Investigation • The Wave

Name ______________________________ **Period** __________ **Date** ______________

You will need: a stopwatch or watch with a second hand

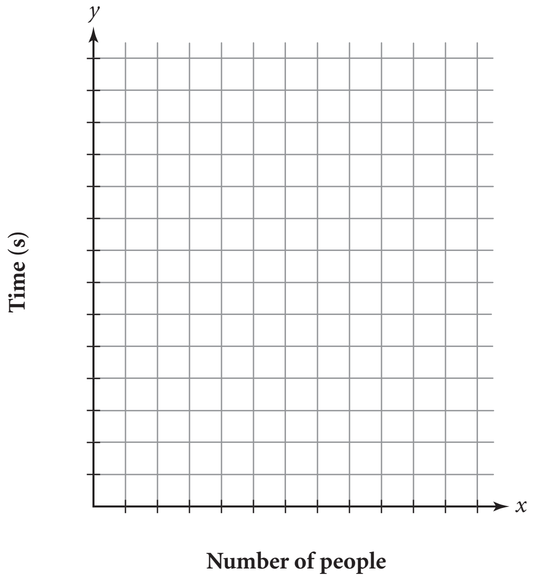
Sometimes at sporting events, people in the audience stand up quickly in succession with their arms upraised and then sit down again. The continuous rolling motion that this creates through the crowd is called "the wave." You and your class will investigate how long it takes different-size groups to do the wave.

Step 1 Using different-size groups, determine the time for each group to complete the wave. Collect at least nine pieces of data of the form (*number of people, time*), and record them in this table.

Number of people	Time (s)

Step 2 Plot the points, and find the equation of a reasonable line of fit. Write a paragraph about your results. Be sure to answer these questions:

- What is the slope of your line, and what is its real-world meaning?
- What are the x- and y-intercepts of your line, and what are their real-world meanings?
- What is a reasonable domain for this equation? Why?



Step 3 Can you use your line of fit to predict how long it would take to complete the wave if everyone at your school participated? Everyone in a large stadium? Explain why or why not.

Name ________________________ **Period** __________ **Date** ____________

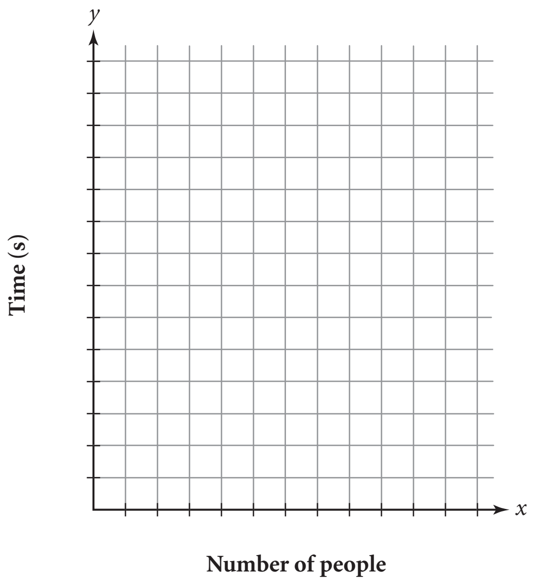
Sometimes at sporting events, people in the audience stand up quickly in succession with their arms upraised and then sit down again. The continuous rolling motion that this creates through the crowd is called "the wave." You and your class will investigate how long it takes different-size groups to do the wave.

Step 1 Using different-sized groups, a mathematics class determined the time for each group to complete the wave. They collected data of the form (*number of people, time*), and recorded them in this table.

Number of people	Time (s)
2	2.1
5	4.4
6	5.2
8	5.8
9	4.7
10	6.7
15	7.5
16	10.4
22	11.0

Step 2 Plot the points, and find the equation of a reasonable line of fit. Write a paragraph about these results. Be sure to answer these questions:

- What is the slope of your line, and what is its real-world meaning?
- What are the x- and y-intercepts of your line, and what are their real-world meanings?
- What is a reasonable domain for this equation? Why?



Step 3 Can you use your line of fit to predict how long it would take to complete the wave if everyone at the school participated? Everyone in a large stadium? Explain why or why not.

E X P L O R A T I O N

Residual Plots and Least Squares

You have learned a couple of different ways to fit a line to data. You've also used residuals to judge how well a line fits and to give a range for predictions made with that line. In this activity you will learn two graphical methods to judge how well your line fits the data. You will also use these methods to identify outliers.

