mortgage loan example, all the formulas use references to cells, not actual values.

You *could* use the value 360, for example, for the loan term argument of the PMT function in cell C11. Using a cell reference has two advantages. First, you have no doubt about the values that the formula uses (they aren't buried in the formula). Second, you can easily change the value.

Using values in formulas may not seem like much of an issue when only one formula is involved, but just imagine what would happen if this value were hard coded into several hundred formulas that were scattered throughout a worksheet.

Types of What-If Analyses

As you may expect, Excel can handle much more sophisticated models than the preceding example. To perform a what-if analysis using Excel, you have four basic options:

- **Manual what-if analysis:** Plug in new values and observe the effects on formula cells.
- **Macro-assisted what-if analysis:** Create macros to plug in variables for you.
- **Data tables:** Create a table that displays the results of selected formula cells as you systematically change one or two input cells.
- **Scenario Manager:** Create named scenarios and generate reports that use outlines or pivot tables.

Manual What-If Analysis

This method doesn't require too much explanation. In fact, the example that opens this chapter is a good one. It's based on the idea that you have one or more input cells that affect one or more key formula cells. You change the value in the input cells and see what happens to the formula cells. You may want to print the results or save each scenario to a new workbook. The term *scenario* refers to a specific set of values in one or more input cells.

This is how most people perform what-if analysis. Manual what-if analysis certainly has nothing wrong with it, but you should be aware of some other techniques.

Macro-Assisted What-If Analysis

A slightly more sophisticated form of manual what-if analysis uses macros. As is discussed in later chapters, a *macro* is a program that performs several operations automatically. Rather than change the input cells manually, you can create a macro to make the changes for you. For example, you may have three macros named BestCase, WorstCase, and MostLikelyCase. Running the BestCase macro enters the appropriate values into the input cells. Executing the WorstCase or MostLikelyCase macros enters other values.

If you understand how to create macros, this technique can be simple to set up. You can attach the macros to buttons to make running the macros as easy as clicking the button.

Figure 26-2 shows a worksheet designed for what-if analysis. This simple production model contains two input cells: the hourly cost of labor and the unit cost for materials. The company produces three products, and each product requires a different number of hours and a different amount of materials to produce. Excel calculates the combined total profit. Management is trying to predict the total profit but is uncertain what the hourly labor cost and material costs are going to be. They've identified three scenarios, as listed in Table 26-1.

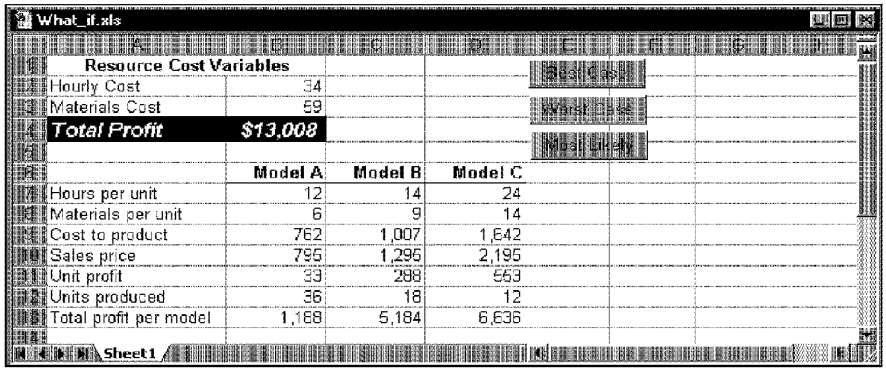


Figure 26-2: This worksheet uses macros to display three different combinations of values for the input cells.

Table 26-1
Three Scenarios for the Production Model

Scenario	*Hourly Cost*	*Materials Cost*
Best Case	30	57
Worst Case	38	62
Most Likely	34	59

I developed three simple macros and attached one to each of the three buttons on the worksheet. Figure 26-3 shows the VBA macros (also known as subroutines) that Excel executes when you click a worksheet button. These macros simply place values into the named cells on the worksheet. To change the values that any one of the scenarios uses, you must edit the macros.

If you like the idea of instantly displaying a particular scenario, you may be interested in learning about Excel's Scenario Manager, which is described later in this chapter. The Scenario Manager does not require macros.

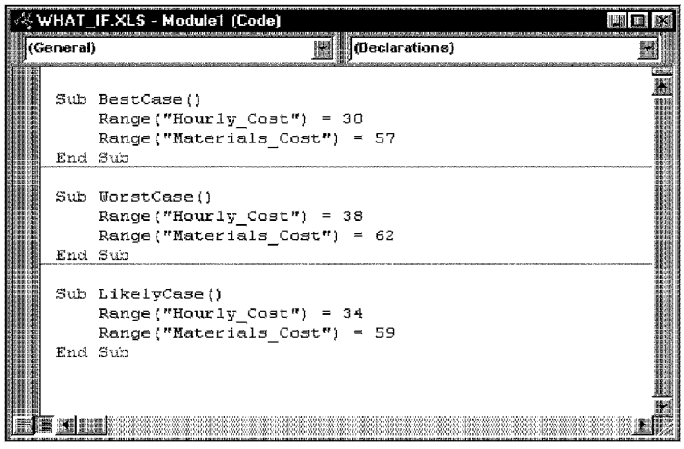


Figure 26-3: These macros simply place different values in the input cells in the worksheet.

Creating Data Tables

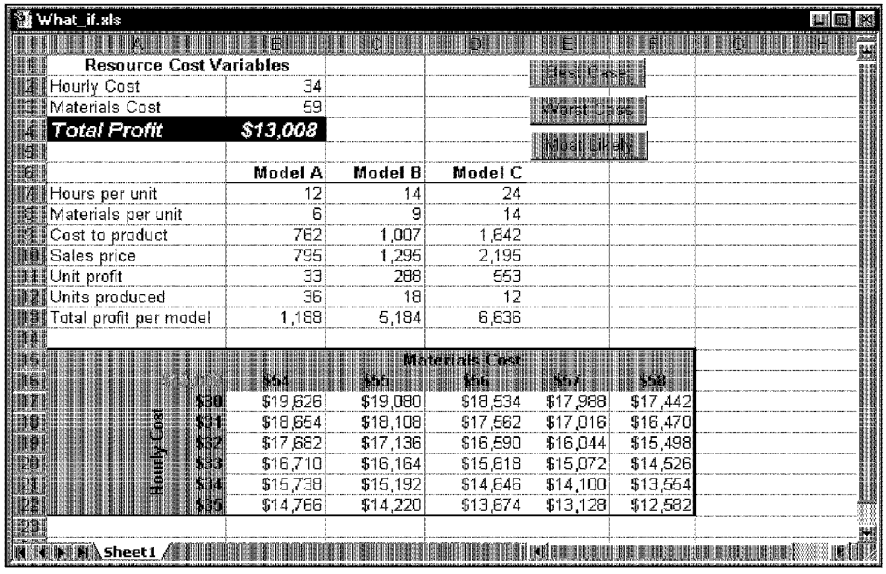
When you're working with a what-if model, Excel displays only one scenario at a time. But you can compare the results of various scenarios by using any of the following techniques:

- Print multiple copies of the worksheet, each displaying a different scenario.
- Copy the model to other worksheets and set it up so that each worksheet displays a different scenario.
- Manually create a table that summarizes key formula cells for each scenario.
- Use Excel's Data • Table command to create a summary table automatically.

This section discusses the last option—the Data • Table command, which enables you to create a handy data table that summarizes formula cells for various values of either of the following:

- A single input cell
- Various combinations of two input cells

For example, in the production model example, you may want to create a table that shows the total profit for various combinations of hourly cost and materials cost. Figure 26-4 shows a two-input data table that shows these combinations.



What_if.xls

	Resource Cost Variables					
	Hourly Cost	34				
	Materials Cost	69				
	Total Profit	*$13,008*				
		Model A	Model B	Model C		
	Hours per unit	12	14	24		
	Materials per unit	6	9	14		
	Cost to product	762	1,007	1,642		
	Sales price	795	1,295	2,195		
	Unit profit	33	288	553		
	Units produced	36	18	12		
	Total profit per model	1,188	5,184	6,636		
			Materials Cost			
		$54	$55	$56	$57	$58
Hourly Cost	$30	$19,626	$19,080	$18,534	$17,988	$17,442
	$31	$18,654	$18,108	$17,562	$17,016	$16,470
	$32	$17,682	$17,136	$16,590	$16,044	$15,498
	$33	$16,710	$16,164	$15,618	$15,072	$14,526
	$34	$15,738	$15,192	$14,646	$14,100	$13,554
	$35	$14,766	$14,220	$13,674	$13,128	$12,582

Sheet1

Figure 26-4: This data table summarizes the total profit for various combinations of the input values.

You can create a data table fairly easily, but data tables have some limitations. In particular, a data table can deal with only one or two input cells at a time. In other words, you can't create a data table that uses a combination of three or more input cells.

The Scenario Manager, discussed later in this chapter, can produce a report that summarizes any number of input cells and result cells.

Creating a One-Input Data Table

A one-input data table displays the results of one or more formulas when you use multiple values in a single input cell. Figure 26-5 shows the general layout for a one-input data table. You can place the table anywhere in the workbook. The left column contains various values for the single input cell. The top row contains formulas or, more often, references to formulas located elsewhere in the worksheet. You can use a single formula reference or any number of formula references. The upper-left cell of the table remains empty. Excel calculates the values that result from each level of the input cell and places them under each formula reference.

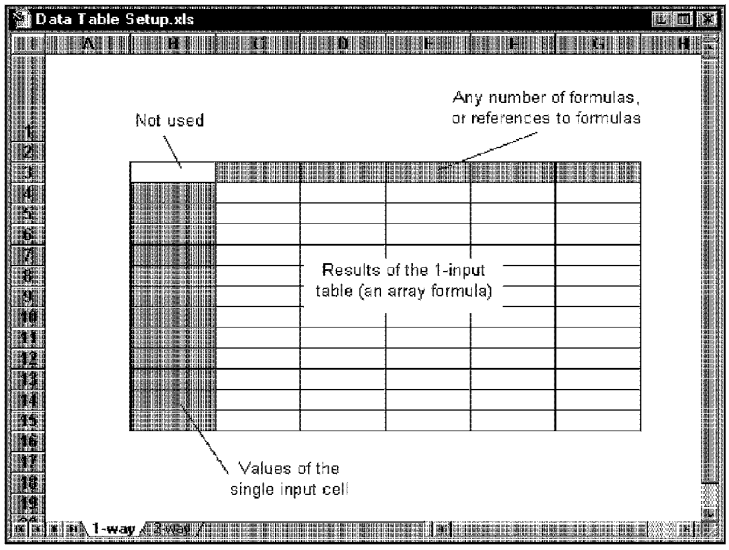


Figure 26-5: How a one-input data table is set up.

This example uses the mortgage loan worksheet from earlier in the chapter, which is shown again in Figure 26-6. The goal of this example is to create a table that shows the values of the four formula cells (loan amount, monthly payment, total payments, and total interest) for various interest rates ranging from 7 to 9 percent, in 0.25 percent increments.

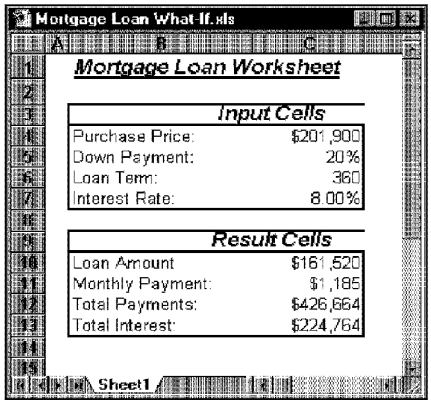


Figure 26-6: This example uses the mortgage loan worksheet to generate a one-input data table.

Figure 26-7 shows the setup for the data table area. Row 2 consists of references to the formulas in the worksheet. For example, cell F3 contains the formula =C10. Column E contains the values of the single input cell (interest rate) that Excel will use in the table. Borders also are added, to indicate where the calculated values go.

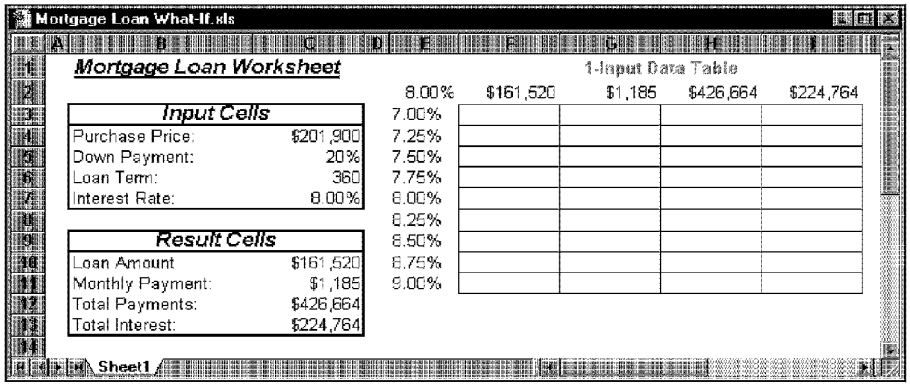


Figure 26-7: Preparing to create a one-input data table.

To create the table, select the range (in this case, E2:I11) and then choose Data • Table. Excel displays the Table dialog box, shown in Figure 26-8. You must specify the worksheet cell that contains the input value. Because variables for the input cell appear in a column in the data table rather than in a row, you place this cell reference in the text box called Column input cell. Enter **Interest_Rate** (the name for cell C7) or point to the cell in the worksheet. Leave the Row input cell field blank. Click OK, and Excel fills in the table with the appropriate results (see Figure 26-9).

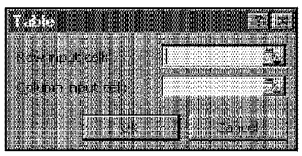


Figure 26-8: The Table dialog box.

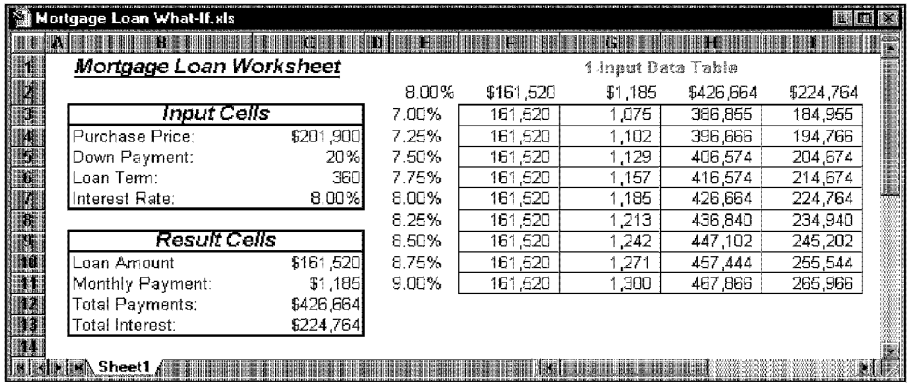


Figure 26-9: The result of the one-input data table.

Examine the contents of the cells that Excel entered as a result of this command, and notice that Excel filled in formulas — more specifically, array formulas that use the TABLE function. As discussed in Chapter 20, an array formula is a single formula that produces results in multiple cells. Because the table uses formulas, Excel updates the table that you produce if you change the cell references in the first row or plug in different interest rates in the first column.

You can arrange a one-input table vertically (as in this example) or horizontally. If you place the values of the input cell in a row, you enter the input cell reference in the text box labeled Row input cell in the Table dialog box.

Creating a Two-Input Data Table

As the name implies, a two-input data table lets you vary *two* input cells. You can see the setup for this type of table in Figure 26-10. Although it looks similar to a one-input table, the two-input table has one critical difference: it can show the results of only one formula at a time. With a one-input table, you can place any number of formulas, or references to formulas, across the top row of the table. In a two-input table, this top row holds the values for the second input cell. The upper-left cell of the table contains a reference to the single result formula.

