CHAPTER
MEASURES OF CENTRAL TENDENCY

Meaning:

In the study of a population with respect to one in which we are interested we may get a large number of observations. It is not possible to grasp any idea about the characteristic when we look at all the observations. So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic. Such representative number can be a central value for all these observations. This central value is called a measure of central tendency or an average. Measures of Central Tendency are also called the measures of location since they enable us to locate the position or place of the distribution. The general idea behind this measure of central tendency is that to look for a common measure that best describe or represents the characteristics of the entire group. This typical central value is a focal point around which large amount of data try to concentrate.

There are five averages. Among them mean, median and mode are called simple averages and the other two averages geometric mean and harmonic mean are called special averages. The meaning of average is nicely given in the following definitions.

"A measure of central tendency is a typical value around which other figures congregate."
"An average stands for the whole group of which it forms a part yet represents the whole."
"One of the most widely used set of summary figures is known as measures of location."

Requisites for a good or an ideal average:

The following properties should possess for a good average:

1. It should be rigidly defined.
2. It should be easy to understand and compute.
3. It should be based on all observations.
4. Its definition shall be in the form of a mathematical formula.
5. It should be suitable for further mathematical treatment.
6. It should be fluctuated least from sample to sample drawn from the same population.
7. It should be least affected by the extreme observations.
8. A good average should represent maximum characteristics of the data, its value should be nearest to the most items of the given series.

Types of Measures of Central Tendency

There are five types of the most commonly used measures of Central tendency which are as follows:

1. Arithmetic mean (A.M.)
   (i) Simple Arithmetic mean
   (ii) Weighted Arithmetic mean
2. Median (M_d)
3. Mode (M_o)
4. Geometric Mean (G.M.)
5. Harmonic Mean (H.M.)

Arithmetic mean or mean:

Arithmetic mean or simply the mean of a variable is defined as the sum of all observations divided by total number of observations. It is the most popular and widely applicable measures of central Tendency. It can be divided into two Categories:

- Simple Arithmetic mean
- Weighted Arithmetic mean

Simple Arithmetic mean:

The sum of all the observations divided by the number of observations is called simple arithmetic mean. In simple arithmetic mean, all the observations or items are equally important.

(a) **A.M for individual series:** Let x_1 x_2 .... x_n be a variate values of the variable X, then their arithmetic mean (A.M.) is defined as

A.M. $(\bar{x}) = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$ or simply $\frac{\sum x}{n}$ (**Direct method**)

Where, $\sum xi = \sum x$ = Sum of all observations.

n = Total number of observations.

Let, $d = x - A$ then

$\boxed{\text{A.M } (\bar{x}) = A + \frac{\sum d}{n}}$ ` (**Deviation Method**)

Where, A = Assumed mean

d = Deviation from assumed mean A.

(b) **A.M for discrete series:** Let x_1 x_2 ....x_n be the variate values of the variable X with their respective frequencies $f_1, f_2, \ldots, f_n$ then their A.M. is defined as

A.M $(\bar{x}) = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_n x_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum f_i x_i}{N}$ or simply $\frac{\sum fx}{N}$ (**Direct method**)

Let $d = x - A$ then

$A.M.(\bar{X}) = A + \frac{\sum fd}{N}$ (**Deviation Method**)

Where, A = Assumed mean

d = Deviation from the assumed mean A.

(c) **A.M for continuous series:** Let x_1 x_2 .... x_n be the mid-value of the continuous variable x with their respective frequencies $f_1, f_2, \ldots, f_n$ then

A.M. $(\bar{x}) = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_n x_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum f_i x_i}{N}$ or simply $\frac{\sum fx}{N}$ (**Direct method**)

Where $\sum f$ = Total frequency = N

Let $d' = \frac{x - A}{h}$, Where h = Class - interval.

Then, A.M. $(\bar{x}) = A + \frac{\sum fd'}{N} \times h$ (**Step – deviation method**)

Weight Arithmetic Mean:

For calculating simple arithmetic mean, we suppose that all the values or the sizes of items in the distribution have equal importance. But, in practical life this may not be always true. In case some items are more important than others, a simple average computed is not representative of the distribution. Proper weightage has to be given to the various items. . For example, to have an idea of the change in cost of living of a certain group of persons, the simple average of the prices of the commodities consumed by them will not do because all the commodities are not equally important, e.g. rice, wheat and pulses are more important than tea, confectionery etc., It is the weighted arithmetic average which helps in finding out the average value of the series after giving proper weight to each group.

Let $x_1, x_2, \ldots, x_n$ be a variate values with their respective weight $w_1, w_2, \ldots, w_n$ then their weighted arithmetic mean is denoted by $\bar{x}_w$ and is defined as

$$\bar{x}_w = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_n x_n}{w_1 + w_2 + \ldots + w_n}$$

$$\boxed{\bar{x}_w = \frac{\sum wx}{\sum w}}$$

Example: Calculate the mean for 2, 4, 6, 8, 10.

Solution:

$$\textbf{Mean} = \bar{x} = \frac{2+4+6+8+10}{5} = 6$$

Example: A student's marks in 5 subjects are 75, 68, 80, 92, and 56. Find his average mark using deviation method.

Solution:

Let X be the marks, and assumed mean (A) = 68

Calculation of Mean

X	d = x – A (68)
75	7
68	0
80	12
92	24
56	-12
Total	**31**

Here, n = 5, A = 68,

We know $\boxed{\text{A.M}\ (\bar{x}) = A + \frac{\sum d}{n}}$`

$= 68 + \frac{31}{5}$

$= 68 + 6.2$

$= 74.2$

Hence, mean marks = 74.2

Example: From the following data, determine the average age by using deviation method.

Age(years)	10	20	30	40	50	60	70
No. of People	20	15	12	10	4	8	6

Solution:

Let X be the age of the people and 40 be the assumed mean i.e. A.

