CHAPTER 2

Analytics on Spreadsheets

S. Dashkevych/Shutterstock.com

Learning Objectives

After studying this chapter, you will be able to:

- Find buttons and menus in the Excel 2013 ribbon.
- Write correct formulas in an Excel worksheet.
- Apply relative and absolute addressing in Excel formulas.
- Copy formulas from one cell to another or to a range of cells.
- Use Excel features such as split screen, paste special, show formulas, and displaying grid lines and headers in your applications.
- Use basic and advanced Excel functions.
- Use Excel functions for business intelligence queries in databases.

Many commercial software packages are available to facilitate the application of business analytics. Although they often have unique features and capabilities, they can be expensive, generally require advanced training to understand and apply, and may work only on specific computer platforms. Spreadsheet software, on the other hand, is widely used across all areas of business and is standard on nearly every employee's computer. Spreadsheets are an effective platform for manipulating data and developing and solving models; they support powerful commercial add-ins and facilitate communication of results. Spreadsheets provide a flexible modeling environment and are particularly useful when the end user is not the designer of the model. Teams can easily use spreadsheets and understand the logic upon which they are built. Information in spreadsheets can easily be copied from Excel into other documents and presentations. A recent survey identified more than 180 commercial spreadsheet products that support analytics efforts, including data management and reporting, data- and model-driven analytical techniques, and implementation.[1] Many organizations have used spreadsheets extremely effectively to support decision making in marketing, finance, and operations. Some illustrative applications include the following:[2]

- Analyzing supply chains (Hewlett-Packard)
- Determining optimal inventory levels to meet customer service objectives (Procter & Gamble)
- Selecting internal projects (Lockheed Martin Space Systems Company)
- Planning for emergency clinics in response to a sudden epidemic or bioterrorism attack (Centers for Disease Control)
- Analyzing the default risk of a portfolio of real estate loans (Hypo International)
- Assigning medical residents to on-call and emergency rotations (University of Vermont College of Medicine)
- Performance measurement and evaluation (American Red Cross)

The purpose of this chapter is to provide a review of the basic features of Microsoft Excel that you need to know to use spreadsheets for analyzing and

[1]Thomas A. Grossman, "Resources for Spreadsheet Analysts," *Analytics* (May/June 2010): 8. analyticsmagazine.com

[2]Larry J. LeBlanc and Thomas A. Grossman, "Introduction: The Use of Spreadsheet Software in the Application of Management Science and Operations Research," *Interfaces*, 38, 4 (July–August 2008): 225–227.

solving problems with techniques of business analytics. In this text, we use Microsoft Excel 2013 for Windows to perform spreadsheet calculations and analyses. Excel files for all text examples and data used in problems and exercises are provided with this book (see the Preface). This review is not intended to be a complete tutorial; many good Excel tutorials can be found online, and we also encourage you to use the Excel help capability (by clicking the question mark button at the top right of the screen). Also, for any reader who may be a Mac user, we caution you that Mac versions of Excel do not have the full functionality that Windows versions have, particularly statistical features, although most of the basic capabilities are the same. In particular, the Excel add-in that we use in later chapters, *Analytic Solver Platform*, only runs on Windows. Thus, if you use a Mac, you should either run Bootcamp with Windows or use a third-party software product such as Parallels or VMWare.

Basic Excel Skills

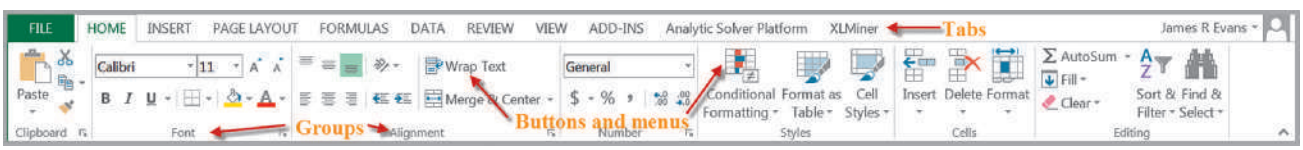
To be able to apply the procedures and techniques that you will learn in this book, it is necessary for you to be relatively proficient in using Excel. We assume that you are familiar with the most elementary spreadsheet concepts and procedures, such as

- opening, saving, and printing files;
- using workbooks and worksheets;
- moving around a spreadsheet;
- selecting cells and ranges;
- inserting/deleting rows and columns;
- entering and editing text, numerical data, and formulas in cells;
- formatting data (number, currency, decimal places, etc.);
- working with text strings;
- formatting data and text; and
- modifying the appearance of the spreadsheet using borders, shading, and so on.

Menus and commands in Excel 2013 reside in the "ribbon" shown in Figure 2.1. Menus and commands are arranged in logical *groups* under different *tabs* (*File, Home, Insert,* and so on); small triangles pointing downward indicate *menus* of additional choices. We often refer to certain commands or options and where they may be found in the ribbon.

Figure 2.1

Excel 2013 Ribbon



Excel Formulas

Formulas in Excel use common mathematical operators:

- addition (+)
- subtraction (−)
- multiplication (*)
- division (/)

Exponentiation uses the ^ symbol; for example, 2^5 is written as 2^5 in an Excel formula.

