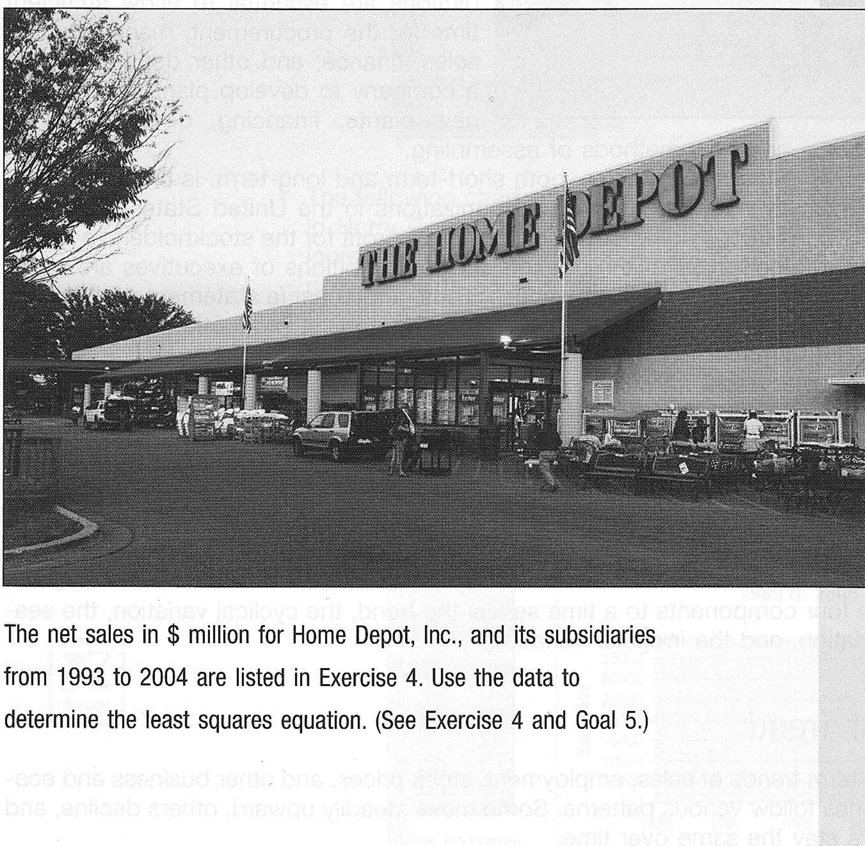
16

Time Series and Forecasting



The net sales in $ million for Home Depot, Inc., and its subsidiaries from 1993 to 2004 are listed in Exercise 4. Use the data to determine the least squares equation. (See Exercise 4 and Goal 5.)

GOALS

When you have completed this chapter, you will be able to:

1 Define the components of a *time series.*

2 Compute a *moving average.*

3 Determine a *linear trend equation.*

4 Compute a trend equation for a nonlinear trend.

5 Use a trend equation to forecast future time periods and to develop seasonally adjusted forecasts.

6 Determine and interpret a set of *seasonal indexes.*

7 Deseasonalize data using a seasonal index.

8 Test for autocorrelation.

Introduction

What is a time series?

This chapter discusses time series analysis and forecasting. A **time series** is a collection of data recorded over a period of time—weekly, monthly, quarterly, or yearly. A few examples of time series are Microsoft Corporation sales by quarter since 1985, the annual production of sulfuric acid since 1970, annual fall enrollment at the University of Missouri, and average number of associates employed annually since 1991 at Home Depot.

An analysis of history—a time series—can be used by management to make current decisions and plans based on long-term forecasting. We usually assume past patterns will continue into the future. Long-term forecasts extend more than 1 year into the future; 2-, 5-, and 10-year projections are common. Long-range predictions are essential to allow sufficient time for the procurement, manufacturing, sales, finance, and other departments of a company to develop plans for possible new plants, financing, development of new products, and new methods of assembling.

Forecasting the level of sales, both short-term and long-term, is practically dictated by the very nature of business organizations in the United States. Competition for the consumer's dollar, stress on earning a profit for the stockholders, a desire to procure a larger share of the market, and the ambitions of executives are some of the prime motivating forces in business. Thus, a forecast (a statement of the goals of management) is necessary to have the raw materials, production facilities, and staff to meet demand.

This chapter deals with the use of data to forecast future events. First, we look at the components of a time series. Then, we examine some of the techniques used in analyzing data. Finally, we forecast future events.

Components of a Time Series

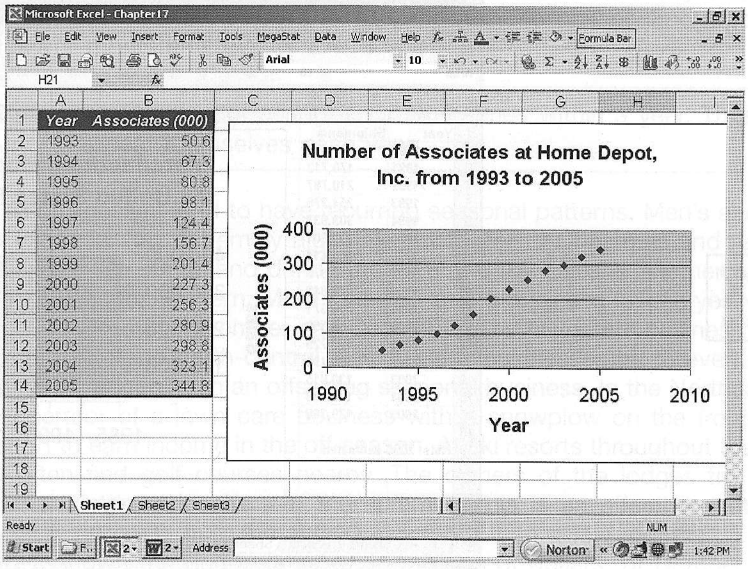
There are four components to a time series: the trend, the cyclical variation, the seasonal variation, and the irregular variation.

Secular Trend

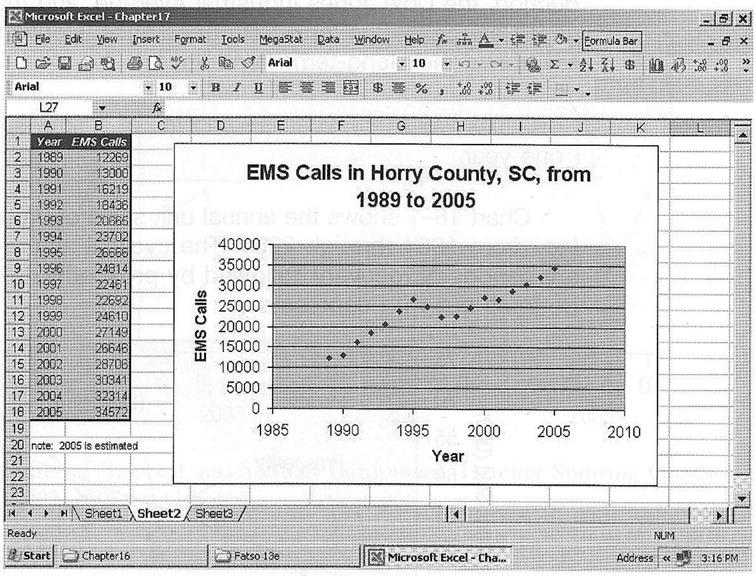
The long-term trends of sales, employment, stock prices, and other business and economic series follow various patterns. Some move steadily upward, others decline, and still others stay the same over time.
The following are several examples of a secular trend.

SECULAR TREND The smooth long-term direction of a time series.

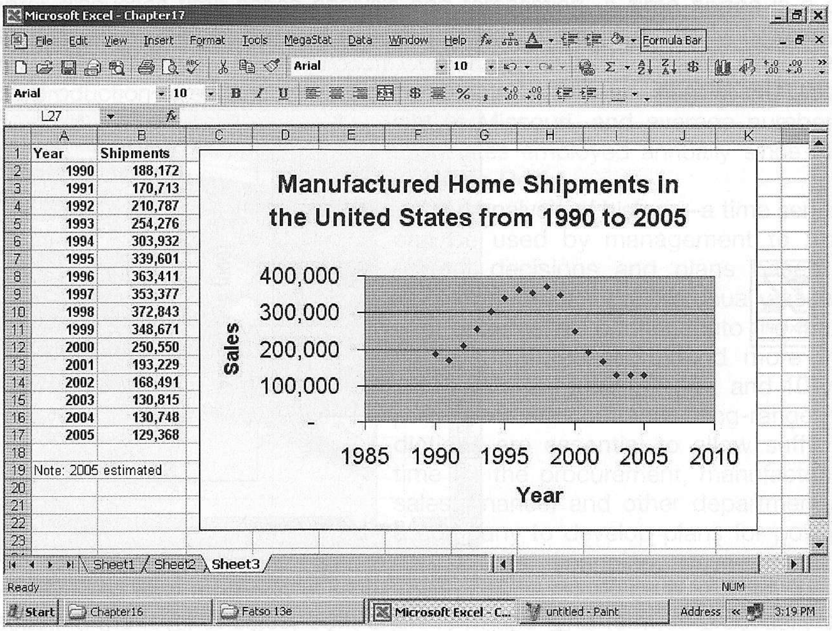
- Home Depot was founded in 1978 and is the second largest retailer (Wal-Mart is the largest) in the United States. The following chart shows the number of employees at Home Depot, Inc. You can see this number has increased rapidly over the last 12 years. In 1993 there were just over 50,000 associates and by 2005 that number had increased to over 340,000.



- The following chart shows the number of emergency medical service (EMS) calls in Horry County, South Carolina, since 1989. The number of calls for EMS has increased nearly 3 times from 12,269 in 1989 to 34,572 in 2005. Notice that the number of calls increased from 1989 to 1995. From 1995 to 2000 the number of calls stayed about the same, and then in 2000 it began another increase to over 30,000. The long-run direction of the trend is increasing.



- The number of manufactured homes shipped in the United States showed a steady increase from 1990 to 1996, then remained about the same until 1999, when the number began to decline. By the year 2002 the number shipped was less than it had been in 1990. This information is shown in the following chart.



Cyclical Variation

The second component of a time series is cyclical variation. A typical business cycle consists of a period of prosperity followed by periods of recession, depression, and then recovery. There are sizable fluctuations unfolding over more than one year in time above and below the secular trend. In a recession, for example, employment, production, the Dow Jones Industrial Average, and many other business and economic series are below the long-term trend lines. Conversely, in periods of prosperity they are above their long-term trend lines.

> **CYCLICAL VARIATION** The rise and fall of a time series over periods longer than one year.

Chart 16–1 shows the annual unit sales of batteries by National Battery Retailers, Inc., from 1984 through 2004. The cyclical nature of business is highlighted. There are periods of recovery, followed by prosperity, then recession, and finally the cycle bottoms out with depression.

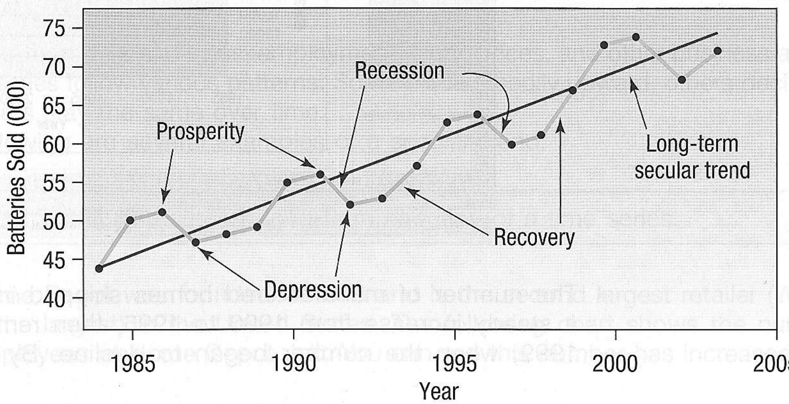


CHART 16–1 Batteries Sold by National Battery Retailers, Inc., from 1984 to 2004

Seasonal Variation

The third component of a time series is the **seasonal variation.** Many sales, production, and other series fluctuate with the seasons. The unit of time reported is either quarterly or monthly.

SEASONAL VARIATION Patterns of change in a time series within a year. These patterns tend to repeat themselves each year.

Almost all businesses tend to have recurring seasonal patterns. Men's and boy's clothing, for example, has extremely high sales just prior to Christmas and relatively low sales just after Christmas and during the summer. Toy sales is another example with an extreme seasonal pattern. More than half of the business for the year is usually done in the months of November and December. The lawn care business is seasonal in the northeast and north-central states. Many businesses try to even out the seasonal effects by engaging in an offsetting seasonal business. In the Northeast you will see the operator of a lawn care business with a snowplow on the front of the truck in an effort to earn income in the off-season. At ski resorts throughout the country, you will often find golf courses nearby. The owners of the lodges try to rent to skiers in the winter and golfers in the summer. This is an effective method of spreading their fixed costs over the entire year rather than a few months.

Chart 16–2 shows the quarterly sales, in millions of dollars, of Hercher Sporting Goods, Inc. The Chicago area sporting goods company specializes in selling baseball and softball equipment to high schools, colleges, and youth leagues. It also has several retail outlets in some of the larger shopping malls. There is a distinct seasonal pattern to its business. Most of its sales are in the first and second quarters of the year, when schools and organizations are purchasing equipment for the upcoming season. During the early summer, it keeps busy by selling replacement equipment. It does some business during the holidays (fourth quarter). The late summer (third quarter) is its slow season.

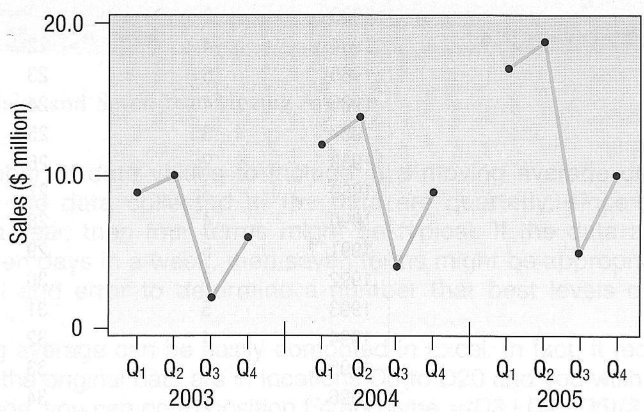
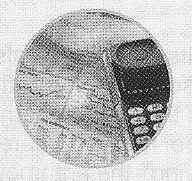


CHART 16–2 Sales of Baseball and Softball Equipment, Hercher Sporting Goods, 2003–2005 by Quarter



Statistics in Action

Statisticians, economists, and business executives are constantly looking for variables that will forecast the country's economy. The production of crude oil, price of gold on world markets, and the Dow Jones average, as well as many published government indexes are variables that have been used with some success. Variables such as the length of hemlines and the winner of the Super Bowl have also been tried. The variable that seems overall to be the most successful is the price of scrap metal. Why? Scrap metal is the beginning of the manufacturing chain. When its demand increases, this is an indication that manufacturing is also increasing.

Irregular Variation

Many analysts prefer to subdivide the **irregular variation** into *episodic* and *residual* variations. Episodic fluctuations are unpredictable, but they can be identified. The initial impact on the economy of a major strike or a war can be identified, but a strike or war cannot be predicted. After the episodic fluctuations have been removed, the remaining variation is called the residual variation. The residual fluctuations, often called chance fluctuations, are unpredictable, and they cannot be identified. Of course, neither episodic nor residual variation can be projected into the future.

A Moving Average

Moving-average method smooths out fluctuations

A **moving average** is useful in smoothing a time series to see its trend. It is also the basic method used in measuring the seasonal fluctuation, described later in the chapter. In contrast to the least squares method, which expresses the trend in terms of a mathematical equation ($\hat{Y} = a + bt$), the moving-average method merely smooths the fluctuations in the data. This is accomplished by "moving" the arithmetic mean values through the time series.

