

Geometry of the Mean

<i>What is this?</i>	<i>Sketchpad activity</i>
<i>Topic</i>	<i>What the Mean Means Geometrically</i>
<i>Paradigm</i>	<i>computer-intensive</i>
<i>Materials</i>	<i>Sketchpad, Means.gsp sketch, Geometry of the Mean worksheet</i>

There are many ways of defining a “center” of a set of numbers—the mean, median, and mode being the most commonly used. Here we are interested in discovering the particular characteristics of the mean. A geometric view will highlight two of these characteristics: 1) the algebraic sum of the differences between the given numbers and the mean is zero; 2) the mean is the number that minimizes the sum of the squares of the differences.

There are geometric properties of the mean to discover; for example, at least one number will lie on each side of the mean, changing the value of one number by an amount s will change the mean by an amount s/n where n is the number of points, and the mean is sensitive to outliers.

Introduction

Ask six students to stand in front of the class in something like a line. Get one student to stand quite far from the others (an outlier). Ask seated students in pairs to decide on where the “center” of this group is. Explain that there is no one right answer, but you will ask for volunteers to position themselves (perhaps holding a flag or some other special marker) and give an explanation of why they chose that location.

Choose volunteers (at least three) until there are no more explanations to be offered.

Explain that you are going to ask one of the original six students to move. Each of the “center” students should move (or not move) accordingly. Do it and get students to comment on whether the centers moved as they should have. Try again with the “outlier” student.

Conceptual Development

Worksheet in hand, students go through a series of investigations using *The Geometer's Sketchpad* to look at two numerical criteria for a center. No special Sketchpad knowledge aside from dragging and multiple selection is assumed.

The worksheet has five sets of questions that students should answer and which can form the basis of class or small group discussion.

The conclusion we're headed for is that the mean is a measure of center that both sets the algebraic sum of the residuals to zero and minimizes the sum of squares of residuals; i.e., the mean is a least squares measure.

The Geometry of the Mean

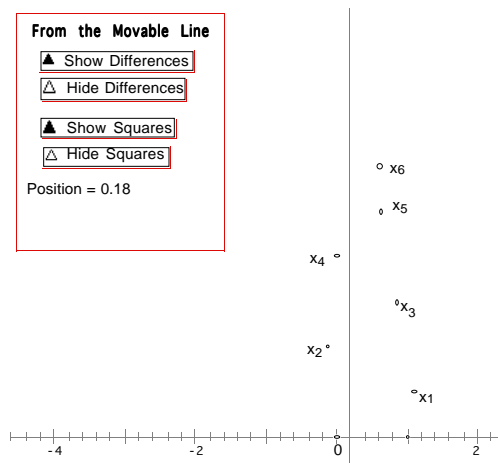
In this activity you'll investigate the question "What does the mean mean?"

- Open the sketch named Mean. It should look similar to the figure at right.

You can think of the six points labeled x_1 through x_6 as representing six data values such as six heights, six incomes, or six cholesterol levels.

- Try dragging one or more of the points.

(Recall that with Sketchpad, whenever you do something that you don't like, you can choose Undo from the Edit menu to get things back the way they were.)



Notice that you cannot change the vertical position of the points. We are only interested in the points' x-coordinates.

The vertical dashed line is movable. We'll use it to look for a good place for the "center" of the x points.

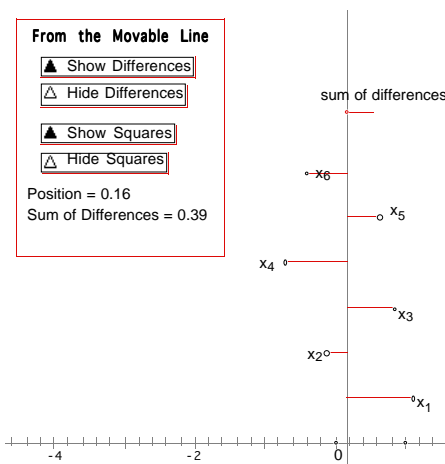
- Try dragging the dashed line. Think about where to position the line to be in the center of the data points.

The Sum of Differences

- Double-click the Show Differences button.

You should see horizontal segments connecting the data points with the dashed line, and a vertical segment labeled "sum of differences."

- Experiment with dragging the dashed line and the data points until you can answer the questions below.



Questions

- Describe at least two ways to get the sum of differences to be negative.
- What can you do to get the sum of differences to be zero?
- No matter what the positions of the data points, is it always possible to adjust the dashed line so that the sum of differences is zero? Why or why not?
- Write a definition for the center of data values that uses the sum of differences.
- The horizontal segments, the differences, are often called "residuals." Explain why that is an appropriate term for them.

Let's call this center of the data points the "sum of differences is zero" center. (Don't worry, we'll soon have a shorter name.)

