

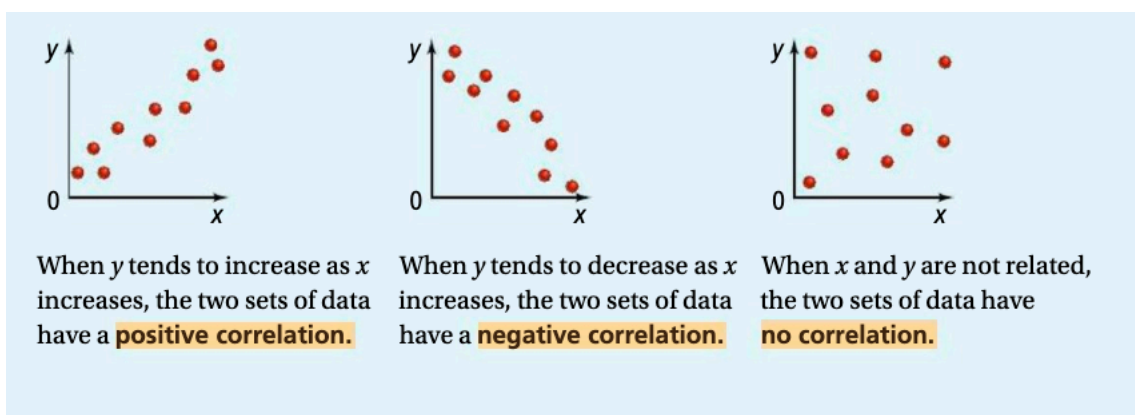
5-7

Scatter Plots and Trend Lines

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.



PROBLEM 1: MAKING A SCATTER PLOT AND DESCRIBING ITS CORRELATION

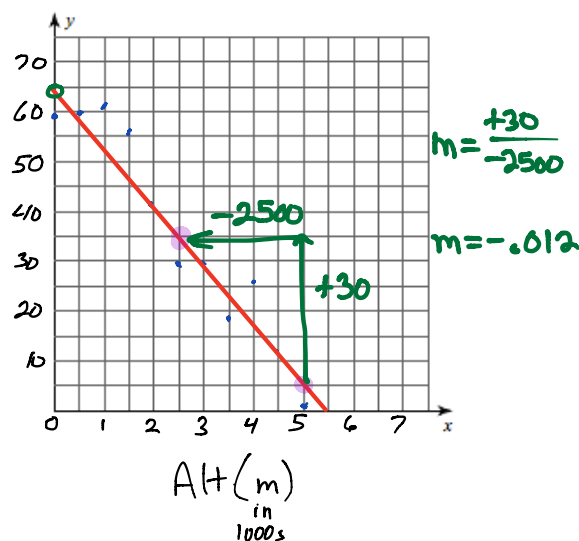
1. The table shows the altitude of an airplane and the temperature outside the plane.

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

- Make a scatter plot of the data.
- What type of relationship does the scatter plot show?

$$y = -.012x + 64.5$$

Temp
(°F)

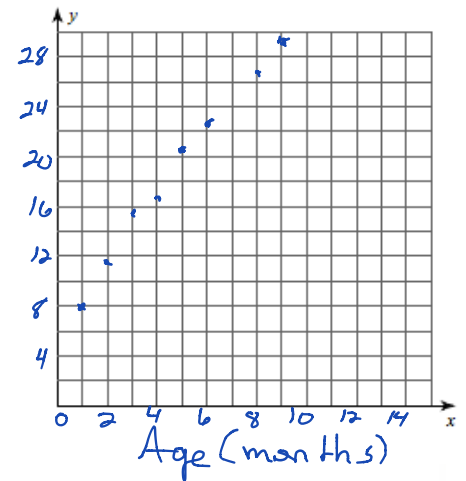


2. Make a scatter plot of the data in the table below. What type of relationship does the scatter plot show?

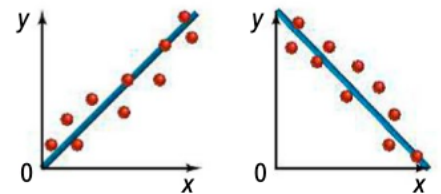
Body Length of a Panda								
Age (months)	1	2	3	4	5	6	8	9
Body Length (in.)	8.0	11.75	15.5	16.7	20.1	22.2	26.5	29.0

Positive correlation

Body Length (in)



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 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.

