13. **Calculate weighted mean of the following :**

Size :	6	12	18	24	30	36
Weights :	8	15	22	16	25	14

14. **Compare the results of two colleges from the following data and comment which college gives the better result ?**

	College A		College B	
Faculty	**No. of Students**		**No of Students**	
	Appeared	**Passed**	**Appeared**	**Passed**
Arts	600	480	100	70
Science	150	135	30	24
Commerce	250	200	70	56
Total	1000	885	200	150

15. **An average of twenty items is 30. The average of first 19 items is 25. Find out the value of the last item.**

□□□

Lesson–16

GEOMETRIC MEAN

In the previous lesson we have discussed the simple and weighted arithmetic mean in detail. Now, in this lesson we shall discuss simple and weighted geometric mean in detail.

Geometric Mean–Geometric mean is the nth root of the product of n items of a series. If we have to calculate geometric mean of 3, 4, 8, we should have to compute the 3rd root of the product of these three figures. Geometric mean is generally written as G. M. The geometric mean of the above figures will be—

$$GM = \sqrt[3]{3 \times 4 \times 8}$$

Similarly the geometric mean of 4, 8, 6 and 5 would be the 4th root of the product of these four figures.

Symbolically—

$$GM = \sqrt[n]{X_1, X_2, X_3 \ldots\ldots X}$$

Here–GM is Geometric mean, X_1, X_2,.....X_n are the observed values of the variable X and n is the total number of values in the series.

Example–1.

Find out the geometric mean of following–

1, 4, 5, 50 and 100.

Solution—

Here n = 5

Thus, GM $= \sqrt[5]{1\times 4\times 5\times 50\times 100}$

$= \sqrt[5]{10000}$

$= \sqrt[5]{10\times 10\times 10\times 10\times 10} = 10$

GM = 10

The calculation of geometric mean by the above process becomes most difficult with the increase of items and their values in the series. In such conditions calculations are done with the help of logs. Logtarthmic value of each item is computed with the help of logarithmic table. A separate column of log values are constructed. They are added together and the sum total is divided by the number of items in the series. The anti-log of the quotient so obtained is the geometric mean.

Symbolically—

$$GM = \text{Anti-log of } \left| \frac{\sum_{i=1}^{n} \log Xi}{n} \right|$$

Here, $\sum_{i=1}^{n} \log xi$ is the sum total of log values of all the items in the series and N is the total number of items in the series.

Log and Anti-log, these two items have been mentioned in the above formula for computation of geometric mean. Now, we shall discuss these terms before discussing any other point.

Logarithm–Logarithm of a number, with a base 10 is the power to which 10 must be raised to give the number. For example—

10 = 10^1. Thus log of 10 is 1.

100 = 10^2. Thus log of 100 is 2.

1000 = 10^3. Thus log of 1000 is 3.

Such log numbers can be found without any difficulty. We have used 10 as base. Any other number can be used as base but due to case in calculation the base of 10 popularly used. The base of 10 is always understood where the base is not clearly mentioned.

It is clear from the above that log of 10 is 1 and log of 100 is 2. It shows that numbers between 10 and 100 have a logarithm in decimal between 1 and 2. Thus we can say that the log value of a number consists of two parts. The first is the integral value and is the second is the decimal value. Both the parts of a log are known as characteristics and mantissa respectively.

Thus to find out the log-value of any number one has to find out characteristics and mantissa. There are separate procedures for finding out the two values, characteristics is determined by applying the following rules—

1. If the number given is greater than or equal to one the charateristics will be 1 less than the number of digits to the left of the decimal point.

Symbolically—

Characteristics = n – 1

Here, n = number of digits left to the decimal point. i.e. the number of digits in the intergral part of the number. This can be explained with the help of the following.

Example–

Number	Characteristics
8675	(4 – 1) = 3
546	(3 – 1) = 2
30	(2 – 1) = 1
5	(1 – 1) = 0
27.3	(2 – 1) = 1
158.46	(3 – 1) = 2
5312.325	(4 – 1) = 3
4120.035	(4 – 1) = 3

2. If the number given is less than one, the characteristics will be negative and 1 greater than the number of zeros between the decimal point and the first significant figure of the given number.

Characteristics = – (n + 1)

Here, n is the number of zero between the decimal point and the first significant digit of the given number.

This will be clear from the following example–

Numbers		Characteristics
.524	(nil + 1)	$\overline{1}$.
.0325	(1 + 1)	$\overline{2}$.
.00659	(2 + 1)	$\overline{3}$.
.05032	(1 + 1)	$\overline{2}$.
.0004	(3 + 1)	$\overline{4}$.

The first number given in the above example is.524. There is no zero between decimal point and 5 which is the first significant digit in the given number. In order to find out characteritics of the number we have added 1 in the number of zero which is nil. Thus characteristics is (nil + 1) 1. The second number given is .0325 which consists of one zero between the decimal point and 3 the first significant digit. Thus its characteristics = (1 + 1)= 2. Characteristics for other numbers have been found out similarly.

Mantissa–Mantissa values are located with the help of log-table. They are always positive. Log value for any negative figure can never be computed. We ignore the decimal point in the given number while locating mantissa value. Generally, we use a table consisting of various

numbers of four digits for finding out mantissa value. This, if we have a number consisting of five or more digits we should first change them into a number of four digits by approximation method before reading mantissa value from the log table. The first two digits of the number so approximated are seen in the left-hand vartical column and we read the figures given before it in the column heading of the third digit. To this figure, thus obtained, we add the quantity appearing under the 'mean difference under column heading of the fourth digit. For example, if we are to find out the log value of 3645 then its characteristics will be 3 and for matissa the first two digits will be located in the first column of the log-table. After locating 36 in the first column of the log-table we will read the number written against 36 in the column heading four because the third digit in the given number i.e. 3645 is four. The required number is .5611. For the fourth digit in the given member (3645) which is 5 we will read the value written fourth digit in the given column in the mean difference section of the log-table which is 6. This will be added to .5611. Thus mantissa for 3645 will be .5617. Therefore, the log value for the number 3645 will be 3.5617.

If the given number is 680,056 and we are asked to find out the log-value, the characteristics in this case will be 2 and mastissa will be found out for 680056. As there are 6 digits and we can read mantissa value of a number of four digits only. the said number will be approximated to four digits which will be 6801. Thus finally mantissa will be found out for 6801.

After locating the first two digits (68) in the first column we will note the value written in the column heading zero which comes as 8335. We will locate the value for the fourth digit in the given number (6801) and its value will be located in the respective column in the mean difference section against 68. The required value is 1. This will be added to .8325. This way the mantissa would be .8325 + 1 = 8326.

Finally :

Log-value for 680.056 = 2.8326.

Suppose we are asked to find out the log of 87. This is a number of two digits. Its characteristics will be 1. For its mantissa the value in the zero column will be read after locating 87 in the first column of the log-table which is 9395.

Thus log of 87 = 1.9395.

In the above case we have assumed a zero after 87 because there is no third digit in the given number. In the same way if the number given is 7 and we are to find out its mantissa value we will have to make it a three digit number by assuming two zeroes after it. Thus mantissa value can be read a number consisting less three digits. The mantissa values for 70, 700, 7000.....will be identical.

If the number given is valued less than one, for example. 004336, its characteristics will be 3. and mantissa will be seen for 436 which will be 6395. Therefore, log value for .00436 = $\overline{3}$. 6395.

