



Math Objectives

- Students will analyze the relationship between two quantities by plotting data on a scatterplot.
- Students will informally assess the fit of a function by plotting and analyzing residuals.
- Students will recognize that, if a linear model is a good fit for the data, the residual plot should not have a pattern or be systematic in any way.
- Students will analyze and interpret the significance of the slope and y -intercept of a linear model.

Vocabulary

- explanatory variable
- least-squares linear regression line
- scatterplot
- response variable
- residual
- slope

About the Lesson

- This lesson involves collecting data of student pulse rates before and after exercise and then creating a scatterplot and fitting a line to the data.
- As a result, students will:
 - Fit a movable line to a scatterplot of pulse rates before and after exercise.
 - Interpret the slope and y -intercept of the line in terms of the pulse rates.
 - Identify large residuals from the scatterplot and line and from the residual plot.
 - Drag the line, and observe how the residuals change.
 - Inspect the least-squares regression line and the corresponding residual plot.

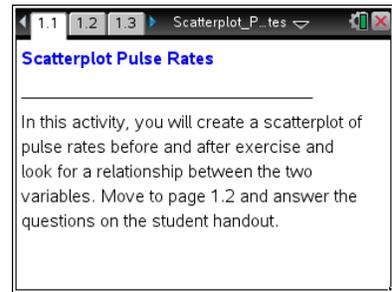


TI-Nspire™ Navigator™ System

- Send out the *Scatterplot_Pulse_Rates.tns* file.
- Monitor student progress using Class Capture.
- Use Live Presenter to spotlight student answers.

Activity Materials

- Compatible TI Technologies: TI-Nspire™ CX Handhelds, TI-Nspire™ Apps for iPad®, TI-Nspire™ Software



Tech Tips:

- This activity includes screen captures taken from the TI-Nspire CX handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at <http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials>

Lesson Files:

Student Activity

- Scatterplot_Pulse_Rates_Student.pdf
- Scatterplot_Pulse_Rates_Student.doc

TI-Nspire document

- Scatterplot_Pulse_Rates.tns



Discussion Points and Possible Answers



Tech Tip: Moving the cursor over a point in a scatterplot will display the coordinates of the point.

Teacher Tip: The Create document contains student instructions for creating a scatterplot and a movable line. You might choose to have students enter the data themselves. In the *Scatterplot_Pulse_Rates.tns* document, the data have already been entered into the spreadsheet. Another option is to have students collect similar data from their classmates outside of class a few days before you do the activity during class.

The data in the tables below represent the pulse rates for 30 students. The data in the "before" columns represent the mean pulse rate over a period of five days at the beginning of class for a student, and the entry in the corresponding "after" columns represents the mean pulse rate of that student after one minute of jumping jacks.

Have students create a scatterplot of the data below using the directions in the *Scatterplot_Pulse_Rates_Create.doc* document. Once students have created the scatterplot, have them follow the directions to add a movable line. Then, they should drag it to model the relationship between the "before" and "after" pulse rates.

	Pulse Beats/Min before exercise	Pulse Beats/Min <u>after</u> exercise		Pulse Beats/Min before exercise	Pulse Beats/Min <u>after</u> exercise
1	86	164	16	73	144
2	88	164	17	73	136
3	75	140	18	79	145
4	88	160	19	79	141
5	64	137	20	86	150
6	84	136	21	98	156
7	85	154	22	74	136
8	93	140	23	76	122
9	93	164	24	96	160
10	86	156	25	92	145
11	87	153	26	58	140
12	96	154	27	84	148
13	72	145	28	92	150
14	67	134	29	88	141
15	62	129	30	80	130



Move to page 1.2.

1. a. Write your equation. Describe what the slope represents in terms of the pulse rates.

Sample Answers: Answers will vary. A positive linear slope indicates that, while the average initial heart rates vary, they all tend to increase by the same amount during one minute of exercise.

- b. Is the y-intercept of your line meaningful? Why or why not?

Sample Answers: The y-intercept is not meaningful because it would be the point where the average pulse rate before exercise was 0.

- c. Identify a point for which your line would not be a good predictor and a point for which your line would be a good predictor. Explain your reasoning.