PROBLEM 2: WRITING AN EQUATION OF A TREND LINE

3. Make a scatter plot of the data pairs (year, attendance). Use a trend line to estimate the attendance at U.S. theme parks in 2019.

Attendance and Revenue at U.S. Theme Parks									
Year	1990	1992	1994	1996	1998	2000	2002	2004	2006
Attendance (millions)	253	267	267	290	300	317	324	328	335
Revenue (billions of dollars)	5.7	6.5	7.0	7.9	8.7	9.6	9.9	10.8	11.5

SOURCE: International Association of Amusement Parks and Attractions

$$y - 320 = 5.6(x - 102)$$

$$y - 320 = 5.6x - 571.2$$

Attendance (millions)

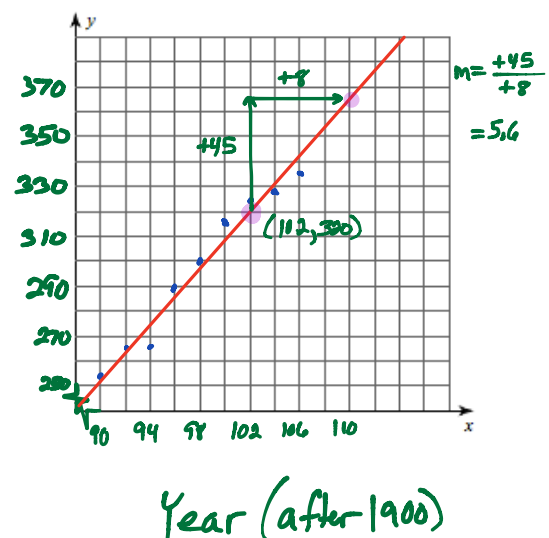
$$y = 5.6x - 251.2$$

2019: $x = 119$;

$$y = 5.6(119) - 251.2$$

$$= 415.2$$

415.2 million



4. Make a scatter plot of the data below. Draw a trend line and write its equation. Use the equation to approximate the body length of a 7-month old panda. 2019?

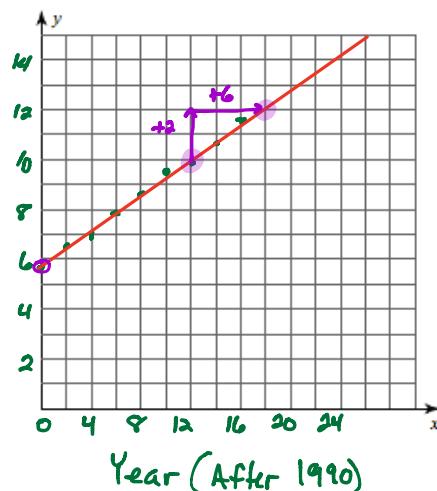
Attendance and Revenue at U.S. Theme Parks									
Year	1990	1992	1994	1996	1998	2000	2002	2004	2006
Attendance (millions)	253	267	267	290	300	317	324	328	335
Revenue (billions of dollars)	5.7	6.5	7.0	7.9	8.7	9.6	9.9	10.8	11.5

SOURCE: International Association of Amusement Parks and Attractions

$$m = \frac{+2}{+6} = .33$$

$$b = 5.8$$

Revenue
(in \$billions)



$$y = .33x + 5.8$$

2019:

$$x = 19,$$

$$y = .33(19) + 5.8 = 12.07$$

\$12.07 billion

5. Do you think you can use your model to extrapolate the revenue in 2050?

It would not be reliable to predict that far away from the data in the table.

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

The technology 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.