Calculation of average age

Age (X)	d = x –A(40)	f	fd
10	–30	20	–600
20	–20	15	–300
30	–10	12	–120
40	0	10	0
50	10	4	40
60	20	8	160
70	30	6	180
		N = 75	$\sum$fd = –640

We know, $A.M.(\bar{X}) = A + \frac{\sum fd}{N}$

$\therefore$ A.M $(\bar{x}) = 40 + \frac{(-640)}{75} = 40 - 8.53$

A.M $(\bar{x}) = 31.47$

Hence, average age = 31.47 years

Example: Data below represents the wages (Rs) received by workers in construction site. Calculate the average wages using step deviation method.

Wages:	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70
No. of workers:	10	13	18	21	24	28	20	11	8

Solution: Calculation of average wages using changing origin and scale of data

Wages (Rs)	Mid-Value (X)	$d' = \frac{x-47.5}{5}$	f	fd'
25-30	27.5	–4	10	–40
30-35	32.5	–3	13	–39
35-40	37.5	–2	18	–36
40-45	42.5	–1	21	–21

45-50	47.5	0	24	0
50-55	52.5	+1	28	+28
55-60	57.5	+2	20	+40
60-65	62.5	+3	11	+33
65-70	67.5	+4	8	+32
			N = 153	$\sum fd' = -3$

We know, A.M. $(\bar{x}) = A + \frac{\sum fd'}{N} \times h$, where, A = 47.5, h = 5,

$$= 47.5 + \frac{(-3)}{153} \times 5 = 47.40$$

Hence, mean wages is Rs 47.40.

Example : Given the following frequency distribution, calculate the arithmetic mean.

Wages (Rs)	64	63	62	61	60	59
Number of People	8	18	12	9	7	6

Solution:

Let X be the wages and f be the number of students.

x	f	fx	d=x-A(62)	fd
64	8	512	2	16
63	18	1134	1	18
62	12	744	0	0
61	9	549	-1	-9
60	7	420	-2	-14
59	6	354	-3	-18
	N=60	**3713**		**-7**

Direct method

$A.M.(\bar{X}) = \frac{\sum fx}{N} = \frac{3713}{60} = 61.88$

Short-cut method

$A.M.(\bar{X}) = A + \frac{\sum fd}{N} = 61.88$

Example: For a certain frequency table of number of accident and number of days which is only partly reproduced here, the average number of accidents was found to be 1.46.

Number of accident	Number of days
0	46
1	?
2	?
3	25
4	10
5	5

Calculate the missing frequencies. Where N = 200.

Solution:

Computation of missing frequencies

No. of accidents (X)	No. of days (f)	fx
0	46	0
1	f_1	f_1
2	f_2	$2f_2$
3	25	75
4	10	40
5	5	25
	$N = 86 + f_1 + f_2$	$\sum fx = 140 + f_1 + 2f_2$

$\therefore$ Here, $N = 86 + f_1 + f_2 = 114$

Or, $200 = 86 + f_1 + f_2$

Or, $f_1 + f_2 = 114$

Or, $f_1 = 114 - f_2$ .... (i)

Now, $\bar{x} = \frac{\sum fx}{N} \Rightarrow 1.46 = \frac{140 + f_1 + 2f_2}{200}$

Or, $140 + f_1 + 2f_2 = 200 \times 1.46$

$\therefore$ $f_1 + 2f_2 = 152$ .... (ii)

Putting the value of f_1 from (i) to (ii)

We get, $114 - f_2 + 2f_2 = 152$

Or $f_2 = 38$

Putting the value of f_2 in equation (i),

We get, $f_1 = 114 - 38 = 76$

Hence, $f_1 = 76$, $f_2 = 38$

Example: The mean of 200 items was 50. Later on it was discovered that two items were misread as 92 and 8 instead of 192 and 88. Find out the correct mean.

Solution: We have given,

Incorrect mean ($\bar{x}$) = 50, and Number of observations (n) = 200

We know, $\bar{x} = \frac{\sum x}{n} \Rightarrow \sum x = n\bar{x} = 200 \times 50$

$\sum x = 10000$

Now,

Correct $\sum x = 10000 + 192 + 88 - 92 - 8$

Correct $\sum x = 10180$

$\therefore$ Correct $\bar{x} = \frac{\text{Correct } \sum x}{n} = \frac{10180}{200}$

Correct $\bar{x} = 50.9$

Therefore, required correct mean is 50.9

Merits and Demerits of Arithmetic Mean:

Merits:

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. If the number of items is sufficiently large, it is more accurate and more reliable.
4. It is a calculated value and is not based on its position in the series.
5. It is possible to calculate even if some of the details of the data are lacking.
6. Of all averages, it is affected least by fluctuations of sampling.
7. It provides a good basis for comparison.
8. Arrangement of data is not required for computing Arithmetic mean.
9. It is suitable for further mathematical treatment.

Demerits:

1. It cannot be obtained by inspection nor located through a frequency graph.
2. It cannot be in the study of qualitative phenomena not capable of numerical measurement i.e. Intelligence, beauty, honesty etc.,
3. It is not a suitable measure of central value in case of highly skewed distribution.
4. It is very much affected by extreme values.
5. It cannot be calculated for open-end classes.

Harmonic mean (H.M) :

Harmonic mean of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values.

H.M for individual series: If $x_1, x_2 \ldots . x_n$ are n non-zero observations, then H.M. is given by

$$H.M. = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$

H.M. for discrete series: Let $x_1, x_2, \ldots\ldots, x_n$ be non-zero variate values with their correspondings frequencies $f_1, f_2, \ldots\ldots, f_n$, then their H.M. is defined as

$$H.M. = \frac{N}{\sum_{i=1}^{n} f_i\left(\frac{1}{x_i}\right)}, \text{ where } \sum f = N$$

H.M. for continuous series:

Let $x_1, x_2, \ldots\ldots, x_n$ be the mid-values of a variable of the classes with their corresponding frequencies $f_1, f_2, \ldots\ldots, f_n$, then their H. M. is given by

$$H.M = \frac{N}{\sum f \frac{1}{x}}, \text{ where } \sum f = N$$

Example: From the given data, calculate H.M.

