- **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
```

Note

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.

Shipping Costs Problem.xls

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

	L.A.	St. Louis	Boston	
Starting Inventory:	400	350	500	
No. Remaining:	130	0	125	
Shipping Costs:	$16,140	$15,000	$24,375	$55,515 Total

SHIPPING

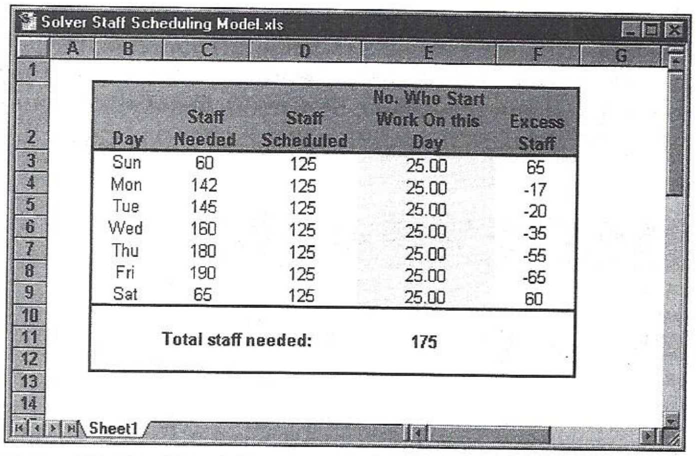
Figure 27-13: The solution that was created by Solver.

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.



Solver Staff Scheduling Model.xls

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.

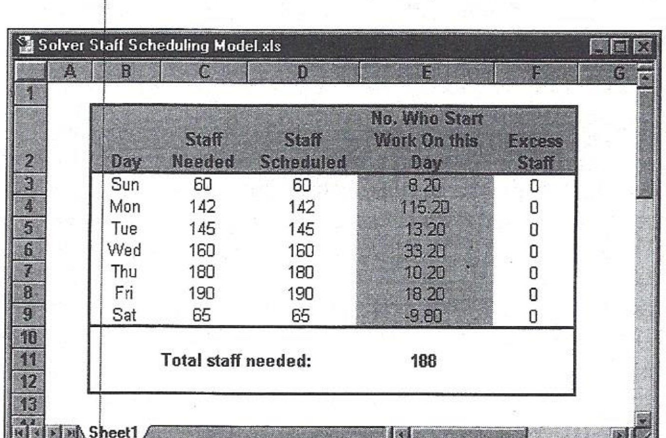
✦ **Total staff needed:** Cell E11 contains a formula that sums the number of people who start on each day. The formula is =SUM(E3:E9). This is the value that Solver minimizes.

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	115.20	0
Tue	145	145	13.20	0
Wed	160	160	33.20	0
Thu	180	180	10.20	0
Fri	190	190	18.20	0
Sat	65	65	-9.80	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.

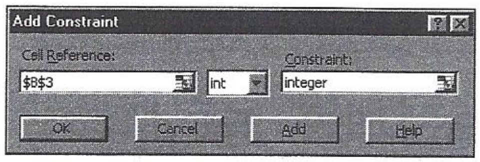


Figure 27-16: With many problems, you have to limit the solution to integers. You can do this by selecting the integer option in the Constraint box of the Add Constraint dialog box.

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.

Tip

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.

	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	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.

Material	Toy A	Toy B	Toy C	Toy D	Toy E	Amt. Avail.	Amt. Used	Amt. Left
	Materials Needed							
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.

Solver Investment Optimization.xls

	A	B	C	D	E
1	**Portfolio Amount:**	$5,000,000			
2					
3					
4	**Investment**	*Pct Yield*	*Amount Invested*	*Yield*	*Pct. of Portfolio*
5	New Car Loans	6.90%	1,000,000	69,000	20.00%
6	Used Car Loans	8.25%	1,000,000	82,500	20.00%
7	Real Estate Loans	8.90%	1,000,000	89,000	20.00%
8	Unsecured Loans	13.00%	1,000,000	130,000	20.00%
9	Bank CDs	4.60%	1,000,000	46,000	20.00%
10	**TOTAL**		$5,000,000	$416,500	100.00%
11					
12			Total Yield:	8.33%	
13					
14			Auto Loans	40.00%	
15					

Sheet1

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.