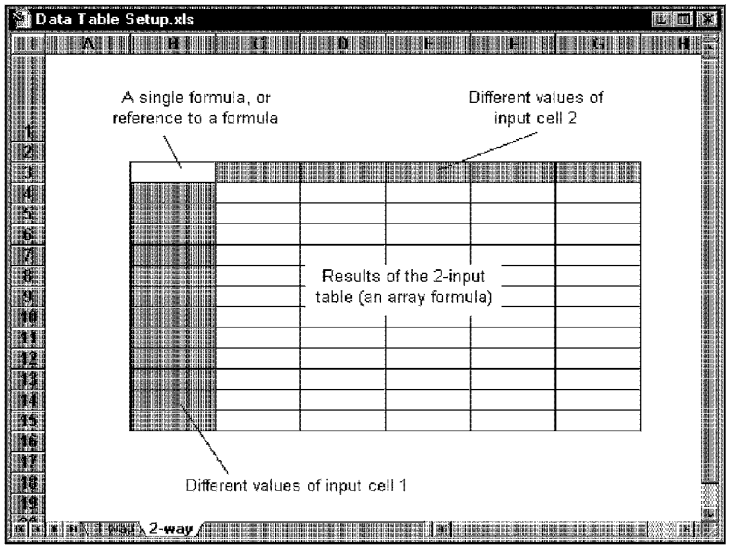


Figure 26-10: The setup for a two-input data table.

In the preceding example, you could create a two-input data table that shows the results of a formula (say, monthly payment) for various combinations of two input cells (such as interest rate and down-payment percent). To see the effects on other formulas, you simply create multiple data tables — one for each formula cell that you want to summarize.

The worksheet that is shown in Figure 26-11 demonstrates a two-input data table. In this example, a company wants to conduct a direct-mail promotion to sell its product. The worksheet calculates the net profit from the promotion.

This model uses two input cells: the number of promotional pieces mailed and the anticipated response rate. The following items appear in the results area:

- **Printing costs per unit:** The cost to print a single mailer. The unit cost varies with the quantity: $0.20 each for quantities less than 200,000; $0.15 each for quantities of 200,001 through 300,000; and $0.10 each for quantities of more than 300,000. The following formula is used:

  ```
  =IF(Number_mailed<200000,0.2,IF(Number_mailed<300000,0.15,0.1
  ))
  ```

- **Mailing costs per unit:** This is a fixed cost, $0.32 per unit mailed.

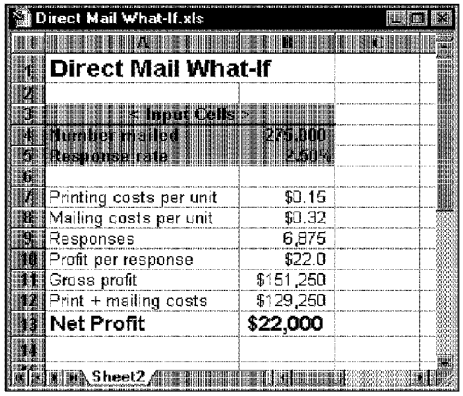


Figure 26-11: This worksheet calculates the net profit from a direct-mail promotion.

- **Responses:** This is the number of responses, calculated from the response rate and the number mailed. The formula in this cell is the following:

  ```
  =Response_rate*Number_mailed
  ```

- **Profit per response:** This is a fixed value. The company knows that it will realize a profit of $22 per order.
- **Gross profit:** This is a simple formula that multiplies the profit per response by the number of responses:

  ```
  =Profit_per_response*Responses
  ```

- **Print + mailing costs:** This formula calculates the total cost of the promotion:

  ```
  =Number_mailed*(Printing_costs_per_unit+Mailing_costs_per_unit)
  ```

- **Net Profit:** This formula calculates the bottom line—the gross profit minus the printing and mailing costs.

If you plug in values for the two input cells, you see that the net profit varies widely—often going negative to produce a net loss.

Figure 26-12 shows the setup of a two-input data table that summarizes the net profit at various combinations of quantity and response rate; the table appears in the range A15:I25.

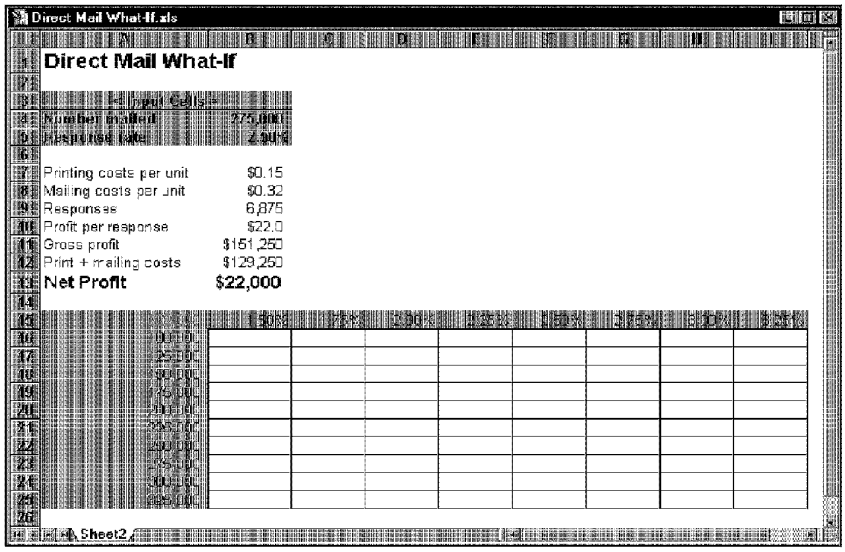


Figure 26-12: Preparing to create a two-input data table.

To create the data table, select the range and choose Data • Table. The Row input cell is Number_Mailed (the name for cell B4), and the Column input cell is Response_Rate (the name for cell B5). Figure 26-13 shows the result of this command.

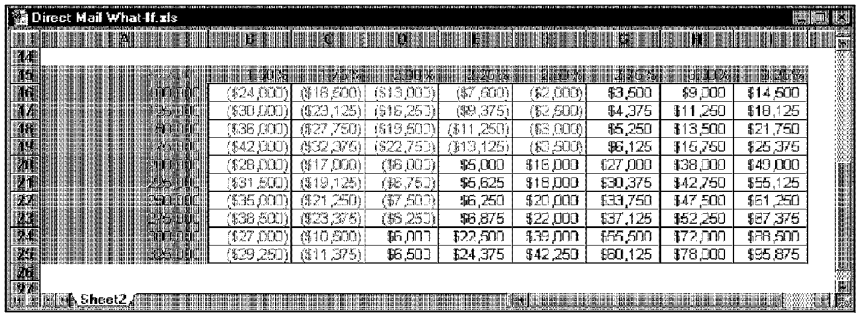


Figure 26-13: The result of the two-input data table.

Two-input data tables often make good 3D charts. An example of such a chart for the direct-mail example appears in Figure 26-14.

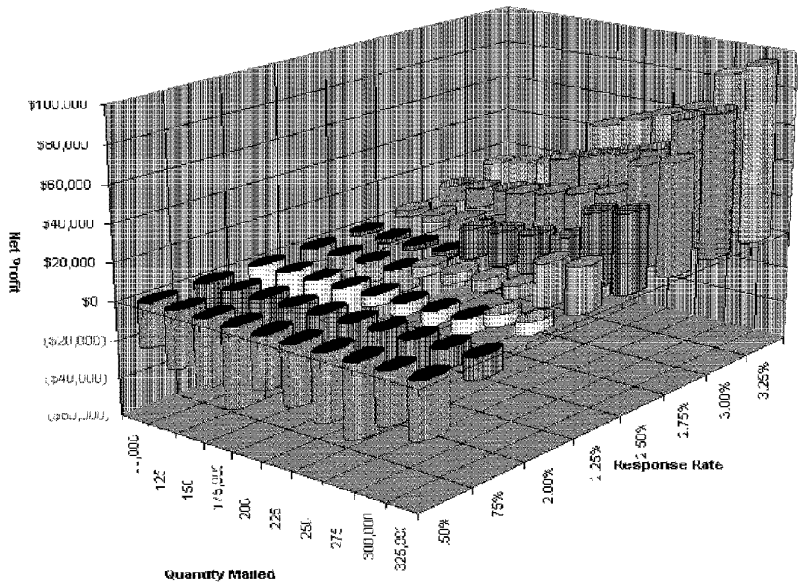


Figure 26-14: Viewing the two-input data table graphically.

Using Scenario Manager

Data tables are useful, but they have a few limitations:

- You can vary only one or two input cells at a time.
- The process of setting up a data table is not all that intuitive.
- A two-input table shows the results of only one formula cell (although you can create additional tables for more formulas).
- More often than not, you're interested in a few select combinations—not an entire table that shows all possible combinations of two input cells.

Excel's Scenario Manager feature makes it easy to automate your what-if models. You can store different sets of input values (called *changing cells* in the terminology of Scenario Manager) for any number of variables and give a name to each set. You can then select a set of values by name, and Excel displays the worksheet by using those values. You can also generate a summary report that shows the effect of various combinations of values on any number of result cells. These summary reports can be an outline or a pivot table.

Your sales forecast for the year, for example, may depend on several factors. Consequently, you can define three scenarios: best case, worst case, and most likely case. You then can switch to any of these scenarios by selecting the named scenario from a list. Excel substitutes the appropriate input values in your worksheet and recalculates the formulas. This process is similar, in some respects, to the macro-assisted what-if technique described earlier. The Scenario Manager is easier to use, however.

Defining Scenarios

To introduce you to the Scenario Manager, this section starts with a simple example: the production model used earlier in the chapter.

This example defines three scenarios, as depicted in Table 26-2. The Best Case scenario has the lowest hourly cost and materials cost. The Worst Case scenario has high values for both the hourly cost and the materials cost. The third scenario, Most Likely Case, has intermediate values for both of these input cells (this represents the management's best estimate). The managers need to be prepared for the worst case, however — and they are interested in what would happen under the Best Case scenario.

Table 26-2
Three Scenarios for the Production Model

Scenario	*Hourly Cost*	*Materials Cost*
Best Case	30	57
Worst Case	38	62
Most Likely Case	34	59

Access the Scenario Manager by selecting Tools • Scenarios to display the Scenario Manager dialog box, shown in Figure 26-15.

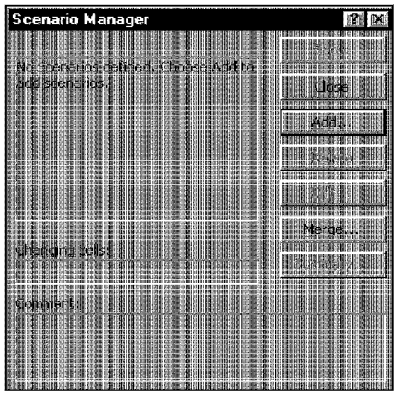


Figure 26-15: The Scenario Manager dialog box lets you assign names to different sets of assumptions.

When you first open this dialog box, it tells you that no scenarios are defined — which is not too surprising, because you're just starting. As you add named scenarios, they appear in this dialog box.

I strongly suggest that you create names for the changing cells and all the result cells that you want to examine. Excel uses these names in the dialog boxes and in the reports that it generates. If you use names, you'll find that keeping track of what's going on is much easier; names also make your reports more readable.

To add a scenario, click the Add button in the Scenario Manager dialog box. Excel displays its Add Scenario dialog box, shown in Figure 26-16.

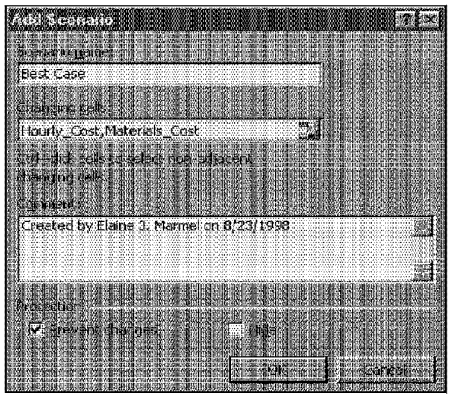


Figure 26-16: The Add Scenario dialog box lets you create a named scenario.

This dialog box consists of four parts:

- **Scenario name:** The name for the scenario. You can give it any name that you like — preferably something meaningful.
- **Changing cells:** The input cells for the scenario. You can enter the cell addresses directly or point to them. Multiple selections are allowed, so the input cells need not be adjacent. Each named scenario can use the same set of changing cells or different changing cells. The number of changing cells for a scenario is limited to 32.
- **Comment:** By default, Excel displays the name of the person who created the scenario and the date that it was created. You can change this text, add new text to it, or delete it.
- **Protection:** The two options (preventing changes and hiding a scenario) are in effect only when you protect the worksheet and choose the Scenario option in the Protect Sheet dialog box. Protecting a scenario prevents anyone from modifying it; a hidden scenario doesn't appear in the Scenario Manager dialog box.

In this example, define the three scenarios that are listed in the preceding table. The changing cells are Hourly_Cost (B4) and Materials_Cost (B5).

After you enter the information in the Add Scenario dialog box, click OK. Excel then displays the Scenario Values dialog box, shown in Figure 26-17. This dialog box displays one field for each changing cell that you specified in the previous dialog box. Enter the values for each cell in the scenario. If you click OK, you return to the Scenario Manager dialog box — which then displays your named scenario in its list. If you have more scenarios to create, click the Add button to return to the Add Scenario dialog box.

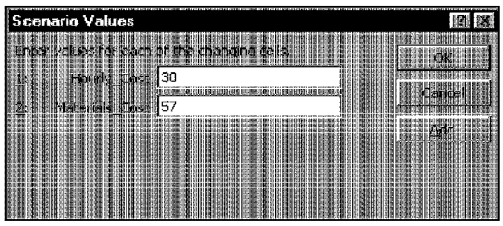


Figure 26-17: You enter the values for the scenario in the Scenario Values dialog box.

Using the Scenarios Tool

Excel has a Scenarios tool, which is a drop-down list that shows all the defined scenarios and enables you to display a scenario or create a new scenario. Oddly, this useful tool doesn't appear on any of the prebuilt toolbars. But, if you use the Scenario Manager, you may want to add the Scenarios tool to one of your toolbars, using the following procedure:

1. Choose Tools • Customize.
2. In the Customize dialog box, click the Commands tab.
3. Select the Tools category.
4. In the Commands tab, locate the Scenarios tool and drag it to any toolbar.
5. Click the Close button.

Refer to Chapter 33 for additional details on customizing toolbars.

Using the Scenarios tool may be more efficient than bringing up the Scenario Manager dialog box to create or view a different scenario.

To create a scenario by using the Scenarios tool, enter the scenario's values, select the changing cells, and then enter the name for the scenario in the Scenario drop-down box. To view a named scenario, just choose it from the list. Scenarios that you define in this manner also appear in the Scenario Manager dialog box. So, if you want to perform any operations on your scenarios (add comments, edit values, or generate reports), you need to select Tools • Scenarios, to display the Scenario Manager dialog box.

Displaying Scenarios

After you define all the scenarios and return to the Scenario Manager dialog box, the dialog box displays the names of your defined scenarios. Select one of the scenarios and then click the Show button. Excel inserts the corresponding values into the changing cells and calculates the worksheet to show the results for that scenario.

Modifying Scenarios

The Edit button in the Scenario Manager dialog box lets you change one or more of the values for the changing cells of a scenario. Select the scenario that you want to change, click the Edit button, choose OK to access the Scenario Values dialog box, and then make your changes. Notice that Excel automatically updates the Comments box with new text that indicates when the scenario was modified.

Merging Scenarios

In workgroup situations, you may have several people working on a spreadsheet model, and several people may have defined various scenarios. The marketing department, for example, may have its opinion of what the input cells should be, the finance department may have another opinion, and your CEO may have yet another opinion.

Excel makes it easy to merge these various scenarios into a single workbook by using the Merge button in the Scenario Manager dialog box. Clicking this button displays the dialog box shown in Figure 26-18.

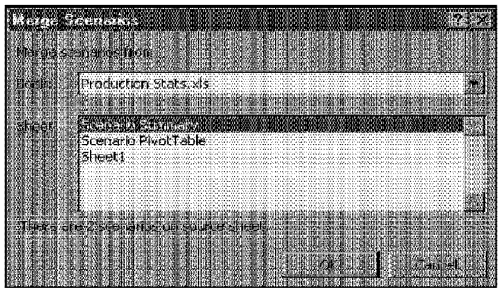


Figure 26-18: The Merge Scenarios dialog box lets you merge scenarios that are defined by others into your workbook.

Before you merge scenarios, make sure that the workbook from which you're merging is open. Then, click the Merge button in the Scenario Manager dialog box. Excel displays its Merge Scenarios dialog box. Choose the workbook from which you're merging in the Book drop-down list. Then, choose the sheet that contains the scenarios you want to merge from the Sheet list box (notice that the dialog box displays the number of scenarios in each sheet as you scroll through the Sheet list box). Click OK, and you return to the previous dialog box, which now displays the scenario names that you merged from the other workbook.