5, 10, 17, 24, 30

Solution:

x	1/x
5	0.2000
10	0.1000
17	0.0588
24	0.0417
30	0.0333
Total	**0.4338**

We know,

$$H.M. = \frac{n}{\sum \frac{1}{x}} = 5/0.4338 = 11.526$$

Example: Find the harmonic mean for the following data:

x:	5	10	15	20	25
f:	4	6	8	5	2

Calculation of H.M

X	**f**	$f\frac{1}{x}$
5	4	0.800
10	6	0.600
15	8	0.534
20	5	0.250
25	2	0.080
	N = 25	$\sum f \frac{1}{x} = 2.264$

Here, n = 25, $\sum f \frac{1}{x} = 2.2644$

$$H.M. = \frac{N}{\sum f \frac{1}{x}} = \frac{25}{2.264}$$

$\therefore$ H.M. = 11.042

Example: Ages of some people of a village are given below. Calculate the harmonic mean of the age of these people.

Age(years)	20	21	22	23	24	25
Number of People	4	2	7	1	3	1

Solution:

Age(x)	**Number of People(f)**	$\frac{1}{x}$	$f\left(\frac{1}{x}\right)$
20	4	0.0500	0.2000
21	2	0.0476	0.0952
22	7	0.0454	0.3178
23	1	0.435	0.0435
24	3	0.0417	0.1251
25	1	0.0400	0.0400
Total	**N=18**		**0.8216**

We know,

$$H.M. = \frac{N}{\sum f\frac{1}{x}}, \text{ where } \sum f = N$$

$$\mathbf{=21.9}$$

Merits of H.M. :

1. It is rigidly defined.
2. It is defined on all observations.
3. It is amenable to further algebraic treatment.
4. It is the most suitable average when it is desired to give greater weight to smaller observations and less weight to the larger ones.

Demerits of H.M. :

1. It is not easily understood.
2. It is difficult to compute.
3. It is only a summary figure and may not be the actual item in the series
4. It gives greater importance to small items and is therefore, useful only when small items have to be given greater weightage.

Geometric mean:

G.M. for individual series:

The geometric mean of a series containing n observations is the n^{th} root of the product of the values. If $x_1, x_2, \ldots, x_n$ are n observations then

$$G.M. = \sqrt[n]{x_1.x_2 \ldots\ldots x_n} = (x_1.x_2 \ldots\ldots x_n)^{\frac{1}{n}}$$

$$log(G.M.) = \frac{1}{n}\, log(x_1.x_2 \ldots\ldots x_n)$$

$$= \frac{1}{n}(\log\ x_1 + \log x_2 + \cdots\ldots + log\ x_n)$$

$$= \frac{\sum log\, x_i}{n}$$

$$G.M. = Antilog\ \frac{\sum log\, x_i}{n}$$

G.M. for discrete series:

Let $x_1, x_2, \ldots\ldots, x_n$ be the variate values with their respective frequency $f_1, f_2, \ldots\ldots, f_n$ then their G.M. is given by

$$G.M. = Antilog\ \frac{\sum f\ logx}{N},\ \text{where } \sum f = N$$

G.M. for continuous series:

Let $x_1, x_2, \ldots\ldots, x_n$ be the mid - values of the classes with their respective frequencies $f_1, f_2, \ldots\ldots, f_n$ then their G.M. is given by

$$G.M. = Antilog\ \frac{\sum f\ logx}{N},\ \text{where } \sum f = N$$

Example: Calculate the geometric mean of the following series of monthly income of a batch of families 180, 250, 490, 1400, and 1050.

Solution:

Let X be the monthly income

x	Log x
180	2.2553
250	2.3979
490	2.6902
1400	3.1461
1050	3.0212
Total	**13.5107**

We know

$$G.M. = Antilog\ \frac{\sum log\, x_i}{n}$$

$$= Antilog\ \frac{13.5107}{5}$$

$$= Antilog\ 2.7021$$

$$= 503.6$$

Example: Find the geometric mean of the following data:

Age	3	6	8	11	13
Number of Students	4	8	5	3	2

Solution:

Let X be the age of the students.

Calculation of Geometric mean

x	f	log x	f log x
3	4	0.4771	1.9084
6	8	0.7781	6.2248
8	5	0.9030	4.5150
11	3	1.0414	3.1242
13	2	1.1140	2.2280
Total	N = 22		$\sum$ f logx = 18.0004

Here, N = 22, $\sum$f. log x = 18.0004

$$\text{G.M.} = \text{Antilog}\left[\frac{\sum f.\log x}{N}\right] = \text{Antilog}\left[\frac{18.0004}{22}\right] = \text{Antilog}\ (0.8182) = 6.578$$

$\therefore$ G.M. = 6.578 years

Merits of Geometric mean:

1. It is rigidly defined
2. It is based on all items
3. It is very suitable for averaging ratios, rates and percentages
4. It is capable of further mathematical treatment.
5. Unlike AM, it is not affected much by the presence of extreme values

Demerits of Geometric mean:

1. It cannot be used when the values are negative or if any of the observations is zero
2. It is difficult to calculate particularly when the items are very large or when there is a frequency distribution.
3. It brings out the property of the ratio of the change and not the absolute difference of change as the case in arithmetic mean.
4. The GM may not be the actual value of the series.

Combined Mean:

If the arithmetic averages and the number of items in two or more related groups are known, the combined or the composite mean of the entire group can be obtained by

Combined mean = $\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Where, n_1 and n_2 are the number of items in two groups and $\bar{x}_1$ & $\bar{x}_2$ are the arithmetic averages of items in two groups.

The advantage of combined arithmetic mean is that, we can determine the over, all mean of the combined data without going back to the original data.

Example: Find the combined mean for the data given below

n_1=20, n_2 = 30, $\bar{x}_1$ = 4 and $\bar{x}_2$ = 3.

Solution:

Combined mean = $\bar{x}_{12}$

$$= \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$$

$$= \frac{20\times 4 + 30 \times 3}{20 + 30}$$

$$= 3.4$$

Positional Averages:

These averages are based on the position of the given observation in a series, arranged in an ascending or descending order. The magnitude or the size of the values does matter as was in the case of arithmetic mean. It is because of the basic difference that the median and mode are called the positional measures of an average.

Median:

The median is that value of the variate which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median.

