

# Perform Linear Regression



Use appropriate tools strategically.

**QUESTION** How can you model data with the best-fitting line?

The line that most closely follows a trend in data is the *best-fitting line*. The process of finding the best-fitting line to model a set of data is called *linear regression*. This process can be tedious to perform by hand, but you can use a graphing calculator to make a scatter plot and perform linear regression on a data set.

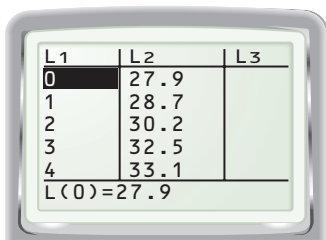
**EXAMPLE 1** Create a scatter plot

The table shows the total sales from women’s clothing stores in the United States from 1997 to 2002. Make a scatter plot of the data. Describe the correlation of the data.

Year	1997	1998	1999	2000	2001	2002
Sales (billions of dollars)	27.9	28.7	30.2	32.5	33.1	34.3

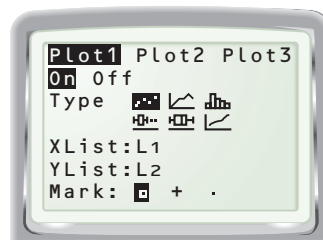
**STEP 1** Enter data

Press **STAT** and select Edit. Enter years since 1997 (0, 1, 2, 3, 4, 5) into List 1 ( $L_1$ ). These will be the  $x$ -values. Enter sales (in billions of dollars) into List 2 ( $L_2$ ). These will be the  $y$ -values.



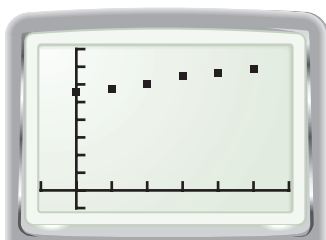
**STEP 2** Choose plot settings

Press **2nd** **Y=** and select Plot1. Turn Plot1 On. Select scatter plot as the type of display. Enter  $L_1$  for the Xlist and  $L_2$  for the Ylist.



**STEP 3** Make a scatter plot

Press **ZOOM** 9 to display the scatter plot so that the points for all data pairs are visible.



**STEP 4** Describe the correlation

Describe the correlation of the data in the scatter plot.

The data have a positive correlation. This means that with each passing year, the sales of women’s clothing tended to increase.

**MODELING DATA** The *correlation coefficient*  $r$  for a set of paired data measures how well the best-fitting line fits the data. You can use a graphing calculator to find a value for  $r$ .

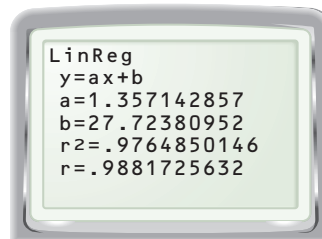
For  $r$  close to 1, the data have a strong positive correlation. For  $r$  close to  $-1$ , the data have a strong negative correlation. For  $r$  close to 0, the data have relatively no correlation.

**EXAMPLE 2** Find the best-fitting line

Find an equation of the best-fitting line for the scatter plot from Example 1. Determine the correlation coefficient of the data. Graph the best-fitting line.

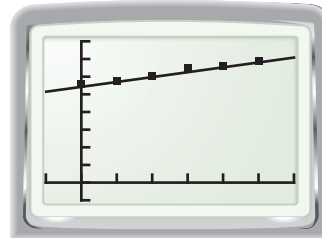
**STEP 1** Perform regression

Press **STAT**. From the CALC menu, choose LinReg(ax+b). The  $a$ - and  $b$ -values given are for an equation of the form  $y = ax + b$ . Rounding these values gives the equation  $y = 1.36x + 27.7$ . Because  $r$  is close to 1, the data have a strong positive correlation.



**STEP 2** Draw the best-fitting line

Press **Y=** and enter  $1.36x + 27.7$  for  $y_1$ . Press **GRAPH**.



**PRACTICE**

In Exercises 1–5, refer to the table, which shows the total sales from men's clothing stores in the United States from 1997 to 2002.

Year	1997	1998	1999	2000	2001	2002
Sales (billions of dollars)	10.1	10.6	10.5	10.8	10.3	9.9

1. Make a scatter plot of the data. *Describe* the correlation.
2. Find the equation of the best-fitting line for the data.
3. Draw the best-fitting line for the data.

**DRAW CONCLUSIONS**

4. What does the value of  $r$  for the equation in Exercise 2 tell you about the correlation of the data?
5. **PREDICT** How could you use the best-fitting line to predict future sales of men's clothing? *Explain* your answer.