

Statistics and Probability

In eighth grade, students build on their previous statistics and probability understanding to work with **bivariate data**, which, in mathematics, means two-variable data. The data is commonly represented using a scatter plot that helps reveal whether there might be a relationship between the variables.

Students will construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. They will know that straight lines are widely used to model relationships between two quantitative variables. They will use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept.

Students also will understand that patterns of association also can be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Additionally, they will construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects.

The Grade 8 Common Core State Standards for Statistics and Probability specify that students should—

- Investigate patterns of association in bivariate data.

The following hands-on activities will enable students to experience opportunities to become more familiar with how bivariate data is collected, analyzed, and interpreted. Teachers will want to coach students to apply what they know about data collection, analysis, and interpretation to the mathematical and real-world problems they solve.

Mathematically proficient students apply mathematical concepts to solve problems in daily life. They are able to identify important quantities in a practical situation and map their relationships using tools such as tables and graphs. They can analyze these relationships mathematically to draw conclusions. They also can interpret their results in the context of the situation and reflect on whether their results make sense. Teachers will want to use teachable moments to make connections with what students are learning to do with bivariate data in relation to what others in the “real world” do with such data.

Statistics and Probability

Contents

Lesson 1 Scatter Plot Diagrams 112

Objective: Graph a set of data and understand the basics of scatter plots.

Manipulative: XY Coordinate Pegboard

Lesson 2 Line of Best Fit 116

Objective: Determine the line of best fit for a set of experimental data.

Manipulative: XY Coordinate Pegboard

Lesson 3 Make a Conjecture Using a Scatter Plot 120

Objective: Create and use a scatter plot to make a conjecture.

Manipulative: XY Coordinate Pegboard

Statistics and Probability

Scatter Plot Diagrams

Objective

Graph a set of data and understand the basics of scatter plots.

Common Core State Standards

- **8.SP.1** Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

Try It! Perform the Try It! activity on the next page.

Talk About It

Discuss the Try It! activity.

- **Say:** If one variable gets larger as the other variable gets larger, the two variables have a positive correlation. The graph will appear to go “uphill” from the y-axis. If one variable gets smaller as the other gets larger, the two variables have a negative correlation. The graph will appear to go “downhill” from the y-axis. If you can’t determine which way the data point, or even that there is any trend, the graph shows no correlation.
- **Ask:** What does the first scatter plot show? Elicit that it shows a positive correlation between the number of hours the team has practiced and the number of hits they get. **Ask:** What does the second scatter plot show?

Solve It

Reread the problem with students. After students determine that there is a correlation between the scattered coordinate pairs, have them estimate values greater than $x = 14$ that could reasonably appear on each graph.

More Ideas

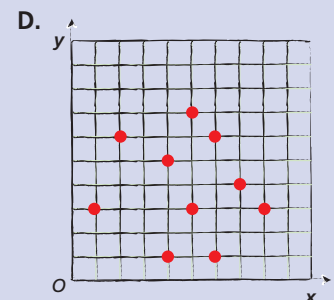
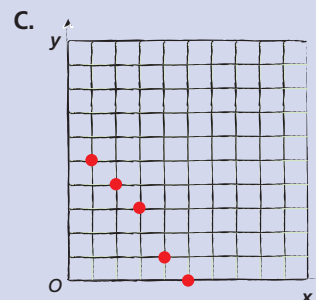
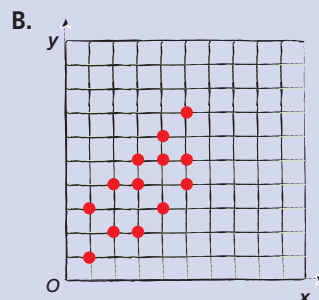
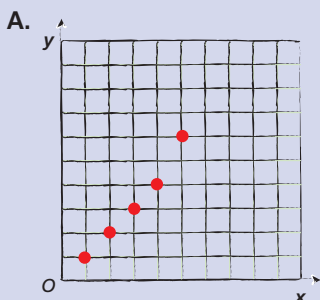
For another way to teach about scatter plots—

- Have students repeat the exercise using the XY Coordinate Pegboard and statistics from their favorite sports team.

Formative Assessment

Have students try the following problem.