Ungrouped or Raw data or Individual data:

Arrange the given values in the increasing or decreasing order. If the number of values are odd, median is the middle value. If the number of values are even, median is the mean of middle two values.

By formula, Median = Value of $(n+1)/2^{th}$ item

When odd number of values are given:

Example: Find median for the following data 25, 18, 27, 10, 7, 30, 42, 20, 53

Solution:

Arranging the data in the increasing order, we get 7, 10, 18, 20, 25, 27, 30, 42, 53

The middle value is the 5^{th} item i.e., 25 is the median.

Using formula, Median = Value of $(n+1)/2^{th}$ item

= Value of $(9+1)/2^{th}$ item

= Value of 5^{th} item

= 25

When even number of values are given:

Example: Find the median for the following data 4, 8, 12, 30, 18, 10, 2, 22

Solution:

Arranging the data in the increasing order, we get 2, 4, 8, 10, 12, 18, 22, 30

Here median is the mean of the middle two items (i.e.) mean of (10, 12) i.e. = (10+12)/2= 11

Using formula, Median = Value of $(n+1)/2^{th}$ item

= Value of $(8+1)/2^{th}$ item

= Value of 4.5^{th} item

= Mean of 4^{th} and 5^{th} item

= (10+12)/2

= 11

Example*:* The following table represents the marks obtained by a batch of 10 students in certain class tests in statistics and Economics.

Serial No.	1	2	3	4	5	6	7	8	9	10
Marks (Statistics)	53	55	52	32	30	60	47	46	35	28
Marks (Economics)	57	45	24	31	25	84	43	80	32	72

Indicate in which subject is the level of knowledge higher?

Solution*:*
For such question, median is the most suitable measure of central tendency. The marks in the two subjects are first arranged in increasing order as follows:

Serial No.	1	2	3	4	5	6	7	8	9	10
Marks (Statistics)	28	30	32	35	46	47	52	53	55	60
Marks (Economics)	24	25	31	32	43	45	57	72	80	84

Using formula, Median = Value of $(n+1)/2^{th}$ item
= Value of $(10+1)/2^{th}$ item
= Value of 5.5^{th} item
= Mean of 5^{th} and 6^{th} item

So,
Median for statistics = (46+47)/2
= 46.5
Median for Economics = (43+45)/2
=44

Since the median for Statistics is greater than the median for Economics, the level of knowledge in Statistics is higher than that in Economics.

Grouped Data:
In a grouped distribution, values are associated with frequencies. Grouping can be in the form of a discrete frequency distribution or a continuous frequency distribution. Whatever may be the type of distribution, cumulative frequencies have to be calculated to know the total number of items.

Cumulative frequency (c.f.):
Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the pervious classes, i.e. adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

Discrete Series:

- Find cumulative frequencies.
- Find (N+1)/2

- See in the cumulative frequencies the value just greater than (N+1)/2
- Then the corresponding value of x is median.

Example: The following data pertaining to the number of members in a family. Find median size of the family.

Number of Members (x)	1	2	3	4	5	6	7	8	9	10	11	12
Frequency(f)	1	3	5	6	10	13	9	5	3	2	2	1

Solution:

X	**F**	**c. f.**
1	1	1
2	3	4
3	5	9
4	6	15
5	10	25
6	13	38
7	9	47
8	5	52
9	3	55
10	2	57
11	2	59
12	1	60
Total	**60**	

Using formula, Median = Size of $(N+1)/2^{th}$ item
= Size of $(60+1)/2^{th}$ item
= Size of 30.5^{th} item

The cumulative frequency just greater than 30.5 is 38 and the value of x corresponding to 38 is 6. Hence the median size is 6 members per family.

Continuous Series:

The steps given below are followed for the calculation of median in continuous series.
Step1: Find cumulative frequencies.
Step2: Find N/2
Step3: See in the cumulative frequency the value first greater than N/2

Then the corresponding class interval is called the Median class. Then apply the formula

$$Median(M_d) = l + \frac{\frac{N}{2} - m}{f} \times c$$

Where, l = Lower limit of the median class
m = cumulative frequency preceding the median class
c = width of the median class
f =frequency in the median class.
N=Total frequency.

Note: If the class intervals are given in inclusive type convert them into exclusive type.

Example: The following table gives the frequency distribution of 325 workers of an industry, according to their average monthly income in a certain year.

Income group (in Rs)	Number of Workers
Below 100	1
100 – 150	20
150 – 200	42
200 – 250	55
250 – 300	62
300 – 350	45
350 – 400	30
400 – 450	25
450 – 500	15
500 – 550	18
550 – 600	10
600 and above	2
Total	325

Calculate median income.

Solution:

Income group (in Rs)	Number of Workers	c. f.
Below 100	1	1
100 – 150	20	21
150 – 200	42	63
200 – 250	55	118
250 – 300	62	180
300 – 350	45	225
350 – 400	30	255
400 – 450	25	280
450 – 500	15	295
500 – 550	18	313
550 – 600	10	323
600 and above	2	325
Total	325	

Here, N/2 = 325/2 = 162.5, so the median class is 250 – 300.
l = 250, m = 118, c = 50, f = 62
Hence,

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$= 250 + (162.5-118) \times 50/62 = 285.89$$

Example: Calculate median from the following data

Value	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-50
Frequency	5	8	10	12	7	6	3	2

Solution:
To convert the inclusive class interval into exclusive, correction factor = (10-9)/2=0.5
So, we get the exclusive frequency distribution as follows:

Value	**Frequency**	**Exclusive class interval**	**c.f.**
5-9	5	4.5-9.5	5
10-14	8	9.5-14.5	13
15-19	10	14.5-19.5	23
20-24	12	19.5-24.5	35
25-29	7	24.5-29.5	42
30-34	6	29.5-34.5	48
35-39	3	34.5-39.5	51
40-44	2	39.5-44.5	53
	53		

Here, N/2 = 53/2 = 26.5. So, Median class = 19.5-24.5
l= 19.5, m = 23, f = 12, c = 5

So,

$$Median(M_d) = l + \frac{\frac{N}{2} - m}{f} \times c$$ = 19.5+ (26.5-23) x 5/12 = 20.96

Example: The following table shows the wage distribution of person in particular region:

Wage Below (Rs.):	10	20	30	40	50	60	70	80
No. of persons (00):	2	5	9	12	14	15	15.5	15.6

Find the median wage.