Generating a Scenario Report

You are ready to take the Scenario Manager through its final feat — generating a summary report. When you click the Summary button in the Scenario Manager dialog box, Excel displays the Scenario Summary dialog box, shown in Figure 26-19.

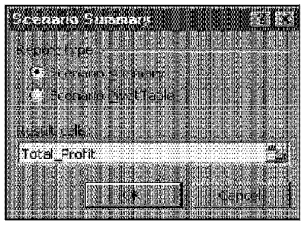


Figure 26-19: The Scenario Summary dialog box enables you to choose a report type and specify the result cells in which you're interested.

You have a choice of report types:

- **Scenario Summary:** The summary report appears in the form of an outline.
- **Scenario PivotTable:** The summary report appears in the form of a pivot table (see Chapter 25).

For simple cases of scenario management, a standard Scenario Summary report is usually sufficient. If you have many scenarios defined with multiple result cells, however, you may find that a Scenario Pivot Table provides more flexibility.

The Scenario Summary dialog box also asks you to specify the result cells (the cells that contain the formulas in which you're interested). For this example, select B15:D15 and B17 (a multiple selection) to make the report show the profit for each product, plus the total profit.

Excel creates a new worksheet to store the summary table. Figure 26-20 shows the Scenario Summary form of the report, and Figure 26-21 shows the Scenario Pivot Table form. If you gave names to the changing cells and result cells, the table uses these names. Otherwise, it lists the cell references.

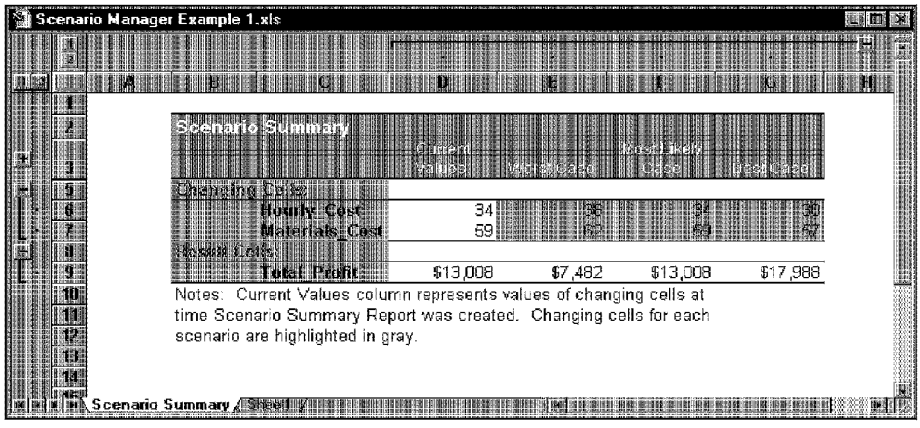


Figure 26-20: A Scenario Summary report produced by the Scenario Manager.

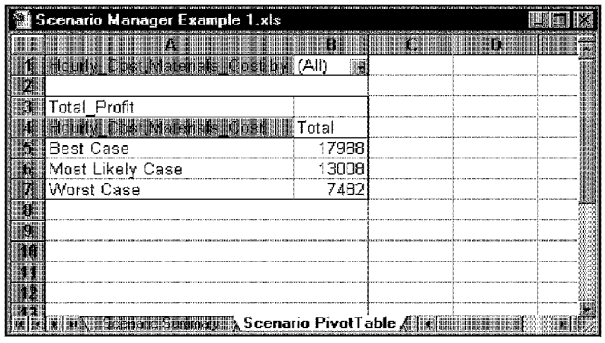


Figure 26-21: A Scenario Pivot Table report produced by the Scenario Manager.

Scenario Manager Limitations

As you work with the Scenario Manager, you may discover its main limitation: a scenario can use no more than 32 changing cells. If you attempt to use more, you get the message that is shown in Figure 26-22.

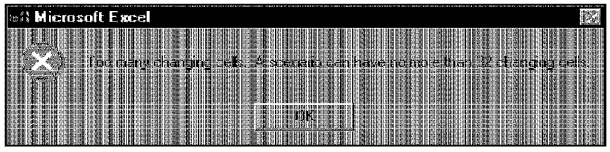


Figure 26-22: The Scenario Manager is limited to 32 changing cells.

You can get around this limitation by splitting your scenarios into parts. For example, assume that you have a worksheet with monthly sales projections for three years (36 changing cells). You may want to define various scenarios for these projections. But, because the number of changing cells exceeds the 32-cell limit, you can break it down into two or three scenarios — each of which uses a different set of changing cells. For example, you can define a scenario for the first 12 months, another for the second 12 months, and yet another for the third 12 months. Then, to display a particular scenario, you must display all three subscenarios. Writing simple macros makes this easy. If you use this technique, be aware that Excel includes superfluous information in summary reports.

Summary

This chapter discusses the concept of spreadsheet what-if analysis. What-if analysis is the process of systematically changing input cells and observing the effects on one or more formula cells. You can perform what-if analysis manually by plugging in different values. You also can use macros to automate this process. Excel's data table feature enables you to summarize the results of various values of a single input cell or various combinations of two input cells. The Scenario Manager feature makes it easy to create scenarios and generate summary reports.

CHAPTER 27

Analyzing Data Using Goal Seeking and Solver

In This Chapter

The preceding chapter discusses *what-if analysis*—the process of changing input cells to observe the results on other dependent cells. This chapter looks at that process from the opposite perspective—finding the value of one or more input cells that produces a desired result in a formula cell.

What-If Analysis—In Reverse

Consider the following what-if question: "What is the total profit if sales increase by 20 percent?" If you set up your worksheet properly, you can change the value in one cell to see what happens to the profit cell. Goal seeking takes the opposite approach. If you know what a formula result *should* be, Excel can tell you the values that you need to enter in one or more input cells to produce that result. In other words, you can ask a question such as, "How much do sales need to increase to produce a profit of $1.2 million?" Excel provides two tools that are relevant:

- **Goal seeking:** Determines the value that you need to enter in a single input cell to produce a result that you want in a dependent (formula) cell.
- **Solver:** Determines the values that you need to enter in multiple input cells to produce a result that you want. Moreover, because you can specify certain constraints to the problem, you gain significant problem-solving ability.

Single-Cell Goal Seeking

Single-cell goal seeking (also known as *backsolving*) is a rather simple concept. Excel determines what value in an input cell produces a desired result in a formula cell. Walk through the following example to understand how single-cell goal seeking works.

A Goal-Seeking Example

Figure 27-1 shows the mortgage loan worksheet that was used in the preceding chapter. This worksheet has four input cells and four formula cells. Originally, this worksheet was used for a what-if analysis example. In this section, the opposite approach is taken — rather than supply different input cell values to look at the calculated formulas, this example lets Excel determine one of the input values.

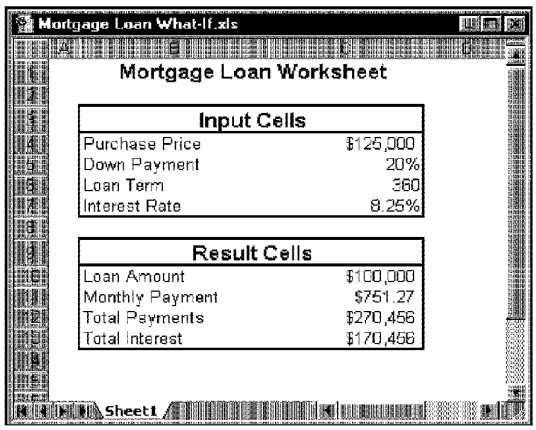


Figure 27-1: This worksheet is a good demonstration of goal seeking.

Assume that you're in the market for a new home and you know that you can afford $1,200 per month in mortgage payments. You also know that a lender can issue a fixed-rate mortgage loan for 8.25 percent, based on an 80 percent loan-to-value (that is, a 20-percent down payment). The question is, "What is the maximum purchase price I can handle?" In other words, what value in cell C4 causes the formula in cell C11 to result in $1,200? You could plug values into cell C4 until C11 displays $1,200; however, Excel can determine the answer much more efficiently.

To answer the question posed in the preceding paragraph, select Tools • Goal Seek. Excel displays the dialog box shown in Figure 27-2. Completing this dialog box is similar to forming a sentence. You want to set cell C11 to 1200 by changing cell C4. Enter this information in the dialog box either by typing the cell references or by pointing with the mouse. Click OK to begin the goal-seeking process.

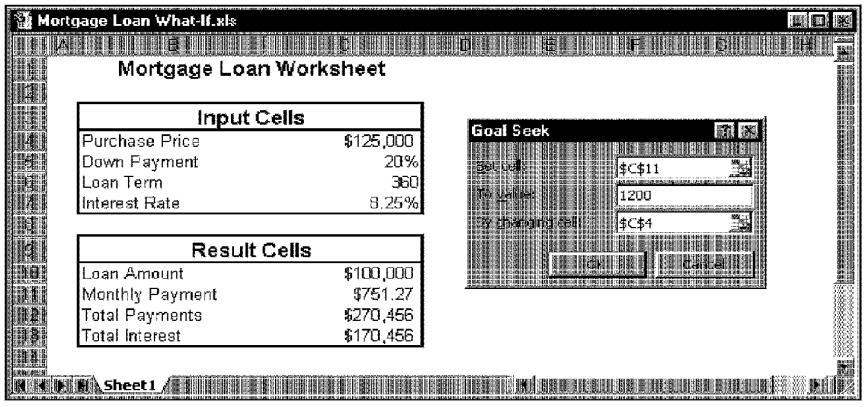


Figure 27-2: The Goal Seek dialog box.

In about a second, Excel announces that it has found the solution and displays the Goal Seek Status box, which shows the target value and the value that Excel calculated. In this case, Excel found an exact value. The worksheet now displays the found value in cell C4 ($199,663). As a result of this value, the monthly payment amount is $1,200. At this point, you have two options:

- Click OK to replace the original value with the found value.
- Click Cancel to restore your worksheet to the form that it had before you chose Tools • Goal Seek.

More About Goal Seeking

Excel can't always find a value that produces the result for which you're looking—sometimes, a solution simply doesn't exist. In such a case, the Goal Seek Status box informs you of that fact (see Figure 27-3).

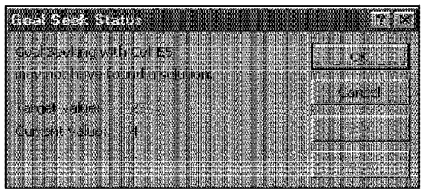


Figure 27-3: When Excel can't find a solution to your goal-seeking problem, it tells you so.

Other times, however, Excel may report that it can't find a solution, but you're pretty sure that one exists. If that's the case, you can try the following options:

- Change the current value of the By changing cell box in the Goal Seek dialog box to a value that is closer to the solution, and then reissue the command.
- Adjust the Maximum iterations setting in the Calculation tab of the Options dialog box. Increasing the number of iterations makes Excel try more possible solutions.
- Double-check your logic and make sure that the formula cell does, indeed, depend on the specified changing cell.

Like all computer programs, Excel has limited precision. To demonstrate this limitation, enter **=A1^2** into cell A2. Then, select Tools • Goal Seek to find the value in cell A1 (which is empty) that makes the formula return 16. Excel comes up with a value of 4.00002269 (you may need to widen the column to see the complete value), which is close to the square root of 16, but certainly not exact. You can adjust the precision in the Calculation tab of the Options dialog box (make the Maximum change value smaller).

In some cases, multiple values of the input cell produce the same desired result. For example, the formula =A1^2 returns 16 if cell A1 contains either –4 or +4. If you use goal seeking when two solutions are possible, Excel gives you the solution that has the same sign as the current value in the cell.

Perhaps the main limitation of the Tools • Goal Seek command is its inability to find the value for more than one input cell. For example, it can't tell you what purchase price *and* what down-payment percent will result in a particular monthly payment. If you want to change more than one variable at a time, use Solver (discussed later in this chapter).

Graphical Goal Seeking

Excel provides another way to perform goal seeking — by manipulating a graph. Figure 27-4 shows a worksheet that projects sales for a startup company. The CFO knows from experience that companies in this industry can grow exponentially according to a formula such as this one:

```
y*(bx)
```

Table 27-1 lists and describes the variables.

Table 27-1
Variables Used in the Sales Growth Formula

Variable	*Description*
y	A constant equal to the first year's sales
b	A growth coefficient
x	A variable relating to time

The company managers know that sales during the first year are going to be $250,000, and they want to increase the company's sales to $10 million by the year 2005. The financial modelers want to know the exact growth coefficient that meets this goal. The worksheet that is shown in Figure 27-4 uses formulas to forecast the annual sales, based on the growth coefficient in cell B1. The worksheet has an embedded chart that plots the annual sales.

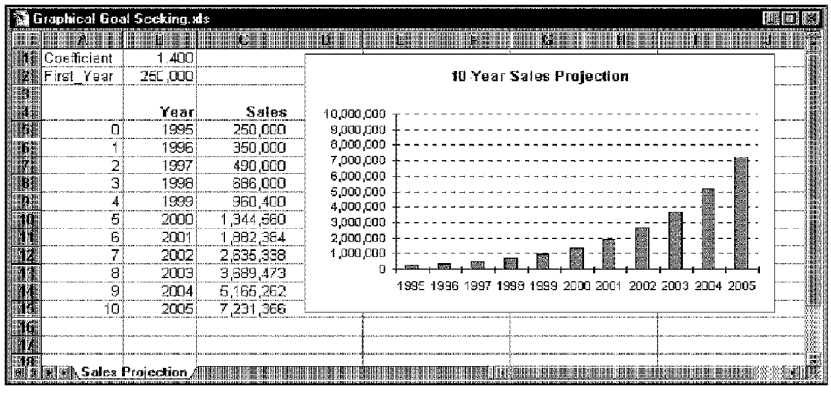


Figure 27-4: This sales projection predicts exponential growth, based on the growth coefficient in cell B1.

The initial guess for the growth coefficient is 1.40. As you can see, this number is too low — it results in sales of only $7.231 million for the year 2005. Although you can select Tools • Goal Seek to arrive at the exact coefficient, you have another way to do it.

Click the chart so that you can edit it and then select the chart series. Now, click the last data column to select only that column in the series. Point to the top of the column, and the mouse pointer changes shape. Drag the column upward and watch the value change in the small box displayed next to the mouse pointer. When the value is exactly $10 million, release the mouse button.

Excel responds with the Goal Seek dialog box, with two fields completed, as shown in Figure 27-5. Excel just needs to know which cell to use for the input cell. Specify cell B1 or enter **Coefficient** in the By changing cell edit box. Excel calculates the value of Coefficient that is necessary to produce the result that you pointed out on the chart. If you want to keep that number (which, by the way, is 1.44612554959182), click OK. Excel replaces the current value of Coefficient with the new value, and the chart is updated automatically. You can probably appreciate the fact that it would take quite a while to arrive at this number by plugging in successive approximations.

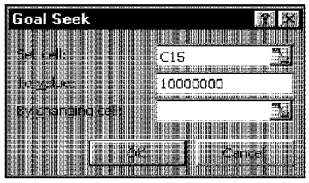


Figure 27-5: The Goal Seek dialog box appears when you directly manipulate a point on a chart that contains a formula.

You don't want to use this graphical method all the time, however, because the normal Tools • Goal Seek command is more efficient. But, it does demonstrate another way to approach problems that is helpful for those who are more visually oriented.

As you may expect, goal seeking can get much more impressive when it's used with complex worksheets that have many dependent cells. In any event, it sure beats trial and error.

Introducing Solver

Excel's goal-seeking feature is a useful tool, but it clearly has limitations. It can solve for only one adjustable cell, for example, and it returns only a single solution. Excel's powerful Solver tool extends this concept by enabling you to do the following:

- Specify multiple adjustable cells.
- Specify constraints on the values that the adjustable cells can have.
- Generate a solution that maximizes or minimizes a particular worksheet cell.
- Generate multiple solutions to a problem.

Although goal seeking is a relatively simple operation, using Solver can be much more complicated. In fact, Solver is probably one of the most difficult (and potentially frustrating) features in Excel. I'm the first to admit that Solver isn't for

everyone. In fact, most Excel users have no use for this feature. However, many users find that having this much power is worth spending the extra time to learn about it.

Appropriate Problems for Solver

Problems that are appropriate for Solver fall into a relatively narrow range. They typically involve situations that meet the following criteria:

- A target cell depends on other cells and formulas. Typically, you want to maximize or minimize this target cell or set it equal to some value.
- The target cell depends on a group of cells (called *changing cells*) that Solver can adjust to affect the target cell.
- The solution must adhere to certain limitations, or *constraints*.