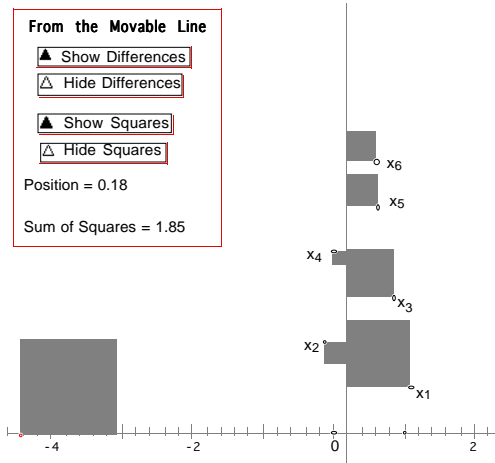
The Sum of Squares

There are many other ways to define a center of values. Let's experiment with a second way.

- Double-click the Hide Differences button.
- Double-click the Show Squares button.

The difference segments are replaced with squares. The area of the seventh square equals the sum of the areas of the smaller squares.

- Experiment with dragging the data points and with dragging the dashed line until you can answer the questions below.



Questions

- Q1. Under what circumstances will the sum of squares be zero?
- Q2. No matter what the positions of the data points, is there always a position of the dashed line that makes the sum of squares zero? Why or why not?
- Q3. How can you use the sum of squares to define a center?

Let's call this center of the data points the "least squares" center.

Comparing the Two Methods

- Show both the differences *and* the squares.
- Adjust the dashed line so that it is at the least squares center.

What do you notice about the sum of differences?

- Adjust the dashed line so that it is at the "sum of differences is zero" center.

What do you notice about the sum of squares?

- Change the locations of some of the data points.
- Adjust the dashed line so that it is at the "sum of differences is zero" center.

What do you notice about the sum of the squares?

Conjecture

State a conjecture relating the two centers you have investigated.

The Mean as a Center

Now it is time to compare the two centers you've investigated above with one you are familiar with—the *mean* of a group of values. You know that to compute the mean of six numbers you add them up and divide by size. That is,

$$\text{mean} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6}$$

We often indicate the mean with a bar over the variable, so we say the mean of the x 's is \bar{x} . Then we have

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6}$$

Double-click the Show Mean button.

The differences and squares related to the moveable dashed line will disappear. The moveable dashed line itself will remain. (You may find it helpful to move this line to one side for now.)

A new vertical dashed line will appear. Its position corresponds to the mean of the x -values.

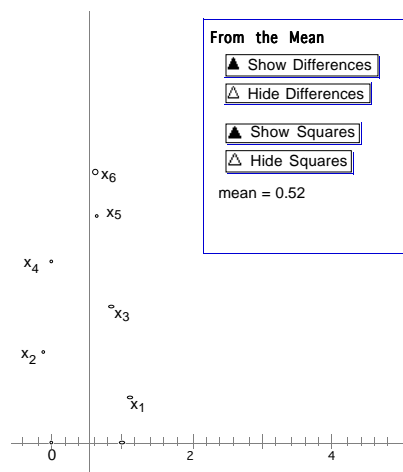
Try dragging the mean line.

Notice that all the x points and the axis come with it.

Drag one of the x points.

Notice that the mean moves as well, but not as far.

Experiment with moving the x points until you can answer the questions below.



Questions

- Q1. How many points can you get to lie to the right of the mean? Explain your finding.
- Q2. How far do you have to drag one point to get the mean to move one small tick along the axis?
- Q3. Select and drag two data points. How far do you have to drag them to get the mean to move one small tick along the axis?

The Mean Compared to “Sum of Differences Is Zero” and Least Squares

Now we want to know whether the mean is the same or different than the center defined by the two previous criteria.

- Double-click the Show Differences buttons for both the moveable dashed line and for the mean.
- Drag the moveable line until it matches the mean.

What do you notice about the sum of differences when the two lines match?

- Drag the moveable line away from the mean. (Or choose Undo.)

- Double-click the Show Squares buttons for both the moveable dashed line and for the mean.

What do you notice about the sum of squares when the two lines match?

- Experiment further with this sketch until you can answer the questions below.

Questions

- Q1. Write a conjecture about the mean and the sum of differences.
- Q2. Write a conjecture about the mean and the sum of squares.
- Q3. Explain the statement, “The mean is a least squares measure of center.”

Further Study

- Q1. Prove algebraically that the sum of differences from the mean is zero.
- Q2. Consider three points in the plane. Find the mean of their x-coordinates and the mean of their y-coordinates. Relate this to the triangle defined by the three points.

Extend this line of investigation to greater numbers of points.
- Q3. Prove that the mean minimizes the sum of squares of the residuals.