Antilogarithm–Antilograithm table is used to get real number from the log-values. In order to find out antilog of a given number we take up only the given number. The procedure for consulting anti-log table is the same as for log table. The place of decimal in the real number is determined by the characteristics part of the given numbers i.e., the value of the digits before the decimal point in the given number. In the case of positive characteristics decimal point will be placed n + 1 digits from

left to right in the antilog of the mantissa value. In the case of negative characteristics the decimal point will be placed n – 1 digits left from its place in the antilog of the mantissa value. This will be clear from the examples given below—

log	Antilog
2.6342	430.7
1.6342	43.07
3.6342	4307.0
0.6342	4.307
1.6342	.4307
3.63242	.004307

Calculation of Geometirc Mean

Individual Series–The following steps are taken to calculate geometric mean of an individual series–

1. Log value for each number is found out two with the help of a log table.
2. Summation of log values (Σlog) is found out.
3. Summation of logs is divided by the number of items (N) in the series.
4. The aniti-log of the quotient is found out with the help of anti-log-table. This serves as geometirc mean of the series.

The following example may explain the procedures–

Example–2

Com pute geometric mean of the following–

X : 5, 25, 45, 8, 100, 150, 175, 4129, 6275, 101564.

Solution–

X	logs
5	0.6090
25	1.3979
45	1.6532
8	0.9031
100	2.0000
150	2.1761
175	2.2430
4129	3.6158
6275	3.7976
101564	5.0069
N = 10	Σlogs = 23.4926

$$GM = \text{Antilog of } \left(\frac{\Sigma \log}{N}\right)$$

$$= \text{Antilog of } \frac{33.4926}{10}$$

= Antilog of 2.34926 = 223.6

∴ G. M. = 223.6.

Example 3.

Calculate geometirc mean of the following–

X : 0.62, .0783, .005467, 1.234, 54.65, 674.82, 3403.83, 3403.21, 54302.10, .000823, .0000602.

Solution :

X	logs
.62	$\bar{1}.7924$
.0783	$\bar{2}.8938$
.005467	$\bar{3}.7387$
1.234	0.0913
54.65	1.7476
674.82	2.8292
3403.2	3.6319
54302.1	4.7348
.000823	$\bar{4}.9154$
.0000602	$\bar{5}.7796$
N = 10	Σlogs = 2.0438

G. M. = Antilog of $(\frac{\Sigma \log}{N})$

= Antilog of $\frac{2.0438}{10}$

= Antilog of 0.20438 = 1.601

∴ G. M. = 1601 Ans.

Example 4.

Calculate geometric mean of the following–

.8974, .0570, .0081, .5677, .0001, .0984, .0854 and .5672

Solution :

X	log
.8974	$\bar{1}.9530$
.0570	$\bar{2}.7559$
.0081	$\bar{3}.9085$
.5677	$\bar{1}.7541$
.0002	$\bar{4}.4010$
.0984	$\bar{2}.9930$
.0854	$\bar{2}.9315$
.5672	$\bar{1}.7538$
N = 8	Σlogs = 10.3508

G. M. = Antilog of $(\frac{\Sigma \log s}{N})$

= Antilog of $\frac{\overline{10}.3508}{8}$

= Antilog of $\frac{-16}{8}\ \frac{+6.3508}{8}$

= Antilog of $\overline{2}.7938 = .0622$

Explanation–In the above example we find that integral part of the summation logs or characteristics is negative and not divisible by N. In such cases we make a modification in the summation logs and make the characteristic fully divisible by N. For this we deduct a minimum number from the characteristic which can make it divisible by N. This minimum number is added in the mantissa side. After doing that we divide both the characteristics and the mantissa by N separately. In the problem stated above we have substracted 6 from the characteristic and have added 6 in the mantissa for making the charcteristic full divisible by 8.

(ii) Discrete Series–In order to calculate Geometric mean in a discrete series, the following steps are taken—

(A) Logarithms of all the values are found out.

(B) Logarithms are multiplied by their respective frequencies and their products are added together. In other words $\Sigma(\log \overline{Xi})$ fi is found out.

(C) In the last, Geometirc mean is calculated with the use of the following formula—

G. M. = Antilog of $(\frac{\Sigma \log sXi}{N} fi)$

Examples 5.

The following example will explain the procedures involved in it–

X	:	2	4	6	8	10
f	:	5	7	8	3	2

Solution :

X	f	logX	logXf
2	5	0.3010	1.5056
4	7	0.6021	4.2147
6	8	0.7782	6.2256
8	3	0.9031	2.7093
10	2	1.0000	2.0000
	N = 25		Σlog × f = 16,6546

G. M. = Antilog of $(\frac{\Sigma(\log X)f.}{N})$

= Antilog of $\frac{16.6546}{25}$

= Antilog of 0.6662 = 4.636

(iii) Continuous Series–In order to calculate geometirc mean in a continuous series mid values for all the class intervals are found out. After this the process of calculation is the same which we have already discussed in the diserete series. Σ log f is obtained by multiplying the log value of the different mid points with their respective frequencies. Products so obtained are added together and divided by N.

Symbolically—

G. M. = Antilog of $(\frac{\Sigma(\log X)f}{N})$

Example 6

Compute geometric mean of the following series—

Class :	0-10	10-20	20-30	30-40	40-50	
f	:	2	3	4	3	2

Solution—

Class	f.	Mid-values (X)	logs	log Xf
0-10	2	5	0.6910	1.3980
10-20	3	15	1.1761	3.5283
20-30	4	25	1.3979	5.5916
30-40	3	35	1.5441	4.6323
40-50	2	45	1.6532	3.3064
	N = 14			Σlog f = 18.4566

G. M. = Antilog of $(\frac{\Sigma(\log X)f}{N})$

= Antilog of $\frac{18.4565}{14}$

= Antilog of 1.3183 = 20.81

∴ G. M. = 20.81 Ans.

Weighted Geometric Mean

Weighted geometric mean is denoted as Wgm or W. Gm. Process of calculation is same as in the frequency series i.e. discrete or continuous series we use weight (w) in place of frequency (f) in discrete or continuous series. First of all we find out the log values of all the items. Log values are multiplies with their respective weight (w) and the sum total of such products (Σ log w) is found out. this total is divided by sum total of weights.

W. Gm = Antilog of $(\frac{\Sigma(\log X)w}{\Sigma w})$

Example 7.

Calculate weighted geometric mean from the following–

X–	5	10	15	20	25	30
W–	8	12	10	10	6	4

Solution :

X	W	log (X)	Log X W
5	8	0.6990	5.5920
10	12	1.0000	12.0000
15	10	1.1761	11.7610
20	10	1.3010	13.0100
25	6	1.3979	8.3874
30	4	1.4771	5.9084
	50		56.6588

W. Gm = Antilog of $(\frac{\Sigma(\log X)w}{\Sigma w})$

= Antilog of $\frac{56.6588}{50}$

= Antilog of 1.1332 = 13.59

W. Gm = 10.59 Ans.

Combined Geometric Mean :

If geometric mean of the two or more series and number of items in each series are known then a combined geometric mean of all such series taken together can be calculated with the help of the following formula—

G. M_{12} = Antilog of $(\frac{N_1.\log Gm_1X)+(N_2\log Gm_2)}{N_1+N_2})$

Here–G.M_{12} = Combined Geometric Mean.

Gm_1 and Gm_2 = Geometric means of different series.

N1 and N2 = Number of items in different series.

The following example will make the position clear–

Example 8.

Calculate the combined geometric mean of two series whose geometric meean are 8 and 6 respoectively. The number of items in the first series is 4 and in the second series is 2.

Solution :

Here–$N_1 = 4$ G.$M_1 = 8$

$N_2 = 2$ G. $M_2 = 5$

We know that—

$$Gm_{12} = \text{Antilog of } \left(\frac{(N_1.\log Gm_1)+(N_2 \log Gm_2)}{N_1+N_2}\right)$$

By putting the values in the above formula.

$$Gm_{12} = \text{Antilog of } \left(\frac{(4\times \log 8)+(2\times \log 5)}{4+2}\right)$$

$$= \text{Antilog of } \left(\frac{(4\times 0.9031)+(2\times 0.6990)}{6}\right)$$

$$= \text{Antilog of } \left(\frac{3.6124+1.3980}{6}\right)$$

$$= \text{Antilog of } \left(\frac{5.0104}{6}\right)$$

= Antilog of 0.8351 = 6.841

Combined geometric mean = 6.841 Ans.

