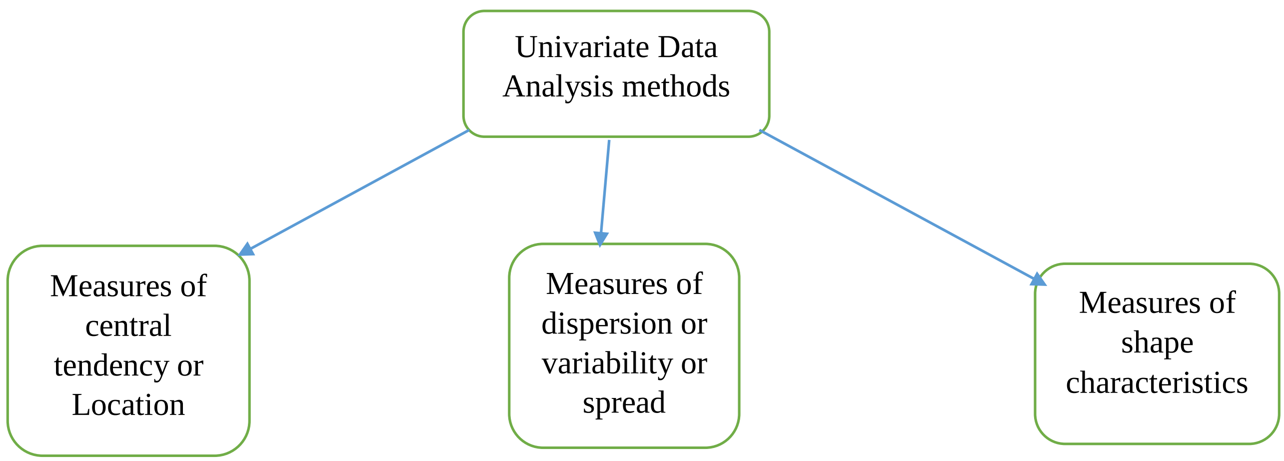
Measures of Central Tendency

Introduction:

Step: 1- Data sets are usually either a sample or a population.

Step: 2- Our target is to infer something about a whole (population) by a part (sample).

Step: 3- Most methods measure one of two data characteristics.



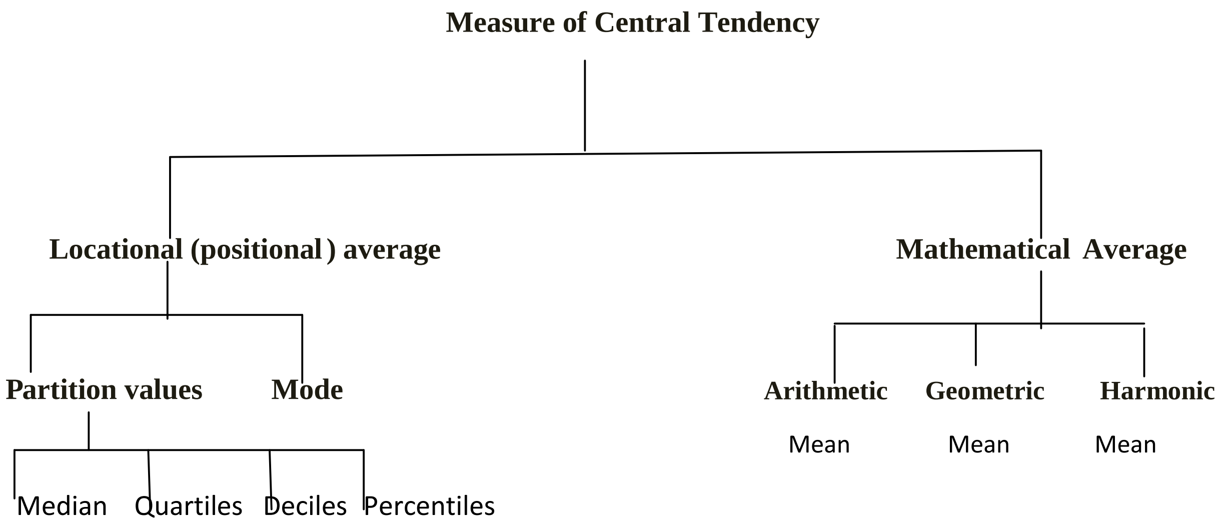
<u>Necessity/objective of measuring the central tendency</u>

I. To get a single value that describes the whole characteristics of the entire data.
II. It's gives us an idea about the concentration of the values in the central part of the distribution & describes the distribution in a concise manner.
III. To facilitate comparison of different data/distribution
IV. It represents all relevant information contained in the data in as few numbers as possible.
V. They give precise information, not information of a vague general type.
VI. For computing various other statistical measures such as dispersion, skewness, kurtosis, and various other basic characteristics of a mass of data.

<u>Characteristics of a good measure of central tendency:</u> According to Yule and Kendall, a good measure of central tendency should have the following characteristics-

i. It should be easy to understand.
ii. It should be easy to calculate/compute.
iii. It should be based on all observations.
iv. It should be rigidly defined.
v. It should not be unduly affected by extreme values.
vi. It should be suitable for further algebraic treatment.
vii. It should be less affected by sampling fluctuation i.e. should have sampling stability.

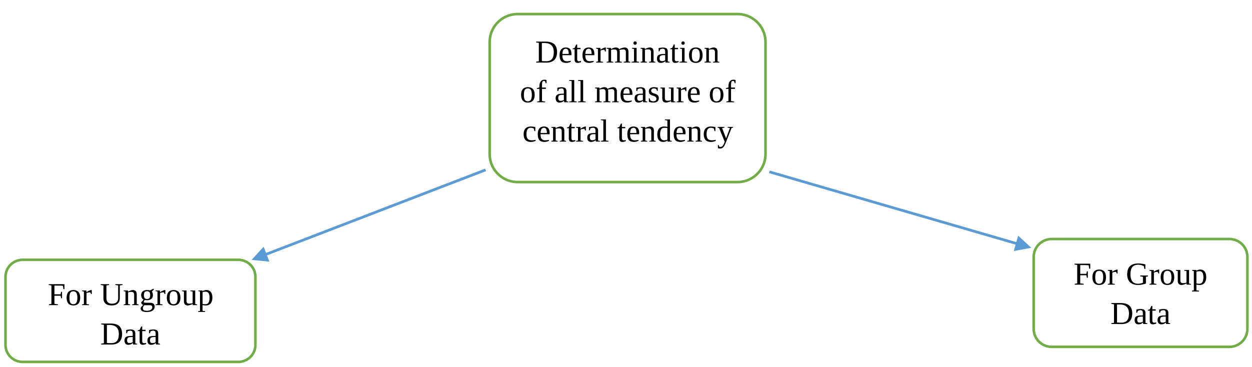
Different measures of central tendency:



In addition to these common measures, a few special measures occasionally used and most important in our daily life. These are-

- Generalized mean, Weighted mean, Pooled mean, Truncated mean or trimmed mean, Interquartile mean, Midrange, Midhinge, Quasi-arithmetic mean, Trimean, Winsorized mean, Geometric median, Quadratic mean, Simplicial depth, Tukey median, Progressive average, Moving average, etc.

Determination of all measures of central tendency is two types.



Arithmetic Mean: The most popular and widely used measure for representing the entire data by one value is the arithmetic mean. There are two methods of finding arithmetic mean-

1) Direct method & 2) Indirect method or short-cut method

We can obtain the arithmetic mean of a series of observations by adding the values of the observations and then dividing the sum by the number of observations. Arithmetic mean (AM) for

- Sample mean is denoted by $\bar{x}$
- Population mean is denoted by μ

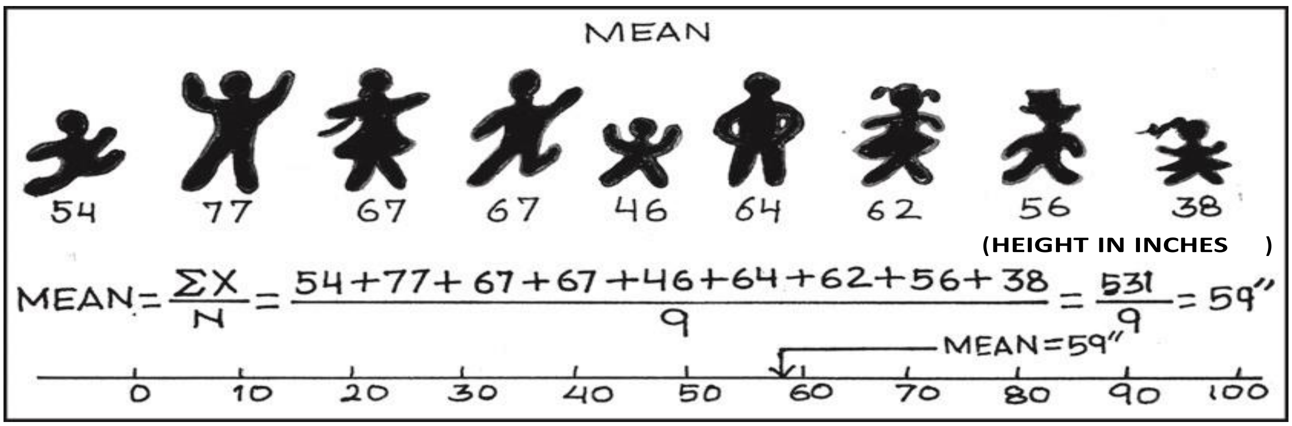
$$\text{A.M} = \frac{\text{Sum of all values in the data}}{\text{Total Number of observation}}$$

1) Direct Method: Population Mean: If there are N values $x_1, x_2, \ldots, \ldots, \ldots, x_N$ for a variable X in a finite population, then the population mean defined by-

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots + \ldots + \ldots + x_N}{N}$$

$$= \frac{\sum x_i}{N} \, (i = 1,\ 2, \ldots, \ldots, N)$$

Here, N is the population size or the total number of observations in the population. It is customary to represent the parameter by Greek letters "μ".