Cell references in formulas can be written either with *relative addresses* or *absolute addresses*. A **relative address** uses just the row and column label in the cell reference (for example, A4 or C21); an **absolute address** uses a dollar sign ($ sign) before either the row or column label or both (for example, $A2, C$21, or $B$15). Which one we choose makes a critical difference if you copy the cell formulas. If only relative addressing is used, then copying a formula to another cell changes the cell references by the number of rows or columns in the direction that the formula is copied. So, for instance, if we would use a formula in cell B8, =B4−B5*A8, and copy it to cell C9 (one column to the right and one row down), all the cell references are increased by one and the formula would be changed to =C5−C6*B9.

Using a $ sign before a row label (for example, B$4) keeps the reference fixed to row 4 but allows the column reference to change if the formula is copied to another cell. Similarly, using a $ sign before a column label (for example, $B4) keeps the reference to column B fixed but allows the row reference to change. Finally, using a $ sign before both the row and column labels (for example, $B$4) keeps the reference to cell B4 fixed no matter where the formula is copied. You should be very careful to use relative and absolute addressing appropriately in your models, especially when copying formulas.

EXAMPLE 2.1 Implementing Price-Demand Models in Excel

In Chapter 1, we described two models for predicting demand as a function of price:

$$D = a - bP$$

and

$$D = cP^{-d}$$

Figure 2.2 shows a spreadsheet (Excel file *Demand Prediction Models*) for calculating demand for different prices using each of these models. For example, to calculate the demand in cell B8 for the linear model, we use the formula

= $B$4 – $B$5*A8

To calculate the demand in cell E8 for the nonlinear model, we use the formula

= $E$4*D8^ – $E$5

Note how the absolute addresses are used so that as these formulas are copied down, the demand is computed correctly.

Copying Formulas

Excel provides several ways of copying formulas to different cells. This is extremely useful in building decision models, because many models require replication of formulas for different periods of time, similar products, and so on. One way is to select the cell with the formula to be copied, click the *Copy* button from the *Clipboard* group under the *Home* tab (or simply press Ctrl-C on your keyboard), click on the cell you wish to copy to, and then click the *Paste* button (or press Ctrl-V). You may also enter a formula directly in a range of cells without copying and pasting by selecting the range, typing in the formula, and pressing Ctrl-Enter.

Figure 2.2

Excel Models for Demand Prediction

	A	B	C	D	E
1	**Demand Prediction Models**				
2					
3	**Linear Model**			**Nonlinear Model**	
4	**a**	20,000		**c**	20,000
5	**b**	10		**d**	0.0111382
6					
7	**Price**	**Demand**		**Price**	**Demand**
8	$80.00	$19,200		$70.00	$19,075.63
9	$90.00	$19,100		$80.00	$19,047.28
10	$100.00	$19,000		$90.00	$19,022.31
11	$110.00	$18,900		$100.00	$19,000.00
12	$120.00	$18,800		$110.00	$18,979.84
13				$120.00	$18,961.45
14				$130.00	$18,944.56

To copy a formula from a single cell or range of cells down a column or across a row, first select the cell or range, click and hold the mouse on the small square in the lower right-hand corner of the cell (the "fill handle"), and drag the formula to the "target" cells to which you wish to copy.

Other Useful Excel Tips

- ***Split Screen.*** You may split the worksheet horizontally and/or vertically to view different parts of the worksheet at the same time. The vertical splitter bar is just to the right of the bottom scroll bar, and the horizontal splitter bar is just above the right-hand scroll bar. Position your cursor over one of these until it changes shape, click, and drag the splitter bar to the left or down.
- ***Paste Special.*** When you normally copy (one or more) cells and paste them in a worksheet, Excel places an exact copy of the formulas or data in the cells (except for relative addressing). Often you simply want the *result* of formulas, so the data will remain constant even if other parameters used in the formulas change. To do this, use the *Paste Special* option found within the *Paste* menu in the *Clipboard* group under the *Home* tab instead of the *Paste* command. Choosing *Paste Values* will paste the result of the formulas from which the data were calculated.
- ***Column and Row Widths.*** Many times a cell contains a number that is too large to display properly because the column width is too small. You may change the column width to fit the largest value or text string anywhere in the column by positioning the cursor to the right of the column label so that it changes to a cross with horizontal arrows and then double-clicking. You may also move the arrow to the left or right to manually change the column width. You may change the row heights in a similar fashion by moving the cursor below the row number label. This can be especially useful if you have a very long formula to display. To break a formula within a cell, position the cursor at the break point in the formula bar and press *Alt-Enter*.
- ***Displaying Formulas in Worksheets.*** Choose *Show Formulas* in the *Formula Auditing* group under the *Formulas* tab. You often need to change the column width to display the formulas properly.
- ***Displaying Grid Lines and Row and Column Headers for Printing.*** Check the *Print* boxes for gridlines and headings in the *Sheet Options* group under the *Page*

Layout tab. Note that the *Print* command can be found by clicking on the *Office* button.

- ***Filling a Range with a Series of Numbers.*** Suppose you want to build a worksheet for entering 100 data values. It would be tedious to have to enter the numbers from 1 to 100 one at a time. Simply fill in the first few values in the series and highlight them. Then click and drag the small square (fill handle) in the lower right-hand corner down (Excel will show a small pop-up window that tells you the last value in the range) until you have filled in the column to 100; then release the mouse.

