

Grade 7 Effect of outlier on mean, median, mode

7.SP.2	
Determine the effect on the mean, median and mode when an outlier is included in a data set.	<ol style="list-style-type: none">1. Analyze a given set of data to identify any outliers.2. Explain the effect of outliers on the measures of central tendency for a data set.3. Identify outliers in a set of data and justify whether or not they are to be included in the reporting of the measures of central tendency.4. Provide examples of situations in which outliers would and would not be used in reporting the measures of central tendency.

Clarification of the outcome:

- ◆ An outlier is datum that is far outside the rest of the data. For example, consider the data: 3, 145, 187, 162, 139, 155, 191, 180. '3' is the outlier of the data. The rest of the data lie in the hundred zone.
- ◆ The outcome concerns what an outlier is, whether to include it or drop it, and its effect on the three measures of central tendency (mean, median, mode) when an outlier is present. An outlier does not affect the median or mode in any important way but an outlier can create a significant change to the mean.

Required close-to-at-hand prior knowledge:

- ❖ Understand mean, median, and mode (grade 7 outcome 7.SP.1).
- ❖ Comfortability with a spreadsheet (the DEVELOP lesson makes use of a spreadsheet). A calculator can be used to replace the spreadsheet in the activities that involve it.

SET SCENE stage

The problem task to present to students:

Organize students into groups of 2. Provide them with data on test scores (e. g. mathematics scores) and have them enter the data into a spreadsheet.

Note:

Ensure that the data has one outlier (for example, a test score of 7%).

Have students use the spreadsheet to order the data and to determine the median and mode. Have students use the spreadsheet to determine the mean. Ask students to explain how the spreadsheet is determining the mean, median, and mode. [e.g.: For the mean, the spreadsheet is adding up all the numbers in the list and dividing by how many numbers were added.]

Comments:

The main purpose of the task is to refresh students' understanding of the measures of central tendency. Understanding the effect of an outlier on these measures requires a good understanding of the measures themselves.

Using a spreadsheet should encourage engagement with the task. It also sets the stage for using a spreadsheet for investigating the effect of an outlier on the measures of central tendency (in the DEVELOP lesson).

DEVELOP stage

Activity 1: Revisits SET SCENE, and addresses achievement indicator 1.

- ◆ Ask selected groups to describe the mean, median, and mode. ENSURE students understand these measures of central tendency.
- ◆ Ask students to identify a test score that seems out of line with the rest of the scores. Discuss why it might be unusual.

Activity 2: Addresses achievement indicators 1 and 2.

- ◆ Provide students with measurement data on the height of grade 7 students that is in centimetres (about 10 numbers). [Arrange it so that one of the numbers is way off - very high or very low.] Ask students to examine the data and identify the height that seems way off. Discuss the term 'outlier' - an extreme value (far beyond or below the other numbers in the data set).
- ◆ Organize students into groups. Ask the groups to determine the mean, median, and mode of the height data without using a spreadsheet. Discuss results.
- ◆ Ask them to determine the mean, median, and mode when the outlier is removed from the height data.
- ◆ Compare the result for the mean, median, and mode when the outlier was included and when it was not included. Come to a conclusion about which measure of central tendency (mean, median, or mode) seems to be most affected by an outlier. Ask them to explain why that might be. Ensure students begin to realize that because the mean involves addition of all data, an extreme value (an outlier) can distort the sum. Median and mode do not involve adding up the numbers.

Activity 3: Addresses achievement indicators 1, 2, and 3.

- ◆ Provide students with hourly pay data for ten MacDonald's employees. [**The rates should be fairly close to each other - around minimum wage, \$10 per hour.**] Have them enter the data into a spreadsheet and use the spreadsheet to determine the mean, median, and mode. Ask students to predict what happens to the mean, median, and mode if one of the employees receives a \$9 increase in the hourly rate.
- ◆ Organize students into groups with access to a spreadsheet. Have each group add \$9 to the hourly pay of one of the employees (ensure that each group does this for a different employee). Have each group determine the mean, median, and mode of the changed data. Ask each group to present their results. Discuss the term 'outlier' in relation to the changed data and whether the employee with the increase in hourly pay should be included in the data.
- ◆ Have each group subtract \$9 from the hourly pay of one of the employees (ensure that each group does this for a different employee). Have each group determine the mean, median, and mode of the changed data. Ask each group to present the results. Discuss 'outlier' in relation to the changed data and whether the employee with the decrease in hourly pay should be included in the data.
- ◆ Change the hourly pay data so that there are 50 employees, each of which have an hourly pay around \$10. Have students calculate the mean of the data. Have students change the hourly pay of one of the employees to \$19 per hour. Ask them to calculate the mean again. Discuss the effect of the outlier on mean for a data set of 50 numbers compared to a data set of 10 numbers.

