STANDARD DEVIATION

Step 1: Find the mean.

Step 2: For each data point, find the square of its distance to the mean.

Step 3: Sum the values from Step 2.

Step 4: Divide by the number of data points.

Step 5: Take the square root.

$$\mathrm{SD} = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

MEAN

$$\text{Mean} = \frac{\text{Sum of all data values}}{\text{Number of data values}}$$

Symbolically,

$$\bar{x} = \frac{\sum x}{n}$$

where $\bar{x}$ (read as 'x bar') is the mean of the set of x values,

$\sum x$ is the sum of all the x values, and

n is the number of x values.

MEDIAN (n is odd)

The median tells you where the middle of a data set is.

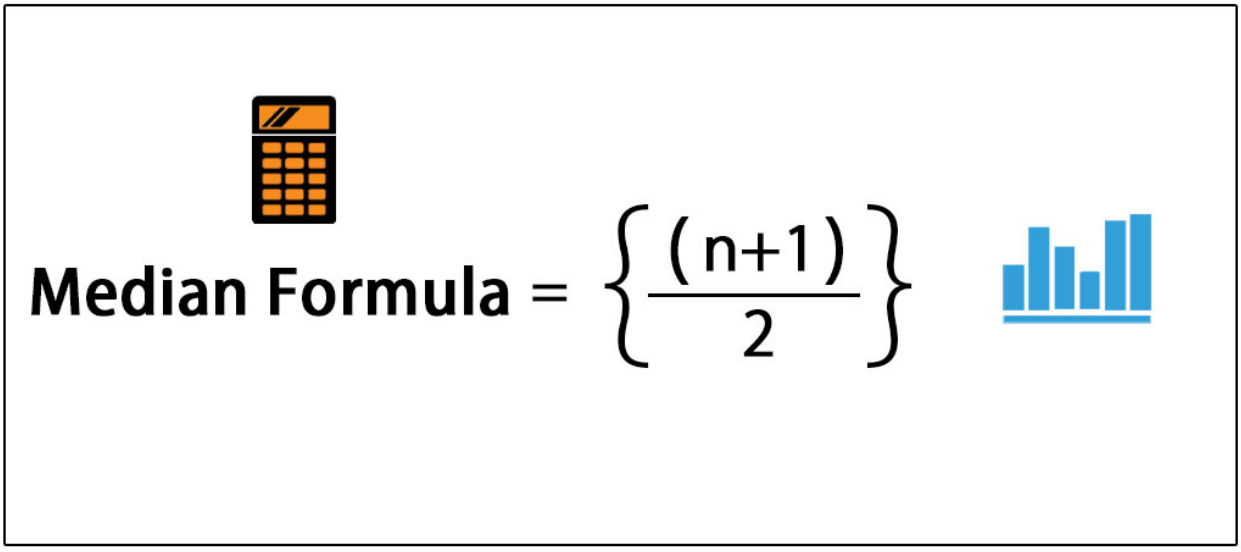
The median formula is {(n + 1) ÷ 2}th, where “n” is the number of items in the set and “th” just means the (n)th number.

To find the median, first order the numbers from smallest to largest. Then find the middle number. For example, the middle for this set of numbers is 5, because 5 is right in the middle: 1, 2, 3, 5, 6, 7, 9.

You get the same result with the formula. There are 7 numbers in the set, so n = 7:

{(7 + 1) ÷ 2}th
= {(8) ÷ 2}th
= {4}th

The 4th number in 1, 2, 3, 5, 6, 7, 9 is 5. If this were in a number array, the median number would be in the 4 - 1 index of the array.



Tip: If you have a large data set, divide the number in the set by 2. That tells you how many numbers should be above and how many numbers should be below. For example, 1001/2 = 500.5. Ignore the decimal; 500 numbers should be above and 500 below.

MEDIAN (n is even)

How to find the median for an even set of numbers:

Example question: Find the median for the following data set:
102, 56, 34, 99, 89, 101, 10, 54.

Step 1: Place the data in ascending order (smallest to highest).
10, 34, 54, 56, 89, 99, 101, 102.

Step 2: Find the TWO numbers in the middle (where there are an equal number of data points above and below the two middle numbers).
10, 34, 54, 56, 89, 99, 101, 102

Step 3: Add the two middle numbers and then divide by two, to get the average:

56 + 89 = 145
145 / 2 = 72.5.
The median is 72.5.

Tip: For large data sets, divide the number of items by 2, then subtract 1 to find the number that should be above and the number that should be below.

For example, if you have 1000 data values then: 1000/2 = 500. 500 – 1 = 499.
The middle two numbers will have 499 items above and 499 below.

Quartiles

Quartiles mark each 25% of a set of data:

- The first quartile Q_1 is the 25th percentile
- The second quartile Q_2 is the 50th percentile
- The third quartile Q_3 is the 75th percentile

The second quartile Q_2 is easy to find. It is the median of any data set and it divides an ordered data set into upper and lower halves.

The first quartile Q_1 is the median of the lower half not including the value of Q_2. The third quartile Q_3 is the median of the upper half not including the value of Q_2.

How to Calculate Quartiles

1. Order your data set from lowest to highest values
2. Find the median. This is the second quartile Q_2.
3. At Q_2 split the ordered data set into two halves.
4. The lower quartile Q_1 is the median of the lower half of the data.
5. The upper quartile Q_3 is the median of the upper half of the data.

If the size of the data set is odd, do not include the median when finding the first and third quartiles.

If the size of the data set is even, the median is the average of the middle 2 values in the data set. Add those 2 values, and then divide by 2. The median splits the data set into lower and upper halves and is the value of the second quartile Q_2.

Interquartile Range

The interquartile range IQR is the range in values from the first quartile Q_1 to the third quartile Q_3. Find the IQR by subtracting Q_1 from Q_3.

- $IQR = Q_3 - Q_1$

Minimum

The minimum is the smallest value in a sample data set.

Ordering a data set from lowest to highest value, $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$, the minimum is the smallest value x_1.

Maximum

The maximum is the largest value in a sample data set.

Ordering a data set from lowest to highest value, $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$, the maximum is the largest value x_n.

Range

The range of a data set is the difference between the minimum and maximum. To find the range, calculate x_n minus x_1.

COUNT

The number of rows of data that have non missing data values in them.