Excel Functions

Functions are used to perform special calculations in cells and are used extensively in business analytics applications. All Excel functions require an equal sign and a function name followed by parentheses, in which you specify arguments for the function.

Basic Excel Functions

Some of the more common functions that we will use in applications include the following:

MIN(*range*)—finds the smallest value in a range of cells
MAX(*range*)—finds the largest value in a range of cells
SUM(*range*)—finds the sum of values in a range of cells
AVERAGE(*range*)—finds the average of the values in a range of cells
COUNT(*range*)—finds the number of cells in a range that contain numbers
COUNTIF(*range, criteria*)—finds the number of cells within a range that meet a specified criterion.

The COUNTIF function counts the number of cells within a range that meet a criterion that you specify. For example, you can count all the cells that start with a certain letter, or you can count all the cells that contain a number that is larger or smaller than a number you specify. Examples of criteria are 100, ">100", a cell reference such as A4, a text string such as "Facebook." Note that text and logical formulas must be enclosed in quotes. See Excel Help for other examples.

Excel has other useful COUNT-type functions: COUNTA counts the number of nonblank cells in a range, and COUNTBLANK counts the number of blank cells in a range. In addition, COUNTIFS(*range1, criterion1, range2, criterion2,… range_n, criterion_n*) finds the number of cells within multiple ranges that meet specific criteria for each range.

We illustrate these functions using the *Purchase Orders* data set in Example 2.2.

EXAMPLE 2.2 Using Basic Excel Functions

In the *Purchase Orders* data set, we will find the following:

- smallest and largest quantity of any item ordered
- total order costs
- average number of months per order for accounts payable
- number of purchase orders placed
- number of orders placed for O-rings
- number of orders with A/P terms shorter than 30 months
- number of O-ring orders from Spacetime Technologies

The results are shown in Figure 2.3. In this figure, we used the split-screen feature in Excel to reduce the number of rows shown in the spreadsheet. To find the smallest and largest quantity of any item ordered, we use the MIN and MAX functions for the data in column F. Thus, the formula in cell B99 is =MIN(F4:F97) and the formula in cell B100 is =MAX(F4:F97). To find the total order costs, we sum the data in column G using the SUM function: =SUM(G4:G97); this is the formula in cell B101. To find the average number of A/P months, we use the AVERAGE function for the data in column H. The formula in cell B102 is =AVERAGE(H4:H97). To find the number of purchase orders placed, use the COUNT function. Note that the COUNT function counts only the number of cells in a range that contain numbers, so we could not use it in columns A, B, or D; however, any other column would be acceptable. Using the item numbers in column C, the formula in cell B103 is =COUNT(C4:C97). To find the number of orders placed for O-rings, we use the COUNTIF function. For this example, the formula used in cell B104 is =COUNTIF(D4:D97, "O-Ring"). We could have also used the cell reference for any cell containing the text O-Ring, such as =COUNTIF(D4:D97,D12). To find the number of orders with A/P terms less than 30 months, use the formula =COUNTIF(H4:H97,"<30") in cell B105. Finally, to count the number of O-Ring orders for Spacetime Technologies, we use =COUNTIFS(D4:D97,"O-Ring", A4:A97,"Spacetime Technologies").

IF-type functions are also available for other calculations. For example, the functions SUMIF, AVERAGEIF, SUMIFS, and AVERAGEIFS can be used to embed IF logic within mathematical functions. For instance, the syntax of SUMIF is SUMIF(*range, criterion, [sum range]*). "Sum range" is an optional argument that allows you to add cells in a different range. Thus, in the *Purchase Orders* database, to find the total cost of all airframe fasteners, we would use

SUMIF(D4:D97, "Airframe fasteners", G4:G97)

This function looks for Airframe fasteners in the range D4:D97, but then sums the associated values in column G (cost per order).

Functions for Specific Applications

Excel has a wide variety of other functions for statistical, financial, and other applications, many of which we introduce and use throughout the text. For instance, some financial models that we develop require the calculation of net present value (NPV). **Net present value** (also called **discounted cash flow**) measures the worth of a stream of cash flows, taking into

Figure 2.3

Application of Excel Functions to *Purchase Orders* Data

	A	B	C	D	E	F	G	H	I	J
1	Purchase Orders									
2										
3	Supplier	Order No.	Item No.	Item Description	Item Cost	Quantity	Cost per order	A/P Terms (Months)	Order Date	Arrival Date
4	Hulkey Fasteners	Aug11001	1122	Airframe fasteners	$ 4.25	19,500	$ 82,875.00	30	08/05/11	08/13/11
5	Alum Sheeting	Aug11002	1243	Airframe fasteners	$ 4.25	10,000	$ 42,500.00	30	08/08/11	08/14/11
6	Fast-Tie Aerospace	Aug11003	5462	Shielded Cable/ft.	$ 1.05	23,000	$ 24,150.00	30	08/10/11	08/15/11
7	Fast-Tie Aerospace	Aug11004	5462	Shielded Cable/ft.	$ 1.05	21,500	$ 22,575.00	30	08/15/11	08/22/11
8	Steelpin Inc.	Aug11005	5319	Shielded Cable/ft.	$ 1.10	17,500	$ 19,250.00	30	08/20/11	08/31/11
9	Fast-Tie Aerospace	Aug11006	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/20/11	08/26/11
10	Steelpin Inc.	Aug11007	4312	Bolt-nut package	$ 3.75	4,250	$ 15,937.50	30	08/25/11	09/01/11
11	Durrable Products	Aug11008	7258	Pressure Gauge	$ 90.00	100	$ 9,000.00	45	08/25/11	08/28/11
12	Fast-Tie Aerospace	Aug11009	6321	O-Ring	$ 2.45	1,300	$ 3,185.00	30	08/25/11	09/04/11
96	Steelpin Inc.	Nov11009	5677	Side Panel	$ 195.00	110	$ 21,450.00	30	11/05/11	11/17/11
97	Manley Valve	Nov11010	9955	Door Decal	$ 0.55	125	$ 68.75	30	11/05/11	11/10/11
98										
99	Minimum Quantity	90								
100	Maximum Quantity	25,000								
101	Total Order Costs	$ 2,471,760.00								
102	Average Number of A/P Months	30.63829787								
103	Number of Purchase Orders	94								
104	Number of O-ring Orders	12								
105	Number of A/P Terms < 30	17								
106	Number of O-ring Orders Spacetime	3								

