CHAPTER-5 FORECASTING

5.1 INTRODUCTION

Every day, managers make decisions without knowing what will happen in the future. Inventory is ordered though no one knows what sales will be, new equipment is purchased though no one knows the demands for products and investments are made though no one knows what profits will be. Managers are always trying to reduce this uncertainty and to make better estimates of what will happen in the future. Accomplishing this is the main purpose of forecasting.

There are many ways to forecast the future. In numerous firms, the entire process is subjective, involving seat-of-the-pants methods, intuition and years of experience. There are also many quantitative forecasting models, such as moving averages, exponential smoothing, trend projections and least squares regression analysis.

The following steps can help in the development of a forecasting system. While steps 5 and 6 may not be relevant if a qualitative model is selected in step 4, data are certainly necessary for the quantitative forecasting models.

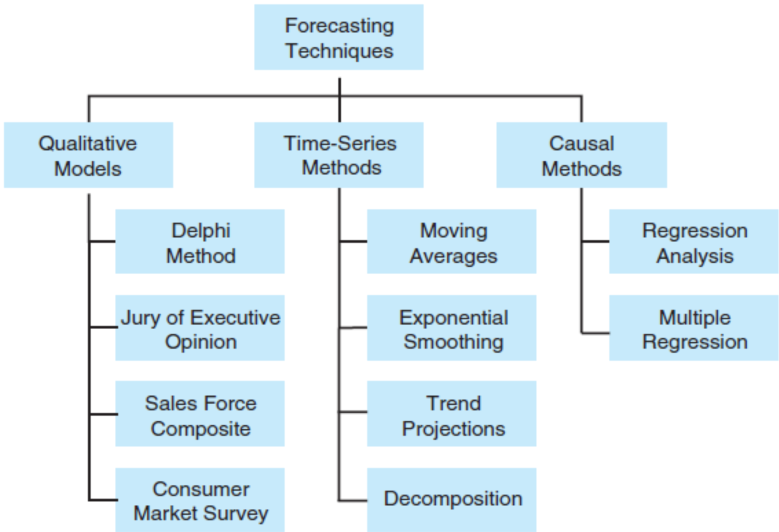
Eight Steps to Forecasting

1. Determine the use of the forecast—what objective are we trying to obtain?
2. Select the items or quantities that are to be forecasted.
3. Determine the time horizon of the forecast—is it 1 to 30 days (short term), 1 month to 1 year (medium term), or more than 1 year (long term)?
4. Select the forecasting model or models.
5. Gather the data needed to make the forecast.
6. Validate the forecasting model.
7. Make the forecast.
8. Implement the results.

5.2TYPES OF FORECASTS

In this chapter, we consider forecasting models that can be classified into one of three categories: time-series models, casual models and qualitative models (see Fig 5.1)

FIGURE 5.1
Forecasting Models



Qualitative Models

Whereas time-series and casual models rely on quantitative data, qualitative models attempt to incorporate judgmental or subjective factors into the forecasting model. Opinions by experts, individual experiences and judgments and other subjective factors may be considered. Here is a brief overview of 4 different qualitative forecasting techniques:

1. **Delphi Method:** This iterative group process allows experts, who may be located in different places, to make forecasts. There are three different types of participants in the Delphi process: decision makers, staff personnel and respondents. The decision making group usually consists of 5 to 10 experts who will be making the actual forecast. The staff personnel assist the decision makers by preparing, distributing, collecting and summarizing a series of questionnaires and survey results. The respondents are a group of people whose judgments are valued and are being sought. This group provides inputs to the decision makers before the forecast is made.
2. **Jury of executive opinion:** This method takes the opinions of a small group of high-level managers, often in combination with statistical models and results in a group estimate of demand.
3. **Sales force composite:** In this approach, each salesperson estimates what sales will be in his or her region; these forecasts are reviewed to ensure that they are realistic and are then combines at the district and national levels to reach an overall forecast.
4. **Consumer market survey:** This method solicits input from customers regarding their future purchasing plans. It can help not only in preparing a forecast but also in improving product design and planning for new products.

Time-Series Models

Time-Series Models attempt to predict the future by using historical data. These models make the assumption that what happens in the future is a function of what has happened in the past. In other words, time-series models look at what has happened over a period of time and use a series of past data to make a forecast

Causal Models

Causal Models incorporate the variables or factors that might influence the quantity being forecasted into the forecasting model. For example, daily sales of a cola drink might depend on the season, the average temperature, the average humidity, whether it is a weekend or a weekday and so on. Thus, a causal model would attempt to include factors for temperature, humidity, season, day of the week and so on. Causal models may also include past sales data as time series models do, but they include other factors as well. The most common quantitative causal model is regression analysis.

5.3 SCATTER DIAGRAMS AND TIME SERIES

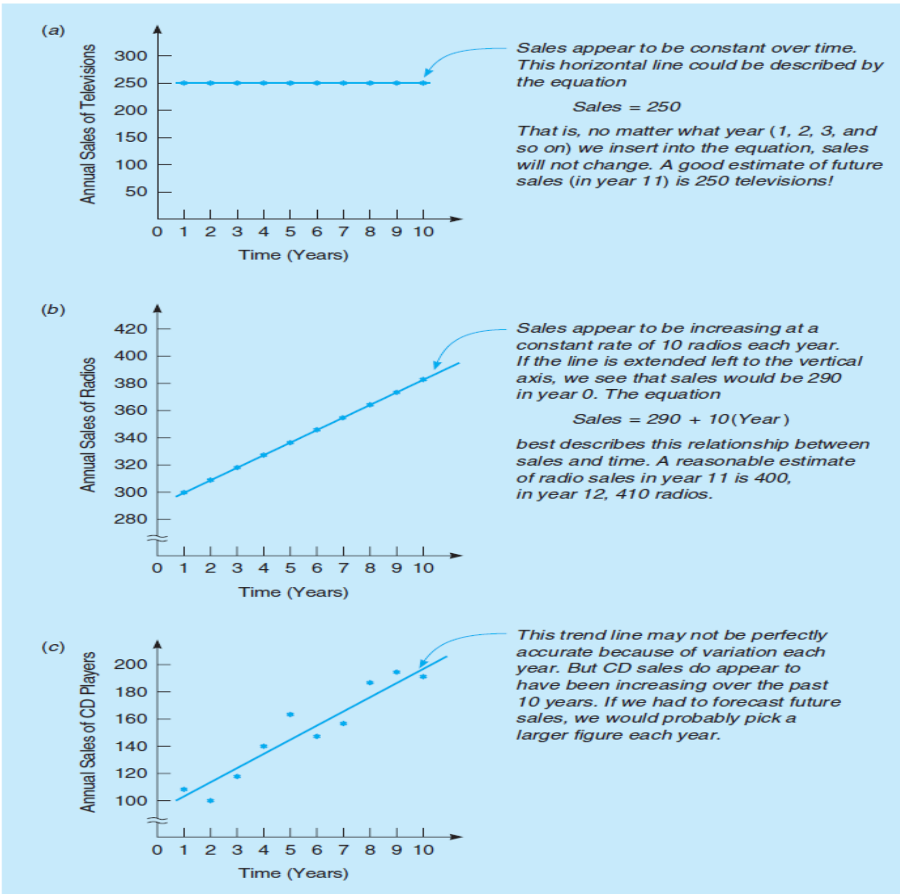
As with regression models, **scatter diagrams** are very useful when forecasting time series. A scatter diagram for a time series may be plotted on a two-dimensional graph with the horizontal axis representing the time period. The variable to be forecast (such as sales) is placed on vertical axis. Let us consider the example of a firm that needs to forecast sales for three different products.