Solution:

Calculation of median

Wage (X)	**c. f.**	**f**
Below 10	2	2
10-20	5	3
20-30	9	4
30-40	12	3
40-50	14	2
50-60	15	1
60-70	15.5	0.5
70-80	15.6	0.1
		N = 15.6

Here, $\frac{N}{2} = \frac{15.6}{2} = 7.8$

The c. f. just greater than 7.8 is 9. Therefore, median lies in the class 20-30.

Then, $l = 20$, h = 10, c.f. = 5, f = 4

$$\therefore \quad \text{Md.} = l + \frac{\frac{N}{2} - \text{c.f}}{\text{f}} \times \text{h}$$

$$= 20 + \frac{7.8 - 5}{4} \times 10$$

= Rs. 27

∴ Median Wage = Rs. 27

Example: Compute median for the following data.

Mid-value	5	15	25	35	45	55	65	75
Frequency	7	10	15	17	8	4	6	7

Solution: Here, correction factor = (15-5)/2 = 5. Hence we change the given distribution with mid value as follows:

Mid-value	**Class Interval**	**Frequency**	**c.f.**
5	0-10	7	7
15	10-20	10	17
25	20-30	15	32
35	30-40	17	49
45	40-50	8	57
55	50-60	4	61
65	60-70	6	67
75	70-80	7	74
Total		**74**	

N/2 = 74/2 = 37, so median class is 30 - 40.

l= 30, m = 32, f = 17, c = 10

So,

$Median(M_d) = l + \frac{\frac{N}{2} - m}{f} \times c$ = 19.5+ (37-32) x 10/17 = 32.94

Graphic method for Location of median:
Median can be located with the help of the cumulative frequency curve or ‘ ogive’ . The procedure for locating median in a grouped data is as follows:

Step1: The class boundaries, where there are no gaps between consecutive classes, are represented on the horizontal axis (x-axis).
Step2: The cumulative frequency corresponding to different classes is plotted on the vertical axis (y-axis) against the upper limit of the class interval (or against the variate value in the case of a discrete series.)
Step3: The curve obtained on joining the points by means of freehand drawing is called the ‘ ogive’ . The ogive so drawn may be either a (i) less than ogive or a (ii) more than ogive.
Step4: The value of N/2 or (N+1)/2 is marked on the y-axis, where N is the total frequency.
Step5: A horizontal straight line is drawn from the point N/2 or (N+1)/2 on the y-axis parallel to x-axis to meet the ogive.
Step6: A vertical straight line is drawn from the point of intersection perpendicular to the horizontal axis.
Step7: The point of intersection of the perpendicular to the x-axis gives the value of the median.

Remarks :

1. From the point of intersection of ‘ less than’ and ‘more than’ ogives, if a perpendicular is drawn on the x-axis, the point so obtained on the horizontal axis gives the value of the median.

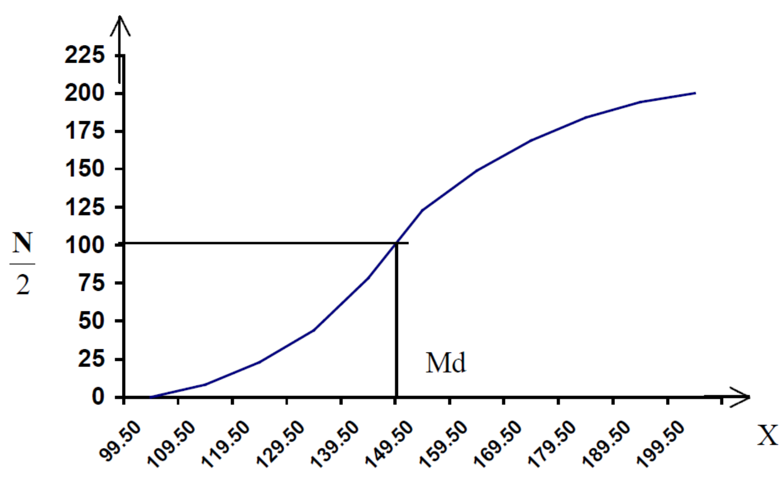
2. If ogive is drawn using cumulated percentage frequencies, then we draw a straight line from the point intersecting 50 percent cumulated frequency on the y-axis parallel to the x-axis to intersect the ogive. A perpendicular drawn from this point of intersection on the horizontal axis gives the value of the median.

Example: Draw an ogive of ‘ less than’ type on the data given below and hence find median.

Weight(lbs)	Number of persons
100-109	8
110-119	15
120-129	21
130-139	34
140-149	45
150-159	26
160-169	20
170-179	15
180-189	10
190-199	6

Solution:

Class interval	No of persons	True class interval	Less than c.f
100-109	8	99.5-109.5	8
110-119	15	109.5-119.5	23
120-129	21	119.5-129.5	44
130-139	34	129.5-139.5	78
140-149	45	139.5-149.5	123
150-159	26	149.5-159.5	149
160-169	20	159.5-169.5	169
170-179	15	169.5-179.5	184
180-189	10	179.5-189.5	194
190-199	6	189.5-199.5	200

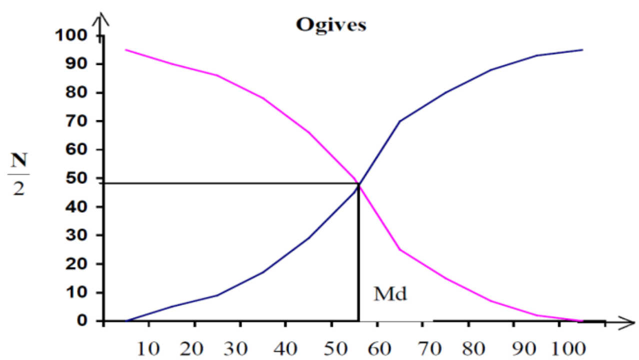


Example: Draw an ogive for the following frequency distribution and hence find median.