To apply the moving average to a time series, the data should follow a fairly linear trend and have a definite rhythmic pattern of fluctuations (repeating, say, every three years). The data in the following example have three components—trend, cycle, and irregular, abbreviated *T, C,* and *I*. There is no seasonal variation, because the data are recorded annually. What the moving average accomplishes is to average out *C* and *I*. What is left is the trend.

If the duration of the cycles is constant, and if the amplitudes of the cycles are equal, the cyclical and irregular fluctuations are removed entirely using the moving average. The result is a line. For example, in the following time series the cycle repeats itself every seven years, and the amplitude of each cycle is 4; that is, there are exactly four units from the trough (lowest time period) to the peak. The seven-year moving average, therefore, averages out the cyclical and irregular fluctuations perfectly, and the residual is a linear trend.

Compute mean of first seven years

The first step in computing the seven-year moving average is to determine the seven-year moving totals. The total sales for the first seven years (1981–87 inclusive) are $22 million, found by 1 + 2 + 3 + 4 + 5 + 4 + 3. (See Table 16–1.) The total

TABLE 16–1 The Computations for the Seven-Year Moving Average

Year	Sales ($ millions)	Seven-Year Moving Total	Seven-Year Moving Average
1981	$1		
1982	2		
1983	3		
1984	4	22	3.143
1985	5	23	3.286
1986	4	24	3.429
1987	3	25	3.571
1988	2	26	3.714
1989	3	27	3.857
1990	4	28	4.000
1991	5	29	4.143
1992	6	30	4.286
1993	5	31	4.429
1994	4	32	4.571
1995	3	33	4.714
1996	4	34	4.857
1997	5	35	5.000
1998	6	36	5.143
1999	7	37	5.286
2000	6	38	5.429
2001	5	39	5.571
2002	4	40	5.714
2003	5	41	5.857
2004	6		
2005	7		
2006	8		

of $22 million is divided by 7 to determine the arithmetic mean sales per year. The seven-year total (22) and the seven-year mean (3.143) are positioned opposite the middle year for that group of seven, namely, 1984, as shown in Table 16–1. Then the total sales for the next seven years (1982–88 inclusive) are determined. (A convenient way of doing this is to subtract the sales for 1981 [$1 million] from the first seven-year total [$22 million] and add the sales for 1988 [$2 million], to give the new total of $23 million.) The mean of this total, $3.286 million, is positioned opposite the middle year, 1985. The sales data and seven-year moving average are shown graphically in Chart 16–3.

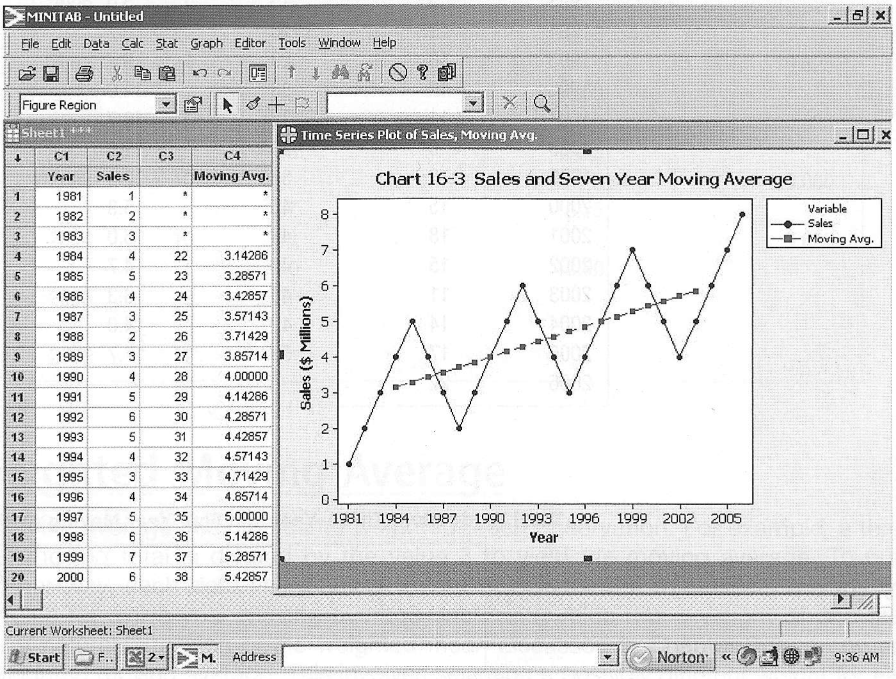


CHART 16–3 Sales and Seven-Year Moving Average

The number of data values to include in a moving average depends on the character of the data collected. If the data are quarterly, since there are four quarters in a year, then four terms might be typical. If the data are daily, since there are seven days in a week, then seven terms might be appropriate. You might also use trial and error to determine a number that best levels out the chance fluctuations.

A moving average can be easily computed in Excel. In fact, it requires only one command. If the original data are in locations D3 to D20 and you wish a three-period moving average, you can go to position E4 and type *=(D3+D4+D5)/3* and then copy that same formula down to position E19. A three-year and a five-year moving average for some production data are shown in Table 16–2 and depicted in Chart 16–4.

Sales, production, and other economic and business series usually do not have (1) periods of oscillation that are of equal length or (2) oscillations that have identical amplitudes. Thus, in practice, the application of a moving average does not result precisely in a line. For example, the production series in Table 16–2 repeats about every five years, but the amplitude of the data varies from one oscillation to another. The trend appears to be upward and somewhat linear. Both moving averages—the three-year and the five-year—seem to adequately describe the trend in production since 1988.

TABLE 16–2 A Three-Year Moving Average and a Five-Year Moving Average

Year	Production, *Y*	Three-Year Moving Total	Three-Year Moving Average	Five-Year Moving Total	Five-Year Moving Average
1988	5				
1989	6	19	6.3		
1990	8	24	8.0	34	6.8
1991	10	23	7.7	32	6.4
1992	5	18	6.0	33	6.6
1993	3	15	5.0	35	7.0
1994	7	20	6.7	37	7.4
1995	10	29	9.7	43	8.6
1996	12	33	11.0	49	9.8
1997	11	32	10.7	55	11.0
1998	9	33	11.0	60	12.0
1999	13	37	12.3	66	13.2
2000	15	46	15.3	70	14.0
2001	18	48	16.0	72	14.4
2002	15	44	14.7	73	14.6
2003	11	40	13.3	75	15.0
2004	14	42	14.0	79	15.8
2005	17	53	17.7		
2006	22				

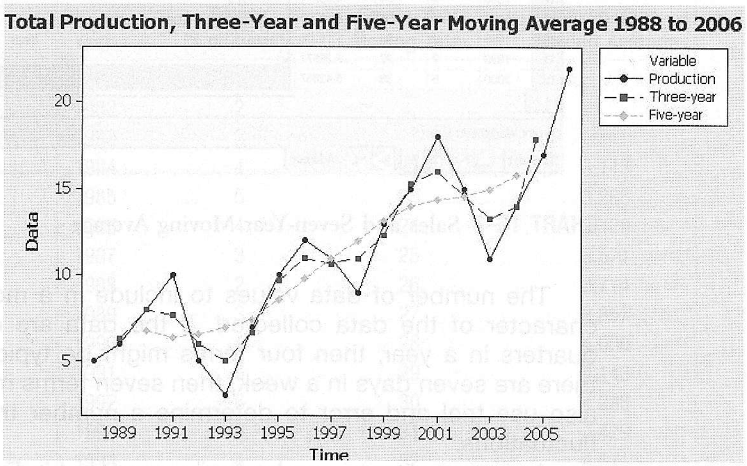


CHART 16–4 A Three-Year and Five-Year Moving Average 1988 to 2006

Determining a moving average for an even-numbered period, such as four years

Four-year, six-year, and other even-numbered-year moving averages present one minor problem regarding the centering of the moving totals and moving averages. Note in Table 16–3 that there is no center time period, so the moving totals are positioned *between* two time periods. The total for the first four years ($42) is positioned between 1999 and 2000. The total for the next four years is $43. The averages of the first four years and the second four years ($10.50 and $10.75, respectively) are averaged, and the resulting figure is centered on 2000. This procedure is repeated until all possible four-year averages are computed.

TABLE 16–3 A Four-Year Moving Average

Year	Sales, *Y*	Four-Year Moving Total	Four-Year Moving Average	Centered Four-Year Moving Average
1998	$ 8			
1999	11			
		$42 (8 + 11 + 9 + 14)	$10.50 ($42 ÷ 4)	
2000	9			10.625
		43 (11 + 9 + 14 + 9)	10.75 ($43 ÷ 4)	
2001	14			10.625
		42	10.50	
2002	9			10.625
		43	10.75	
2003	10			10.000
		37	9.25	
2004	10			9.625
		40	10.00	
2005	8			
2006	12			

Weighted Moving Average

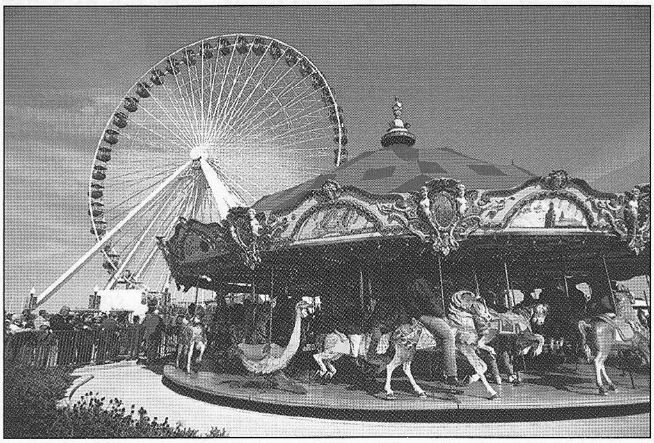
A moving average uses the same weight for each observation. For example, a three-year moving total is divided by the value 3 to yield the moving average. To put it another way, each data value has a weight of one-third in this case. Similarly, you can see for a five-year moving average each data value has a weight of one-fifth.

A natural extension of the weighted mean discussed in Chapter 3 is to compute a weighted moving average. This involves selecting a possibly different weight for each data value and then computing a weighted average of the most recent *n* values as the smoothed value. In the majority of applications we use the smoothed value as a forecast of the future. So, the most recent observation receives the most weight, and the weight decreases for older data values. Notice that for both the simple moving average and for the weighted moving average the sum of the weights is equal to 1.

Suppose, for example, we compute a two-year weighted moving average for the data in Table 16–3 giving twice as much weight to the most recent value. In other words give a weight of 2/3 to the last year and 1/3 to the value immediately before that. Then "forecast" sales for 2000 would be found by (1/3)($8) + (2/3)($11) = $10. The next moving average would be computed as (1/3)($11) + (2/3)($9) = $9.667. Proceeding in the same fashion, the final or 2007 weighted moving average would be (1/3)($8) + (2/3)($12) = $10.667. To summarize the technique of using moving averages, its purpose is to help identify the long-term trend in a time series (because it smooths out short-term fluctuations). It is used to reveal any cyclical and seasonal fluctuations.

Example

Cedar Fair operates seven amusement parks and five separately gated water parks. Its combined attendance (in thousands) for the last 12 years is given in the following table. A partner asks you to study the trend in attendance. Compute a three-year moving average and a three-year weighted moving average with weights of 0.2, 0.3, and 0.5 for successive years.



Year	Attendance (000)
1993	5,761
1994	6,148
1995	6,783
1996	7,445
1997	7,405
1998	11,450
1999	11,224
2000	11,703
2001	11,890
2002	12,380
2003	12,181
2004	12,557

Solution

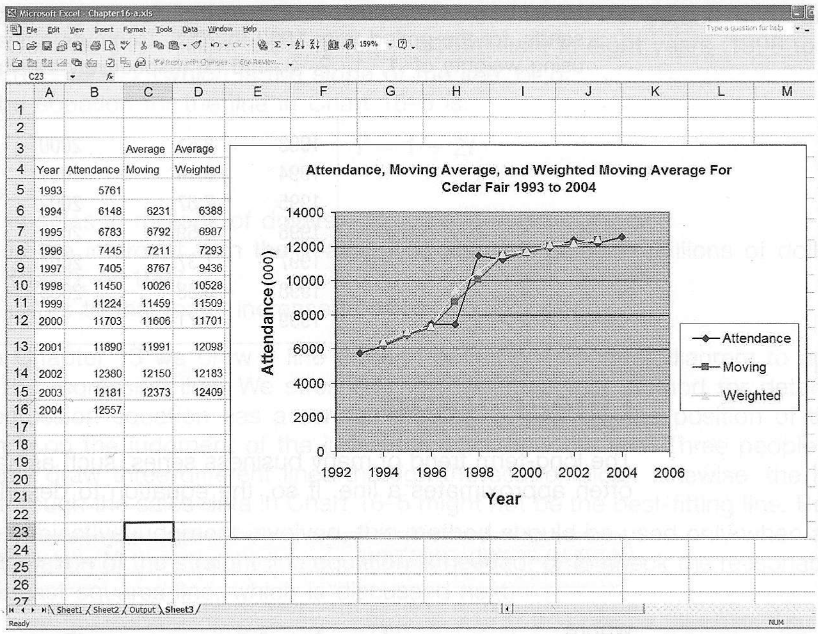
The three-year moving average is:

Year	Attendance (000)	Moving Average	Found by
1993	5,761		
1994	6,148	6,231	(5,761 + 6,148 + 6,783)/3
1995	6,783	6,792	(6,148 + 6,783 + 7,445)/3
1996	7,445	7,211	(6,783 + 7,445 + 7,405)/3
1997	7,405	8,767	(7,445 + 7,405 + 11,450)/3
1998	11,450	10,026	(7,405 + 11,450 + 11,224)/3
1999	11,224	11,459	(11,450 + 11,224 + 11,703)/3
2000	11,703	11,606	(11,224 + 11,703 + 11,890)/3
2001	11,890	11,991	(11,703 + 11,890 + 12,380)/3
2002	12,380	12,150	(11,890 + 12,380 + 12,181)/3
2003	12,181	12,373	(12,380 + 12,181 + 12,557)/3
2004	12,557		

The three-year *weighted* moving average is:

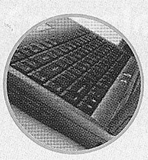
Year	Attendance (000)	Weighted Moving Average	Found by
1993	5,761		
1994	6,148	6,388	.2(5,761) + .3(6,148) + .5(6,783)
1995	6,783	6,987	.2(6,148) + .3(6,783) + .5(7,445)
1996	7,445	7,293	.2(6,783) + .3(7,445) + .5(7,405)
1997	7,405	9,436	.2(7,445) + .3(7,405) + .5(11,450)
1998	11,450	10,528	.2(7,405) + .3(11,450) + .5(11,224)
1999	11,224	11,509	.2(11,450) + .3(11,224) + .5(11,703)
2000	11,703	11,701	.2(11,224) + .3(11,703) + .5(11,890)
2001	11,890	12,098	.2(11,703) + .3(11,890) + .5(12,380)
2002	12,380	12,183	.2(11,890) + .3(12,380) + .5(12,181)
2003	12,181	12,409	.2(12,380) + .3(12,181) + .5(12,557)
2004	12,557		

Excel



Study the graph carefully. You will see that the attendance trend is evenly upward with 360,000 added visitors each year. However, there is a "hop" of approximately 3 million per year between 1997 and 1998. That probably reflects the fact Cedar Fair acquired Knott's Berry Farm in late 1997, leading to a one-time boost in attendance. The weighted moving average follows the data more closely than the moving average. This reflects the additional influence given to the most recent period. In other words, the weighted method, where the most recent period is given the largest weight, won't be quite as smooth. However, it may be more accurate as a forecasting tool.

