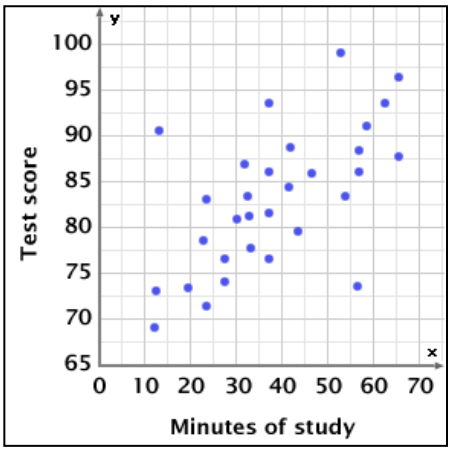
Name: _____ Date: _____

# Student Exploration: Least-Squares Best Fit Lines

**Vocabulary:** least-squares best fit line, outlier, residual, scatter plot, trend line



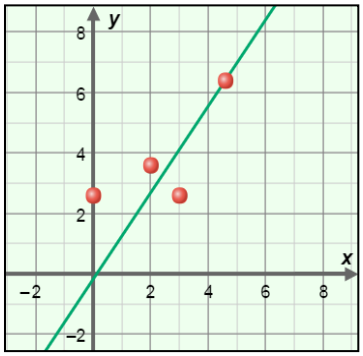

**Prior Knowledge Questions** (Do these BEFORE using the Gizmo.)
Mr. Moore asks his students how long they studied for their last history test. He plots the study times and test scores on a scatter plot, shown to the right. (Each point represents a student's information.)

1. Based on the scatter plot, does study time seem to have an effect on test scores? _____

   Explain. _____

2. Draw a line through the scatter plot that best "fits" the data. Based on this line, about what test score would you predict for a student who studied for 50 minutes? _____

**Gizmo Warm-up**
When statisticians look at data on a **scatter plot**, it is often useful to fit a **trend line** to the data to approximate how one variable is related to the other. There are several ways of plotting trend lines, but the most common is the **least-squares best fit line**, as illustrated in the *Least-Squares Best Fit Lines* Gizmo.



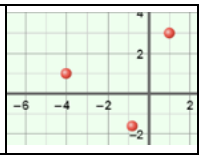

Turn on **Show least-squares fit line**. Notice the green least-squares best-fit line plotted on the graph. Click **New data set** several times.

1. Does the least-squares fit line always go through every point in the scatter plot? _____

2. Does the least-squares fit line always go through at least one point in the plot? _____

3. In general, how does the least-squares fit line relate to the points on the scatter plot?

   _____

| **Activity A:**<br>**Estimating trend lines** | Get the Gizmo ready:<br>• Turn off **Show least-squares fit line**.<br>• Click **New data set**. |  |
|---|---|---|

1. Turn on **Fit a line**. Use the ***m*** and ***b*** sliders to try to fit the line to the data. Write the equation for your line in the left column of the table below. Then, turn on **Show least-squares fit line** and write the actual equation of the line.

| Data set | Equation of estimated best-fit line | Equation of actual least-squares line |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |

Next, turn off **Show least-squares fit line**, click **New data set**, and fit the line again. Continue for two more data sets to complete the table.

In general, how well were you able to fit a line, compared to the actual least-squares fit line?

_____

2. To see how the least-squares method works, turn off **Show least-squares fit line** and check that **Fit a line** is on. Set ***m*** to 1 and ***b*** to 0, to graph $y = x$. (To quickly set the value of a slider, type the number in the text field to the right of the slider, and hit **Enter**.) Drag the red points to these coordinates: (–6, –6), (–4, –4), (–2, –2), (2, 2), (4, 4), and (6, 6).

   A. Turn on **Show error squares**. The **Total error** should be 0.00. (If not, check the coordinates of each point and adjust as necessary.)

   B. Move the point at (2, 2) to (2, 0). An "error square" is now shown. What is the side length of the error square? _____ What is its area? _____

   C. What is the total error? _____

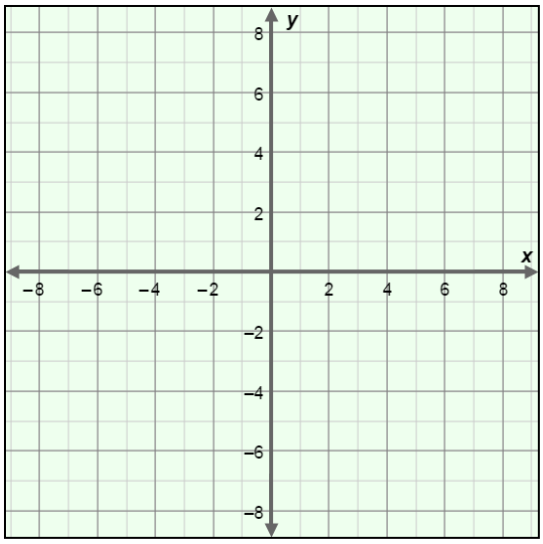
   D. Move the point at (6, 6) to (6, 3). What are the side length and area of the error square for this point? Side length = _____ Area = _____

   E. What is the **Total error** now? _____

The vertical distance between a point and the best-fit line is called the **residual**. The least-squares method finds the total error by summing the squares of the residuals. The best fit line is the line with the smallest possible total error.