Which graph shows no correlation?



Try It! 20 minutes | Pairs

Here is a problem about scatter plots and correlations.

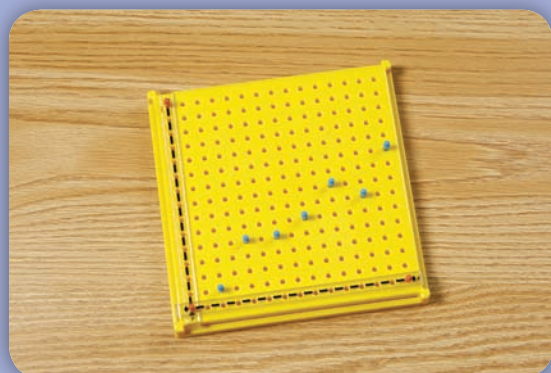
The coach of the softball team is reviewing some statistics for the first seven games. Is there a positive correlation, a negative correlation, or no correlation between hours of practice and the number of hits the team got, or between hours of practice and the number of errors the players made?

Hours of Practice	2	4	6	8	10	12	14
Number of Hits	1	4	4	5	7	6	9
Number of Errors	8	6	5	5	3	2	2

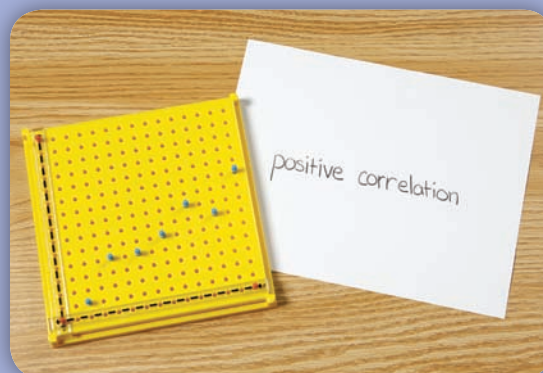
Introduce the problem. Then have students do the activity to solve the problem. Distribute the materials.

Materials

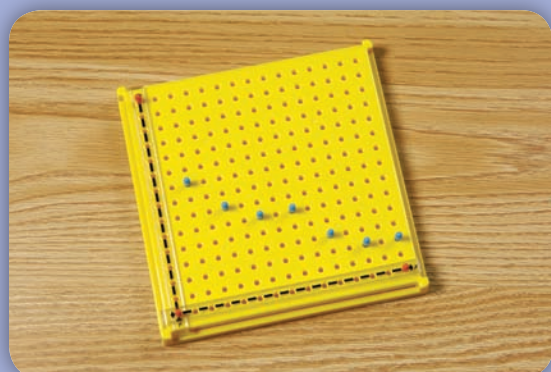
- XY Coordinate Pegboard



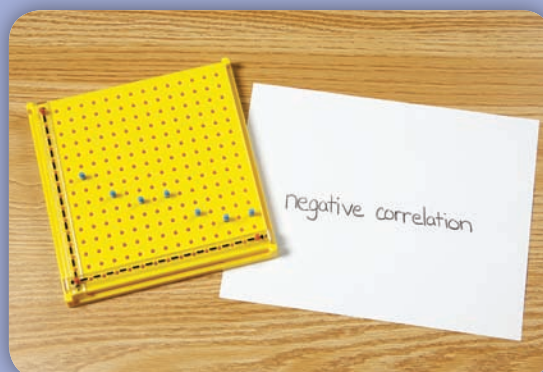
1. Have students plot the coordinate pairs for Hours of Practice (x) and Number of Hits (y) onto the pegboard.



2. Have students examine the graph.
Ask: Which way do the data points go? Do the data show a positive correlation, a negative correlation, or no correlation?



3. Have students plot the coordinate pairs for Hours of Practice (x) and Number of Errors (y).



4. Have students examine the graph.
Ask: Do the data show a positive correlation, a negative correlation, or no correlation?

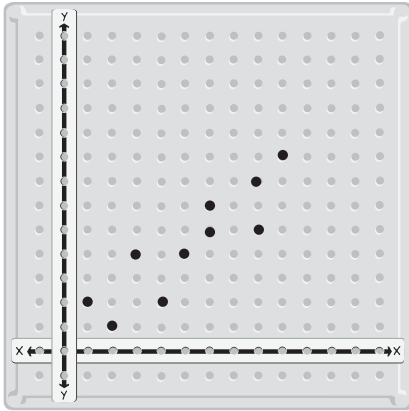
Look Out!