Solver Investment Optimization.xls

	A	B	C	D	E
1	Portfolio Amount:	$5,000,000			
2					
3					
4	Investment	Pct Yield	Amount Invested	Yield	Pct. of Portfolio
5	New Car Loans	6.90%	562,500	38,813	11.25%
6	Used Car Loans	8.25%	187,500	15,469	3.75%
7	Real Estate Loans	8.90%	2,500,000	222,500	50.00%
8	Unsecured Loans	13.00%	1,250,000	162,500	25.00%
9	Bank CDs	4.60%	500,000	23,000	10.00%
10	TOTAL		$5,000,000	$462,281	100.00%
11					
12			Total Yield:	9.25%	
13					
14			Auto Loans	15.00%	
15					

Sheet1

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.

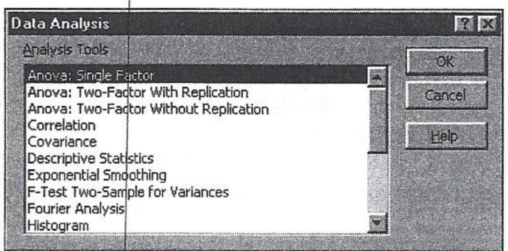


Figure 28-1: The Data Analysis dialog box enables you to select the tool in which you're interested.

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.

Caution

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.

Note

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.

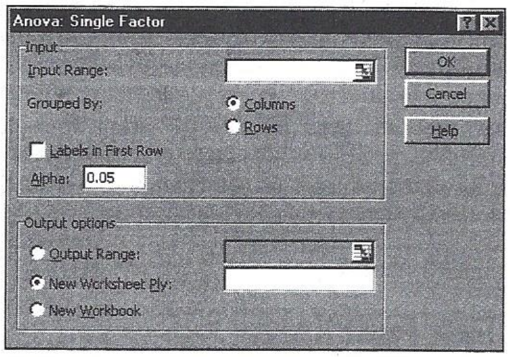
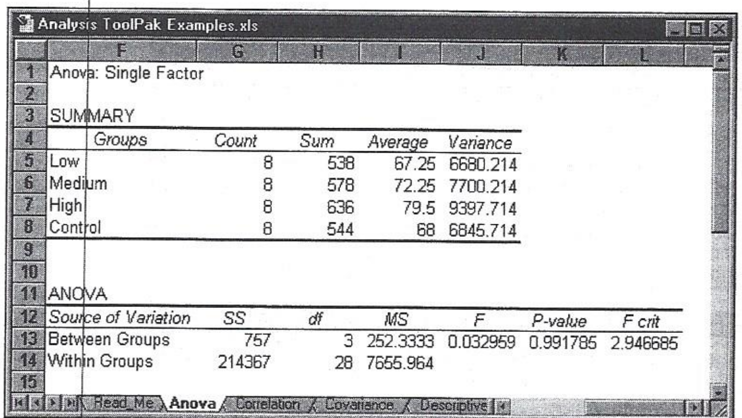


Figure 28-2: Specifying parameters for a single-factor analysis of variance.

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.



Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Low	8	538	67.25	6680.214
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	0.991785	2.946685
Within Groups	214367	28	7655.964			

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.

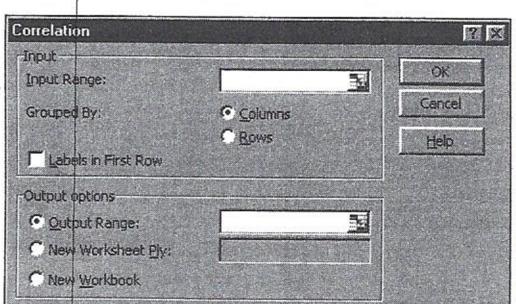


Figure 28-4: The Correlation dialog box.