**(Activity A continued on next page)**

**Activity A (continued from previous page)**




3. Turn off **Fit a line**, and be sure **Show least-squares fit line** is still off. Click **New data set**. Sketch the data set on the grid to the right.

   Next, turn on **Fit a line**. Adjust ***m*** and ***b*** until you create a line with the smallest total error.

   A. What is the equation of your line?

      $y =$ _____

   B. What is the total error? _____

   C. Turn on **Show least-squares fit line**.

      What are the equation and total error of the actual least-squares best fit line?

      Equation: $y =$ _____ Total error: _____

   D. Use the same procedure to estimate the least-squares best fit line for several other data sets. How well were you able to estimate the best-fit line when you could see the total error, compared to when you could not? Explain your answer.

      _____

      _____


4. The scatter plot to the right shows low-density lipoprotein (LDL) levels vs. weekly hours of exercise for patients in a fictional study. ("LDL's" are often called "bad cholesterol.") The least-squares best fit line is shown.

   A. Estimate the equation of the least-squares line:

      $y =$ _____

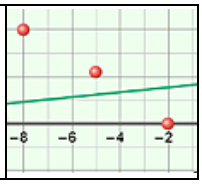
   B. Based on the equation, what LDL level would you expect if you exercised 0 hours per week? _____

   C. What LDL level would you expect if you exercised 11 hours per week? _____

   D. High levels of LDLs in the blood have been associated with greater risk of heart disease. What does this graph indicate about the possible benefits of exercise?

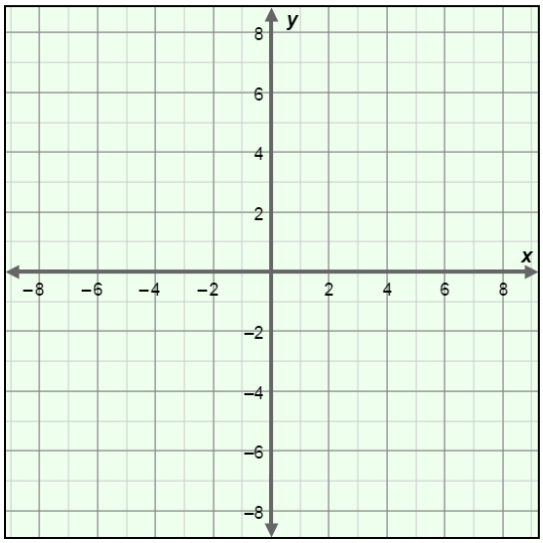
      _____

      _____

| **Activity B:**<br>**Limitations of least-squares** | Get the Gizmo ready:<br>• If necessary, turn off **Fit a line** and **Show least-squares fit line**. |  |
|---|---|---|

1. Drag the points on the scatter plot to these coordinates: (–8, 4), (–5, 2), (–2, 0), (1, –2), (4, –4), and (7, –6). Turn on **Show least-squares fit line**. The points should all be on a line with the equation $y = –0.67x – 1.33$.

   Next, move the point at (7, –6) to (8, 8). This point, which does not follow the trend of the rest of the data, is called an **outlier**. Sketch the new scatter plot and trend line in the grid to the right.




   A. How did the outlier affect the least-squares best fit line? _____

   _____

   B. In your opinion, does the least-squares line describe this data set well? _____

      Explain. _____

      _____

2. Turn off **Show least-squares fit line**. Drag the points to the following coordinates: (–6, 6), (–4, 4), (–2, 2), (2, 2), (4, 4), (6, 6). This data represents $y = |x|$ (absolute value of $x$).

   A. How would you describe this data distribution? _____

   _____

   B. Do you think of best-fit line will be useful in describing this data set? Why or why not?

   _____

   _____

   C. Turn on **Show least-squares fit line**. What is its equation? _____

   D. Do you think the least-squares best fit line describes this data set well? _____

      Explain. _____

Lines of best fit are most useful when data shows a linear relationship. Many relationships, including $y = |x|$, are not linear and are not described well by linear trend lines.