Sample mean or sample arithmetic mean: If there are n values $x_1, x_2, \ldots, \ldots, \ldots, x_n$ for a variable X, then the AM denoted by $\bar{x}$ is defined as-

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + \ldots + \ldots + x_n}{n}$$

$$= \frac{\sum x_i}{n} \, (i = 1,\ 2, \ldots, \ldots, n)$$

2) **Indirect/Step deviation method:** The arithmetic mean can also be calculated by taking deviations from any arbitrary point in which case the formula shall be:

$$\bar{x} = A + \frac{\sum d}{N}$$

Where, d=(X-A) and A= Arbitrary point (or assumed mean)

It should be noted that any value can be taken as arbitrary point and answer would be the same as obtained by the direct method.

Illustration 1. The monthly income (in rupees) of 10 employees working in a firm is as follows :

4487 4493 4502 4446 4475 4492 4572 4516 4468 4489

Find the average monthly income.

Solution. Let income be denoted by X.

$\Sigma X = 4487+4493+4502+4446+4475+4492+4572+4516+4468+4489 = 44,940$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{44940}{10} = 4494$$

Hence the average monthly income is Rs. 4,494.

CALCULATION OF AVERAGE INCOME

X (Rs.)	$(X-4460)$ d
4487	+27
4493	+33
4502	+42
4446	−14
4475	+15
4492	+32
4572	+112
4516	+56
4468	+ 8
4489	+29
	$\Sigma d = +340$

$$\bar{X} = A + \frac{\Sigma d}{N} = 4460 + \frac{340}{10} = 4460 + 34 = \text{Rs. } 4494.$$

Example: Banglatel is studying the number of minutes used by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.

90 77 94 89 119 112 91 110 92 100 113 83

What is the mean (arithmetic mean) number of minutes used?

Solution: Average use of the rate plan-

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + \ldots + \ldots + x_n}{n} = \frac{90 + 77 + \ldots + \ldots + 112 + 91 + \ldots + \ldots + 113 + 83}{12} = 97.5$$

Thus the arithmetic mean number of minutes used last month by the sample of cell phone users is 97.5 minutes.

<u>Calculation of arithmetic mean (for group data):</u> For grouped data, arithmetic mean computed by applying the any of the following methods:

1) Direct method & 2) Indirect method or step deviation method

For a group data as given in the following table

Mid values/Values:	x_1	x_2	...	...	x_n
Frequencies :	f_1	f_2	...	...	f_n

Such that $f_1 + f_2 + f_3 + \ldots + \ldots + \ldots + f_k = n$, then the AM (denoted by $\bar{x}$) is defined as

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots + f_kx_k}{n}; (i = 1, 2, \ldots, \ldots, k)$$

Steps for the Computation of Arithmetic Mean

1. Multiply each value of X or the mid-value of the class (in case of grouped or continuous frequency distribution) by the corresponding frequency f.

2. Obtain the total of the products obtained in step 1 above to get ΣfX.

3. Divide the total obtained in step 2 by $n = \Sigma f$, the total frequency.

The resulting value gives the arithmetic mean.

Example 5·1. *The intelligence quotients (IQ's) of 10 boys in a class are given below :*

70, 120, 110, 101, 88, 83, 95, 98, 107, 100

Find the mean I.Q.

Solution. Mean I.Q. ($\overline{X}$) of the 10 boys is given by :

$$\overline{X} = \frac{\Sigma X}{n} = \frac{1}{10}(70 + 120 + 110 + 101 + 88 + 83 + 95 + 98 + 107 + 100) = \frac{972}{10} = 97{\cdot}2$$

Example 5·2. *The following is the frequency distribution of the number of telephone calls received in 245 successive one-minute intervals at an exchange :*

Number of Calls	:	0	1	2	3	4	5	6	7
Frequency	:	14	21	25	43	51	40	39	12

Obtain the mean number of calls per minute.

Solution. Let the variable X denote the number of calls received per minute at the exchange.

COMPUTATION OF MEAN NUMBER OF CALLS

No. of Calls (X)	0	1	2	3	4	5	6	7	Total
Frequency (f)	14	21	25	43	51	40	39	12	$N = 245$
fX	0	21	50	129	204	200	234	84	$\Sigma fX = 922$

Mean number of calls per minute at the exchange is given by :

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{922}{245} = 3{\cdot}763$$

Step Deviation Method for Computing Arithmetic Mean. It may be pointed out that the direct formula can be used conveniently if the values of *X* or/and *f* are small. However, if the values of *X* or/and *f* are large, the calculation of mean by the direct formula is quite tedious and time consuming. In such a case the calculations can be reduced to a great extent by using the step deviation method which consists in taking the deviations (differences) of the given observations from any arbitrary value *A*. Let $d = X - A$ then,

$$\bar{X} = A + \frac{\sum fd}{N}$$

This formula is much more convenient to use for numerical problems than the direct formula. In case of grouped or continuous frequency distribution, with class intervals of equal magnitude, the calculations are further simplified by taking:

$$d = \frac{X - A}{h}$$

where *X* is the mid-value of the class and *h* is the common magnitude of the class intervals. Then

$$\bar{X} = A + h\frac{\sum fd}{N}$$

Steps for Computation of Mean by Step Deviation Method.
Step 1. Compute $d = (X - A)/h$, *A* being any arbitrary number and *h* is the common magnitude of the classes. Algebraic signs + or – are to be taken with the deviations.
Step 2. Multiply *d* by the corresponding frequency *f* to get *fd*.
Step 3. Find the sum of the products obtained in step 2 to get Σfd.
Step 4. Divide the sum obtained in step 3 by *N*, the total frequency.
Step 5. Multiply the value obtained in step 4 by *h*.
Step 6. Add *A* to the value obtained in step (5).
The resulting value gives the arithmetic mean of the given distribution. If we take $h = 1$, then the both formula are same.

Example 5·3. *Calculate the mean for the following frequency distribution :*

Marks :	*0—10*	*10—20*	*20—30*	*30—40*	*40—50*	*50—60*	*60—70*
Number of students :	*6*	*5*	*8*	*15*	*7*	*6*	*3*

(i) By the direct formula. ; *(ii) By the step deviation method.*

Solution.

COMPUTATION OF ARITHMETIC MEAN

Marks	*Mid-value (X)*	*Number of Students (f)*	*fX*	$d = \frac{X-35}{10}$	*fd*
0—10	5	6	30	–3	–18
10—20	15	5	75	–2	–10
20—30	25	8	200	–1	–8
30—40	35	15	525	0	0
40—50	45	7	315	1	7
50—60	55	6	330	2	12
60—70	65	3	195	3	9
		$N = \sum f = 50$	$\sum fX = 1670$		$\sum fd = -8$

(*i*) **Direct Formula** : $\text{Mean } (\bar{X}) = \frac{\sum fX}{\sum f} = \frac{1670}{50} = 33{\cdot}4$ marks.

(*ii*) **Step Deviation Method** : In the usual notations we have $A = 35$ and $h = 10$.

$$\therefore \quad \bar{X} = A + \frac{h\sum fd}{N} = 35 + \frac{10 \times (-8)}{50} = 35 - 1{\cdot}6 = 33{\cdot}4 \text{ marks.}$$

Example 5·4. *The numbers 3·2, 5·8, 7·9 and 4·5, have frequencies x, (x + 2), (x – 3) and (x + 6) respectively. If the arithmetic mean is 4·876, find the value of x.*

Solution.

COMPUTATION OF MEAN

Number (X)	*Frequency (f)*	*fX*
3·2	x	$3{\cdot}2x$
5·8	$x+2$	$5{\cdot}8\,(x+2)$
7·9	$x-3$	$7{\cdot}9\,(x-3)$
4·5	$x+6$	$4{\cdot}5\,(x+6)$

we have :

$$\sum f = x + (x+2) + (x-3) + (x+6) = 4x + 5$$

$$\sum fX = 3{\cdot}2x + 5{\cdot}8(x+2) + 7{\cdot}9(x-3) + 4{\cdot}5(x+6)$$

$$= (3{\cdot}2 + 5{\cdot}8 + 7{\cdot}9 + 4{\cdot}5)x + 11{\cdot}6 - 23{\cdot}7 + 27{\cdot}0$$

$$= 21{\cdot}4x + 14{\cdot}9$$

$$\therefore \quad \text{Mean} = \frac{\sum fX}{\sum f} = \frac{21{\cdot}4x + 14{\cdot}9}{4x+5} = 4{\cdot}876 \text{ (Given)}$$

$$\Rightarrow \quad 21{\cdot}4x + 14{\cdot}9 = 4{\cdot}876\,(4x+5)$$

$$\Rightarrow \quad 21{\cdot}4x + 14{\cdot}9 = 19{\cdot}504x = 24{\cdot}380$$

$$\Rightarrow \quad (21{\cdot}400 - 19{\cdot}504)x = 24{\cdot}380 - 14{\cdot}900$$

$$\Rightarrow \quad 1{\cdot}896x = 9{\cdot}480 \qquad \Rightarrow \qquad x = \frac{9{\cdot}480}{1{\cdot}896} = 5$$

Properties of arithmetic mean:

Property 1. The algebraic sum of the deviations of the given set of observations from their arithmetic mean is zero.

Property 2. Mean of the Combined Series. If we know the sizes and means of two component series, then we can find the mean of the resultant series obtained on combining the given series.

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

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + \ldots + n_k\bar{X}_k}{n_1 + n_2 + \ldots + n_k}$$

Example 5·13. *The mean of marks in Statistics of 100 students in a class was 72. The mean of marks of boys was 75, while their number was 70. Find out the mean marks of girls in the class.*

Solution. In the usual notations we are given :

$$n_1 = 70, \quad \bar{x}_1 = 75; \quad n_1 + n_2 = 100, \quad \bar{x} = 72 \; ; \quad \therefore \quad n_2 = 100 - 70 = 30. \text{ We want } \bar{x}_2.$$

$$\text{We have} \quad \bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \quad \Rightarrow \quad 72 = \frac{70 \times 75 + 30\bar{x}_2}{100}$$

$$\therefore \quad 72 \times 100 = 5250 + 30\bar{x}_2 \quad \Rightarrow \quad \bar{x}_2 = \frac{7200 - 5250}{30} = \frac{1950}{30} = 65$$

Hence, the mean of marks of girls in the class is 65.