account the time value of money. That is, a cash flow of F dollars t time periods in the future is worth $F/(1 + i)^t$ dollars today, where i is the **discount rate**. The discount rate reflects the opportunity costs of spending funds now versus achieving a return through another investment, as well as the risks associated with not receiving returns until a later time. The sum of the present values of all cash flows over a stated time horizon is the net present value:

$$\text{NPV} = \sum_{t=0}^{n} \frac{F_t}{(1 + i)^t} \tag{2.1}$$

where F_t = cash flow in period t. A positive NPV means that the investment will provide added value because the projected return exceeds the discount rate.

The Excel function NPV(*rate, value1, value2,…*) calculates the net present value of an investment by using a discount rate and a series of future payments (negative values) and income (positive values). *Rate* is the value of the discount rate i over the length of one period, and *value1*, *value2*,… are 1 to 29 arguments representing the payments and income for each period. The values must be equally spaced in time and are assumed to occur at the end of each period. The NPV investment begins one period before the date of the *value1* cash flow and ends with the last cash flow in the list. The NPV calculation is based on future cash flows. If the first cash flow (such as an initial investment or fixed cost) occurs at the beginning of the first period, then it must be added to the NPV result and *not* included in the function arguments.

EXAMPLE 2.3 Using the NPV Function

A company is introducing a new product. The fixed cost for marketing and distribution is $25,000 and is incurred just prior to launch. The forecasted net sales revenues for the first six months are shown in Figure 2.4. The formula in cell B8 computes the net present value of these cash flows as =NPV(B6,C4:H4) – B5. Note that the fixed cost is not a future cash flow and is not included in the NPV function arguments.

Insert Function

The easiest way to locate a particular function is to select a cell and click on the *Insert function* button [f_x], which can be found under the ribbon next to the formula bar and also in the *Function Library* group in the *Formulas* tab. You may either type in a description in the search field, such as "net present value," or select a category, such as "Financial," from the drop-down box.

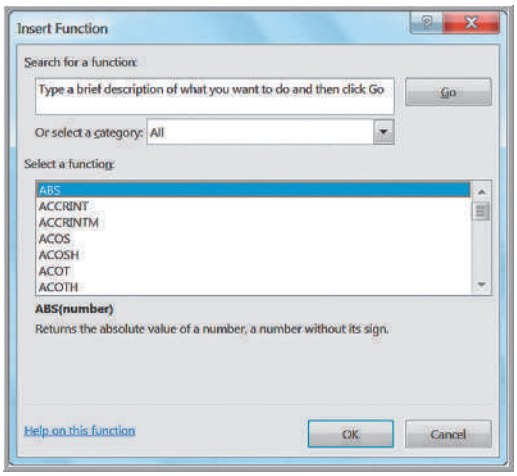
This feature is particularly useful if you know what function to use but are not sure of what arguments to enter because it will guide you in entering the appropriate data for the function arguments. Figure 2.5 shows the dialog from which you may select the function you wish

Figure 2.4
Net Present Value Calculation

	A	B	C	D	E	F	G	H
1	Net Present Value							
2								
3		Month	January	February	March	April	May	June
4		Sales Revenue Forecast	$2,500	$4,000	$5,000	$8,000	$10,000	$12,500
5	Fixed Cost	$25,000.00						
6	Discount Rate	3%						
7								
8	NPV	$11,975.81						

Figure 2.5

Insert Function Dialog



to use. For example, if we would choose the COUNTIF function, the dialog in Figure 2.6 appears. When you click in an input cell, a description of the argument is shown. Thus, if you are not sure what to enter for the range, the explanation in Figure 2.6 will help you. For further information, you could click on the *Help* button in the lower left-hand corner.

