Write

Complete each sentence with the appropriate vocabulary term.

1. A _____ can be used to organize and display the values of two variables in a data set.

2. _____ models the relationship between two variables in a data set by producing a line of best fit.

3. A(n) _____ is a line that best approximates the linear relationship between two variables in a data set.

4. The _____ is used to approximate a line of best fit by minimizing the squares of the distances of the points from the line.

5. _____ is using a linear regression to make predictions within the data set.

6. Using a linear regression to make predictions outside of the data set is _____.

7. After a scatter plot is created, the _____ is a point with an *x*-value that is the mean of all the *x*-values of the points on the plot and a *y*-value that is the mean of all the *y*-values of the points on the plot.
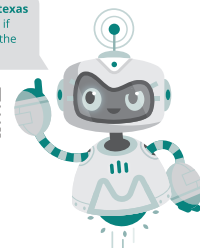
Remember

Patterns in data can be modeled with lines of best fit. The Least Squares Method is one way to create a linear regression equation, and it is the method that graphing calculators tend to use.

Practice

Visit **livehint.com/texas** or use this QR code if you need a hint on the Practice questions.

1. One of the jobs of the National Center for Education Statistics is to gather information about public high schools and their dropout rates. This includes anyone who leaves school without a high school diploma or an equivalent credential. The table shows the average percent of high school dropouts from the year 2002 through the year 2014.

   a. Create a scatter plot of the high school dropout data. What information can you gather about the dropout rates from the scatter plot?

   b. Use the data table and graphing technology to generate a line of best fit.

   c. Interpret the slope and *y*-intercept of the linear regression equation. What do these values represent in terms of the problem situation?

   d. Determine the dropout rate for the year 2010. Is this the same as the dropout rate recorded in the table? If not, explain the difference.

| Year	High School Dropout Rate (percent)
2002	3.5
2003	4.0
2004	4.7
2005	3.8
2006	3.8
2007	3.5
2008	3.5
2009	3.4
2010	3.0
2011	3.4
2012	3.4
2013	4.7
2014	5.2

2. Mr. Li is a math teacher at Pinkston High School and is preparing his students to take the SAT test. He collected data from 10 students who took the test last year and presented this information to the students in a table. The highest math SAT score a student can achieve is 800. Analyze the data in the table.

| Time Spent Studying (hours)	Math SAT Score
1	350
22	780
12	600
14	700
4	380
10	650
9	580
3	400
7	530
4	410

   a. Construct a scatter plot of the data and describe any patterns you see in the data.
   b. Determine the equation of a line passing through (14, 700) and (7, 530). Then determine the equation of a line passing through (22, 780) and (4, 380). Graph both lines on the same graph as the scatter plot.
   c. Which line seems to best fit the data? Would you use either one of these lines to make predictions about a student's math SAT score based on the amount of studying they do? Why or why not?
   d. Use graphing technology to determine the regression equation.
   e. Interpret the least squares regression equation in terms of the problem situation.
   f. Use the regression line to predict the math SAT score for a student who studies for 17 hours. Did you use interpolation or extrapolation to make this prediction? Is this prediction reasonable for this problem situation? Explain your reasoning.
   g. Use the line of best fit to predict the math SAT score for a student who studies for 40 hours. Did you use interpolation or extrapolation to make this prediction? Is this prediction reasonable for this problem situation? Explain your reasoning.
   h. One of Mr. Li's students comes back to him the following year and says that he studied for 15 hours for the math SAT and got a score of 610. He argues that the equation predicted that he would have scored a 682. What do you think explains the discrepancy?

Stretch

Consider the two sets of data shown in the graphs.
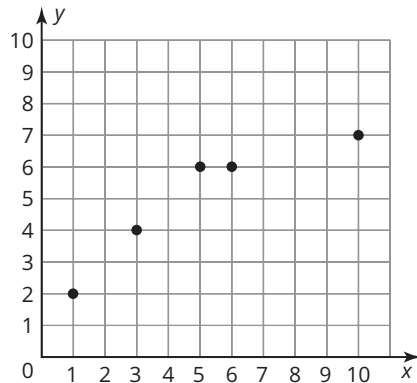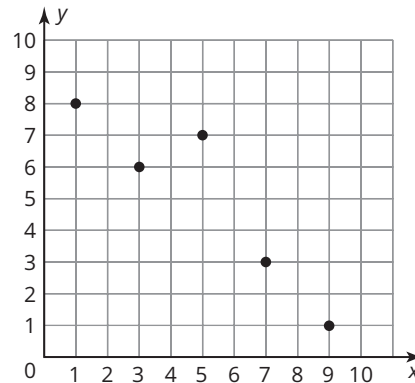
Graph A　　　　　　　　　　　**Graph B**

1. Calculate the mean of the x-values, $\bar{x}$, and the mean of the y-values, $\bar{y}$, for each graph.
2. Complete the tables for each graph.

Graph A				
x	$x - \bar{x}$	y	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
1	−4	2	−3	12
3		4		
5		6		
6		6		
10		7		
				SUM =

Graph B				
x	$x - \bar{x}$	y	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
1	−4	8	3	−12
3		6		
5		7		
7		3		
9		1		
				SUM =

3. Compare the two sums in the last column of each table. Determine if there seems to be a connection between the sums and the graphs of the data set.

Review

1. A maintenance worker in a factory notices that a water tank is leaking. She records the amount of water in the tank each day in a table.

Day	Volume of Water (L)
1	16,000
2	12,000
3	9000
4	6750

   a. Write a recursive formula to represent the pattern shown in the table. What predictions does this formula make for the amount of water in the tank on the 5th day?

   b. Write an explicit formula to represent the pattern shown in the table. What predictions does this formula make for the amount of water in the tank on the 10th day?

2. The graph represents a linear relationship between x and y.

   a. Describe whether the graph is increasing or decreasing. Justify your reasoning.

   b. Determine the x- and y-intercept.