Geometric mean is used in calculation of average percentage rate of increase or decrease e.g. Growth rate of population, cumulative interest or percentage change in prices etc. The following example can explain the use of geometric mean–

Example 9.

The population of a village was 1950 in the Year 1930. It rose to 3467 in the Year 1990. What was the average percentage rate of growth of population per year ?

Solution :

As the growth in population is of cumulative effect. geometric mean will be used to calculate the annual average percentage rate of growth. Following formula will be used—

$$Pn = Po\ (1 + r)n \text{ or } r + n\sqrt{\frac{Pn}{Po}} - 1$$

Here—

Pn = Population at the end of the period = 3467

Po = Population at the beginning = 1950

r = Rate of growth per unit.

n = Time in years = (1990–1930 = 60) years

By putting the value in the above formula–

$$r = 60\sqrt{\frac{3467}{1950}} - 1$$

$$= \text{Antilog of } \left(\frac{\log 3467-\log 1650}{60}\right) - 1$$

$$= \text{Antilog of } \left(\frac{3.4503 - 3.2900}{60}\right) - 1$$

$$= \text{Antilog of } \left(\frac{.2503}{60}\right) - 1$$

= Antilog of (0.00472)–1

=1.009-1 = .009

Percentage growth = .009 × 100 = 0.9 %

Example 10.

The annual rates of output of a factory in five years are 5.0, 7.5,2.5, 5.0 and 10.0 percent respectively. What is the average rate of growth of output per annum for the period ?

Solution :

Year	output	logs
I	100 + 5 = 105	2.0212
II	100 + 7.5 = 107.5	5.0314
III	100 + 2.4 = 102.4	2.0106
IV	100 + 5.0 = 105.0	2.0212
V	100 + 10.0 = 110.0	2.0414
		10.1058

$$\text{G. M.} = \text{Antilog of } \left(\frac{\Sigma \log s}{N}\right)$$

$$= \text{Antilog of } \frac{10.1258}{5}$$

= Antilog of 2.02516 = 105.9

Rate of growth = 105.9–100 = 5.9%

Merits of Geometric Mean :

1. Geometric mean is based on all the observations. This makes it more representative average.
2. It is capable of further algebraic treatment.
3. As it is less affected by extreme values in the series. It is more typical average than the arithmetic mean.

Demerits of Geometric Mean :

1. Its calculation is very complicated. It requires high level skill to understand it.
2. Knowledge of all the items is a must for its calculation. If some of the values in a series are not known geometric mean for the series cannot be calculated.
3. It can not be calculated if the size of any of the variable is either 'O' or 'n' is in negative.

Use of Geomeric Mean :

1. Gemoetric mean is the most appropriate average to calculate average of ratio or percentage rates of growth etc.
2. It is an average most suitable where large weights have to be given to small items and small weights to large items.
3. Geometric mean provides a satisfactory measure of computing rate of growth of economic phenomena, specialy that phenomena which grows in geometric progression or which have cumulative effect.
4. Geometric mean is suitable for highly skewed distributions.

Model Questions

1. **Calculate geometric mean of the following—**

   4328, 592.35, 83.65, 5.832, 0.7698, 0.05263, 0.00506 and 50.062

2. **Find the geometric mean of the following–**

Size	:	**4**	**8**	**12**	**16**	**20**	**24**
f	:	**15**	**10**	**3**	**20**	**2**	**18**

3. **Compute geometric mean of the following–**

Class	:	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**
f	:	**5**	**40**	**10**	**50**	**5**

4. **If the price of a commodity doubles in four years, then what is the annual average rate of growth ?**
5. **Depreciation has been charged on a machine @ 40% in the first year, 25% in the second year, 15% in the third year and 8% in the forthcoming two years. Each percentage has been calculated on diminishing balance. Find out the average percentage rate of depreciation for the whole period.**

□□□

Lesson–17

HARMONIC MEAN

Harmonic mean is a special type of average. It is popularly used in the calculation of average speed. Harmonic mean is denoted by Hm. It is the reciprocal of the arithmetic mean of the reciprocals of all the items in a series. In order to know the harmonic mean we find out reciprocal values for all the items, get then added together and divide the aggregate (Σ Reciprocal) by number of items (N) In other words we get the arithmetic mean of reciprocals. Lastly, we find out the reciprocal of the arithmetic mean and get the harmonic. Mean.

Symbolically,

$$Hm = \text{Reciprocal of } \left(\frac{\Sigma Reciprocals}{N}\right)$$

$$\text{Or, } Hm = \left(\frac{N}{\Sigma Reciprocals}\right)$$

We know that reciprocal of any number is one divided by that number e.g. reciprocal of 5 would be 1 ÷ 5, reciprocal of 9 = and so on.

Calculation of Harmonic Mean :

(i) Individual Series–It is very simple to calculate a harmonic mean in an individual series. First of all we find out reciprocal values for each and every item in the series. Secondly we find out the sum total of the reciprocals (Σ Reciprocals) and lastly we get the harmonic mean with the help of any one of the formula mentioned above. The following example would clear the process of calculation.

Example 1. Calculate harmonic mean of the following–

X : 10, 8, 5, 2, 3

Solution :

X	Reciprocals
10	1/10
8	1/8
5	1/5
2	1/2
3	1/3

$$\Sigma \text{ Reciprocals} = \frac{1}{10}+\frac{1}{8}+\frac{1}{5}+\frac{1}{2}+\frac{1}{3}=\frac{151}{120}$$

N = 5

We know that–

$$Hm = \left(\frac{\Sigma Reciprocals}{N}\right)$$

By putting the value in the formula

$$Hm = \text{Reciprocal of } \frac{\frac{151}{120}}{5}$$

$$= \text{Reciprocal of } \frac{151}{120} \times \frac{1}{5}X$$

$$= \text{Reciprocal of } \frac{151}{600}$$

$$= \frac{\frac{1}{151}}{\frac{1}{600}} = \frac{1}{1} \times \frac{600}{151} = \frac{600}{151} = 3.97$$

Here Hm = 3.97

Example 2. A man travels from Danapur to Patna City at an average speed of 30 km. per hour and returns along the same route at an average speed of 60 km. per hour. What is the average speed for the entire trip ?

Solution :

$$\text{Hm} = \text{Reciprocal of } \left(\frac{\Sigma \text{Reciprocals}}{N}\right)$$

Here–

$$\Sigma \text{ Reciprocals} = \frac{1}{30} + \frac{1}{60}$$

$$N = 2$$

$$\text{Thus Hm} = \frac{N}{\Sigma \text{Reciprocals}} = \frac{2}{\frac{2}{30} + \frac{1}{20}} = \frac{1}{\frac{2+1}{60}}$$

$$= \frac{\frac{2}{3}}{\frac{1}{60}} = \frac{2}{1} \times \frac{60}{30} = \frac{120}{3} = 40 \text{ km. per hour.}$$

Therefore, average speed for the entire trip = 40 km. per hour.

Example 3. In a certain factor a unit of work is completed by A in 4 minutes, by B in five minutes by C in 6 minutes, by D in 10 minutes and by E in 12 minutes. What is their average rate of working ?

Solution :

As some amount of work has been completed by different workers with different speeds the average speed of work will be calculated with the help of harmonic mean.

We know–

$$\text{Hm} = \frac{N}{\Sigma \text{Reciprocals}}$$

Here N = 5

$$\Sigma \text{ Reciprocals} = \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{10} + \frac{1}{12} = \frac{48}{60}$$

By putting the values in the formula

$$Hm = \frac{5}{\frac{48}{60}} = \frac{5}{1} x \frac{5}{\frac{60}{48}} = \frac{25}{4} = 6.25$$

$$\frac{60}{48}$$

$$4$$

Thus average speed of work = 6.25 minutes per unit.