After you set up your worksheet appropriately, you can use Solver to adjust the changing cells and produce the result that you want in your target cell—and, simultaneously meet all the constraints that you have defined.

You can find all the Solver examples in this chapter on this book's CD-ROM.

A Simple Solver Example

I start with a simple example to introduce Solver and then present some increasingly complex examples to demonstrate what it can do.

Figure 27-6 shows a worksheet that is set up to calculate the profit for three products. Column B shows the number of units of each product, column C shows the profit per unit for each product, and column C contains formulas that calculate the profit for each product by multiplying the units by the profit per unit.

Solver Production Model.xls

	A	B	C	D
1				
2		Units	Profit/Unit	Profit
3	Product A	100	$13	$1,300
4	Product B	100	$18	$1,800
5	Product C	100	$22	$2,200
6	Total	300		$5,300
7				
8				
9				
10				
11				

Sheet1

Figure 27-6: Use Solver to determine the number of units to maximize the total profit.

It doesn't take an MBA degree to realize that the greatest profit per unit comes from Product C. Therefore, the logical solution is to produce only Product C. If things were really this simple, you wouldn't need tools such as Solver. As in most situations, this company has some constraints to which it must adhere:

- The combined production capacity is 300 total units per day.
- The company needs 50 units of Product A to fill an existing order.
- The company needs 40 units of Product B to fill an anticipated order.
- Because the market for Product C is relatively limited, the company doesn't want to produce more than 40 units of this product.

These four constraints make the problem more realistic and challenging. In fact, it's a perfect problem for Solver.

The basic procedure for using Solver is as follows:

1. Set up the worksheet with values and formulas. Make sure that you format cells logically; for example, if you cannot produce portions of your products, format those cells to contain numbers with no decimal values.
2. Bring up the Solver dialog box.
3. Specify the target cell.
4. Specify the changing cells.
5. Specify the constraints.
6. Change the Solver options, if necessary.
7. Let Solver solve the problem.

To start Solver, select Tools • Solver. Excel displays its Solver Parameters dialog box, shown in Figure 27-7.

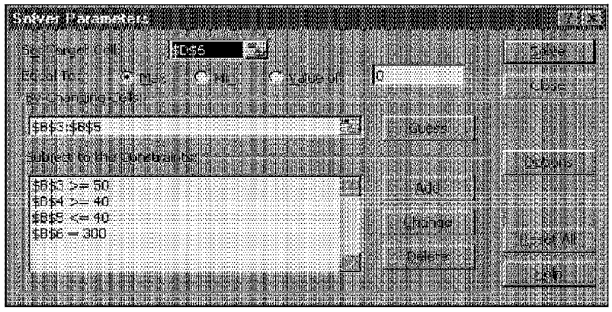


Figure 27-7: The Solver Parameters dialog box.

No Tools • Solver Command?

Solver is an add-in, so it's available only when the add-in is installed. If the Tools menu doesn't show a Solver command, you need to install the add-in before you can use it.

Select Tools • Add-Ins. Excel displays its Add-Ins dialog box. Scroll down the list of add-ins and place a check mark next to the item named Solver Add-In. Click OK, and Excel installs the add-in and makes the Tools • Solver command available. If Solver isn't available on your computer, you'll be asked if you want to install it.

In this example, the target cell is D6—the cell that calculates the total profit for three products. Enter (or point to) cell D6 in the Set Target Cell field of the Solver Parameters dialog box. Because the objective is to maximize this cell, click the Max option. Next, specify the changing cells, which are in the range B3:B5, in the By Changing Cells box.

The next step is to specify the constraints on the problem. The constraints are added one at a time and appear in the box labeled Subject to the Constraints. To add a constraint, click the Add button. Excel displays the Add Constraint dialog box, shown in Figure 27-8. This dialog box has three parts: a cell reference, an operator, and a value. To set the first constraint—that the total production capacity is 300 units—enter B6 as the Cell Reference, choose equal (=) from the drop-down list of operators, and enter 300 as the Constraint value. Click Add to add the remaining constraints. Table 27-2 summarizes the constraints for this problem.

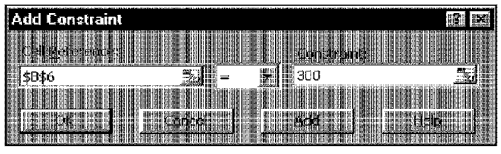


Figure 27-8: The Add Constraint dialog box.

Table 27-2
Constraints Summary

Constraint	*Expressed As*
Capacity is 300 units	B6=300
At least 50 units of Product A	B3>=50
At least 40 units of Product B	B4>=40
No more than 40 units of Product C	B5<=40

After you enter the last constraint, click OK to return to the Solver Parameters dialog box—which now lists the four constraints.

At this point, Solver knows everything about the problem. Click the Solver button to start the solution process. You can watch the progress onscreen, and Excel soon announces that it has found a solution. The Solver Results dialog box is shown in Figure 27-9.

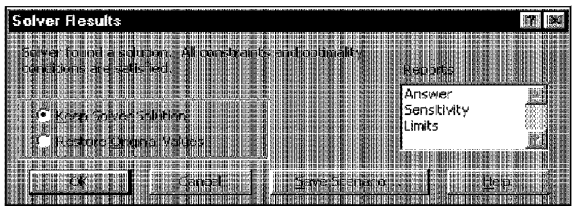


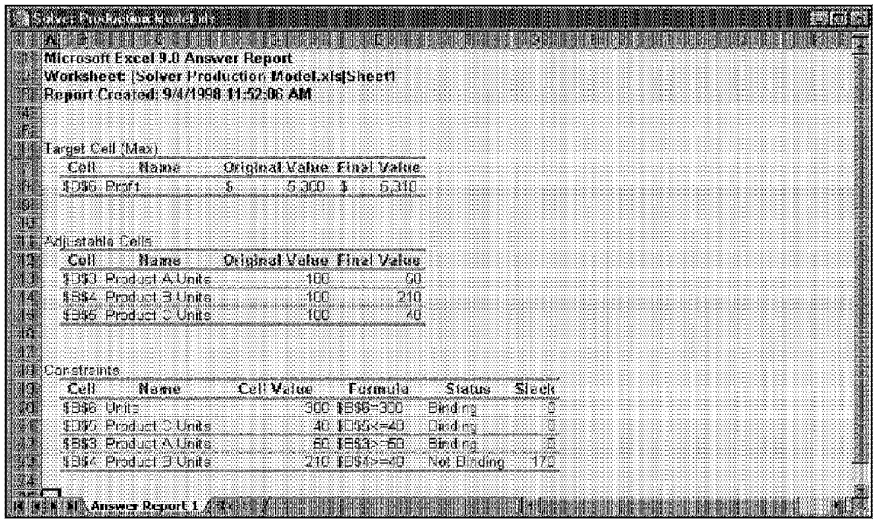
Figure 27-9: Solver displays this dialog box when it finds a solution to the problem.

At this point, you have the following options:

- Replace the original changing cell values with the values that Solver found
- Restore the original changing cell values
- Create any or all three reports that describe what Solver did (press Shift to select multiple reports from this list)
- Click the Save Scenario button to save the solution as a scenario, so that the Scenario Manager can use it (see Chapter 26)

If you specify any report options, Excel creates each report on a new worksheet, with an appropriate name. Figure 27-10 shows an Answer Report. In the Constraints section of the report, all the constraints except one are *binding*, which means that the constraint was satisfied at its limit, with no more room to change.

This simple example illustrates how Solver works. The fact is, you could probably solve this particular problem manually just as quickly. That, of course, isn't always the case.


Microsoft Excel 9.0 Answer Report
Worksheet: [Solver Production Model.xls]Sheet1
Report Created: 9/4/1998 11:52:06 AM

Target Cell (Max)

Cell	Name	Original Value	Final Value
$D$6	Profit	$ 5,300	$ 5,310

Adjustable Cells

Cell	Name	Original Value	Final Value
$D$3	Product A Units	100	60
$E$4	Product B Units	100	210
$F$5	Product C Units	100	40

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$B$6	Units	300	$B$6<=300	Binding	0
$D$7	Product C Units	40	$D$7>=40	Binding	0
$B$3	Product A Units	60	$B$3>=60	Binding	0
$B$4	Product B Units	210	$B$4>=40	Not Binding	170

Figure 27-10: One of three reports that Solver can produce.

More About Solver

Before presenting complex examples, this section discusses the Solver Options dialog box—one of the more feature-packed dialog boxes in Excel. From this dialog box, you control many aspects of the solution process, as well as load and save model specifications in a worksheet range.

Having Solver report to you that it can't find a solution isn't unusual—even when you know that one should exist. Often, you can change one or more of the Solver options and try again. When you choose the Options button in the Solver Parameters dialog box, Excel displays the Solver Options dialog box shown in Figure 27-11.

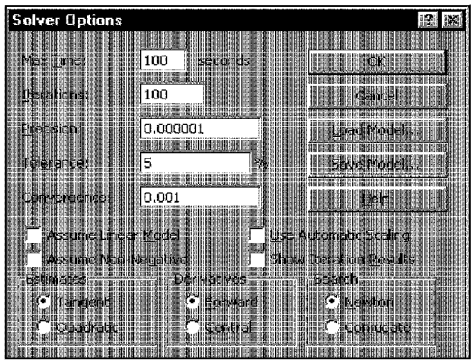


Figure 27-11: You can control many aspects of how Solver solves a problem.

This list describes Solver's options:

- **Max Time:** Specify the maximum amount of time (in seconds) that you want Solver to spend on a problem. If Solver reports that it exceeded the time limit, you can increase the amount of time that it spends searching for a solution.
- **Iterations:** Enter the maximum number of trial solutions that you want Solver to perform.
- **Precision:** Specify how close the Cell Reference and Constraint formulas must be to satisfy a constraint. Excel may solve the problem more quickly if you specify less precision.
- **Tolerance:** Designate the maximum percentage of error allowed for integer solutions (relevant only if an integer constraint is used).
- **Assume Linear Model:** Choose this option to speed the solution process, but you can use it only if all the relationships in the model are linear. You can't use this option if the adjustable cells are multiplied or divided, or if the problem uses exponents.
- **Use Automatic Scaling:** Use when the problem involves large differences in magnitude—when you attempt to maximize a percentage, for example, by varying cells that are very large.
- **Show Iteration Results:** Instruct Solver to pause and display the results after each iteration, by checking this box.
- **Estimates, Derivatives, and Search group boxes:** Use these options to control some technical aspects of the solution. In most cases, you don't need to change these settings.
- **Load Model:** Click this button to make Excel display the Load Model dialog box, in which you specify a range containing the model that you want to load.
- **Save Model:** Click this button to make Excel display the Save Model dialog box, in which you specify a range where Excel should save the model parameters.

Usually, you want to save a model only when you're using more than one set of Solver parameters with your worksheet, because Excel saves the first Solver model automatically with your worksheet (using hidden names). If you save additional models, Excel stores the information in the form of the formulas that correspond to the specification that you make (the last cell in the saved range is an array formula that holds the options settings).

Solver Examples

The remainder of this chapter consists of examples of using Solver for various types of problems.

Minimizing Shipping Costs

This example involves finding alternative options for shipping materials while keeping total shipping costs at a minimum (see Figure 27-12). A company has warehouses in Los Angeles, St. Louis, and Boston. Retail outlets throughout the United States place orders, which the company then ships from one of the warehouses. Ideally, the company wants to meet the product needs of all six retail outlets from available inventory in the warehouses — and keep total shipping charges as low as possible.

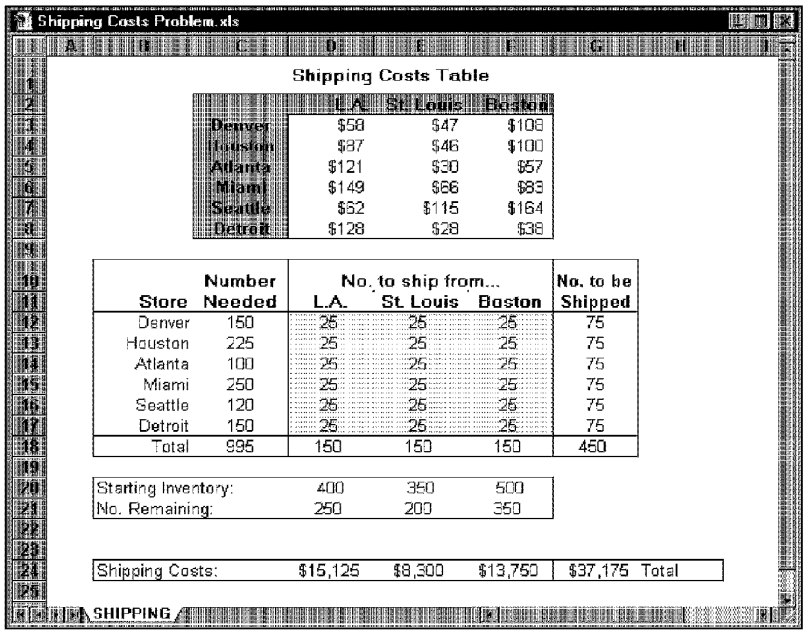


Figure 27-12: This worksheet determines the least expensive way to ship products from warehouses to retail outlets.

This workbook is rather complicated, so each part is explained individually:

- **Shipping Costs Table:** This table, at the top of the worksheet, contains per-unit shipping costs from each warehouse to each retail outlet. The cost to ship a unit from Los Angeles to Denver, for example, is $58.
- **Product needs of each retail store:** This information appears in C12:C17. For example, Denver needs 150 units, Houston needs 225, and so on. C18 holds the total needed.

- **Number to ship:** The shaded range (D12:F17) holds the adjustable cells that Solver varies (they are all initialized with a value of 25, to give Solver a starting value.) Column G contains formulas that total the number of units the company needs to ship to each retail outlet.
- **Warehouse inventory:** Row 20 contains the amount of inventory at each warehouse, and row 21 contains formulas that subtract the amount shipped (row 18) from the inventory. For example, cell D21 contains the following formula: =D20–D18.
- **Calculated shipping costs:** Row 24 contains formulas that calculate the shipping costs. Cell D24 contains the following formula, which is copied to the two cells to the right of cell D24:

```
=SUMPRODUCT(D3:D8,D12:D17)
```

This formula calculates the total shipping cost from each warehouse. Cell G24 is the bottom line, the total shipping costs for all orders.

Solver fills in values in the range D12:F17 in such a way that minimizes shipping costs while still supplying each retail outlet with the desired number of units. In other words, the solution minimizes the value in cell C24 by adjusting the cells in D12:F17, subject to the following constraints:

- The number of units needed by each retail outlet must equal the number shipped (in other words, all the orders are filled). These constraints are represented by the following specifications:

```
C12=G12     C14=G14     C16=G16
C13=G13     C15=G15     C17=G17
```

- The adjustable cells can't be negative, because shipping a negative number of units makes no sense. These constraints are represented by the following specifications:

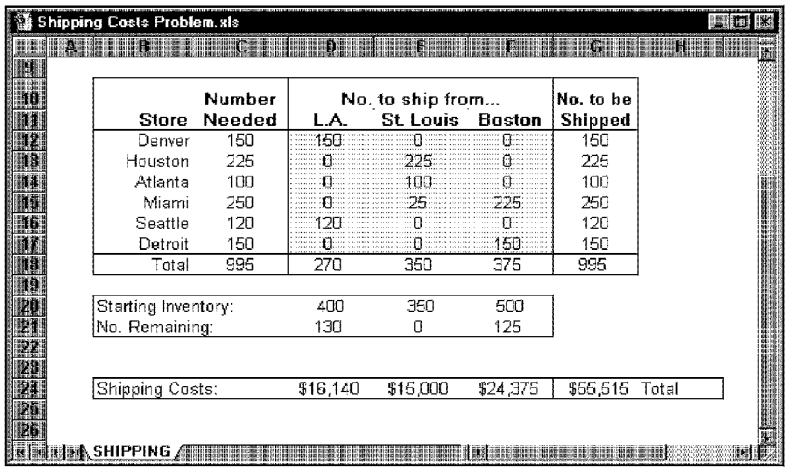
```
D12>=0      E12>=0      F12>=0
D13>=0      E13>=0      F13>=0
D14>=0      E14>=0      F14>=0
D15>=0      E15>=0      F15>=0
D16>=0      E16>=0      F16>=0
D17>=0      E17>=0      F17>=0
```

- The number of units remaining in each warehouse's inventory must not be negative (that is, they can't ship more than what is available). This is represented by the following constraint specifications:

```
D21>=0      E21>=0      F21>=0
```

Before you solve this problem with Solver, you may try your hand at minimizing the shipping cost manually by entering values in D12:F17. Don't forget to make sure that all the constraints are met. This is often a difficult task—and you can better appreciate the power behind Solver.