Sample Answers: The point (76, 122) is not a good predictor because it represents a student whose average pulse rate was 76 beats per minute before exercising and 122 beats per minute after exercising. The line predicts an average pulse rate of more than 140 beats per minute after exercising for this student. The difference of approximately 18 beats per minute between the observed and predicted values is larger than most of the other differences in the scatter plot. This can be seen in the graph because the vertical distance from the point to the line is one of the larger distances. Any point that lies close to the line such as (75, 140) would be a good predictor.

Teacher Tip: Be sure students recognize that the vertical segment between (before, observed) and (before, predicted) represents the residual and that segment will be the side of the square when they look at the sum of the squared residuals.

2. Follow the directions in the Create document (Step 9) to show the residuals for your line.
 - a. Move the cursor to the point you identified in 1c as the one for which the line would not be a good predictor. Move the cursor over the point to see the coordinates. What do these represent?

Sample Answers: The point (76, 122) means that for an initial average pulse rate of 76 beats

	A	B	C	D
1	86	164		
2	88	164		
3	75	140		
4	88	160		
5	64	137		



per minute, the observed average pulse rate after exercise was 122 beats per minute.

- b. Select this point, and observe the associated point in the residual plot. Move the cursor over the point in the residual plot to see the coordinates. What do these represent?

Sample Answers: The coordinates of the associated point in the residual plot are (76, -19.83). These indicate that for an initial pulse rate of 76 beats per minute, the line predicts a pulse rate that is 19.83 or about 20 beats per minutes higher than the observed number of beats, 122.

- c. The y -coordinates of the points in a residual plot are called residuals. By moving the cursor over the points in the residual plot, identify the points that seem to have the largest residuals.

Sample Answers: (58, 21.19) and (93, -21.5) are two other points with large residuals.

- d. Describe any pattern you can see in the residuals. What does this indicate about your line?

Sample Answers: The residuals for lower initial average pulse rates are all positive, and the ones for the largest initial average pulse rates are negative, which suggests this line is not a good model for (before, after) pulse rates. The pattern suggests that the slope has to be less positive for a good fit, which indicates the line should be rotated clockwise.

Teacher Tip: Ask students to think about how they might move the line in terms of slope and y -intercept to address a pattern in the residuals. This can reinforce the relationship between the equation and the line fitted to the scatter plot data.

3. Move your line, and observe how the residuals change. (Note: You might have to select **MENU > Window/Zoom > Zoom – Data** in order to see all of the residuals.)

- a. Find a line where the residuals have a clear pattern. Describe the pattern and the relationship between the line and the scatterplot that led to the pattern.

Sample Answers: Students might describe a residual plot where the residuals are nearly all negative, which indicates the line was above most of the points, or they might describe a residual plot where the only negative residuals occur in the middle, which suggests the line is above most of the points for small and large initial average pulse rates.

- b. Find a line where the residuals seem to have no pattern. Compare your line to the lines several



other students have created. Whose line seems to be the "best fit" for the trend in the data? Explain your thinking.

Sample Answers: Reasons for selecting the "best fit" for the data might include equal number of positive and negative residuals. Alternatively, students might think that not seeing a pattern in the residuals, or ignoring the residuals, suggests that a particular line best represents the trend in the data.



TI-Nspire Navigator Opportunity: *Live Presenter and Class Capture*

See Note 1 at the end of this lesson.

4. Return to the .tns document (Step 10), and add a new page with a scatterplot of the data. Fit the least-squares linear regression line to the data. Look back at your answer to question 1a.
 - a. Compare the slope of your line in question 1a to the slope of the least-squares linear regression line.

Sample Answer: The slope was smaller for the regression line, which had equation $y = 89.79 + 0.68x$. My original line was $y = 1.22x + 48$, so the slopes differ by 0.54.

- b. Describe the residual plot for the regression line.

Sample Answers: The least-squares regression line has no pattern in the residual plot.

According to certain experts, the average resting heart rate for an adult is between 60 and 100 beats per minute, while athletes might have between 40 and 60 beats per minute. The maximum pulse rate should be 220 minus your age, and a healthy pulse rate during, or just after, exercise, should be about 60-80% of the maximum value.