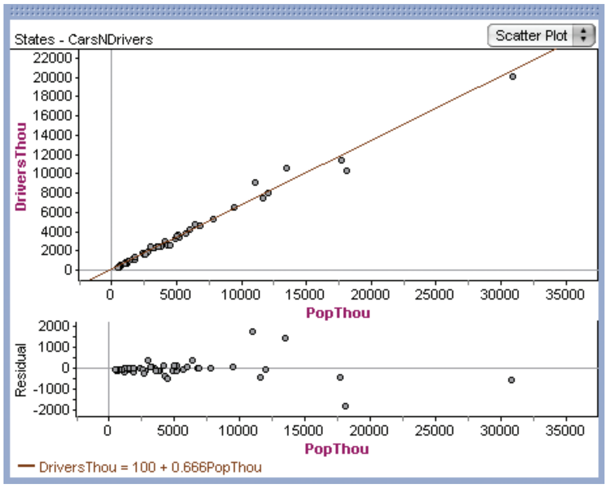
Activity

A Good Fit?

Step 1 Start Fathom and open the sample document titled **States–CarsNDrivers.ftm.** When the file opens, you will see a collection and a case table. This collection gives you various data about the population, drivers, vehicles, and roadways in each U.S. state and the District of Columbia in the year 1992.

Step 2 Create a new graph. Drag the attribute PopThou (population in thousands) to the *x*-axis and drag the attribute DriversThou (licensed drivers in thousands) to the *y*-axis. Your graph will automatically become a scatter plot. Describe the trend in the data and give a possible explanation for the trend.

Step 3 With the graph window selected, choose **Add Movable Line** from the Graph menu. Drag the movable line until it fits the data well. What is the equation of your estimated line of fit? Based on your line, which states are outliers?



Step 4 Choose **Make Residual Plot** from the Graph menu. What does the residual plot show you? From the residual plot, which states are outliers? Are they the same states you selected in Step 3?

Step 5 Experiment by moving the movable line and observing how the residual plot changes.

Step 6 Return to your scatter plot and residual plot. With the graph window selected, choose **Show Squares** from the Graph menu. Move the movable line and observe how the squares change. Explain how each square is drawn.

Step 7 Move your line back to a position where it is a good fit. (You might want to turn off the squares while you do this.) Notice the size of the squares for the outliers that you identified in Step 4. How do they compare to the other squares?

Step 8 In addition to the movable line, Fathom will graph a median-median line and a least squares line. (You'll learn more about the least squares line in Chapter 11.) Try adding a median-median line and a least squares line to your graph, either alone or with your estimate of a line of fit. How do the equations compare? How do the residual plots compare? How do the squares compare?

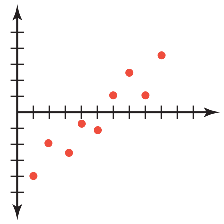
Questions

1. How can you identify outliers from the squares? How did you identify outliers from looking at the line of fit? How did you identify outliers from the residual plot? How are the approaches interrelated? Do you find one approach easier than the other?

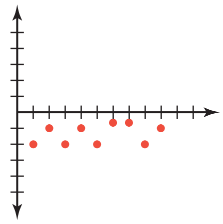
2. As you change the slope of your line, what happens to the residual plot? As you change the *y*-intercept of your line, what happens to the residual plot? Explain how you can use the residual plot to adjust the fit of your line.

3. Graphs A, B, and C are residual plots for different lines of fit for the same data set. How would you adjust each line to be a better fit?

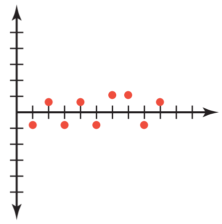
Graph A



Graph B



Graph C



4. Explain how you can use the squares to adjust your movable line to a better fit. Based on your explanation, how do you think the least squares line got its name?

5. When you experimented with all three lines of fit in Step 8, did you get the same equation for all three? Give some reasons why this may or may not have happened.

EXPLORATION

Residual Plots and Least Squares

You have learned a couple of different ways to fit a line to data. You've also used residuals to judge how well a line fits and to give a range for predictions made with that line. In this activity you will learn two graphical methods to judge how well your line fits the data. You will also use these methods to identify outliers.

Step 2 The data are approximately linear. Sample answer: The percentage of people who drive is approximately the same everywhere, so as population increases, the number of drivers will increase too.