Property 3. The arithmetic mean depends on the change of origin and scale of measurement.

Property 4. The arithmetic mean of the first n natural number is (n+1)/2

Advantages/merits of arithmetic mean:

1. Rigidly defined.
2. Easy to understand and calculate.
3. Based upon all observation.
4. Most amenable to algebraic treatment.
5. Not based on position in the series.

Dis-advantages/Demerits of arithmetic mean:

1. Cannot be defined graphically.
2. Cannot be used in case of qualitative data.
3. Affected very much by extreme values.
4. Cannot be calculated for qualitative data and sometimes provide wrong information.
5. Difficult to calculate in the case of the data with open-end class

Example 5·4. *The numbers 3·2, 5·8, 7·9 and 4·5, have frequencies x, (x + 2), (x − 3) and (x + 6) respectively. If the arithmetic mean is 4·876, find the value of x.*

Solution.

COMPUTATION OF MEAN

Number (X)	*Frequency (f)*	*fX*
3·2	x	$3{\cdot}2x$
5·8	$x+2$	$5{\cdot}8\,(x+2)$
7·9	$x-3$	$7{\cdot}9\,(x-3)$
4·5	$x+6$	$4{\cdot}5\,(x+6)$

we have :

$$\sum f = x + (x+2) + (x-3) + (x+6) = 4x + 5$$

$$\begin{aligned}\sum fX &= 3{\cdot}2x + 5{\cdot}8(x+2) + 7{\cdot}9(x-3) + 4{\cdot}5(x+6)\\ &= (3{\cdot}2 + 5{\cdot}8 + 7{\cdot}9 + 4{\cdot}5)x + 11{\cdot}6 - 23{\cdot}7 + 27{\cdot}0\\ &= 21{\cdot}4x + 14{\cdot}9\end{aligned}$$

$$\therefore \quad \text{Mean} = \frac{\sum fX}{\sum f} = \frac{21{\cdot}4x + 14{\cdot}9}{4x+5} = 4{\cdot}876 \text{ (Given)}$$

$$\Rightarrow \quad 21{\cdot}4x + 14{\cdot}9 = 4{\cdot}876\,(4x+5)$$

$$\Rightarrow \quad 21{\cdot}4x + 14{\cdot}9 = 19{\cdot}504x = 24{\cdot}380$$

$$\Rightarrow \quad (21{\cdot}400 - 19{\cdot}504)x = 24{\cdot}380 - 14{\cdot}900$$

$$\Rightarrow \quad 1{\cdot}896x = 9{\cdot}480 \qquad \Rightarrow \qquad x = \frac{9{\cdot}480}{1{\cdot}896} = 5$$

Example 5·9. *For a certain frequency table which has only been partly reproduced here, the mean was found to be 1·46.*

No. of accidents	:	*0*	*1*	*2*	*3*	*4*	*5*	*Total*
Frequency (No. of days)	:	*46*	*?*	*?*	*25*	*10*	*5*	*200*

Calculate the missing frequencies.

Solution. Let X denote the number of accidents and let the missing frequencies corresponding to $X = 1$ and $X = 2$ be f_1 and f_2 respectively.

We have

$$200 = 86 + f_1 + f_2$$

$$\Rightarrow \quad f_1 + f_2 = 200 - 86 = 114 \qquad \ldots(*)$$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{f_1 + 2f_2 + 140}{200} = 1{\cdot}46 \ (Given)$$

$$\Rightarrow \quad f_1 + 2f_2 + 140 = 1{\cdot}46 \times 200 = 292$$

$$\Rightarrow \quad f_1 + 2f_2 = 292 - 140 = 152 \qquad \ldots(**)$$

Subtracting (*) from (**), we get

$$f_2 = 152 - 114 = 38$$

COMPUTATION OF ARITHMETIC MEAN

No. of accidents (X)	*Frequency* (f)	fX
0	46	0
1	f_1	f_1
2	f_2	$2f_2$
3	25	75
4	10	40
5	5	25
Total	$86 + f_1 + f_2 = 200$	$f_1 + 2f_2 + 140$

Substituting in (*), we get $f1 = 114 - f2 = 114 - 38 = 76$

WEIGHTED ARITHMETIC MEAN: The arithmetic mean is based on the assumption that all the items in the distribution are of equal importance. However, in practice, the relative importance of all the items of the distribution is not same. If some items in a distribution are more important than others, then this point must be borne in mind, in order that average computed is representative of the distribution. In such cases, proper weightage is to be given to various items - the weights attached to each item being proportional to the importance of the item in the distribution. For example, if we want to have an idea of the change in cost of living of a certain group of people, then the simple mean of the prices of the commodities consumed by them will not do, since all the commodities are not equally important, *e.g.*, wheat, rice, pulses, housing, fuel and lighting are more important than cigarettes, tea, confectionery, cosmetics, etc. Let $W1$, $W2$,…, Wn be the weights attached to variable values $X1$, $X2$,…, Xn respectively. Then the weighted arithmetic mean, defined by:

$$\bar{X}_w = \frac{W_1X_1 + W_2X_2 + \ldots + W_nX_n}{W_1 + W_2 + \ldots + W_n} = \frac{\sum WX}{\sum W}$$

Example 5·16. *A candidate obtained the following percentages of marks in an examination : English 60; Hindi 75; Mathematics 63; Physics 59 ; Chemistry 55. Find the candidate's weighted arithmetic mean if weights 1, 2, 1, 3, 3 respectively are allotted to the subjects.*

Solution. Let the variable X denote the percentage of marks in the examination.

COMPUTATION OF WEIGHTED MEAN

Subject	*Marks (%) (X)*	*Weight (W)*	*WX*
English	60	1	60
Hindi	75	2	150
Mathematics	63	1	63
Physics	59	3	177
Chemistry	55	3	165
		$\sum W = 10$	$\sum WX = 615$

$$\therefore \quad \text{Weighted Arithmetic Mean (in\%)} = \frac{\sum WX}{\sum W} = \frac{615}{10} = 61{\cdot}5.$$

Application of arithmetic mean: It is extensively used in economics (per capita income, Consumption, price, sales, GDP, & GNP), Business forecasting, social science, demography, and time series, etc.

Application of weighted mean: When the importance of all the items in a series is not equal and it is difficult for finding the mean of combined set.

Example: Performance analysis of employees, Quality control of same product by different company, Results of students.

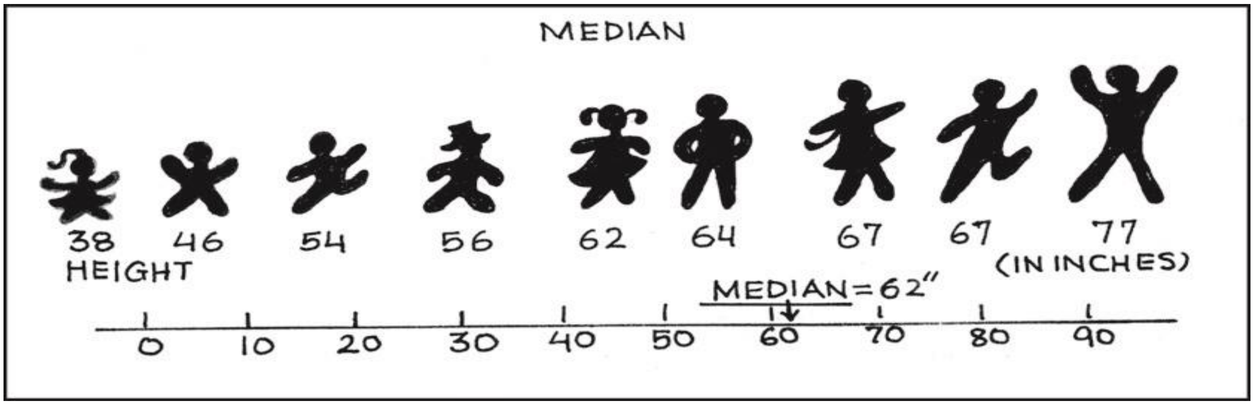
MEDIAN: In the words of L.R. Connor: "The median is that value of the variable which divides the group in two equal parts, one part comprising all the values greater and the other, all the values less than median". Thus median of a distribution may be defined as that value of the variable which exceeds and is exceeded by the same number of observations *i.e.*, it is the value such that the number of observations above it is equal to the number of observations below it. Thus, we see that as against arithmetic mean which is based on all the items of the distribution, the median is only *positional* average *i.e.*, its value depends on the position occupied by a value in the frequency distribution.

Calculation of Median. Case (I): Ungrouped Data. Let n be the number of observations

Rule-1: When ***n is odd*** the value of the $\frac{n+1}{2}th$ observation will be the median.

Rule-2: When ***n is even*** the median will be the AM of the values of $\frac{n}{2}th$ and $(\frac{n}{2}+1)th$ observation in the series. i.e, $\frac{[\frac{n}{2}th + \left(\frac{n}{2}+1\right)th] \text{ observation}}{2}$

If the number of observations is odd, then the median is the *middle value* after the observations have been arranged in ascending or descending order of magnitude. For example, the median of 5 observations 35, 12, 40, 8, 60 *i.e.*, 8, 12, 35, 40, 60, is 35. In the case of an even number of observations, the median is obtained as the arithmetic mean of the two middle observations after they are arranged in ascending or descending order of magnitude. Thus, if one more observation, say, 50 is added to the above five observations then the six observations in ascending order of magnitude are: 8, 12, 35, 40, 50, 60. Thus, Median = Arithmetic mean of two middle terms = (35+40)/2 = 37·5.



