
Finding mean, median and mode in mathcad

[Descriptive statistics](#) is commonly used in the field of medical research studies. It is used to quantitatively describe the main features of a set of data. Inferential statistics differs in that it is used to reach conclusions that generalise beyond the immediate data. Descriptive statistics are used to present quantitative descriptions of large amounts of data in a clear and understandable way. [MathCAD](#) offers a variety of descriptive statistical functions to reduce large amounts of data into a much simpler summary. Descriptive statistics are generally presented along with more formal analyses, to give the audience an overall sense of the data being analysed.

I plan to go through the following basic functions:

1. `mean(A,B,C,...)`
2. `median(A,B,C,...)`
3. `mode(A,B,C,...)`
4. `percentile(v,p)`

These functions take single or multiple scalars or arrays, and return the mean, median, and mode, respectively, giving measures of the location of a data point relative to the rest of the distribution. The best choice of location estimator depends on the general dispersion or distribution of your data.

Mean

In statistics we refer to this also as the "arithmetic mean" and is the most commonly used type of average. To calculate the mean of a set of numbers this involves simply adding the total sum of the numbers in a set divided by the number of items in the set. The other types of averages such as the median and the mode will be discussed later on.

Example:

Consider the following numeric data:

$$x := \begin{pmatrix} 79 \\ 85 \\ 46 \\ 86 \\ 81 \end{pmatrix}$$

The arithmetic mean or average of N values is given by the following formula:

$$\frac{1}{N} \cdot \left(\sum_{i=0}^{N-1} x_i \right) = 75.4$$

`mean(x) = 75.4`

The mean is sensitive to changes in values of one or more data points:

$$x_2 := 1.2 \cdot x_2$$

`mean(x) = 77.24`

The mean is greatly affected by significant outliers. So you may find that the mean is a poor description of the central location if this is the case.

You may choose to trim the outliers and find the "trimmed mean" for a better estimate.

Consider the "trimmed" numeric data:

$$x := \begin{pmatrix} 79 \\ 85 \\ 86 \\ 81 \end{pmatrix}$$

As you can see I have chosen to leave out the value 46 which was a significant outlier in this set.

$$\frac{1}{N} \cdot \left(\sum_{i=0}^{N-1} x_i \right) = 82.75$$

MathCAD automatically readjusts all the values of the formulas and recalculate the new mean for this data set.

Median

The [median](#), or "middle value," of a set of data is another description of central location. The median depends on the relative positions of the data, not on the actual values of every data point, and so is relatively insensitive to small changes in individual data values. Mathcad's median function does not accept [complex numbers](#).

A median is the value falling in the middle when data are sorted in ascending order (smallest to largest).

Example:

$$\text{sort}(x) = \begin{pmatrix} 55.2 \\ 79 \\ 81 \\ 85 \\ 86 \end{pmatrix}$$

If there are an odd number of data, there is one data value which is the median.

e.g. if $a < b < c$ then the median for the list $\{a, b, c\}$ is b

Example:

$$a := \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$\text{median}(a) = 2$$

If there is an even number of data, the median is taken to be the mean of the two middle numbers.

e.g. if $a < b < c < d$ then the median for the list $\{a, b, c, d\}$ is the mean of b and c i.e. $(b+c)/2$.

Example:

$$a := \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$\text{median}(a) = 2.5$$

Mode

The `mode` refers to the value that occurs the most frequently in a data set.

Example:

In a data set where there are no repeated values, you see an error message like the one shown below.

$x := \begin{pmatrix} 79 \\ 85 \\ 46 \\ 86 \\ 81 \end{pmatrix}$

`mode(x) =`

No value occurs more frequently than any others.

Example:

In the case where more than one data value is repeated with the same frequency, the mode function also gives an error, with the message "multimodal."

$x := \begin{pmatrix} 78 & 75 \\ 79 & 78 \\ 81 & 80 \\ 79 & 77 \\ 75 & 76 \end{pmatrix}$

`mode(x) =`

Multimodal - more than one value occurs at the highest frequency.