Step 3 Sample answer: $DriversThou = 0.670PopThou - 0.12$; Ohio, Florida, and New York appear to be outliers.

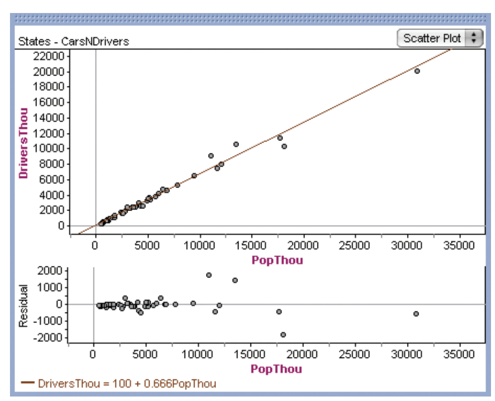
Activity

A Good Fit?

Step 1 Start Fathom and open the sample document titled **States–CarsNDrivers.ftm.** When the file opens, you will see a collection and a case table. This collection gives you various data about the population, drivers, vehicles, and roadways in each U.S. state and the District of Columbia in the year 1992.

Step 2 Create a new graph. Drag the attribute PopThou (population in thousands) to the *x*-axis and drag the attribute DriversThou (licensed drivers in thousands) to the *y*-axis. Your graph will automatically become a scatter plot. Describe the trend in the data and give a possible explanation for the trend.

Step 3 With the graph window selected, choose **Add Movable Line** from the Graph menu. Drag the movable line until it fits the data well. What is the equation of your estimated line of fit? Based on your line, which states are outliers?



Step 4 The residual plot graphs the independent variable against the residuals of the dependent variable; it shows the vertical distance from the line of fit for each data point. Ohio, Florida, and New York still appear to be outliers.

Step 4 Choose **Make Residual Plot** from the Graph menu. What does the residual plot show you? From the residual plot, which states are outliers? Are they the same states you selected in Step 3?

Step 5 Experiment by moving the movable line and observing how the residual plot changes.

DIFFERENTIATING INSTRUCTION

ELL

This activity is a visual representation of the concepts of residuals, which will aid in conceptual understanding.

Extra Support

Although this exploration may look complicated, students will be able to manipulate the data quite easily and effectively.

Advanced

This is an excellent activity to prepare students for the least squares regression line, which is a required objective in advanced statistics courses.

EXPLORATION

OBJECTIVES

- Use statistics software to make residual plots
- Use statistics software to examine least squares
- Evaluate a line of fit based on residual plots and least squares

OUTLINE

One day:

30 min Activity

15 min Discuss Activity

MATERIALS

- Fathom
- Fathom demonstration A Good Fit?, *optional*

TEACHING THE EXPLORATION

ONGOING ASSESSMENT

Check students' understanding of residuals and lines of fit.

Guiding the Activity

MODIFYING THE ACTIVITY

Whole Class Demonstrate the data analysis with the class. Discuss questions in Steps 2 through 4.

Shortened Skip Step 5.

One Step There is no One Step alternative for this activity.

FACILITATING STUDENT WORK

Step 1 You might need to give these instructions: Choose **Open Sample Document** from the File menu. Find the document within the Fathom application folder: **Social Science | United States | States | States – CarsNDrivers.ftm.**

Step 2 You might need to give these instructions: Choose **New | Graph** from the Object menu. Drag the corner of the graph window to increase the size of

Facilitating Student Work (continued)

the graph. **[Ask]** "Does it make sense that these data are linear?"

Step 3 After students have a satisfactory line of fit, **[Ask]** "Does your *y*-intercept make sense?" [Theoretically the *y*-intercept should be 0.] "What does your slope mean in terms of the problem?" [If the *y*-intercept is 0, the slope is the portion $\left(\text{about } \frac{2}{3}\right)$ of the population that drives.]

[Alert] Students might consider the point far to the right an outlier; don't correct them yet. Students can determine the state by double-clicking on a point to open the inspector or by pointing the cursor to the point and looking in the bottom-left corner of the Fathom window.

Step 4 Now students can see that outliers are points with large or small residuals. **[Ask]** "Why do you think these states are outliers?" [Nearly everyone drives in Ohio and Florida; many people in New York City use public transportation.] California is far to the right because it has a large population, not because it has an unusual percentage of drivers.