Self-Review 16–1



Determine a three-year moving average for the sales of Waccamaw Machine Tool, Inc. Plot both the original data and the moving average.

Year	Number Produced (thousands)	Year	Number Produced (thousands)
2000	2	2003	5
2001	6	2004	3
2002	4	2005	10

Exercises

1. Calculate a four-quarter weighted moving average for the number of shares outstanding for the Boxley Box Company for the nine quarters of data. The data are reported in thousands. Apply weights of .1, .2, .3 and .4, respectively, for the quarters. In a few words describe the trend in the number of subscribers.

31-Mar-04	28,766	30-Jun-05	35,102
30-Jun-04	30,057	30-Sep-05	35,308
30-Sep-04	31,336	31-Dec-05	35,203
31-Dec-04	33,240	31-Mar-06	34,386
31-Mar-05	34,610		

2. Listed below is the number of movie tickets sold at the Library Cinema-Complex, in thousands, for the period from 1993 to 2005. Compute a five-year weighted moving average using weights of .1, .1, .2, .3, and .3, respectively. Describe the trend in yield.

1993	8.61	2000	6.61
1994	8.14	2001	5.58
1995	7.67	2002	5.87
1996	6.59	2003	5.94
1997	7.37	2004	5.49
1998	6.88	2005	5.43
1999	6.71		

Linear Trend

The long-term trend of many business series, such as sales, exports, and production, often approximates a line. If so, the equation to describe this growth is:

LINEAR TREND EQUATION $\hat{Y} = a + bt$ **[16–1]**

where:

$\hat{Y}$ read Y hat, is the projected value of the Y variable for a selected value of t.

a is the Y-intercept. It is the estimated value of Y when $t = 0$. Another way to put it is: a is the estimated value of Y where the line crosses the Y-axis when t is zero.

Slope of trend line is b

b is the slope of the line, or the average change in $\hat{Y}$ for each increase of one unit in t.

t is any value of time that is selected.

To illustrate the meaning of $\hat{Y}$, a, b, and t in a time-series problem, a line has been drawn in Chart 16–5 to represent the typical trend of sales. Assume that this company started in business in 1998. This beginning year (1998) has been arbitrarily designated as year 1. Note that sales increased $2 million on the average every year; that is, based on the straight line drawn through the sales data, sales increased from $3 million in 1998 to $5 million in 1999, to $7 million in 2000, to $9 million in 2001, and so on. The slope, or b, is therefore 2. Note too that the line intercepts the Y-axis (when $t = 0$) at $1 million. This point is a. Another way of determining b is to locate the starting place of the straight line in year (1). It is 3 for 1998 in this

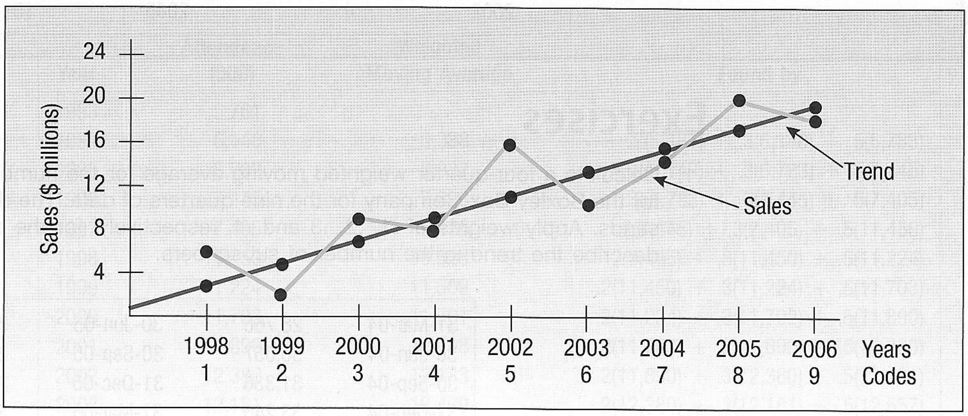


CHART 16–5 A Straight Line Fitted to Sales Data

problem. Then locate the value on the straight line for the last year. It is 19 for 2006. Sales went up \$19 million − \$3 million = \$16 million, in eight years (1998 to 2006). Thus, 16 ÷ 8 = 2, which is the slope of the line, or *b*.

The equation for the line in Chart 16–5 is:

$$\hat{Y} = 1 + 2t$$

where

$\hat{Y}$ is sales in millions of dollars.

1 is the intercept with the *Y*-axis. It is also the sales in millions of dollars for year 0, or 1997.

t refers to the yearly increase in sales.

In Chapter 13 we drew a line through points on a scatter diagram to approximate the regression line. We stressed, however, that this method for determining the regression equation has a serious drawback—namely, the position of the line depends on the judgment of the individual who drew the line. Three people would probably draw three different lines through the scatter plots. Likewise, the line we drew through the sales data in Chart 16–5 might not be the best-fitting line. Because of the subjective judgment involved, this method should be used only when a quick approximation of the straight-line equation is needed, or to check the reasonableness of the least squares line, which is discussed next.

Least Squares Method

In the discussion of simple linear regression in Chapter 13, we showed how the least squares method is used to find the best linear relationship between two variables. In forecasting methods, time is the independent variable and the value of the time series is the dependent variable. Furthermore, we often code the independent variable time to make the equations easier to interpret. In other words, we let *t* be 1 for the first year, 2 for the second, and so on. If a time series includes the sales of General Electric for five years starting in 2002 and continuing through 2006, we would code the year 2002 as 1, 2003 as 2, and 2006 as 5.

Example

The sales of Jensen Foods, a small grocery chain located in southwest Texas, since 2002 are:

Year	Sales ($ million)
2002	7
2003	10
2004	9
2005	11
2006	13

Determine the regression equation. How much are sales increasing each year? What is the sales forecast for 2009?

Solution

To determine the trend equation we could use formula (13–4) to find the slope, or *b* value, and formula (13–5) to locate the intercept, or *a* value. We would substitute *t*, the coded values for the year, for *X* in these equations. Another approach is to use a software package, such as MINITAB or Excel. Chart 16–6 is output from MINITAB. The values for year, coded year, sales, and fitted sales are shown in the lower right portion of the output. The left half is a scatter plot of the data and the fitted regression line.

From the output the trend equation is $\hat{Y} = 6.1 + 1.3t$. How do we interpret this equation? The sales are in millions of dollars. So the value 1.3 tells us that sales

Minitab

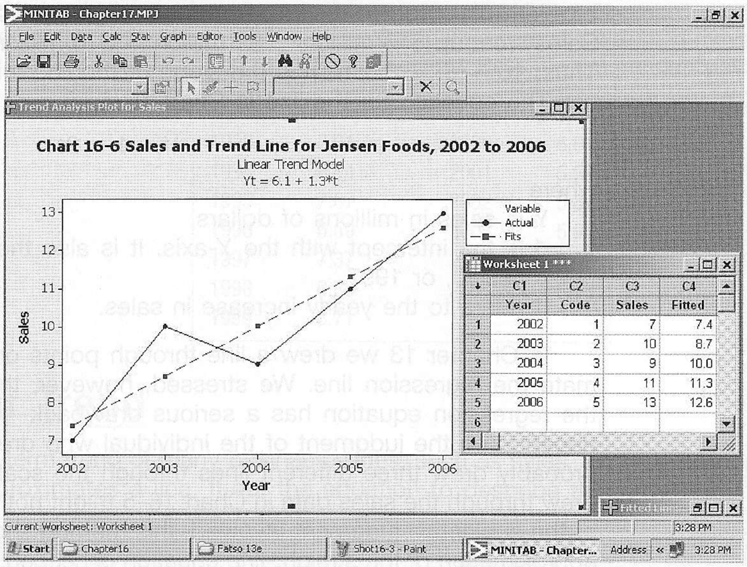


CHART 16–6 Sales and Trend Line, 2002–2006

Statistics in Action

Investors frequently use regression analysis to study the relationship between a particular stock and the general condition of the market. The dependent variable is the monthly percentage change in the value of the stock, and the independent variable is the monthly percentage change in a market index, such as the Standard & Poor's 500 Composite Index. The value of b in the regression equation is the particular stock's *beta coefficient* or just the *beta*. If b is greater than 1, the implication is that the stock is sensitive to market changes. If b is between 0 and 1, the implication is that the stock is not sensitive to market changes.

increased at a rate of 1.3 million dollars per year. The value 6.1 is the estimated value of sales in the year 0. That is the estimate for 2001, which is called the base year. For example, to determine the point on the line for 2005, insert the t value of 4 in the equation. Then $\hat{Y} = 6.1 + 1.3(4) = 11.3$.

If sales, production, or other data approximate a linear trend, the equation developed by the least squares technique can be used to estimate future values. It is reasonable that the sales for Jensen Foods follow a linear trend. So we can use the trend equation to forecast future sales.

See Table 16–4. The year 2002 is coded 1, the year 2004 is coded 3, and 2006 is coded 5. Logically we code 2008 as 7 and 2009 as 8. So we substitute 8 into the trend equation and solve for $\hat{Y}$.

$$\hat{Y} = 6.1 + 1.3t = 6.1 + 1.3(8) = 16.5$$

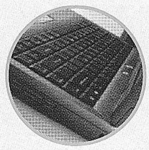
Thus, on the basis of past sales, the estimate for 2009 is $16.5 million.

TABLE 16–4 Calculations for Determining the Points on the Least Squares Line Using the Coded Values

Year	Sales ($ millions), Y	t	$\hat{Y}$	Found by
2002	7	1	7.4	6.1 + 1.3(1)
2003	10	2	8.7	6.1 + 1.3(2)
2004	9	3	10	6.1 + 1.3(3)
2005	11	4	11.3	6.1 + 1.3(4)
2006	13	5	12.6	6.1 + 1.3(5)

In this time series example, there were five years of sales data. Based on those five sales figures, we estimated sales for 2009. Many researchers suggest that we do not project sales, production, and other business and economic series more than $n/2$ time periods into the future where n is the number of data points. If, for example, there are 10 years of data, we would make estimates only up to 5 years into the future ($n/2 = 10/2 = 5$). Others suggest the forecast may be for no longer than 2 years, especially in rapidly changing economic times.

Self-Review 16–2 Annual production of king-size rockers by Wood Products, Inc., since 1999 follows.



Year	Production (thousands)	Year	Production (thousands)
1999	4	2003	11
2000	8	2004	9
2001	5	2005	11
2002	8	2006	14

(a) Plot the production data.
(b) Determine the least squares equation using a software package.
(c) Determine the points on the line for 1999 and 2005. Connect the two points to arrive at the line.
(d) Based on the linear trend equation, what is the estimated production for 2009?

Exercises

3. Listed below are the net sales for Schering-Plough Corporation (a pharmaceutical company) and its subsidiaries for the six years from 1997 to 2004. The net sales are in millions of dollars.

Year	Net Sales	Year	Net Sales
1997	\$ 6,714	2001	\$ 9,762
1998	7,991	2002	10,180
1999	9,075	2003	8,334
2000	9,775	2004	8,272

Determine the least squares equation. According to this information, what are the estimated sales for 2005?

4. Listed below is the net sales in \$ million for Home Depot, Inc., and its subsidiaries from 1993 to 2004.

Year	Net Sales	Year	Net Sales
1993	9,239	1999	38,434
1994	12,477	2000	45,738
1995	15,470	2001	53,553
1996	19,535	2002	58,247
1997	24,156	2003	64,816
1998	30,219	2004	73,094

Determine the least squares equation. On the basis of this information, what are the estimated sales for 2005 and 2006?

5. The following table lists the annual amounts of glass cullet produced by Kimble Glass Works, Inc.

Year	Code	Scrap (tons)
2002	1	2.0
2003	2	4.0
2004	3	3.0
2005	4	5.0
2006	5	6.0

Determine the least squares trend equation. Estimate the amount of scrap for the year 2008.

6. The amounts spent in vending machines in the United States, in billions of dollars, for the years 1999 through 2005 are given below. Determine the least squares trend equation, and estimate vending sales for 2007.

Year	Code	Vending Machine Sales ($ billions)
1999	1	17.5
2000	2	19.0
2001	3	21.0
2002	4	22.7
2003	5	24.5
2004	6	26.7
2005	7	27.3

Nonlinear Trends

The emphasis in the previous discussion was on a time series whose growth or decline approximated a line. A linear trend equation is used to represent the time series when it is believed that the data are increasing (or decreasing) by *equal amounts,* on the average, from one period to another.

Data that increase (or decrease) by *increasing amounts* over a period of time appear *curvilinear* when plotted on paper having an arithmetic scale. To put it another way, data that increase (or decrease) by *equal percents* or *proportions* over a period of time appear curvilinear on arithmetic paper. (See Chart 16–7.)

The trend equation for a time series that does approximate a curvilinear trend, such as the one portrayed in Chart 16–7, may be computed by using the logarithms of the data and the least squares method. The general equation for the logarithmic trend equation is:

LOG TREND EQUATION $$\log \hat{Y} = \log a + \log b(t) \qquad [16\text{–}2]$$

The logarithmic trend equation can be determined for the Gulf Shores Importers data in Chart 16–7 using Excel. The first step is to enter the data, then find the log base 10 of each year's imports. Finally, use the regression procedure to find the least squares equation. To put it another way, we take the log of each year's data, then use the logs as the dependent variable and the coded year as the independent variable.

Excel

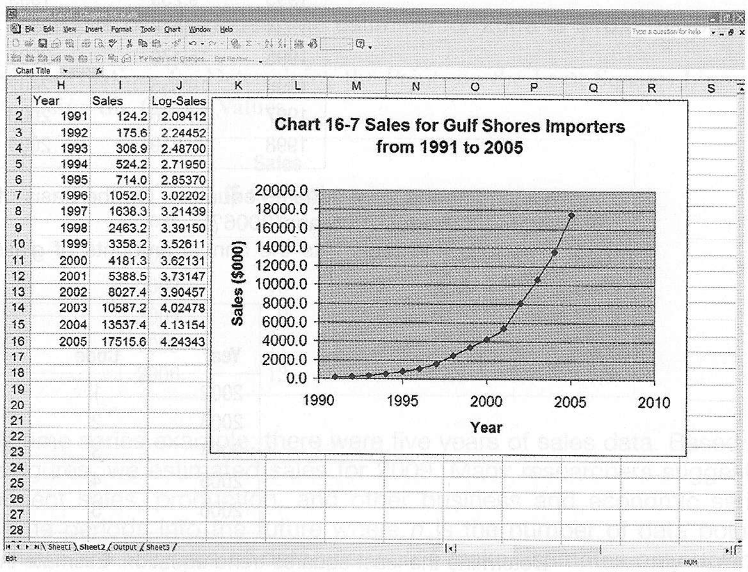


CHART 16–7 Sales for Gulf Shores Importers from 1991–2005

Microsoft Excel - Chapter16-a.xls

Year	Sales	Log-Sales	Code
1991	124.2	2.09412	1
1992	175.6	2.24452	2
1993	306.9	2.48700	3
1994	524.2	2.71950	4
1995	714.0	2.85370	5
1996	1052.0	3.02202	6
1997	1638.3	3.21439	7
1998	2463.2	3.39150	8
1999	3358.2	3.52611	9
2000	4181.3	3.62131	10
2001	5388.5	3.73147	11
2002	8027.4	3.90457	12
2003	10587.2	4.02478	13
2004	13537.4	4.13154	14
2005	17515.6	4.24343	15

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.993953309
R Square	0.98794318
Adjusted R	0.987015732
Standard E	0.078625262
Observatic	15

ANOVA

	df	*SS*	*MS*	*F*
Regressio	1	6.58517	6.58517	1065.23
Residual	13	0.08037	0.00618	
Total	14	6.66553		

	Coefficients	*andard Err*	*t Stat*	*P-value*
Intercept	2.053805	0.04272	48.0741	5E-16
Code	0.153357	0.0047	32.6378	7.4E-14

Excel

The regression equation is $\hat{Y} = 2.053805 + 0.153357t$, which is the log form. We now have a trend equation in terms of percent of change. That is, the value 0.153357 is the percent of change in $\hat{Y}$ for each unit increase in t. This value is similar to the geometric mean described in Chapter 3.

