LECTURE 21

MEDIAN

PROF.SHRUTI

Introduction and definition

Median is the value of the middle item of a series arranged in ascending or descending order of magnitude. Thus, if there are 9 items in a series arranged in ascending or descending order of magnitude, median will be the value of the 5^{th} item. This item would divide the series in two equal parts—one part containing values less than the median value and the other part containing values more than the median value. If, however, there are even number of items in a series, then there is no central item dividing the series in two equal parts. For example if there are 10 items in a series, the median value would be between the values of 5th and 6th items. It would, thus, be the arithmetic average of the values of 5th and 6th items or it would be equal to the value of the 5th item plus the value of the 6th item divided by two.

According to **A.L. Bowley,** "If the numbers of the group are ranked in order according to the measurement under consideration, then the measurement of the number most nearly one half is the median."

The above definitions of the median do not hold good in situations where a median value is surrounded by neighbouring values which are equal in magnitude to it. For example, in a series of values such as 12, 13, 14, 15, 16, 17 and 18, there is no value which is so located that three values are smaller than it and three are greater than it. However, value 15 is designated as median. Keeping in view such situations, **Croxton** and **Cowdon** have given a revised definition of median as, "The median is that value which divides a series so that one half or more of the items are equal to or less than it and one half or more of the items are equal to or greater than it."

Calculation of Median

The calculation of median involves two basic steps, viz. (i) the location of the middle item and (ii) finding out its value.

The middle item in series of individual observations and also in a discrete series is $\left(\frac{n+1}{2}\right)^{th}$ item, where n is the total number of observations. In case of a continuous series $\left(\frac{n}{2}\right)^{th}$ item is the middle item of the series.

Once the middle item is located, its values has to be found out. In a series of individual observations, if the total number of items is an odd figure, the value of the middle item is the median value. If the number of items is even, the median value is the average of the two items in the centre of the distribution. The examples given below would clarify these points.

Computation of Median in a series of individual observations

Example 3. Find out the median of the following items :

5, 7, 9, 12, 10, 8, 7, 15, 21

Solution. These items would first be arranged in ascending order of magnitude. The series then would he as follows:

Calculation of Median

Serial Number	Size of items
1	5
2	7
3	7
4	8
5	9
6	10
7	12

8	15
9	21

If M represents the median and N the number of items.

$$M = \text{Size of } \left(\frac{n+1}{2}\right)^{th} \text{ item} = \text{Size of } \left(\frac{9+1}{2}\right)^{th} \text{ or } 5^{th} \text{ item} = 9.$$

In the above example, the number of items was odd and there was no difficulty in locating the middle items and its value. If the number of items is even, the middle item and its value would be calculated as illustrated in the following example.

Example 4. Find out the value of median from the following data :

Daily wages (R)	10	5	7	11	8
Number of Workers	15	20	15	18	12

Solution

Calculation of median

Wages in ascend-ing order (Rs.)	Number of persons (f)	Cumulative Frequency (c.f.)
5	20	20
7	15	35
8	12	47
10	15	62
11	18	80

Median is the value of $\left(\frac{N+1}{2}\right)^{th}$ or $\left(\frac{80+1}{2}\right)^{th}$ 40.5^{th} items. All items from 35 onwards upto 47 have a value of 8. Thus, the median value would be Rs. 8.

Continuous series descending order : In such series, there is a slight change in the formula for calculating median. Since the series are cumulated in descending order, the cumulative frequency of the class preceding the median class is found out by subtracting the cumulative frequency of the median class from the total of the cumulative frequency. In other words :

$$c = (N - c.f.)$$

where c = cumulative frequency less of the class preceding the median class

N = total cumulative frequency

cf = cumulative frequency of the median class.

The following example would illustrate the point :

Example 5. Calculate median from the following data :

Age	**Number of persons**	**Age**	**Number of persons**
55—60	7	35—40	30
50—55	13	30—35	33
45—50	15	25—30	28
40—45	20	20—25	14
			Total 160

Solution

Computation of Median

Age	Number of persons	Cumulative frequency
55—60	7	7
50—55	13	20
45—50	15	35
40—45	20	55
35—40	30	85
30—35	33	118
25—30	28	146
20—25	14	160

In the above example, median is the value of $\left(\frac{N}{2}\right)^{th}$ or $\left(\frac{160}{2}\right)^{th}$ or 80^{th} item which lies in 35—40 class interval.

