



Calculate Median ^[1]

Assisted Self-Help ^[2]25.9K reads

The median is central to many experimental data sets, and to calculate median in such examples is important, by not falling into the trap of reporting the arithmetic mean.

The median can be seen to be the “middle value” of the distribution, i.e. it separates the upper and lower halves.

Calculation - Examples

To calculate median, consider the following example. Suppose we have the heights of different trees in a garden, and we need an “average” value for this. Say the heights in meters are 1.5, 6.9, 2.8, 1.8 and 2.3.

1. **Arrange order.** Arrange the numbers in either ascending or descending order, such as 1.5, 1.8, 2.3, 2.8 and 6.9.
2. **Choose middle value.** Choose the value which is exactly in the middle of the order. 1.5, 1.8, **2.3**, 2.8, 6.9

The median of this distribution is 2.3, which is the middle value, separating the lower half {1.5, 1.8} from the upper half {2.8, 6.9}.

Median = 2.3

3. **In case of two middle values - Use the Arithmetic Mean ^[3] on the two middle values**

Suppose there are six heights (New value: 1.2), rather than five:

1.2, 1.5, **1.8**, **2.3**, 2.8, 6.9

This leaves two middle values, {1.8, 2.3}. The median for this data set is $(1.8+2.3)/2 = 2.05$

Median = 2.05

One can immediately see that the data is skewed - the 6.9 meter tree makes it so. The arithmetic mean ^[3] of this data is 3.08 meters, which is more than 4 out of 5 data points. Thus the arithmetic mean doesn't make much sense in this case.

If the number of data points is even, unlike the example we previously considered, then to

calculate median we simply take the mean of the middle two elements. Thus if we have 10 numbers arranged in ascending order, the median is the average of the 5th and 6th numbers.

Another Example:

If the salaries of professional passing out of college in thousands of dollars per annum is 60, 64, 71, 73, 73, 77, 82, 85, 160 and 255, then their median salary is $(73+77)/2 = 75$. The mean [3] in this case is 100, which like the previous case, doesn't make much sense and doesn't really tell us about the central tendency [4] of the data.

In cases where the data is skewed, it is the median that makes sense, and not the mean. In these cases, as an experimenter, you need to calculate median [5] and not mean for your experiment. This is especially true in cases where there are outliers [6]. Many scientists calculate their results both in terms of median and mean, to see whether the outcome of their results [7] are the same.

The median is resistant to change with the discovery of outliers [6]. For example, if we want to know the mean weight of all the dinosaurs, it is a very difficult task because we do not yet know all the types of dinosaurs that ever walked the earth. Therefore if a new type of dinosaur bigger than all the others is discovered, it will significantly alter the mean.

However, the median remains almost unchanged in this case. Thus in many cases when the end points of the data set are not known, you need to calculate median and not the mean for that data set.

Source URL: <https://staging.explorables.com/en/calculate-median?gid=1588>

Links

[1] <https://staging.explorables.com/en/calculate-median>

[2] <https://staging.explorables.com/en>

[3] <https://staging.explorables.com/arithmatic-mean>

[4] <https://staging.explorables.com/measures-of-central-tendency>

[5] <http://en.wikipedia.org/wiki/Median>

[6] <https://staging.explorables.com/statistical-outliers>

[7] <https://staging.explorables.com/statistically-significant-results>