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	96.53346	86.15175	-274.67	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).

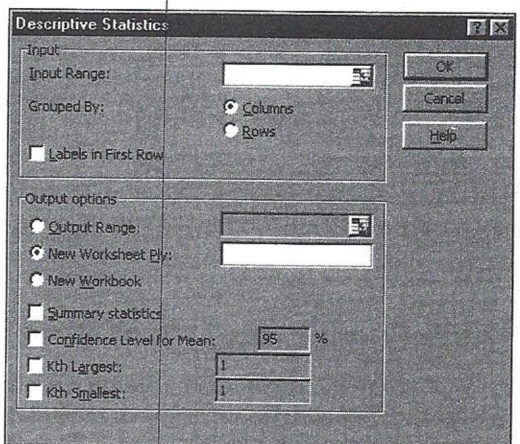


Figure 28-7: The Descriptive Statistics dialog box.

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

	E	F	G	H	I	J
1	*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
16	Confidence Level(95.0%)	3.86793	Confidence Level(95.0%)	4.41132	Confidence Level(95.0%)	3.27531

Correlation / Covariance / **Descriptive** / Exponential Smoothing

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.

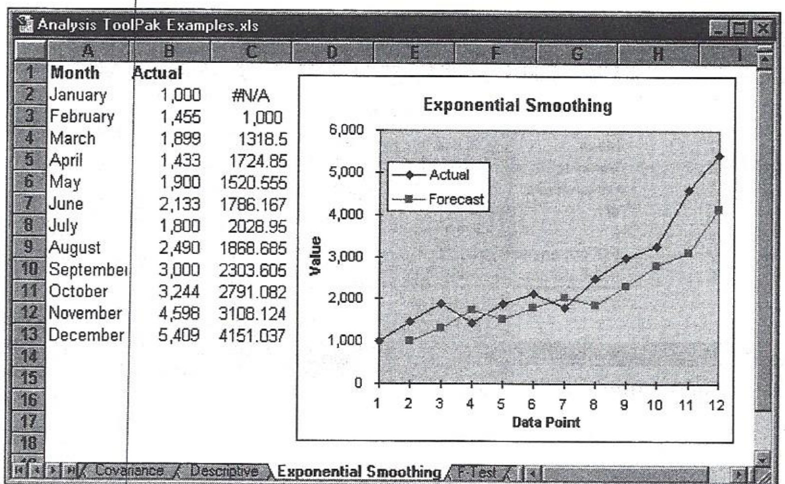


Figure 28-9: Output from the Exponential Smoothing tool.

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.

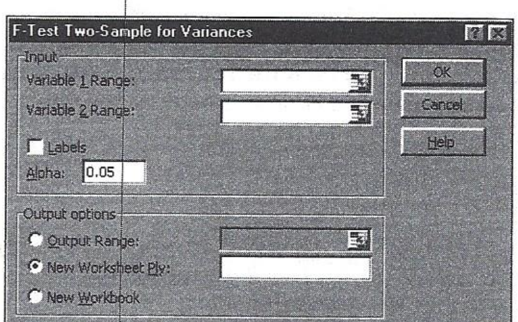
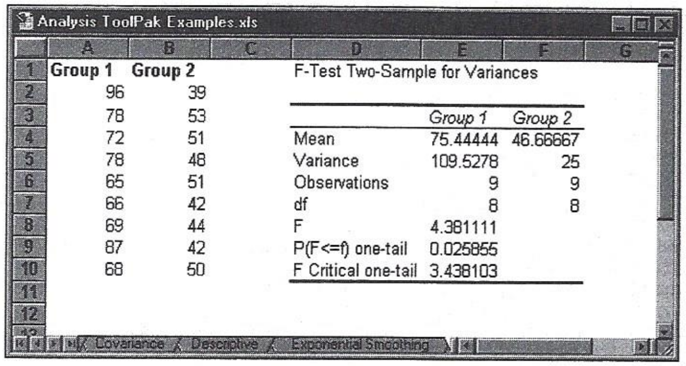


Figure 28-10: The F-Test dialog box.

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.



	A	B	C	D	E	F
1	Group 1	Group 2		F-Test Two-Sample for Variances		
2	96	39				
3	78	53			Group 1	Group 2
4	72	51		Mean	75.44444	46.66667
5	78	48		Variance	109.5278	25
6	65	51		Observations	9	9
7	66	42		df	8	8
8	69	44		F	4.381111	
9	87	42		P(F<=f) one-tail	0.025855	
10	68	50		F Critical one-tail	3.438103	

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.

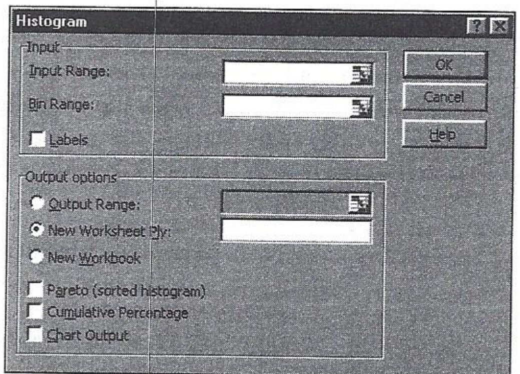


Figure 28-12: The Histogram tool enables you to generate distributions and graphical output.

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.

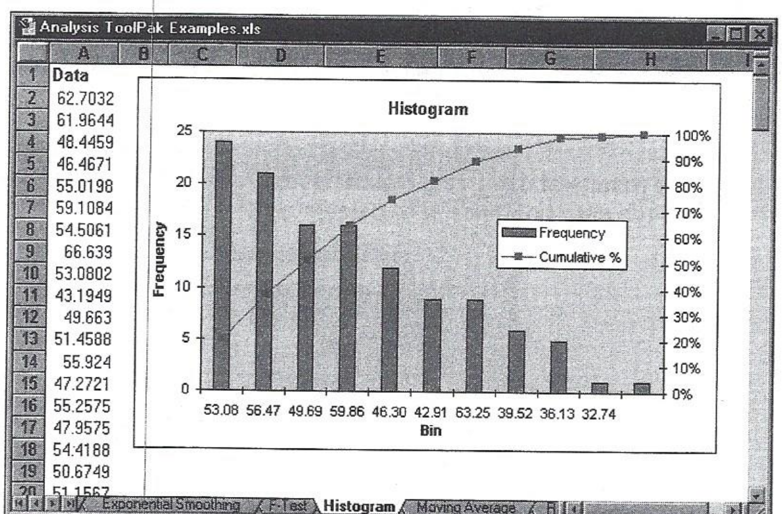


Figure 28-13: Output from the Histogram tool.

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.

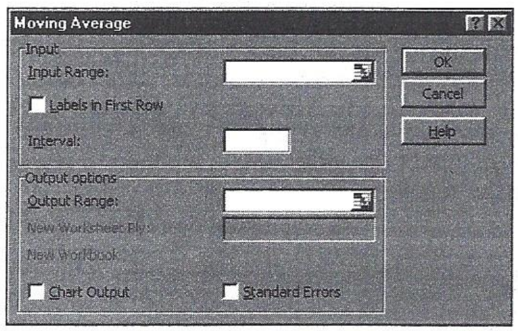


Figure 28-14: The Moving Average dialog box.

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.