Setting up the problem is the difficult part. For example, you must enter 27 constraints. When you have specified all the necessary information, click the Solve button to put Solver to work. This process takes a while (Solver's speed depends on the speed of your computer and the amount of memory installed on your computer), but eventually Solver displays the solution that is shown in Figure 27-13.



Store	Number Needed	No. to ship from... L.A.	St Louis	Boston	No. to be Shipped
Denver	150	150	0	0	150
Houston	225	0	225	0	225
Atlanta	100	0	100	0	100
Miami	250	0	25	225	250
Seattle	120	120	0	0	120
Detroit	150	0	0	150	150
Total	995	270	350	375	995

Starting Inventory:	400	350	500
No. Remaining:	130	0	125

Shipping Costs:	$16,140	$15,000	$24,375	$55,515 Total

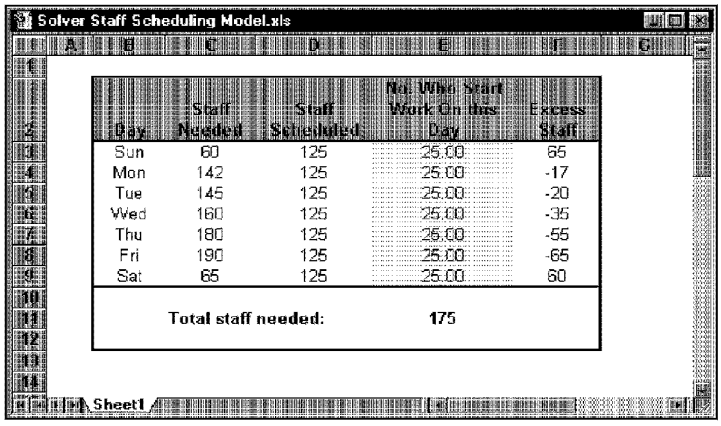
Figure 27-13: The solution that was created by Solver.

The total shipping cost is $55,515, and all the constraints are met. Notice that shipments to Miami come from both St. Louis and Boston.

Scheduling Staff

This example deals with staff scheduling. Such problems usually involve determining the minimum number of people that satisfy staffing needs on certain days or times of the day. The constraints typically involve such details as the number of consecutive days or hours that a person can work.

Figure 27-14 shows a worksheet that is set up to analyze a simple staffing problem. The question is, "What is the minimum number of employees required to meet daily staffing needs?" At this company, each person works five consecutive days. As a result, employees begin their five-day workweek on different days of the week.



Day	Staff Needed	Staff Scheduled	No. Who Start Work On this Day	Excess Staff
Sun	60	125	25.00	65
Mon	142	125	25.00	-17
Tue	145	125	25.00	-20
Wed	160	125	25.00	-35
Thu	180	125	25.00	-55
Fri	190	125	25.00	-65
Sat	65	125	25.00	60
	Total staff needed:		175	

Figure 27-14: This staffing model determines the minimum number of staff members required to meet daily staffing needs.

The key to this problem, as with most Solver problems, is figuring out how to set up the worksheet. This example makes it clear that setting up your worksheet properly is critical to Solver. This worksheet is laid out as follows:

- **Day:** Column B consists of plain text for the days of the week.
- **Staff Needed:** The values in column C represent the number of employees needed on each day of the week. As you see, staffing needs vary quite a bit by the day of the week.
- **Staff Scheduled:** Column D holds formulas that use the values in column E. Each formula adds the number of people who start on that day to the number of people who started on the preceding four days. Because the week wraps around, you can't use a single formula and copy it. Consequently, each formula in column D is different:

```
D3:     =E3+E9+E8+E7+E6
D4:     =E4+E3+E9+E8+E7
D5:     =E5+E4+E10+E9+E8
D6:     =E6+E5+E4+E10+E9
D7:     =E7+E6+E5+E4+E10
D8:     =E8+E7+E6+E5+E4
D9:     =E9+E8+E7+E6+E5
```

- **Adjustable cells:** Column E holds the adjustable cells—the numbers to be determined by Solver. These cells are initialized with a value of 25, to give Solver a starting value. Generally, you should initialize the changing cells to values that are as close as possible to the anticipated answer.

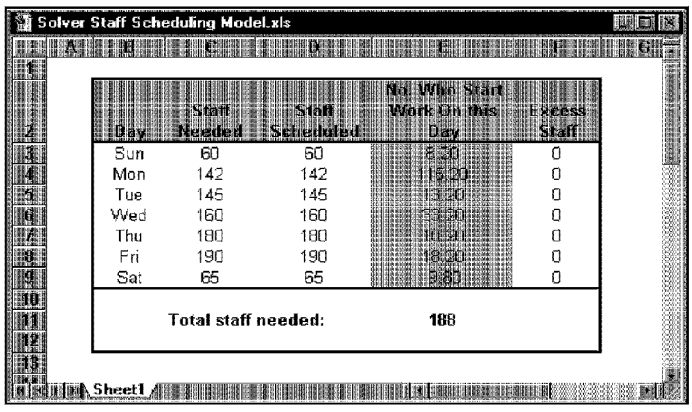
- **Excess Staff:** Column F contains formulas that subtract the number of staff members needed from the number of staff members scheduled, to determine excess staff. Cell F3 contains =D3 – C3, which was copied to the six cells below it.
- **Total staff needed:** Cell E11 contains a formula that sums the number of people who start on each day. The formula is =SUM(E3:E9). This is the value that Solver minimizes.

This problem, of course, has constraints. The number of people scheduled each day must be greater than or equal to the number of people required. If each value in column F is greater than or equal to 0, the constraints are satisfied.

After the worksheet is set up, select Tools • Solver and specify that you want to minimize cell E11 by changing cells E3:E9. Next, click the Add button to begin adding the following constraints:

```
F3>=0
F4>=0
F5>=0
F6>=0
F7>=0
F8>=0
F9>=0
```

Click Solve to start the process. The solution that Solver finds, shown in Figure 27-15, indicates that a staff of 188 meets the staffing needs and that no excess staffing exists on any day.



Day	Staff Needed	Staff Scheduled	No. Who Start Work On this Day	Excess Staff
Sun	60	60	8.20	0
Mon	142	142	116.20	0
Tue	145	145	13.20	0
Wed	160	160	33.20	0
Thu	180	180	10.31	0
Fri	190	190	16.20	0
Sat	65	65	0.80	0
	Total staff needed:		188	

Figure 27-15: This solution offered by Solver isn't quite right—you have to add more constraints.

But wait! If you examine the results carefully, you notice that a few things are wrong here:

- Solver's solution involves partial people — who are difficult to find. For example, 8.2 people begin their workweek on Sunday.
- Even more critical is the suggestion that a negative number of people should begin their workweek on Saturday.

You can correct both of these problems easily by adding more constraints. Fortunately, Solver enables you to limit the solution to integers, by using the integer option in the Add Constraint dialog box. This means that you must add another constraint for each cell in E3:E9. Figure 27-16 shows how you can specify an integer constraint. Avoiding the negative people problem requires seven more constraints of the form **E3>=0**, one for each cell in E3:E9.

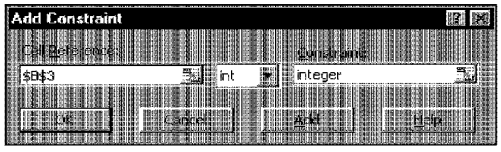
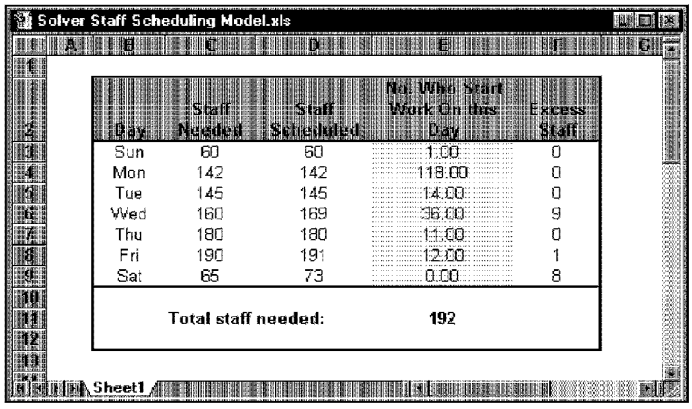


Figure 27-16: With many problems, you have to limit the solution to integers. You can do this by selecting the integer option in the Constraint box of the Add Constraint dialog box.

These two problems (integer solutions and negative numbers) are quite common when using Solver. They also demonstrate that checking the results is important, rather than relying only on Solver's solution.

If you find that adding these constraints is tedious, save the model to a worksheet range. Then, you can add new constraints to the range in the worksheet (and make sure that you don't overwrite the last cell in this range). Next, run Solver again and load the modified model from the range that you edited. The example workbook (available on this book's CD-ROM) has three Solver ranges stored in it.

After adding these constraints, run Solver again. This time it arrives at the solution shown in Figure 27-17. Notice that this solution requires 192 people and results in excess staffing on three days of the week. This solution is the best one possible that uses the fewest number of people — and almost certainly is better than what you would arrive at manually.



Solver Staff Scheduling Model.xls

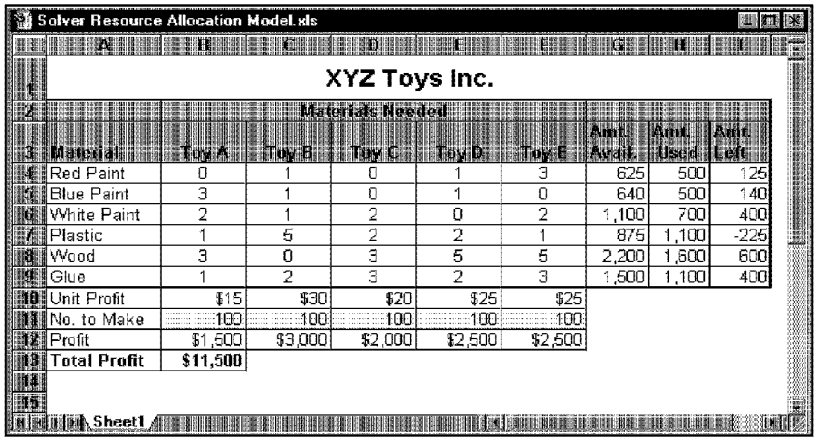
Day	Staff Needed	Staff Scheduled	No. Who Start Work On this Day	Excess Staff
Sun	60	60	1.00	0
Mon	142	142	118.00	0
Tue	145	145	14.00	0
Wed	160	169	36.00	9
Thu	180	180	11.00	0
Fri	190	191	12.00	1
Sat	65	73	0.00	8
	Total staff needed:		192	

Sheet1

Figure 27-17: Rerunning Solver after adding more constraints produces a better solution to the staffing model problem.

Allocating Resources

The example in this section is a common type of problem that's ideal for Solver. Essentially, problems of this sort involve optimizing the volumes of individual production units that use varying amounts of fixed resources. Figure 27-18 shows an example for a toy company.



Solver Resource Allocation Model.xls

XYZ Toys Inc.

Material	Materials Needed Toy A	Toy B	Toy C	Toy D	Toy E	Amt. Avail.	Amt. Used	Amt. Left
Red Paint	0	1	0	1	3	625	500	125
Blue Paint	3	1	0	1	0	640	500	140
White Paint	2	1	2	0	2	1,100	700	400
Plastic	1	5	2	2	1	875	1,100	-225
Wood	3	0	3	5	5	2,200	1,600	600
Glue	1	2	3	2	3	1,500	1,100	400
Unit Profit	$15	$30	$20	$25	$25			
No. to Make	100	100	100	100	100			
Profit	$1,500	$3,000	$2,000	$2,500	$2,500			
Total Profit	$11,500							

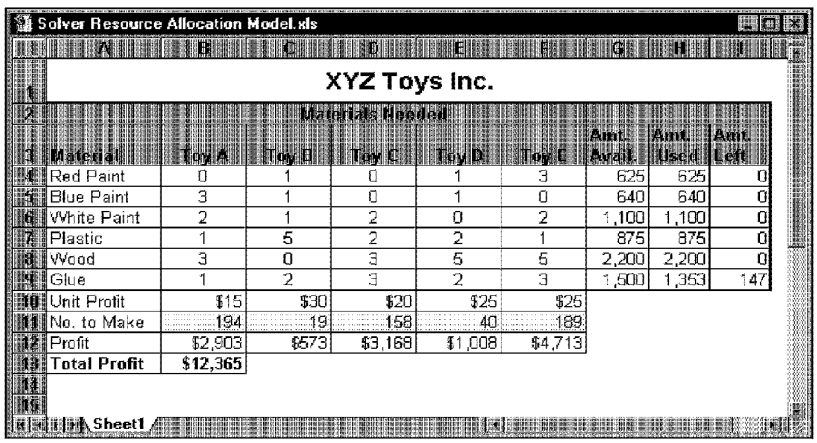
Sheet1

Figure 27-18: Using Solver to maximize profit when resources are limited.

This company makes five different toys, which use six different materials in varying amounts. For example, Toy A requires 3 units of blue paint, 2 units of white paint, 1 unit of plastic, 3 units of wood, and 1 unit of glue. Column G shows the current inventory of each type of material. Row 10 shows the unit profit for each toy. The number of toys to make is shown in the range B11:F11—these are the values that Solver determines. The goal of this example is to determine how to allocate the resources to maximize the total profit (B13). In other words, Solver determines how many units of each toy to make. The constraints in this example are relatively simple:

- Ensure that production doesn't use more resources than are available. This can be accomplished by specifying that each cell in column F is greater than or equal to zero.
- Ensure that the quantities produced aren't negative. This can be accomplished by specifying that each cell in row 11 be greater than or equal to zero.

Figure 27-19 shows the results that are produced by Solver. It shows the product mix that generates $12,365 in profit and uses all resources in their entirety, except for glue.



Solver Resource Allocation Model.xls

XYZ Toys Inc.

	Materials Needed							
Material	Toy A	Toy B	Toy C	Toy D	Toy E	Amt. Avail.	Amt. Used	Amt. Left
Red Paint	0	1	0	1	3	625	625	0
Blue Paint	3	1	0	1	0	640	640	0
White Paint	2	1	2	0	2	1,100	1,100	0
Plastic	1	5	2	2	1	875	875	0
Wood	3	0	3	5	5	2,200	2,200	0
Glue	1	2	3	2	3	1,500	1,353	147
Unit Profit	$15	$30	$20	$25	$25			
No. to Make	194	19	158	40	189			
Profit	$2,903	$573	$3,168	$1,008	$4,713			
Total Profit	**$12,365**							

Sheet1

Figure 27-19: Solver determined how to use the resources to maximize the total profit.

Optimizing an Investment Portfolio

This example demonstrates how to use Solver to help maximize the return on an investment portfolio. Portfolios consist of several investments, each of which has different yields. In addition, you may have some constraints that involve reducing risk and diversification goals. Without such constraints, a portfolio problem becomes a no-brainer: put all of your money in the investment with the highest yield.

This example involves a credit union, a financial institution that takes members' deposits and invests them in loans to other members, bank CDs, and other types of investments. The credit union distributes part of the return on these investments to the members in the form of *dividends*, or interest on their deposits. This hypothetical credit union must adhere to some regulations regarding its investments, and the board of directors has imposed some other restrictions. These regulations and restrictions comprise the problem's constraints. Figure 27-20 shows a workbook set up for this problem.

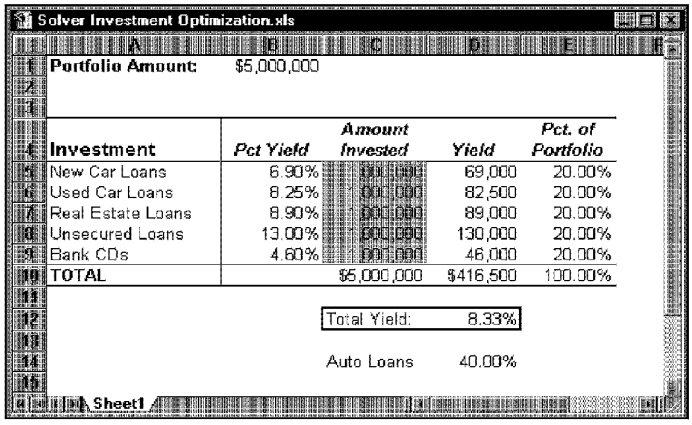


Figure 27-20: This worksheet is set up to maximize a credit union's investments, given some constraints.