Logical Functions

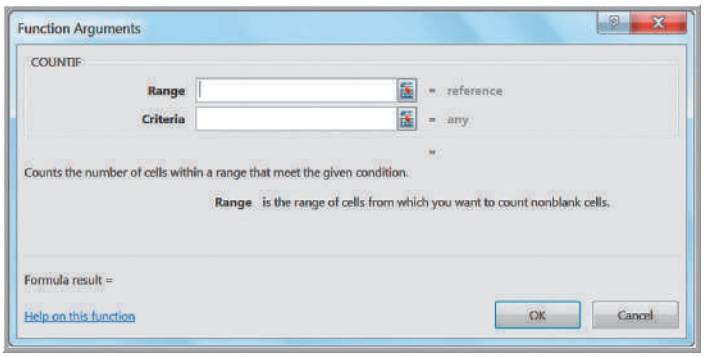
Logical functions return only one of two values: TRUE or FALSE. Three useful logical functions in business analytics applications are

IF(*condition, value if true, value if false*)—a logical function that returns one value if the condition is true and another if the condition is false,
AND(*condition 1, condition 2…*)—a logical function that returns TRUE if all conditions are true and FALSE if not,
OR(*condition 1, condition 2…*)—a logical function that returns TRUE if any condition is true and FALSE if not.

The IF function, *IF(condition, value if true, value if false),* allows you to choose one of two values to enter into a cell. If the specified *condition* is true, *value if true* will be put in

Figure 2.6

Function Arguments Dialog for COUNTIF



the cell. If the condition is false, *value if false* will be entered. *Value if true* and *value if false* can be a number or a text string enclosed in quotes. Note that if a blank is used between quotes, " ", then the result will simply be a blank cell. This is often useful to create a clean spreadsheet. For example, if cell C2 contains the function =IF(A8=2,7,12), it states that if the value in cell A8 is 2, the number 7 will be assigned to cell C2; if the value in cell A8 is not 2, the number 12 will be assigned to cell C2. Conditions may include the following:

= equal to
> greater than
< less than
>= greater than or equal to
<= less than or equal to
<> not equal to

You may "nest" up to seven IF functions by replacing *value-if-true* or *value-if-false* in an IF function with another IF function:

=IF(A8=2,(IF(B3=5,"YES","")),15)

This says that if cell A8 equals 2, then check the contents of cell B3. If cell B3 is 5, then the value of the function is the text string YES; if not, it is a blank space (represented by quotation marks with nothing in between). However, if cell A8 is not 2, then the value of the function is 15 no matter what cell B3 is.

AND and OR functions simply return the values of *true* or *false* if all or at least one of multiple conditions are met, respectively. You may use AND and OR functions as the

EXAMPLE 2.4 Using the IF Function

Suppose that the aircraft-component manufacturer considers any order of 10,000 units or more to be large, whereas any other order size is considered to be small. We may use the IF function to classify the orders. First, create a new column in the spreadsheet for the order size, say, column K. In cell K4, use the formula

=IF(F4>=10000,"Large","Small")

This function will return the value *Large* in cell K4 if the order size in cell F4 is 10,000 or more; otherwise, it returns the value *Small*. Further, suppose that large orders with a total cost of at least $25,000 are considered critical. We may flag these orders as critical by using the function in cell L4:

=IF(AND(K4="Large", G4>=25000),"Critical"," ")

After copying these formulas down the columns, Figure 2.7 shows a portion of the results.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Purchase Orders											
2												
3	Supplier	Order No.	Item No.	Item Description	Item Cost	Quantity	Cost per order	A/P Terms (Months)	Order Date	Arrival Date	Order Size	Type
4	Hulkey Fasteners	Aug11001	1122	Airframe fasteners	$ 4.25	19,500	$ 82,875.00	30	08/05/11	08/13/11	Large	Critical
5	Alum Sheeting	Aug11002	1243	Airframe fasteners	$ 4.25	10,000	$ 42,500.00	30	08/08/11	08/14/11	Large	Critical
6	Fast-Tie Aerospace	Aug11003	5462	Shielded Cable/ft.	$ 1.05	23,000	$ 24,150.00	30	08/10/11	08/15/11	Large	
7	Fast-Tie Aerospace	Aug11004	5462	Shielded Cable/ft.	$ 1.05	21,500	$ 22,575.00	30	08/15/11	08/22/11	Large	
8	Steelpin Inc.	Aug11005	5319	Shielded Cable/ft.	$ 1.10	17,500	$ 19,250.00	30	08/20/11	08/31/11	Large	
9	Fast-Tie Aerospace	Aug11006	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/20/11	08/26/11	Large	
10	Steelpin Inc.	Aug11007	4312	Bolt-nut package	$ 3.75	4,250	$ 15,937.50	30	08/25/11	09/01/11	Small	
11	Durrable Products	Aug11008	7258	Pressure Gauge	$ 90.00	100	$ 9,000.00	45	08/25/11	08/28/11	Small	
12	Fast-Tie Aerospace	Aug11009	6321	O-Ring	$ 2.45	1,300	$ 3,185.00	30	08/25/11	09/04/11	Small	
13	Fast-Tie Aerospace	Aug11010	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/25/11	09/02/11	Large	
14	Steelpin Inc.	Aug11011	5319	Shielded Cable/ft.	$ 1.10	18,100	$ 19,910.00	30	08/25/11	09/05/11	Large	
15	Hulkey Fasteners	Aug11012	3166	Electrical Connector	$ 1.25	5,600	$ 7,000.00	30	08/25/11	08/29/11	Small	

Figure 2.7

Classifying Order Sizes Using the IF Function

condition within an IF function; for example, =IF(AND(B1=3,C1=5),12,22). Here, if cell B1=3 and cell C1=5, then the value of the function is 12; otherwise it is 22.