Example 4. An aeroplane files around a square whose side is 1000 Kilometres long. The plane covers all the four sides of the square with an average speed of 1000, 2000, 3000 and 4000 Kilometres per hour respectively, what is the average speed of the aeroplane ?

Solution—

Here N = 4

$$\Sigma \text{ Reciprocals} = \frac{1}{1000}+\frac{1}{2000}+\frac{1}{3000}+\frac{1}{4000}=\frac{25}{12000}$$

$$Hm = \frac{N}{\Sigma Reciprocals}$$

$$= \frac{4}{\frac{25}{12000}} \frac{12000}{25} = 1920$$

Average speed = 1920 Kilometres per hour.

The method through which we have obtained the reciprocal values in the above mentioned examples is very simple but it becomes difficult as the size of items in a series increases. If the items given are in decimal it becomes more difficult to obtain their reciprocal values through above method. It would be more time-consuming and there will be greater chance of mistake in the calculation also. Thus with a view to overcome these problems we use reciprocal tables for obtaining reciprocal value of any number. Here we shall first discuss the procedure of reading reciprocal table and thereafter explain their applications with the help of different examples.

A reciprocal table has three component parts first part consists of only one calumn. This column ruse from 1.0 to 9.9. Second part of the table consists of 10 columns labelled 0 to 9 separately. The third or the last past of the table is called mean difference part. It consists of 9 columns each for one digit raning 1 to 9.

In order to obtain the reciprocal value of a number consisting of more than 4 digits, we get a number of four digits by using approximation techniques. Then the first two digits from the left side of the given number are located in the first column of the reciprocal table. In order to get the number in this column we change the place of decimal in the given number accordingly. For the third digit of the given number column we locate the place of decimal in the respective column in the second part of the table and read the value written in this column against the first two digits, located in the first column. If there is fourth digit also in the given number we read the respective p value written in the mean difference column and substract it from the value obtained from the second part of the table. Decimal point is placed in the number so obtained in the same way as it has been done in the beginning while locating first two digits in the first column of the reciprocal table. Let us take some examples.

Suppose, we have to obtain the reciprocal value of 5 from the reciprocal table. For this we shall search out 5 or 50 in the first column of the reciprocal table. Here in this case any shifting in the place of decimal point in the given number is not required as it is available in its original form in the first column, Now the number written aginst 5 in the zero column of the second part of the table will read which is 2000. this will be the reciprocal value.

Let us take another example.

Suppose the number given is 55. As this number (55) is not available in its original form in the first column of the table decimal point will be moved one digit left. This lefward movement of the decimal point will convert the original number into 5.5 Now it can be easily located in the first column. After locating it in the first column we read the value written in the zero column in front of it. Which is 1818. Now the decimal point will be shifted one digit left because we have already shifted the decimal point one digit left to get 55. Thus reciprocal value of 55 would be .01818.

If the original number is 1548 and we are to obtain reciprocal value of it. In this case the decimal point will more three digit left. This will convert 5648 into 5.648. Then after locating 5.6 in the first column the number against it in the 4th column will be noted down which is 1773. The fourth digit in the given number is 8. For this we shall read the number given in the 8th column of the mean difference column aginst 5.6 which is 3. This 3 will be substracted from 1773 which comes to 1770. Now the decimal point will move three places left because in the beginning it was moved three digit left and converted 5648 into 5.638. Thus reciprocal value of 5648 will be .0001770 or .0002, after approximation.

Example 5. Calculate harmonic mean of the following—

85, 70, 10, 75, 500, 8, 42, 250, 40 and 36

Solution :

X	Reciprocals
85	.01176
70	.01429
10	.10000
75	.01333
500	.00200
8	.12500
42	.02381
250	.00400
40	.02500
36	.02778
	0.3497

$$Hm = \frac{N}{\Sigma \text{Reciprocals}}$$

$$= \frac{10}{0.34697}$$

= 28.82 Ans:

Example 6. Compute the harmonic means from the following data—

800, 78.3 9.45, .676, .07834, 004656 0003425 and .0002853

Solution—

X	Reciprocals
800	.001250
78.3	.012770
9.45	.05800
.676	1.479000
.07834	12.47940
.004656	21.4800009
.0003425	2920.00000
.00002853	35050.000000
N = 8	Σ Recp = 38199.158820

$$Hm = \text{Recip. of } \frac{\Sigma Reciprocals}{N}$$

$$= \text{Recip of. } \frac{38199.158820}{8}$$

= Recip of .4774.8948525

= .0002094

Hm = 0.0002094 **Ans.**

(2) Discrete Series–The following steps are takn while calculation harmoinc mean of discrate series–

(i) Firstly, recipracal value of all the items are found out.

(ii) Secondly, reciprocal values are multiplied with their corresponding frequencies and products are added together. This gives Σ Recip xf.

(iii) Lastly, with the use of the following formula we get the harmonic mean—

$$Hm = \text{Reciprocal of } \left(\frac{\Sigma(Reciprocal)Xf}{N}\right)$$

Example 6. Calculate harmonic mean of the following distribution—

X	:	2	3	4	5	6
f	:	4	9	16	10	24

Solution :

X	f	Reciprocals (x)	Reciprocal Xf.
2	4	1/2	1/2 × 4 = 2
3	9	1/3	1/3 × 9 = 3
4	16	1/4	1/4 × 16 = 4
5	10	1/5	1/5 × 10 = 2
24	24	1/6	1/6 × 24 = 4
	N = 63		Σ Recip. f = 15

$$Hm = \frac{N}{\Sigma Recipf} = \frac{63}{15} = 4.2$$

Hm = 4.2

Above example can be solved with the help of reciprocal table in the following way—

X	f	Rciprocal (x)	(Recip x) f.
2	4	0.5000	2.0000
3	9	0.3333	2.9997
4	16	0.2000	4.0000
5	10	0.2000	2.0000
6	24	0.1667	4.0008
	63		15.0005

$$Hm = \text{Recip. of } (\frac{\Sigma Recip.f}{N})$$

$$= \text{Recip. of } \frac{15.0005}{63}$$

= Recip. of 0.2381031

= 4.202 or 4.2

(3) Continuous Series–The following steps are taken in calculating harmonic mean in a continuous series.

(i) Mid values for each are every class interval is found out.

(ii) Reciprocal value for each and very mid value is found out.

(iii) Sum total of the products of reciprocals with their respective frequency is found out (Σ Recip. x.f.)

(iv) With the help of the following formula harmonic mean is calculated–

$$Hm = \text{Reciprocal of } (\frac{\Sigma(Reciprocalx)f}{N})$$

Example 7. Calculate harmonic mean from the following.

Class :	0-8	8-16	16-24	24-32	32-40	40-48
Frequency :	12	16	13	10	8	11

Solution :

Class	**f**	**M.V.**	**Reciprocal of N.V. (x)**	**(Reciptocal x) f**
0-1	12	4	0.2500	3.0000
8-16	16	12	0.0833	1.3328
16-24	13	20	0.0500	0.6500
24-32	10	28	0.0357	0.3570
32-40	8	36	0.0278	0.2224
40-48	11	44	0.0227	0.2497
	70			5.8119

$$Hm = \text{Reciprocal of } \left(\frac{\Sigma(\text{Recip.}x)f}{N}\right)$$

$$= \text{Recip. of } \frac{5.8119}{70}$$

= Recip. of 0.0830271 = 12.05 Ans.

Weighted Harmonic Mean.

Weighted harmonic mean is denoted by WHm. In its calculation reciprocals of items are multiplied with their respective frequencies and the sum total of their products, (Σ Recip.w) is divided by the total of weight (Σ w). The reciprocal of the quotient so obtained is called harmonic mean.