Some students may think that the coordinate pairs should form a perfect line. Reassure them that the points do not need to do this in order to show a correlation.

Use an XY Coordinate Pegboard to model the scatter plot shown. Write the ordered pairs that are graphed. Identify a correlation as positive or negative. Explain your answer.

(Check students' work.)

1.



x	1	2	3	4	5	6	6	8	8	9
y	2	1	4	2	4	5	6	5	7	8

This scatter plot shows a positive correlation.

As the x-values increase, the y-values increase.

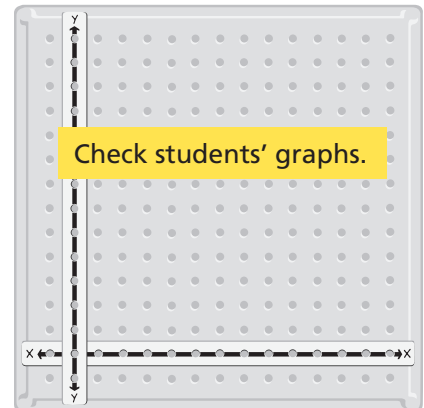
Using an XY Coordinate Pegboard, model the data given in the table. Does the scatter plot show a positive or negative correlation? Explain.

2.

x	1	2	3	4	5	7	9
y	8	5	5	6	3	2	1

This scatter plot shows a negative correlation.

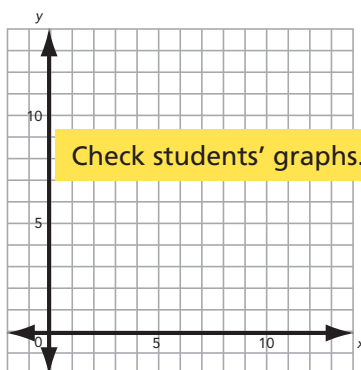
As the x-values increase, the y-values decrease.



Graph each set of ordered pairs. What type of correlation is shown?

3.

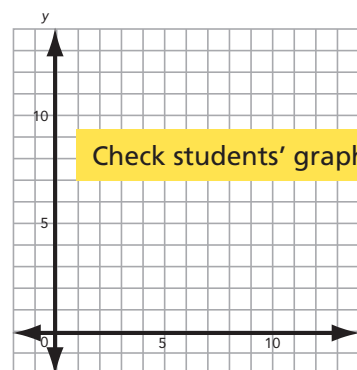
x	1	3	5	6	10	12
y	13	11	6	4	3	1



negative

4.

x	4	5	8	9	11	12
y	3	3	8	9	10	14



positive

Answer Key

Challenge! Describe the appearance of a scatter plot that shows a negative correlation. What does a negative correlation mean in terms of the x- and y-values? Describe the appearance of a scatter plot that shows a positive correlation. What does a positive correlation mean in terms of the x- and y-values?

Challenge: (Sample) A scatter plot that shows a negative correlation will have points that generally begin on the left of the graph and get closer to the x-axis as the x-values increase. The meaning of a negative correlation is that as the x-values increase, the y-values decrease. A scatter plot that shows a positive correlation will have points that generally begin on the left of the graph and get farther from the x-axis as the x-values increase. The meaning of a positive correlation is that as the x-values increase, the y-values increase.

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and extend across the width of the page. There are no margins, text, or other markings on the paper.

Objective

Determine the line of best fit for a set of experimental data.

Common Core State Standards

- **8.SP.1** Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
- **8.SP.2** Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.
- **8.SP.3** Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.*

Statistics and Probability

Line of Best Fit

Previously, students learned how to graph a scatter plot diagram. From their diagram, they were able to determine that the coordinate pairs may have a linear relationship, even though the pairs did not form a perfect line. In this lesson, students will be introduced to the term “line of best fit.” In future mathematics courses, students may hear this line referred to as “the regression line” or “the least squares line.”

Try It! Perform the Try It! activity on the next page.

Talk About It

Discuss the Try It! activity.