Marks	Number of students
0-10	5
10-20	4
20-30	8
30-40	12
40-50	16
50-60	25
60-70	10
70-80	8
80-90	5
90-100	2

Solution:

Class boundary	Cumulative Frequency	
	Less than	More than
0	0	95
10	5	90
20	9	86
30	17	78
40	29	66
50	45	50
60	70	25
70	80	15
80	88	7
90	93	2
100	95	0



Merits of Median :

1. Median is not influenced by extreme values because it is a positional average.
2. Median can be calculated in case of distribution with open end intervals.
3. Median can be located even if the data are incomplete.
4. Median can be located even for qualitative factors such as ability, honesty etc.

Demerits of Median :

1. A slight change in the series may bring drastic change in median value.
2. In case of even number of items or continuous series, median is an estimated value other than any value in the series.
3. It is not suitable for further mathematical treatment except its use in mean deviation.
4. It is not taken into account all the observations.

Quartiles:

The quartiles divide the total number of observations into four equal parts. There are three quartiles. The second quartile divides the distribution into two halves and therefore is the same as the median. The first (lower) quartile (Q_1) covers the first one-fourth (25%), the third (upper) quartile (Q3) covers the three-fourth (75%) of the series. Thus, Q1 < Q2 < Q3.

Quartiles for individual series:

Let $x_1, x_2, \ldots, x_n$ be n observations arranged in ascending order of their magnitude then the quartiles, can be computed as

$$Q_i = \text{Value of } i\left(\frac{n+1}{4}\right)^{th} \text{ items, where i = 1, 2, and 3.}$$

Example: Compute quartiles for the data given below 25,18,30, 8, 15, 5, 10, 35, 40, 45

Solution : arranging the given data in ascending order,

5, 8, 10, 15, 18,25, 30,35,40, 45

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ item}$$
$$= \left(\frac{10+1}{4}\right)^{th} \text{ item}$$
$$= (2.75)^{th} \text{ item}$$
$$= 2^{nd} \text{ item} + \left(\frac{3}{4}\right) (3^{rd} \text{ item-}2^{nd} \text{ item})$$
$$= 8 + \frac{3}{4} \quad (10\text{-}8)$$
$$= 8 + \frac{3}{4} \times 2$$
$$= 8 + 1.5$$
$$= 9.5$$
$$Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{ item}$$
$$= 3 \times (2.75)^{th} \text{ item}$$
$$= (8.25)^{th} \text{ item}$$
$$= 8^{th} \text{ item} + \frac{1}{4} [9^{th} \text{ item-}8^{th} \text{ item}]$$
$$= 35 + \frac{1}{4} [40\text{-}35]$$
$$= 35+1.25=36.25$$

Discrete Series :

Step1: Find cumulative frequencies.

Step2: Find 1(N+1)/4

Step3: See in the cumulative frequencies , the value just greater than 1(N+1)/4, then the corresponding value of *x* is Q_1

Step4: Find 3(N+1)/4

Step5: See in the cumulative frequencies, the value just greater than 3(N+1)/4, then the corresponding value of x is Q_3

Example: Compute quartiles for the data given bellow.

X	5	8	12	15	19	24	30
f	4	3	2	4	5	2	4

Solution:

x	f	c.f
5	4	4
8	3	7
12	2	9
15	4	13
19	5	18
24	2	20
30	4	24
Total	24	

$$Q_1 = \left(\frac{N+1}{4}\right)^{th} \text{item} = \left(\frac{24+1}{4}\right) = \left(\frac{25}{4}\right) = 6.25^{th} \text{ item}$$

$$Q_3 = 3\left(\frac{N+1}{4}\right)^{th} \text{item} = 3\left(\frac{24+1}{4}\right) = 18.75^{th} \text{ item} \therefore Q_1 = 8;\ \ Q_3 = 24$$

Continuous series :

Step1: Find cumulative frequencies

Step2: Find N/4

Step3: See in the cumulative frequencies, the value just greater than N/4, then the corresponding class interval is called first quartile class.

Step4: Find 3(N/4), then see in the cumulative frequencies the value just greater than 3(N/4), then the corresponding class interval is called 3rd quartile class. Then apply the respective formulae

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

$$Q_3 = l_3 + \frac{3\left(\frac{N}{4}\right) - m_3}{f_3} \times c_3$$

Where l_1 = lower limit of the first quartile class

f_1 = frequency of the first quartile class

c_1 = width of the first quartile class

m_1 = c.f. preceding the first quartile class

l_3 = 1ower limit of the 3rd quartile class

f_3 = frequency of the 3rd quartile class
c_3 = width of the 3rd quartile class
m_3 = c.f. preceding the 3rd quartile class

Example: The following series relates to the marks secured by students in an examination.

Marks	No. of students
0-10	11
10-20	18
20-30	25
30-40	28
40-50	30
50-60	33
60-70	22
70-80	15
80-90	12
90-100	10

Find the lower and upper quartiles

Solution:

C.I.	f	cf
0-10	11	11
10-20	18	29
20-30	25	54
30-40	28	82
40-50	30	112
50-60	33	145
60-70	22	167
70-80	15	182
80-90	12	194
90-100	10	204
	204	

N/4 = 204/4 = 51; 3(N/4) = 3x51 = 153, So, lower quartile is

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

$$= 20 + \frac{51-29}{25} \times 10 \qquad = 20 + 8.8 = 28.8$$

Upper quartile is

$$Q_3 = l_3 + \frac{3\left(\frac{N}{4}\right) - m_3}{f_3} \times c_3$$

$$= 60 + \frac{153-145}{22} \times 12 \quad = 60 + 4.36 = 64.36$$

Deciles:

The deciles divide the total number of observations into ten equal parts. There are nine deciles. These are D_1, D_2, …….., D_9. These are called first decile, second decile, …..ninth decile. Thus $D_1 < D_2$ ………$< D_9$.

Deciles for individual series:

Let $x_1, x_2, \ldots, x_n$ be n observations arranged in ascending order of their magnitude then the quartiles, can be computed as

Di = Value of $i\left(\frac{n+1}{10}\right)^{th}$ items, where i = 1, 2, ……, 9.

Example: Compute D_5 for the data 5, 24, 36, 12, 20, 8

Solution : Arranging the given values in the increasing order, we get 5, 8, 12, 20, 24, 36

D_5 = value of $5(n+1)/10^{th}$ item

= value of $5(6+1)/10^{th}$ item

= value of $(3.5)^{th}$ item

= 3^{rd} item + 0.5 × [4^{th} item – 3^{rd} item]

= 12 + 0.5 x [20 – 12] = 12+ 4 = 16

Deciles for Grouped data :

Example: Calculate D_3 and D_7 for the data given below

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency :	5	7	12	16	10	8	4

Solution:

C.I	f	c.f
0-10	5	5
10-20	7	12
20-30	12	24
30-40	16	40
40-50	10	50
50-60	8	58
60-70	4	62
	62	

$$D_3 \text{ item } = \left(\frac{3N}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{3\times 62}{10}\right)^{th} \text{ item}$$

$$= (18.6)^{th} \text{ item}$$

which lies in the interval 20-30

$$\therefore D_3 = l + \frac{3\left(\frac{N}{10}\right) - m}{f} \times c$$

$$= 20 + \frac{18.6 - 12}{12} \times 10$$

$$= 20 + 5.5 = 25.5$$

$$D_7 \text{ item} = \left(\frac{7 \times N}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{7 \times 62}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{434}{10}\right)^{th} \text{ item} = (43.4)^{\text{th}} \text{ item}$$

which lies in the interval(40-50)

$$D_7 = l + \frac{\left(\frac{7N}{10}\right) - m}{f} \times c$$

$$= 40 + \frac{43.4 - 40}{10} \times 10$$

$$= 40 + 3.4 = 43.4$$

Percentiles : The percentile values divide the distribution into 100 parts each containing 1 percent of the cases. The percentile (P_k) is that value of the variable up to which lie exactly k% of the total number of observations.

Relationship : $P_{25} = Q_1$; $P_{50} = D_5 = Q_2$ = Median and $P_{75} = Q_3$

Percentile for Raw Data or Ungrouped Data :

Example: Calculate P_{15} for the data 5, 24 , 36 , 12 , 20 , 8

Arranging the given values in the increasing order, we get 5, 8, 12, 20, 24, 36

P_{15} = Value of 15(n+1)/100 th observation= Value of 15(6+1)/100 th observation

= Value of (1.05) th observation = 1_{st} item + 0.05 (2_{nd} item – 1_{st} item)

= 5 + 0.05 (8-5) = 5 + 0.15 = 5.15

Percentile for grouped data:

Example: Find P_{53} for the following frequency distribution.

Class interval	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	5	8	12	16	20	10	4	3

Solution:

Class Interval	Frequency	C.f
0-5	5	5
5-10	8	13
10-15	12	25
15-20	16	41
20-25	20	61
25-30	10	71
30-35	4	75
35-40	3	78
Total	78	

$$P_{53} = l + \frac{\frac{53N}{100} - m}{f} \times c$$

$$= 20 + \frac{41.34 - 41}{20} \times 5$$

$$= 20 + 0.085 \quad = 20.085.$$

Mode:
The mode is the value in the distribution which occurs most frequently. It is an actual value, which has the highest concentration of items in and around it. According to Croxton and Cowden "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded at the most typical of a series of values".
Its importance is very great in marketing studies where a manager is interested in knowing about the size, which has the highest concentration of items. For example, in placing an order for shoes or ready-made garments the modal size helps because this sizes and other sizes around in common demand.

For ungrouped or Raw Data:
For ungrouped data or a series of individual observations, mode is often found by mere inspection.
Example: 3, 5, 11, 18, 11, 16, 9, 11, 3, 11
$\therefore$ Mode = $M_0 = 11$ since it occurs 4 times i.e. maximum times of occurrence.
In some cases the mode may be absent while in some cases there may be more than one mode.

Example: 1. 12, 10, 15, 24, 30 (no mode)
2. 7, 10, 15, 12, 7, 14, 24, 10, 7, 20, 10
$\therefore$ the modes are 7 and 10

Discrete distribution:
See the highest frequency and corresponding value of X is mode.

Continuous distribution:
See the highest frequency then the corresponding value of class interval is called the modal class. Then apply the formula.

$$\text{Mode} = \text{Mo} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h,$$

Where,
l = Lower limit of the modal class
f_1 = frequency of the modal class
f_0 = frequency of the class preceding the modal class
f_2 = frequency of the class succeeding the modal class
h = size of the modal class

Example: The following table gives the distribution of 100 families according to their age:

Age(years):	0-10	10-20	20-30	30-40	40-50
No. of families:	14	a	27	b	15

Find the values of a and b if mode is 24.

Solution:

Age	Frequency
0-10	14
10-20	A

20-30	27
30-40	B
40-50	15
	N = 56 + a + b

Given that N = 100

$\therefore$ a + b = 44 .... (i)

And b = 44 – a

Mo = 24 which lies in the class 20-30.

Here, $l = 20$, $h = 10$, $f_1 = 27$, $f_0 = a$, $f_2 = b$, $f_2 = 44 - a$

Now,

$$Mo = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

$$\text{Or, } 24 = 20 + \frac{(27 - a)}{54 - a - (44 - a)}$$

$$\text{Or, } 4 = \frac{(27 - a)}{10} \times 10$$

$\therefore$ a = 23 , b = 44 – 23 = 21

Determination of Modal class :

For a frequency distribution modal class corresponds to the maximum frequency. But in any one (or more) of the following cases

- If the maximum frequency is repeated
- If the maximum frequency occurs in the beginning or at the end of the distribution
- If there are irregularities in the distribution

The modal class is determined by the method of grouping.

Method of grouping

Steps for Calculation:

We prepare a grouping table with 6 columns.

1. In column I, we write down the given frequencies.
2. Column II is obtained by combining the frequencies two by two.
3. Leave the 1st frequency and combine the remaining frequencies two by two and write in column III.
4. Column IV is obtained by combining the frequencies three by three.
5. Leave the 1st frequency and combine the remaining frequencies three by three and write in column V.
6. Leave the 1st and 2nd frequencies and combine the remaining frequencies three by three and write in column VI.

