

How to Calculate the Mean with Outliers

Authored by
stats writer

March 6, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Calculate the Mean with Outliers*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=134179>

Outliers are data points that significantly deviate from the overall trend or pattern of a data set. When calculating the mean, or average, of a set of data, the presence of outliers can greatly influence the value of the mean. This is because the mean is calculated by adding up all the data points and dividing by the total number of data points. Outliers, being extreme values, can drastically impact the overall sum and therefore skew the mean towards their direction. In other words, outliers can pull the mean towards their own value, making it an inaccurate representation of the central tendency of the data set. It is important to identify and handle outliers appropriately in order to obtain a more reliable and accurate mean value.

How Do Outliers Affect the Mean?

In statistics, the mean of a dataset is the average value. It's useful to know because it gives us an idea of where the "center" of the dataset is located. It is calculated using the simple formula:

mean = (sum of observations) / (number of observations)

For example, suppose we have the following dataset:

The mean of the dataset is $(1+4+5+6+7) / (5) = 4.6$

But while the mean is a useful and easy to calculate, it does have one drawback: It can be affected by outliers. In particular, the smaller the dataset, the more that an outlier could affect the mean.

To illustrate this, consider the following classic example:

Ten men are sitting in a bar. The average income of the ten men is \$50,000. Suddenly one man walks out and Bill Gates walks in. Now the average income of the ten men in the bar is \$40 million.

This example shows how one outlier (Bill Gates) could drastically affect the mean.

Small & Large Outliers

An outlier can affect the mean by being unusually small or unusually large. In the previous example, Bill Gates had an unusually large income, which caused the mean to be misleading.

However, an unusually small value can also affect the mean. To illustrate this, consider the following example:

Ten students take an exam and receive the following scores:

The mean score is 84.6.

However, if we remove the "0" score from the dataset, then the mean score becomes 94.

The one unusually low score of one student drags the mean down for the entire dataset.

Sample Size & Outliers

The smaller the sample size of the dataset, the more an outlier has the potential to affect the mean.

For example, suppose we have a dataset of 100 exam scores where all of the students scored at least a 90 or higher except for one student who scored a zero:

The mean turns out to be 93.18. If we removed the "0" from the dataset, the mean would be 94.12. This is a relatively small difference. This shows that even an extreme outlier only has a small effect if the dataset is large enough.

How to Handle Outliers

If you're worried that an outlier is present in your dataset, you have a few options:

Make sure the outlier is not the result of a data entry

error. Sometimes an individual simply enters the wrong data value when recording data. If an outlier is present, first verify that the value was entered correctly and that it wasn't an error. Assign a new value to the outlier. If the outlier turns out to be a result of a data entry error, you may decide to assign a new value to it such as of the dataset. Remove the outlier. If the value is a true outlier, you may choose to remove it if it will have a significant impact on your overall analysis. Just make sure to mention in your final report or analysis that you removed an outlier.

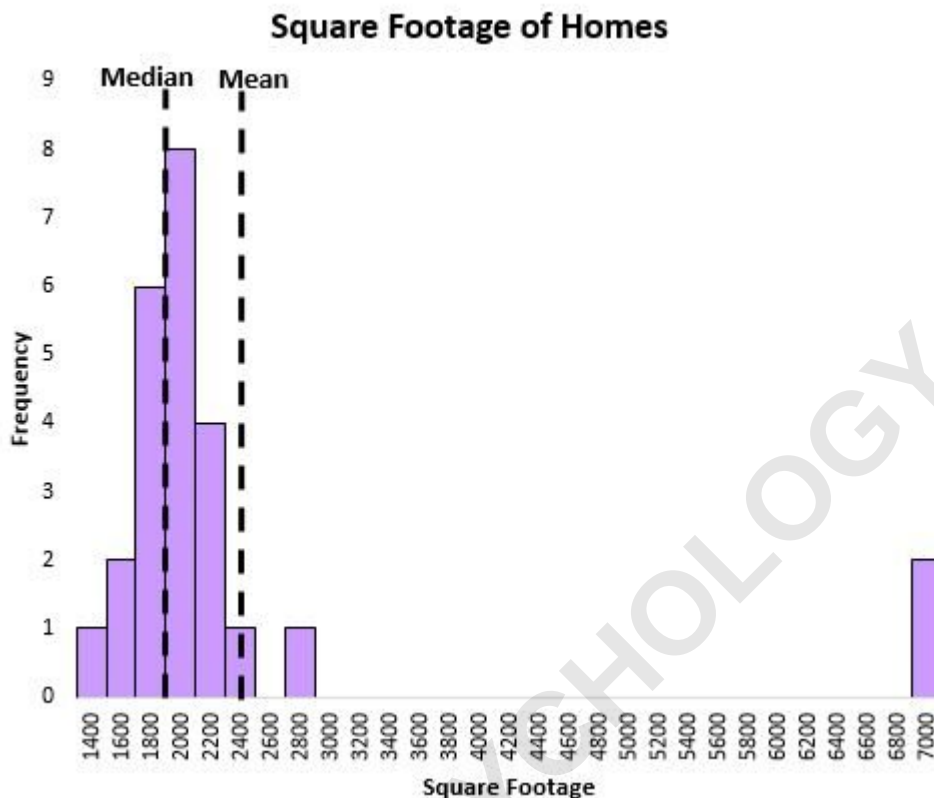
Use the Median

Another way to find the "center" of a dataset is to use the median, which is found by arranging all of the individual values in a dataset from smallest to largest and finding the middle value.

Because of the way it is calculated, the median is less affected by outliers and it does a better job of capturing the central location of a distribution when there are outliers present.

For example, consider the following chart that shows

the square footage of houses in a particular neighborhood:



The mean is heavily influenced by a couple extremely large houses, while the median is not. Thus, the median does a better job of capturing the "typical" square footage of a house in this neighborhood compared to the mean.

Further Reading: