2.5 Applications of Matrix Operations

- Write and use a stochastic matrix.
- Use matrix multiplication to encode and decode messages.
- Use matrix algebra to analyze an economic system (Leontief input-output model).
- Find the least squares regression line for a set of data.

STOCHASTIC MATRICES

Many types of applications involve a finite set of *states* $\{S_1, S_2, \ldots, S_n\}$ of a given population. For instance, residents of a city may live downtown or in the suburbs. Voters may vote Democrat, Republican, or Independent. Soft drink consumers may buy Coca-Cola, Pepsi Cola, or another brand.

The probability that a member of a population will change from the jth state to the ith state is represented by a number p_{ij}, where $0 \le p_{ij} \le 1$. A probability of $p_{ij} = 0$ means that the member is certain *not* to change from the jth state to the ith state, whereas a probability of $p_{ij} = 1$ means that the member is certain to change from the jth state to the ith state.

$$P = \begin{array}{c} \overbrace{\begin{array}{cccc} S_1 & S_2 & \cdots & S_n \end{array}}^{\text{From}} \\ \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \end{array} \left.\begin{array}{c} S_1 \\ S_2 \\ \vdots \\ S_n \end{array}\right\} \text{To}$$

P is called the **matrix of transition probabilities** because it gives the probabilities of each possible type of transition (or change) within the population.

At each transition, each member in a given state must either stay in that state or change to another state. For probabilities, this means that the sum of the entries in any column of P is 1. For instance, in the first column,

$$p_{11} + p_{21} + \cdots + p_{n1} = 1.$$

Such a matrix is called **stochastic** (the term "stochastic" means "regarding conjecture"). That is, an $n \times n$ matrix P is a **stochastic matrix** when each entry is a number between 0 and 1 inclusive, and the sum of the entries in each column of P is 1.

EXAMPLE 1 Examples of Stochastic Matrices and Nonstochastic Matrices

The matrices in parts (a) and (b) are stochastic, but the matrices in parts (c) and (d) are not.

a. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{4} & \frac{2}{3} & 0 \end{bmatrix}$

c. $\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.2 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.5 \end{bmatrix}$ **d.** $\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix}$

Example 2 describes the use of a stochastic matrix to measure consumer preferences.

EXAMPLE 2 A Consumer Preference Model

Two competing companies offer satellite television service to a city with 100,000 households. Figure 2.1 shows the changes in satellite subscriptions each year. Company A now has 15,000 subscribers and Company B has 20,000 subscribers. How many subscribers will each company have in one year?

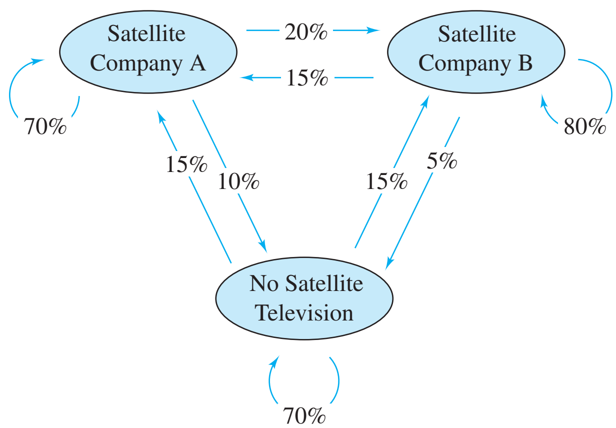


Figure 2.1

SOLUTION

The matrix representing the given transition probabilities is

$$P = \begin{array}{c} \overbrace{\begin{array}{ccc} \text{A} & \text{B} & \text{None} \end{array}}^{\text{From}} \\ \begin{bmatrix} 0.70 & 0.15 & 0.15 \\ 0.20 & 0.80 & 0.15 \\ 0.10 & 0.05 & 0.70 \end{bmatrix} \end{array} \begin{array}{l} \\ \left.\begin{array}{l} \text{A} \\ \text{B} \\ \text{None} \end{array}\right\} \text{To} \end{array}$$

and the **state matrix** representing the current populations in the three states is

$$X = \begin{bmatrix} 15,000 \\ 20,000 \\ 65,000 \end{bmatrix}. \qquad \begin{array}{l} \text{A} \\ \text{B} \\ \text{None} \end{array}$$

To find the state matrix representing the populations in the three states in one year, multiply P by X to obtain

$$\begin{aligned} PX &= \begin{bmatrix} 0.70 & 0.15 & 0.15 \\ 0.20 & 0.80 & 0.15 \\ 0.10 & 0.05 & 0.70 \end{bmatrix} \begin{bmatrix} 15,000 \\ 20,000 \\ 65,000 \end{bmatrix} \\ &= \begin{bmatrix} 23,250 \\ 28,750 \\ 48,000 \end{bmatrix}. \end{aligned}$$

In one year, Company A will have 23,250 subscribers and Company B will have 28,750 subscribers.

