B. MEDIAN

The median by definition refers to the middle value in a distribution. In case of median one-half of the items in the distribution have a value the size of the median value or smaller and one-half have a value the size of the median value or larger. The median is just the 50th percentile value below which 50% of the values in the sample fall. It splits the observation into two halves.

As distinct from the arithmetic mean which is calculated from the *value of every* item in the series, the median is what is called a *positional* average. The term 'position' refers to the place of a value in a series. The place of the median in a series is such that an equal number of items lie on either side of it.

For example, if the income of five employees is Rs. 900, 950, 1020, 1200 and 1280, the median would be 1020.

900
950
1020 Value at middle position of the array
1200
1280

For the above example the calculation of median was simple because of odd number of observations. When an even number of observations are listed, there is no single middle position value and the median is taken to be the arithmetic mean of two middlemost items. For example, if in the above case we are given the income of six employers as 900, 950, 1020, 1200, 1280, 1300, the median income would be :

900
950
$\left.\begin{matrix}\mathbf{1020}\\ \mathbf{1200}\end{matrix}\right\} \leftarrow$ **there are two middle position values**
1280
1300

$$\text{Median} = \frac{1020+1200}{2} = \frac{2220}{2} = \mathbf{1110}$$

Hence, in case of even number of observations median may be found by averaging two middle position values.

Thus, when N is odd, the median is an actual value, with the remainder of the series in two equal parts on either side of it. If N is even, the median is a derived figure, *i.e.*, half the sum of the middle values.

Calculation of Median—Individual Observations

Steps. (*i*) Arrange the data in ascending or descending order of magnitude. (Both arrangements would give the same answer.)

(*ii*) In a group composed of an odd number of values such as 7, add 1 to the total number of values and divide by 2. Thus, 7+1 would be 8 which divided by 2 gives 4—the number of the value starting at either end of the numerically arranged groups will be the median value. In a large group the same method may be followed. In a group of 199 items the middle value would be 100th value. This would be determined by $\frac{199+1}{2}=100$. In the form of formula :

$$\text{Med.*}=\text{Size of } \frac{N+1}{2}\text{th item.}$$

Illustration 9. From the following data of the wages of 7 workers compute the median wage :

Wages (in Rs.)	1100	1150	1080	1120	1200	1160	1400

Solution :

CALCULATION OF MEDIAN

Sl. No.	*Wages arranged in ascending order*	*Sl. No.*	*Wages arranged in ascending order*
1	1080	5	1160
2	1100	6	1200
3	1120	7	1400
4	1150		

$$\text{Median}=\text{Size of } \frac{N+1}{2}\text{th item}=\frac{7+1}{2}=4\text{th item.}$$

Size of 4th item=1150 Hence the median wage=Rs. 1150

We thus find that median is the middlemost item : 3 persons get a wage less than Rs. 1150 and equal number, *i.e.*, 3, get more than Rs. 1150.

The procedure for determining the median of an even-numbered group of items is not as obvious as above. If there were, for instance, different values in a group, the median is really not determinable since both the 5th and 6th values are in the centre. In practice, the median value for a group composed of an even number of items is estimated by finding the arithmetic mean of the two middle values—that is, adding the two values in the middle and dividing by two. Expressed in the form of formula, it amounts to :

$$\text{Median}=\text{Size of } \frac{N+1}{2}\text{ th item}$$

*The abbreviation Med. represents median.

Thus we find that it is both when N is odd as well as even that 1 (one) has to be added to determine median value.

Illustration 10. Obtain the value of median from the following data :

391, 384, 591, 407, 672, 522, 777, 753, 2,488 1,490

Solution. CALCULATION OF MEDIAN

Sl. No.	*Data arranged in ascending order X*	*Sl. No.*	*Data arranged in ascending order X*
1	384	6	672
2	391	7	753
3	405	8	777
4	522	9	1,490
5	591	10	2,488

Median=Size of $\frac{N+1}{2}$th item= $\frac{11}{2}$ =5·5th item

Size of 5·5th item= $\frac{\text{5th item}+\text{6th item}}{2}=\frac{591+672}{2}=\frac{1,263}{2}=631\cdot5$

Computation of Median—Discrete Series

Steps. (*i*) Arrange the data in ascending or descending order of magnitude.

(*ii*) Find out the cumulative frequencies.

(*iii*) Apply the formula : Median=Size of $\frac{N+1}{2}$.

(*iv*) Now look at the cumulative frequency column and find that total which is either equal to $\frac{N+1}{2}$ or next higher to that and determine the value of the variable corresponding to it. That gives the value of median.