Symbolically—

$$WHm = \text{Reciprocal of } \left(\frac{\Sigma(\text{Recip.}x)w}{\Sigma w}\right)$$

Example : Calculate harmonic mean—

Indives :	100	110	105	120	95	80	130
Weights :	5	6	8	9	2	4	6

Solution :

Indices	**Weights**	**Reciprocals**	**Reciprocals X weight**
100	5	.010000	0.050000
110.	6	.089091	0.054546
105	8	.009524	0.075612
120	9	.008333	0.074997
95	2	.010530	0.021060
80	4	0.12500	0.050000
130	6	.007692	0.046152
	40		4272947

$$WHm = \text{Reciprocal of } \frac{\Sigma \text{Recip.}Xw}{\Sigma w}$$

$$= \text{Reciprocal of } \frac{0.372947}{40}$$

= Reciprocal of 0.0093236 = 107.3

WHm = 107.3 **Ans.**

Example 9. A man travels 60 kms with train which runs at the speed of 40 kms per hour, 60 Kms with motor car which runs at the speed of 30 Kms per hour, 20 Kums with cycles at the speed of 8 Kms per hour and 10 Kms on foot at the speed of 5 Kms per hour. Calculate the average speed per hour of the whole journey.

Solution :

As different distances have been covered with varying speeds, the distance covered in each case will be treated as weight. A weighted average would be most appropriate in this case.

Speed in Kms per hour (x)	Distance Covered in Kms (w)	Reciprocal (X)	(Reciprocal x) W
40	60	0.0250	1.5000
30	60	0.0333	1.9980
8	20	0.1250	2.5000
5	10	0.2000	2.0000
	150		7.9880

$$WHm = \text{Recip. of } \left(\frac{\Sigma \text{Recip.} Xw}{\Sigma w}\right)$$

$$= \text{Recip. of } \left(\frac{7.9980}{150}\right)$$

= Recip. of 0.5332

= 18.75 Kms per hour.

Merits of Harmonic Mean :

(1) It is based on all the items of a series.

(2) It assigns greater importance to smaller items of series in comparison to larger items.

(3) It is capable of algebrical analysis.

(4) Average related with motion, speed, velocity in physical sciences are calculated with harmonic mean. It is also used in getting average in problems related with time and work, work, time and distance etc.

Demerits of Harmonic Mean :

(1) Its calculation is difficult.

(2) Knowlege of every item of a series is essential for its calculation.

(3) It has a limited scope.

Empirical Relationship between Averages.

We have discussed five kinds of averages namely mode, Median, Arthmetic Mean, Gometic Mean and Harmonic Mean. They have their own special features. Although they differ from each other, a relationship between them can be established keeping in view the composition of the frequency distribution. Statistical series are divided into three categories according to their construction. They are

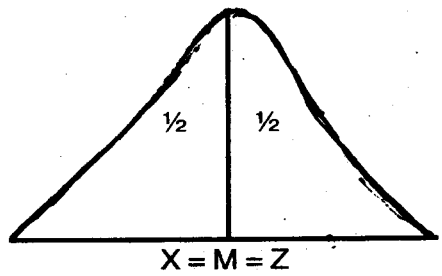
(i) Perfectly symmetrical,

(ii) Moderately Asymmetical and

(iii) Highly symmetrical.

(i) Perfectly Symmetrical–In such distributions a similarity is found in rise and fall of the frequecy. in the beginning frequency increases and after reaching the maximum point it decreases but with a speed similar to its earlier movement. There are equal number of units before and after the maximum points. If plotted n the graph paper, such distribution gives a symmetrical bell-shaped curve. If we draw a perpendicular on the base line from the point of mean, median and mode coincide. The following series form a suitable example of symmetical distribution.

Class :	10-20	20-30	30-40	40-50	50-60
Frequency :	5	10	15	10	5



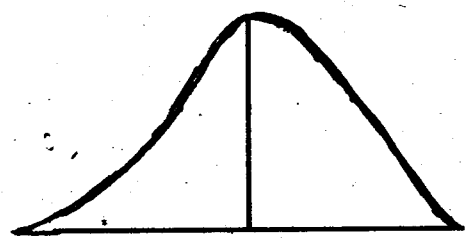
(ii) Moderately Asymmetical Distribution–Unlike symmetrical distribution, number of items lying left and right side of the miximum point are unequal in asn asymmetical distribution. They do not present bell-shaped curve. They may be skewed to the right or skewed to the left. Such skewness pushes the mean, median and mode apart. Such distributions can be shown below—

Class :	10-20	20-30	30-40	40-50	50-60	60-70
f :	5	12	18	25	10	4

A curve drawn as the basis of above data will be as follows—

Class :	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f :	10	20	60	50	30	20	1

A curve drawn on the basis of above data will take the following shape.



The difference between the mean and the median is taken as a measure of skewness (asymmetry). The larger is this difference, the greater is skewness. If the difference (Mena–mode is positive the distribution is said to be positvely skewed and in case it is negative. the distribution is negative skewed.

(iii) Higly asymmetrical Distribution–In such distribution frequency increases and decreases in untidy manner. The following distribution shows the nature and charactersitics of a perfectly asymmetrical distribution–

Calss :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f :	165	396	204	170	98	61	25	3

A curve drawn on the basis of above data may take the following shape–

Relationship between Mean, Median and Mode

(1) In perfectly symmetrical distribution, arithmetic mean, median and mode are identical.

Symbolically–

$\overline{X} = M = Z$

(2) In a moderated asymmetrical distribution difference between arithmetic mean, median and mode is equal to three times of the difference between arithmetic mean and median.

Symbolically–

$$X - Z = (\overline{X} - M)$$

or $$Z = 3M - 2\overline{X}$$

With the use of above formula we can calculate the value of the third average if values of any two of them are given.

Relationship between Arithmetic mean, Geometric mean and Harmonic mean.

(i) If all the items in a distribution are the same the arithmetic mean, the geometric mean and the harmonic mean are equal.

Symbolically—

$\overline{X} = Gm = Hm$

This will be clear form the following illustration–

Example 10. Calculate Arithmetic, Geometric and Harmonic mean of the following–

X : 2, 2, 2, 2, 2

Solution :

X	LogX	Reciprocal X
2	0.3010	0.5000
2	0.3010	0.5000
2	0.3010	0.5000
2	0.3010	0.5000
2	0.3010	0.5000
	1.5050	2.5000

Arithmetic Mean :

$$\overline{X} = \frac{\Sigma X}{N} = \frac{10}{5} = 2$$

Geometiric Mean :

$$Gm = \text{Antilog of } \frac{\Sigma \log X}{N}$$

$$= \text{Antilog of } \frac{1.5050}{5}$$

= Antilog of 0.301 = 2.000

Harmonic Mean :

$$Hm = \frac{N}{\Sigma \text{Reciprocals}} = \frac{5}{2.5000} = 2$$

$\bar{X}$ = Gm = Hm.

(2) In a series of unequal items, as generally found, the arithmetic mean will be greater than the Geometiric mean and Geometric mean will be greater than the Harmonic mean.

Symbolically : X > Gm > Hm.