Wacker Distributers notes that annual sales for three of its products—television sets, radios and compact disk players—over the past 10 years are as shown in Table 5.1. One simple way to examine these historical data and perhaps to use them to establish a forecast, is to draw a scatter diagram for each product (Figure 5.2)

TABLE 5.1
Annual Sales of Three Products

YEAR	TELEVISION SETS	RADIOS	COMPACT DISC PLAYERS
1	250	300	110
2	250	310	100
3	250	320	120
4	250	330	140
5	250	340	170
6	250	350	150
7	250	360	160
8	250	370	190
9	250	380	200
10	250	390	190

FIGURE 5.2
Scatter Diagram for Sales



5.4 MEASURES OF FORECAST ACCURACY

We discuss several different forecasting models in this chapter. To see how well one model works or to compare that model with other models, the forecasted values are compared with the actual or observed values. The forecast error (or deviation) is defined as

$$\text{Forecast error} = \text{Actual value} - \text{Forecast value}$$

One measure of accuracy is the **mean absolute deviation (MAD).** This is computed as

$$\text{MAD} = \frac{\sum|\text{forecast error}|}{\text{no. of errors}} \qquad \ldots\ldots(5.1)$$

Consider the Wacker Distributers sales of CD players shown in Table 5.1. Suppose that in the past, Wacker had forecast sales for each year to be the sales that were actually achieved in the previous year. This is sometimes called a naïve model.

Table 5.2 Computing the Mean Absolute Deviation (MAD)

Year	Actual Sales of CD Players	Forecast Sales	Absolute value of Errors i.e., $\lvert\text{Actual} - \text{Forecast}\rvert$
1	110	--	--
2	100	110	$\lvert 100 - 110\rvert = 10$
3	120	100	$\lvert 120 - 100\rvert = 20$
4	140	120	$\lvert 140 - 120\rvert = 20$
5	170	140	$\lvert 170 - 140\rvert = 30$
6	150	170	$\lvert 150 - 170\rvert = 20$
7	160	150	$\lvert 160 - 150\rvert = 10$
8	190	160	$\lvert 190 - 160\rvert = 30$
9	200	190	$\lvert 200 - 190\rvert = 10$
10	190	200	$\lvert 190 - 200\rvert = 10$
11	--	190	--
			Sum of $\lvert\text{errors}\rvert = 160$

From this, we see that

$$\text{MAD} = \frac{\sum|\text{forecast error}|}{\text{no. of errors}}$$

$$= \frac{160}{9} = 17.8$$

This means that on the average, each forecast missed the actual value by 17.8 units.

Other measures of the accuracy of historical errors in forecasting are sometimes used besides the MAD. One of the most common is the **mean squared error (MSE),** which is the average of the squared errors

$$\text{MSE} = \frac{\sum(\text{error})^2}{\text{n}} \qquad \ldots\ldots(5.2)$$

where n is the no. of errors

5.5 TIME-SERIES FORECASTING MODELS

A time series is based on a sequence of evenly spaced (weekly, monthly, quarterly and so on) data points. Examples include weekly sales of HP personal computers, quarterly earnings reports of Microsoft Corporation, daily shipments of Eveready batteries and annual U.S. consumer price indices.

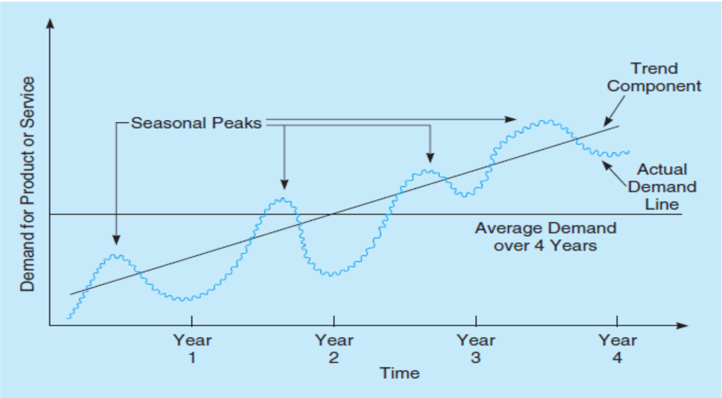
Components of a Time Series

Analyzing time series means breaking down past data into components and then projecting them forward. A time series typically has four components:

1. *Trend* (T) is the gradual upward or downward movement of the data over time.
2. *Seasonality* (S) is a pattern of the demand fluctuation above or below the trend line that repeats at regular intervals.
3. *Cycles* (C) are patterns in annual data that occur every several years. They are usually tied into the business cycle.
4. *Random Variations* (R) are "blips" in the data caused by chance and unusual situations; they follow no discernible pattern.

Figure 5.3 shows a time series and its components.

FIGURE 5.3
Product Demand Charted over 4 Years, with Trend and Seasonality Indicated



There are two general forms of time-series models in statistics. The first is a multiplicative model, which assumes that demand is the product of four components. It is stated as follows:

$$\text{Demand} = T \times S \times C \times R$$

An additive model adds the components together to provide an estimate. Multiple regression is often used to develop additive models. This additive relationship is stated as follows:

$$\text{Demand} = T + S + C + R$$

There are other models that may be a combination of these. For example, one of the components (such as trend) might be additive while another (such as seasonality) could be multiplicative.