One appeal of the matrix solution in Example 2 is that once you have created the model, it becomes relatively easy to find the state matrices representing future years by repeatedly multiplying by the matrix P. Example 3 demonstrates this process.

EXAMPLE 3 A Consumer Preference Model

Assuming the matrix of transition probabilities from Example 2 remains the same year after year, find the number of subscribers each satellite television company will have after (a) 3 years, (b) 5 years, and (c) 10 years. Round each answer to the nearest whole number.

SOLUTION

a. From Example 2, you know that the numbers of subscribers after 1 year are

$$PX = \begin{bmatrix} 23{,}250 \\ 28{,}750 \\ 48{,}000 \end{bmatrix}. \quad \begin{matrix} \text{A} \\ \text{B} \\ \text{None} \end{matrix} \quad \text{After 1 year}$$

Because the matrix of transition probabilities is the same from the first year to the third year, the numbers of subscribers after 3 years are

$$P^3X \approx \begin{bmatrix} 30{,}283 \\ 39{,}042 \\ 30{,}675 \end{bmatrix}. \quad \begin{matrix} \text{A} \\ \text{B} \\ \text{None} \end{matrix} \quad \text{After 3 years}$$

After 3 years, Company A will have 30,283 subscribers and Company B will have 39,042 subscribers.

b. The numbers of subscribers after 5 years are

$$P^5X \approx \begin{bmatrix} 32{,}411 \\ 43{,}812 \\ 23{,}777 \end{bmatrix}. \quad \begin{matrix} \text{A} \\ \text{B} \\ \text{None} \end{matrix} \quad \text{After 5 years}$$

After 5 years, Company A will have 32,411 subscribers and Company B will have 43,812 subscribers.

c. The numbers of subscribers after 10 years are

$$P^{10}X \approx \begin{bmatrix} 33{,}287 \\ 47{,}147 \\ 19{,}566 \end{bmatrix}. \quad \begin{matrix} \text{A} \\ \text{B} \\ \text{None} \end{matrix} \quad \text{After 10 years}$$

After 10 years, Company A will have 33,287 subscribers and Company B will have 47,147 subscribers.

In Example 3, notice that there is little difference between the numbers of subscribers after 5 years and after 10 years. If you continue the process shown in this example, then the numbers of subscribers eventually reach a **steady state.** That is, as long as the matrix P does not change, the matrix product P^nX approaches a limit $\overline{X}$. In Example 3, the limit is the steady state matrix

$$\overline{X} = \begin{bmatrix} 33{,}333 \\ 47{,}619 \\ 19{,}048 \end{bmatrix}. \quad \begin{matrix} \text{A} \\ \text{B} \\ \text{None} \end{matrix} \quad \text{Steady state}$$

Check to see that $P\overline{X} = \overline{X}$, as follows.

$$P\overline{X} = \begin{bmatrix} 0.70 & 0.15 & 0.15 \\ 0.20 & 0.80 & 0.15 \\ 0.10 & 0.05 & 0.70 \end{bmatrix} \begin{bmatrix} 33{,}333 \\ 47{,}619 \\ 19{,}048 \end{bmatrix}$$

$$\approx \begin{bmatrix} 33{,}333 \\ 47{,}619 \\ 19{,}048 \end{bmatrix} = \overline{X}$$

CRYPTOGRAPHY

A **cryptogram** is a message written according to a secret code (the Greek word *kryptos* means "hidden"). The following describes a method of using matrix multiplication to **encode** and **decode** messages.

To begin, assign a number to each letter in the alphabet (with 0 assigned to a blank space), as follows.

0 = __	14 = N
1 = A	15 = O
2 = B	16 = P
3 = C	17 = Q
4 = D	18 = R
5 = E	19 = S
6 = F	20 = T
7 = G	21 = U
8 = H	22 = V
9 = I	23 = W
10 = J	24 = X
11 = K	25 = Y
12 = L	26 = Z
13 = M	

Then convert the message to numbers and partition it into **uncoded row matrices,** each having n entries, as demonstrated in Example 4.