Activity 4: Addresses achievement indicators 1, 2, 3, and 4.

- ◆ Provide students with several data sets (e.g. data on average temperature in various parts of the planet). [**The data must contain outliers.**]
- ◆ For each data set, have students identify outliers and discuss if they should be included in reporting measures of central tendency. Their decision about inclusion or not inclusion must be supported by an explanation. [The explanation should be based on the purpose for using the measure of central tendency.]

Activity 5: Addresses achievement indicators 1, 2, 3, and 4, and practice.

- ◆ Provide students with measurement data that concerns 12 people measuring the length of the SAME object. One of the measurements has to be VERY DIFFERENT (high/low) from the other measurements. Have students identify the outlier and discuss if it should be included in reporting measures of central tendency. Their decision about inclusion or not inclusion must be supported by an explanation. [In this case, the explanation must concern error in measurement. When measuring the same length, all measurements will have to be close to each other. If there is an outlier, it indicates an error was made, either in recording or in measuring.]

Activity 6: Revisits SET SCENE & addresses achievement indicators 1 to 4, and practice.

- ◆ Organize students into groups of 2. Ask each group to randomly create 20 test scores (0% to 100%) for some subject area.
- ◆ Have each group write a report on their randomly created test data that includes: (1) identification of any outliers, (2) the effect of any outlier on the measures of central tendency, and (3) what to do with any outliers (ignore or include) and why.
- ◆ Have selected groups present their report. Discuss results.

Activity 7: Assessment of teaching.

- Provide students with a REAL data set of 20 numbers for which the range is no more than 20% above/below the average of the data set. [REAL refers to data that comes from a real world source. Let students know what the data is about.] Provide the mean, median, and mode of that data.
 - Ask them to describe what happens to the mean, median, and mode if one of the numbers is replaced by a value that is 10 times as large (e.g., 39 is replaced by 390).
 - Ask them to tell you the statistics term for replacement number.
 - Ask them to describe what happens to the mean, median, and mode for the changed data (they do not do any calculations for this).
 - Tell them to imagine that the replacement number was part of the original data set. Ask them to decide whether it should be included in reporting measures of central tendency. Ask them to explain the decision.

If all is well with the assessment of teaching, engage students in PRACTICE (the conclusion to the lesson plan).

An example of a partially well-designed worksheet follows.

The worksheet contains a sampling of question types. More questions of each type are needed.

The MAINTAIN stage follows the sample worksheets.

Question 1.

Identify any outliers in each set of data.

- a) 12, 18, 23, 11, 26, 17, 290, 32
- b) 1005, 2099, 1873, 2354, 4012, 124564, 2978, 6001

Question 2.

- a) Determine the mean, median, and mode for the following data.
12, 34, 51, 32, 89, 45, 56, 13
- b) Determine the mean, median, and mode for the following data.
12, 34, 51, 32, 89, 45, 678, 13
- c) Examine the two sets of data. One of the sets has an outlier. Identify it.
- d) Which measure of central tendency has changed a lot because of the outlier? Explain why that happens if an outlier is present.

Question 3.

- a) Determine the mean for the following data.
12, 34, 56, 25, 45, 38, 47, 33, 21, 50
- b) Pick one of the numbers in the data above and multiply its value by 20.
- c) Calculate the mean for the changed data. What happened and why?

MAINTAIN stage

Mini-task example

Every so often:

- Present data and ask students to identify any outliers and to explain their effect on the measures of central tendency.

Rich-task example

Have students engage in a statistics project that involves reporting measures of central tendency. Instruct students to be sensitive to outliers (identifying them and deciding whether to include them for reporting purposes, and providing an explanation for the decision).

Comments

This is a rich-task because it is a complex problem that integrates outliers with project work.