Moving Averages

Moving averages are useful if we can assume that market demands will stay fairly steady over time. This tends to smooth out short-term irregularities in the data series.

An n-period moving average forecast, which serves as an estimate of the next period's demand, is expressed as follows:

$$\text{Moving average forecast} = \frac{\text{Sum of demands in previous } n \text{ periods}}{n} \qquad \ldots\ldots(5.4)$$

Mathematically, this is written as

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-n+1}}{n} \qquad \ldots\ldots..(5.5)$$

Where

F_{t+1} = forecast for time period $t + 1$

Y_t = actual value in time period t

n = no. of periods to average

A 4-month moving average has $n = 4$; a 5-month moving average has $n = 5$.

WALLACE GARDEN SUPPLY EXAMPLE: Storage shed sales at Wallace Garden Supply are shown in the middle column of Table 5.3. A 3- month moving average is indicated on the right. The forecast for the next January, using this technique, is 16. Were we simply asked to find a forecast for next January, we would only have to make this one calculation. The other forecasts are necessary only if we wish to compute the MAD or another measure of accuracy.

Table 5.3 Wallace Garden Supply Shed Sales

MONTH	Actual Shed Sales	3- month Moving Average
January	10	
February	12	
March	13	
April	16	$\frac{10+12+13}{3} = 11.67$
May	19	$\frac{12+13+16}{3} = 13.67$
June	23	$\frac{13+16+19}{3} = 16.00$
July	26	…………………
August	30	…………………
September	28	
October	18	
November	16	
December	14	………………….
January	--	$\frac{18+16+14}{3} = 16.00$

Example Following table represents the sales data from January to June for certain company:

Month	Automobile Battery Sales
January	28
February	21
March	38
April	34
May	36
June	38

a. Use 3 period moving averages to forecast the batteries sales for April, May & June.
b. Find MAD (Mean Absolute Deviation).

Solution:

(a)

Month	Automobile Battery Sales	Forecast	|Forecast Error|
January	28		
February	21		
March	38		
April	34	(28+21+38)/3= 29.0	$\|34-29\| = 5$
May	36	(21+38+34)/3= 31.0	$\|36-31\| = 5$
June	38	(34+36+38)/3= 36.0	$\|38-36\| = 2$
			$\sum\|\text{Forecast Error}\| = 12$

(b)

$$\text{MAD} = \frac{\sum|\text{forecast error}|}{\text{no. of errors}}$$

$$= \frac{12}{3} = 4$$

WEIGHTED MOVING AVERAGE

A simple moving average gives the same weight ($1/n$) to each of the past observations being used to develop the forecast. On the other hand, a weighted moving average allows different weights to be assigned to the previous observations.

A weighted moving average may be expressed as

$$F_{t+1} = \frac{\sum(\text{Weight in period } i)(\text{Actual value in period } i)}{\sum(\text{Weights})} \qquad \dots.(5.6)$$

Mathematically, this is

$$F_{t+1} = \frac{w_1 Y_t + w_2 Y_{t-1} + \cdots + w_n Y_{t-n+1}}{w_1 + w_2 + \cdots + w_n} \qquad \dots\dots(5.7)$$

Where w_i = weight for ith observation

Wallace Garden Supply decides to use a 3-month weighted moving average forecast with weights of 3 for the most recent observation, 2 for the next observation and 1 for the most distant observation.

The results of the Wallace Garden Supply weighted average forecast are shown in Table 5.4

Table 5.4 Weighted Moving Average Forecast for Wallace Garden Supply

MONTH	Actual Shed Sales	3- month Weighted Moving Average
January	10	
February	12	
March	13	
April	16	$\frac{3\times13+2\times12+1\times10}{3+2+1} = 12.17$
May	19	$\frac{3\times16+2\times13+1\times12}{3+2+1} = 14.33$
June	23	$\frac{3\times19+2\times16+1\times13}{3+2+1} = 17.00$
July	26	…………………
August	30	…………………
September	28	
October	18	
November	16	
December	14	…………………
January	--	$\frac{3\times14+2\times16+1\times18}{3+2+1} = 15.33$

Example Following table represents the sales data from January to June for certain company:

Month	January	February	March	April	May	June
Automobile Battery Sales	28	21	38	34	36	38

Develop a 3-month weighted moving average forecast for April, May & June by weighting three months as follows:

Period	Last Month	Two Months Ago	Three Months Ago	Total
Weight Applied	4	3	1	8

Solution: The 3- month weighted moving average forecasts for April, May & June are given in the following Table

Month	Automobile Battery Sales	Forecast
January	28	
February	21	
March	38	
April	34	$(1\times28+3\times21+4\times38)/8= 30.375$
May	36	$(1\times21+2\times38+4\times34)/8= 33.875$
June	38	$(1\times34+3\times36+4\times38)/8= 36.75$

Example: Bike sales at Sport Xpert are shown below:

Week	1	2	3	4	5	6	7
Bike Sales	4	5	4	6	5	7	--

Calculate a 3-week weighted moving average forecast for 7th week by weighting three weeks as follows:

Period	Last Week	Two Weeks Ago	Three Weeks Ago	Total
Weight Applied	4	3	1	8

Solution: The 3-week weighted moving average forecast for week-7 is given in the table:

Week	Bike Sales	Forecast
1	4	
2	5	
3	4	
4	6	
5	5	
6	7	
7	--	$(4\times7+3\times5+1\times6)/8= 6.125$

Exponential Smoothing

Exponential Smoothing is a forecasting method that is easy to use and is handled efficiently by computers. Although it is a type of moving average technique, it involves little record keeping of past data. The basic exponential smoothing formula can be shown as follows:

New forecast=Last period's forecast+α(Last period's actual demand−Last's period's forecast) 5.8

Or $F_{t+1} = F_t + \alpha(Y_t - F_t)$ (5.9)

where α is a weight (or smoothing constant) that has a value between 0 and 1, inclusive i.e. $0 \leq \alpha \leq 1$

For example, in January, a demand for 142 of a certain car model for February was predicted by a dealer. Actual February demand was 153 autos. Using a smoothing constant of $\alpha = 0.20$, we can forecast the March demand using the exponential smoothing model.