EXAMPLE 4 Forming Uncoded Row Matrices

Write the uncoded row matrices of size 1×3 for the message MEET ME MONDAY.

SOLUTION

Partitioning the message (including blank spaces, but ignoring punctuation) into groups of three produces the following uncoded row matrices.

$$\begin{array}{ccccccccccccccc} [13 & 5 & 5] & [20 & 0 & 13] & [5 & 0 & 13] & [15 & 14 & 4] & [1 & 25 & 0] \\ M & E & E & T & \_ & M & E & \_ & M & O & N & D & A & Y & \_ \end{array}$$

Note the use of a blank space to fill out the last uncoded row matrix.

LINEAR ALGEBRA APPLIED

Because of the heavy use of the Internet to conduct business, Internet security is of the utmost importance. If a malicious party should receive confidential information such as passwords, personal identification numbers, credit card numbers, social security numbers, bank account details, or corporate secrets, the effects can be damaging. To protect the confidentiality and integrity of such information, the most popular forms of Internet security use data *encryption,* the process of encoding information so that the only way to decode it, apart from a brute force "exhaustion attack," is to use a *key*. Data encryption technology uses algorithms based on the material presented here, but on a much more sophisticated level, to prevent malicious parties from discovering the key.

To **encode** a message, choose an $n \times n$ invertible matrix A and multiply the uncoded row matrices (on the right) by A to obtain **coded row matrices.** Example 5 demonstrates this process.

Andrea Danti/Shutterstock.com

EXAMPLE 5 Encoding a Message

Use the following invertible matrix

$$A = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & 3 \\ 1 & -1 & -4 \end{bmatrix}$$

to encode the message MEET ME MONDAY.

SOLUTION

Obtain the coded row matrices by multiplying each of the uncoded row matrices found in Example 4 by the matrix A, as follows.

$$\begin{array}{ccccc} \text{Uncoded} & & \text{Encoding} & & \text{Coded Row} \\ \text{Row Matrix} & & \text{Matrix } A & & \text{Matrix} \end{array}$$

$$[13 \quad 5 \quad 5]\begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & 3 \\ 1 & -1 & -4 \end{bmatrix} = [13 \quad -26 \quad 21]$$

$$[20 \quad 0 \quad 13]\begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & 3 \\ 1 & -1 & -4 \end{bmatrix} = [33 \quad -53 \quad -12]$$

$$[5 \quad 0 \quad 13]\begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & 3 \\ 1 & -1 & -4 \end{bmatrix} = [18 \quad -23 \quad -42]$$

$$[15 \quad 14 \quad 4]\begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & 3 \\ 1 & -1 & -4 \end{bmatrix} = [5 \quad -20 \quad 56]$$

$$[1 \quad 25 \quad 0]\begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & 3 \\ 1 & -1 & -4 \end{bmatrix} = [-24 \quad 23 \quad 77]$$

The sequence of coded row matrices is

$$[13 \quad -26 \quad 21][33 \quad -53 \quad -12][18 \quad -23 \quad -42][5 \quad -20 \quad 56][-24 \quad 23 \quad 77].$$

Finally, removing the matrix notation produces the following cryptogram.

$$13 \quad -26 \quad 21 \quad 33 \quad -53 \quad -12 \quad 18 \quad -23 \quad -42 \quad 5 \quad -20 \quad 56 \quad -24 \quad 23 \quad 77$$

For those who do not know the encoding matrix A, decoding the cryptogram found in Example 5 is difficult. But for an authorized receiver who knows the encoding matrix A, decoding is relatively simple. The receiver just needs to multiply the coded row matrices by A^{-1} to retrieve the uncoded row matrices. In other words, if

$$X = [x_1 \ x_2 \cdots x_n]$$

is an uncoded $1 \times n$ matrix, then $Y = XA$ is the corresponding encoded matrix. The receiver of the encoded matrix can decode Y by multiplying on the right by A^{-1} to obtain

$$YA^{-1} = (XA)A^{-1} = X.$$

Example 6 demonstrates this procedure.

