

7.3

The Effects of Outliers on Average

Focus Understand how mean, median, and mode are affected by outliers.

Explore



Students in a Grade 7 class measured their pulse rates.

Here are their results in beats per minute:

97, 69, 83, 66, 78, 8, 55, 82, 47, 52, 67, 76, 84,
64, 72, 80, 72, 70, 69, 80, 66, 60, 72, 88, 88

- Calculate the mean, median, and mode for these data.
- Are there any numbers that are significantly different from the rest? If so, remove them.
Calculate the mean, median, and mode again.
Explain how the three averages are affected.



Reflect & Share

Compare your results with those of another pair of classmates. How did you decide which numbers were significantly different? Why do you think they are so different?

Connect

A number in a set of data that is significantly different from the other numbers is called an **outlier**.

An outlier is much greater than or much less than most of the numbers in the data set.

Outliers sometimes occur as a result of error in measurement or recording. In these cases, outliers should be ignored.

Sometimes an outlier is an important piece of information that should not be ignored. For example, if one student does much better or much worse than the rest of the class on a test.

Outliers may not always be obvious. Identifying outliers is then a matter of choice.

Example

Here are the marks out of 100 on an English test for students in a Grade 7 class:

21, 23, 24, 24, 27, 29, 29, 29, 32, 37, 37, 38, 39,
40, 50, 50, 51, 54, 56, 57, 58, 59, 61, 71, 80, 99

- How many students were in the class? How do you know?
- What is the outlier? Explain your choice.
- Calculate the mean, median, and mode.
- Calculate the mean, median, and mode without the outlier.
What do you notice?
- Should the outlier be used when reporting the average test mark? Explain.

A Solution

- Count the number of marks to find the number of students in the class.
There are 26 students.
- There is only one number, 99, that is significantly different.
The outlier is 99.
The difference between the outlier and the nearest mark is $99 - 80 = 19$.
This difference is much greater than that between other pairs of adjacent marks.
- There are 26 marks. To find the mean mark, add the marks:
 $21 + 23 + 24 + 24 + 27 + 29 + 29 + 29 + 32 + 37 + 37 + 38 + 39 +$
 $40 + 50 + 50 + 51 + 54 + 56 + 57 + 58 + 59 + 61 + 71 + 80 + 99 = 1175$
Divide the total by the number of marks, 26: $1175 \div 26 \doteq 45.2$
The answer is written to the nearest tenth.
The mean mark is about 45.2.
The median mark is the mean of the 13th and 14th marks.
The 13th mark is 39. The 14th mark is 40.
So, the median is: $\frac{39 + 40}{2} = \frac{79}{2} = 39.5$
The mode is the mark that occurs most often. This is 29.
- Without the outlier, there are 25 marks and the sum of the marks is: $1175 - 99 = 1076$
The mean is: $1076 \div 25 = 43.04$
The median is the 13th mark: 39
The mode is 29.
When the outlier was removed, the mean and median decreased.
The mode remained the same.
- The outlier should be used when reporting the average test mark.
To understand how the class is performing, all test marks should be included.

Practice

- This set of data represents the waiting time, in minutes, at a fast-food restaurant:
5, 5, 5, 6, 5, 7, 0, 5, 1, 7, 7, 5, 6, 5, 5, 5, 8, 5, 0, 5, 4, 5, 2, 7, 9
 - Calculate the mean, median, and mode.
 - Identify the outliers. Explain your choice.
 - Calculate the mean, median, and mode without the outliers.
How is each average affected when the outliers are not included?
- Bryan recorded the time he spent on the school bus each day for one month. Here are the times, in minutes:
15, 21, 15, 15, 18, 19, 14, 20, 95, 18, 21, 14, 15, 20, 16, 14, 22, 21, 15, 19
 - Calculate the mean, median, and mode times.
 - Identify the outlier. How can you explain this time?
 - Calculate the mean, median, and mode times without the outlier.
How is each average affected when the outlier is not included?
 - A classmate asks Bryan, "What is the average time you spend on the bus each day?" How should Bryan answer? Give reasons.
- A clothing store carries pant sizes 28 to 46. A sales clerk records the sizes sold during her 6-h shift:
28, 36, 32, 32, 34, 4, 46, 44, 42, 38, 36, 36, 40, 32, 36
 - Calculate the mean, median, and mode sizes.
 - Is there an outlier? If so, why do you think it is an outlier?
 - Calculate the mean, median, and mode sizes without the outlier.
How is each average affected when the outlier is not included?
 - Should the outlier be used when the sales clerk reports the average pant size sold during her shift? Explain your thinking.
- Here are the science test marks out of 100 for the Grade 7 students in a combined-grades class:
0, 66, 65, 72, 78, 93, 82, 68, 64, 90, 65, 68
 - Calculate the mean, median, and mode marks.
 - Identify the outlier. How might you explain this mark?
 - Calculate the mean, median, and mode marks without the outlier.
How is each average affected when the outlier is not included?
 - Should the outlier be used when reporting the average test mark? Explain.

Remember to arrange the data in order before finding the median.



5. a) Give an example of a situation in which outliers would not be used in reporting the averages. Explain why they would not be included.
- b) Give an example of a situation in which outliers would be used in reporting the averages. Explain why they would be included.

6. **Assessment Focus** A Grade 7 class wanted to find out if a TV advertisement was true. The ad claimed that *Full of Raisins* cereal guaranteed an average of 23 raisins per cup of cereal. Each pair of students tested one box of cereal. Each box contained 20 cups of cereal. The number of raisins in each cup was counted.



- a) Assume the advertisement is true. How many raisins should there be in 1 box of cereal?
- b) Here are the results for the numbers of raisins in 15 boxes of cereal:
473, 485, 441, 437, 489, 471, 400, 453, 465, 413, 499, 428, 419, 477, 467
- Calculate the mean, median, and mode numbers of raisins.
 - Identify the outliers. Explain your choice.
 - Calculate the mean, median, and mode without the outliers. How do the outliers affect the mean?
 - Should the outliers be used when reporting the average number of raisins? Explain.
 - Was the advertisement true? Justify your answer.

7. **Take It Further** Here is a set of data: 2, 3, 5, 5, 7, 8
An outlier has been removed.

- a) Calculate the mean, median, and mode without the outlier.
- b) The outlier is returned to the set.
The averages become:
Mean: 7 Median: 5 Mode: 5
What is the outlier? Show your work.

Reflect

Your friend is having difficulty recognizing outliers in a data set. What advice would you give your friend?