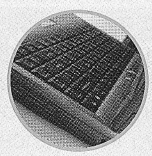
The log of b is 0.153357 and its antilog or inverse is 1.423498. If we subtract 1 from this value, as we did in Chapter 3, the value 0.423498 indicates the geometric mean annual rate of increase from 1991 to 2005. We conclude that imports increased at a rate of 42.35 percent annually during the period.

We can also use the logarithmic trend equation to make estimates of future values. Suppose we want to estimate the imports in the year 2009. The first step is to determine the code for the year 2009, which is 19. How did we get 19? The year 2005 has a code of 15 and the year 2009 is four years later, so $15 + 4 = 19$. The log of imports for the year 2009 is

$$\hat{Y} = 2.053805 + 0.153357t = 2.053805 + 0.153357(19) = 4.967588$$

To find the estimated imports for the year 2009, we need the antilog of 4.967588. It is 92,809. This is our estimate of the number of imports for 2009. Recall that the data were in thousands of dollars, so the estimate is $92,809,000.

Self-Review 16–3



Sales at Tomlin Manufacturing since 2002 are:

Year	Sales ($ millions)
2002	2.13
2003	18.10
2004	39.80
2005	81.40
2006	112.00

(a) Determine the logarithmic trend equation for the sales data.
(b) Sales increased by what percent annually from 2002 to 2006?
(c) What is the projected sales amount for 2007?

Exercises

7. Sally's Software, Inc., is a rapidly growing supplier of computer software to the Sarasota area. Sales for the last five years are given below.

Year	Sales ($ millions)
2002	1.1
2003	1.5
2004	2.0
2005	2.4
2006	3.1

a. Determine the logarithmic trend equation.
b. By what percent did sales increase, on the average, during the period?
c. Estimate sales for the year 2009.

8. It appears that the imports of carbon black have been increasing by about 10 percent annually.

Year	Imports of Carbon Black (thousands of tons)	Year	Imports of Carbon Black (thousands of tons)
1999	92.0	2003	135.0
2000	101.0	2004	149.0
2001	112.0	2005	163.0
2002	124.0	2006	180.0

a. Determine the logarithmic trend equation.
b. By what percent did imports increase, on the average, during the period?
c. Estimate imports for the year 2009.

Seasonal Variation

We mentioned that *seasonal variation* is another of the components of a time series. Business series, such as automobile sales, shipments of soft-drink bottles, and residential construction, have periods of above-average and below-average activity each year.

In the area of production, one of the reasons for analyzing seasonal fluctuations is to have a sufficient supply of raw materials on hand to meet the varying seasonal demand. The glass container division of a large glass company, for example, manufactures nonreturnable beer bottles, iodine bottles, aspirin bottles, bottles for rubber cement, and so on. The production scheduling department must know how many bottles to produce and when to produce each kind. A run of too many bottles of one kind may cause a serious storage problem. Production cannot be based entirely on orders on hand, because many orders are telephoned in for immediate shipment. Since the demand for many of the bottles varies according to the season, a forecast a year or two in advance, by month, is essential to good scheduling.

An analysis of seasonal fluctuations over a period of years can also help in evaluating current sales. The typical sales of department stores in the United States, excluding mail-order sales, are expressed as indexes in Table 16–5. Each index represents the average sales for a period of several years. The actual sales for some months were above average (which is represented by an index over 100.0), and the sales for other months were below average. The index of 126.8 for December indicates that, typically, sales for December are 26.8 percent above an average month; the index of 86.0 for July indicates that department store sales for July are typically 14 percent below an average month.

TABLE 16–5 Typical Seasonal Indexes for U.S. Department Store Sales, Excluding Mail-Order Sales

January	87.0	July	86.0
February	83.2	August	99.7
March	100.5	September	101.4
April	106.5	October	105.8
May	101.6	November	111.9
June	89.6	December	126.8

Suppose an enterprising store manager, in an effort to stimulate sales during December, introduced a number of unique promotions, including bands of carolers strolling through the store singing holiday songs, large mechanical exhibits, and clerks dressed in Santa Claus costumes. When the index of sales was computed for that December, it was 150.0. Compared with the typical December sales of 126.8, it was concluded that the promotional program was a huge success.

Determining a Seasonal Index

Objective: To determine a set of "typical" seasonal indexes

A typical set of monthly indexes consists of 12 indexes that are representative of the data for a 12-month period. Logically, there are four typical seasonal indexes for data reported quarterly. Each index is a percent, with the average for the year equal to 100.0; that is, each monthly index indicates the level of sales, production, or another variable in relation to the annual average of 100.0. A typical index of 96.0 for January indicates that sales (or whatever the variable is) are usually 4 percent below the average for the year. An index of 107.2 for October means that the variable is typically 7.2 percent above the annual average.

Several methods have been developed to measure the typical seasonal fluctuation in a time series. The method most commonly used to compute the typical seasonal pattern is called the **ratio-to-moving-average method.** It eliminates the trend, cyclical, and irregular components from the original data (Y). In the following discussion, T refers to trend, C to cyclical, S to seasonal, and I to irregular variation. The numbers that result are called the *typical seasonal index*.

We will discuss in detail the steps followed in arriving at typical seasonal indexes using the ratio-to-moving-average method. The data of interest might be monthly or quarterly. To illustrate, we have chosen the quarterly sales of Toys International. First, we will show the steps needed to arrive at a set of typical quarterly indexes. Then we use MegaStat Excel and MINITAB software to calculate the seasonal indexes.

Example

Table 16–6 shows the quarterly sales for Toys International for the years 2001 through 2006. The sales are reported in millions of dollars. Determine a quarterly seasonal index using the ratio-to-moving-average method.

TABLE 16–6 Quarterly Sales of Toys International ($ millions)

Year	Winter	Spring	Summer	Fall
2001	6.7	4.6	10.0	12.7
2002	6.5	4.6	9.8	13.6
2003	6.9	5.0	10.4	14.1
2004	7.0	5.5	10.8	15.0
2005	7.1	5.7	11.1	14.5
2006	8.0	6.2	11.4	14.9

Solution

Chart 16–8 depicts the quarterly sales for Toys International over the six-year period. Notice the seasonal nature of the sales. For each year, the fourth-quarter sales are the largest and the second-quarter sales are the smallest. Also, there is a moderate increase in the sales from one year to the next. To observe this feature, look only at the six fourth-quarter sales values. Over the six-year period, the sales in the fourth quarter increased. If you connect these points in your mind, you can visualize fourth-quarter sales increasing for 2007.

Minitab

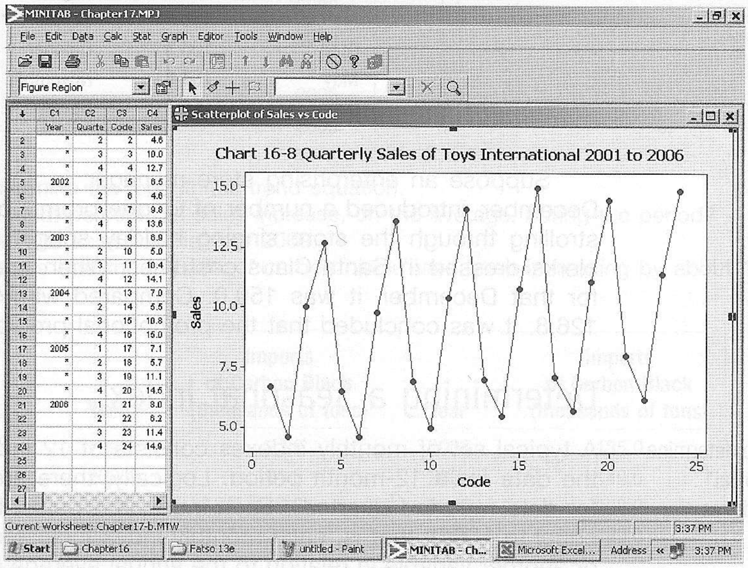


CHART 16–8 Quarterly Sales of Toys International 2001–2006

There are six steps to determining the quarterly seasonal indexes.

Step 1: For the following discussion, refer to Table 16–7. The first step is to determine the four-quarter moving total for 2001. Starting with the winter quarter of 2001, we add \$6.7, \$4.6, \$10.0, and \$12.7. The total is \$34.0 (million). The four-quarter total is "moved along" by adding the spring, summer, and fall sales of 2001 to the winter sales of 2002. The total is \$33.8 (million), found by 4.6 + 10.0 + 12.7 + 6.5. This procedure is continued for the quarterly sales for each of the six years. Column 2 of Table 16–7 shows all of the moving totals. Note that the moving total

TABLE 16–7 Computations Needed for the Specific Seasonal Indexes

Year	Quarter	(1) Sales ($ millions)	(2) Four-Quarter Total	(3) Four-Quarter Moving Average	(4) Centered Moving Average	(5) Specific Seasonal
2001	Winter	6.7				
	Spring	4.6				
			34.0	8.500		
	Summer	10.0			8.475	1.180
			33.8	8.450		
	Fall	12.7			8.450	1.503
			33.8	8.450		
2002	Winter	6.5			8.425	0.772

(continued)

Year	Quarter	(1) Sales ($ millions)	(2) Four-Quarter Total	(3) Four-Quarter Moving Average	(4) Centered Moving Average	(5) Specific Seasonal
			33.6	8.400		
	Spring	4.6			8.513	0.540
			34.5	8.625		
	Summer	9.8			8.675	1.130
			34.9	8.725		
	Fall	13.6			8.775	1.550
			35.3	8.825		
2003	Winter	6.9			8.900	0.775
			35.9	8.975		
	Spring	5.0			9.038	0.553
			36.4	9.100		
	Summer	10.4			9.113	1.141
			36.5	9.125		
	Fall	14.1			9.188	1.535
			37.0	9.250		
2004	Winter	7.0			9.300	0.753
			37.4	9.350		
	Spring	5.5			9.463	0.581
			38.3	9.575		
	Summer	10.8			9.588	1.126
			38.4	9.600		
	Fall	15.0			9.625	1.558
			38.6	9.650		
2005	Winter	7.1			9.688	0.733
			38.9	9.725		
	Spring	5.7			9.663	0.590
			38.4	9.600		
	Summer	11.1			9.713	1.143
			39.3	9.825		
	Fall	14.5			9.888	1.466
			39.8	9.950		
2006	Winter	8.0			9.888	0.801
			40.1	10.025		
	Spring	6.2			10.075	0.615
			40.5	10.125		
	Summer	11.4				
	Fall	14.9				

34.0 is positioned between the spring and summer sales of 2001. The next moving total, 33.8, is positioned between sales for summer and fall 2001, and so on. Check the totals frequently to avoid arithmetic errors.

Step 2: Each quarterly moving total in column 2 is divided by 4 to give the four-quarter moving average. (See column 3.) All the moving averages are still positioned between the quarters. For example, the first moving average (8.500) is positioned between spring and summer 2001.

Step 3: The moving averages are then centered. The first centered moving average is found by $(8.500 + 8.450)/2 = 8.475$ and centered opposite summer 2001. The second moving average is found by $(8.450 + 8.450)/2 = 8.450$. The others are found similarly. Note in column 4 that each centered moving average is positioned on a particular quarter.

Step 4: The **specific seasonal index** for each quarter is then computed by dividing the sales in column 1 by the centered moving average in column 4. The specific seasonal index reports the ratio of the original time series value to the moving average. To explain further, if the time series is represented by *TSCI* and the moving average by *TC,* then, algebraically, if we compute *TSCI*/*TC,* the result is the specified seasonal component *SI*. The specific seasonal index for the summer quarter of 2001 is 1.180, found by 10.0/8.475.

Step 5: The specific seasonal indexes are organized in Table 16–8. This table will help us locate the specific seasonals for the corresponding quarters. The values 1.180, 1.130, 1.141, 1.126, and 1.143 all represent estimates of the typical seasonal index for the summer quarter. A reasonable method to find a typical seasonal index is to average these values in order to eliminate the irregular component. So we find the typical index for the summer quarter by (1.180 + 1.130 + 1.141 + 1.126 + 1.143)/5 = 1.144. We used the arithmetic mean, but the median or a modified mean can also be used.

TABLE 16–8 Calculations Needed for Typical Quarterly Indexes

Year	Winter	Spring	Summer	Fall	
2001			1.180	1.503	
2002	0.772	0.540	1.130	1.550	
2003	0.775	0.553	1.141	1.535	
2004	0.753	0.581	1.126	1.558	
2005	0.733	0.590	1.143	1.466	
2006	0.801	0.615			
Total	3.834	2.879	5.720	7.612	
Mean	0.767	0.576	1.144	1.522	4.009
Adjusted	0.765	0.575	1.141	1.519	4.000
Index	76.5	57.5	114.1	151.9	

Step 6: The four quarterly means (0.767, 0.576, 1.144, and 1.522) should theoretically total 4.00 because the average is set at 1.0. The total of the four quarterly means may not exactly equal 4.00 due to rounding. In this problem the total of the means is 4.009. A *correction factor* is therefore applied to each of the four means to force them to total 4.00.

CORRECTION FACTOR FOR ADJUSTING QUARTERLY MEANS

$$\text{Correction factor} = \frac{4.00}{\text{Total of four means}} \qquad [16\text{–}3]$$

In this example,

$$\text{Correction factor} = \frac{4.00}{4.009} = .9978$$

The adjusted winter quarterly index is, therefore, .767(.9978) = .765. Each of the means is adjusted downward so that the total of the four quarterly means is 4.00. Usually indexes are reported as percentages, so each value in the last row of Table 16–8 has been multiplied by 100. So the index for the winter quarter is 76.5 and for the fall it is 151.9. How are these values interpreted? Sales for the fall quarter are 51.9 percent above the typical quarter, and for winter they are 23.5 below the typical quarter (100.0 − 76.5). These findings should not surprise you. The period prior to Christmas (the fall quarter) is when toy sales are brisk. After Christmas (the winter quarter), sales of the toys decline drastically.

As noted earlier software will perform the calculations and output the results. The MegaStat Excel output is shown below. Use of software will greatly reduce the computational time and the chance of an error in arithmetic, but you should understand the steps in the process, as outlined earlier. There can be slight differences in the answers, due to the number of digits carried in the calculations.