Mark the highest frequency in each column. Then form an analysis table to find the modal class. After finding the modal class, use the formula to calculate the modal value.

Example: Calculate mode for the following frequency distribution.

Class interval	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	9	12	15	16	17	15	10	13

Solution: Grouping Table

<table>
<tr><th>C I</th><th>f</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th></tr>
<tr><td>0- 5</td><td>9</td><td rowspan="2">21</td><td></td><td rowspan="3">36</td><td></td><td rowspan="2"></td></tr>
<tr><td>5-10</td><td>12</td><td rowspan="2">27</td><td rowspan="3">43</td></tr>
<tr><td>10-15</td><td>15</td><td rowspan="2">31</td><td rowspan="3">48</td></tr>
<tr><td>15-20</td><td>16</td><td rowspan="2">33</td><td rowspan="3">48</td></tr>
<tr><td>20-25</td><td>17</td><td rowspan="2">32</td><td rowspan="3">42</td></tr>
<tr><td>25-30</td><td>15</td><td rowspan="2">25</td><td rowspan="3">38</td></tr>
<tr><td>30-35</td><td>10</td><td rowspan="2">23</td><td rowspan="2"></td></tr>
<tr><td>35-40</td><td>13</td><td></td><td></td></tr>
</table>

Analysis Table

Columns	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
1					1			
2					1	1		
3				1	1			
4				1	1	1		
5		1	1	1				
6			1	1	1			
Total		1	2	4	5	2		

The maximum occurred corresponding to 20-25, and hence it is the modal class.

$$\text{Mode} = M_o = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

Here $l = 20$; $\Delta_1 = f_1 - f_0 = 17 - 16 = 1$

$\Delta_2 = f_1 - f_2 = 17 - 15 = 2$

$$\therefore M_0 = 20 + \frac{1}{1+2} \times 5$$

$$= 20 + 1.67 = 21.67$$

Graphic Location of mode:

Steps:

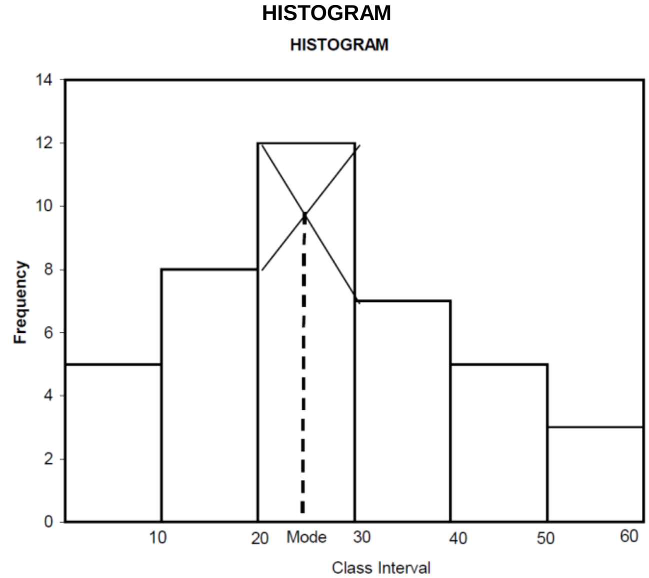
1. Draw a histogram of the given distribution.
2. Join the rectangle corner of the highest rectangle (modal class rectangle) by a straight line to the top right corner of the preceding rectangle. Similarly the top left corner of the highest rectangle is joined to the top left corner of the rectangle on the right.
3. From the point of intersection of these two diagonal lines, draw a perpendicular to the x -axis.
4. Read the value in x-axis gives the mode.

Example:

Locate the modal value graphically for the following frequency distribution.

Class interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	5	8	12	7	5	3

Solution:



Merits of Mode:

1. It is easy to calculate and in some cases it can be located mere inspection
2. Mode is not at all affected by extreme values.
3. It can be calculated for open-end classes.
4. It is usually an actual value of an important part of the series.
5. In some circumstances it is the best representative of data.

Demerits of mode:

1. It is not based on all observations.
2. It is not capable of further mathematical treatment.
3. Mode is ill-defined generally, it is not possible to find mode in some cases.
4. As compared with mean, mode is affected to a great extent, by sampling fluctuations.
5. It is unsuitable in cases where relative importance of items has to be considered.

EMPIRICAL RELATIONSHIP BETWEEN AVERAGES

In a symmetrical distribution the three simple averages mean = median = mode. For a moderately asymmetrical distribution, the relationship between them are brought by Prof. Karl Pearson as
mode = 3median - 2mean.

Example: If the mean and median of a moderately asymmetrical series are 26.8 and 27.9 respectively, what would be its most probable mode?
Solution: Using the empirical formula
Mode = 3 median – 2 mean = $3 \times 27.9 - 2 \times 26.8 = 30.1$

Example: In a moderately asymmetrical distribution the values of mode and mean are 32.1 and 35.4 respectively. Find the median value.
Solution: Median = [2mean+mode]/2 = $[2 \times 35.4 + 32.1]/2 = 34.3$

Example: Following are the daily wages of workers in a textile. Find the median. Ans: 468.75

Wages (in Rs.)	Number of workers
less than 100	5
less than 200	12
less than 300	20
less than 400	32
less than 500	40
less than 600	45
less than 700	52
less than 800	60
less than 900	68
less than 1000	75

$= 400 + 68.75 = 468.75$

Example: Find median for the data given below.(Ans: 43.75)

Marks	Number of students
Greater than 10	70
Greater than 20	62
Greater than 30	50
Greater than 40	38
Greater than 50	30
Greater than 60	24
Greater than 70	17
Greater than 80	9
Greater than 90	4

Example : Following is the distribution of persons according to different income groups. Calculate arithmetic mean. Ans=31

Income Rs(100)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of persons	6	8	10	12	7	4	3

Example: Calculate mode for the following :

C- I	f
0-50	5
50-100	14
100-150	40
150-200	91
200-250	150
250-300	87
300-350	60
350-400	38
400 and above	15