The following constraints are the ones to which you must adhere in allocating the $5 million portfolio:

- The amount that the credit union invests in new-car loans must be at least three times the amount that the credit union invests in used-car loans (used-car loans are riskier investments). This constraint is represented as C5>=C6*3.
- Car loans should make up at least 15 percent of the portfolio. This constraint is represented as D14>=.15.
- Unsecured loans should make up no more than 25 percent of the portfolio. This constraint is represented as E8<=.25.
- At least 10 percent of the portfolio should be in bank CDs. This constraint is represented as E9>=.10.
- All investments should be positive or zero. In other words, the problem requires five additional constraints to ensure that none of the changing cells go below zero.

The changing cells are C5:C9, and the goal is to maximize the total yield in cell D12. Starting values of 1,000,000 have been entered in the changing cells. When you run Solver with these parameters, it produces the solution that is shown in Figure 27-21, which has a total yield of 9.25 percent.

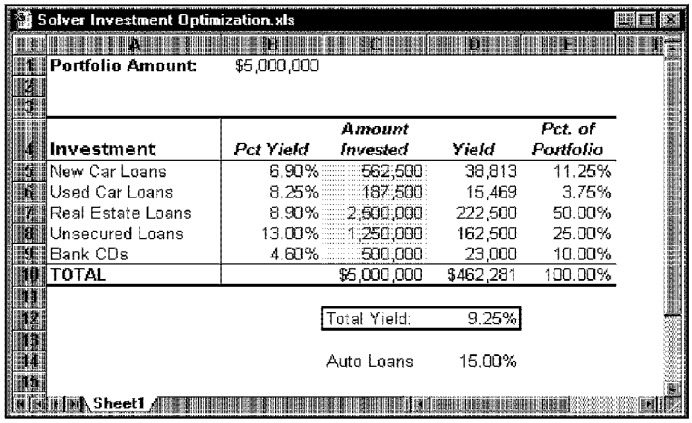


Figure 27-21: The results of the portfolio optimization.

In this example, the starting values of the changing cells are very important. For example, if you use smaller numbers as the starting values (such as 10) and rerun Solver, you find that it doesn't do as well. In fact, it produces a total yield of only 8.35 percent. This demonstrates that you can't always trust Solver to arrive at the optimal solution with one try — even when the Solver Results dialog box tells you that *All constraints and optimality conditions are satisfied.* Usually, the best approach is to use starting values that are as close as possible to the final solution.

The best advice? Make sure that you understand Solver well before you entrust it with helping you make major decisions. Try different starting values, and adjust the options to see whether Solver can do better.

Summary

This chapter discusses two Excel commands: Tools • Goal Seek and Tools • Solver. The latter command is available only if the Solver add-in is installed. Goal seeking is used to determine the value in a single input cell that produces a result that you want in a formula cell. Solver determines values in multiple input cells that produce a result that you want, given certain constraints. Using Solver can be challenging, because it has many options and the result that it produces isn't always the best one.

CHAPTER 28

Analyzing Data with Analysis ToolPak

In This Chapter:

Although spreadsheets such as Excel are designed primarily with business users in mind, these products can be found in other disciplines, including education, research, statistics, and engineering. One way that Excel addresses these nonbusiness users is with its Analysis ToolPak add-in. Many of the features and functions in the Analysis ToolPak are valuable for business applications as well.

The Analysis ToolPak: An Overview

The Analysis ToolPak is an add-in that provides analytical capability that normally is not available. The Analysis ToolPak consists of two parts:

- Analytical procedures
- Additional worksheet functions

These analysis tools offer many features that may be useful to those in the scientific, engineering, and educational communities—not to mention business users whose needs extend beyond the normal spreadsheet fare.

This section provides a quick overview of the types of analyses that you can perform with the Analysis ToolPak. Each of the following tools are discussed in detail in the course of this chapter:

- Analysis of variance (three types)
- Correlation
- Covariance

- Descriptive statistics
- Exponential smoothing
- F-test
- Fourier analysis
- Histogram
- Moving average
- Random number generation
- Rank and percentile
- Regression
- Sampling
- t-test (three types)
- z-test

As you can see, the Analysis ToolPak add-in brings a great deal of new functionality to Excel. These procedures have limitations, however, and in some cases, you may prefer to create your own formulas to do some calculations.

Besides the procedures just listed, the Analysis ToolPak provides many additional worksheet functions. These functions cover mathematics, engineering, unit conversions, financial analysis, and dates. These functions are listed at the end of the chapter.

Using the Analysis ToolPak

This section discusses the two components of the Analysis ToolPak: its tools and its functions.

Using the Analysis Tools

The procedures in the Analysis ToolPak add-in are relatively straightforward. To use any of these tools, you select Tools • Data Analysis, which displays the dialog box shown in Figure 28-1. Scroll through the list until you find the analysis tool that you want to use and then click OK. Excel displays a new dialog box that's specific to the procedure that you select.

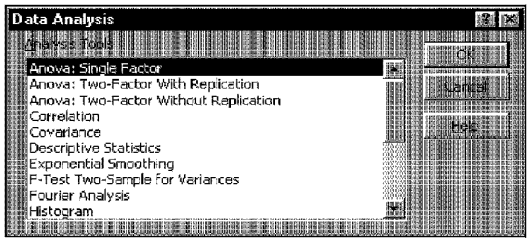


Figure 28-1: The Data Analysis dialog box enables you to select the tool in which you're interested.

Usually, you need to specify one or more input ranges, plus an output range (one cell is sufficient). Alternatively, you can choose to place the results on a new worksheet or in a new workbook. The procedures vary in the amount of additional information that is required. In many dialog boxes, you may be able to indicate whether your data range includes labels. If so, you can specify the entire range, including the labels, and indicate to Excel that the first column (or row) contains labels. Excel then uses these labels in the tables that it produces. Most tools also provide different output options that you can select, based on your needs.

In some cases, the procedures produce their results by using formulas. Consequently, you can change your data, and the results update automatically. In other procedures, Excel stores the results as values, so if you change your data, the results don't reflect your changes. Make sure that you understand what Excel is doing.

Using the Analysis ToolPak Functions

After you install the Analysis ToolPak, you have access to all the additional functions (which are described fully in the online Help system). You access these functions just like any other functions, and they appear in the Function Wizard dialog box, intermixed with Excel's standard functions.

If you plan to share worksheets that use these functions, make sure that the other user has access to the add-in functions. If the other user doesn't install the Analysis ToolPak add-in, formulas that use any of the Analysis ToolPak functions will return #VALUE.

The Analysis ToolPak Tools

This section describes each tool and provides an example. Space limitations prevent a discussion of every available option in these procedures. However, if you need to use some of these advanced analysis tools, then you probably already know how to use most of the options not covered here.

The Analysis of Variance Tool

Analysis of variance is a statistical test that determines whether two or more samples were drawn from the same population. Using tools in the Analysis ToolPak, you can perform three types of analysis of variance:

- **Single-factor:** A one-way analysis of variance, with only one sample for each group of data.
- **Two-factor with replication:** A two-way analysis of variance, with multiple samples (or replications) for each group of data.
- **Two-factor without replication:** A two-way analysis of variance, with a single sample (or replication) for each group of data.

Figure 28-2 shows the dialog box for a single-factor analysis of variance. Alpha represents the statistical confidence level for the test.

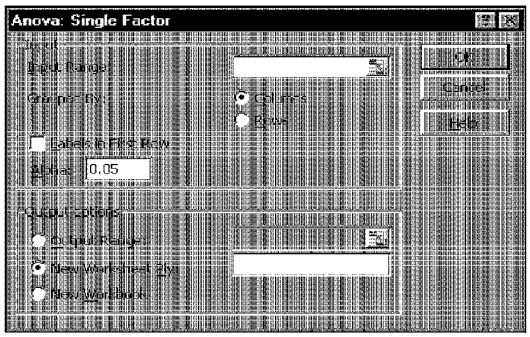
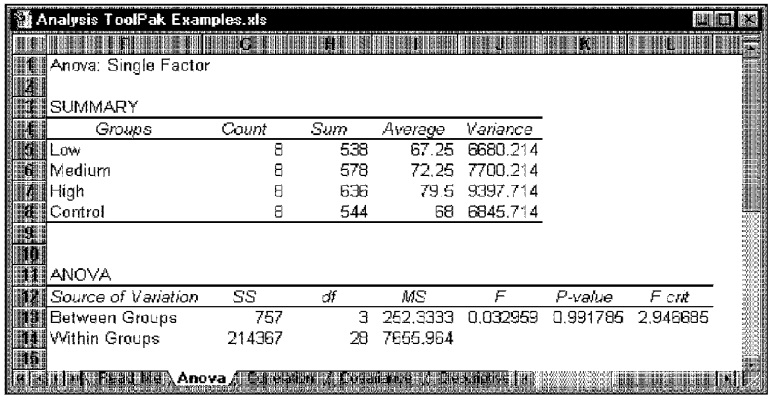


Figure 28-2: Specifying parameters for a single-factor analysis of variance.

Figure 28-3 shows the results of an analysis of variance. The output for this test consists of the means and variances for each of the four samples, the value of F, the critical value of F, and the significance of F (P-value). Because the probability is greater than the Alpha value, the conclusion is that the samples were drawn from the same population.



Analysis ToolPak Examples.xls

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Low	8	538	67.25	8680.214
Medium	8	578	72.25	7700.214
High	8	636	79.5	9397.714
Control	8	544	68	6845.714

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	757	3	252.3333	0.032959	0.991786	2.948685
Within Groups	214367	28	7655.964			

Read Me | Anova | Correlation | Covariance | Descriptive

Figure 28-3: The results of the analysis of variance.

The Correlation Tool

Correlation is a widely used statistic that measures the degree to which two sets of data vary together. For example, if higher values in one data set are typically associated with higher values in the second data set, the two data sets have a positive correlation. The degree of correlation is expressed as a coefficient that ranges from –1.0 (a perfect negative correlation) to +1.0 (a perfect positive correlation). A correlation coefficient of 0 indicates that the two variables are not correlated.

Figure 28-4 shows the Correlation dialog box. Specify the input range, which can include any number of variables, arranged in rows or columns.

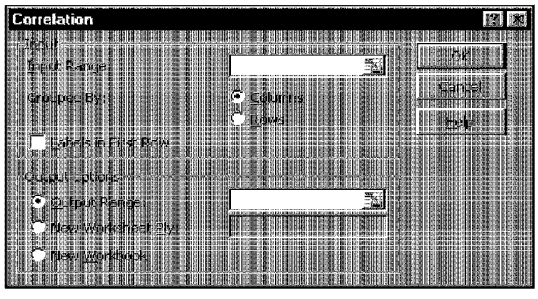
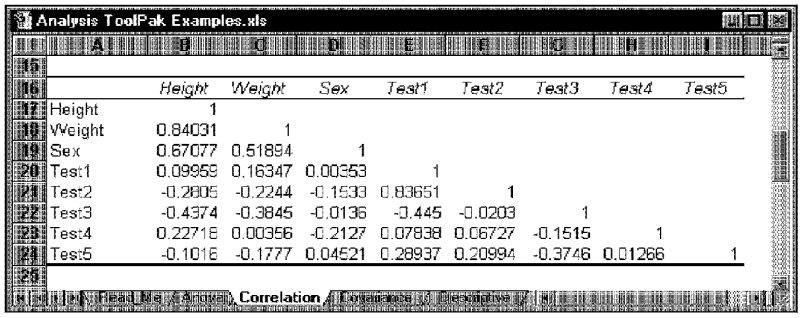


Figure 28-4: The Correlation dialog box.

Figure 28-5 shows the results of a correlation analysis for eight variables. The output consists of a correlation matrix that shows the correlation coefficient for each variable paired with every other variable.


Analysis ToolPak Examples.xls

	Height	Weight	Sex	Test1	Test2	Test3	Test4	Test5
Height	1							
Weight	0.84031	1						
Sex	0.67077	0.51894	1					
Test1	0.09959	0.16347	0.00353	1				
Test2	-0.2805	-0.2244	-0.1533	0.83651	1			
Test3	-0.4374	-0.3845	-0.0136	-0.445	-0.0203	1		
Test4	0.22718	0.00356	-0.2127	0.07838	0.06727	-0.1515	1	
Test5	-0.1016	-0.1777	0.04521	0.28937	0.20994	-0.3746	0.01266	1

Read Me | Anova | Correlation | Covariance | Descriptive

Figure 28-5: The results of a correlation analysis.

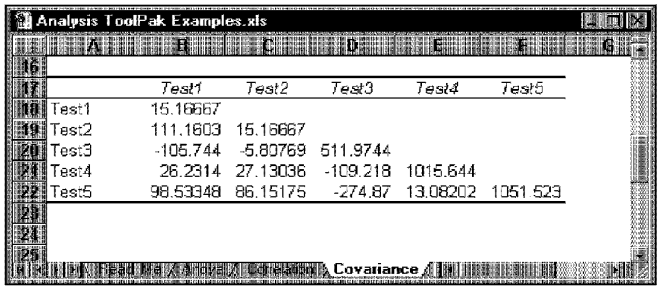

Note

Notice that the resulting correlation matrix doesn't use formulas to calculate the results. Therefore, if any data changes, the correlation matrix isn't valid. You can use Excel's CORREL function to create a correlation matrix that changes automatically when you change data.

The Covariance Tool

The Covariance tool produces a matrix that is similar to the one generated by the Correlation tool. *Covariance*, like correlation, measures the degree to which two variables vary together. Specifically, covariance is the average of the product of the deviations of each data point pair from their respective means.

Figure 28-6 shows a covariance matrix. Notice that the values along the diagonal (where the variables are the same) are the variances for the variable.


Analysis ToolPak Examples.xls

	Test1	Test2	Test3	Test4	Test5
Test1	15.16667				
Test2	111.1603	15.16667			
Test3	-105.744	-5.80769	511.9744		
Test4	26.2314	27.13036	-109.218	1015.644	
Test5	98.53348	86.15175	-274.87	13.08202	1051.523

Read Me | Anova | Correlation | Covariance

Figure 28-6: The results of a covariance analysis.

You can use the COVAR function to create a covariance matrix that uses formulas. The values that are generated by the Analysis ToolPak are *not* the same values that you would get if you used the COVAR function.

The Descriptive Statistics Tool

This tool produces a table that describes your data with some standard statistics. It uses the dialog box that is shown in Figure 28-7. The Kth Largest option and Kth Smallest option each display the data value that corresponds to a rank that you specify. For example, if you check Kth Largest and specify a value of 2, the output shows the second-largest value in the input range (the standard output already includes the minimum and maximum values).

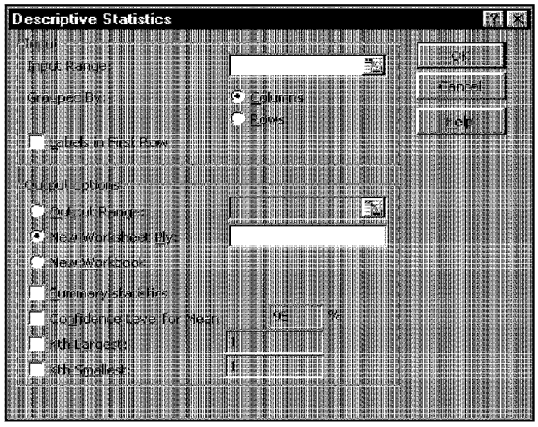
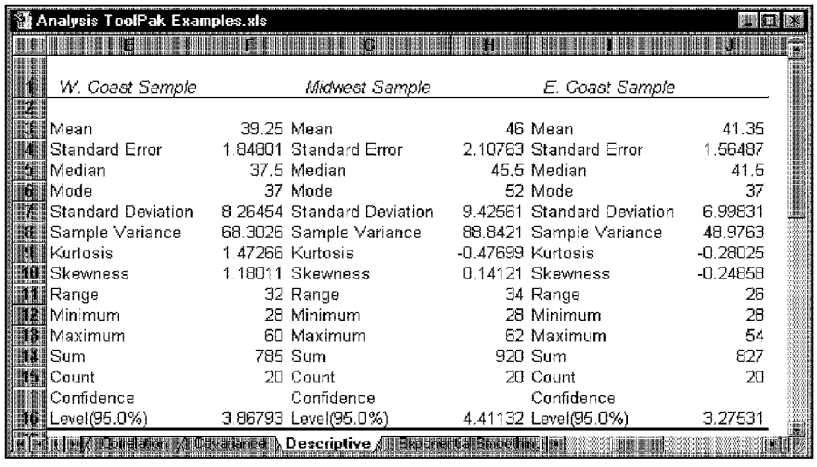


Figure 28-7: The Descriptive Statistics dialog box.