Median in case of Ungrouped Data	
In this case we first arrange the observations in **increasing** or **decreasing** order then we use the following formulae for Median:	
If "n" is odd	$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$
If "n" is even	$Median = \frac{size\ of \left\{\left(\frac{n}{2}\right) th + \left(\frac{n}{2}+1\right) th\right\} observation}{2}$

Determination of Median: Grouped Data

First we have to construct a cumulative frequency table. Then we have to identify the median class. Median class is the most important class for computing median. The class which contains $\frac{n}{2}th$ observation is called the median class. Here we always use $\frac{n}{2}$ instead of $\frac{n+1}{2}$ to locate median because $\frac{n}{2}$ divides the whole area of the curve into two equal parts in case of continuous variable. The formula for computing median is

$$Me = L + \frac{\frac{n}{2} - F_c}{f_m} \times h$$

- Me = Median
- L = Lower limit of the Median class
- F_c = Cumulative frequency of the pre median class.
- f_m = Frequency of the median class.
- h = Width of the median class.

- n = Total number of observation.

- **MEDIAN CLASS:** the class that contains $\frac{n}{2}th$ observation of the given data.

Steps of finding Median for group data-

- Step 1: Compute the cumulative frequencies.
- Step 2: Determine $\frac{n}{2}$th value, one half of the total number of cases.
- Step 3: Locate the median class.
- Step 4: Determine the lower limit (L) of the median class.
- Step 5: Sum the frequencies of all the classes prior to the median class. This is F_c.
- Step 6: Determine the frequency of the median class f_m
- Step 7: Determine the width of the median class, h.

Locating Median graphically: Consider the following data of monthly wages of a group of people. We would like to locate median graphically using OGIVE.

Wages(Rs 1000)	f	less than cf	More than cf
0-10	4	4	100
10-20	6	10	96
20-30	10	20	90
30-40	10	30	80
40-50	25	55	70
50-60	22	77	45
60-70	18	95	23
70-80	5	100	5

- ❖ Less than OGIVE: Plot upper limit and cf (less than cf)
- ❖ More than Ogive: Plot lower limit and more-than-cf

Solution by formula: The formula for computing median is

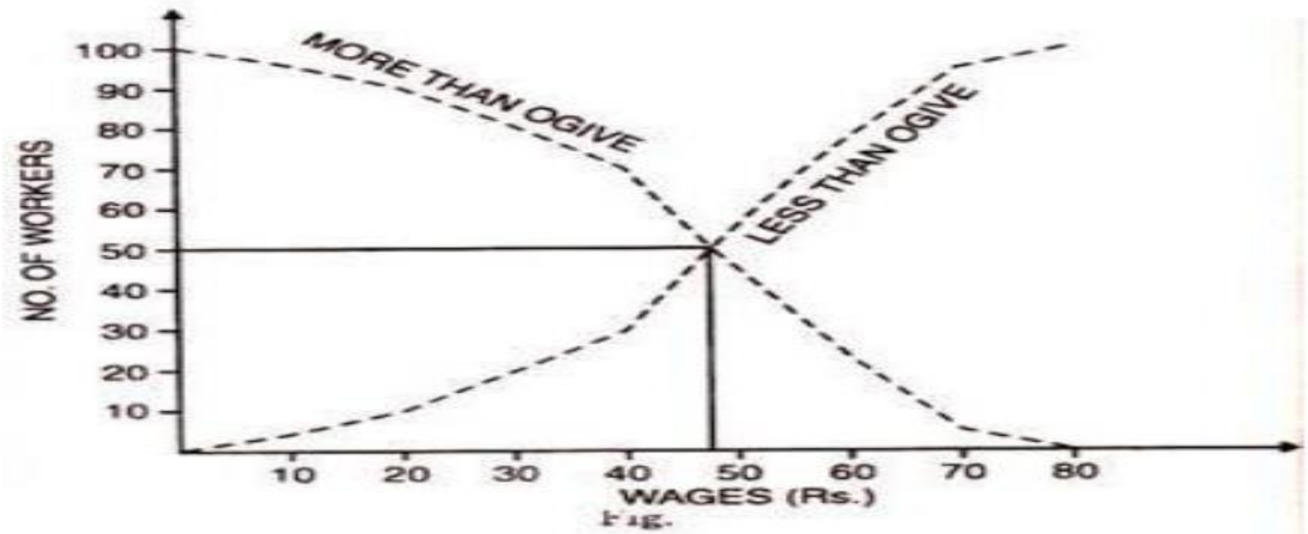
$$Me = L + \frac{\frac{n}{2} - F_c}{f_m} \times h$$

- Here, $\frac{n}{2} = 50$
- L = Lower limit of the Median class=40
- F_c = Cumulative frequency of the pre median class=30
- f_m = Frequency of the median class=25
- h = Width of the median class=10

- n = Total number of observation=100

$$Me = 40 + \frac{50 - 30}{25} \times 10$$

=48



From the avobe graph the median of wages is Rs 48.

Test Yourself

The following table gives the data pertaining to kilowatt hours of electricity consumed by 100 randomly selected flat owners of Japan garden city.

Consumption (in K-watt hours)	**0-100**	**100-200**	**200-300**	**300-400**	**400-500**
No. of users	**6**	**25**	**36**	**20**	**13**

Calculate

i. Mean consumption of electricity
ii. Median use of electricity

Merits of Median-

1. Rigidly defined.
2. Easy to understand and calculate.
3. Not affected very much by extreme values.
4. Can be calculated in the case of the data with open-end class.
5. Can be defined graphically.

Demerits of Median-

1. In case of even number of observations, it is not defined exactly.
2. Not based on all observations.
3. Not easy for algebraic treatment.
4. For calculating median, it is necessary to arrange the data in either ascending or descending order.

Application of Median: Widely used in Business and economic field, Income and wage distribution

Mode- The value of the variable that occurs most frequently; that is for which the frequency is a maximum.

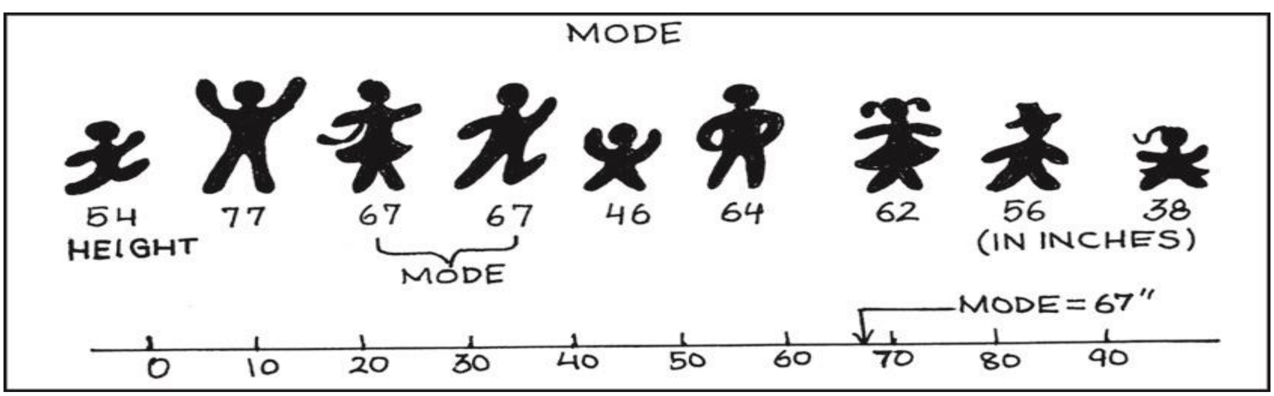
- Generally speaking, mode can be used to describe qualitative data.
- Mode is particularly useful average for discrete data.
- For ungrouped data / categorical variable: mode is the value of the variable for which the **frequency is highest**.

Mode: Ungrouped Data: For the data sets:

i. 7, 8, 6, 7, 9, 7, and 4: Here '7' appears highest 3 times, hence mode is '7'and the data is unimodal.
ii. 6, 4, 8, 5, 8, 1, 2, 5, 4, 7, 5, 2, 4, and 3: here '5' and '4' both occur highest 3 times hence the mode '5' and '4' and the data is bimodal.
iii. 1, 5, 7, 2, 6, 9, and 4: there is no mode.
iv. Consider the following table representing the frequency distribution of religion

Religion	Muslim	Hindu	Buddhist	Christian	Others
Frequency	18	75	12	4	2

Here the highest frequency '75' occurs for the category 'Hindu'. Hence mode for the given data is _ 75_ _ _ _ _.



Determination of Mode: Grouped Data: For group frequency distribution mode is defined as:

$$\text{Mode, Mo} = L + \frac{\triangle_1}{\triangle_1 + \triangle_2} \times h$$

Where,

- Mo/Z = Mode
- L = Lower limit of the modal class for which the frequency is maximum
- $D_1/\triangle_1$ = The difference between the frequency of the modal class and pre-modal class
- $D_2/\triangle_2$ = The difference between the frequency of the modal class and post-modal class
- c/h/i/w = Class interval or the length of the modal class.

We can locate Mode with the help of HISTOGRAM as discussed below with the help of an example.

The process of finding mode from graphical representation is describe below-

GRAPHICAL METHOD

Example 1. Find mode for the following data *Graphically*

X :	0–50	50–100	100–150	150–200	200–250	250–300	300–350
f :	41	171	287	497	382	211	87

and verify numerically.