PROBLEM 3: FINDING THE LINE OF BEST FIT

6. Use an online linear regression calculator to find the equation of the line of best fit for the data in the table. Predict the cost of attending in the 2019-2020 academic year.

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

SOURCE: The College Board

Causation is when a change in one quantity causes a change in a second quantity. A correlation between quantities does not always imply causation.

PROBLEM 4: IDENTIFYING WHETHER RELATIONSHIPS ARE CAUSAL

In the following situations, is there likely to be a correlation? If so, does the correlation reflect a causal relationship?

7. the number of loaves of bread baked and the amount of flour used

causal

8. the number of mailboxes and the number of firefighters in a city

not causal

9. the cost of a family's vacation and the size of their house

not causal

10. the time spent exercising and the number of Calories burned

causal



Lesson Check

Do you know HOW?

Use the table.

Average Maximum Daily Temperature
in January for Northern Latitudes

Latitude (° N)	35	33	30	25	43	40	39
Temperature (°F)	46	52	67	76	32	37	44

SOURCE: U.S. Department of Commerce

1. Make a scatter plot of the data. What type of relationship does the scatter plot show?
2. Draw a trend line and write its equation.
3. Predict the average maximum daily temperature in January at a latitude of 50° N.

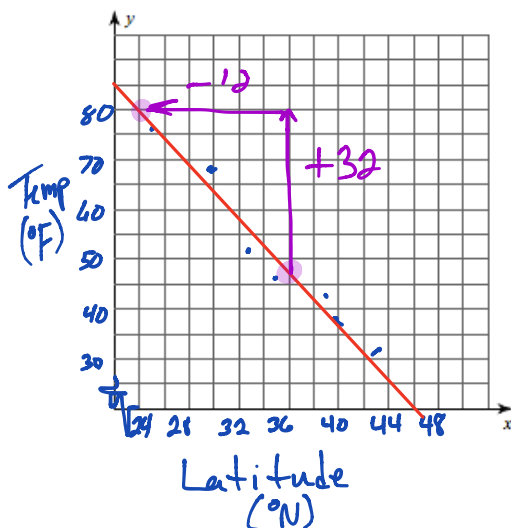
Do you UNDERSTAND?



MATHEMATICAL
PRACTICES

4. **Vocabulary** Given a set of data pairs, how would you decide whether to use interpolation or extrapolation to find a certain value?
5. **Compare and Contrast** How are a trend line and the line of best fit for a set of data pairs similar? How are they different?
6. **Error Analysis** Refer to the table below. A student says that the data have a negative correlation because as x decreases, y also decreases. What is the student's error?

x	10	7	5	4	1	0
y	1	0	-2	-4	-7	-9



$$m = \frac{+32}{-12} = -2.66$$

(24, 80)

$$y - 80 = -2.66(x - 24)$$

$$y - 80 = -2.66x + 63.84$$

$$y = -2.66x + 143.84$$

$$x = 50$$

$$y = -2.66(50) + 143.84$$

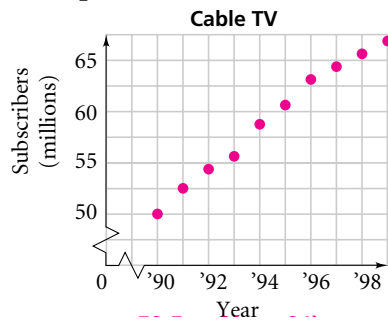
$$= 10.84$$

$$10.84^{\circ}\text{F}$$

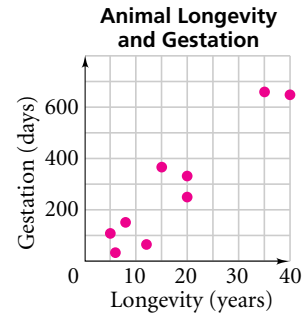
5-7 Practice Worksheet

Find an equation of a reasonable trend line for each scatter plot.