Illustration 11. From the following data find the value of median :

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Solution : CALCULATION OF MEDIAN

Income arranged in ascending order	*No. of persons f*	*c.f.*	*Income arranged in ascending order*	*No. of persons f*	*c.f.*
800	16	16	1800	30	96
1000	24	40	2000	20	116
1500	26	66	2500	6	122

Median=Size of $\frac{N+1}{2}$th item = $\frac{122+1}{2}$ =61·5th item.

Size of 61·5th item=1500.

Calculation of Median—Continuous Series

Steps. Determine the particular class in which the value of median lies. Use $N/2$ as the rank of the median not $(N+1)/2$. Some writers have suggested that while calculating median in continuous series 1 should be added to total frequency if it is odd (say 99) and should not be added if it is even figure (say, 100). However, 1 is to be added in case of individual and discrete series because specific items and individual values are involved. In a continuous frequency distribution all the frequencies lose their individuality. The effort now is not to find the value of one specific item but to find a particular point on a curve—that one value which will have 50 per cent of frequencies on one side of it and 50 per cent of the frequencies on the other. It will be wrong to use the above rule. Hence it is $N/2$ which will divide the area of curve into two equal parts and as such we should use $N/2$ instead of $(N+1)/2$, in continuous series. After ascertaining the class in which median lies, the following formula is used for determining the exact value of median.

$$\text{Median}=L+\frac{N/2-c.f.}{f}\times i$$

L=Lower limit of the median class, *i.e.*, the class in which the middle item of the distribution lies.

$c.f.$=Cumulative frequency of the class preceding the median class or sum of the frequencies of all classes lower than the median class.

f=Simple frequency of the median class.

i=The class interval of the median class.

It should be remembered that while interpolating the median value in a frequency distribution it is assumed that the variable is continuous and that there is an orderly and even distribution of items within each class.

Illustration 12. (*a*) Calculate the median for the following frequency distribution :

Marks	*No. of students*	*Marks*	*No. of students*
45—50	10	20—25	31
40—45	15	15—20	24
35—40	26	10—15	15
30—35	30	5—10	7
25—30	42		

[*B. Com., Madras Univ., 1986*]

Solution. First arrange the data in ascending order and then find out median.

CALCULATION OF MEDIAN

Marks	*f*	*c.f.*	*Marks*	*f*	*c.f.*
5—10	7	7	30—35	30	149
10—15	15	22	35—40	26	175
15—20	24	46	40—45	15	190
20—25	31	77	45—50	10	200
25—30	42	119			

Med.=Size of $\frac{N}{2}$th item$=\frac{200}{2}$=100th item

Illustration 15. Find the missing frequency in the following distribution if N is 100 and median is 30.

Marks	*No. of students*	*Marks*	*No. of students*
0—10	10	30—40	30
10—20	?	40—50	?
20—30	25	50—60	10

(M. Com. HPU. 1985)

Solution. CALCULATION OF MISSING FREQUENCIES

Let the frequencies of the class 10—20 be X

Frequency of the class 40—50 $= 100-(75+X)$

or $25-X$ [75 is the total of $10+25+30+10$]

Marks	*No. of Students*	*c.f.*
0—10	10	10
10—20	X	$10+X$
20—30	25	$35+X$
30—40	30	$65+X$
40—50	$(25-X)$	90
50—60	10	100

$$\text{Median} = L + \frac{N/2 - c.f.}{f} \times i$$

Since median is 30, it lies in the class 30—40*

$$L=30,\ N/2=50,\ c.f.=(35+X),\ f=30.$$

$$30 = 30 + \frac{50-(35+X)}{30} \times 10 = \frac{50-35-X}{30}\ ; \text{ multiplying by 3}$$

or $\quad 90=90+50-35-X=30 \quad$ or $\quad X=15$

Hence the frequency of the class corresponding to 10—20=15 and the frequency of the class corresponding to 40—50 $=25-X=25-15=10$.

Calculation of Median when Class Intervals are Unequal

When the class intervals are unequal, the frequencies need not be adjusted to make the class intervals equal and the same formula for interpolation can be applied as discussed above.

Illustration 16. Calculate median from the following data :

Marks	0—10	10—30	30—60	60—80	80—90
No. of Students	5	15	30	8	2

Solution. Since class intervals are unequal, let us first convert it to a distribution with equal class intervals on the assumption that the frequencies are equally distributed throughout a class.

CALCULATION OF MEDIAN

Marks	*f*	*c.f.*	*Marks*	*f*	*c.f.*
0—10	5	5	60—80	8	58
10—30	15	20	80—90	2	60
30—60	30	50			

* ∵ The class 20—30 would cover up to 29·999 and 30 would be covered in the class 30—40.

$$\text{Med.} = \text{Size of } \frac{N}{2}\text{th item} = \frac{60}{2} = 30\text{th item}$$

Median lies in the class 30—60.

$$\text{Med.} = L + \frac{N/2 - c.f.}{f} \times i$$

$$L=30,\ N/2=30,\ c.f.=20,\ f=30,\ i=30$$

$$\text{Med.} = 30 + \frac{30-20}{30} \times 30 = 30 + 10 = 40$$

If we make the class intervals equal, the same answer will be obtained.