New forecast (for March demand) $= 142 + 0.20(153 - 142)$

$$= 144.2$$

Thus, the demand forecast for the cars in March is 144.

Suppose that actual demand for the cars in March was 136. A forecast for the demand in April, using the exponential smoothing model with a constant of $\alpha = 0.20$, can be made:

New forecast (for April demand) $= 144.2 + 0.20(136 - 144.2)$

$= 142.6$ or 143 autos

Selecting The Smoothing Constant: The exponential smoothing approach is easy to use and has been applied successfully by banks, manufacturing companies, wholesalers and other organizations. The appropriate value of the smoothing constant, α, however, can make the difference between an accurate forecast and an inaccurate forecast. In picking a value for the smoothing constant, the objective is to obtain the most accurate forecast. Several values of the smoothing constant may be tried and the one with the lowest MAD could be selected.

Example: Given an actual demand of 125 for current period when forecast of 129 was anticipated.

a. What is forecast error for current period?
b. For given smoothing constant (α) of 0.5 what would be the forecast for the next period by using simple exponential smoothing?

Solution:

(a) Forecast error = Actual Value – Forecast Value

$$= 125 - 129 = -4$$

(b) Here α=0.5

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$

$$F = (129) + 0.5(125 - 129)$$

$$= 129 + 0.5(-4) = 127$$

Example: Sport Xpert want to use the simple exponential smoothing on the bike sales given as below:

Week	1	2	3	4	5	6	7
Bike Sales	4	5	4	6	5	7	--

Assume that F_1 is perfect.

a) Develop a simple exponential smoothing with α=0.3 and compute the MAD.

b) The MAD of a simple exponential smoothing with α=0.4 is 0.87. What value of α (0.3 or 0.4) should Sport Xpert choose?

Solution: Since F_1 is perfect so $F_1 = Y_1 = 4$

Week	Bike Sales	F using α=0.3 $F_{t+1} = F_t + \alpha(Y_t - F_t)$	Absolute deviations $\lvert Y_t - F_t \rvert$
1	4	4	0
2	5	$4 + 0.3(4-4) = 4$	1.0
3	4	$4 + 0.3(5-4) =4.3$	0.3
4	6	$4.3 + 0.3(4-4.3) = 4.21$	1.79
5	5	$4.21 + 0.3(6-4.21) = 4.75$	0.25
6	7	$4.75 + 0.3(5-4.75) = 4.82$	2.18
7	-	$4.82 + 0.3(7-4.82) = 5.48$	
			$\sum\lvert Y_t - F_t \rvert$ =5.52

By the formula, MAD (corresponding to $\alpha = 0.3$) $= \frac{\sum|Y_t - F_t|}{n} = \frac{5.52}{6} = 0.92$

MAD corresponding to $\alpha = 0.4$ is given 0.87

Since MAD (corresponding to $\alpha = 0.4$)< MAD (corresponding to $\alpha = 0.3$), so Sport Xpert will prefer $\alpha = 0.4$.

Exponential Smoothing with Trend Adjustment The averaging or smoothing forecasting techniques are useful when a time series has only a random component, but these techniques fail to respond to trends. If there is trend present in the data, a forecasting model that explicitly incorporates this into the forecast should be used. One such technique is the exponential smoothing with trend model. The idea is to develop an exponential smoothing forecast and then adjust this for trend. Two smoothing constants, α and β, are used in this model and both of these values must be between 0 and 1.

The exponential smoothing forecast including trend (FIT_t) is developed using 3 steps:

Step 1. Smoothed forecast = Previous forecast including trend + α (last error)

$$F_{t+1} = FIT_t + \alpha(Y_t - FIT_t) \tag{5.10}$$

Step 2. Smoothed trend = Previous trend + β (error or excess in trend)

$$T_{t+1} = T_t + \beta(F_{t+1} - FIT_t) \tag{5.11}$$

Step 3. Forecast including trend = Smoothed forecast + Smoothed trend

$$FIT_{t+1} = F_{t+1} + T_{t+1} \tag{5.12}$$

Where

T_t = smoothed trend for time period t

F_t = smoothed forecast for time period t

FIT_t = forecast including trend for time period t

α = smoothing constant for forecasts

β = smoothing constant for trend

Consider the case of Midwestern Manufacturing Company, which has a demand for electrical generators over the period 2004 to 2010 as shown in Table 5.7.

Table 5.7 Midwestern Manufacturing Demand

Years	Electrical Generators Sold
2004	74
2005	79
2006	80
2007	90
2008	105
2009	142
2010	122

To use the trend-adjusted exponential smoothing method, first set initial conditions (Previous values for F and T) and choose α and β. Assuming that F_1 is perfect and T_1 is 0 and picking 0.3 and 0.4 for the smoothing constants, we have

$$F_1 = Y_1 = 74\ ,\ T_1 = 0,\ \alpha = 0.3,\ \ \beta = 0.4$$

This results in $FIT_1 = F_1 + T_1 = 74 + 0 = 74$

Following the three steps to get the forecast for 2005 (time period 2), we have

Step 1. $F_2 = FIT_1 + 0.3(Y_1 - FIT_1)$, using equation (5.10)

$$= 74 + 0.3(74 - 74) = 74$$

Step 2. $T_2 = T_1 + 0.4(F_2 - FIT_1)$, using equation (5.11)

$$= 0 + 0.4(74 - 74) = 0$$

Step 3.Trend-adjusted exponential smoothing forecast is given as

$$FIT_2 = F_2 + T_2 = 74 + 0 = 74$$

For 2006 (time period 3), we have

Step 1. $F_3 = FIT_2 + 0.3(Y_2 - FIT_2)$, using equation (5.10)

$$= 74 + 0.3(79 - 74) = 75.5$$

Step 2. $T_3 = T_2 + 0.4(F_3 - FIT_2)$, using equation (5.11)

$$= 0 + 0.4(75.5 - 74) = 0.6$$

Step 3.Trend-adjusted exponential smoothing forecast is given as

$$FIT_3 = F_3 + T_3 = 75.5 + 0.6 = 76.1$$

In the same way, we can find the results for other years.