Centered Moving Average and Deseasonalization

t	Year	Quarter	Sales	Centered Moving Average	Ratio to CMA	Seasonal Indexes	Sales Deseasonalized
1	2001	1	6.70			0.765	8.759
2	2001	2	4.60			0.575	8.004
3	2001	3	10.00	8.475	1.180	1.141	8.761
4	2001	4	12.70	8.450	1.503	1.519	8.361
5	2002	1	6.50	8.425	0.772	0.765	8.498
6	2002	2	4.60	8.513	0.540	0.575	8.004
7	2002	3	9.80	8.675	1.130	1.141	8.586
8	2002	4	13.60	8.775	1.550	1.519	8.953
9	2003	1	6.90	8.900	0.775	0.765	9.021
10	2003	2	5.00	9.038	0.553	0.575	8.700
11	2003	3	10.40	9.113	1.141	1.141	9.112
12	2003	4	14.10	9.188	1.535	1.519	9.283
13	2004	1	7.00	9.300	0.753	0.765	9.151
14	2004	2	5.50	9.463	0.581	0.575	9.570
15	2004	3	10.80	9.588	1.126	1.141	9.462
16	2004	4	15.00	9.625	1.558	1.519	9.875
17	2005	1	7.10	9.688	0.733	0.765	9.282
18	2005	2	5.70	9.663	0.590	0.575	9.918
19	2005	3	11.10	9.713	1.143	1.141	9.725
20	2005	4	14.50	9.888	1.466	1.519	9.546
21	2006	1	8.00	9.988	0.801	0.765	10.459
22	2006	2	6.20	10.075	0.615	0.575	10.788
23	2006	3	11.40			1.141	9.988
24	2006	4	14.90			1.519	9.809

MegaStat

Calculation of Seasonal Indexes

	1	2	3	4	
2001			1.180	1.503	
2002	0.772	0.540	1.130	1.550	
2003	0.775	0.553	1.141	1.535	
2004	0.753	0.581	1.126	1.558	
2005	0.733	0.590	1.143	1.466	
2006	0.801	0.615			
mean:	0.767	0.576	1.144	1.522	4.009
adjusted:	0.765	0.575	1.141	1.519	4.000

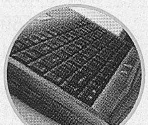
Now we briefly summarize the reasoning underlying the preceding calculations. The original data in column 1 of Table 16–7 contain trend (*T*), cyclical (*C*), seasonal (*S*), and irregular (*I*) components. The ultimate objective is to remove seasonal (*S*) from the original sales valuation.

Columns 2 and 3 in Table 16–7 are concerned with deriving the centered moving average given in column 4. Basically, we "average out" the seasonal and irregular fluctuations from the original data in column 1. Thus, in column 4 we have only trend and cyclical (*TC*).

Next, we divide the sales data in column 1 (*TCSI*) by the centered four-quarter moving average in column 4 (*TC*) to arrive at the specific seasonals in column 5 (*SI*). In terms of letters, $TCSI/TC = SI$. We multiply *SI* by 100.0 to express the typical seasonal in index form.

Finally, we take the mean of all the winter typical indexes, all the spring indexes, and so on. This averaging eliminates most of the irregular fluctuations from the specific seasonals, and the resulting four indexes indicate the typical seasonal sales pattern.

Self-Review 16–4



Teton Village, Wyoming, near Grand Teton Park and Yellowstone Park, contains shops, restaurants, and motels. The village has two peak seasons—winter, for skiing on the 10,000-foot slopes, and summer, for tourists visiting the parks. The number of visitors (in thousands) by quarter for five years follows.

	Quarter			
Year	**Winter**	**Spring**	**Summer**	**Fall**
2002	117.0	80.7	129.6	76.1
2003	118.6	82.5	121.4	77.0
2004	114.0	84.3	119.9	75.0
2005	120.7	79.6	130.7	69.6
2006	125.2	80.2	127.6	72.0

(a) Develop the typical seasonal pattern for Teton Village using the ratio-to-moving-average method.
(b) Explain the typical index for the winter season.

Exercises

9. Victor Anderson, the owner of Anderson Belts, Inc., is studying absenteeism among his employees. His workforce is small, consisting of only five employees. For the last three years he recorded the following number of employee absences, in days, for each quarter.

	Quarter			
Year	**I**	**II**	**III**	**IV**
2004	4	10	7	3
2005	5	12	9	4
2006	6	16	12	4

Determine a typical seasonal index for each of the four quarters.

10. Appliance Center sells a variety of electronic equipment and home appliances. For the last four years the following quarterly sales (in $ millions) were reported.

	Quarter			
Year	**I**	**II**	**III**	**IV**
2003	5.3	4.1	6.8	6.7
2004	4.8	3.8	5.6	6.8
2005	4.3	3.8	5.7	6.0
2006	5.6	4.6	6.4	5.9

Determine a typical seasonal index for each of the four quarters.

Deseasonalizing Data

A set of typical indexes is very useful in adjusting a sales series, for example, for seasonal fluctuations. The resulting sales series is called **deseasonalized sales** or **seasonally adjusted sales.** The reason for deseasonalizing the sales series is to remove the seasonal fluctuations so that the trend and cycle can be studied. To illustrate the procedure, the quarterly sales totals of Toys International from Table 16–6 are repeated in column 1 of Table 16–9.

TABLE 16–9 Actual and Deseasonalized Sales for Toys International

Year	Quarter	(1) Sales	(2) Seasonal Index	(3) Deseasonalized Sales
2001	Winter	6.7	0.765	8.76
	Spring	4.6	0.575	8.00
	Summer	10.0	1.141	8.76
	Fall	12.7	1.519	8.36
2002	Winter	6.5	0.765	8.50
	Spring	4.6	0.575	8.00
	Summer	9.8	1.141	8.59
	Fall	13.6	1.519	8.95
2003	Winter	6.9	0.765	9.02
	Spring	5.0	0.575	8.70
	Summer	10.4	1.141	9.11
	Fall	14.1	1.519	9.28
2004	Winter	7.0	0.765	9.15
	Spring	5.5	0.575	9.57
	Summer	10.8	1.141	9.47
	Fall	15.0	1.519	9.87
2005	Winter	7.1	0.765	9.28
	Spring	5.7	0.575	9.91
	Summer	11.1	1.141	9.73
	Fall	14.5	1.519	9.55
2006	Winter	8.0	0.765	10.46
	Spring	6.2	0.575	10.79
	Summer	11.4	1.141	9.99
	Fall	14.9	1.519	9.81

To remove the effect of seasonal variation, the sales amount for each quarter (which contains trend, cyclical, irregular, and seasonal effects) is divided by the seasonal index for that quarter, that is, *TSCI/S*. For example, the actual sales for the first quarter of 2001 were $6.7 million. The seasonal index for the winter quarter is 76.5 percent, using the MegaStat results on page 623. The index of 76.5 indicates that sales for the first quarter are typically 23.5 percent below the average for a typical quarter. By dividing the actual sales of $6.7 million by 76.5 and multiplying the result by 100, we find the *deseasonalized sales* value—that is, removed the seasonal effect on sales—for the first quarter of 2001. It is $8,758,170, found by ($6,700,000/76.5)100. We continue this process for the other quarters in column 3 of Table 16–9, with the results reported in millions of dollars. Because the seasonal component has been removed (divided out) from the quarterly sales, the deseasonalized sales figure contains only the trend (*T*), cyclical (*C*), and irregular (*I*) components. Scanning the deseasonalized sales in column 3 of Table 16–9, we see that the sales of toys showed a moderate increase over the six-year period. Chart 16–9 shows both the actual sales and the deseasonalized sales. It is clear that removing the seasonal factor allows us to focus on the overall long-term trend of sales. We will also be able to determine the regression equation of the trend data and use it to forecast future sales.

Using Deseasonalized Data to Forecast

The procedure for identifying trend and the seasonal adjustments can be combined to yield seasonally adjusted forecasts. To identify the trend, we determine the least squares trend equation on the deseasonalized historical data. Then we project this

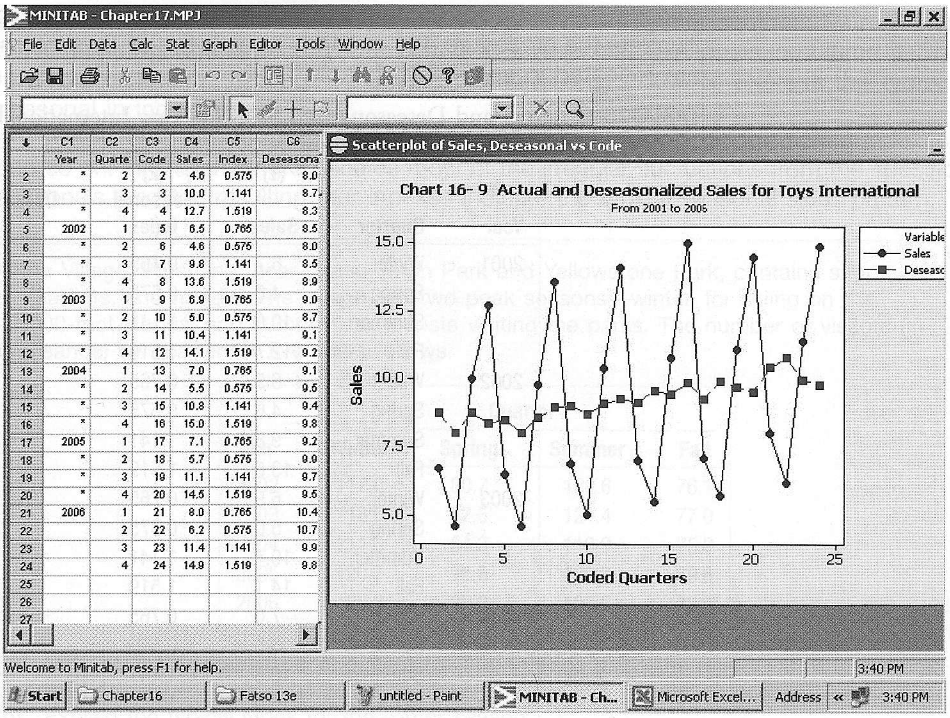


CHART 16–9 Actual and Deseasonalized Sales for Toys International from 2001 to 2006

trend into future periods, and finally we adjust these trend values to account for the seasonal factors. The following example will help to clarify.

Example

Toys International would like to forecast its sales for each quarter of 2007. Use the information in Table 16–9 to determine the forecast.

Solution

The deseasonalized data depicted in Chart 16–9 seems to follow a straight line. Hence it is reasonable to develop a linear trend equation based on these data. The deseasonalized trend equation is:

$$\hat{Y} = a + bt$$

where

$\hat{Y}$ is the estimated trend value for Toys International sales for the period t.
a is the intercept of the trend line at time 0.
b is the slope of the line.
t is the coded time period.

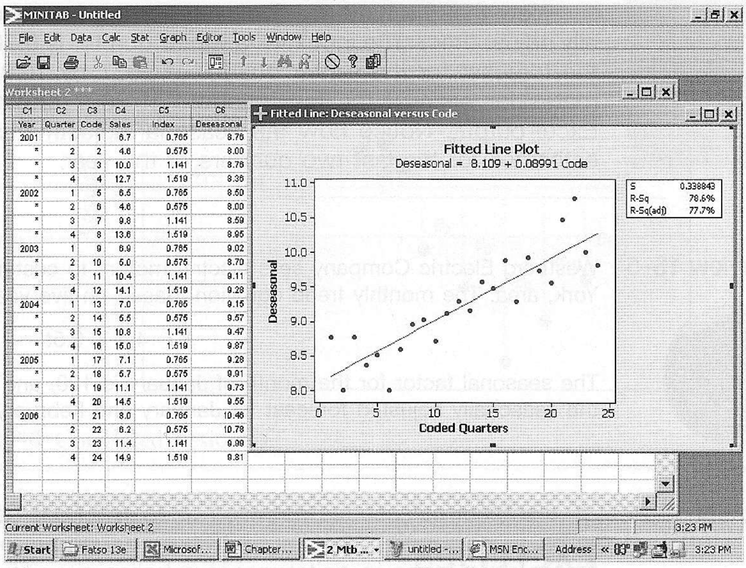
The winter quarter of 2001 is the first quarter, so it is coded 1, the spring quarter of 2001 is coded 2, and so on. The last quarter of 2006 is coded 24. These coded values are shown in the data section of the MINITAB output associated with Chart 16–9.

We use MINITAB to find the regression equation. The output follows. The output includes a scatter diagram of the coded time periods and the deseasonalized sales as well as the regression line.

The equation for the trend line is:

$$\hat{Y} = 8.109 + .08991t$$

The slope of the trend line is .08991. This shows that over the 24 quarters the deseasonalized sales increased at a rate of 0.08991 ($ million) per quarter, or $89,910 per quarter. The value of 8.109 is the intercept of the trend line on the Y-axis (i.e., for $t = 0$).



Year	Quarter	Code	Sales	Index	Deseasonal
2001	1	1	6.7	0.765	8.76
*	2	2	4.6	0.575	8.00
*	3	3	10.0	1.141	8.76
*	4	4	12.7	1.519	8.36
2002	1	5	6.5	0.765	8.50
*	2	6	4.6	0.575	8.00
*	3	7	9.8	1.141	8.59
*	4	8	13.6	1.519	8.95
2003	1	9	6.9	0.765	9.02
*	2	10	5.0	0.575	8.70
*	3	11	10.4	1.141	9.11
*	4	12	14.1	1.519	9.28
2004	1	13	7.0	0.765	9.15
*	2	14	5.5	0.575	9.57
*	3	15	10.8	1.141	9.47
*	4	16	15.0	1.519	9.87
2005	1	17	7.1	0.765	9.28
*	2	18	5.7	0.575	9.91
*	3	19	11.1	1.141	9.73
*	4	20	14.5	1.519	9.55
2006	1	21	8.0	0.765	10.46
	2	22	6.2	0.575	10.78
	3	23	11.4	1.141	9.99
	4	24	14.9	1.519	9.81

Statistics in Action

Forecasts are not always correct. The reality is that a forecast may just be a best guess as to what will happen. What are the reasons forecasts are not correct? One expert lists eight common errors: (1) Failure to carefully examine the assumptions, (2) Limited expertise, (3) Lack of imagination, (4) Neglect of constraints, (5) Excessive optimism, (6) Reliance on mechanical extrapolation, (7) Premature closure, and (8) Overspecification.

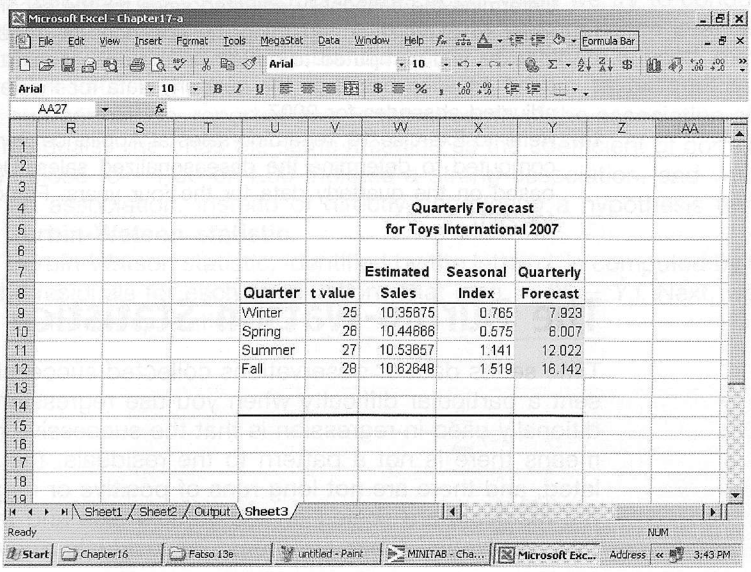
The MINITAB system also outputs the coefficient of determination. This value, called R^2, is 78.6 percent. It is shown in the upper right of the MINITAB output. We can use this value as an indication of the fit of the data. Because this is *not* sample information, technically we should not use R^2 for judging a regression equation. However, it will serve to quickly evaluate the fit of the deseasonalized sales data. In this instance, because R^2 is rather large, we conclude the deseasonalized sales of Toys International are effectively explained by a linear trend equation.

If we assume that the past 24 periods are a good indicator of future sales, we can use the trend equation to estimate future sales. For example, for the winter quarter of 2007 the value of t is 25. Therefore, the estimated sales of that period is 10.35675, found by

$$\hat{Y} = 8.109 + .08991t = 8.109 + .08991(25) = 10.35675$$

The estimated deseasonalized sales for the winter quarter of 2007 are $10,356,750. This is the sales forecast before we consider the effects of seasonality.

