

EXAMPLE (a):

The American Express Company has long believed that its cardholders tend to travel more extensively than others—both on business and for pleasure. As part of a comprehensive research effort undertaken by a New York marketing research firm on behalf of American Express, a study was conducted to determine the relationship between travel and charges on the American Express card. The research firm selected a random sample of 25 cardholders from the American Express computer file and recorded their total charges over a specified period of time. For the selected cardholders, information was also obtained, through a mailed questionnaire, on the total number of miles traveled by each cardholder during the same period of time. The data for this study are given in Table 9–1. Figure 9–11 is a scatter plot of the data.

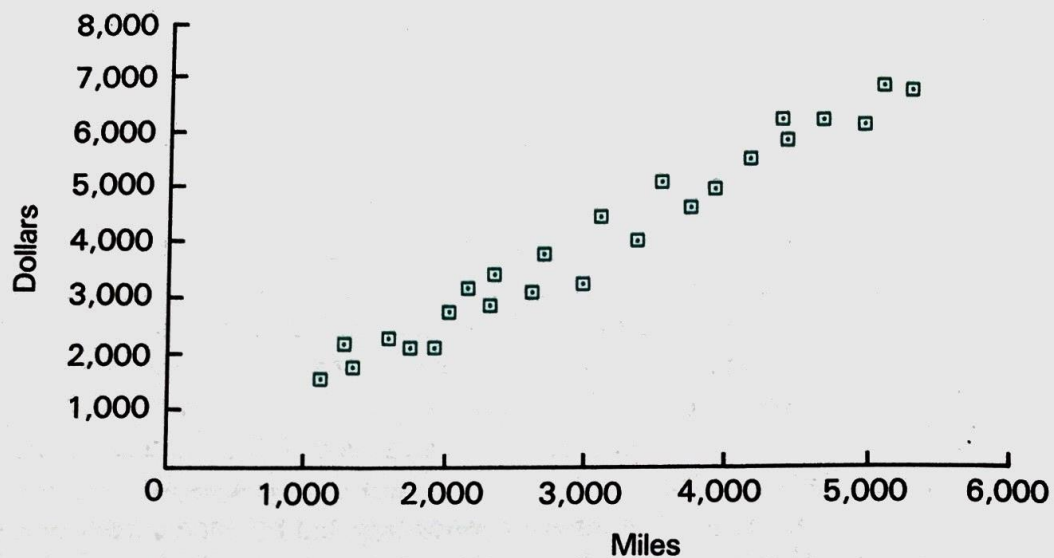
SOLUTION:

As can be seen from the figure, it seems likely that a straight line will describe the trend of increase in dollar amount charged with increase in number of miles traveled. The least-squares line that fits these data is shown in Figure 9–12.

TABLE 9–1 American Express Study Data

Miles	Dollars
1,211	1,802
1,345	2,405
1,422	2,005
1,687	2,511
1,849	2,332
2,026	2,305
2,133	3,016
2,253	3,385
2,400	3,090
2,468	3,694
2,699	3,371
2,806	3,998
3,082	3,555
3,209	4,692
3,466	4,244
3,643	5,298
3,852	4,801
4,033	5,147
4,267	5,738
4,498	6,420
4,533	6,059
4,804	6,426
5,090	6,321
5,233	7,026
5,439	6,964

FIGURE 9-11 The Data for the American Express Study



We will now show how the least squares regression line in Figure 9-12 is obtained. Table 9-2 shows the necessary computations. From Equations 9-9, using sums at the bottom of Table 9-2, we get:

$$SS_x = \sum x^2 - \frac{(\sum x)^2}{n} = 293,426,946 - \frac{79,448^2}{25} = 40,947,557.84$$

and

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 390,185,014 - \frac{(79,448)(106,605)}{25} = 51,402,852.4$$