Sample output for the Descriptive Statistics tool appears in Figure 28-8. This example has three groups. Because the output for this procedure consists of values (not formulas), you should use this procedure only when you're certain that your data isn't going to change; otherwise, you will need to re-execute the procedure. You can generate all of these statistics by using formulas.



Analysis ToolPak Examples.xls

	W. Coast Sample		Midwest Sample		E. Coast Sample	
2						
3	Mean	39.25	Mean	46	Mean	41.35
4	Standard Error	1.84801	Standard Error	2.10763	Standard Error	1.56487
5	Median	37.5	Median	45.5	Median	41.5
6	Mode	37	Mode	52	Mode	37
7	Standard Deviation	8.26454	Standard Deviation	9.42561	Standard Deviation	6.99831
8	Sample Variance	68.3026	Sample Variance	88.8421	Sample Variance	48.9763
9	Kurtosis	1.47266	Kurtosis	-0.47699	Kurtosis	-0.28025
10	Skewness	1.18011	Skewness	0.14121	Skewness	-0.24858
11	Range	32	Range	34	Range	26
12	Minimum	28	Minimum	28	Minimum	28
13	Maximum	60	Maximum	62	Maximum	54
14	Sum	785	Sum	920	Sum	827
15	Count	20	Count	20	Count	20
	Confidence		Confidence		Confidence	
16	Level(95.0%)	3.86793	Level(95.0%)	4.41132	Level(95.0%)	3.27531

Descriptive

Figure 28-8: Output from the Descriptive Statistics tool.

The Exponential Smoothing Tool

Exponential smoothing is a technique for predicting data that is based on the previous data point and the previously predicted data point. You can specify the *damping factor* (also known as a *smoothing constant*), which can range from 0 to 1. This determines the relative weighting of the previous data point and the previously predicted data point. You also can request standard errors and a chart.

The exponential smoothing procedure generates formulas that use the damping factor that you specify. Therefore, if the data changes, Excel updates the formulas. Figure 28-9 shows sample output from the Exponential Smoothing tool.

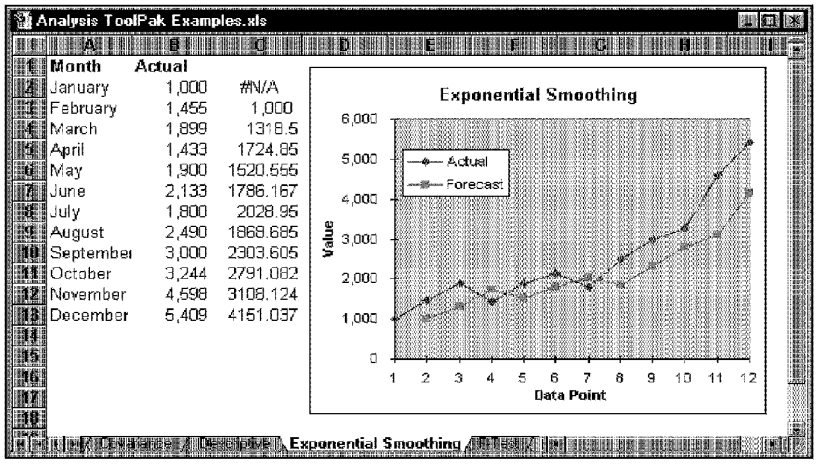


Figure 28-9: Output from the Exponential Smoothing tool.

The F-Test (Two-Sample Test for Variance) Tool

The *F-test* is a commonly used statistical test that enables you to compare two population variances. Figure 28-10 shows the dialog box for this tool.

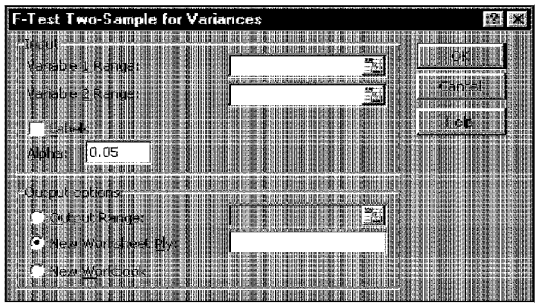


Figure 28-10: The F-Test dialog box.

The output for this test consists of the means and variances for each of the two samples, the value of F, the critical value of F, and the significance of F. Sample output appears in Figure 28-11.

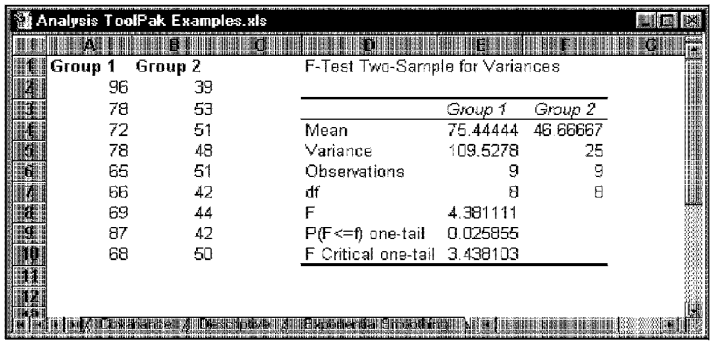


Figure 28-11: Sample output for the F-test.

The Fourier Analysis Tool

This tool performs a "fast Fourier" transformation of a range of data. Using the Fourier Analysis tool, you can transform a range limited to the following sizes: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, or 1,024 data points. This procedure accepts and generates complex numbers, which are represented as labels (not values).

The Histogram Tool

This procedure is useful for producing data distributions and histogram charts. It accepts an input range and a bin range. A *bin* range is a range of values that specifies the limits for each column of the histogram. If you omit the bin range, Excel creates ten equal-interval bins for you. The size of each bin is determined by a formula of the following form:

```
=(MAX(input_range)-MIN(input_range))/10
```

The Histogram dialog box appears in Figure 28-12. As an option, you can specify that the resulting histogram be sorted by frequency of occurrence in each bin.

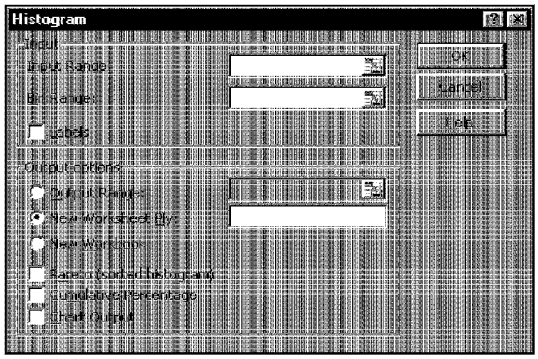


Figure 28-12: The Histogram tool enables you to generate distributions and graphical output.

If you specify the Pareto (sorted histogram) option, the bin range must contain values and can't contain formulas. If formulas appear in the bin range, Excel doesn't sort properly, and your worksheet displays error values.

Figure 28-13 shows a chart generated from this procedure. The Histogram tool doesn't use formulas, so if you change any of the input data, you need to repeat the histogram procedure to update the results.

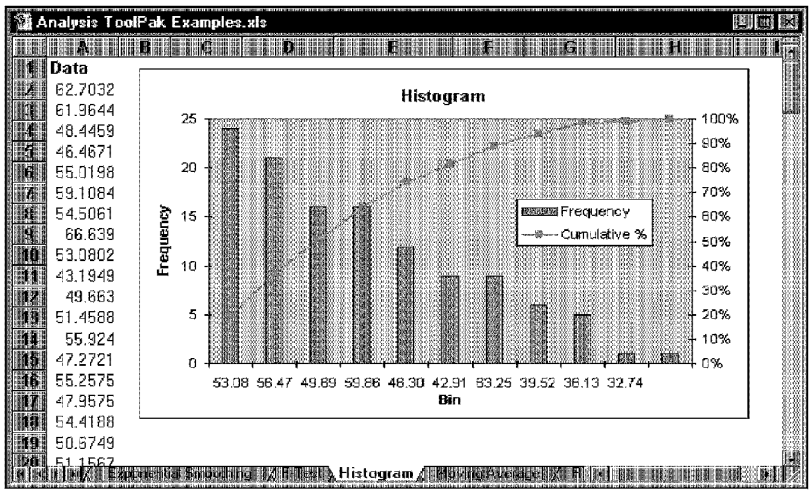


Figure 28-13: Output from the Histogram tool.

The Moving Average Tool

The Moving Average tool helps you to smooth out a data series that has a lot of variability. This is best done in conjunction with a chart. Excel does the smoothing by computing a moving average of a specified number of values. In many cases, a moving average enables you to spot trends that otherwise would be obscured by noise in the data.

Figure 28-14 shows the Moving Average dialog box. You can, of course, specify the number of values that you want Excel to use for each average. If you place a check in the Standard Errors check box, Excel calculates standard errors and places formulas for these calculations next to the moving average formulas. The standard error values indicate the degree of variability between the actual values and the calculated moving averages. When you close this dialog box, Excel creates formulas that reference the input range that you specify.

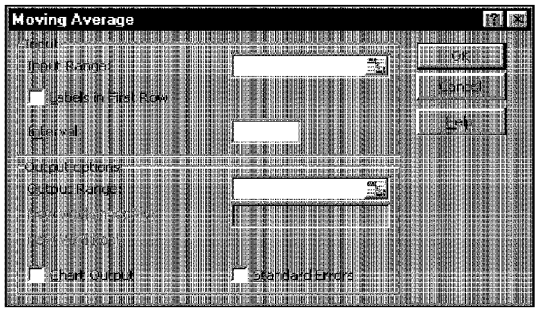


Figure 28-14: The Moving Average dialog box.

Figure 28-15 shows the results of using this tool. The first few cells in the output are #N/A because not enough data points exist to calculate the average for these initial values.

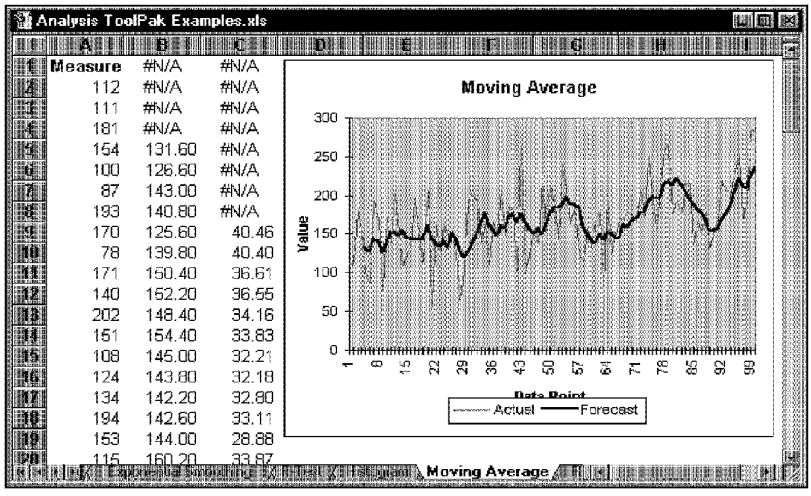


Figure 28-15: Output from the Moving Average tool.

The Random Number Generation Tool

Although Excel contains a built-in function to calculate random numbers, the Random Number Generation tool is much more flexible, because you can specify what type of distribution you want the random numbers to have. Figure 28-16 shows the Random Number Generation dialog box. The Parameters box varies, depending on the type of distribution that you select.

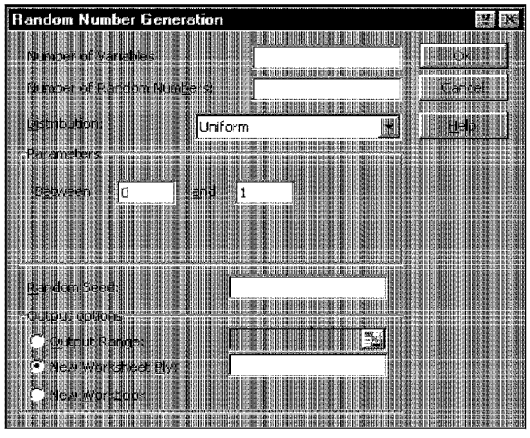


Figure 28-16: This dialog box enables you to generate a wide variety of random numbers.

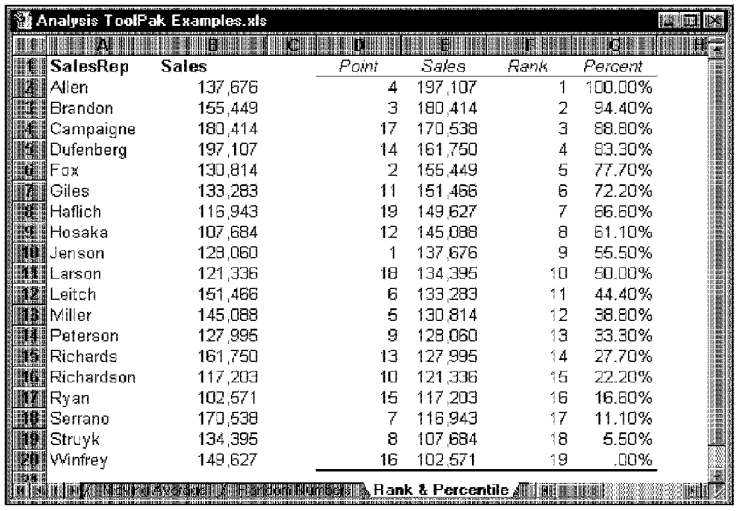
The Number of Variables refers to the number of columns that you want, and the Number of Random Numbers refers to the number of rows that you want. For example, if you want 200 random numbers arranged in 10 columns of 20 rows, you specify 10 and 20, respectively, in these text boxes.

The Random Seed box enables you to specify a starting value that Excel uses in its random number-generating algorithm. Usually, you leave this blank. If you want to generate the same random number sequence, however, you can specify a seed between 1 and 32,767 (integer values only). You can create the following types of distributions:

- **Uniform:** Every random number has an equal chance of being selected. You specify the upper and lower limits.
- **Normal:** The random numbers correspond to a normal distribution. You specify the mean and standard deviation of the distribution.
- **Bernoulli:** The random numbers are either 0 or 1, determined by the probability of success that you specify.
- **Binomial:** This returns random numbers based on a Bernoulli distribution over a specific number of trials, given a probability of success that you specify.
- **Poisson:** This option generates values in a Poisson distribution. This is characterized by discrete events that occur in an interval, where the probability of a single occurrence is proportional to the size of the interval. The *lambda* parameter is the expected number of occurrences in an interval. In a Poisson distribution, lambda is equal to the mean, which also is equal to the variance.
- **Patterned:** This option doesn't generate random numbers. Rather, it repeats a series of numbers in steps that you specify.
- **Discrete:** This option enables you to specify the probability that specific values are chosen. It requires a two-column input range; the first column holds the values, and the second column holds the probability of each value being chosen. The sum of the probabilities in the second column must equal 100 percent.

The Rank and Percentile Tool

This tool creates a table that shows the ordinal and percentile ranking for each value in a range. Figure 28-17 shows the results of this procedure. You can also generate ranks and percentiles by using formulas.



Analysis ToolPak Examples.xls

	A	B	C	D	E	F	G	H
1	SalesRep	Sales		*Point*	*Sales*	*Rank*	*Percent*	
2	Allen	137,676		4	197,107	1	100.00%	
3	Brandon	155,449		3	180,414	2	94.40%	
4	Campaigne	180,414		17	170,538	3	88.80%	
5	Dufenberg	197,107		14	161,750	4	83.30%	
6	Fox	130,814		2	155,449	5	77.70%	
7	Giles	133,283		11	151,466	6	72.20%	
8	Haflich	116,943		19	149,627	7	66.60%	
9	Hosaka	107,684		12	145,088	8	61.10%	
10	Jenson	129,060		1	137,676	9	55.50%	
11	Larson	121,336		18	134,395	10	50.00%	
12	Leitch	151,466		6	133,283	11	44.40%	
13	Miller	145,088		5	130,814	12	38.80%	
14	Peterson	127,995		9	129,060	13	33.30%	
15	Richards	161,750		13	127,995	14	27.70%	
16	Richardson	117,203		10	121,336	15	22.20%	
17	Ryan	102,571		15	117,203	16	16.60%	
18	Serrano	170,538		7	116,943	17	11.10%	
19	Struyk	134,395		8	107,684	18	5.50%	
20	Winfrey	149,627		16	102,571	19	.00%	

Rank & Percentile

Figure 28-17: Output from the rank and percentile procedure.