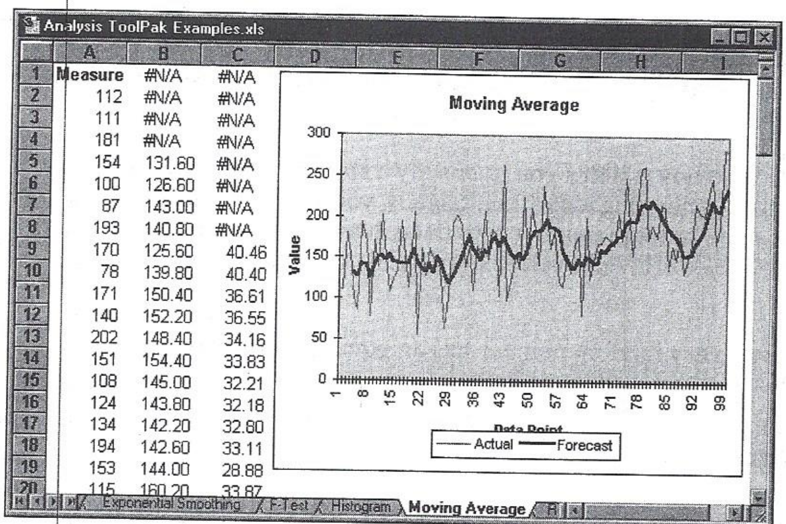


Figure 28-15: Output from the Moving Average tool.

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.

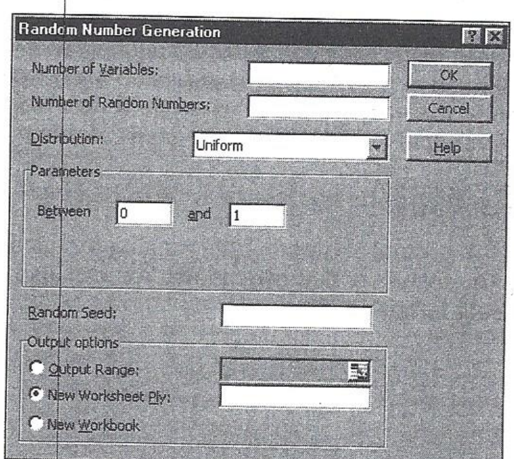


Figure 28-16: This dialog box enables you to generate a wide variety of random numbers.

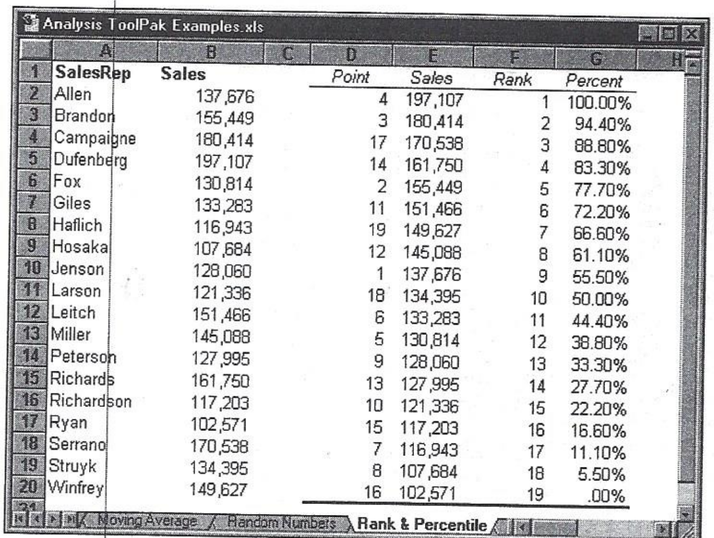
The Number of Variables refers to the number of columns that you want, and the Number of Random Numbers refers to the number of rows that you want. For example, if you want 200 random numbers arranged in 10 columns of 20 rows, you specify 10 and 20, respectively, in these text boxes.

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.



	A	B	C	D	E	F	G
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	128,060		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	128,060	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%

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.

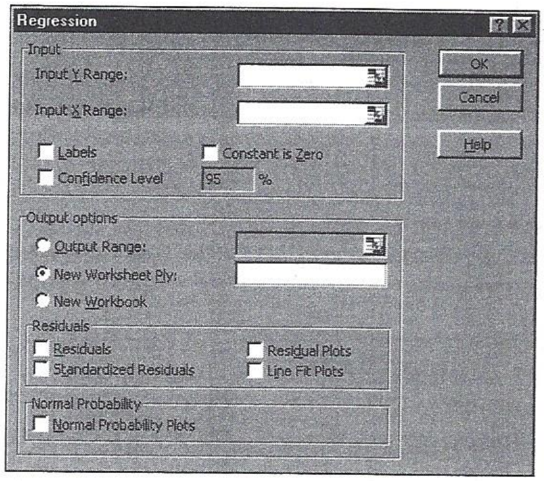


Figure 28-18: The Regression dialog box.