The frequency of the class preceding the median class or the value of c = Total frequency minus the cumulative frequency of the median class or

160—85 = 75

$$\text{Median } M = l_1 + \frac{l_2 - l_1}{f_1}(m - c) = 35 + \frac{40 - 35}{30}(80 - 75) = 35 + \left(\frac{5}{30} \times 5\right) = 35.83$$

Example 6. Compute median from the following data :

Mid-values:	115	125	135	145	155	165	175	185	195
Frequency:	6	25	48	72	116	60	38	22	3

Solution. Here, we are given the mid-values of the class-intervals of a continuous frequency distribution. The difference between two mid-values is 10, hence, 10/2 = 5 is reduced from each mid-value to find the lower limit and the same is added to find the upper limit of a class. The classes are, thus, 110–120, 120–130 ..... and so on upto 190–200.

Computation of Median

Class-intervals	**Frequency**	**Cumulative frequency**
110—120	6	6
120—130	25	31
130—140	48	79
140—150	72	151
150—160	116	267
160—170	60	327
170—180	38	365
180—190	22	387
190—200	3	390
Total	**390**	

The middle item is $\frac{390}{2}$ or 195 which lies in the 150—160 group.

$$M = l_1 + \frac{l_2 - l_1}{f_1}(m - c)$$

$$= 150 + \frac{160 - 150}{116}(195 - 151) = 150 + \frac{10}{116}(44)$$

$$= 153.8$$

Summary:

Median:

The arithmetic mean is unsuccessful when distribution has extremely high or low values; in such condition it will either over estimate or under estimate. The both conditions are not considerable. A.M. is not possible to calculate, while the distribution has open end class. Due to these circumstances median is best suitable measure. Median of the distributions the value of the variable which partitions its into two equal halves. Median is a positional average. Mathematically it is indicated as M_d.

- For ungrouped data, let $x_1, x_2, \ldots, x_n$ be n values, If n is odd, median is $(n/2)^{th}$ value but If n is even, median is $\{(n+1)/2\}^{th}$ value.
- For grouped data or in case of frequency distribution x_i/f_i, if $x_1, x_2, \ldots, x_n$ be n values of any variable X with corresponding frequencies $f_1, f_2, \ldots, f_n$, then median is,

$$M_d = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

Where, the class which is corresponding to the cumulative frequency (*cf*) just greater than *N/2* is known as the *median class*, l is lower limit of median class, f frequency of median class, h is magnitude of median class and c is cumulative frequency just preceding the median class.

Advantages of Median: An important advantage of median is that it is less sensitive than the mean to extreme scores. For skewed data, the median is a better choice because it is usually not affected by a few outlier. Median is also a desirable measure when the distribution has to be truncated for some reasons. If the purpose is to describe the central tendency of a set of scores, the median is preferable to other measures. It gives an undistorted picture of central tendency whether the data are skewed or not.

Disadvantages of Median: Under usual circumstances the median is more vulnerable to sampling variable than the arithmetic mean. This makes median less stable than the mean from sample to sample and therefore it is not very useful in inferential statistics. For ordinal data median also ignores the actual values of observations and simply takes into account their positions.

Partition values:

Partition values are those values, which separates the entire data into number of equivalent parts. For mathematical calculations the given table gives an idea.

Median (Two equal parts)	One point	N/2	$M_d = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$
Quartiles **Q_i** (Four equal parts)	Q_1, Q_2, Q_3 (Three points)	iN/4 where i=1,2,3	$Q_i = l + \frac{h}{f}\left(\frac{iN}{4} - c\right)$
Deciles **D_i** (Ten equal parts)	(D_1, D_2, D_3,…D_9) (Nine points)	iN/10 where i=1,2,3,…, 8,9	$D_i = l + \frac{h}{f}\left(\frac{iN}{10} - c\right)$
Percentile **P_i**	(P_1, P_2, P_3,…..P_{99})	iN/100	$P_i = l + \frac{h}{f}\left(\frac{iN}{100} - c\right)$

(Hundred equal parts)	(Ninety Nine points)	where $i=1,2,3,\ldots\ldots,98,99$	