We use the same procedure and an Excel spreadsheet to determine a forecast for each of the four quarters of 2007. A partial Excel output follows.

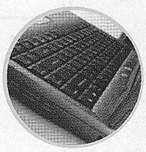


Quarterly Forecast
for Toys International 2007

Quarter	t value	Estimated Sales	Seasonal Index	Quarterly Forecast
Winter	25	10.35675	0.765	7.923
Spring	26	10.44666	0.575	6.007
Summer	27	10.53657	1.141	12.022
Fall	28	10.62648	1.519	16.142

Now that we have the forecasts for the four quarters of 2007, we can seasonally adjust them. The index for a winter quarter is 0.765. So we can seasonally adjust the forecast for the winter quarter of 2007 by 10.35675(0.765) = 7.923. The estimates for each of the four quarters of 2007 are in the right-hand column of the Excel output. Notice how the seasonal adjustments drastically increase the sales estimates for the last two quarters of the year.

Self-Review 16–5



Westberg Electric Company sells electric motors to customers in the Jamestown, New York, area. The monthly trend equation, based on five years of monthly data, is

$$\hat{Y} = 4.4 + 0.5t$$

The seasonal factor for the month of January is 120, and it is 95 for February. Determine the seasonally adjusted forecast for January and February of the sixth year.

Exercises

11. The planning department of Padget and Kure Shoes, the manufacturer of an exclusive brand of women's shoes, developed the following trend equation, in millions of pairs, based on five years of quarterly data.

$$\hat{Y} = 3.30 + 1.75t$$

The following table gives the seasonal factors for each quarter.

	Quarter			
	I	II	III	IV
Index	110.0	120.0	80.0	90.0

Determine the seasonally adjusted forecast for each of the four quarters of the sixth year.

12. Team Sports, Inc., sells sporting goods to high schools and colleges via a nationally distributed catalog. Management at Team Sports estimates it will sell 2,000 Wilson Model A2000 catcher's mitts next year. The deseasonalized sales are projected to be the same for each of the four quarters next year. The seasonal factor for the second quarter is 145. Determine the seasonally adjusted sales for the second quarter of next year.

13. Refer to Exercise 9, regarding the absences at Anderson Belts, Inc. Use the seasonal indexes you computed to determine the deseasonalized absences. Determine the linear trend equation based on the quarterly data for the three years. Forecast the seasonally adjusted absences for 2007.

14. Refer to Exercise 10, regarding sales at Appliance Center. Use the seasonal indexes you computed to determine the deseasonalized sales. Determine the linear trend equation based on the quarterly data for the four years. Forecast the seasonally adjusted sales for 2007.

The Durbin-Watson Statistic

Time series data or observations collected successively over a period of time present a particular difficulty when you use regression. One of the assumptions traditionally used in regression is that the successive residuals are independent. This means there is not a pattern to the residuals, the residuals are not highly correlated, and there are not long runs of positive or negative residuals. In Chart 16–10 the residuals are scaled on the vertical axis and the $\hat{Y}$ values along the horizontal axis. Notice there are "runs" of residuals above and below the 0 line. If we

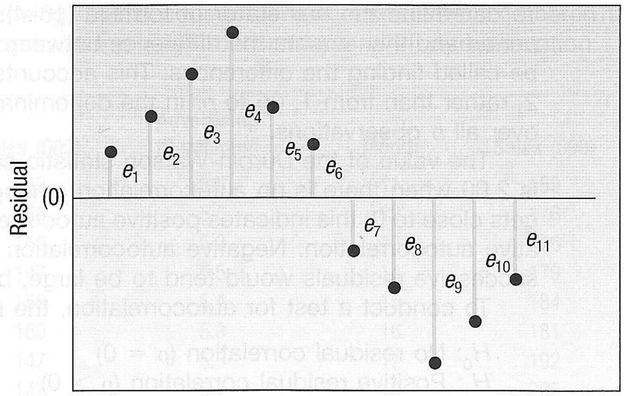


CHART 16–10 Correlated Residuals

computed the correlation between successive residuals, it is likely the correlation would be strong.

This condition is called autocorrelation or serial correlation.

AUTOCORRELATION Successive residuals are correlated.

Successive residuals are correlated in time series data because an event in one time period often influences the event in the next period. To explain, the owner of a furniture store decides to have a sale this month and spends a large amount of money advertising the event. We would expect a correlation between sales and advertising expense, but all the results of the increase in advertising are not experienced this month. It is likely that some of the effect of the advertising carries over into next month. Therefore, we expect correlation among the residuals.

The regression relationship in a time series is written

$$Y_t = \alpha + \beta_1 X_t + \varepsilon_t$$

where the subscript t is used in place of i to suggest the data were collected over time.

If the residuals are correlated, problems occur when we try to conduct tests of hypotheses about the regression coefficients. Also, a confidence interval or a prediction interval, where the multiple standard error of estimate is used, may not yield the correct results.

The autocorrelation, reported as r, is the strength of the association among successive residuals. The r has the same meaning as the coefficient of correlation. That is, values close to −1.00 or 1.00 indicate a strong association, and values near 0 indicate no association. Instead of directly conducting a hypothesis test on r, we use the **Durbin-Watson statistic.**

The Durbin-Watson statistic, identified by the letter d, is computed by first determining the residuals for each observation. That is, $e_t = (Y_t - \hat{Y}_t)$. Next, we compute d using the following relationship.

DURBIN-WATSON STATISTIC

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} (e_t)^2} \qquad [16\text{–}4]$$

To determine the numerator of formula (16–4), we "lag" each of the residuals one period and then square the difference between consecutive residuals. This may also be called finding the differences. This accounts for summing the observations from 2, rather than from 1, up to *n.* In the denominator we square the residuals and sum over all *n* observations.

The value of the Durbin-Watson statistic can range from 0 to 4. The value of *d* is 2.00 when there is no autocorrelation among the residuals. When the value of *d* gets close to 0, this indicates positive autocorrelation. Values beyond 2 indicate negative autocorrelation. Negative autocorrelation seldom exists in practice. To occur, successive residuals would tend to be large, but would have opposite signs.

To conduct a test for autocorrelation, the null and alternate hypotheses are:

H_0: No residual correlation ($\rho = 0$)
H_1: Positive residual correlation ($\rho > 0$)

Recall from the previous chapter that *r* refers to the sample correlation and that ρ is the correlation coefficient in the population. The critical values for *d* are reported in Appendix B.10. To determine the critical value, we need α (the significance level), *n* (the sample size), and *k* (the number of independent variables). The decision rule for the Durbin-Watson test is altered from what we are used to. As usual there is a range of values where the null hypothesis is rejected and a range where it is not rejected. However, there is also a range of values where the test is inconclusive. That is, in the inconclusive range the null hypothesis is neither rejected nor not rejected. To state this more formally:

- Values less than d_l cause the rejection of the null hypothesis.
- Values greater then d_u will result in the null hypothesis not being rejected.
- Values of *d* between d_l and d_u yield inconclusive results.

The subscript *l* refers to the lower limit of *d* and the subscript *u* the upper limit.

How do we interpret the various decisions for the test for residual correlation? If the null hypothesis is not rejected, we conclude that autocorrelation is not present. The residuals are not correlated, there is no autocorrelation present, and the regression assumption has been met. There will not be any problem with the estimated value of the standard error of estimate. If the null hypothesis is rejected, then we conclude that autocorrelation is present.

The usual remedy for autocorrelation is to include another predictor variable that captures time order. For example, we might use the square root of *Y* instead of *Y.* This transformation will result in a change in the distribution of the residuals. If the result falls in the inconclusive range, more sophisticated tests are needed, or conservatively, we treat the conclusion as rejecting the null hypothesis.

An example will show the details of the Durbin-Watson test and how the results are interpreted.

Example

Banner Rocker Company manufactures and markets rocking chairs. The company developed a special rocker for senior citizens, which it advertises extensively on TV. Banner's market for the special chair is the Carolinas, Florida, and Arizona where there are many senior citizens and retired people. The president of Banner Rocker is studying the association between his advertising expense (*X*) and the number of rockers sold over the last 20 months (*Y*). He collected the following data. He would like to create a model to forecast sales, based on the amount spent on advertising, but is concerned that, because he gathered these data over consecutive months, there might be problems with autocorrelation.

Month	Sales (000)	Advertising ($ millions)	Month	Sales (000)	Advertising ($ millions)
1	153	$5.5	11	169	$6.3
2	156	5.5	12	176	5.9
3	153	5.3	13	176	6.1
4	147	5.5	14	179	6.2
5	159	5.4	15	184	6.2
6	160	5.3	16	181	6.5
7	147	5.5	17	192	6.7
8	147	5.7	18	205	6.9
9	152	5.9	19	215	6.5
10	160	6.2	20	209	6.4

Determine the regression equation. Is advertising a good predictor of sales? If the owner were to increase the amount spent on advertising by $1,000,000, how many additional chairs can he expect to sell? Investigate the possibility of autocorrelation.

Solution

The first step is to determine the regression equation.

```
Regression Analysis: Chairs (000) versus Advertising ($mil)

The regression equation is
Chairs (000) = -43.8 + 36.0 Advertising ($mil)

Predictor             Coef   SE Coef      T      P
Constant            -43.80     34.44  -1.27  0.220
Advertising ($mil)  35.950     5.746   6.26  0.000

S = 12.3474 R-Sq = 68.5% R-Sq(adj) = 66.8%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  5967.7  5967.7  39.14  0.000
Residual Error  18  2744.3   152.5
```

The coefficient of determination is 68.5 percent. So we know there is a strong positive association between the variables. We conclude that, as we increase the amount spent on advertising, we can expect to sell more chairs. Of course this is what we had hoped.

How many more chairs can we expect to sell if we increase advertising by $1,000,000? We must be careful with the units of the data. Sales are in thousands of chairs and advertising expense is in millions of dollars. The regression equation is:

$$\hat{Y} = -43.80 + 35.950X$$

This equation indicates that an increase of 1 in X will result in an increase of 35.95 in Y. So an increase of $1,000,000 in advertising will increase sales by 35,950 chairs. To put it another way, it will cost $27.82 in additional advertising expense per chair sold, found by $1,000,000/35,950.

What about the potential problem with autocorrelation? Many software packages, such as MINITAB, will calculate the value of the Durbin-Watson test and output the results. To understand the nature of the test and to see the details of formula (16–4) we use an Excel spreadsheet.

	A	B	C	D	E	F	G	H
1		Chairs (000)	Advertising ($mil)	Predicted Chairs (000)	Residuals	Lagged		
2	Month	Y	X	$\hat{Y}$	$e_t = Y - \hat{Y}$	e_{t-1}	$(e_t - e_{t-1})^2$	e_t^2
3	1	153	5.5	153.9237	-0.9237			0.8531
4	2	156	5.5	153.9237	2.0763	-0.9237	9.0000	4.3112
5	3	153	5.3	146.7336	6.2664	2.0763	17.5564	39.2675
6	4	147	5.5	153.9237	-6.9237	6.2664	173.9771	47.9371
7	5	159	5.4	150.3286	8.6714	-6.9237	243.2046	75.1925
8	6	160	5.3	146.7336	13.2664	8.6714	21.1142	175.9968
9	7	147	5.5	153.9237	-6.9237	13.2664	407.6376	47.9371
10	8	147	5.7	161.1137	-14.1137	-6.9237	51.6966	199.1965
11	9	152	5.9	168.3037	-16.3037	-14.1137	4.7963	265.8118
12	10	160	6.2	179.0888	-19.0888	-16.3037	7.7565	364.3820
13	11	169	6.3	182.6838	-13.6838	-19.0888	29.2138	187.2467
14	12	176	5.9	168.3037	7.6963	-13.6838	457.1076	59.2325
15	13	176	6.1	175.4938	0.5062	7.6963	51.6966	0.2563
16	14	179	6.2	179.0888	-0.0888	0.5062	0.3540	0.0079
17	15	184	6.2	179.0888	4.9112	-0.0888	25.0000	24.1200
18	16	181	6.5	189.8738	-8.8738	4.9112	190.0278	78.7452
19	17	192	6.7	197.0639	-5.0639	-8.8738	14.5158	25.6430
20	18	205	6.9	204.2539	0.7461	-5.0639	33.7557	0.5566
21	19	215	6.5	189.8738	25.1262	0.7461	594.3881	631.3234
22	20	209	6.4	186.2788	22.7212	25.1262	5.7839	516.2515
23							2338.5829	2744.2685

To investigate the possible autocorrelation we need to determine the residuals for each observation. We find the fitted values, that is the $\hat{Y}$, for each of the 20 months. This information is shown in the fourth column, column D. Next we find the residual, which is the difference between the actual value and the fitted values. So for the first month:

$$\hat{Y} = -43.80 + 35.950X = -43.80 + 35.950(5.5) = 153.925$$

$$e_1 = Y_1 - \hat{Y}_1 = 153 - 153.925 = -0.925$$

The residual, reported in column E, is slightly different due to rounding in the software. Notice in particular the string of five negative residuals in rows 9 through 13. In column F we lag the residuals one period. In column G we find the difference between the current residual and the residual in the previous and square this difference. Using the values from the software:

$$(e_t - e_{t-1})^2 = (e_2 - e_{2-1})^2 = [2.0763 - (-0.9237)]^2 = (3.0000)^2 = 9.0000$$

The other values in column G are found the same way. The values in column H are the squares of those in column E.

$$(e_1)^2 = (-0.9237)^2 = 0.8531$$

To find the value of d we need the sums of columns G and H. These sums are noted in yellow in the spreadsheet.

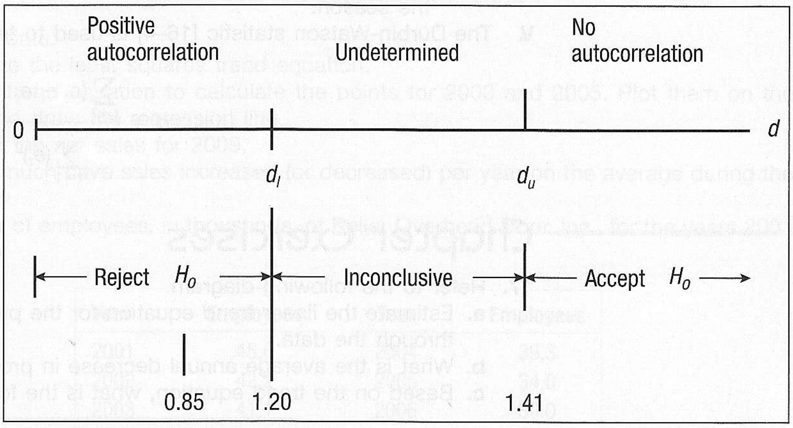
$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} (e_t)^2} = \frac{2,338.5829}{2,744.2685} = 0.8522$$

Now to answer the question as to whether there is significant autocorrelation. The null and the alternate hypotheses are stated as follows.

H_0: No residual correlation
H_1: Positive residual correlation

The critical value of d is found in Appendix B.10, a portion of which is shown below. There is one independent variable, so $k = 1$, the level of significance is 0.05, and the sample size is 20. We move to the .05 table, the columns where $k = 1$, and the row of 20. The reported values are $d_l = 1.20$ and $d_u = 1.41$. The null hypothesis is rejected if $d < 1.20$ and not rejected if $d > 1.41$. No conclusion is reached if d is between 1.20 and 1.41.

n	k = 1		k = 2	
	d_l	d_u	d_l	d_u
15	1.08	1.36	0.95	1.54
16	1.10	1.37	0.98	1.54
17	1.13	1.38	1.02	1.54
18	1.16	1.39	1.05	1.53
19	1.18	1.40	1.08	1.53
20	1.20	1.41	1.10	1.54
21	1.22	1.42	1.13	1.54
22	1.24	1.43	1.15	1.54
23	1.26	1.44	1.17	1.54
24	1.27	1.45	1.19	1.55
25	1.29	1.45	1.21	1.55



Because the computed value of d is 0.8522, which is less than the d_l, we reject the null hypothesis and accept the alternate hypothesis. We conclude that the residuals are autocorrelated. We have violated one of the regression assumptions. What do we do? The presence of autocorrelation usually means that the regression model has not been correctly specified. It is likely we need to add one or more independent variables that have some time-ordered effects on the dependent variable. The simplest independent variable to add is one that represents the time periods.