Using Excel Lookup Functions for Database Queries

In Chapter 1 we noted that business intelligence was instrumental in the evolution of business analytics. Organizations often need to extract key information from a database to support customer service representatives, technical support, manufacturing, and other needs. Excel provides some useful functions for finding specific data in a spreadsheet. These are:

VLOOKUP(*lookup_value, table_array, col_index_num, [range lookup]*) looks up a value in the leftmost column of a table (specified by the *table_array*) and returns a value in the same row from a column you specify (*col_index_num*).

HLOOKUP(*lookup_value, table_array, row_index_num, [range lookup]*) looks up a value in the top row of a table and returns a value in the same column from a row you specify.

INDEX(*array, row_num, col_num*) returns a value or reference of the cell at the intersection of a particular row and column in a given range.

MATCH(*lookup_value, lookup_array, match_type*) returns the relative position of an item in an array that matches a specified value in a specified order.

In the VLOOKUP and HLOOKUP functions, *range lookup* is optional. If this is omitted or set as *True,* then the first column of the table must be sorted in ascending numerical order. If an exact match for the *lookup_value* is found in the first column, then Excel will return the value the *col_index_num* of that row. If an exact match is not found, Excel will choose the row with the largest value in the first column that is less than the *lookup_value*. If *range lookup* is *false*, then Excel seeks an exact match in the first column of the table range. If no exact match is found, Excel will return #N/A (not available). We recommend that you specify the range lookup to avoid errors.

EXAMPLE 2.5 Using the VLOOKUP Function

In Chapter 1, we introduced a database of sales transactions for a firm that sells instructional fitness books and DVDs (Excel file *Sales Transactions*). The database is sorted by customer ID, and a portion of it is shown in Figure 2.8. Suppose that a customer calls a representative about a payment issue. The representative finds the customer ID—for example, 10007—and needs to look up the type of payment and transaction code. We may use the VLOOKUP function to do this. In the function VLOOKUP(*lookup_value, table_array, col_index_num*), *lookup_value* represents the customer ID. The *table_array* is the range of the data in the spreadsheet; in this case, it is the range A4:H475. The value for *col_index_num* represents the column in the table range we wish to retrieve. For the type of payment, this is column 3; for the transaction code, this is column 4. Note that the first column is already sorted in ascending numerical order, so we can either omit the *range lookup* argument or set it as *true*. Thus, if we enter the formula below in any blank cell of the spreadsheet:

=VLOOKUP(10007,$A$4:$H$475,3)

returns the payment type, *Credit*. If we use the following formula:

=VLOOKUP(10007,$A$4:$H$475,4)

the function returns the transaction code, 80103311.

Now suppose the database was sorted by transaction code so that the customer ID column is no longer in ascending numerical order as shown in Figure 2.9. If we use the function =VLOOKUP(10007,$A$4:$H$475,4, True), Excel returns #N/A. However, if we change the range lookup argument to False, then the function returns the correct value of the transaction code.

Figure 2.8

Portion of *Sales Transactions* Data Sorted by Çustomer ID

	A	B	C	D	E	F	G	H
1	Sales Transactions: July 14							
2								
3	Cust ID	Region	Payment	Transaction Code	Source	Amount	Product	Time Of Day
4	10001	East	Paypal	93816545	Web	$20.19	DVD	22:19
5	10002	West	Credit	74083490	Web	$17.85	DVD	13:27
6	10003	North	Credit	64942368	Web	$23.98	DVD	14:27
7	10004	West	Paypal	70560957	Email	$23.51	Book	15:38
8	10005	South	Credit	35208817	Web	$15.33	Book	15:21
9	10006	West	Paypal	20978903	Email	$17.30	DVD	13:11
10	10007	East	Credit	80103311	Web	$177.72	Book	21:59
11	10008	West	Credit	14132683	Web	$21.76	Book	4:04
12	10009	West	Paypal	40128225	Web	$15.92	DVD	19:35
13	10010	South	Paypal	49073721	Web	$23.39	DVD	13:26

Figure 2.9

Portion of *Sales Transactions* Data Sorted by Transaction Code

	A	B	C	D	E	F	G	H
1	Sales Transactions: July 14							
2								
3	Cust ID	Region	Payment	Transaction Code	Source	Amount	Product	Time Of Day
4	10391	West	Credit	10325805	Web	$22.79	Book	0:00
5	10231	North	Paypal	10400774	Web	$216.20	Book	10:33
6	10267	West	Paypal	10754185	Web	$23.01	DVD	17:44
7	10228	West	Credit	10779898	Web	$15.33	DVD	5:05
8	10037	South	Paypal	11165609	Web	$217	Book	0:00
9	10297	North	Credit	11175481	Web	$22.65	Book	6:06
10	10294	West	Paypal	11427628	Web	$15.40	Book	17:16
11	10081	North	Credit	11673210	Web	$16.14	DVD	4:04
12	10129	West	Credit	11739665	Web	$22.03	DVD	14:49
13	10406	East	Credit	12075708	Web	$22.99	Book	9:09
14	10344	East	Credit	12222505	Web	$15.55	DVD	6:06

The HLOOKUP function works in a similar fashion. For most spreadsheet databases, we would normally need to use the VLOOKUP function. In some modeling situations, however, the HLOOKUP function can be useful if the data are arranged column by column rather than row by row.