EXAMPLE 6 Decoding a Message

Simulation
Explore this concept further with an electronic simulation available at *www.cengagebrain.com.*

Use the inverse of the matrix

$$A = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & 3 \\ 1 & -1 & -4 \end{bmatrix}$$

to decode the cryptogram

$$13 \quad -26 \quad 21 \quad 33 \quad -53 \quad -12 \quad 18 \quad -23 \quad -42 \quad 5 \quad -20 \quad 56 \quad -24 \quad 23 \quad 77.$$

SOLUTION

Begin by using Gauss-Jordan elimination to find A^{-1}.

$$\begin{array}{ccc} [A \quad I] & & [I \quad A^{-1}] \\ \left[\begin{array}{rrrrrr} 1 & -2 & 2 & 1 & 0 & 0 \\ -1 & 1 & 3 & 0 & 1 & 0 \\ 1 & -1 & -4 & 0 & 0 & 1 \end{array}\right] & \longrightarrow & \left[\begin{array}{rrrrrr} 1 & 0 & 0 & -1 & -10 & -8 \\ 0 & 1 & 0 & -1 & -6 & -5 \\ 0 & 0 & 1 & 0 & -1 & -1 \end{array}\right] \end{array}$$

Now, to decode the message, partition the message into groups of three to form the coded row matrices

$$[13 \ \ -26 \ \ 21][33 \ \ -53 \ \ -12][18 \ \ -23 \ \ -42][5 \ \ -20 \ \ 56][-24 \ \ 23 \ \ 77].$$

To obtain the decoded row matrices, multiply each coded row matrix by A^{-1} (on the right).

$$\begin{array}{ccc} \text{Coded Row Matrix} & \text{Decoding Matrix } A^{-1} & \text{Decoded Row Matrix} \end{array}$$

$$[13 \ \ -26 \ \ 21]\begin{bmatrix} -1 & -10 & -8 \\ -1 & -6 & -5 \\ 0 & -1 & -1 \end{bmatrix} = [13 \ \ 5 \ \ 5]$$

$$[33 \ \ -53 \ \ -12]\begin{bmatrix} -1 & -10 & -8 \\ -1 & -6 & -5 \\ 0 & -1 & -1 \end{bmatrix} = [20 \ \ 0 \ \ 13]$$

$$[18 \ \ -23 \ \ -42]\begin{bmatrix} -1 & -10 & -8 \\ -1 & -6 & -5 \\ 0 & -1 & -1 \end{bmatrix} = [5 \ \ 0 \ \ 13]$$

$$[5 \ \ -20 \ \ 56]\begin{bmatrix} -1 & -10 & -8 \\ -1 & -6 & -5 \\ 0 & -1 & -1 \end{bmatrix} = [15 \ \ 14 \ \ 4]$$

$$[-24 \ \ 23 \ \ 77]\begin{bmatrix} -1 & -10 & -8 \\ -1 & -6 & -5 \\ 0 & -1 & -1 \end{bmatrix} = [1 \ \ 25 \ \ 0]$$

The sequence of decoded row matrices is

$$[13 \ \ 5 \ \ 5][20 \ \ 0 \ \ 13][5 \ \ 0 \ \ 13][15 \ \ 14 \ \ 4][1 \ \ 25 \ \ 0]$$

and the message is

$$\begin{array}{ccccccccccccccc} 13 & 5 & 5 & 20 & 0 & 13 & 5 & 0 & 13 & 15 & 14 & 4 & 1 & 25 & 0. \\ M & E & E & T & \_ & M & E & \_ & M & O & N & D & A & Y & \_ \end{array}$$

LEONTIEF INPUT-OUTPUT MODELS

In 1936, American economist Wassily W. Leontief (1906–1999) published a model concerning the input and output of an economic system. In 1973, Leontief received a Nobel prize for his work in economics. A brief discussion of Leontief's model follows.

Suppose that an economic system has n different industries $I_1, I_2, \ldots, I_n$, each of which has **input** needs (raw materials, utilities, etc.) and an **output** (finished product). In producing each unit of output, an industry may use the outputs of other industries, including itself. For example, an electric utility uses outputs from other industries, such as coal and water, and also uses its own electricity.

