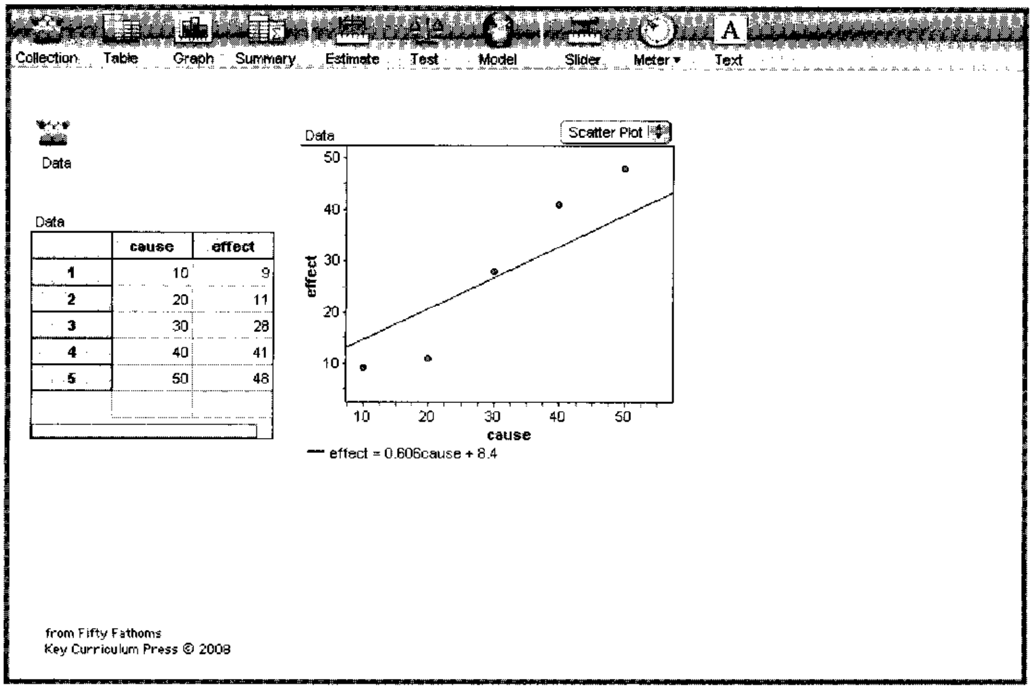
Demo 6: Least-Squares Linear Regression

Exploring the squares in least squares • Minimizing the areas of the squares built on residuals

This demo is essential for anyone learning about statistics.

This demo clarifies what the least-squares line means by connecting the idea behind the calculation with a dynamic visual representation. It's so easy and quick, we've added an extension that explores the startling effect of influential points.



What To Do

▹ Open **Least Squares.ftm.** It shows made-up "data" that are correlated. It should look something like the illustration.

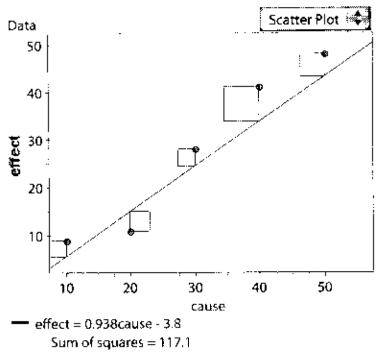
▹ Play with the brown *movable line.* Drag its middle to change the intercept; drag the ends to change the slope. Notice how the equation for the line (below the graph) changes as you move the line, and that the variables in the equation are the attributes **cause** and **effect** rather than x and y.

▹ Move the line to make as good an "eyeball" fit to the points as you can.

There may be several lines that appear to be "best" fits. In fact, different people will choose different best lines. So, is there an objective way to determine how good a fit is? We'll explore one such way—looking at the squares of the residuals.

▹ Choose **Show Squares** from the **Graph** menu. Squares appear, built on vertical segments extending from each point to the line. In addition, you can see the **Sum of squares** by the equation. It's the sum of the *areas* of the five squares.

⇨ If there is no **Graph** menu, click once on the graph to select it. Then the **Graph** menu appears.



▹ Play with the line, watching as the sum of squares changes. Try to minimize it. It will end up looking something like the illustration (but you can do better).

⇨ If this is a whole-class demo, you can invite students to the computer to move the line.

▹ Test your minimum; have Fathom display the true *least-squares line.* Choose **Least-Squares Line** from the **Graph** menu; it appears with your movable line.

▹ You can try minimizing again if you wish: Unchoose **Least-Squares Line** to make the answer go away, then move any points you want by dragging them. Now minimize the sum of squares again by dragging the movable line, and test your result by making the least-squares line reappear.

Questions

1 How did you go about minimizing the sum of squares?

2 It should look as if some points contribute more to the sum of squares than others do. Which ones?

3 Why do you suppose they used *squares* for this?

Sol

Extension

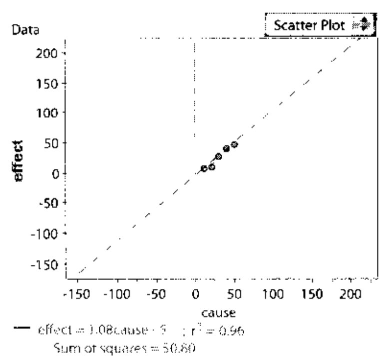
▹ Make the least-squares line appear and make the movable line disappear (choose **Remove Movable Line** from the **Graph** menu)—that is, just show the least-squares line and its associated squares.

▹ Drag a point and watch what happens. Then return the points to their original places.

▹ Drag an end point up and down a little. What happens to the least-squares slope? What happens to the intercept? Does it matter which end point you move?

▹ Drag the middle point up and down a little. What happens to the slope? What happens to the intercept?

▹ Now we want to drag points farther, so we need a bigger range on both axes. Rescale the axes (see "Rescaling Graph Axes" in the introductory material) to give more space, as in this illustration:



⇨ To zoom out automatically, hold down **Option+Shift** (Mac) or **Control+Shift** (Windows) and point at the graph. A magnifying glass with a minus sign appears. Click to zoom out.

▹ Now drag a point and watch the line move (its equation updates, too).

▹ See if you can make the least-squares slope *negative.*

▹ See if you can make **r^2** (its value appears at the bottom of the graph, to the right of the equation) equal to zero. What do you suppose that means?