The INDEX function works as a lookup procedure by returning the value in a particular row and column of an array. For example, in the *Sales Transactions* database, INDEX($A$4:$H$475, 7, 4) would retrieve the transaction code, 80103311 that is in the 7th row and 4th column of the data array (see Figure 2.8), as the VLOOKUP function did in Example 2.5. The difference is that it relies on the row number rather than the actual value of the customer ID.

In the MATCH function, *lookup_value* is the value that you want to match in *lookup_array*, which is the range of cells being searched. The *match_type* is either −1, 0, or 1. The default is 1. If *match_type* = 1, then the function finds the largest value that is less than or equal to *lookup_value*. The values in the *lookup_array* must be placed in ascending order. If *match_type* = 0, MATCH finds the first value that is exactly equal to *lookup_value*. The values in the *lookup_array* can be in any order. If *match_type* = −1, then the function finds the smallest value that is greater than or equal to *lookup_value*. The values in the *lookup_array* must be placed in descending order. Example 2.6 shows how the INDEX and MATCH functions can be used.

The VLOOKUP function will not work if you want to look up something to the left of a specified range (because it uses the first column of the range to find the lookup value). However, we can use the INDEX and MATCH function easily to do this, as Example 2.7 shows.

EXAMPLE 2.6 Using INDEX and MATCH Functions for Database Queries

Figure 2.10 shows the data in the Excel file *Monthly Product Sales Queries*. Suppose we wish to design a simple query application to input the month and product name, and retrieve the corresponding sales. The three additional worksheets in the workbook show how to do this in three different ways. The *Query1* worksheet (see Figure 2.11) uses the VLOOKUP function with embedded IF statements. The formulas in cell I8 is:

=VLOOKUP(I5,A4:F15,IF(I6="A",2,IF(I6="B",3,
IF(I6="C",4,IF(I6="D",5,IF(I6="E",6))))),FALSE)

The IF functions are used to determine the column in the lookup table to use, and, as you can see, is somewhat complex, especially if the table were much larger.

The *Query2* worksheet (not shown here; see the Excel workbook) uses the VLOOKUP and MATCH functions in cell I8. The formula in cell I8 is:

=VLOOKUP(I5,A4:F15,MATCH(I6,B3:F3,0)+1,FALSE)

In this case, the MATCH function is used to identify the column in the table corresponding to the product name in cell I6. Note the use of the "+1" to shift the relative column number of the product to the correct column number in the lookup table.

Finally, the *Query3* worksheet (also not shown here) uses only INDEX and MATCH functions in cell I8. The formula in cell I8 is:

=INDEX(A4:F15,MATCH(I5,A4:A15,0), MATCH(I6,A3:F3,0))

The MATCH functions are used as arguments in the INDEX function to identify the row and column numbers in the table based on the month and product name. The INDEX function then retrieves the value in the corresponding row and column. This is perhaps the cleanest formula of the three. By studying these examples carefully, you will better understand how to use these functions in other applications.

Figure 2.10

***Monthly Product Sales Queries* Workbook**

	A	B	C	D	E	F
1	Sales Units					
2				Product		
3	Month	A	B	C	D	E
4	January	7,792	5,554	3,105	3,168	10,350
5	February	7,268	3,024	3,228	3,751	8,965
6	March	7,049	5,543	2,147	3,319	6,827
7	April	7,560	5,232	2,636	4,057	8,544
8	May	8,233	5,450	2,726	3,837	7,535
9	June	8,629	3,943	2,705	4,664	9,070
10	July	8,702	5,991	2,891	5,418	8,389
11	August	9,215	3,920	2,782	4,085	7,367
12	September	8,986	4,753	2,524	5,575	5,377
13	October	8,654	4,746	3,258	5,333	7,645
14	November	8,315	3,566	2,144	4,924	8,173
15	December	7,978	5,670	3,071	6,563	6,088

Data | Query1 | Query2 | Query3

Figure 2.11

***Query1* Worksheet in *Monthly Product Sales Queries* Workbook**

	A	B	C	D	E	F	G	H	I
1	Sales Units							Using VLOOKUP + IF	
2				Product					
3	Month	A	B	C	D	E		Sales Lookup	
4	January	7,792	5,554	3,105	3,168	10,350			
5	February	7,268	3,024	3,228	3,751	8,965		Month	April
6	March	7,049	5,543	2,147	3,319	6,827		Product	E
7	April	7,560	5,232	2,636	4,057	8,544			
8	May	8,233	5,450	2,726	3,837	7,535		Sales	8,544
9	June	8,629	3,943	2,705	4,664	9,070			
10	July	8,702	5,991	2,891	5,418	8,389			
11	August	9,215	3,920	2,782	4,085	7,367			
12	September	8,986	4,753	2,524	5,575	5,377			
13	October	8,654	4,746	3,258	5,333	7,645			
14	November	8,315	3,566	2,144	4,924	8,173			
15	December	7,978	5,670	3,071	6,563	6,088			

Data | Query1 | Query2 | Query3

EXAMPLE 2.7 Using INDEX and MATCH for a Left Table Lookup

Suppose that, in the *Sales Transactions* database, we wish to find the customer ID associated with a specific transaction code. Refer back to Figure 2.8 or the Excel workbook. Suppose that we enter the transaction code in cell K2, and want to display the customer ID in cell K4. Use the formula in cell K4:

=INDEX(A4:A475,MATCH(K2,D4:D475,0),1)

Here, the MATCH function is used to identify the row number in the table range that matches the transaction code exactly, and the INDEX function uses this row number and column 1 to identify the associated customer ID.