Marks	*f*	*c.f.*	*Marks*	*f*	*c.f.*
0—10	5	5	50—60	10	50
10—20	7·5	12·5	60—70	4	54
20—30	7·5	20	70—80	4	58
30—40	10	30	80—90	2	60
40—50	10	40			

$$\text{Med.} = \text{Size of } \frac{N}{2}\text{th item} = \frac{60}{2}\text{th} = 30 \text{ item.}$$

Median lies in the class 30—40.

$$\text{Med.} = L + \frac{N/2 - c.f.}{f} \times i$$

$$L=30,\ N/2=30,\ c.f.=20,\ f=10,\ i=10$$

$$\text{Med.} = 30 + \frac{30-20}{10} \times 10 = 30 + 10 = 40.$$

Mathematical Property of Median I. mp

The sum of the deviations of the items from median, ignoring signs, is the least. For example, the median of 4, 6, 8, 10, 12 is 8. The deviations from 8 ignoring signs are 4, 2, 0, 2, 4 and the total is 12. This total is smaller than the one obtained if deviations are taken from any other value. Thus if deviations are taken from 7, values ignoring signs would be 3, 1, 1, 3, 5, and the total 13.

Merits and Limitations of Median

Merits. 1. It is especially useful in case of open-end classes since only the position and not the values of items must be known. The median is also recommended if the distribution has unequal classes, since it is easier to compute than the mean.

2. Extreme values do not affect the median as strongly as they do the mean. For example, the median of 10, 20, 30, 40 and 150 would be 30 whereas the mean 50. Hence very often when extreme values are present in a set of observations, the median is a more satisfactory measure of the central tendency than the mean.

3. In markedly skewed distributions such as income distributions or price distributions where the arithmetic mean would be distorted by extreme values, the median is especially useful. Consequently, the median income for some purposes be regarded as a more representative figure, for half the income earners must be receiving at least the median income. One can say as many receive the median income and as many do not.

4. It is the most appropriate average in dealing with qualitative data, *i.e.*, where ranks are given or there are other types of items that are not counted or measured but are scored.

5. The value of median can be determined graphically whereas the value of mean cannot be graphically ascertained.

6. Perhaps the greatest advantage of median is, however, the fact that the median actually does indicate what many people incorrectly believe the arithmetic mean indicates. The median indicates the value of the middle item in the distribution. This is a clear-cut meaning and makes the median a measure that can be easily explained

Limitations. 1. For calculating median it is necessary to arrange the data ; other averages do not need any arrangement.

2. Since it is a positional average, its value is not determined by each and every observation.

3. It is not capable of algebraic treatment. For example, median cannot be used for determining the combined median of two or more groups as is possible in case of mean. Similarly, the median wage of a skewed distribution times the number of workers will not give the total payroll. Because of this limitation the median is much less popular as compared to the arithmetic mean.

4. The value of median is affected more by sampling fluctuations than the value of the arithmetic mean.

5. The median, in some cases, cannot be computed exactly as the mean. When the number of items included in a series of data is even, the median is determined approximately as the mid-point of the two middle items.

6. It is erratic if the number of items is small.

Usefulness. The median is useful for distributions containing open-end intervals since these intervals do not enter its computation. Also since the median is affected by the number rather than the size of items, it is frequently used instead of the mean as a measure of central tendency in cases where such values are likely to distort the mean.

Related Positional Measures

Besides median, there are other measures which divide a series into equal parts. Important amongst these are quartiles, deciles and percentiles. Quartiles are those values of the variate which divide the total frequency into four equal parts, deciles divide the total frequency into 10 equal parts and the percentiles divide the total frequency into 100 equal parts. Just a one point divides a series into two parts, three points would divide it into four parts, 9 points into 10 parts and 99 points into 100 parts. Consequently, there are only 3 quartiles, 9 deciles and 99 percentiles for a series. The quartiles are denoted by symbol Q, deciles by D and percentiles by P. The subscripts 1, 2, 3, etc., beneath Q, D, etc., would refer to the particular value that we want to compute. Thus Q_1 would denote first quartile, Q_2 second quartile, Q_3 third quartile, D_1 first decile, D_8 8th decile, P_1 first percentile and P_{60} 60th percentile, etc.

Graphically any set of these partition values divides the area of the frequency curve or histogram into equal parts. If vertical lines are drawn