- **Ask:** Does the line show the general direction of the data? Are there approximately an equal number of pegs (points) above the line as below? What is the slope of the line? What does it mean? What is the y-intercept? What does it mean?
- **Say:** Scatter plots may show an association between two variables. **Ask:** What association does this scatter plot suggest? **Say:** It’s important to know that an association is not the same as a cause-and-effect relationship, however. There may be other factors involved that we are not aware of.

Solve It

Reread the problem with students. Have students use their knowledge of graphing to determine the line of best fit. Have them estimate values greater than $x = 7$ that might be on their graph.

More Ideas

For another way to teach about the line of best fit—

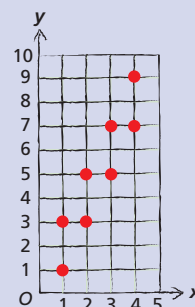
- Have students find scatter plots in the media (newspapers, magazines, online). Have them draw the line of best fit and write an equation that expresses it in y-intercept form ($y + mx = b$).

Formative Assessment

Have students try the following problem.

Which is the line of best fit?

- A. $y = x + 2$
- B. $y = 2x$
- C. $y = x + 1$
- D. $y = 2x + 1$



Try It! 30 minutes | Pairs

Here is a problem about the line of best fit.

Widget Works is a shop that produces handmade custom widgets. The table shows the number of hours worked and the number of widgets produced by ten skilled craftspeople.

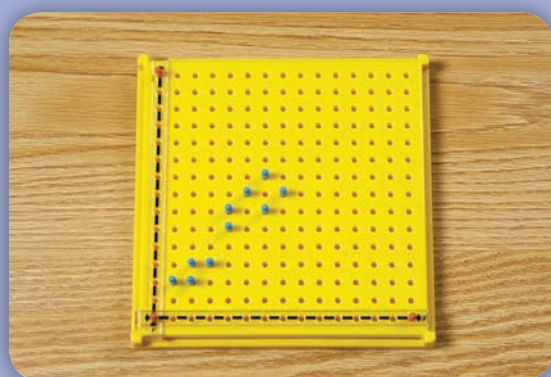
Craftsperson	1	2	3	4	5	6	7	8	9	10
x-Hours Worked	2	7	1	5	2	6	3	4	6	4
y-Widgets Produced	3	7	2	7	2	8	3	5	6	6

Determine the line of best fit that relates to the number of hours worked and the number of widgets produced. Write the equation for the line in y-intercept form ($y = mx + b$).

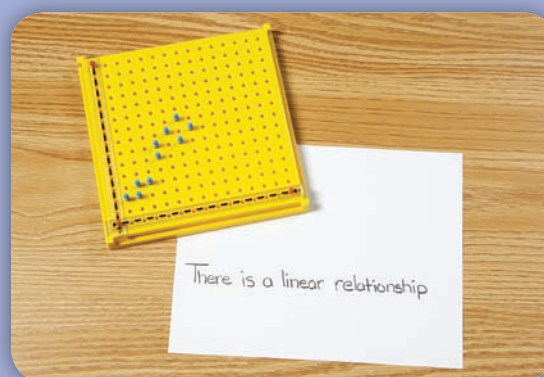
Introduce the problem. Then have students do the activity to solve the problem. Distribute the materials.

Materials

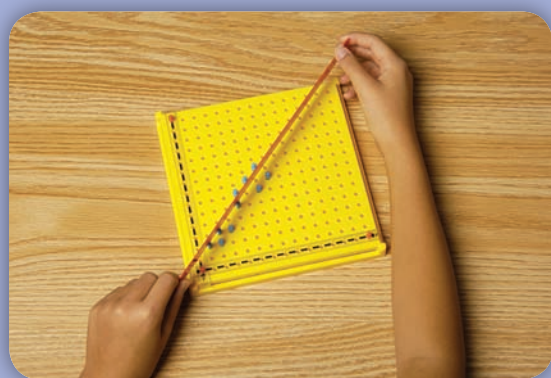
- XY Coordinate Pegboard
- ruler or straightedge



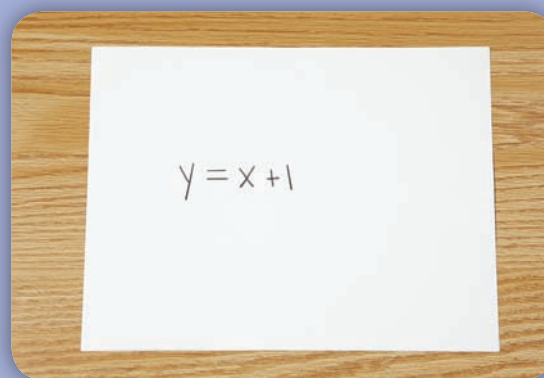
1. Have students peg the coordinate pairs on their XY Coordinate Pegboard.