FIGURE 9-12 Least-Squares Line for the American Express Study

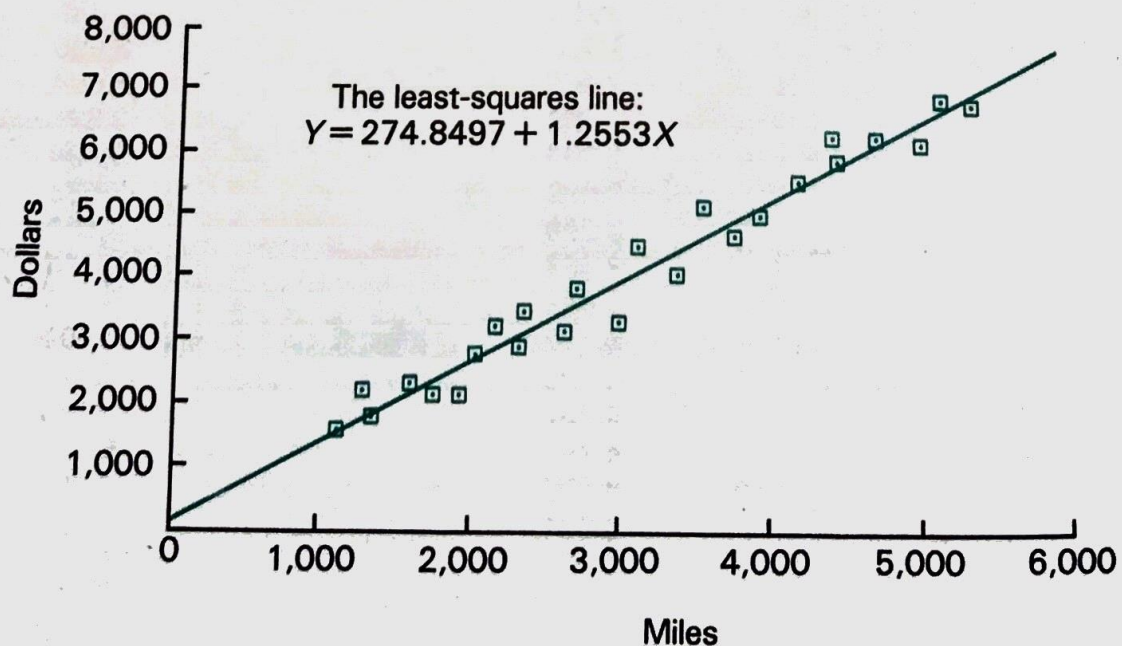


TABLE 9-2 The Computations Required for the American Express Study

Miles (X)	Dollars (Y)	X ²	Y ²	XY
1,211	1,802	1,466,521	3,247,204	2,182,222
1,345	2,405	1,809,025	5,784,025	3,234,725
1,422	2,005	2,022,084	4,020,025	2,851,110
1,687	2,511	2,845,969	6,305,121	4,236,057
1,849	2,332	3,418,801	5,438,224	4,311,868
2,026	2,305	4,104,676	5,313,025	4,669,930
2,133	3,016	4,549,689	9,096,256	6,433,128
2,253	3,385	5,076,009	11,458,225	7,626,405
2,400	3,090	5,760,000	9,548,100	7,416,000
2,468	3,694	6,091,024	13,645,636	9,116,792
2,699	3,371	7,284,601	11,363,641	9,098,329
2,806	3,998	7,873,636	15,984,004	11,218,388
3,082	3,555	9,498,724	12,638,025	10,956,510
3,209	4,692	10,297,681	22,014,864	15,056,628
3,466	4,244	12,013,156	18,011,536	14,709,704
3,643	5,298	13,271,449	28,068,804	19,300,614
3,852	4,801	14,837,904	23,049,601	18,493,452
4,033	5,147	16,265,089	26,491,609	20,757,851
4,267	5,738	18,207,289	32,924,644	24,484,046
4,498	6,420	20,232,004	41,216,400	28,877,160
4,533	6,059	20,548,089	36,711,481	27,465,447
4,804	6,426	23,078,416	41,293,476	30,870,504
5,090	6,321	25,908,100	39,955,041	32,173,890
5,233	7,026	27,384,289	49,364,676	36,767,058
5,439	6,964	29,582,721	48,497,296	37,877,196
<u>79,448</u>	<u>106,605</u>	<u>293,426,946</u>	<u>521,440,939</u>	<u>390,185,014</u>

Using Equations 9-10 for the least-squares estimates of the slope and intercept parameters, we get:

$$b_1 = \frac{SS_{XY}}{SS_X} = \frac{51,402,852.40}{40,947,557.84} = 1.255333776$$

and

$$b_0 = \bar{y} - b_1\bar{x} = \frac{106,605}{25} - (1.255333776)\frac{79,448}{25} = 274.8496866$$

It is important to carry out as many significant digits as you can in these computations. Here we carried out the computations by hand, for demonstration purposes. Usually, all computations are done by computer or by calculator. There are many hand calculators with a built-in routine for simple linear regression. From now on, we will present only the computed results, the least-squares estimates. The estimated least-squares relationship for Example (a) is, reporting estimates to the second significant decimal:

$$Y = 274.85 + 1.26X + e \quad (9-11)$$

The equation of the line itself, that is, the predicted value of Y for a given X , is:

$$\hat{Y} = 274.85 + 1.26X \quad (9-12)$$