Let d_{ij} be the amount of output the jth industry needs from the ith industry to produce one unit of output per year. The matrix of these coefficients is called the **input-output matrix.**

$$\begin{array}{c} \overbrace{\begin{array}{cccc} I_1 & I_2 & \cdots & I_n \end{array}}^{\text{User (Output)}} \\ D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \left.\begin{array}{c} I_1 \\ I_2 \\ \vdots \\ I_n \end{array}\right\} \text{Supplier (Input)} \end{array}$$

To understand how to use this matrix, consider $d_{12} = 0.4$. This means that for Industry 2 to produce one unit of its product, it must use 0.4 unit of Industry 1's product. If $d_{33} = 0.2$, then Industry 3 needs 0.2 unit of its own product to produce one unit. For this model to work, the values of d_{ij} must satisfy $0 \le d_{ij} \le 1$ and the sum of the entries in any column must be less than or equal to 1.

EXAMPLE 7 Forming an Input-Output Matrix

Consider a simple economic system consisting of three industries: electricity, water, and coal. Production, or output, of one unit of electricity requires 0.5 unit of itself, 0.25 unit of water, and 0.25 unit of coal. Production of one unit of water requires 0.1 unit of electricity, 0.6 unit of itself, and 0 units of coal. Production of one unit of coal requires 0.2 unit of electricity, 0.15 unit of water, and 0.5 unit of itself. Find the input-output matrix for this system.

SOLUTION

The column entries show the amounts each industry requires from the others, and from itself, to produce one unit of output.

$$\begin{array}{c} \overbrace{\begin{array}{ccc} \text{E} & \text{W} & \text{C} \end{array}}^{\text{User (Output)}} \\ \begin{bmatrix} 0.5 & 0.1 & 0.2 \\ 0.25 & 0.6 & 0.15 \\ 0.25 & 0 & 0.5 \end{bmatrix} \left.\begin{array}{c} \text{E} \\ \text{W} \\ \text{C} \end{array}\right\} \text{Supplier (Input)} \end{array}$$

The row entries show the amounts each industry supplies to the others, and to itself, for that industry to produce one unit of output. For instance, the electricity industry supplies 0.5 unit to itself, 0.1 unit to water, and 0.2 unit to coal.

To develop the Leontief input-output model further, let the total output of the ith industry be denoted by x_i. If the economic system is **closed** (meaning that it sells its products only to industries within the system, as in the example above), then the total output of the ith industry is given by the linear equation

$$x_i = d_{i1}x_1 + d_{i2}x_2 + \cdots + d_{in}x_n. \qquad \text{Closed system}$$

On the other hand, if the industries within the system sell products to nonproducing groups (such as governments or charitable organizations) outside the system, then the system is **open** and the total output of the ith industry is given by

$$x_i = d_{i1}x_1 + d_{i2}x_2 + \cdots + d_{in}x_n + e_i \qquad \text{Open system}$$

where e_i represents the external demand for the ith industry's product. The following system of n linear equations represents the collection of total outputs for an open system.

$$\begin{aligned} x_1 &= d_{11}x_1 + d_{12}x_2 + \cdots + d_{1n}x_n + e_1 \\ x_2 &= d_{21}x_1 + d_{22}x_2 + \cdots + d_{2n}x_n + e_2 \\ &\vdots \\ x_n &= d_{n1}x_1 + d_{n2}x_2 + \cdots + d_{nn}x_n + e_n \end{aligned}$$

The matrix form of this system is $X = DX + E$, where X is the **output matrix** and E is the **external demand matrix.**

EXAMPLE 8 Solving for the Output Matrix of an Open System

An economic system composed of three industries has the following input-output matrix.

$$\begin{array}{cc} & \overbrace{\begin{array}{ccc} \text{A} & \text{B} & \text{C} \end{array}}^{\text{User (Output)}} \\ D = & \begin{bmatrix} 0.1 & 0.43 & 0 \\ 0.15 & 0 & 0.37 \\ 0.23 & 0.03 & 0.02 \end{bmatrix} \end{array} \left.\begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array}\right\} \text{Supplier (Input)}$$

Find the output matrix X when the external demands are

$$E = \begin{bmatrix} 20{,}000 \\ 30{,}000 \\ 25{,}000 \end{bmatrix}. \quad \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array}$$

Round each matrix entry to the nearest whole number.