This can be explained with the help of the following example–

Example 11. Calculate Arithmetic, Geometric and Harmonic mean of the following–

X	:	2	4	6	8	10
f	:	5	10	15	10	5

Solutions :

X	f	X.f	log x	log x.f	Reciprocal (x)	Recip (X.f.)
2	5	10	0.3010	1.5050	0.5000	2.5000
4	10	40	0.6021	6.0210	0.2500	2.5000
6	15	90	0.7712	11.6730	0.1667	2.5005
8	10	80	0.9031	9.0310	0.1250	1.2500
10	5	50	1.0000	5.0000	0.1000	0.5000
	45	270		33.2300		9.2505

Geometric Mean :

$$Gm = \text{Antilog of } \left(\frac{\Sigma \log . f .}{N}\right)$$

$$= \text{Antilog of } \frac{33.2300}{45}$$

= Antilog of 0.738444 = 5.48

Arithmetic Mean :

$$\bar{X} = \frac{\Sigma X . f .}{N} = \frac{270}{45} = 6$$

$\bar{X}$ = 6, Gm = 5.48 and Hm = 4.86

6 > 5.48 > 4.86

Thus $\bar{X}$ > Gm > Hm

Harmonic Mean :

$$Hm = \frac{N}{\Sigma Recip.f.} = \frac{45}{9.2504} = 4.86$$

(3) The geometric mean of series of two values is equal to the geometric mean of their arithmetic mean and Harmonic mean; Symbolically.

$Gm = \sqrt{X \times Hm}$ or $Gm^2 = X \times Hm$.

Example 12. If a series of two values has a mean equal to 4.67 and harmonic mean equal to 3.43 then fid out the value of its geometric mean.

Solution :

$Gm^2 = \overline{X} \cdot Hm$

$Gm = \sqrt{X.Hm} = \sqrt{4.67 \times 3.43} = \sqrt{16.0181} = 4.00$

$Gm = 4.$

Section of a suitable average :

After discussing different statistical averages naturally the question arises as two which one is the best among them ? What type of average we should use ? There is no one answer of these questions which is acceptable to all and in all circumstances. The different statistical averages have different characteristics. There is no all purpose average. The choice of a particular average is usually determined by the prupose of investigation, conditions of investigation and the nature of distribution etc. An average fit for one purpose or distribution may be unfit for another. Generally the following points should be ketp in mind while making a choice of an average for a particular investigation–

(1) Objective of Study–Choice of an average is highly influenced by the object of study. If all the values in a series are to be given equal importance, then arithmetic mean will be a suitable choice. To determine the most popular item, mode should be found out. If the object is to determine an average that would indicate its position or ranking in relation to all the values median should be the best choice. Phenomenon changing with geometric rate or cumulative ways are best represented by geometric mean. Harmonic mean is best suited to the problems related with average spped, work and time etc.

(2) Nature of data–Arithmetic mean is more suitable than medians or mode in a distribution of few values. If the difference between different values of a series is very large, geometric mean will bet he best average. In an opened class intervals mean cannot be calculated accurately. In a frequencty distribution of unequal class intervals mode cannot be determined accurately. Mean will not give a representative picture, it there are a few extermely large or small values at either end of the series.

(3) Easy Calculation–The average selected should be easy to calculate-Geometric mean and Harmonic mean are difficult as they require knowledge of logatithm and Reciprocal. Thus individual merits and demerits are also taken into considerations while making a choice of a particular average.

MODEL QUESTIONS

1. **Explain the meaning of Harmonic mean and give its merits and demertis.**
2. **A driver drives his car at a speed of 80 Km. per hours and covers a distance of**

50 Kilometers, at a speed of 60 Km. Per hour he covers the next 50 Kilometers and the last 50 Kilometers of his journey he covers with a speed of 40 Km. per hour. What is his average speed per hour ?

3. Ram's car runs 40 Kilometers in one gallon patrol and Mohan's car runs 60 Kilometers in one gallon petrol. If both the cars cover a distance of 120 Kilometers individually what is the average speed per gallon for the entire journey ?

4. A can do a piece of work in 5 days, B in 10 days and C in 16 days. Find out the average required to finish the work if all the three persons work for equal number of days ?

5. A taxi ride in Calcutta costs two rupees for first Kilometer and 60 paise for each additional kilometer. The cost of each Kilometer is incurred at the beginning of the km. So that the rider pays for the whole Km. What is the average cost for 2.3/4 Km ?

6. Compute Harmonic mean from the following :

Class :	0-10	10-20	20-30	30-40	40-50
Frequency :	8	10	6	5	3

7. (a) If $\overline{X} - M = 5$, and $Z = 50$, find out $\overline{X}$ and M.

(b) The mean of 40 items is 17. If the mean of 10 items among them is 20. Find the mean of the rest.

8. In a moderately asymmetrical series mean is 15.6 and median is 15.73, find the probable value of mode.

9. The arthmetic and geometric mean of the two characters are 15 and 12 respectively. Find out the two characters and the harmonic mean.

10. You take a trip which entails travelling 900 Kms. by train at an average speed of 60 Kms. per hour, 3000 Kms, by boat at an average speed of 25 Kms. per hours. 400 Kms. by plane at 350 Kms. per hour and finally 15 Kms. by a taxi at 20 Km. per hour, What is the average speed for the entire distance ?

11. Calculate Harmonic mean of the following :

X : 20, 125.61680.82 0.7254 0.073

12. A taxi driver from a plain town to a hill town station, covers 60 Kms. at a 10 Km. Per gollan of petrol and on the return trip at 15 Km. per gallon. Find the harmonic rate of Km. per gallon.

13. A men travels from Patna to Delhi at an average speed of 50 Km. per hour and returns along the same route at an average speed of 75 Km. per hour. Find the average speed for the entire trip.

14. Find out the Harmonic mean.

1578 15.81.58 0.16 0.0160.0075 0.0009

❐❐❐

DISPERSION – I

We have studied different measures of central tendency in the previous lesson. Measures of central tendency give us a single representative number for a huge mass of numbers which helps in estimation of different values of the distribution. But we find that a measure of central tendency alone fails to reveal all characteristic of a statistical series. It also falls in throwing light on the composition of the statistical series. This can be illustrated with the help of the following example :

X	Y	Z
10	6	20
10	8	12
10	10	8
10	12	6
10	14	4
50	50	50

$$\overline{X} = \frac{\Sigma X}{N} = \frac{50}{5} = 10$$

$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{50}{5} = 10$$

$$\overline{Z} = \frac{\Sigma Z}{N} = \frac{50}{5} = 10$$

The above example sets before us that each series has similar value as their mean i.e. $\overline{X} = \overline{Y} = \overline{Z}$. But although they have identical values as measure of central tendency nevertheless similar inference can not be drawn for all the three distributions. Each and every item of the series X is being cent-percent represented by its average : The inference drawn on the basis of average about different items of series X is highly accurate but it is not so for series Y and Z. There is stability among the items of series X but items in the series Y are increasing whereas items in the series Z are decreasing. Thus it is evident that mean happens to be the representative value of the series but it is not competent enough to divulge all details about the distribution. We have seen that series having identical mean may have different composition. Therefore, it would not be proper to draw inference on the basis of mean alonge. Apart from mean some other measures are also essential to get complete knowledge about a phenomenon or a statisatical series. Dispersion is one of the such measures.

Meaning of Dispersion–Dispersion is related with the homogeneity of data and measures of dispersion are measures of scatter or spread about an average. In the words of Dr. A. Bowley "Dispersion is the measure of the variation of the item." Generally the word dispersion is used in two senses. In one sense it means the range of scatter of data in a series. In other words dispersion

expresses the limt within which different values of a series are scattired. In this sense we can say that difference between the largest and smallest values of a series is the dispersion expresses the limt within which different values of a series are scattired. In this sense we can say that difference between the largest and smallest values of a series is the dispersion.

In its second sense dispersion means the arithemetic mean of the deviations taken from an average of the different items of a series. As it is based on the average of the deviations of the different values from their mean, it is called the Average Deviation. According to John I. Griffin, "A measure of variation or dispersion describes the degree of scatter shown by observations and is usually measured as an average deviation about some central value." In simple words we can say that dispersion shows us the average fluctuation of individual items of a series from any measure of central tendency of the series. In the words of Spriengel "The degree to which numberical data tend to spread about an average value is called the variation or dispersion of the data."