Solution

Numerically ; By Inspection, Modal Interval is 150 – 200

$\therefore$ L = 150 i = 50

$D_1 = 497 - 287 = 210$;

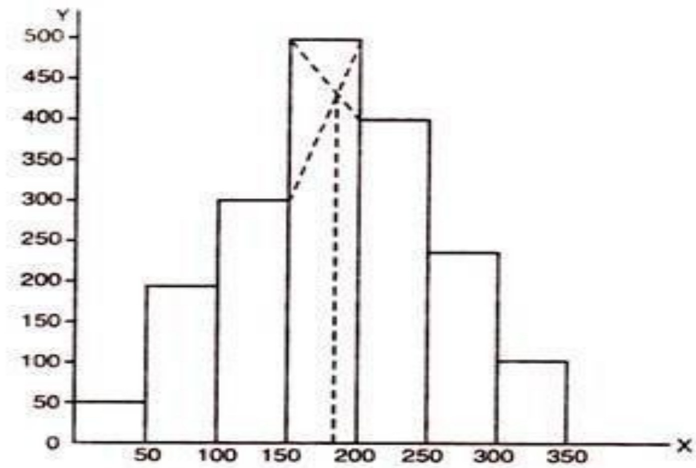
$D_2 = 497 - 382 = 115$

As $Z = L + \frac{D_1}{D_1 + D_2} \times i$

$\therefore$ $Z = 150 + \frac{210}{210 + 115} \times 50$

$= 150 + \frac{10500}{325}$

$= 150 + 32.3 = \mathbf{182.3.}$



Graphical Representation :

Steps 1. Draw Histogram

2. Draw Two lines diagonally inside Modal interval rectangle to upper corner of adjacent rectangles.

3. Draw line from intersection of two diagonals on X-axis.

4. It gives value of Mode.

5. Mode = 182 (Approx.) in the present example.

EXERCISE I

1. Find Z graphically and verify your result numerically also. (1–2)

X :	0–10	10–20	20–30	30–40	40–50	50–60	60–70
f :	4	17	21	44	37	18	3

(**Ans.** 37.7)

2.

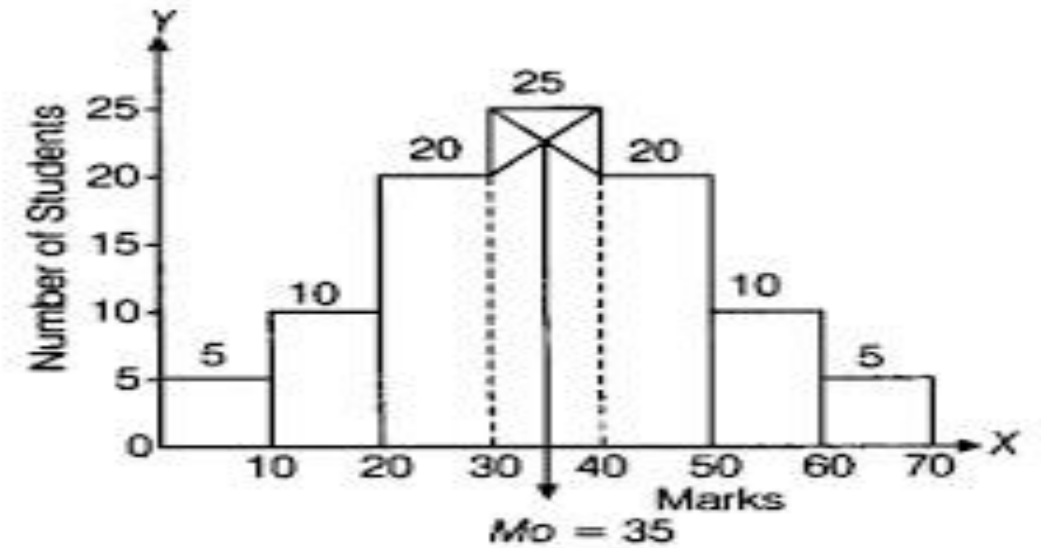
X :	0–50	50–100	100–150	150–200	200–250	250–300	300–350	350–400
f :	27	84	189	296	242	143	77	34

(**Ans.** 183.2)

Q2. Consider this data of marks of students in a class of 95 students.

We have to find mode in this data. Graphically, we draw a histogram for the data. In this data modal class is 30-40 and highest frequency is 25.

Marks	No. of students
0 - 10	5
10 - 20	10
20 - 30	20
30 - 40	25
40 - 50	20
50 - 60	10
60 - 70	5



Example 5·34. *Below is given the frequency distribution of weights of a group of 60 students of a class in a school* :

Weight in kg.	*Number of students*	*Weight in kg.*	*Number of students*
30–34	3	50–54	14
35–39	5	55–59	6
40–44	12	60–64	2
45–49	18		

(*a*) *Draw histogram for this distribution and find the modal value.*

(*b*) (*i*) *Prepare the cumulative frequency* (*both less than and more than types*) *distribution, and* (*ii*) *represent them graphically on the same graph paper. Hence, find the* (*iii*) *median, and* (*iv*) *co-efficient of quartile deviation.*

(*c*) *With the modal and the median values as obtained in* (*a*) *and* (*b*), *use an appropriate empirical formula to find the arithmetic mean of this distribution.*

Application of mode: Mode widely used in science, biological and meteorological condition, production analysis, business forecasting, stock exchange, output analysis, market research etc.

Empirical relationship between Mean, Median and Mode-

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}$$

Merits of Mode-

1. Most typical and representative value of a distribution.
2. Not at all affected by extreme values.
3. Can be calculated in the case of the data with open-end class.
4. Easy to understand and calculate.
5. Can be defined graphically.

Demerits of Mode-

1. Not clearly defined in case of bimodal or multi modal distribution.
2. Not based on all observation.
3. Not suitable for further algebraic treatment.
4. Affected by sampling fluctuations.

Geometric mean: The geometric mean of n non-zero positive observations is the nth root of their product. It is usually denoted by G.M or G and expressed as,

$$\text{Geometric mean, G.M} = (\text{Product of n positive items})^{\frac{1}{n}}$$

For ungroup Data- If there are n non-zero values $x_1, x_2, \ldots, \ldots, \ldots, x_n$ for a variable X, then the

$$\textit{Geometric Mean, } \mathbf{GM} = \sqrt[n]{x_1 * x_2 * \ldots * x_n} = (x1 * x2 * \ldots * xn)^{\frac{1}{n}}$$

- $\log GM = \log\,(x_1 * x_2 * \ldots * x_n)^{\frac{1}{n}}$
- $\log GM = \frac{1}{n}(\log x_1 + \log x_2 + \cdots + \log x_n)$

  $[as,\ \log x^n = n\log x;\ \&\ \log(AB) = \log A + \log B]$
- $GM = \text{Antilog}\left(\frac{\sum \log x_i}{n}\right)$

Find the Geometric Mean of 2, 4, 8, 12, 16 and 24. (Ans. 8·158)

For Group data: - If $x_1, x_2, \ldots, \ldots, \ldots, x_n$ are taken as the values/mid-values of various classes like

Values/Mid-Values:	x_1	x_2	…	…	x_n
Frequencies :	f_1	f_2	…	…	f_n

Such that $f_1 + f_2 + f_3 + \ldots + \ldots + \ldots + f_k = n$, then the GM is defined as G

$$= (X_1^{f_1} \times X_2^{f_2} \times \ldots \times X_n^{f_n})^{1/N}$$

$$\log \text{ G.M.} = \frac{1}{N}\left[\log\left(X_1^{f_1} . X_2^{f_2} \dots X_n^{f_n}\right]\right.$$

$$= \frac{1}{N}\left[\log X_1^{f_1} + \log X_2^{f_2} + \dots + \log X_n^{f_n}\right]$$

$$= \frac{1}{N}\left[f_1 \log X_1 + f_2 \log X_2 + \dots + f_n \log X_n\right]$$

$$= \frac{1}{N}\sum f \log X$$

$$\text{G.M.} = \text{Antilog}\left[\frac{1}{N}\sum f \log X\right]$$

Ungrouped data	Grouped data
$G = Antilog\left(\frac{\sum logx}{n}\right)$	$G = Antilog\left(\frac{\sum f\ logx}{n}\right)$;Here $n = \sum f$

Steps of finding Geometric mean-

1. Find log X, where X is the value of the variable or the mid-value of the class (in case of grouped or continuous frequency distribution).

2. Compute $f \times \log X$ *i.e.*, multiply the values of log X obtained in step 1 by the corresponding frequencies.

3. Obtain the sum of the products $f \log X$ obtained in step 2 to get $\sum f \log X$.

4. Divide the sum obtained in step 3 by N, the total frequency.

5. Take the Antilog of the value obtained in step 4. The resulting figure gives the value of G.M.

If $G_1, G_2, \dots G_k$ are the geometric means of the k groups of sizes $n_1, n_2, \dots, n_k$ respectively, then the geometric mean G of the combined group of size $n_1 + n_2 + \dots + n_k$ is given by :

$$\log G = \frac{n_1 \log G_1 + n_2 \log G_2 + \dots + n_k \log G_k}{n_1 + n_2 + \dots + n_k} \qquad \dots(5{\cdot}23a)$$

Example 5·36. *Find the geometric mean for the following distribution :*

Marks	:	*0—10*	*10—20*	*20—30*	*30—40*	*40—50*
No. of students	:	*5*	*7*	*15*	*25*	*8*

Solution.

Marks	*Mid-Point (X)*	*No. of Students (f)*	*log X*	*f. log X*
0—10	5	5	0·6990	3·4950
10—20	15	7	1·1761	8·2327
20—30	25	15	1·3979	20·9685
30—40	35	25	1·5441	38·6025
40—50	45	8	1·6532	13·2256
		$N = 60$		84·5243

$$\text{Geometric mean} = \text{Antilog}\left[\frac{\sum f \log X}{N}\right] = \text{Antilog}\left[\frac{84{\cdot}5243}{60}\right] = \text{Antilog}\,[1{\cdot}40874] = 25{\cdot}64 \text{ marks.}$$

Weighted Geometric Mean. If the different values *X*1, *X*2,…, *Xn* of the variable are not of equal importance and are assigned different weights, say, *W*1, *W*2, … ,*Wn* respectively according to their degree of importance then their weighted geometric mean G.M. (*W*) is given by:

$$\text{G.M.}(W) = (X_1^{W_1} \times X_2^{W_2} \times \ldots \times X_n^{W_n})^{1/N}$$

where $N = W_1 + W_2 + \ldots + W_n = \sum W$, is the sum of weights.

$$\log [\text{G.M.}(W)] = \frac{1}{N}\left[W_1 \log X_1 + W_2 \log X_2 + \ldots + W_n \log X_n\right] = \frac{1}{N}\sum W \log X$$

$$\text{G.M.}(W) = \text{Antilog}\left[\frac{1}{N}\sum \square W \log X\right] = \text{Antilog}\left[\frac{\sum W \log X}{\sum W}\right]$$

Example 5·45. *The weighted geometric mean of the four numbers 8, 25, 19 and 28 is 22·15. If the weights of the first three numbers are 3, 5, 7 respectively, find the weight (positive integer) of the fourth number.*

Solution. Let the weight of the fourth number be *w*.