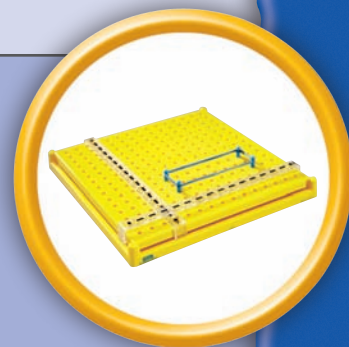
2. Have students determine whether or not there is a linear relationship between the coordinate values. **Ask:** Can you draw a straight line that will lie near all of the data points?



3. Have students take a ruler or straightedge and see if they can find a "line of best fit." The line should "split" the pegs, leaving the same number of pegs above the line as below it (or as close to this as possible).



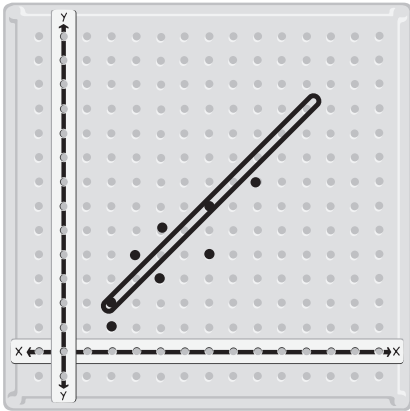
4. Have students use the slope and the y-intercept to write the equation for this line of best fit. (The y-intercept should be estimated to the nearest whole number.)



Use an XY Coordinate Pegboard to model the scatter plot shown. Write the equation for the line of best fit in the form $y = mx + b$.

(Check students' work.)

1.



If the line extended to the y-axis, where would it intersect? 0

What is the slope of the line? 1

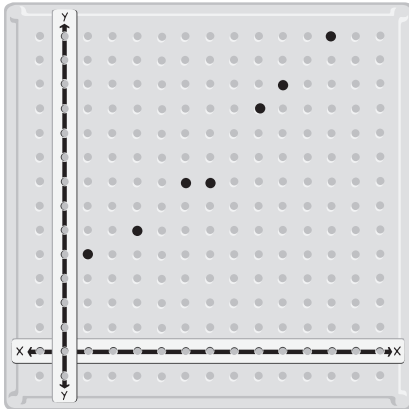
The equation for the line of best fit is

$y = x$

Using an XY Coordinate Pegboard, model the scatter plot shown. Write the equation for the line of best fit in the form $y = mx + b$.

(Check students' models.)

2.



For the line of best fit, $b =$ 2

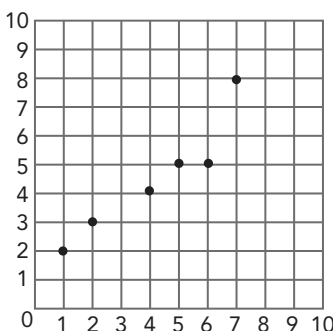
For the line of best fit, $m =$ 1

The equation for the line of best fit is

$y = x + 2$

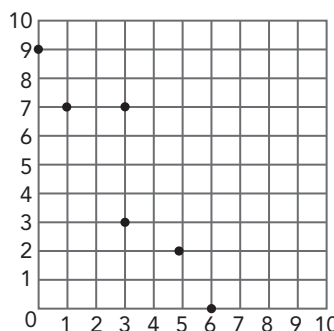
Write the equation for the line of best fit for each scatter plot.

3.



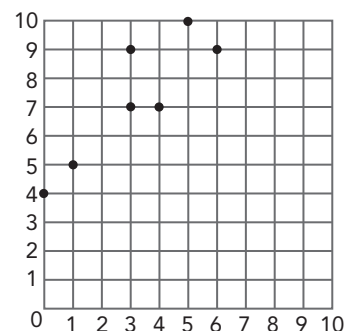
$y = x + 1$

4.