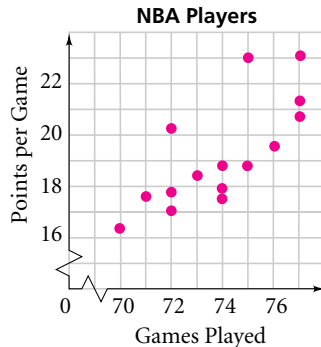
1.



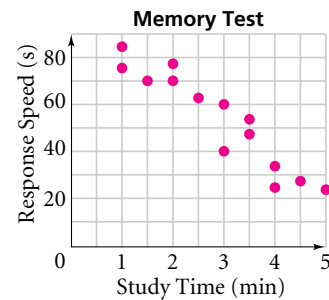
2.



3.

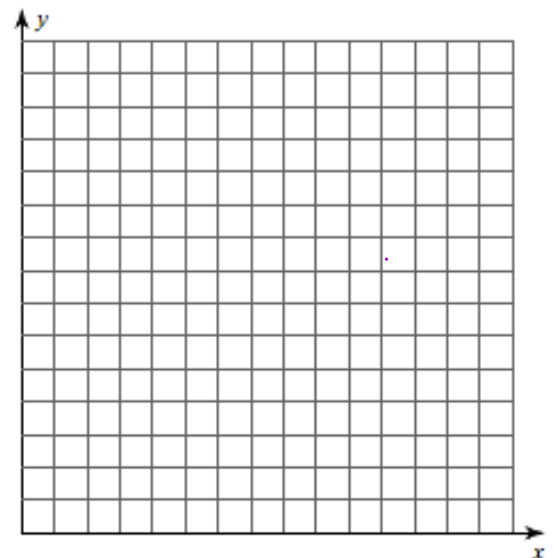


4.



5. Graph the data in the table below for the body length of a panda. Find an equation for a trend line of the data. Use the equation to approximate the length of a 11-month-old panda.

Body Length of a Panda								
Age (month)	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



Do you think you could use your trend line to approximate the length of a 3-year-old panda?

6. Graph the data for the average July temperature and the annual precipitation of the cities in the table below.

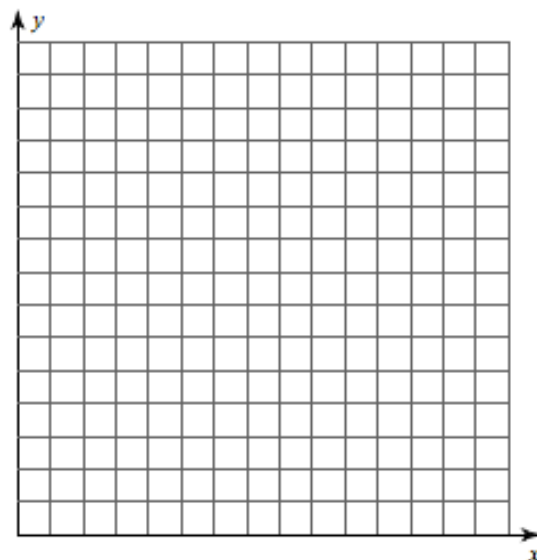
a) Find an equation for the line of best fit of the data.

b) Estimate the average rainfall for a city with average July temperatures of 75° F.

Precipitation and Temperature in Selected Eastern Cities

City	Average July Temperature ($^{\circ}$ F)	Average Annual Precipitation (in.)
New York	76.4	42.82
Baltimore	76.8	41.84
Atlanta	78.6	48.61
Jacksonville	81.3	52.76
Washington, D.C.	78.9	39.00
Boston	73.5	43.81
Miami	82.5	57.55

SOURCE: Time Almanac



TECHNOLOGY

7. Use a graphing calculator or online linear regression calculator to find the equation of the line of best fit for the data below. Predict sales of greeting cards in the year 2019.

Greeting Card Sales

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Sales (billions)	\$4.2	\$4.6	\$5.0	\$5.3	\$5.6	\$5.9	\$6.3	\$6.8	\$7.3	\$7.5

SOURCE: Greeting Card Association