Weighted Geometric Mean (*G*) = 22·15 (*Given*)

Also $\log G = \frac{\sum W \log X}{\sum W} \Rightarrow \log 22{\cdot}15 = \frac{\sum W \log X}{\sum W}$

COMPUTATION OF WEIGHTED G.M.

X	*log X*	*W*	*W log X*
8	0·9031	3	2·7093
25	1·3979	5	6·9895
19	1·2788	7	8·9516
28	1·4472	*w*	1·4472*w*
Total		15 + *w*	18·6504 + 1·4472*w*

$$\Rightarrow \log(22{\cdot}15) = \frac{18{\cdot}6504 + 1{\cdot}4472w}{15 + w}$$

$$\Rightarrow (15 + w) \times 1{\cdot}3454 = 18{\cdot}6504 + 1{\cdot}4472w$$

$$\Rightarrow 15 \times 1{\cdot}3454 + 1{\cdot}3454w = 18{\cdot}6504 + 1{\cdot}4472w$$

$$\Rightarrow 20{\cdot}1810 + 1{\cdot}3454w = 18{\cdot}6504 + 1{\cdot}4472w$$

$$\Rightarrow 1{\cdot}4472w - 1{\cdot}3454w = 20{\cdot}1810 - 18{\cdot}6504$$

$$\Rightarrow 0{\cdot}1018w = 1{\cdot}5306$$

$$\Rightarrow w = \frac{1{\cdot}5306}{1{\cdot}1018} = 15 \textit{ approx.}$$

Application of Geometric mean: Geometric mean is usually used for dealing with data related to growth rates (like population growth etc.) or interest rates, index number etc. In many cases, the geometric mean is the best measure to determine the average growth rate of some quantity like-

1. Proportional growth in share prices, interest
2. Determining aspect ratios
3. UN Human Development Index
4. Average of Price fluctuation

Merits of geometric mean:

1. It is rigidly defined.
2. It is based on all the observations of the series.
3. It is suitable for measuring the relative changes.
4. It gives more weights to the small values and less weights to the large values.
5. It is used in averaging the ratios,
   percentages and in determining the rate gradual increase and decrease.
6. It is capable of further algebraic treatment.

Demerits of geometric mean:

1. It is not easy to understand by a man of ordinary prudence as it involves logarithmic operations. As such it is not popular like that of arithmetic average.
2. It is difficult to calculate as it involves finding out of the root of the products of certain values either directly, or through logarithmic operations.
3. It cannot be calculated, if the number of negative values is odd.
4. It cannot be calculated, if any value of a series is zero.

Harmonic Mean: Harmonic Mean is defined as the reciprocal of the arithmetic mean of the reciprocal of the values in a series/observation. If there are n non-zero values $x_1, x_2, \ldots,\ldots,\ldots, x_n$ for a variable X, then harmonic mean is denoted by H.M and defined as-

$$\text{H.M} = \frac{\text{Number of items or observations}}{\text{Sum of reciprocal of items}}$$

For ungroup data: If $X1, X2, \ldots, Xn$ is a given set of *n non-zero* observations, then their harmonic mean, abbreviated as H.M. or simply H is given by :

$$H = \frac{1}{\frac{1}{n}\left[\frac{1}{X_1} + \frac{1}{X_2} + \ldots + \frac{1}{X_n}\right]} = \frac{1}{\frac{1}{n}\Sigma\left(\frac{1}{X}\right)} = \frac{n}{\Sigma\left(\frac{1}{X}\right)}$$

For group data:

In case of frequency distribution, we have

$$\frac{1}{H} = \frac{1}{N}\left[\frac{f_1}{X_1} + \frac{f_2}{X_2} + \ldots + \frac{f_n}{X_n}\right] = \frac{1}{N}\Sigma\left(\frac{f}{X}\right) \qquad \Rightarrow \qquad H = \frac{N}{\Sigma\,(f/X)}$$

where $N = \Sigma f$, is the total frequency, X is the value of the variable or the mid-value of the class (in case of grouped or continuous frequency distribution) and f is the corresponding frequency of X.

The following are formulae of harmonic mean:

Ungrouped data	Grouped data
$H = \frac{n}{\Sigma\left(\frac{1}{x}\right)}$	$H = \frac{n}{\Sigma\left(\frac{f}{x}\right)}$; Here $n = \Sigma f$

Example 5·46. *The following table gives the weights of 31 persons in a sample enquiry. Calculate the mean weight using (i) Geometric mean and (ii) Harmonic mean.*

Weight (lbs.) :	*130*	*135*	*140*	*145*	*146*	*148*	*149*	*150*	*157*
No. of persons :	*3*	*4*	*6*	*6*	*3*	*5*	*2*	*1*	*1*

Solution.

COMPUTATION OF G.M. AND H.M.

Weight (lbs.) (X)	*No. of persons* (f)	$\log X$	$f \log X$	$\frac{1}{X}$	$\frac{f}{X}$
130	3	2·1139	6·3417	0·00769	0·02307
135	4	2·1303	8·5212	0·00741	0·02964
140	6	2·1461	12·8766	0·00714	0·04284
145	6	2·1614	12·9684	0·00690	0·04140
146	3	2·1644	6·4932	0·00685	0·02055
148	5	2·1703	10·8515	0·00676	0·03380
149	2	2·1732	4·3464	0·00671	0·01342
150	1	2·1761	2·1761	0·00667	0·00667
157	1	2·1959	2·1959	0·00637	0·00637
	$\sum f = N = 31$		$\sum f \log X = 66{\cdot}7710$		$\sum (f/X) = 0{\cdot}21776$

$$\text{G.M.} = \text{Antilog}\left(\frac{1}{N}\sum f \log X\right) = \text{Antilog}\left(\frac{66{\cdot}7710}{31}\right) = \text{Antilog}\,(2{\cdot}1539) = 142{\cdot}5$$

$$\text{H.M.} = \frac{N}{\sum (f/X)} = \frac{31}{0{\cdot}21776} = 142{\cdot}36$$

Hence, the mean weight of 31 persons using (*i*) geometric mean is 142·5 lbs. and (*ii*) harmonic mean is 142·36 lbs.

If H_1 and H_2 are the harmonic means of two groups of sizes N_1 and N_2 respectively, then the harmonic mean H of the combined group of size $N_1 + N_2$ is given by :

$$\frac{1}{H} = \frac{1}{N_1 + N_2}\left[\frac{N_1}{H_1} + \frac{N_2}{H_2}\right] \qquad \ldots(5{\cdot}28c)$$

Application of Harmonic Mean:

- Harmonic mean is used for calculating mean data for values obtained by combining two scales, like distance and time for speed.
- Harmonic means are often used in **averaging ratios** (e.g., the average travel speed given a duration of several trips, combined work rates).
- The weighted harmonic mean is used in finance to average multiples like the price-earnings ratio because it gives equal weight to each data point.
- Other uses: Thin lens equation in optics, parallel resistors in electricity etc.

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.
5. It is rarely used in grouped data.

Test Yourself

Evan was driving up a mountainous road, with some flat stretches among the steep climbs. As a result, his speed varied for each 10 kilometers, as seen below:

Distance Travelled (km)	10	10	10	10	10
Speed (km/h)	10	25	6	15	8

What was his average speed throughout the drive?

Weighted Harmonic Mean: If the different values $X1$, $X2$,…, Xn of the variable are not of equal importance and are assigned different weights, say, $W1$, $W2$,…,Wn respectively according to their degree of importance then their weighted Harmonic mean H.M. (W) is given by:

$$H.M_w = \frac{W1 + W2 + \cdots + Wn}{\frac{W_1}{x_1} + \frac{W_2}{x_2} + \cdots + \frac{W_n}{x_n}}$$

$$= \frac{\sum W_i}{\sum \frac{w_i}{x_i}}$$

Example 5·53. *You make a trip which entails travelling 900 kms. by train at an average speed of 60 km. p.h.; 3000 kms. by boat at an average of 25 km. p.h.; 400 kms. by plane at 350 km. p.h., and finally, 15 kms by taxi at 25 km. p.h. What is your average speed for the entire distance ?*

Solution. Since different distances are covered with varying speeds, the required average speed is given by the weighted harmonic mean of the speeds (in km. p.h) 60, 25, 350 and 25; the corresponding weights being the distances covered (in kms.) *viz.*, 900, 3000, 400 and 15 respectively.

$$\therefore \quad \text{Average Speed} = \frac{\sum W}{\sum (W/X)} = \frac{4315}{137{\cdot}03} = 31{\cdot}49 \text{ km. p.h.}$$