SOLUTION

Letting I be the identity matrix, write the equation $X = DX + E$ as $IX - DX = E$, which means that $(I - D)X = E$. Using the matrix D above produces

$$I - D = \begin{bmatrix} 0.9 & -0.43 & 0 \\ -0.15 & 1 & -0.37 \\ -0.23 & -0.03 & 0.98 \end{bmatrix}.$$

Finally, applying Gauss-Jordan elimination to the system of linear equations represented by $(I - D)X = E$ produces

$$\begin{bmatrix} 0.9 & -0.43 & 0 & 20{,}000 \\ -0.15 & 1 & -0.37 & 30{,}000 \\ -0.23 & -0.03 & 0.98 & 25{,}000 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 46{,}616 \\ 0 & 1 & 0 & 51{,}058 \\ 0 & 0 & 1 & 38{,}014 \end{bmatrix}.$$

So, the output matrix is

$$X = \begin{bmatrix} 46{,}616 \\ 51{,}058 \\ 38{,}014 \end{bmatrix}. \quad \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array}$$

To produce the given external demands, the outputs of the three industries must be 46,616 units for industry A, 51,058 units for industry B, and 38,014 units for industry C.

REMARK

The economic systems described in Examples 7 and 8 are, of course, simple ones. In the real world, an economic system would include many industries or industrial groups. A detailed analysis using the Leontief input-output model could easily require an input-output matrix greater than 100 × 100 in size. Clearly, this type of analysis would require the aid of a computer.

LEAST SQUARES REGRESSION ANALYSIS

You will now look at a procedure used in statistics to develop linear models. The next example demonstrates a visual method for approximating a line of best fit for a given set of data points.

EXAMPLE 9 A Visual Straight-Line Approximation

Determine a line that appears to best fit the points $(1, 1)$, $(2, 2)$, $(3, 4)$, $(4, 4)$, and $(5, 6)$.

SOLUTION

Plot the points, as shown in Figure 2.2. It appears that a good choice would be the line whose slope is 1 and whose y-intercept is 0.5. The equation of this line is

$$y = 0.5 + x.$$

An examination of the line in Figure 2.2 reveals that you can improve the fit by rotating the line counterclockwise slightly, as shown in Figure 2.3. It seems clear that this line, whose equation is $y = 1.2x$, fits the given points better than the original line.

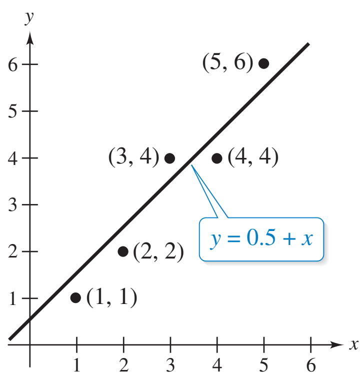


Figure 2.2

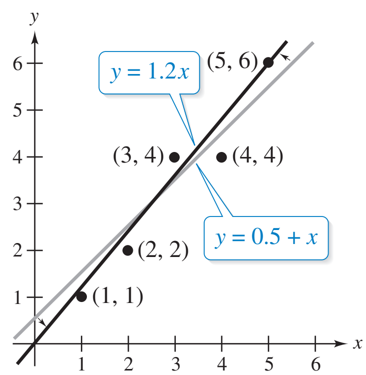


Figure 2.3

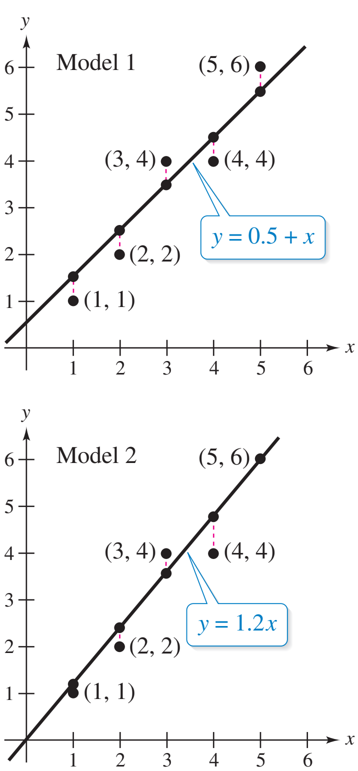


Figure 2.4

One way of measuring how well a function $y = f(x)$ fits a set of points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

is to compute the differences between the values from the function $f(x_i)$ and the actual values y_i. These values are shown in Figure 2.4. By squaring the differences and summing the results, you obtain a measure of error called the **sum of squared error.** The table shows the sums of squared errors for the two linear models.

