Mean Meets the Median: Measures of Central Tendency

For GSP5 ACTIVITY NOTES

INTRODUCE

Project the sketch for viewing by the class. Expect to spend about 5 minutes.

1. Open **Mean Meets the Median.gsp** and go to page "Median 5."
2. Explain, ***Today you're going to use Sketchpad to compare the behavior of the mean and the median. We'll start with a given data set and explore how changes to the values of that data set affect both the mean and the median. This will help us understand the difference between the two measures of central tendency and when it might be more appropriate to use one or the other. We'll look at questions such as "When is the mean greater than the median?" and "Can the mean and the median ever have the same value?" Before you begin, I'll demonstrate how to change a data set represented in the sketch.***
3. Point out that the vertical orange line represents the median value of the five data points. The line currently goes right through the data point at 4.0. Drag one of the data points along the axis until the median changes value. Drag a different point along the axis until the median changes again.
   - Ask students what they notice about the median value. They might describe it as always being in the middle of the data values, and they might also notice that the vertical line always passes through one of the data points. Make sure such intuitive ideas about the median are shared with the whole class in this introduction portion of the activity.
   - Draw attention to the fact that there are five data points. Later, students will investigate how to calculate the median of a data set that has an even number of data points.

DEVELOP

Expect students at computers to spend about 25 minutes.

4. Assign students to computers and tell them where to find **Mean Meets the Median.gsp.** Distribute the worksheet. Tell students to work through step 5, which wraps up the exploration of the median of five data points.
5. After most students have completed worksheet step 5, ask for examples of data sets with a median of 4.0. Illustrate them in Sketchpad. Ask students for different data sets (these might include data sets where there are two or more values on 4.0, as well as data sets that are either

clustered or spread out). The variety of data sets should show that even though the median value is always the middle value, it does not always look visually centered.

6. Ask students to work through step 13 and do the Explore More question if they have time. Let pairs work at their own pace. As you circulate, here are some things to notice.
   - Make sure students can articulate the difference between the way the median is calculated with five or six data points.
   - Encourage students to drag the different data points to different locations and not just focus on one data point.
   - For steps where students are asked to provide more than one example, encourage them to construct examples that are different. For example, the data sets {1, 2, 3, 4.1, 5}, {1, 2.1, 3, 4, 5}, and {1, 2, 3, 4, 5} are similar examples with medians of 3. However, {0, 0, 3, 3, 10} is quite different because it contains both repeated and extreme data points.
   - Encouraging students to see what their data sets have in common might help them construct different data sets. (Sometimes it's hard for them to think outside a particular strategy they have developed to generate new data sets.)
   - In worksheet step 13, you might ask students to try some sample data values so that they see that the waiting times for the patients who stayed all morning are much longer than for the other patients. Therefore the corresponding data points might be considered outliers.

SUMMARIZE

Project the sketch. Expect to spend about 15 minutes.

7. Gather the class. Students should have their worksheets with them. Begin the discussion by opening **Mean Meets the Median.gsp** and use it to support the class discussion.
8. Ask students to offer explanations for worksheet step 8. If students do not explicitly mention the words *odd* and *even,* probe their understanding by asking them what might happen in the case of data sets with 11 or 12 data points. They should be able to generalize that to find the median of an odd data set one simply locates the middle number, and to find the median of an even data set one must find the mean of the two middle numbers.

9. Now discuss the question in step 10, and ask students to give examples of three different ways to make the mean and median equal to each other. One might be to change the data point representing the median value, whereas another might be to increase or decrease one of the other values to push the mean closer to the median. Help students notice that there are many different data sets that will have the same mean and median, so that these two measures of central tendency do not completely describe a data set.

10. Students have now worked quite a bit on constructing different sets of data. ***Now we'll try to characterize the major differences between the mean and the median. Which measure of central tendency is more sensitive? Which might be more useful in different circumstances? Using the sketch, on page "Mean and Median" set the data values to 5.1, 5.4, 5.9, 5.9, and 6.2. Imagine these are the heights of five people lined up to go through security at the airport. The mean is lower than the median. Now suppose that one of the people lined up was a toddler, who measured 2.3. This new data value can be thought of as an outlier. How would the mean and median change?*** Drag one of the first values to 2.3. Now the mean will be much smaller than the median. This illustrates the way in which outliers can affect the mean without affecting the median at all.

11. Continue on to worksheet step 13. Ask for volunteers to describe their answers. Students should be able to recognize that the waiting times that are outliers influence the decision about hiring a new doctor, so the median might be a better measure. ***Are there other situations you can think of where you would want to consider the outliers?*** Encourage students to propose a variety of situations.

EXTEND

Suppose a teacher has just marked a mathematics test she gave to her students. Can you find a reason why she might prefer to know the mean of the test scores? Can you find a reason why she might want to know the median? Can you find a reason why she might want to know both? What other pieces of information might be useful to her? Encourage students to play with specific examples if they need to. If we want to know whether a test was too hard for most students, then the mean is helpful because it is sensitive to extreme values (in this case, to very low test scores).

However, the median might be a good measure if we don't care about extreme values (maybe some students were away, so we don't want their marks of 0 to affect the measure). It might also be helpful to know the range of the values. If the range is relatively wide, we might conclude that the test was not a very good one, because some students found it much too easy and others, much too hard.

You might also ask what data sets with equal means and medians have in common, and whether there are other measures of central tendency that have the advantages of each (for instance, dropping the highest and lowest values and finding the mean of the remaining values, as they do in the Olympics).

ANSWERS

1. Answers will vary. All data sets should include the value 4.0.
2. The median won't change as long as the data value remains greater than 4.0.
3. Once the data value becomes smaller than 4.0, the median will change because 4.0 is no longer the middle value.
4. Answers will vary. The data set should include two points with the same value as the median.
5. The median is the middle value when the data points are placed in order, but sometimes the middle value will not be visually centered among the data points.
6. Answers will vary.
7. Answers will vary. Some students might start by placing two data points on 7.5 and then two data points on each side of 7.5. Others might start with two data points whose average is 7.5 and then place two data points on each side of 7.5.
8. In step 1, with five data points the median is the middle value, but now the median is the average of the two middle values.
10. Answers will vary.
11. Answers will vary. Increasing the largest data value will not affect the median, but it will increase the mean.

12. Only the mean. Increasing the value of one data point does not change the middle value, but it does change the sum, and therefore the mean, of the values.

13. It's probably better to use the median because we don't want to include the extra waiting time of the patients who arrive early.

14. Answers will vary. The sum of the values of the two data points remains constant, so the mean is unaffected.