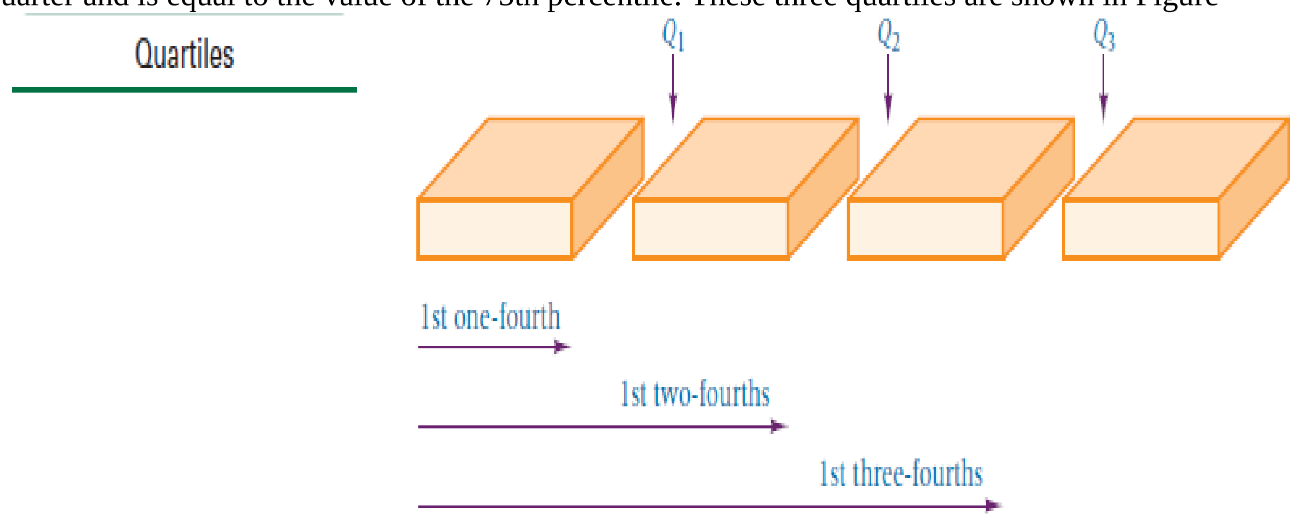
COMPUTATION OF WEIGHTED H.M.

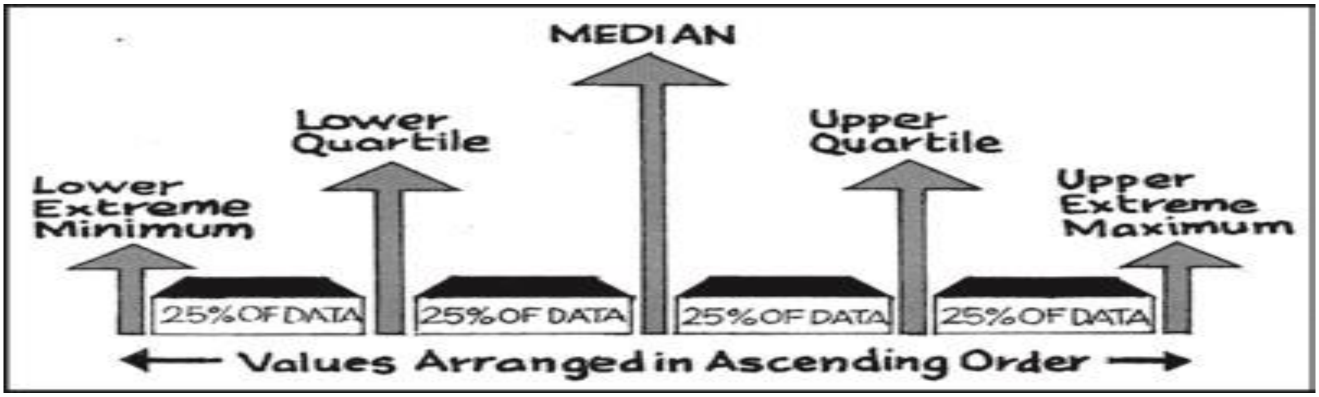
X	W	W/X
60	900	15
25	3000	120
350	400	1·43
25	15	0·60
	$\sum W = 4315$	$\sum (W/X) = 137{\cdot}03$

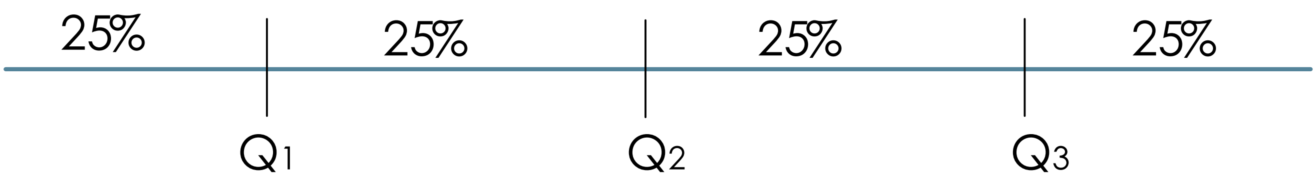
Some positional measure (Quartiles, Deciles, Percentiles):

Partition values: The points which divide the data in to equal parts are called Partition values.

Quartiles: If the items in a series are arranged in ascending order of their magnitudes, then those values of the variable that divide the total frequency in to four equal parts are called quartiles. Quartiles are measures of central tendency that divide a group of data into four subgroups or parts. There are three quartiles denoted as Q1, Q2, and Q3. The first quartile, Q1, separates the first, or lowest, one-fourth of the data from the upper three-fourths and is equal to the 25th percentile. The second quartile, Q2, separates the second quarter of the data from the third quarter. Q2 is located at the 50th percentile and equals the median of the data. **The second quartile (Q_2) coincides with the median.** The third quartile, Q3, divides the first three-fourth/quarters of the data from the last quarter and is equal to the value of the 75th percentile. These three quartiles are shown in Figure







Determination of Quartile for ungroup data:

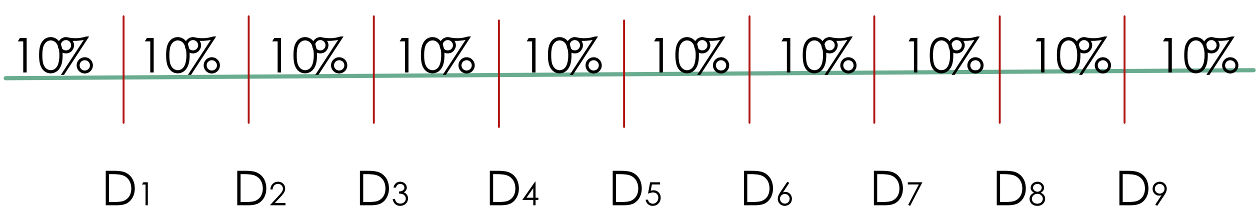
- For raw data,
  - If $i = 1,2,3$
  - $n = number\ of\ values$

Then, $i^{th} quartile,\ Q_i = \frac{1}{2}\ [\left(\frac{in}{4}\right)^{th}\ value + (\frac{in}{4} + 1)^{th}\ value)],\ when\ \frac{in}{4}\ is\ an\ integer$

Or, $i^{th} quartile,\ Q_i = next\ integer^{th} value\ of\ \frac{in}{4},\ when\ \frac{in}{4}\ is\ not\ an\ integer$

Deciles

- Deciles (Denoted by $D_1, D_2, \dots, \dots, \dots, D_9$) are the quantities of the variable that divides the total frequencies into 10 equal parts.



Note: Median divides the dataset into 2 equal halves, thus coinciding with D_5

Percentiles: Percentiles are measures of central tendency that divide a group of data into 100 parts. There are 99 percentiles because it takes 99 dividers to separate a group of data into 100 parts. Percentiles denoted by $P_1, P_2, \dots, \dots, \dots, P_{99}$

Note: Median = $\boldsymbol{Q_2 = D_5 = P_{50}}$ >> All these quantities divide the total frequencies in to two equal halves.

Determination of Decile and Percentile for ungroup data:

- For raw data,
  - If $i = 1,2,3,\ldots\ldots 9$ (For Decile) or $1,2,3,\ldots\ldots\ldots\ldots,99$
  - $n = number\ of\ values$

Then, $i^{th} decile,\ D_i = \frac{1}{2}\ [\ \left(\frac{in}{10}\right)^{th}\ value + (\frac{in}{10} + 1)^{th}\ value)],\ when\ \frac{in}{10}\ is\ an\ integer$

Or, $i^{th}\ decile,\ D_i = next\ integer^{th} value\ of\ \frac{in}{10},\ when\ \frac{in}{10} is\ not\ an\ integer$

And, $i^{th}\ percentile,\ P_i = \frac{1}{2}\ [\ \left(\frac{in}{100}\right)^{th}\ value + (\frac{in}{100} + 1)^{th}\ value)],\ when\ \frac{in}{100}\ is\ an\ integer$

Or, $i^{th}\ percentile,\ P_i = next\ integer^{th} value\ of\ \frac{in}{100},\ when\ \frac{in}{100} is\ not\ an\ integer$

Example: Following is the marks in STA-201 obtained by 20 students in Summer 2020.

99	75	84	33	45	66	97	69	55	61
72	91	74	93	54	76	62	91	77	68

Find out:

- 1st and 3rd Quartiles
- 3rd, 6th, 8th Deciles
- 20th, 37th, 60th, 86th Percentiles

Solution:

- First, arrange the data.

33	45	54	55	61	62	66	68	69	72
74	75	76	77	84	91	91	93	97	99

Here, n =20

1st Quartile:

For *i=1*, $\frac{in}{4} = \frac{1*20}{4}$ = 5, an integer

So, $1^{st}\ quartile,\ Q_1 = \frac{1}{2}\ [\ 5^{th}\ value + (5+1)^{th}\ value)]$ = ½ (61+62) = 61.5

6th Decile:

For i=6, $\frac{in}{10} = \frac{6*20}{10}$ = 12, an integer

So, $6^{th}\ decile,\ D_6 = \frac{1}{2}\ [\ 12^{th}\ value + 13^{th}\ value] =$ ½ (75+76) = 75.5

37th Percentile:

For *i=37*, $\frac{in}{100} = \frac{37*20}{100}$ = 7.4, not an integer

Next integer = 8

So, $37^{th}\ percentile,\ P_{37} = 8^{th}\ value = 68$

For grouped data: First obtain the cumulative frequencies for the data. Then mark the class corresponding to which a cumulative frequency is greater than (in)/4 for the first time. (Where n is total number of observations.). Then that class is Qi class. Then Qi is evaluated by interpolation formula.

$$Q_i = L + \frac{\frac{in}{4} - F_c}{f_q} \times h \qquad i = 1, 2, 3$$

Where, L= lower limit of the Qi class.

n = Number of observations.

previous class

Fc = cumulative frequency of the class proceeding to the Qi class.

f_q = frequency of the Qi class.

h = Class interval

Deciles are nine points which divided the data in to ten equal parts and denoted by Di.