Objects of Dispersion :

Measures of dispersion or variation are needed for the following purpose :

(1) To gauge the reliability of an average–Dispersion reveals the representative character of a measure of central tendency. It shows the degree to which extent one can depend upon average value of a series. A measure of dispersion along with a measure of central tendency gives a more accurate knowledge of the construction of the distribution. A knowledge of dispersion increases the degree of reliability of a measure of cental tendency. William A Spurr and Charles P. Bonini observe "when dispersion is small, the average is a typical value in that it closely represents the individual values, and its is reliable in that it is a good estmate of the corresponding average in the population. On the other hand, when the dispersion is great, the average is not so typical and unless the sample is very large, the average may be quite unreliable."

(2) Knowledge of scatteredness–Dispersion shows the average distance at which the items of a series lie from their representative value, kimite to which they are scattered on both sides of their average value.

(3) Comparative study–On the basis of measures od dispersion, the extent of variability between two or more series can be compered. Dispersion is useful to find out degree of uniformity or consistency in the two or more sets of data. A greater amount of dispersion means lack of uniformty or consistency in the data.

Method of Measuring Dispersion.

Based on the different meanings of dispersion, following methods are used to measure dispersion :

(1) Method of Limits–(A) Range, (B) Inter Quaritile Range and (C) Deciles Percentiles range.

(2) Method of Averaging Deviation–(A) Quartile Deviation, (B) Average Deviation and (C) Standard Deviation.

(3) Graphical Methods–Lorenze curve.

Absolute and Relative Dispersion

Dispersion can be measured both in absolute as well as relative terms. An absolute measurement of dispersion presents an absolute picture of the scatter of the items of the series. It provides the idea about the size of the series. It is experssed in the unit of series.

Relative measures of dispersion facilitate comparative study of variation between two or more series. Absolute measures are not fit for comperative study. Relative measures of dispersion are not expressed in the units of the series. They are expressed as ratios and percentage. Such measures are computed by dividing the absolute measure by the average which has been used in the calculation of absolute measure. Such measure are known as coefficients.

Range :

Range is the simplest measure of dispersion. it refers to the difference between largest and smallest values in distribution. Greater amount of range denotes higher degree of dispersion or inconsistency among different values in the distribution and vice-versa. Range in a distribution is measured with the help of following formula :

Range = Largest value – Smallest value.

Or, R = L – S

In a discreate or continuous series we do not take into account the frequency while calculating range. We consider only largest and smallest values of the variable.

Co-efficient of Range

Co-efficient of range is the relative measure of range. It is calculated with the khelp of the following formula :

$$\text{C of R} = \frac{L-S}{L+S}$$

Here, C or R = Co-efficient of range, L = largest value and S = smallest value.

Example 1.

Find out the range and its co-efficient from the following data :

140, 370, 80, 73, 185, 30, 90, 80, 203 and 150.

Solution :

Largest value = 370, Smallest value = 30

Rang (R) = L – S

=370 – 30 = 340.

$$\text{Co-efficient of Range (C of R)} = \frac{L-S}{L+S}$$

$$= \frac{370-30}{370+30} = \frac{340}{400} = 0.85$$

Range = 340 and C of R = 0.85.

Example 2.

Calculate range and co-efficient of range from the following data :

x	:	5	10	15	20	25	30
f	:	2	3	5	10	4	2

Solution :

Range (R) $= L - S$

$= 30 - 5 \quad = 25$

Co-efficient of Range (C of R) $= \dfrac{L-S}{L+S} = \dfrac{30-5}{30+5} = \dfrac{25}{35} \quad = 0.71$

Example 3.

Calculate the range and its co-efficient from the following distribution :

Marks	:	10 - 20,	20 - 30,	30 - 40,	40 - 50,	50 - 60
f	:	5	10	20	10	5

Solution :

Here Largest value = 60

Smallest vallue = 10

Thus Range (R) = $L - S$

= $60 - 10 = 50$ Marks

Co- efficient of Range (C of R) $= \dfrac{L - S}{L + S} = \dfrac{60 - 10}{60 + 10} = \dfrac{50}{70} \quad = 0.70$

Merits of Range :

(1) It is very simple to calculate and easy to understand.

(2) It is used in quality central measures. The control charts are prepared on the basis of range for controlling the quality of products. Products lying within a definite range are accepted otherwise they are rejected.

(3) Range is a very useful measure of fluctuation. Variation in rate of interest, rate of exchange, prices of securities, tecmperature etc; are studies with the help of range.

Demerits of Range :

(i) Range is an uncertain measure of dispersion. It is based on the extreme value of the distribution. It is attracted by a slight change in the extreme values of the distribution.

(ii) It does not utilise all the information in the set of observation. It does not reflect the distribution of the observation between the extremes.

(iii) Range does not geve any idea about composition of the series. Two series may have the same range but of varied composition and shape.

(iv) It is not satisfactorily used in frequency series.

Inter Quartile Range

This measure of range is based on quartile values of the distribution. It shows the simple difference between third quartile and first quartile.

Symbolically :

I. Q. R. = $Q_3 - Q_1$

Here, I. Q. R. = Inter Quartile Range, Q_3 = Third quartile and Q_1 = First quartile.

Similarly :

Co - efficient of I. Q. R. = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

The following example can explain the process of calculation of inter quartile range and its co-efficent.

Example 4.

Calculate inter quartile range and its co-efficient from the following data :

X : 66, 58, 45, 65, 48, 80, 72, 55, 78, 76 and 79.

Solution :

X (arranged in ascending order)

45

48

55

58

65

66

72

76

78

79

80

Here : N = 11

Q_1 = Value of $\frac{N + 1}{4}$ th unit.

= Value of $\frac{11 + 1}{4}$ th unit.

= Value of 3 rd unit.

= 55

Q_2 = Value of th unit.

= Value of $\frac{3(11 + 1)}{4}$ th unit.

= Value of $\frac{3}{4} \times 12$ th unit.

= Value of 9th unit. = 78

Inter Quartile Range (I. Q. R.) = $Q_3 - Q_1$

= 78 – 55 = 23

Co-efficient of I.Q.R = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{78 - 55}{78 + 55} = \frac{23}{133}$ = 0.173

Example 5.

Find out the inter quartile range and its co-efficient from the following data :

X :	1	2	3	4	5	6	7	8	9	10
f :	2	3	3	9	12	10	6	2	2	1

Solution :

X	f	c.f.
1	2	2
2	3	5
3	3	8
4	9	17
5	12	29
6	10	39
7	6	45
8	2	47
9	2	49
10	1	50

$= \text{Value of } \frac{N + 1}{4} \text{ th unit.}$

$= \text{Value of } \frac{50 + 1}{4} \text{th unit.}$

= Value of 12.75th unit. This is included in cumulative frequency 17. Therefore, the variable against it i.e. 4 is the first quartile.

$Q_1 = 4$

$Q_4 = \text{Value of } \frac{3(N + 1)}{4} \text{ th unit}$

$= \text{Value of } \frac{3}{4} \times 51 \text{ th unit}$

= Value of 38.25 th unit.

This is included in cumulative frequency 39. Therefore, the value against it i.e. 6 is the fourth quartile. $Q_3 = 6$

Inter Quartile Range = $Q_3 - Q_1 = 6 - 4 = 2$

Co- efficient of I. Q.R = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{6 - 4}{6 + 4} = \frac{2}{10} = 0.2$

Example 6.

Find I. Q. R. and its co-efficient from the following :

Class	:	0 - 10,	10 - 20,	20 - 30,	30 - 40,	40 - 50
f	:	5	10	20	10	5

Solution :

Class	f	c.f.
0 - 10	5	5
10 - 20	10	15
20 - 30	20	35
30 - 40	10	45
40 - 50	5	50

Calculation of Q_3

Size of $q_3 = \frac{5(N)}{4}$ th unit.