TABLE 5.8 Midwestern Manufacturing Exponential Smoothing with Trend Forecasts

TIME (t)	DEMAND (Y_t)	$F_{t+1} = FIT_t + 0.3(Y_t - FIT_t)$	$T_{t+1} = T_t + 0.4(F_{t+1} - FIT_t)$	$FIT_{t+1} = F_{t+1} + T_{t+1}$
1	74	74	0	74
2	79	$74 = 74 + 0.3(74 - 74)$	$0 = 0 + 0.4(74 - 74)$	$74 = 74 + 0$
3	80	$75.5 = 74 + 0.3(79 - 74)$	$0.6 = 0 + 0.4(75.5 - 74)$	$76.1 = 75.5 + 0.6$
4	90	$77.270 = 76.1 + 0.3(80 - 76.1)$	$1.068 = 0.6 + 0.4(77.27 - 76.1)$	$78.338 = 77.270 + 1.068$
5	105	$81.837 = 78.338 + 0.3(90 - 78.338)$	$2.468 = 1.068 + 0.4(81.837 - 78.338)$	$84.305 = 81.837 + 2.468$
6	142	$90.514 = 84.305 + 0.3(105 - 84.305)$	$4.952 = 2.468 + 0.4(90.514 - 84.305)$	$95.466 = 90.514 + 4.952$
7	122	$109.426 = 95.466 + 0.3(142 - 95.466)$	$10.536 = 4.952 + 0.4(109.426 - 95.466)$	$119.962 = 109.426 + 10.536$
8		$120.573 = 119.962 + 0.3(122 - 119.962)$	$10.780 = 10.536 + 0.4(120.573 - 119.962)$	$131.353 = 120.573 + 10.780$

Trend Projections

Another method for forecasting time series with trend is called **trend projection**. This technique fits a trend line to a series of historical data points and then projects the line into the future for medium- to long-range forecasts. There are several mathematical trend equations that can be developed (e.g., exponential and quadratic), but in this section we look at linear (straight line) trends only. A trend line is simply a linear regression equation in which the independent variable (X) is the time period. The form of this is

$$\hat{Y} = b_0 + b_1X$$

where

$$\begin{aligned} \hat{Y} &= \text{predicted value} \\ b_0 &= \text{intercept} \\ b_1 &= \text{slope of the line} \\ X &= \text{time period (i.e., } X = 1, 2, 3, \ldots, \text{n)} \end{aligned}$$

The **least squares** regression method may be applied to find the coefficients that minimize the sum of the squared errors, thereby also minimizing the mean squared error (MSE).

MIDWESTERN MANUFACTURING COMPANY EXAMPLE Let us consider the case of Midwestern Manufacturing Company. That firm's demand for electrical generators over the period 2004–2010 was shown in Table 5.7. A trend line to predict demand (Y) based on the time period can be developed using a regression model. If we let 2004 be time period 1 ($X = 1$), then 2005 is time period 2 ($X = 2$), and so forth.

From this we get

$$\hat{Y} = 56.71 + 10.54X$$

To project demand in 2011, we first denote the year 2011 in our new coding system as $X = 8$:

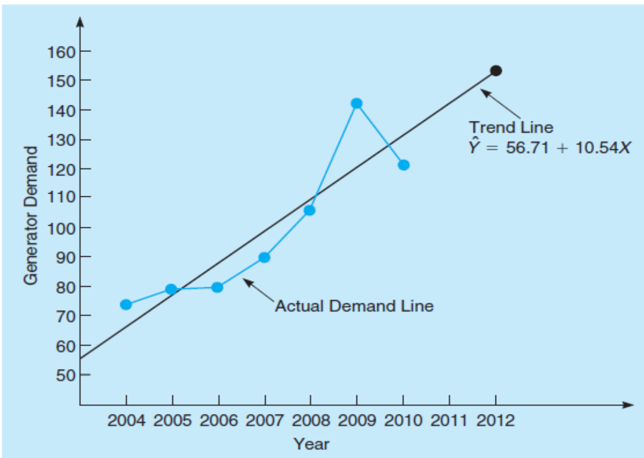
$$\begin{aligned} (\text{sales in } 2011) &= 56.71 + 10.54(8) \\ &= 141.03, \text{ or } 141 \text{ generators} \end{aligned}$$

We can estimate demand for 2012 by inserting $X = 9$ in the same equation:

$$\begin{aligned} (\text{sales in } 2012) &= 56.71 + 10.54(9) \\ &= 151.57, \text{ or } 152 \text{ generators} \end{aligned}$$

A plot of historical demand and the trend line is provided in Figure 5.4. In this case, we may wish to be cautious and try to understand the 2009–2010 swings in demand.

FIGURE 5.4
Electrical Generators and the Computed Trend Line



Seasonal Variations

Time-series forecasting such as that in example of Midwestern Manufacturing involves looking at the *trend* of data over a series of time observations. Sometimes, however, recurring variations at certain seasons of the year make a *seasonal* adjustment in the trend line forecast necessary. Demand for coal and fuel oil, for example, usually peaks during cold winter months. Demand for golf clubs or suntan lotion may be highest in summer. Analyzing data in monthly or quarterly terms usually makes it easy to spot seasonal patterns. A seasonal index is often used in multiplicative time series forecasting models to make an adjustment in the forecast when a seasonal component exists. An alternative is to use an additive model such as a regression model that will be introduced in a later section.

A **seasonal index** indicates how a particular season (e.g., month or quarter) compares with an average season. When no trend is present, the index can be found by dividing the average value for a particular season by the average of all the data. Thus, an index of 1 means the season is average. The example illustrates how to compute seasonal indices from historical data and to use these in forecasting future values.

TABLE 5.9
Answering Machine Sales and Seasonal Indices

MONTH	SALES DEMAND YEAR 1	SALES DEMAND YEAR 2	AVERAGE 2-YEAR DEMAND	MONTHLY DEMAND[a]	AVERAGE SEASONAL INDEX[b]
January	80	100	90	94	0.957
February	85	75	80	94	0.851
March	80	90	85	94	0.904
April	110	90	100	94	1.064
May	115	131	123	94	1.309
June	120	110	115	94	1.223
July	100	110	105	94	1.117
August	110	90	100	94	1.064
September	85	95	90	94	0.957
October	75	85	80	94	0.851
November	85	75	80	94	0.851
December	80	80	80	94	0.851
		Total average demand =	1,128		

[a]Average monthly demand $= \frac{1{,}128}{12 \text{ months}} = 94$

[b]Seasonal index $= \frac{\text{Average 2 year demand}}{\text{Average monthly demand}}$

Monthly sales of one brand of telephone answering machine at Eichler Supplies are shown in Table 5.9, for the two most recent years. The average demand in each month is computed and these values are divided by the overall average (94) to find the seasonal index for each month. We then use the seasonal indices from Table 5.9 to adjust future forecasts. For example, suppose we expected the third year's annual demand for answering machines to be 1,200 units, which is 100 per month. We should not forecast each month to have a demand of 100, but we should adjust these based on the seasonal indices as follows:

January $\frac{1200}{12} \times 0.957 = 96$

February $\frac{1200}{12} \times 0.851 = 85$

March $\frac{1200}{12} \times 0.904 = 90$ and so on

5.6 Monitoring and Controlling Forecasts

After a forecast has been completed, it is important that it not be forgotten. No manager wants to be reminded when his or her forecast is horribly inaccurate, but a firm needs to determine why the actual demand (or whatever variable is being examined) differed significantly from that projected.