$y = -2x + 9$

5.



$y = x + 4$

Answer Key

Challenge! How can you tell from looking at the points in a scatter plot if the line of best fit has a positive or negative slope? Draw a picture to help.

Challenge: (Sample) If the points tend to move higher up the grid as they go from left to right, then the line of best fit has a positive slope. If the points tend to move lower down the grid as they go from left to right, then the line of best fit has a negative slope.

[illegible]

LESSON 3

Objective

Create and use a scatter plot to make a conjecture.

Common Core State Standards

- **8.SP.1** Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
- **8.SP.2** Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.

Statistics and Probability

Make a Conjecture Using a Scatter Plot

Students apply the skills of organizing and representing data to examine relationships between characteristics in a population. Scatter plots are used to test for trends in data—that is, to test for correlations between characteristics. The lesson prepares students to explore other relationships in or between populations.

Try It! Perform the Try It! activity on the next page.

Talk About It

Discuss the Try It! activity.

- **Ask:** Will Keenan's data be appropriate for all students in the class? Have students explain their answers.
- **Ask:** Why might more than one point have the same x-coordinate? The same y-coordinate?
- Discuss the positive trend shown by the scatter plot. **Ask:** How would a scatter plot show a negative trend? How would it show no trend?

Solve It

Reread the problem with students. Have them discuss the trend in the data. Ask students to make a conjecture about how quiz score depends on hours of study. Stress, however, that a trend does not prove a cause-and-effect relationship.

More Ideas

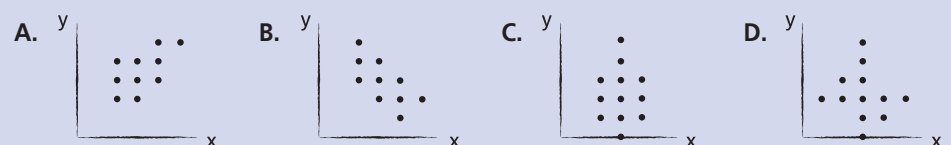
For other ways to teach about scatter plots—

- Extend this problem by having students create tables of values from their lines of best fit. Have them use their tables to predict scores based on number of hours of study.
- Encourage students to use real data, such as baseball or basketball statistics from their local newspaper, to create a scatter plot. Have students plot the data on a pegboard. Then have them transfer each scatter plot to graph paper and draw a line of best fit.

Formative Assessment

Have students try the following problem.

Which scatter plot shows a negative trend?



Try It! 20 minutes | Pairs

Here is a problem about making and using a scatter plot.

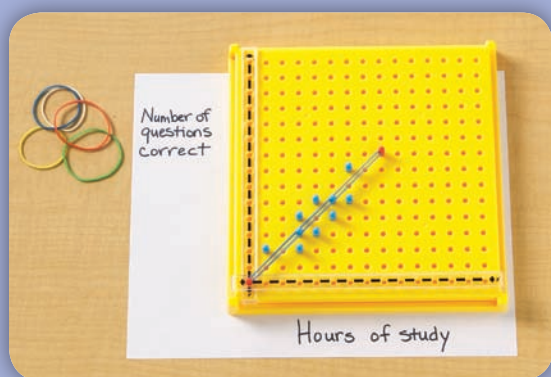
Keenan wants to see if there is a relationship between the number of hours a classmate studies and the number of questions he or she gets correct on a quiz. He takes a sample of ten classmates and makes the table shown below. Does there appear to be a relationship between the number of hours of study and the number of questions answered correctly?

Hours of Study	3	6	1	4	3	3	5	4	5	6
Number of Questions Correct	4	7	2	5	3	2	5	3	4	5

Introduce the problem. Then have students do the activity to solve the problem. Distribute the materials.



1. Ask students to set up their pegboards for Quadrant I graphing and have them plot the data from the table. **Say:** Let the x-axis represent the hours of study and let the y-axis represent the number of questions correct.



3. Discuss how to form the line of best fit through the data. Tell students that they should make a straight line, using a rubber band, such that there are about the same number of data points above the line as there are below it, and such that the points above the line are the same average distance from the line as the points below. **Ask:** Which coordinates does your line pass through?