Exercises

15. Recall Exercise 9 from Chapter 14 and the regression equation to predict job performance. See page 543.
 a. Plot the residuals in the order in which the data are presented.
 b. Test for autocorrelation at the .05 significance level.

16. Consider the data in Exercise 10 from Chapter 14 and the regression equation to predict commissions earned. See page 544.
 a. Plot the residuals in the order in which the data are presented.
 b. Test for autocorrelation at the .01 significance level.

Chapter Summary

I. A time series is a collection of data over a period of time.
 A. The trend is the long-run direction of the time series.
 B. The cyclical component is the fluctuation above and below the long-term trend line over a longer period of time.
 C. The seasonal variation is the pattern in a time series within a year. These patterns tend to repeat themselves from year to year for most businesses.
 D. The irregular variation is divided into two components.
  1. The episodic variations are unpredictable, but they can usually be identified. A flood is an example.
  2. *The residual variations are random in nature.*

II. A moving average is used to smooth the trend in a time series.

III. *The linear trend equation is $\hat{Y} = a + bt$, where a is the Y-intercept, b is the slope of the* line, and t is the coded time.
 A. The trend equation is determined using the least squares principle.

B. If the trend is not linear, but rather the increases tend to be a constant percent, the Y values are converted to logarithms, and a least squares equation is determined using the logarithms.

IV. A seasonal factor can be estimated using the ratio-to-moving-average method.

A. The six-step procedure yields a seasonal index for each period.

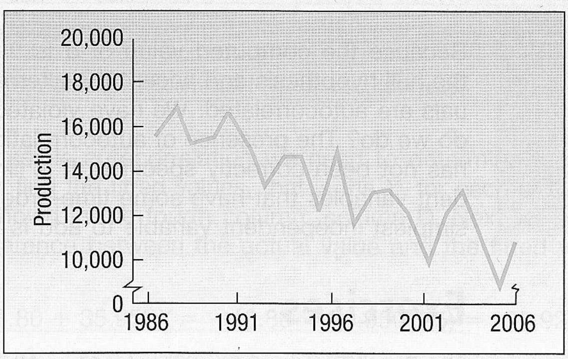
1. Seasonal factors are usually computed on a monthly or a quarterly basis.
2. The seasonal factor is used to adjust forecasts, taking into account the effects of the season.

V. The Durbin-Watson statistic [16–4] is used to test for autocorrelation.

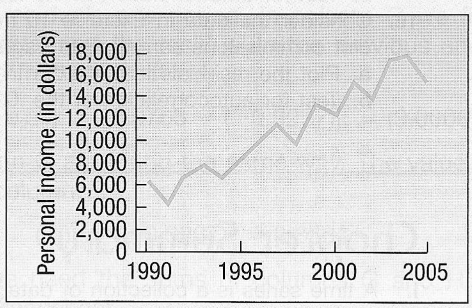
$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} (e_t)^2} \quad \textbf{[16–4]}$$

Chapter Exercises

17. Refer to the following diagram.
    a. Estimate the linear trend equation for the production series by drawing a straight line through the data.
    b. What is the average annual decrease in production?
    c. Based on the trend equation, what is the forecast for the year 2010?



18. Refer to the following diagram.
    a. Estimate the linear trend equation for the personal income series.
    b. What is the average annual increase in personal income?



19. The asset turnovers, excluding cash and short-term investments, for RNC Company from 1996 to 2006 are:

1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
1.11	1.28	1.17	1.10	1.06	1.14	1.24	1.33	1.38	1.50	1.65

    a. Plot the data.
    b. Determine the least squares trend equation.
    c. Calculate the points on the trend line for 1999 and 2004, and plot the line on the graph.
    d. Estimate the asset turnover for 2011.
    e. How much did the asset turnover increase per year, on the average, from 1996 to 2006?

20. The sales, in billions of dollars, of Keller Overhead Door, Inc., for 2001 to 2006 are:

Year	Sales	Year	Sales
2001	7.45	2004	7.94
2002	7.83	2005	7.76
2003	8.07	2006	7.90

a. Plot the data.
b. Determine the least squares trend equation.
c. Use the trend equation to calculate the points for 2003 and 2005. Plot them on the graph and draw the regression line.
d. Estimate the net sales for 2009.
e. By how much have sales increased (or decreased) per year on the average during the period?

21. The number of employees, in thousands, of Keller Overhead Door, Inc., for the years 2001 to 2006 are:

Year	Employees	Year	Employees
2001	45.6	2004	39.3
2002	42.2	2005	34.0
2003	41.1	2006	30.0

a. Plot the data.
b. Determine the least squares trend equation.
c. Use the trend equation to calculate the points for 2003 and 2005. Plot them on the graph and draw the regression line.
d. Estimate the number of employees in 2009.
e. By how much has the number of employees increased (or decreased) per year on the average during the period?

22. Listed below is the selling price for a share of PepsiCo, Inc., at the close of each year.

Year	Price	Year	Price	Year	Price
1990	12.9135	1995	27.7538	2000	49.5625
1991	16.8250	1996	29.0581	2001	48.6800
1992	20.6125	1997	36.0155	2002	42.2200
1993	20.3024	1998	40.6111	2003	46.6200
1994	18.3160	1999	35.0230	2004	52.2000

a. Plot the data.
b. Determine the least squares trend equation.
c. Calculate the points for the years 1995 and 2000.
d. Estimate the selling price in 2008. Does this seem like a reasonable estimate based on the historical data?
e. By how much has the stock price increased or decreased (per year) on average during the period?

23. If plotted, the following sales series would appear curvilinear. This indicates that sales are increasing at a somewhat constant annual rate (percent). To fit the sales, therefore, a logarithmic equation should be used.

Year	Sales ($ millions)	Year	Sales ($ millions)
1996	8.0	2002	39.4
1997	10.4	2003	50.5
1998	13.5	2004	65.0
1999	17.6	2005	84.1
2000	22.8	2006	109.0
2001	29.3		

a. Determine the logarithmic equation.
b. Determine the coordinates of the points on the logarithmic straight line for 1995 and 2004.
c. By what percent did sales increase per year, on the average, during the period from 1996 to 2006?
d. Based on the equation, what are the estimated sales for 2007?

24. Reported below are the amounts spent on advertising ($ millions) by a large firm from 1996 to 2006.

Year	Amount	Year	Amount
1996	88.1	2002	132.6
1997	94.7	2003	141.9
1998	102.1	2004	150.9
1999	109.8	2005	157.9
2000	118.1	2006	162.6
2001	125.6		

a. Determine the logarithmic trend equation.
b. Estimate the advertising expenses for 2009.
c. By what percent per year did advertising expense increase during the period?

25. Listed below is the selling price for a share of Oracle, Inc., stock at the close of the year.

Year	Price	Year	Price	Year	Price
1990	0.1944	1995	3.1389	2000	29.0625
1991	0.3580	1996	4.6388	2001	13.8100
1992	0.7006	1997	3.7188	2002	10.8000
1993	1.4197	1998	7.1875	2003	13.2300
1994	2.1790	1999	28.0156	2004	13.7200
				2005	12.2100

a. Plot the data.
b. Determine the least squares trend equation. Use both the actual stock price and the logarithm of the price. Which seems to yield a more accurate forecast?
c. Calculate the points for the years 1993 and 1998.
d. Estimate the selling price in 2007. Does this seem like a reasonable estimate based on the historical data?
e. By how much has the stock price increased or decreased (per year) on average during the period? Use your best answer from part (b).

26. The production of Reliable Manufacturing Company for 2002 and part of 2003 follows.

Month	2002 Production (thousands)	2003 Production (thousands)	Month	2002 Production (thousands)	2003 Production (thousands)
January	6	7	July	3	4
February	7	9	August	5	
March	12	14	September	14	
April	8	9	October	6	
May	4	5	November	7	
June	3	4	December	6	

a. Using the ratio-to-moving-average method, determine the specific seasonals for July, August, and September 2002.
b. Assume that the specific seasonal indexes in the following table are correct. Insert in the table the specific seasonals you computed in part (a) for July, August, and September 2002, and determine the 12 typical seasonal indexes.

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
2002							?	?	?	92.1	106.5	92.9
2003	88.9	102.9	178.9	118.2	60.1	43.1	44.0	74.0	200.9	90.0	101.9	90.9
2004	87.6	103.7	170.2	125.9	59.4	48.6	44.2	77.2	196.5	89.6	113.2	80.6
2005	79.8	105.6	165.8	124.7	62.1	41.7	48.2	72.1	203.6	80.2	103.0	94.2
2006	89.0	112.1	182.9	115.1	57.6	56.9						

c. Interpret the typical seasonal index.

27. The sales of Andre's Boutique for 2002 and part of 2003 are:

Month	2002 Sales (thousands)	2003 Sales (thousands)	Month	2002 Sales (thousands)	2003 Sales (thousands)
January	78	65	July	81	65
February	72	60	August	85	61
March	80	72	September	90	75
April	110	97	October	98	
May	92	86	November	115	
June	86	72	December	130	

a. Using the ratio-to-moving-average method, determine the specific seasonals for July, August, September, and October 2002.

b. Assume that the specific seasonals in the following table are correct. Insert in the table the specific seasonals you computed in part (a) for July, August, September, and October 2002, and determine the 12 typical seasonal indexes.

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
2002							?	?	?	?	123.6	150.9
2003	83.9	77.6	86.1	118.7	99.7	92.0	87.0	91.4	97.3	105.4	124.9	140.1
2004	86.7	72.9	86.2	121.3	96.6	92.0	85.5	93.6	98.2	103.2	126.1	141.7
2005	85.6	65.8	89.2	125.6	99.6	94.4	88.9	90.2	100.2	102.7	121.6	139.6
2006	77.3	81.2	85.8	115.7	100.3	89.7						

c. Interpret the typical seasonal index.

28. The quarterly production of pine lumber, in millions of board feet, by Northwest Lumber since 2002 is:

	Quarter			
Year	Winter	Spring	Summer	Fall
2002	7.8	10.2	14.7	9.3
2003	6.9	11.6	17.5	9.3
2004	8.9	9.7	15.3	10.1
2005	10.7	12.4	16.8	10.7
2006	9.2	13.6	17.1	10.3

a. Determine the typical seasonal pattern for the production data using the ratio-to-moving-average method.

b. Interpret the pattern.

c. Deseasonalize the data and determine the linear trend equation.

d. Project the seasonally adjusted production for the four quarters of 2007.

29. Work Gloves Corp. is reviewing its quarterly sales of Toughie, the most durable glove it produces. The numbers of pairs produced (in thousands) by quarter are:

	Quarter			
	I	II	III	IV
Year	Jan.–Mar.	Apr.–June	July–Sept.	Oct.–Dec.
1999	142	312	488	208
2000	146	318	512	212
2001	160	330	602	187
2002	158	338	572	176
2003	162	380	563	200
2004	162	362	587	205

a. Using the ratio-to-moving-average method, determine the four typical quarterly indexes.
b. Interpret the typical seasonal pattern.

30. Sales of roof material, by quarter, since 2000 by Carolina Home Construction, Inc., are shown below (in $000).

	Quarter			
Year	I	II	III	IV
2000	210	180	60	246
2001	214	216	82	230
2002	246	228	91	280
2003	258	250	113	298
2004	279	267	116	304
2005	302	290	114	310
2006	321	291	120	320

a. Determine the typical seasonal patterns for sales using the ratio-to-moving-average method.
b. Deseasonalize the data and determine the trend equation.
c. Project the sales for 2007, and then seasonally adjust each quarter.

31. Blueberry Farms Golf and Fish Club of Hilton Head, South Carolina, wants to find monthly seasonal indexes for package play, nonpackage play, and total play. The package play refers to golfers who visit the area as part of a golf package. Typically the greens fees, cart fees, lodging, maid service, and meals are included as part of a golfing package. The course earns a certain percentage of this total. The nonpackage play includes play by local residents and visitors to the area who wish to play golf. The following data, beginning with July 2002, report the package and nonpackage play by month, as well as the total amount, in thousands of dollars.

Year	Month	Package	Local	Total
2002	July	$ 18.36	$43.44	$ 61.80
	August	28.62	56.76	85.38
	September	101.34	34.44	135.78
	October	182.70	38.40	221.10
	November	54.72	44.88	99.60
	December	36.36	12.24	48.60
2003	January	25.20	9.36	34.56
	February	67.50	25.80	93.30
	March	179.37	34.44	213.81
	April	267.66	34.32	301.98
	May	179.73	40.80	220.53
	June	63.18	40.80	103.98
	July	16.20	77.88	94.08

(continued)

Year	Month	Package	Local	Total
	August	23.04	76.20	99.24
	September	102.33	42.96	145.29
	October	224.37	51.36	275.73
	November	65.16	25.56	90.72
	December	22.14	15.96	38.10
2004	January	30.60	9.48	40.08
	February	63.54	30.96	94.50
	March	167.67	47.64	215.31
	April	299.97	59.40	359.37
	May	173.61	40.56	214.17
	June	64.98	63.96	128.94
	July	25.56	67.20	92.76
	August	31.14	52.20	83.34
	September	81.09	37.44	118.53
	October	213.66	62.52	276.18
	November	96.30	35.04	131.34
	December	16.20	33.24	49.44
2005	January	26.46	15.96	42.42
	February	72.27	35.28	107.55
	March	131.67	46.44	178.11
	April	293.40	67.56	360.96
	May	158.94	59.40	218.34
	June	79.38	60.60	139.98

Using statistical software:

a. Develop a seasonal index for each month for the package sales. What do you note about the various months?

b. Develop a seasonal index for each month for the nonpackage sales. What do you note about the various months?

c. Develop a seasonal index for each month for the total sales. What do you note about the various months?

d. Compare the indexes for package sales, nonpackage sales, and total sales. Are the busiest months the same?

32. The following is the number of retirees receiving benefits from the State Teachers Retirement System of Ohio from 1991 until 2004.

Year	Service	Year	Service	Year	Service
1991	58,436	1996	70,448	2001	83,918
1992	59,994	1997	72,601	2002	86,666
1993	61,515	1998	75,482	2003	89,257
1994	63,182	1999	78,341	2004	92,574
1995	67,989	2000	81,111		

a. Plot the data.

b. Determine the least squares trend equation. Use a linear equation.

c. Calculate the points for the years 1993 and 1998.

d. Estimate the number of retirees that will be receiving benefits in 2006. Does this seem like a reasonable estimate based on the historical data?

e. By how much has the number of retirees increased or decreased (per year) on average during the period?

33. Ray Anderson, owner of Anderson Ski Lodge in upstate New York, is interested in forecasting the number of visitors for the upcoming year. The following data are available, by quarter, since 2000. Develop a seasonal index for each quarter. How many visitors would you expect for each quarter of 2007, if Ray projects that there will be a 10 percent increase from the total number of visitors in 2006? Determine the trend equation, project the number of visitors for 2007, and seasonally adjust the forecast. Which forecast would you choose?