One way to monitor forecasts to ensure that they are performing well is to employ a **tracking signal.** A tracking signal is a measurement of how well the forecast is predicting actual values.

A tracking signal is computed as

$$\text{Tracking signal} = \frac{\text{RSFE}}{\text{MAD}} \qquad (5.13)$$

$$= \frac{\sum(\text{forecast error})}{\text{MAD}}$$

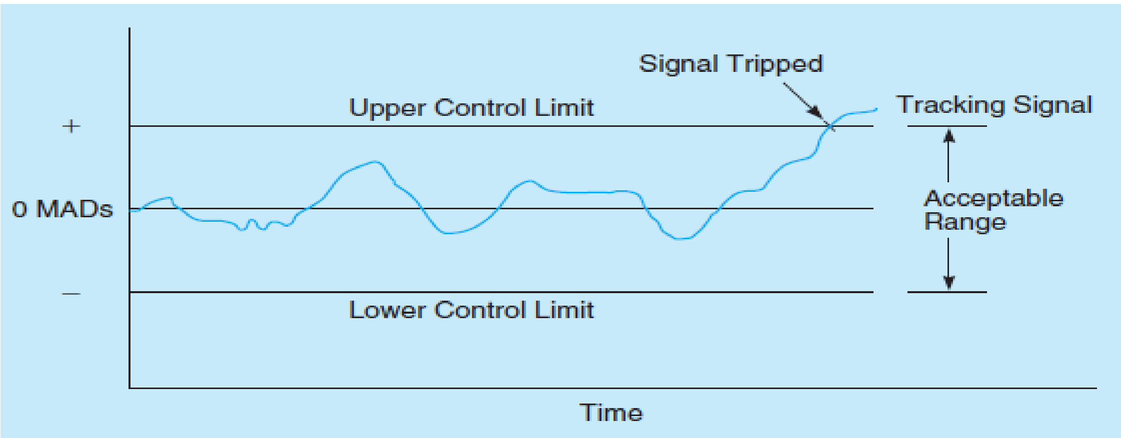
Where

$$\text{MAD} = \frac{\sum|\text{forecast error}|}{n}$$

Positive tracking signals indicate that demand is greater than the forecast. Negative signals mean that demand is less than forecast. A good tracking signal—that is, one with a low RSFE—has about as much positive error as it has negative error. In other words, small deviations are okay, but the positive and negative deviations should balance so that the tracking signal centers closely around zero.

When tracking signals are calculated, they are compared with predetermined control limits. When a tracking signal exceeds an upper or lower limit, a signal is tripped. This means that there is a problem with the forecasting method and management may want to reevaluate the way it forecasts demand. Figure 5.6 shows the graph of a tracking signal that is exceeding the range of acceptable variation. If the model being used is exponential smoothing, perhaps the smoothing constant needs to be readjusted.

Figure 5.6 Plot of Tracking Signals



KIMBALL'S BAKERY EXAMPLE: Here is an example that shows how the tracking signal and RSFE can be computed. Kimball's Bakery's quarterly sales of croissants (in thousands), as well as forecast demand and error computations, are in the following table. The objective is to compute the tracking signal and determine whether forecasts are performing adequately.

Time Period	Actual Demand	Forecast Demand	Error	RSFE	|Forecast Error|	Cumulative Error	MAD	Tracking Signal
1	90	100	−10	−10	10	10	10.0	−1
2	95	100	−5	−15	5	15	7.5	−2
3	115	100	+15	0	15	30	10.0	0
4	100	110	−10	−10	10	40	10.0	−1
5	125	110	+15	+5	15	55	11.0	+0.5
6	140	110	+30	+35	30	85	14.2	+2.5

In period 6, the calculations are

$$\text{MAD} = \frac{\sum|\text{forecast error}|}{\text{n}}$$

$$= \frac{85}{6} = 14.2$$

$$\text{Tracking signal} = \frac{\text{RSFE}}{\text{MAD}}$$

$$= \frac{35}{14.2} = 2.5\text{MADs}$$

This tracking signal is within acceptable limits. We see that it drifted from −2.0 MADs to +2.5 MADs

SECTION 4.1 INTRODUCTION

Regression analysis is a very valuable tool for today's manager. Regression has been used to model such things as the relationship between level of education and income, the price of a house and the square footage and the sales volume for a company relative to the dollars spent on advertising. Cost estimation models are often regression models.

There are generally two purposes for regression analysis. The first is to understand the relationship between variables such as advertising expenditures and sales. The second purpose is to predict the value of one variable based on the value of the other.

In any regression model, the variable to be predicted is called the **dependent variable** or **response variable**. The value of this is said to be dependent upon the value of an independent variable, which is sometimes called an **explanatory variable** or a **predictor variable**.

SECTION 4.2 SCATTER DIAGRAMS

To investigate the relationship between variables, it is helpful to look at a graph of the data. Such a graph is often called a **scatter diagram** or a **scatter plot.** Normally the independent variable is plotted on the horizontal axis and the dependent variable is plotted on the vertical axis. The following example will illustrate this:

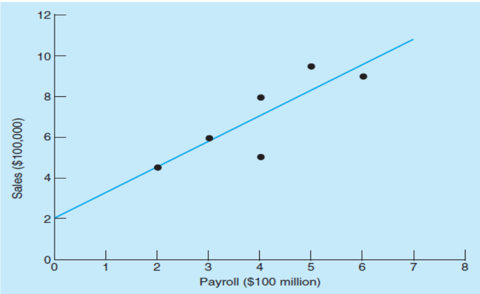
Triple A Construction Company renovates old homes in Albany. Over time, the company has found that its dollar volume of renovation work is dependent on the Albany area payroll. The figures for Triple A's revenues and the amount of money earned by wage earners in Albany for the past six years are presented in Table 4.1. Economists have predicted the local area payroll to be $600 million next year and Triple A wants to plan accordingly.