5. Assume the students were between 16 and 17 years old. Which of the students had a healthy pulse rate after the exercise?

Sample Answers: Healthy pulse rates for 16-year olds would be between 122 and 163 beats per minute, while for 17-year olds, they would be between 122 and 162 beats per minute. One student had a pulse rate of 122 beats per minute, while three students had a pulse rate of 164 beats per minute. All of the others were within the target range.



Tech Tip: Students might want to work in a Scratchpad, which they can access by selecting . Or, if they are familiar with spreadsheets, they might want to use the spreadsheet on Page 1.2 to perform the calculations.

Teacher Tip: The lesson can be stopped here, or students can generate and inspect the sum of the squared residuals.

Patterns in the residuals suggest that the line with those residuals might not be an appropriate model for the relationship between two variables.

Move back to page 1.3.

6. Follow the directions in the Create document to display the sum of the squared residuals for your movable line.
 - a. Write down the sum of the squared residuals for your line.

Sample Answers: For the equation $y = 1.22x + 48$, the sum of the squared residuals is 3093.4. (Note that the sum might appear to be slightly different for the same line because the decimals have been set to **fix2** and the line's equation might actually be $y = 1.2235x + 48$ or $y = 1.2199x + 48$.)

- b. The area of a square is the square of the residual for a given average pulse rate. Select a square, and describe its dimensions and area.

Sample Answers: The side of the square is the residual determined by: [observed] minus [the predicted after exercise mean pulse rate for a given initial average pulse rate]. If the residual is 11.5, then the side of the square is 11.5, and the area of the square is 132.25 square units.

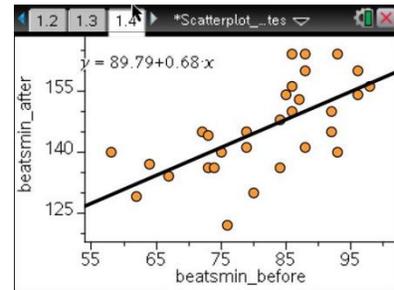
- c. Drag your line to minimize the sum of the squared residuals. What is the smallest sum you can find?

Sample Answers: Answers will vary; for example, the smallest sum I could find was 2069.7.



Move to page 1.4.

7. One special attribute of the least-squares linear regression line is that the line minimizes the sum of the squared residuals. Display the sum of squared residuals for the regression line. Given your work with the movable lines on Page 1.3, does this statement seem to be true?



Sample Answers: The sum of squared residuals for my movable line was always bigger than the sum of squared residuals for the least-squares regression line, which is 2065.5.



TI-Nspire Navigator Opportunity: **Quick Poll**

See Note 2 at the end of this lesson.

Extension

As an extension, students might investigate how the sum of squared residuals varies for the movable line and how the sums of squared residuals for those lines compare to the sum for the least-squares regression line.

Wrap Up

Upon completion of the lesson, the teacher should ensure that students understand:

- What a residual is and how it is represented on a graph.
- If the residuals have a pattern, the model is not a good fit for the data.



Assessment

Decide whether each statement is true or false. If it is false, be prepared to give an example to justify your claim.

1. The least-squares regression line for a set of data will always have the same number of points on both sides of the line.

Answer: False. Finding the minimal sum of squares is not dependent on having an equal number of points on both sides of the line.

2. The least-squares regression line for a set of data will never have a pattern in the residuals.

Answer: False. When a pattern exists, it suggests that the least-squares regression line is not a good model for the relationship between the explanatory and response variables.

3. A residual is always positive.

Answer: False. A residual will be either positive or negative depending on whether the line over or under predicts for a given explanatory value.



TI-Nspire Navigator

Note 1

Question 3, Live Presenter and Class Capture

Use Live Presenter or Class Capture with question 3 to display the different lines students have created, the residuals, and the sum of squares as they try to minimize them.

Note 2

Question 7, Quick Poll

A *Quick Poll* can be given at the conclusion of the lesson. Save the results and before the next class, go to the Portfolio, and open the saved Quick Poll file in the Review Workspace. Discuss possible misunderstandings students may have.