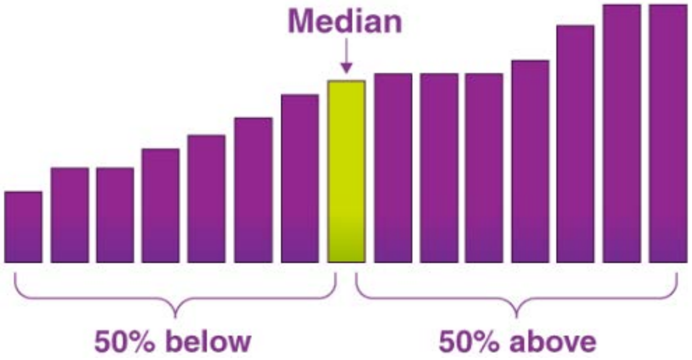
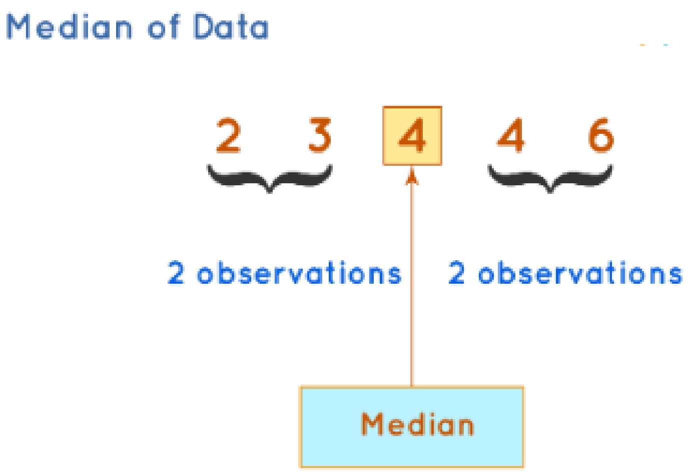
MEDIAN

By- Dr. Ekta Khare

Median

- Median, in statistics, is the middle value of the given list of data when arranged in an order. The arrangement of data or observations can be made either in ascending order or descending order.
- Example: The median of 2,3,4 is 3.
- Median is one of the three measures of central tendency.
- Among the statistical summary metrics, the median is an easy metric to calculate.
- Median is also called the **Place Average**, as the data placed in the middle of a sequence is taken as the median.





Median Formula

- The median formula is different for even and odd numbers of observations.

Odd Number of Observations

If the total number of observations given is odd, then the formula to calculate the median is:

$$Median = (\frac{n+1}{2})^{th}\, term$$

where n is the number of observations

Even Number of Observations

If the total number of observation is even, then the median formula is:

$$Median = \frac{(\frac{n}{2})^{th}\, term + (\frac{n}{2}+1)^{th}\, term}{2}$$

where n is the number of observations

Example: The age of the members of a weekend poker team has been listed below. Find the median of the above set.

{42, 40, 50, 60, 35, 58, 32}

Solution:

Step 1: Arrange the data items in ascending order.

Original set: {42, 40, 50, 60, 35, 58, 32}

Ordered Set: {32, 35, 40, 42, 50, 58, 60}

Step 2: Count the number of observations. If the number of observations is odd, then we will use the following formula: Median = $[(n + 1)/2]^{th}$ term

Step 3: Calculate the median using the formula.

Median = $[(n + 1)/2]^{th}$ term

= $(7 + 1)/2^{th}$ term = 4^{th} term = 42

Median = 42

Example: The height (in centimeters) of the members of a school football team have been listed below.

{142, 140, 130, 150, 160,135, 158,132}

Find the median of the above set.

Solution:

Step 1:

Arrange the data items in ascending order.

Original set: {142, 140, 130, 150, 160, 135, 158,132}

Ordered Set: {130, 132, 135, 140, 142, 150, 158, 160}

Step 2:

Count the number of observations.

Number of observations, n = 8

If number of observations is even, then we will use the following formula:

Median = $[(n/2)^{th}$ term + $((n/2) + 1)^{th}$ term$]/2$

Step 3:

Calculate the median using the formula.

Median = $[(n/2)^{th}$ term + $((n/2) + 1)^{th}$ term$]/2$

Median = $[(8/2)^{th}$ term + $((8/2) + 1)^{th}$ term$]/2$

= $(4^{th}$ term + 5^{th} term$)/2$

= (140 + 142)/2

= 141

Median of Grouped Data

To find the median class, we have to find the cumulative frequencies of all the classes and n/2. After that, locate the class whose cumulative frequency is greater than (nearest to) n/2. The class is called the median class.

After finding the median class, use the below formula to find the median value.

$$Median = l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$$

Where

l is the lower limit of the median class

n is the number of observations

f is the frequency of median class

h is the class size

cf is the cumulative frequency of class preceding the median class.

Median of Grouped Data Example

Example:

The following data represents the survey regarding the heights (in cm) of 51 girls of Class x. Find the median height.