Table 4.1 Triple A Construction Company Sales and Local Payroll

Triple A's Sales ($100,000)	Local Payroll($100,000,000)
6	3
8	4
9	6
5	4
4.5	2
9.5	5

Figure 4.1 provides a scatter diagram for the Triple A Construction data given in Table 4.1. This graph indicated that higher values for the local payroll seem to result in higher sales for the company. There is not a perfect relationship because not all the points lie in a straight line, but there is a relationship. A line has been drawn through the data to help show the relationship that exists between the payroll and sales.

FIGURE 4.1
Scatter Diagram of Triple A Construction Company Data



SECTION 4.3 SIMPLE LINEAR REGRESSION

In any regression model, there is an implicit assumption (which can be tested) that a relationship exists between the variables. There is also some random error that cannot be predicted. The underlying simple linear regression model is

$$Y = \beta_0 + \beta_1 X + \epsilon \qquad (4.1)$$

Where Y = dependent variable (response variable)

X = independent variable (predictor variable or explanatory variable)

β_0 = intercept (value of Y when $X = 0$)

β_1 = slope of regression line

ϵ = random error

The true values of the intercept and slope are not known and therefore they are estimated using sample data. The regression equation based on sample data is given as

$$\hat{Y} = b_0 + b_1 X \qquad (4.2)$$

where $\hat{Y}$ = predicted value of Y

b_0 = estimate of β_0, based on sample results

b_1 = estimate of β_1, based on sample results

Error is defined as

$$\text{Error} = (\text{Actual value}) - (\text{Predicted value})$$

$$e = Y - \hat{Y} \qquad (4.3)$$

Some errors may be positive or negative, the average error could be zero even though there are extremely large errors—both positive and negative. To eliminate the difficulty of negative errors canceling positive errors, the errors can be squared. The best regression line will be defined as one with the minimum sum of the squared errors. For this reason, regression analysis is sometimes called **least-squares** regression.

Statisticians have developed formulas that we can use to find the equation of a straight line that would minimize the sum of the squared errors. The simple linear regression equation is

$$\hat{Y} = b_0 + b_1 X$$

The following formulas can be used to compute the intercept and the slope:

$$\bar{X} = \frac{\sum X}{n} = \text{average (mean)of } X \text{ values}$$

$$\bar{Y} = \frac{\sum Y}{n} = \text{average (mean)of } Y \text{ values}$$

$$b_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} \qquad (4.4)$$

$$b_0 = \bar{Y} - b_1 \bar{X} \qquad (4.5)$$

Example: Computing the Regression equation for the Triple A construction Company

Local Payroll ($100,000,000)	Triple A's Sales ($100,000)
3	6
4	8
6	9
4	5
2	4.5
5	9.5

First we calculate the mean values of X & Y as follows

$$\bar{X} = \frac{\sum X}{n} = \frac{24}{6} = 4$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{42}{6} = 7$$

Table 4.2 Regression Calculations for Triple A Construction

X	Y	$(X-\bar{X})^2$	$(X-\bar{X})(Y-\bar{Y})$
3	6	$(3-4)^2 = 1$	$(3-4)(6-7) = 1$
4	8	$(4-4)^2 = 0$	$(4-4)(8-7) = 0$
6	9	$(6-4)^2 = 4$	$(6-4)(9-7) = 4$
4	5	$(4-4)^2 = 0$	$(4-4)(5-7) = 0$
2	4.5	$(2-4)^2 = 4$	$(2-4)(4.5-7) = 5$
5	9.5	$(5-4)^2 = 1$	$(5-4)(9.5-7) = 2.5$
$\sum X = 24$	$\sum Y = 42$	$\sum(X-\bar{X})^2 = 10$	$\sum(X-\bar{X})(Y-\bar{Y}) = 12.5$

Now slope and intercept of the regression equation can be calculated as

$$b_1 = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2} = \frac{12.5}{10} = 1.25$$

$$b_0 = \bar{Y} - b_1\bar{X} = 7 - (1.25)(4) = 2$$

The estimated regression equation therefore is

$$\hat{Y} = 2 + 1.25X$$

or $\text{sales} = 2 + 1.25(\text{payroll})$

If the payroll next year is $600 million ($X = 6$), then the predicted value would be

$$\hat{Y} = 2 + 1.25(6) = 9.5$$

or $950,000

One of the purposes of regression is to understand the relationship among variables. This model tells us that for each $100 million (represented by X) increase in the payroll, we would expect the sales to increase by $125,000 since $b_1 = 1.25$ ($100,000s). This model helps Triple A Construction see how the local economy and company sales are related.

Example: Fit a regression curve to the following data

X	1	3	5	7	9
Y	15	18	21	24	22

Solution:

X	Y	(X-$\bar{X}$)	(Y-$\bar{Y}$)	(X-$\bar{X}$)(Y-$\bar{Y}$)	(X-$\bar{X}$)^2
1	15	-4	-5	20	16
3	18	-2	-2	4	4
5	21	0	1	0	0
7	24	2	4	8	4
9	22	4	2	8	16
$\sum$X = 25	$\sum$Y = 100			$\sum$ (X-$\bar{X}$)(Y-$\bar{Y}$) = 40	$\sum$ (X-$\bar{X}$)^2 = 40

First we calculate the mean values of X & Y as follows

$$\bar{X} = \frac{\sum X}{n} = \frac{25}{5} = 5$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{100}{5} = 20$$

Now slope and intercept of the regression equation can be calculated as

$$b_1 = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2} = \frac{40}{40} = 1$$

$$b_0 = \bar{Y} - b_1\bar{X} = 20 - (1)(5) = 15$$

The estimated regression equation therefore is

$$\hat{Y} = b_0 + b_1 X$$

or $\hat{Y} = 15 + (1)X$

SECTION 4.4 MEASURING THE FIT OF THE REGRESSION MODEL

A regression equation can be developed for any variables X and Y, even random numbers. We certainly would not have any confidence in the ability of one random number to predict the value of another random number. How do we know that the model is actually helpful in predicting Y based on X? Should we have confidence in this model? Does the model provide better predictions (smaller errors) than simply using the average of the Y values?

