# Mean, Median and Outliers

A short original statistics lesson | GitMind learning example

## 1. What the measures describe

The mean is the sum of all observations divided by the number of observations. Include every observation in both the total and the count. The mean uses each numerical value, so a very large or very small observation can change it substantially.

The median is the middle value after sorting the observations. With an odd number of observations, use the single middle value. With an even number, average the two middle values. Sort first; the middle position in an unsorted list is not a median.

## 2. A worked example

Six learners report these daily reading times, in minutes: 2, 3, 4, 5, 6, 10. The list is already sorted. The sum is 30, so the mean is 30 / 6 = 5 minutes. There are six values, so the median is (4 + 5) / 2 = 4.5 minutes.

## 3. When a value changes

Replace 10 with 40. The new list is 2, 3, 4, 5, 6, 40. The sum is 60 and the mean is 10 minutes. The two middle values are still 4 and 5, so the median remains 4.5 minutes. This illustrates sensitivity to an extreme observation; it does not establish that every large observation is an error.

Learning targets: calculate both measures; explain the effect of an extreme value; select a measure for a stated purpose; distinguish a suspicious record from a confirmed error.

Original educational example. May be reused for study and teaching.

# Interpretation and Practice

## 4. Choose a measure for a purpose

For the original data, both 5 minutes (mean) and 4.5 minutes (median) describe a center. After replacing 10 with 40, the median better reflects the middle of the group if the aim is to describe a typical learner. The mean is still useful when every value contributes to a total: mean multiplied by the number of observations recovers that total. Neither measure describes all differences between learners.

## 5. Check an unusual value

An outlier is an observation unusually far from the rest of a dataset in its context. Before removing a value, check units, transcription and the original record. A 40-minute reading time could be a real behavior. Removing it only because it changes the mean would hide valid information. This lesson uses an illustrative extreme value, not a formal outlier-detection rule.

## 6. Practice questions

A. For 3, 4, 4, 5, 9, calculate the mean and median.

B. For 2, 3, 4, 5, 6, 40, which measure would you choose to describe the middle learner, and why?

C. A classmate says that the median is always one of the observed values. Give a counterexample from page 1.

D. Should the 40-minute value automatically be deleted? State what you would check first.

## Reference answers

A: sum 25, count 5, mean 5, median 4. B: median 4.5 minutes for this purpose; the extreme value raises the mean. C: the even-sized page-1 dataset has median 4.5, which is not an observed value. D: no; check the source record and units before deciding whether it is an error.

Original educational example. May be reused for study and teaching.