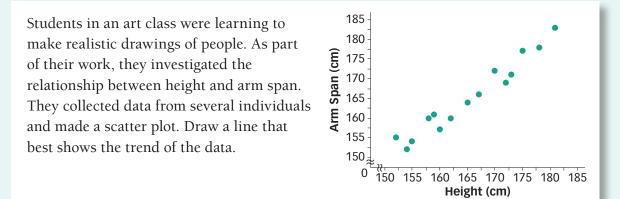# Lesson 29   Part 1: Introduction

Scatter Plots and Linear Models

In the previous lesson, you learned how to analyze scatter plots. In this lesson, you will explore more tools that can help with scatter-plot analysis.

Students in an art class were learning to make realistic drawings of people. As part of their work, they investigated the relationship between height and arm span. They collected data from several individuals and made a scatter plot. Draw a line that best shows the trend of the data.

🔍 Explore It

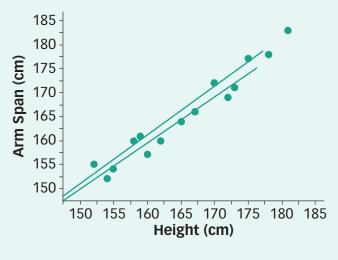**Use the math you already know to solve the problem.**

◻ Sketch a line on the scatter plot that you think captures the trend of the data.

_____

_____

◻ Explain how you decided where to place your line.

_____

◻ Which points seem to be closest to the line? Which seem to be farthest?

_____

◻ Take another look at the line you sketched earlier. Is it possible to improve it so there will be more points close to the line and fewer points far from it? If so, sketch the improved line. If not, explain why it is not possible.

_____

_____

_____

Find Out More

Many bivariate data sets have trends that can be described with straight lines. In the case of height and arm span, the points in the scatter plot show a positive association between the two variables. The data points are also close to falling along the same line.

In most statistical scatter plots, points will not all fall as close to the same line as they do in the height and arm-span plot. So, when placing a line in the scatter plot, aim to have the smallest possible total distance between the points in the plot and the line.

Here are two different trend lines. If you look at the points close to 160 on both axes, you'll see that one line is closer to the upper points while the other is closer to the lower points. The difference between the coordinates of points on the line and the actual points on the scatter plot are errors, or **residuals**.
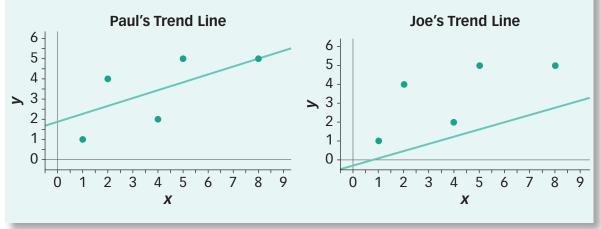
You can't always eliminate error when fitting lines to data, but you can try to make the residuals as small as possible. Work to find a **best-fit line** that minimizes the distance between the actual data points and the line placed on the graph.

Reflect

1 Linear trends sometimes exist in bivariate data, but not always. On a separate piece of paper, draw a picture of a scatter plot for a data set where it would NOT make sense to try to fit a straight line to the data. On the lines below, explain why it would NOT make sense to use a straight line for the scatter plot you drew.

_____

_____

_____

_____

_____

Read the problem below. Then explore more about fitting a line to data to minimize errors in estimates with trend lines.

Paul and Joe placed their trend lines in a data set at different locations. The data set includes the following (x, y) ordered pairs: $(1, 1)$, $(2, 4)$, $(4, 2)$, $(5, 5)$, $(8, 5)$. Whose line better captures the trend of the data?

Paul's Trend Line

Joe's Trend Line
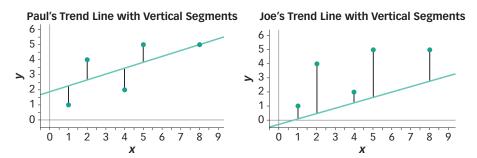
Model It

You can test the accuracy of a trend line by visualizing the residuals.

The trend lines Paul and Joe sketched both try to follow the trend, but they are positioned differently in relation to the points in the scatter plot.

To visualize the residuals for Paul and Joe's trend lines, you can insert vertical segments in the scatter plot.

Paul's Trend Line with Vertical Segments

Joe's Trend Line with Vertical Segments

The vertical segments give an estimate of the distance between each point in the scatter plot and the two trend lines.

Connect It

Now you will decide which trend line is more suitable for the data set that Paul and Joe gathered.

2 Look at each trend line. Do you think Paul or Joe did a better job capturing the trend of the data? Explain.

_____

_____

3 Look at the vertical segments connecting each trend line to each of the data points. Estimate the length of each vertical segment in Paul's scatter plot, then add the lengths to find an approximate sum. Do the same with Joe's scatter plot.

_____

_____

4 Compare the sum of the vertical segment lengths in both graphs. How can this comparison help you decide which trend line is more suitable for the data?

_____

_____

5 Do the sums of vertical segment lengths support your answer to problem 2? Explain.

_____

_____

_____

Try It

Use what you just learned to solve this problem. Show your work on a separate sheet of paper.

6 Sketch a line on a grid. Place five data points to create a scatter plot in which the sum of the vertical segment lengths between the data points and the line is less than 1. How would you describe this situation?