In the Triple A Construction example, sales figures (Y) varied from a low of 4.5 to a high of 9.5 and the mean was 7. If each sales value is compared with the mean, we see how far they deviate from the mean and we could compute a measure of the total variability in sales. Because Y is sometimes higher and sometimes lower than the mean, there may be both positive and negative deviations. Simply summing these values would be misleading because the negatives would cancel out the positives, making it appear that the numbers are closer to the mean than they actually are. To prevent this problem, we will use the sum of the **squares total (SST)** to measure the total variability in Y:

$$\text{SST}= \sum(Y-\bar{Y})^2 \qquad (4.6)$$

If we did not use X to predict Y, we would simply use the mean of Y as the prediction and the SST would measure the accuracy of our predictions. However, a regression line may be used to predict the value of Y and while there are still errors involved, the sum of these squared errors will be less than the total sum of squares just computed. The sum of squares error (SSE) is

$$\text{SSE} = \sum e^2 = \sum(Y-\hat{Y})^2 \qquad (4.7)$$

Table 4.3 Sum of Squares for Triple A Conctruction

Y	X	$(Y-\bar{Y})^2$	$\hat{Y}$	$(Y-\hat{Y})^2$	$(\hat{Y}-\bar{Y})^2$
6	3	$(6-7)^2=1$	$2+1.25(3)=5.75$	0.0625	1.563
8	4	$(8-7)^2=1$	$2+1.25(4)=7.00$	1	0
9	6	$(9-7)^2=4$	$2+1.25(6)=9.50$	0.25	6.25
5	4	$(5-7)^2=4$	$2+1.25(4)=7.00$	4	0
4.5	2	$(4.5-7)^2=6.25$	$2+1.25(2)=4.50$	0	6.25
9.5	5	$(9.5-7)^2=6.25$	$2+1.25(5)=8.25$	1.5625	1.563
		$\sum(Y-\bar{Y})^2=22.5$ SST= 22.5		$\sum(Y-\hat{Y})^2=6.875$ SSE = 6.875	$\sum(\hat{Y}-\bar{Y})^2=15.625$ SSR = 15.625

Table 4.3 provides the calculations for the Triple A Construction example. The mean ($\bar{Y}=7$) is compared to each value and we get

$$\text{SST} = 22.5$$

The prediction ($\hat{Y}$) for each observation is computed and compared to the actual value. This results in

$$\text{SSE} = 6.875$$

The SSE is much lower than the SST. Using the regression line has reduced the variability in the sum of squares by $22.5 - 6.875 = 15.625$. This is called the **sum of squares due to regression (SSR)** and indicates how much of the total variability in Y is explained by the regression model. Mathematically, this can be calculated as

$$\text{SSR} = \sum(\hat{Y} - \bar{Y})^2 \tag{4.8}$$

Table 4.3 indicates

$$\text{SSR} = 15.625$$

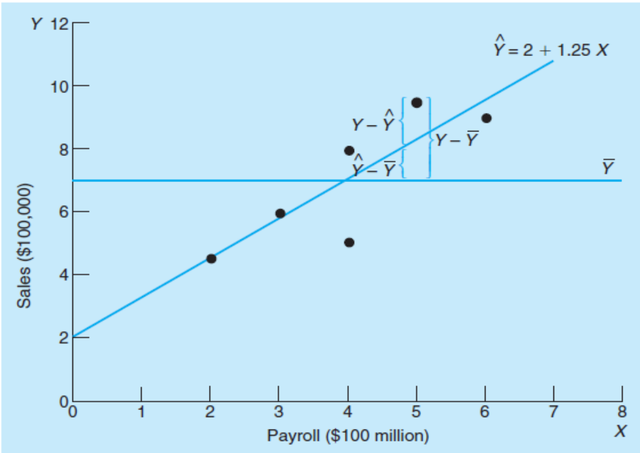
There is a very important relationship between the sums of squares that we have computed:

Sum of squares total = Sum of squares due to regression + Sum of squares error

i.e., $$\text{SST} = \text{SSR} + \text{SSE} \tag{4.9}$$

Figure 4.2 displays the data for Triple A Construction. The regression line is shown, as is a line representing the mean of the Y values. The errors used in computing the sums of squares are shown on this graph. Notice how the sample points are closer to the regression line than they are to the mean.

FIGURE 4.2
Deviations from the Regression Line and from the Mean



Coefficient of Determination

The SSR is sometimes called the explained variability in Y while the SSE is the unexplained variability in Y. The proportion of the variability in Y that is explained by the regression equation is called the **coefficient of determination** and is denoted by r^2. Thus,

$$r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{4.10}$$

For Triple A Construction, we have

$$r^2 = \frac{15.625}{22.5} = 0.6944$$

This means that about 69% of the variability in sales (Y) is explained by the regression equation based on payroll (X).

If every point in the sample were on the regression line (meaning all errors are 0), then 100% of the variability in Y could be explained by the regression equation, so $r^2 = 1$ and SSE = 0. The lowest possible value of r^2 is 0, indicating that X explains 0% of the variability in Y. Thus, r^2 can range from a low of 0 to a high of 1. In developing regression equations, a good model will have an r^2 value close to 1.

Example: In a simple regression model study, the following results are found:

Given $\sum(Y - \bar{Y})^2 = 20$ and $\sum(Y - \hat{Y})^2 = 2$.

Complete the following table:

SST	SSE	SSR	r^2

Solution:

SST	SSE	SSR	r^2
$\sum(Y - \bar{Y})^2$	$\sum(Y - \hat{Y})^2$	SST-SSE	SSR/SST
20	2	18	0.90

Correlation Coefficient

Another measure related to the coefficient of determination is the **coefficient of correlation**. This measure also expresses the degree or strength of the linear relationship. It is usually expressed as r and can be any number between and including +1 and −1. Figure 4.3 illustrates possible scatter diagrams for different values of r. The value of r is the square root of r^2. It is negative if the slope is negative, and it is positive if the slope is positive. Thus,

$$r = \pm\sqrt{r^2} \tag{4-11}$$

For the Triple A Construction example with $r^2 = 0.6944$,

$$r = \sqrt{0.6944} = 0.8333$$

We know it is positive because the slope is +1.25.

FIGURE 4.3
Four Values of the Correlation Coefficient