The Regression Tool

The Regression tool calculates a regression analysis from worksheet data. Use regression to analyze trends, forecast the future, build predictive models, and, often, to make sense out of a series of seemingly unrelated numbers.

Regression analysis enables you to determine the extent to which one range of data (the dependent variable) varies as a function of the values of one or more other ranges of data (the independent variables). This relationship is expressed mathematically, using values that Excel calculates. You can use these calculations to create a mathematical model of the data and predict the dependent variable by using different values of one or more independent variables. This tool can perform simple and multiple linear regressions and calculate and standardize residuals automatically.

Figure 28-18 shows the Regression dialog box.

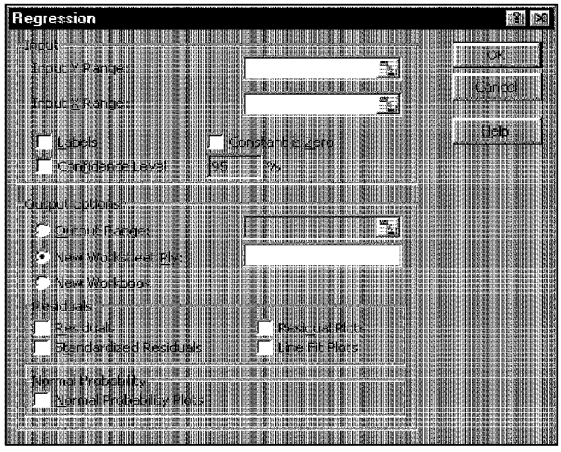
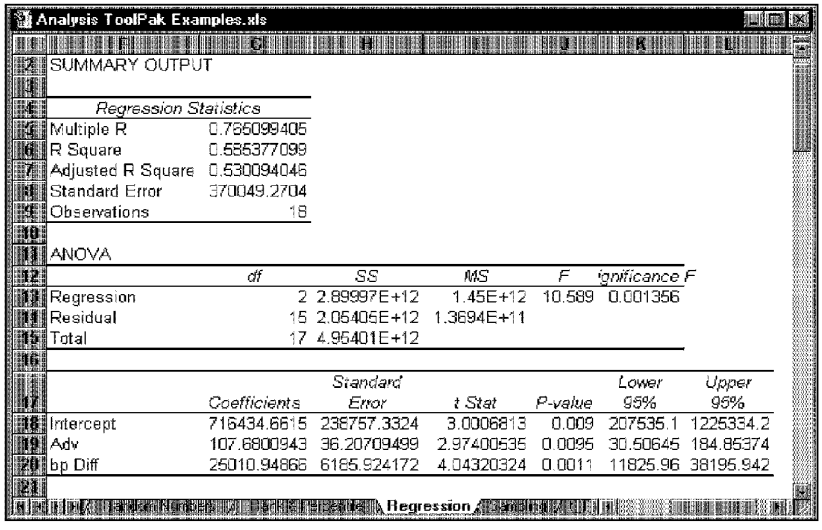


Figure 28-18: The Regression dialog box.

As you can see, the Regression dialog box offers many options:

- **Input Y Range:** The range that contains the dependent variable.
- **Input X Range:** One or more ranges that contain independent variables.
- **Confidence Level:** The confidence level for the regression.
- **Constant is Zero:** If checked, this forces the regression to have a constant of zero (which means that the regression line passes through the origin; when the X values are 0, the predicted Y value is 0).
- **Residuals:** These options specify whether to include residuals in the output. *Residuals* are the differences between observed and predicted values.
- **Normal Probability:** This generates a chart for normal probability plots.

The results of a regression analysis appear in Figure 28-19. If you understand regression analysis, the output from this procedure is familiar.



Analysis ToolPak Examples.xls

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.735099405
R Square	0.585377099
Adjusted R Square	0.530094046
Standard Error	370049.2704
Observations	18

ANOVA

	df	SS	MS	F	ignificance F
Regression	2	2.89997E+12	1.45E+12	10.589	0.001356
Residual	15	2.05406E+12	1.3694E+11		
Total	17	4.95401E+12			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	716434.6615	238757.3324	3.0006813	0.009	207535.1	1225334.2
Adv	107.6800943	36.20709499	2.97400535	0.0095	30.50645	184.85374
bp Diff	25010.94866	6165.924172	4.04320324	0.0011	11825.96	38195.942

Regression

Figure 28-19: Sample output from the Regression tool.

The Sampling Tool

The Sampling tool generates a random sample from a range of input values. The Sampling tool can help you to work with a large database by creating a subset of it. The Sampling dialog box appears in Figure 28-20. This procedure has two options: periodic and random. If you choose a periodic sample, Excel selects every *n*th value from the input range, where *n* equals the period that you specify. With a random sample, you simply specify the size of the sample you want Excel to select, and every value has an equal probability of being chosen.

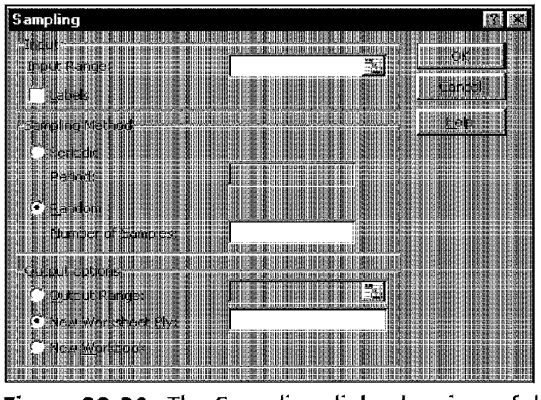


Figure 28-20: The Sampling dialog box is useful for selecting random samples.

The t-Test Tool

Use the *t-test* to determine whether a statistically significant difference exists between two small samples. The Analysis ToolPak can perform three types of t-tests:

- **Paired two-sample for means:** For paired samples in which you have two observations on each subject (such as a pretest and a posttest). The samples must be the same size.
- **Two-sample assuming equal variances:** For independent, rather than paired, samples. Excel assumes equal variances for the two samples.
- **Two-sample assuming unequal variances:** For independent, rather than paired, samples. Excel assumes unequal variances for the two samples.

Figure 28-21 shows the dialog box for the Paired Two Sample for Means t-test. You specify the significance level (alpha) and the hypothesized difference between the two means (that is, the *null hypothesis*).

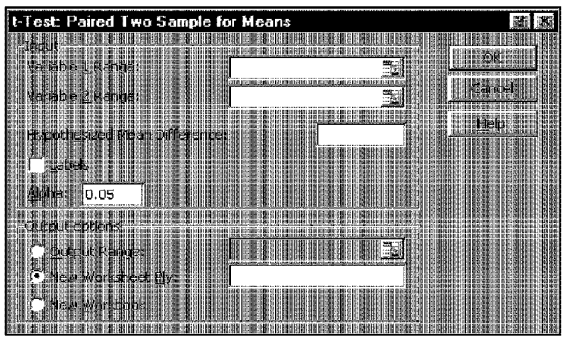
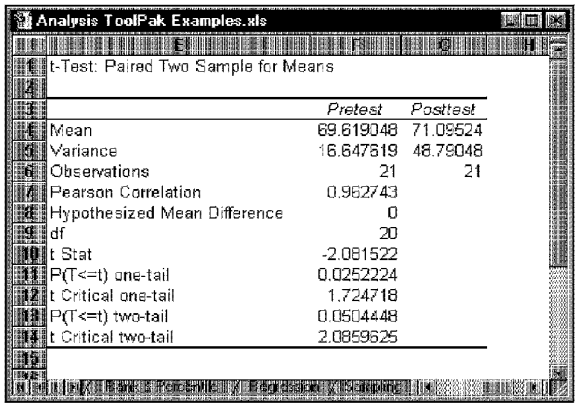


Figure 28-21: The paired t-Test dialog box.

Figure 28-22 shows sample output for the paired two sample for means t-test. Excel calculates *t* for both a one-tailed and two-tailed test.

The z-Test (Two-Sample Test for Means) Tool

The t-test is used for small samples; the z-test is used for larger samples or populations. You must know the variances for both input ranges.



Analysis ToolPak Examples.xls

t-Test: Paired Two Sample for Means

	Pretest	Posttest
Mean	69.619048	71.09524
Variance	16.647619	48.79048
Observations	21	21
Pearson Correlation	0.962743	
Hypothesized Mean Difference	0	
df	20	
t Stat	-2.081522	
P(T<=t) one-tail	0.0252224	
t Critical one-tail	1.724718	
P(T<=t) two-tail	0.0504448	
t Critical two-tail	2.0859625	

Figure 28-22: Results of a paired two sample for means t-test.

Analysis ToolPak Worksheet Functions

This section lists the worksheet functions that are available in the Analysis ToolPak. For specific information about the arguments required, click the Help button in the Paste Function dialog box.

Remember, the Analysis ToolPak add-in must be installed to use these functions in your worksheet. If you use any of these functions in a workbook that you distribute to a colleague, make clear to your colleague that the workbook requires the Analysis ToolPak.

These functions appear in the Paste Function dialog box in the following categories:

- Date & Time
- Engineering (a new category that appears when you install the Analysis ToolPak)
- Financial
- Information
- Math & Trig

Date & Time Category

Table 28-1 lists the Analysis ToolPak worksheet functions that you'll find in the Date & Time category.

Table 28-1
Date & Time Category Functions

Function	*Purpose*
EDATE	Returns the serial number of the date that is the indicated number of months before or after the start date
EOMONTH	Returns the serial number of the last day of the month before or after a specified number of months
NETWORKDAYS	Returns the number of whole workdays between two dates
WEEKNUM	Returns the week number in the year
WORKDAY	Returns the serial number of the date before or after a specified number of workdays
YEARFRAC	Returns the year fraction representing the number of whole days between start_date and end_date

Engineering Category

Table 28-2 lists the Analysis ToolPak worksheet functions that you'll find in the Engineering category. Some of these functions are quite useful for nonengineers as well. For example, the CONVERT function converts a wide variety of measurement units.

Table 28-2
Engineering Category Functions

Function	*Purpose*
BESSELI	Returns the modified Bessel function In(x)
BESSELJ	Returns the Bessel function Jn(x)
BESSELK	Returns the modified Bessel function Kn(x)
BESSELY	Returns the Bessel function Yn(x)
BIN2DEC	Converts a binary number to decimal
BIN2HEX	Converts a binary number to hexadecimal
BIN2OCT	Converts a binary number to octal
COMPLEX	Converts real and imaginary coefficients into a complex number
CONVERT	Converts a number from one measurement system to another

Function	*Purpose*
DEC2BIN	Converts a decimal number to binary
DEC2HEX	Converts a decimal number to hexadecimal
DEC2OCT	Converts a decimal number to octal
DELTA	Tests whether two numbers are equal
ERF	Returns the error function
ERFC	Returns the complementary error function
FACTDOUBLE	Returns the double factorial of a number
GESTEP	Tests whether a number is greater than a threshold value
HEX2BIN	Converts a hexadecimal number to binary
HEX2DEC	Converts a hexadecimal number to decimal
HEX2OCT	Converts a hexadecimal number to octal
IMABS	Returns the absolute value (modulus) of a complex number
IMAGINARY	Returns the imaginary coefficient of a complex number
IMARGUMENT	Returns the argument q, an angle expressed in radians
IMCONJUGATE	Returns the complex conjugate of a complex number
IMCOS	Returns the cosine of a complex number
IMDIV	Returns the quotient of two complex numbers
IMEXP	Returns the exponential of a complex number
IMLN	Returns the natural logarithm of a complex number
IMLOG10	Returns the base-10 logarithm of a complex number
IMLOG2	Returns the base-2 logarithm of a complex number
IMPOWER	Returns a complex number raised to an integer power
IMPRODUCT	Returns the product of two complex numbers
IMREAL	Returns the real coefficient of a complex number
IMSIN	Returns the sine of a complex number
IMSQRT	Returns the square root of a complex number
IMSUB	Returns the difference of two complex numbers
IMSUM	Returns the sum of complex numbers
OCT2BIN	Converts an octal number to binary
OCT2DEC	Converts an octal number to decimal
OCT2HEX	Converts an octal number to hexadecimal

Financial Category

Table 28-3 lists the Analysis ToolPak worksheet functions that you'll find in the Financial category.

Table 28-3
Financial Category Functions

Function	*Purpose*
ACCRINT	Returns the accrued interest for a security that pays periodic interest
ACCRINTM	Returns the accrued interest for a security that pays interest at maturity
AMORDEGRC	Returns the prorated linear depreciation of an asset for each accounting period. Similar to the AMORLINC function, except that this function uses a depreciation coefficient that depends on the life of the assets
AMORLINC	Returns the prorated linear depreciation of an asset for each accounting period
COUPDAYBS	Returns the number of days from the beginning of the coupon period to the settlement date
COUPDAYS	Returns the number of days in the coupon period that contain the settlement date
COUPDAYSNC	Returns the number of days from the settlement date to the next coupon date
COUPNCD	Returns the next coupon date after the settlement date
COUPNUM	Returns the number of coupons payable between the settlement date and maturity date
COUPPCD	Returns the previous coupon date before the settlement date
CUMIPMT	Returns the cumulative interest paid between two periods
CUMPRINC	Returns the cumulative principal paid on a loan between two periods
DISC	Returns the discount rate for a security
DOLLARDE	Converts a dollar price, expressed as a fraction, into a dollar price, expressed as a decimal number
DOLLARFR	Converts a dollar price, expressed as a decimal number, into a dollar price, expressed as a fraction
DURATION	Returns the annual duration of a security with periodic interest payments
EFFECT	Returns the effective annual interest rate

Function	*Purpose*
FVSCHEDULE	Returns the future value of an initial principal after applying a series of compound interest rates
INTRATE	Returns the interest rate for a fully invested security
MDURATION	Returns the Macauley modified duration for a security with an assumed par value of $100
NOMINAL	Returns the annual nominal interest rate
ODDFPRICE	Returns the price per $100 face value of a security with an odd first period
ODDFYIELD	Returns the yield of a security with an odd first period
ODDLPRICE	Returns the price per $100 face value of a security with an odd last period
ODDLYIELD	Returns the yield of a security with an odd last period
PRICE	Returns the price per $100 face value of a security that pays periodic interest
PRICEDISC	Returns the price per $100 face value of a discounted security
PRICEMAT	Returns the price per $100 face value of a security that pays interest at maturity
RECEIVED	Returns the amount received at maturity for a fully invested security
TBILLEQ	Returns the bond-equivalent yield for a Treasury bill
TBILLPRICE	Returns the price per $100 face value for a Treasury bill
TBILLYIELD	Returns the yield for a Treasury bill
XIRR	Returns the internal rate of return for a schedule of cash flows
XNPV	Returns the net present value for a schedule of cash flows
YIELD	Returns the yield on a security that pays periodic interest
YIELDDISC	Returns the annual yield for a discounted security (for example, a Treasury bill)
YIELDMAT	Returns the annual yield of a security that pays interest at maturity

Information Category

Table 28-4 lists the two Analysis ToolPak worksheet functions that you'll find in the Information category.

Table 28-4
Information Category Functions

Function	*Purpose*
ISEVEN	Returns TRUE if the number is even
ISODD	Returns TRUE if the number is odd

Math & Trig Category

Table 28-5 lists the Analysis ToolPak worksheet functions that you'll find in the Math & Trig category.

Table 28-5
Math & Trig Category Functions

Function	*Purpose*
GCD	Returns the greatest common divisor
LCM	Returns the least common multiple
MROUND	Returns a number rounded to the desired multiple
MULTINOMIAL	Returns the multinomial of a set of numbers
QUOTIENT	Returns the integer portion of a division
RANDBETWEEN	Returns a random number between the numbers that you specify
SERIESSUM	Returns the sum of a power series based on the formula
SQRTPI	Returns the square root of pi

Summary

This chapter discusses the Analysis ToolPak, an add-in that extends the analytical powers of Excel. It includes 19 analytical procedures and 93 functions. Many of the tools are useful for general business applications, but many are for more specialized uses, such as statistical tests.

Other Topics

PART V

In This Part