Year	Quarter	Visitors	Year	Quarter	Visitors
2000	I	86	2004	I	188
	II	62		II	172
	III	28		III	128
	IV	94		IV	198
2001	I	106	2005	I	208
	II	82		II	202
	III	48		III	154
	IV	114		IV	220
2002	I	140	2006	I	246
	II	120		II	240
	III	82		III	190
	IV	154		IV	252
2003	I	162			
	II	140			
	III	100			
	IV	174			

34. The enrollment in the College of Business at Midwestern University by quarter since 2001 is:

	Quarter			
Year	Winter	Spring	Summer	Fall
2001	2,033	1,871	714	2,318
2002	2,174	2,069	840	2,413
2003	2,370	2,254	927	2,704
2004	2,625	2,478	1,136	3,001
2005	2,803	2,668	—	—

Using the ratio-to-moving-average method:

a. Determine the four quarterly indexes.

b. Interpret the quarterly pattern of enrollment. Does the seasonal variation surprise you?

c. Compute the trend equation, and forecast the 2006 enrollment by quarter.

35. The Jamie Farr Kroger Classic is an LPGA (women's professional golf) tournament played in Toledo, Ohio, each year. Listed below are the total purse and the prize for the winner for the 19 years from 1987 through 2005. Develop a trend equation for both variables. Which variable is increasing at a faster rate? Project both the amount of the purse and the prize for the winner in 2007. Find the ratio of the winner's prize to the total purse. What do you find? Which variable can we estimate more accurately, the size of the purse or the winner's prize?

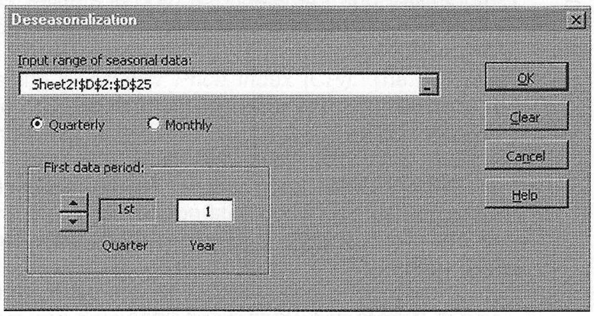
Year	Purse	Prize	Year	Purse	Prize
1987	$225,000	$33,750	1997	$ 700,000	$105,000
1988	275,000	41,250	1998	800,000	120,000
1989	275,000	41,250	1999	800,000	120,000
1990	325,000	48,750	2000	1,000,000	150,000
1991	350,000	52,500	2001	1,000,000	150,000
1992	400,000	60,000	2002	1,000,000	150,000
1993	450,000	67,500	2003	1,000,000	150,000
1994	500,000	75,000	2004	1,200,000	180,000
1995	500,000	75,000	2005	1,200,000	180,000
1996	575,000	86,250			

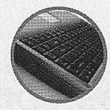
exercises.com

36. Go to the Bureau of Labor Statistics website, www.bls.gov, and click on the **Consumer Price Index** option, select **Consumer Price Index—All Urban Consumers (Current Series),** select **U.S. All items, 1982–84 = 100** and click **Retrieve data** at the bottom. Ask for the yearly output for the last 10 to 20 years. Develop a regression equation for the annual Consumer Price Index for the selected period. Use both the linear and the log approach. Which do you think is best?
37. Develop a trend line for a large or well-known company, such as GM, General Electric, or Microsoft, for the last 10 years. You could go to the company website. Most companies have a section called "Financial Information" or similar. Go to that location and look for sales over the last 10 years. If you do not know the website of the company, go to the financial section of Yahoo or *USA Today,* where there is a location for "symbol lookup." Type in the company name, which should then give you the symbol. Look up the company via the symbol, and you should find the information. The symbol for GM is just *GM,* the symbol for General Electric is *GE*. Comment on the trend line of the company you selected over the period. Is the trend increasing or decreasing? Does the trend follow a linear or log equation?
38. Select one of the major economic indicators, such as the Dow Jones Industrial Average, Nasdaq, or the S&P 500. Develop a trend line for the index over the last 10 years by using the value of the index at the end of the year, or for the last 30 days by selecting the closing value of the index for the last 30 days. You can locate this information in many places. For example, go to http://finance.yahoo.com, click on **Nasdaq** on the left hand, select **Historical Prices,** and a period of time, perhaps the last 30 days, and you will find the information. You should be able to download it directly to Excel to create your trend equation. Comment on the trend line you created. Is it increasing or decreasing? Does the trend line follow a linear or log equation?

Software Commands

1. The MegaStat commands for creating the seasonal indexes on page 623 are:
   - **a.** Enter the coded time period and the value of the time series in two columns. You may also want to include information on the years and quarters.
   - **b.** Select **MegaStat, Time Series/Forecasting,** and **Deseasonalization,** and hit **Enter.**
   - **c.** Input the range of the data, indicate the data are in the first quarter, and click **OK.**

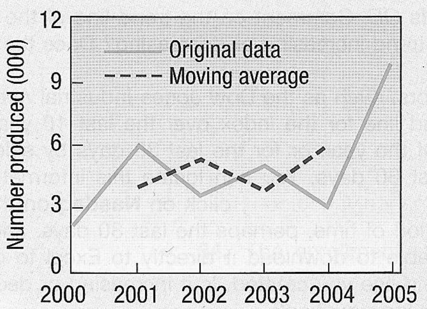




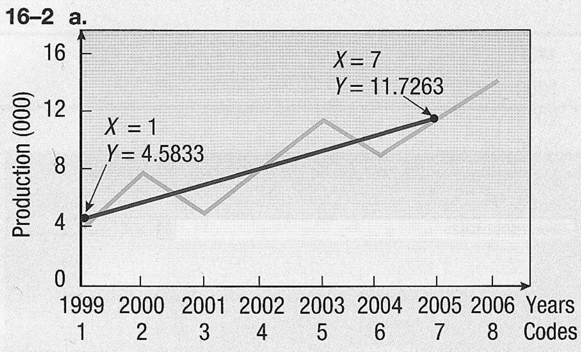
Chapter 16 Answers to Self-Review

16–1

Year	Production (thousands)	Three-Year Moving Total	Three-Year Moving Average
2000	2	—	—
2001	6	12	4
2002	4	15	5
2003	5	12	4
2004	3	18	6
2005	10	—	—



16–2 a.



b. $\hat{Y} = a + bt = 3.3928 + 1.1905t$ (in thousands)

c. For 1999:

$$\hat{Y} = 3.3928 + 1.1905(1) = 4.5833$$

for 2005:

$$\hat{Y} = 3.3928 + 1.1905(7) = 11.7263$$

d. For 2009, $t = 11$, so

$$\hat{Y} = 3.3928 + 1.1905(11) = 16.4883$$

or 16,488 king-size rockers.

16–3 a.

Year	Y	log Y	t
2002	2.13	0.3284	1
2003	18.10	1.2577	2
2004	39.80	1.5999	3
2005	81.40	1.9106	4
2006	112.00	2.0492	5

$$b = 0.40945$$

$$a = 0.20081$$

b. About 156.7 percent. The antilog of 0.40945 is 2.567. Subtracting 1 yields 1.567.

c. About 454.5, found by $\hat{Y} = 0.20081 + .40945(6) = 2.65751$. The antilog of 2.65751 is 454.5.

16–4 a. The following values are from a software package. Due to rounding, your figures might be slightly different.

	Winter	Spring	Summer	Fall
Mean	119.35	81.66	125.31	74.24
Typical seasonal	119.18	81.55	125.13	74.13

The correction factor is 0.9986.

b. Total sales at Teton Village for the winter season are typically 19.18 percent above the annual average.

16–5 The forecast value for January of the sixth year is 34.9, found by

$$\hat{Y} = 4.4 + 0.5(61) = 34.9$$

Seasonally adjusting the forecast, 34.9(120)/100 = 41.88. For February, $\hat{Y} = 4.4 + 0.5(62) = 35.4$. Then (35.4)95/100 = 33.63.

A Review of Chapters 15 and 16

Chapter 15 presents index numbers. An *index number* describes the relative change in value from one period, called the base period, to another called the given period. It is actually a percent, but the percent sign is usually omitted. Indexes are used to compare the change in unlike series over time. For example a company might wish to compare the change in sales with the change in the number of sales representatives employed over the same period of time. A direct comparison is not meaningful because the units for one set of data are dollars and the other people. Index numbers also facilitate the comparison of very large values, where the amount of change in the actual values is very large and therefore difficult to interpret.

There are two types of price indexes. In an *unweighted price index* the quantities are not considered. To form an unweighted index we divide the base period value into the current period (also called the given period) and report the percent change. So if sales were $12,000,000 in 2000 and $18,600,000 in 2006 the simple unweighted price index for 2006 is:

$$P = \frac{p_t}{p_0}(100) = \frac{\$18{,}600{,}000}{\$12{,}000{,}000}(100) = 155.0$$

We conclude there is a 55 percent increase in the sales during the six-year period.

In a *weighted price index quantities* are considered. The most widely used weighted index is the *Laspeyres' Price Index.* It uses the base period quantities as weights to compare changes in prices. It is computed by multiplying the base period quantities by the base period price for each product considered and summing the result. This result is the denominator of the fraction. The numerator of the fraction is the product of the base period quantities and the current price. For example, an appliance store sold 50 computes at $1,000 and 200 DVDs at $150 each in year 2000. In 2006 the same store sold 60 computers at $1,200 and 230 DVDs at $175. The Laspeyres' Price Index is:

$$P = \frac{\Sigma p_t q_0}{\Sigma p_0 q_0}(100) = \frac{\$1{,}200 \times 50 + \$175 \times 200}{\$1{,}000 \times 50 + \$150 \times 200}(100) = \frac{\$95{,}000}{\$80{,}000}(100) = 118.75$$

Notice the same base period quantities are used as weights in both the numerator and the denominator. The index indicates there has been an 18.75 percent increase in the value of sales during the six-year period.

The most widely used and reported index is the *Consumer Price Index (CPI).* The CPI is a Laspeyres type index. It is reported monthly by the U.S. Department of Labor and is often used to report the rate of inflation in the prices of goods and services in the United States. The current base period is 1982–84.

In Chapter 16 we studied time series and forecasting. A *time series* is a collection of data over a period of time. The earnings per share of General Electric common stock over the last ten years is an example of a time series. There are four components to a time series: the trend, cyclic effects, seasonal effects, and irregular effects.

Trend is the long run direction of the time series. It can be either increasing or decreasing.

The *cyclical component* is the fluctuation above and below the trend line over a period of several years. Economic cycles are examples of the cyclical component. Most businesses shift between relative expansion and reduction periods over a cycle of several years.

Seasonal variation is the recurring pattern of the time series within a year. The consumption of many products and services is seasonal. Beach homes along the Gulf Coast are seldom rented during the winter and ski lodges in Wyoming are not used in the summer months. Hence we say the rental of beach front properties and ski lodges are seasonal.

The *irregular component* includes any unpredictable events. In other words, the irregular component includes events that cannot be forecast. There are two types of irregular components. Episodic variations are unpredictable, but can usually be identified. A flood is an example. The residual variation is random in nature and not predicted or identified.

The linear trend for a time series is given by the equation $\hat{Y} = a + bt$, where $\hat{Y}$ is the estimated trend value, a is the intercept with the Y axis, b is the slope of the trend line (the rate of change), and t refers to the coded values for the time periods. We use the least squares method described in Chapter 13 to determine the trend line. Autocorrelation is often a problem when using the trend equation. Autocorrelation means that successive values of the time series are correlated.

Glossary

Chapter 15

Consumer Price Index An index reported monthly by the U.S. Department of Labor. It describes the change in a market basket of goods and services from the base period of 1982–84 to the present.

Simple index The value in the given period divided by the value in the base period. The result is usually multiplied by 100 and reported as a percent.

Weighted index The prices in the base period and the given period are multiplied by quantities (weights).

Chapter 16

Cyclical variation The rise and fall of a time series over periods longer than one year.

Episodic variation It is variation that is random in nature, but a cause can be identified.

Irregular variation Variation in a times series that is random in nature and does not regularly repeat itself.

Residual variation It is variation that is random in nature and cannot be identified or predicted.

Seasonal variation Patterns of change in a time series within a year. These patterns of change repeat themselves each year.

Secular trend The smoothed long-term direction of a time series.

Exercises

Part I—Multiple Choice

1. An index number is
a. Actually a percent but the percent sign is usually omitted.
b. Useful for comparing data with different units.
c. Useful for evaluating the change in very large numbers.
d. All of the above.

2. The sales at Labate Sporting Goods in 2000 were $400,000. In 2006 the sales increased to $450,000.
a. The index for 2006 is 112.5.
b. There has been a 12.5 percent increase in sales during the 6-year period.
c. The index is unweighted.
d. All of the above are correct.

3. Which of the following weighted indexes uses current period or given period quantities to form the denominator of the index?
a. Laspeyres' Price Index
b. Paasche's Price Index
c. Fisher's Ideal Index
d. None of the above.

4. A major disadvantage of the Laspeyres' Price index is
a. It does not reflect changes in buying habits over time.
b. It is too sensitive to small changes during early periods.
c. It requires the denominator to be recalculated each period.
d. None of the above.

5. Which of the following are correct statements about the Consumer Price Index?
a. It is reported monthly by the Bureau of Labor Statistics.
b. It is often used to report the rate of inflation in the United States.
c. The current base period of the index is 1982–84.
d. All of the above.

6. The smoothed long-term direction of a time series is called the
a. Cyclical variation.
b. Seasonal variation.
c. Trend.
d. Irregular variation.

7. The rise and fall of a time series over periods longer than one year is called the
a. Cyclical variation.
b. Seasonal variation.
c. Trend.
d. Irregular variation.

8. Which of the following are correct statements about the seasonal component of a time series?
a. It refers to changing patterns within a year.
b. One of its components is the episodic variation.

c. It is always greater than 100 percent.
d. All of the above are correct.

9. The linear trend for the number of vehicles sold per year at Trythall Motor Sports, Inc. is given by the equation $\hat{Y} = 30 + 125t$. The base period, that is year 1, is 2000. Which of the following statements is correct?
   a. The estimated sales for 2008 are 1030.
   b. Sales are increasing at a rate of 125 per year.
   c. The estimated sales for 1999 would be 30.
   d. All of the above are correct.
10. If the rate of change from one period to the next is a constant *percent*
   a. A linear tend equation is used.
   b. A log transformation is used.
   c. The slope of the trend line will be negative.
   d. The episodic variation will always be less than 1.00.

Part II—Problems

11. Listed below is the consolidated revenue (in $ billions) for the General Electric for the period from 2001 to 2005.

Year	Consolidated Revenues ($ billions)
2001	108
2002	114
2003	113
2004	134
2005	150

   a. Determine the index for 2005, using 2001 as the base period.
   b. Use the period 2001 to 2003 as the base period and find the index for 2005.
   c. With 2001 as the base year, use the least squares method to find the trend equation. What is the estimated consolidated revenue for 2008? What is the rate of increase per year?
12. The table below shows the unemployment rate and the available workforce for three counties in northwest Pennsylvania for the years 2002 and 2005.

	2002		2005	
County	Labor Force	Unemployed %	Labor Force	Unemployed %
Erie	141,500	6.7	141,800	5.6
Warren	22,700	5.8	21,300	5.3
McKean	22,200	6.0	21,900	5.7

   a. Find the overall unemployment rate for this region of northwest Pennsylvania for 2002?
   b. Use the data for this region of northwest Pennsylvania to create an unweighted index of percent unemployed for the year 2002.
   c. Use the data for this region of northwest Pennsylvania to create a weighted unemployment index using the Laspeyres method. Use the year 2002 as the base period.
13. Based on five years of monthly data (the period from January 2001 to December 2005) the trend equation for a small company is $\hat{Y} = 3.5 + 0.7t$. The seasonal index for January is 120 and for June it is 90. What is the seasonally adjusted sales forecast for January 2006 and June 2006?