Step 8 As needed, help students realize that to turn off a movable line, they choose **Remove Movable Line** from the Graph menu. To turn off a median-median line or least squares line, they choose **Median-Median Line** or **Least-Squares Line,** again from the Graph menu. Students might be confused because the residual plot doesn't change when a second or third line is added. Actually, the residual plot is based on the first line for which it is created—probably the movable line. To get a residual plot for another line, students can select that line or its equation. To see all the residuals again, they should once more choose **Scatter Plot** from the Graph menu.

See page XXX for answers to Questions 3 and 4.

Step 6 Each square is drawn by connecting each point to the line of fit with a segment, then constructing a square with this segment as a side.

Step 7 The squares for the outliers are much larger (i.e., have a greater area).

Step 6 Return to your scatter plot and residual plot. With the graph window selected, choose **Show Squares** from the Graph menu. Move the movable line and observe how the squares change. Explain how each square is drawn.

Step 7 Move your line back to a position where it is a good fit. (You might want to turn off the squares while you do this.) Notice the size of the squares for the outliers that you identified in Step 4. How do they compare to the other squares?

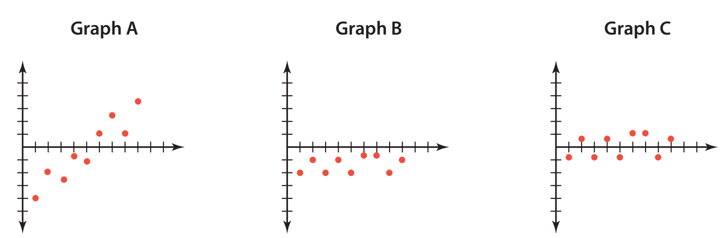
Step 8 In addition to the movable line, Fathom will graph a median-median line and a least squares line. (You'll learn more about the least squares line in Chapter 11.) Try adding a median-median line and a least squares line to your graph, either alone or with your estimate of a line of fit. How do the equations compare? How do the residual plots compare? How do the squares compare?

Questions

1. From the movable line, outliers are those points that have a vertical distance from the line far greater than the other points have. From the residual plot, outliers are those residual points that have a vertical distance from the line $y = 0$ far greater than the other residuals have. The residual plot usually makes it easier to identify outliers. Outliers are those points that have squares with area far greater than the other points have. This is related to the other two approaches because the length of the side of each square is the vertical distance from the line, or the residual.

2. As the slope of the line of fit increases/decreases, the slope of the residual plot decreases/increases (inversely), but the residuals stay equidistant with respect to each other. As the *y*-intercept of the line moves up/down, the residual plot moves down/up (inversely). Because the residuals should have a sum approximately equal to 0 and be equally scattered positive and negative, students can adjust the line of fit by watching how the residual plot changes.

1. How can you identify outliers from the squares? How did you identify outliers from looking at the line of fit? How did you identify outliers from the residual plot? How are the approaches interrelated? Do you find one approach easier than the other?
2. As you change the slope of your line, what happens to the residual plot? As you change the *y*-intercept of your line, what happens to the residual plot? Explain how you can use the residual plot to adjust the fit of your line.
3. Graphs A, B, and C are residual plots for different lines of fit for the same data set. How would you adjust each line to be a better fit?



4. Explain how you can use the squares to adjust your movable line to a better fit. Based on your explanation, how do you think the least squares line got its name?
5. When you experimented with all three lines of fit in Step 8, did you get the same equation for all three? Give some reasons why this may or may not have happened. Answers will vary. For this data set at least, the least squares line and the median-median line differ slightly ($DriversThou = 0.655PopThou + 120$ and $DriversThou = 0.678PopThou + 18$, respectively). Outliers, and how each method of finding a line of fit deals with them, is the primary reason that the equations differ. If the data had been perfectly linear, the median-median line and the least squares line would have been equivalent.

ASSESSING PROGRESS

Students should begin to talk about residuals and outliers in a context of the fit of the line.

DISCUSSING THE ACTIVITY

[Critical Question] "What does rotating the movable line change about its equation? What does shifting the line up and down change?" **[Big Idea]** Rotating the line changes its slope, or the coefficient of *x* in its equation. Translating the line vertically changes its *y*-intercept, or the constant in its equation.

QUESTION NOTES

Question 1, 2 **[Ask]** "What happens to the residual plot if you move the line dramatically, beyond being a good fit?"

Question 4 If students say "You want to make the squares as small as possible," **[Ask]** "How can you take into account all the squares at once?"

DAA2TE_985_03_b.indd 160 9/11/08 1:48:34 PM