

5-7

Scatter Plots and Trend Lines

OBJECTIVE: I can write an equation of a trend line and of a line of best fit to use a trend line and a line of best fit to make predictions



Warm-Up

The table shows the number of digital albums downloaded per year and the number of CDs sold by manufacturers per year. What relationship exists between the two sets of data? Predict the number of CDs sold and the number of albums downloaded in 2010. Explain your reasoning.

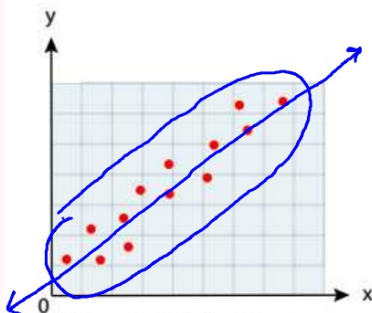
Year	Albums Downloaded (millions)	CDs Sold (millions)
2004	4.6	767
2005	13.6	705.4
2006	27.6	619.7
2007	42.5	511.1

Essential Understanding

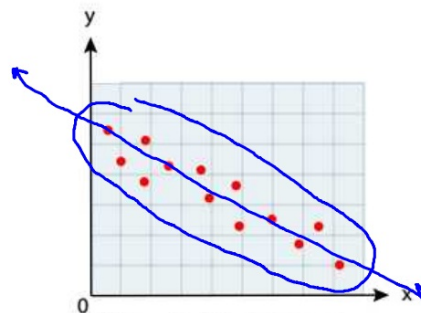
Essential Understanding You can determine whether two sets of numerical data are related by graphing them as ordered pairs. If the two sets of data are related, you may be able to use a line to estimate or predict values.

A scatter plot is a graph that relates two different sets of data by displaying them as ordered pairs. Most scatter plots are in the first quadrant of the coordinate plane because the data are usually positive numbers.

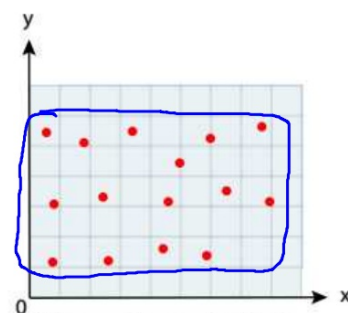
You can use scatter plots to find trends in data. The scatter plots below show the three types of relationships that two sets of data may have.



When y tends to increase as x increases, the two sets of data have a positive correlation.



When y tends to decrease as x increases, the two sets of data have a negative correlation.



When x and y are not related, the two sets of data have no correlation.



Example

#1 Making a Scatter Plot and Describing Its Correlation



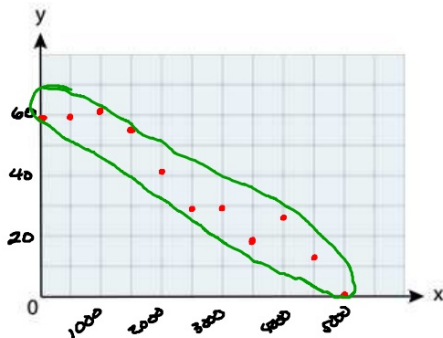
Temperature The table shows the altitude of an airplane and the temperature outside the plane.

Plane Altitude and Outside Temperature

Altitude (m)	0	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
Temperature (°F)	59.0	59.2	61.3	55.5	41.6	29.8	29.9	18.1	26.2	12.4	0.6

A Make a scatter plot of the data

B What type of relationship does the scatter plot show?



The temperature outside the plane tends to decrease as the altitude increases.

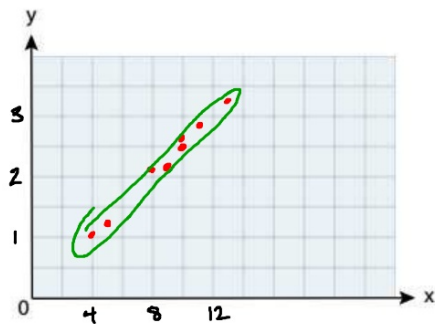
this graph has a negative correlation

Your Turn to Work it Out



1. Make a scatter plot of the data in the table below.
What type of relationship does the scatter plot show?

Gasoline Purchases								
Dollars Spent	10	11	9	10	13	5	8	4
Gallons Bought	2.5	2.8	2.3	2.6	3.3	1.3	2.2	1.1



The more money spent the more gallons were bought

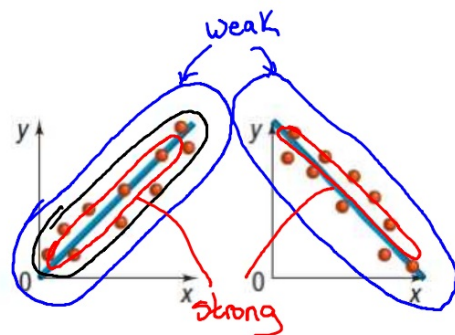
This is a positive correlation

Concept Understanding



When two sets of data have a positive or negative correlation, you can use a trend line to show the correlation more clearly. A trend line is a line on a scatter plot, drawn near the points, that shows a correlation.

You can use a trend line to estimate a value between two known data values or to predict a value outside the range of known data values. Interpolation is estimating a value between two known values. Extrapolation is predicting a value outside the range of known values.