Spreadsheet Add-Ins for Business Analytics

Microsoft Excel will provide most of the computational support required for the material in this book. Excel (Windows only) provides an add-in called the *Analysis Toolpak,* which contains a variety of tools for statistical computation, and *Solver*, which is used for optimization. These add-ins are not included in a standard Excel installation. To install them, click the *File* tab and then *Options* in the left column. Choose *Add-Ins* from the left column. At the bottom of the dialog, make sure *Excel Add-ins* is selected in the *Manage:* box and click *Go*. In the *Add-Ins* dialog, if *Analysis Toolpak*, *Analysis Toolpak VBA,* and *Solver Add-in* are not checked, simply check the boxes and click *OK*. You will not have to repeat this procedure every time you run Excel in the future.

In addition, many third-party add-ins are available to support analytic procedures in Excel. One add-in, Frontline Systems' *Analytic Solver Platform*, offers many other capabilities for both predictive and prescriptive analytics. See the Preface for instructions on how to download and install this software. We will use both the included Excel add-ins and *Analytic Solver Platform* throughout this book, so we encourage you to download and set up these add-ins on your computer at this time.

Key Terms

Absolute address
Discount rate
Net present value (discounted cash flow)
Relative address

Problems and Exercises

1. The Excel file *Firm Data* shows the prices charged and different product sizes. Prepare a worksheet using VLOOKUP function that will compute the invoice to be sent to a customer when any product type, size, and order quantity are entered.

2. The Excel file *Store and Regional Sales Database* provides sales data for computers and peripherals showing the store identification number, sales region, item number, item description, unit price, units sold, and month when the sales were made during the fourth quarter of last year.[3] Modify the

[3]Based on Kenneth C. Laudon and Jane P. Laudon, *Essentials of Management Information Systems*, 9th ed. (Upper Saddle River, NJ: Prentice Hall, 2011).

spreadsheet to calculate the total sales revenue for each of the eight stores as well as each of the three sales regions.

3. The Excel file *President's Inn Guest Database* provides a list of customers, rooms they occupied, arrival and departure dates, number of occupants, and daily rate for a small bed-and-breakfast inn during one month.[4] Room rates are the same for one or two guests; however, additional guests must pay an additional $20 per person per day for meals. Guests staying for seven days or more receive a 10% discount. Modify the spreadsheet to calculate the number of days that each party stayed at the inn and the total revenue for the length of stay.

4. The worksheet *Base Data* in the Excel file *Credit Risk Data* provides information about 425 bank customers who had applied for loans. The data include the purpose of the loan, checking and savings account balances, number of months as a customer of the bank, months employed, gender, marital status, age, housing status and number of years at current residence, job type, and credit-risk classification by the bank.[5]
   a. Use the COUNTIF function to determine (1) how many customers applied for new-car, used-car, business, education, small-appliance, and furniture loans and (2) the number of customers with checking account balances less than $500.
   b. Modify the spreadsheet using IF functions to include new columns, classifying the checking and savings account balances as low if the balance is less than $250, medium if between $250 but less than $2000, and high otherwise.

5. A manager needs to identify some information from the *Purchase Orders* Excel file but has only the order number. Modify the Excel file to use the VLOOKUP function to find the item description and cost per order for the following order numbers: Aug11008, Sep11023, and Oct11020.

6. A pharmaceutical manufacturer has projected net profits for a new drug that is being released to the market over the next five years:

Year	Net Profit
1	$(300,000,000)
2	$(145,000,000)
3	$50,000,000
4	$125,000,000
5	$530,000,000

   Use a spreadsheet to find the net present value of these cash flows for a discount rate of 3%.

7. Example 1.4 in Chapter 1 described a scenario for new product sales that can be characterized by a formula called a Gompertz curve: $S = ae^{be^{ct}}$. Develop a spreadsheet for calculating sales using this formula for $t = 0$ to 160 in increments of 10 when $a = 15000$, $b = -8$, and $c = -0.05$.

8. Example 1.8 in Chapter 1 provided data from an experiment to identify the relationship between sales and pricing, coupon, and advertising strategies. Enter the data into a spreadsheet and implement the model in the example within your spreadsheet to estimate the sales for each of the weekly experiments. Compute the average sales for the three stores, and find the differences between the averages and the model estimates for each week.

9. The following exercises use the *Purchase Orders* database. Use MATCH and/or INDEX functions to find the following:
   a. The row numbers corresponding to the first and last instance of item number 1369 in column C (be sure column C is sorted by order number).
   b. The order cost associated with the first instance of item 1369 that you identified in part (a).
   c. The total cost of all orders for item 1369. Use the answers to parts (a) and (b) along with the SUM function to do this. In other words, you should use the appropriate INDEX and MATCH functions within the SUM function to find the answer. Validate your results by applying the SUM function directly to the data in column G.

[4]Based on Kenneth C. Laudon and Jane P. Laudon, *Essentials of Management Information Systems*.
[5]Based on Efraim Turban, Ranesh Sharda, Dursun Delen, and David King, *Business Intelligence: A Managerial Approach*, 2nd ed. (Upper Saddle River NJ: Prentice Hall, 2011).