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.

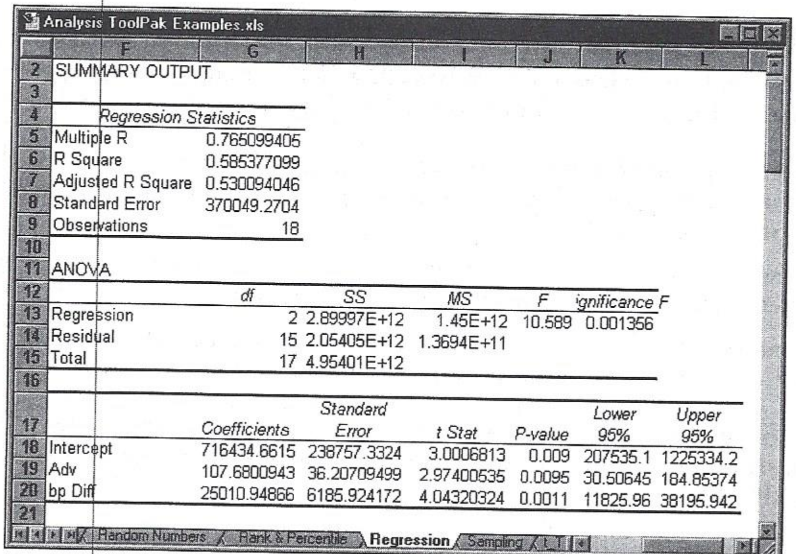


Figure 28-19: Sample output from the Regression tool.

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.

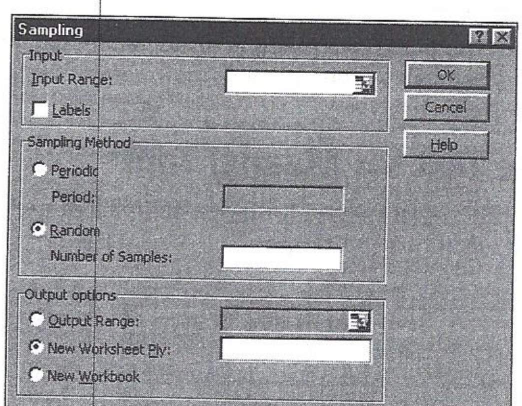


Figure 28-20: The Sampling dialog box is useful for selecting random samples.

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*).

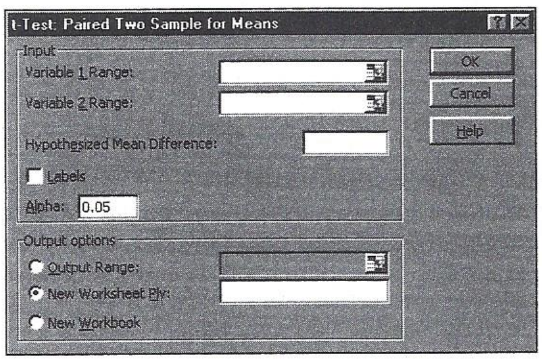


Figure 28-21: The paired t-Test dialog box.

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

	E	F	G
1	t-Test: Paired Two Sample for Means		
2			
3		Pretest	Posttest
4	Mean	69.619048	71.09524
5	Variance	16.647619	48.79048
6	Observations	21	21
7	Pearson Correlation	0.962743	
8	Hypothesized Mean Difference	0	
9	df	20	
10	t Stat	-2.081522	
11	P(T<=t) one-tail	0.0252224	
12	t Critical one-tail	1.724718	
13	P(T<=t) two-tail	0.0504448	
14	t Critical two-tail	2.0859625	
15			

Rank & Percentile / Regression / Sampling

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.

✦ ✦ ✦