Height (in cm)	Number of Girls
Less than 140	4
Less than 145	11
Less than 150	29
Less than 155	40
Less than 160	46
Less than 165	51

Solution:

To find the median height, first, we need to find the class intervals and their corresponding frequencies.

The given distribution is in the form of being less than type,145, 150 ...and 165 gives the upper limit. Thus, the class should be below 140, 140-145, 145-150, 150-155, 155-160 and 160-165.

From the given distribution, it is observed that,

4 girls are below 140. Therefore, the frequency of class intervals below 140 is 4.

11 girls are there with heights less than 145, and 4 girls with height less than 140

Hence, the frequency distribution for the class interval 140-145 = 11-4 = 7

Likewise, the frequency of 145 -150= 29 – 11 = 18

Frequency of 150-155 = 40-29 = 11

Frequency of 155 – 160 = 46-40 = 6

Frequency of 160-165 = 51-46 = 5

Therefore, the frequency distribution table along with the cumulative frequencies are given below:

Class Intervals	Frequency	Cumulative Frequency
Below 140	4	4
140 – 145	7	11
145 – 150	18	29
150 – 155	11	40
155 – 160	6	46
160 – 165	5	51

Here, n= 51.

Therefore, n/2 = 51/2 = 25.5

Thus, the observations lie between the class interval 145-150, which is called the median class.

Therefore,

Lower class limit = 145

Class size, h = 5

Frequency of the median class, f = 18

Cumulative frequency of the class preceding the median class, cf = 11.

We know that the formula to find the median of the grouped data is:

$$Median = l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$$

Now, substituting the values in the formula, we get

$$Median = 145 + \left(\frac{25.5 - 11}{18}\right) \times 5$$

Median = 145 + (72.5/18)

Median = 145 + 4.03

Median = 149.03.

Therefore, the median height for the given data is 149. 03 cm.

Practice Problems

1. The median of the following data set is 525. Find the values of x and y, if the total frequency is 100.

Class Interval	Frequency
0 – 100	2
100 – 200	5
200 – 300	x
300 – 400	12
400 – 500	17
500 – 600	20
600 – 700	y
700 – 800	9
800 – 900	7
900 – 1000	4

2. The following frequency distribution table shows the monthly consumption of electricity of 68 consumers of a locality. Find the median of the given data.

Monthly consumption of electricity (in units)	Number of consumers
65 – 85	4
85 – 105	5
105 – 125	13
125 – 145	20
145 – 165	14
165 – 185	8
185 – 205	4

When should you use the median?

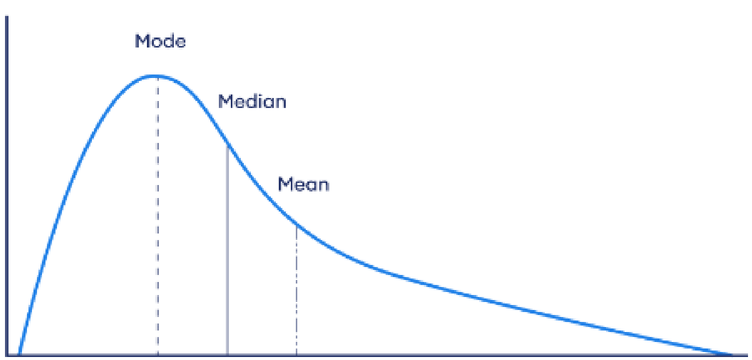
The median is the most informative measure of central tendency for skewed distributions or distributions with outliers.

In skewed distributions, more values fall on one side of the center than the other, and the mean, median and mode all differ from each other.

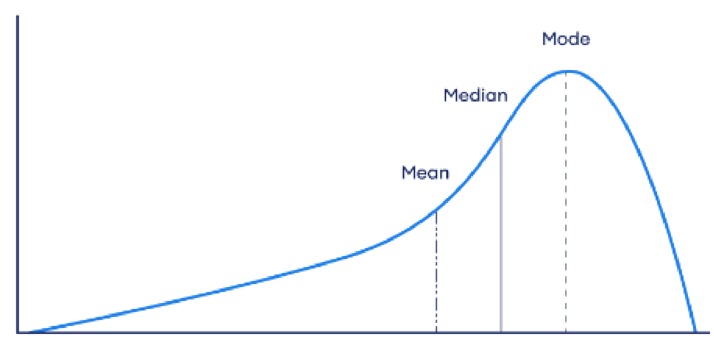
In a positively skewed distribution, there's a cluster of lower scores and a spread out tail on the right.

In a negatively skewed distribution, there's a cluster of higher scores and a spread out tail on the left.

Positively skewed distribution



Negatively skewed distribution



Because the median only uses one or two values from the middle of a dataset, it's unaffected by extreme outliers or non-symmetric distributions of scores. In contrast, the positions of the mean and mode can vary in skewed distributions.

For this reason, the median is often reported as a measure of central tendency for variables such as income, because these distributions are usually positively skewed.

The level of measurement of your variable also determines whether you can use the median. The median can only be used on data that can be ordered – that is, from ordinal, interval and ratio levels of measurement.