Materials

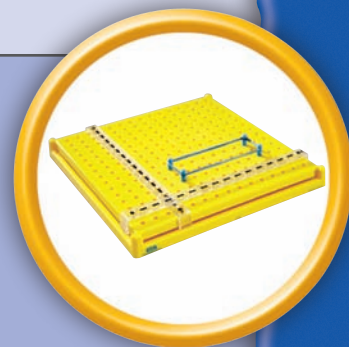
- XY Coordinate Pegboard (1 per pair)
- paper (1 sheet per pair)
- pencils (1 per pair)



2. Introduce the term *scatter plot*. **Say:** A scatter plot is a plot of data points. If the data points rise from left to right, then they show a positive trend. If the data points fall from left to right, then they show a negative trend. If the data points neither rise nor fall, then they show no trend. **Ask:** Do you see a trend? What kind? What conjecture would you make from the data? Elicit that the data show a positive trend between the number of hours of study and the number of questions answered correctly.

Look Out!

Some students might connect the points to make a line graph. Increase the data size, if necessary, to demonstrate that it is not meaningful to connect all the points. Point out that the line of best fit only helps to show the trend in the data—nothing more.

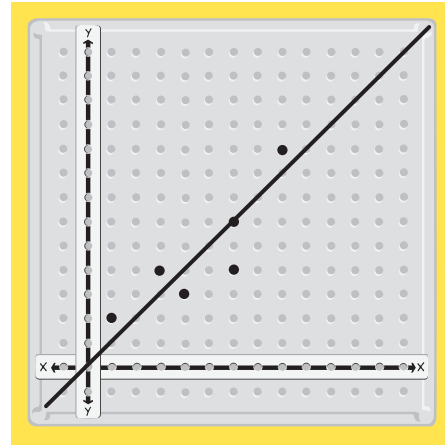


Use an XY Coordinate Pegboard to plot the points and create a scatter plot. Find the line of best fit on your pegboard. Sketch the line of best fit.

(Check students' work.)

1.

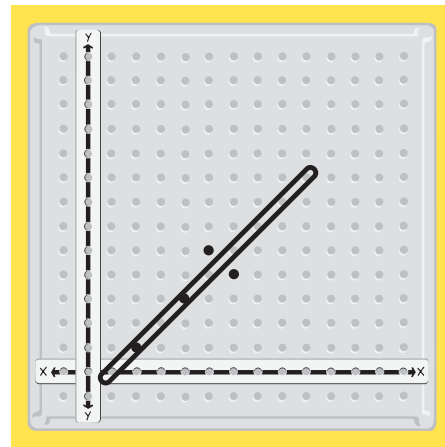
x	y
1	2
3	4
4	3
6	4
6	6
8	9



Using an XY Coordinate Pegboard, plot the data in the table. Find the line of best fit. Make predictions to complete the table.

2.

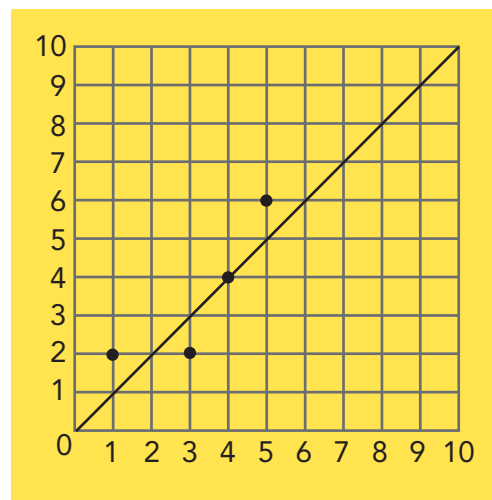
x	y
2	1
4	3
5	5
6	4
7	6
8	7



Make a scatter plot of the data. Find the line of best fit. Make predictions to complete the table.

3.

x	y
1	2
3	2
4	4
5	6
8	8
10	10



Answer Key

Challenge! How do you decide where to place the line of best fit? Can there be more than one line of best fit? Explain.

Challenge: (Sample) The line is placed so as many data points as possible are as close to it as possible. While there may be several lines that are good fits, there is only one line of best fit.

[illegible]