_____

_____

_____

Outliers in bivariate data sets often pull trend lines in their direction.

Pair/Share

Give your own real-world example of a bivariate data set. Where might outliers come from in your data set?

Be careful about trying to fit a trend line to bivariate data. It doesn't always make sense to fit a trend line to all data.
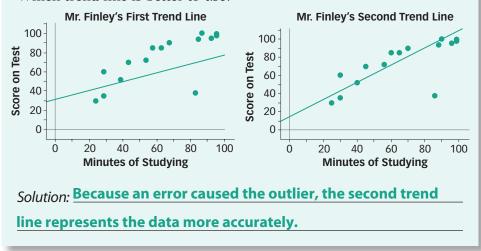
Pair/Share

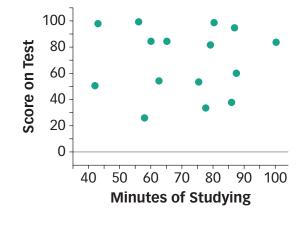Do you think the data shown in the scatter plot are realistic? Why or why not?

Read the situation described below. Then solve problems 7–9.

Student Model

Mr. Finley drew two different trend lines for his data showing students' algebra test scores and time spent studying on the practice website. The first trend line accounts for the value that appears to be an outlier. This outlier was caused when one student logged into the wrong work station and added study time to the wrong person. Which trend line is better to use?
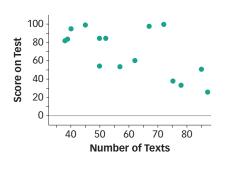
Mr. Finley's First Trend Line

Mr. Finley's Second Trend Line

Solution: **Because an error caused the outlier, the second trend line represents the data more accurately.**

7 This scatter plot shows students' algebra test scores and the time they spent studying on a different website. Do you think a trend line is appropriate for this data? Explain.

_____

_____

8 Mr. Finley is analyzing the data on students' algebra test scores and the number of texts they sent the day before the test. He puts the data in a scatter plot and is about to draw a trend line. What would a trend line show?
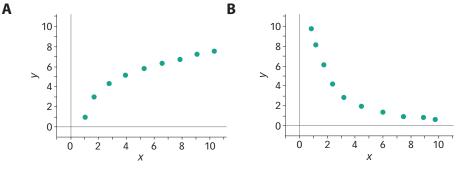

When fitting a line to a data set, keep the residuals as small as possible.

Pair/Share

Is a negative association the same thing as a zero association between variables? Why or why not?

9 Which scatter plot lends itself to having a trend line fit through the data?

A

B

C   neither scatter plot   **D**   both scatter plots

Marnie chose **A** as the correct answer. How is the graph different from a linear trend with an outlier?

A line is just one kind of pattern in a scatter plot. Scatter plots can show lots of other patterns as well.

Pair/Share

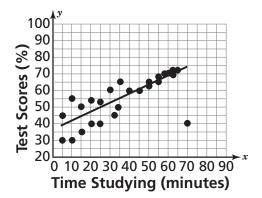How would you help Marnie understand her error?

Solve the problems.

1   Suppose two trend lines are sketched for the same data set in a scatter plot. How can you determine which of the two lines is the better fit?
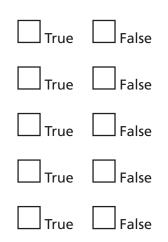
    A   determine which line is closest to the points in the scatter plot

    B   determine which line is the steepest by finding out which one has the greatest slope

    C   determine which line has the greatest *y*-intercept

    D   determine which line intersects the *x*-axis closest to the origin

2   A teacher began collecting data for a particular test to see if an association exists between a student's time spent studying for the test and a student's test score. The data for most of her students is presented in the scatter plot below. A line of best fit shows the relationship.

Time Studying (minutes)

Choose True or False for each statement.

    A   There is a negative association between study times and test scores.   ☐ True   ☐ False

    B   The trend line predicts that a student who studies 80 minutes will get a test score of at least 75%.   ☐ True   ☐ False

    C   The data values are more clustered around the trend line for study times of 40 minutes and higher.   ☐ True   ☐ False

    D   The trend line predicts that a student who spends no time studying will get a 0% on the test.   ☐ True   ☐ False

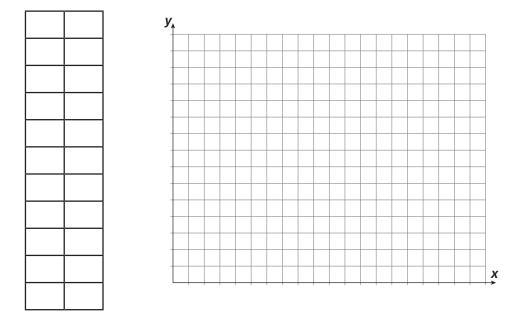   **E**   The student who studied for 70 minutes but received a test score of 40 can be considered an outlier.   ☐ True   ☐ False

3   Invent your own 10-point set of bivariate data. List all the values in the data set in the table below and write a story to describe where you may have collected such data. Also produce a scatter plot for the data and a trend line. The data set should meet the following conditions:

• There should be a linear association between the two variables.

• The linear association should be negative.

• The sum of the distances from the data points to the trend line should be less than 10.

Story about where you collected your data: _____

_____

_____

_____

✓ **Self Check**   *Go back and see what you can check off on the Self Check on page 239.*

©Curriculum Associates, LLC   Copying is not permitted.