Di is the value corresponding to $(in/10)^{th}$ observation after arranging the data in the increasing order.

For grouped data: First obtain the cumulative frequencies for the data. Then mark the class corresponding to which a cumulative frequency is greater than (in/10) for the first time. (Where n is total number of observations.). Then that class is Di class. Then Di is evaluated by interpolation formula.

$$D_i = L + \frac{\frac{in}{10} - F_c}{f_d} \times h \qquad i = 1, 2, \ldots\ldots\ldots\ldots 10.$$

Where, L = lower limit of the Di class

n= Number of observations.

Fc = cumulative frequency of the class proceeding to the Di class.

f_d = frequency of the Di class.

Percentiles are ninety-nine points which divided the data in to hundred equal parts denoted by Pi. Pi is the value corresponding to $(in/100)^{th}$ observation after arranging the data in the increasing order.

For grouped data: First obtain the cumulative frequencies for the data. Then mark the class corresponding to which a cumulative frequency is greater than (in/100) for the first time. (Where n is total number of observations.) Then that class is Pi class. Then Pi is evaluated by interpolation formula.

$$P_i = L + \frac{\frac{in}{100} - F_c}{f_p} \times h \qquad i=1,2, \dots , 99$$

Where, L = lower limit of the Pi class

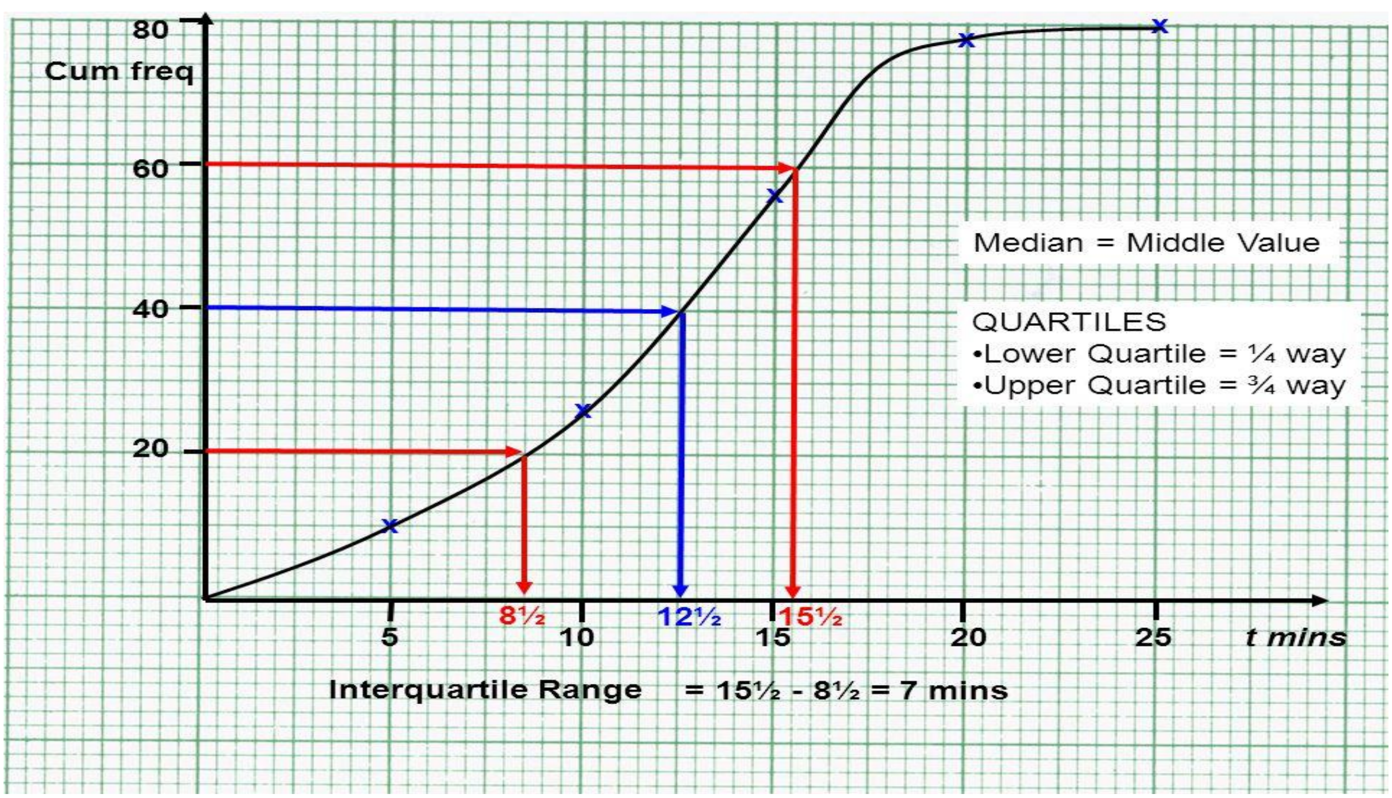
n= Number of observations.

Fc = cumulative frequency of the class proceeding to the Pi class.

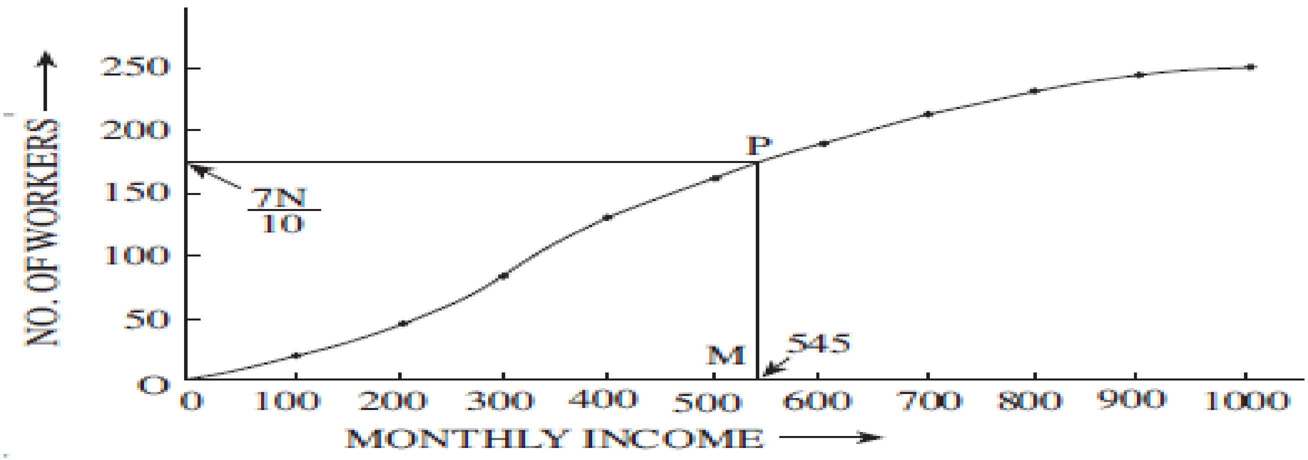
f_p = frequency of the Pi class.

Graphical method for locating partition values: These partition values can be located graphically by using ogives. The point of intersection of both ogives is median.

To locate quartiles, mark (in/4) on Y- axis, from that point draw a line parallel to X-axis, it cuts less than type ogive at Q_1 and intersects greater than or equal to curve at Q_3.



To locate Di mark in/10 on Y-axis, from that point draw line parallel to X-axis, it intersects less than type curve at Di.



Similarly, to locate Pi mark in/100 on Y-axis, from that point draw line parallel to X-axis, it intersects less than type curve at Pi.

Example: Find the median, Q_1, D8, P65 from the following data.

Marks	0-10	10-30	30-50	50-80	80-90	90-100
No of Students	4	12	20	8	4	2

Solution: To locate median class we have to calculate cumulative frequencies.

Marks	0-10	10-30	30-50	50-80	80-90	90-100
No of Students	4	12	20	8	4	2
Cumulative freq	4	16	36	44	48	50

Here N=50 so N/2=25, Hence median class is 30-50

$$Me = L + \frac{\frac{n}{2} - F_c}{f_m} \times h$$

$$= 30 + \frac{25-16}{20} \times 20 = 39$$

Here N=50 so N/4=12.5, hence Q1 class is 10-30

$$Q_1 = L + \frac{\frac{n}{4} - F_c}{f_q} \times h \quad = 10 + \frac{12.5-4}{12} \times 20 = 24.16$$

Here N=50 so 8*N/10=40 , hence D8 class is 50-80

$$D_8 = L + \frac{\frac{8n}{10} - F_c}{f_d} \times h \quad = 50 + \frac{40-36}{8} \times 30 = 65$$

Here N=50 so 65*N/100=32.5, hence P65 class is 30-50

$$P_{65} = L + \frac{\frac{65n}{100} - F_c}{f_p} \times h = 30 + \frac{32.5-16}{20} \times 20 = 46.5$$

Assignment: 005

The number of days of absenteeism of 80 workers of a factory over a particular year are recorded as follows:

Days	No. of workers
0-3	5
4-7	14
8-11	17
12-15	25
16-19	10
20-23	6
24-27	3

a) Calculate the average days of absenteeism by A.M, G.M, & H.M.
b) Calculate median, mode, and Q_3 of days of absenteeism and comment.
c) If the authority decides to terminate 30% of the most irregular workers, how can you help the authority to take decision in this case?
d) On the other hand, if the authority decides to reward 30% of the most regular workers how can you help the authority to take decision?
e) Locate D_3, median, P_{70} and Q_3 graphically.