Model 1: $f(x) = 0.5 + x$				*Model 2:* $f(x) = 1.2x$			
x_i	y_i	$f(x_i)$	$[y_i - f(x_i)]^2$	x_i	y_i	$f(x_i)$	$[y_i - f(x_i)]^2$
1	1	1.5	$(-0.5)^2$	1	1	1.2	$(-0.2)^2$
2	2	2.5	$(-0.5)^2$	2	2	2.4	$(-0.4)^2$
3	4	3.5	$(+0.5)^2$	3	4	3.6	$(+0.4)^2$
4	4	4.5	$(-0.5)^2$	4	4	4.8	$(-0.8)^2$
5	6	5.5	$(+0.5)^2$	5	6	6.0	$(0.0)^2$
Sum			1.25	Sum			1.00

The sums of squared errors confirm that the second model fits the given points better than the first model.

Of all possible linear models for a given set of points, the model that has the best fit is defined to be the one that minimizes the sum of squared error. This model is called the **least squares regression line,** and the procedure for finding it is called the **method of least squares.**

Definition of Least Squares Regression Line

For a set of points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

the **least squares regression line** is given by the linear function

$$f(x) = a_0 + a_1x$$

that minimizes the sum of squared error

$$[y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \cdots + [y_n - f(x_n)]^2.$$

To find the least squares regression line for a set of points, begin by forming the system of linear equations

$$\begin{aligned} y_1 &= f(x_1) + [y_1 - f(x_1)] \\ y_2 &= f(x_2) + [y_2 - f(x_2)] \\ &\vdots \\ y_n &= f(x_n) + [y_n - f(x_n)] \end{aligned}$$

where the right-hand term,

$$[y_i - f(x_i)]$$

of each equation is the error in the approximation of y_i by $f(x_i)$. Then write this error as

$$e_i = y_i - f(x_i)$$

and write the system of equations in the form

$$\begin{aligned} y_1 &= (a_0 + a_1x_1) + e_1 \\ y_2 &= (a_0 + a_1x_2) + e_2 \\ &\vdots \\ y_n &= (a_0 + a_1x_n) + e_n. \end{aligned}$$

Now, if you define Y, X, A, and E as

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad A = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}, \quad E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

then the n linear equations may be replaced by the matrix equation

$$Y = XA + E.$$

Note that the matrix X has a column of 1's (corresponding to a_0) and a column containing the x_i's. This matrix equation can be used to determine the coefficients of the least squares regression line, as follows.

REMARK

You will learn more about this procedure in Section 5.4.

Matrix Form for Linear Regression

For the regression model $Y = XA + E$, the coefficients of the least squares regression line are given by the matrix equation

$$A = (X^TX)^{-1}X^TY$$

and the sum of squared error is

$$E^TE.$$

Example 10 demonstrates the use of this procedure to find the least squares regression line for the set of points from Example 9.

EXAMPLE 10 Finding the Least Squares Regression Line

Find the least squares regression line for the points (1, 1), (2, 2), (3, 4), (4, 4), and (5, 6).

SOLUTION

The matrices X and Y are

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 4 \\ 6 \end{bmatrix}.$$

This means that

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix}$$

and

$$X^TY = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 4 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 17 \\ 63 \end{bmatrix}.$$

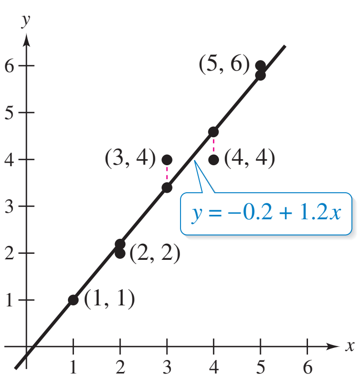
Now, using $(X^TX)^{-1}$ to find the coefficient matrix A, you have

$$\begin{aligned} A &= (X^TX)^{-1}X^TY \\ &= \tfrac{1}{50}\begin{bmatrix} 55 & -15 \\ -15 & 5 \end{bmatrix}\begin{bmatrix} 17 \\ 63 \end{bmatrix} \\ &= \begin{bmatrix} -0.2 \\ 1.2 \end{bmatrix}. \end{aligned}$$

So, the least squares regression line is

$$y = -0.2 + 1.2x$$

as shown in Figure 2.5. The sum of squared error for this line is 0.8 (verify this), which means that this line fits the data better than either of the two experimental linear models determined earlier.



Least Squares Regression Line

Figure 2.5