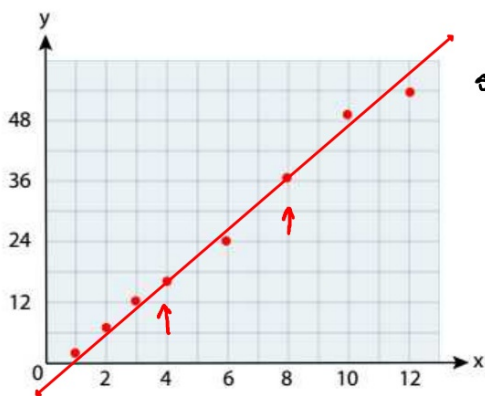


Example

#2 Writing an Equation of a Trend Line



Biology Make a scatter plot of the data at the right. What is the approximate weight of a 7-month-old panda?



Line of "Best fit"

STEP ① Find slope

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$m = \frac{37.6 - 17.9}{8 - 4}$$

$$m = \frac{19.7}{4}$$

$$m = 4.925$$

STEP ② Point slope

$$y - y_1 = m(x - x_1)$$

$$y - 17.9 = 4.925(x - 4)$$

$$y - 17.9 = 4.925x - 19.7$$

$$\begin{array}{r} y - 17.9 = 4.925x - 19.7 \\ +17.9 \qquad \qquad +17.9 \\ \hline y = 4.925x - 1.8 \end{array}$$

Age (months)	Weight (lb)
1	2.5
2	7.6
3	12.5
① 4	17.9
6	24.3
② 8	37.6
10	49.2
12	54.9

STEP ③ Calculate

$$y = 4.925x - 1.8$$

$$y = 4.925(7) - 1.8$$

$$y = 32.675 \text{ lbs}$$

Your Turn to Work it Out



2. Calculate a trend line and write its equation using the table below. What is the approximate body length of a 7-month-old panda?

Body Length of a Panda								
Age (months)	1	2	3	4	5	6	8	9
Body Length (in.)	8.0	11.75	15.5	16.7	20.1	22.2	26.5	29.0

STEP ① Find the slope of "two" points

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$m = \frac{20.1 - 15.5}{5 - 3}$$

$$m = \frac{4.6}{2}$$

$$m = 2.3$$

STEP ② Use point slope

$$y - y_1 = m(x - x_1)$$

$$y - 15.5 = 2.3(x - 3)$$

$$y - 15.5 = 2.3x - 6.9$$

$$\underline{\quad +15.5 \qquad \qquad +15.5 \quad}$$

$$y = 2.3x + 8.6$$

STEP ③ Calculate the panda's length

$$y = 2.3x + 8.6$$

$$y = 2.3(7) + 8.6$$

$$y = 24.7$$

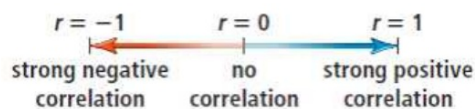
After 7 months the panda is 24.7"

Concept Understanding



The trend line that shows the relationship between two sets of data most accurately is called the line of best fit. A graphing calculator computes the equation of the **line of best fit** using a method called linear regression.

The graphing calculator also gives you the **correlation coefficient** r , a number from -1 to 1, that tells you how closely the equation models the data.



The nearer r is to 1 or -1, the more closely the data cluster around the line of best fit. If r is near 1, the data lie close to a line of best fit with positive slope. If r is near -1, the data lie close to a line of best fit with negative slope.

Example

#3 Finding the Line of Best Fit



Average Tuition and Fees
at Public 4-Year Colleges

Academic Year	Cost (\$)
2000–2001	3508
2001–2002	3766
2002–2003	4098
2003–2004	4645
2004–2005	5126
2005–2006	5492
2006–2007	5836



College Tuition Use a graphing calculator to find the equation of the line of best fit for the data at the right. What is the correlation coefficient to three decimal places? Predict the cost of attending in the 2025–2026 academic year.

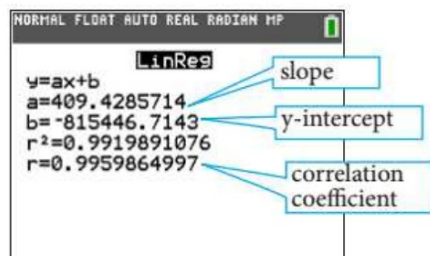
Step 1

Press LIST STAT . From the **EDIT** menu, choose **Edit**. Enter the years into L1.

- Let $x = 2000$ represent academic year 2000–2001,
- $x = 2001$ represent 2001–2002, and so on.
- Enter the costs into L2.

Step 2

Press LIST STAT . Choose **LinReg(ax + b)** from the **CALC** menu. Press enter to find the equation of the line of best fit and the correlation coefficient. The calculator uses the form $y = ax + b$ for the equation.



→ $y = 409.43x - 815,446.71$

Step 3 Predict the cost of attending in the 2025–2026 academic year.

$$y = 409.43x - 815,446.71$$

$$y = 409.43(2025) - 815,446.71$$

$$y \approx 13649.04$$

Your Turn to Work it Out



Entertainment Use a graphing calculator to find the equation of the line of best fit for the data in the table. Find the value of the correlation coefficient r to three decimal places. Then predict the number of movie tickets sold in the U.S. in 2026.

Movie Tickets Sold in U.S. by Year										
Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Tickets Sold (millions)	1289	1311	1340	1339	1406	1421	1470	1415	1472	1470