$= \frac{3}{4} \times 50$ th unit = 37.5 th unit This size of q_3 is included in cumulative frequency 45. Thus corresponding class i.e. 30 –40 is the Q_3 class.

Now,

$$Q_3 = l_1 + \frac{l_2 + l_1}{f}(q_2 - c)$$

Here, l_1 = lower limit of the Q3 class = 30

l_2 = upper limit of the Q3 class = 40

f = Frequency of the Q3 class = 10

q_3 = size of third quartile = 37.5

C = Cumulative frequency of the class preceding Q_3 class = 35

By putting the values in the above formula

$$Q_3 = 30 + \frac{40 - 30}{10}(37.5 - 35)$$

$$= 30 + \frac{10}{10} \times 2.5$$

$$= 30 + 2.5 = 32.5$$

Calculation of Q1

Size of $q_1 = \frac{N}{4}$ th unit.

$= \frac{50}{40}$ th unit = 12.5th unit.

This size of q1, i.e. 12.5 lies in c.f. 15. Thus the 10–20 corresponding to this size is the Q1 class.

Now,

$$Q_1 = l_1 \frac{l_2 + l_1}{f}(q_1 - C)$$

Here l_1 = 10, l2 = 20, f = 10, q_1 = 12.5 and C = 5

Thus,

$$Q_1 = 10 + \frac{20 - 10}{10}(1.25 - 5)$$

$$= 10 + \frac{10}{10} \times 7.5$$

$$10 + \frac{75}{10} = 10 + 7.5 = 17.5$$

$\therefore$ $Q_1 = 17.5$

Inter Quartile Range (I. Q. R.) $= Q_3 - Q_1$

$= 32.5 - 17.5 = 15$

Co-efficient of I. Q. R. $= \frac{Q_3 + Q_1}{Q_3 + Q_1} = \frac{32.5 - 17.5}{32.5 + 17.5} = \frac{15}{50} = .03$

Hence.

I. Q. R. = 15 and Co-eff. I. Q. R. = 0.3

Merits of Inter-Quartile Range

(1) It is superior to range as a measure of dispersion.

(2) It is not based on the extreme items of the distribution. Unlike range, it considers middle fifty percent items in a distribution.

(3) It is simple and easy to calculate and less affected by items. It is helpful in measuring dispersion in open end distributions.

Demerits of Inter-Quartile Range

(1) Inter-quartile range is not capable of mathematical manipulation.

(2) Its value is affected by sampling fluctuations.

(3) Like range, it is not based on all the items of the distribution. It does not throw light on the construction of the series. Apart from these, it ignores 50% items i.e the first 25% and the last 25%.

Decile Range and Percentile Range

These measures of disperson are similar to range and inter quartile range. They are based respectively on decile and percentile values of the distribution. Decile range refers to the difference between 9th and 1st deciles. Whereas the percentile range is the difference between the 99th percentile and the first percentile.

Symbolically, Decile Range (D. R.) $= D_2 - D1$

Co-efficient of Decile Range $= \frac{D_2 - D_1}{D_2 + D_1}$

Similarly, the percentile range (PR) and the co-efficent of (PR) are given by

$PR = P_{99} - P_1$

and co-efficient of PR $= \frac{P_{99} - P_1}{P_{99} + P_1}$

It may be noted that $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{P_{75} - P_{25}}{P_{75} + P_{26}}$

and $\frac{D_9 - D_1}{D_5 + D_1} = \frac{P_{90} - P_{10}}{P_{90} + P_{10}}$

Similary, $Q_2 = D_5 = P_{50}$

These measures of dispersion possess all merits and demerits of range and inter quartile range we have discussed over pages. Apart from they there are considered superior to range and inter-quartile range because these ignore the existence of maximum and minimum items.

Quartile Deviation

This measure of dispersion is also on third and first quartiles. It is also known as semi inter quartile range. It refers to the half of the difference between third and first quartiles.

Symbolically :

Quartile Deviation (Q. D.)

Co- efficient of Quartile Deviation =

The following example can explain its calculation procedures.

Example 7.

Calculate thequartile deviation and its co-efficient from the following :

Marks : 15, 20, 20, 21, 22, 22, 24, 25, 28, 28, 29, 30, 32, 33, 35.

Solution :

S. No.	Marks	
1	15	$Q1$= Value of $\frac{N+1}{4}$ th unit.
2	20	
3	20	= Value of $\frac{15+1}{4}$ th unit.
4	21	
5	22	= Value of the 4th unit.
6	22	= 21 Marks.
7	24	
8	24	$Q3$ = Value of $\frac{3(15+1)}{4}$ th unit.
9	28	
10	28	= Value of $\frac{3(15+1)}{4}$ th unit
11	29	= Value of 12th unit.
12	30	= 30 Marks.
13	32	
14	33	
15	35	

Quartile Deviation (Q. D.) $= \frac{Q_3 - Q_1}{2}$

$$= \frac{30 - 21}{2} = \frac{9}{2} = 4.5 \text{marks.}$$

$$\text{Co-efficient of Q.D.} = \frac{Q_9 - Q_1}{Q_5 + Q_1} = \frac{30 - 21}{30 + 21} = \frac{9}{51} = 0.176$$

Q. D. = 4.5 Mark

Co-eff. of Q. D. = 0.176.

Example 8.

Find the quartile deviation and its co-efficeint from the following data :

No. of Children :	0	1	2	3	4	5	6
No.of family :	4	6	10	15	12	9	4

Solution :

No. of Children (X)	No. of Family (f.)	Cumulative Frequency (c.f.)
0	4	4
1	6	10
2	10	20
3	15	35
4	12	47.
5	9	56
6	4	60

Q_1 = Value of $\frac{N + 1}{4}$ th unit

= Value of $\frac{60 + 1}{4}$ th unit.

= Value of 15.25 th unit. Which is included in cumulative frequency 20. Thus Q1 = 2 children

Q_3 = Value of $\frac{3(N + 1)}{4}$ th unit

= Value of $\frac{3(60 + 1)}{4}$ th unit.

=Value of 45.75th unit. Which is included in the cumulative frequency 47. Thus Q_3 = 4 children.

$$\text{Thus Q. D} = \frac{Q_3 - Q_1}{2} = \frac{4 - 2}{2} = \frac{2}{2} = 1 \text{ Child.}$$

$$\text{Co- efficient of Q. D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{4 - 2}{4 + 2} = \frac{2}{6} = .33$$

Example 9.

Calculate Q. D. and its co-efficient from the following distribution.

Class :	10 - 20,	20 - 30,	30 - 40,	40 - 50,	50 - 60
f :	5	10	15	10	5

Solution :

Class	f	c.f
10 - 20	5	5
20 - 30	10	15
30 - 40	15	30
40 - 50	10	40
50 - 60	5	45

First Quartile

Size of q. $= \frac{N}{4}$ th unit.

$= \frac{45}{4}$ th unit.

= 11.25th unit. Which lies in c.f. 15. Thus 20 - 30 is the Q class.

$$Q_1 = L_1 + \frac{l_1 + l_1}{f} (q_1 - c)$$

(Here, $l_1 = 20$, $l_2 = 30$, $f = 10$, $q_1 = 11.25$ and $c = 5$)

$$Q_1 = 20 + \frac{30 - 20}{10} (11.25 - 5)$$

$$= 20 + \frac{10}{10} \times 6.25$$

$$= 20 + \frac{62.5}{10} = 20 + 6.25 = 26.25$$

Third Quartile

Size of $q_3 = \frac{3(N)}{4}$ th unit

$= \frac{3}{4} \times 45$ th unit.

= 33.75th unit, which lies in the cumulative frequency 40. Therefore, 40 - 50 is the Q3 class.

$$Q_3 = l_2 + \frac{l_2 - l_1}{f} (q_3 - c)$$

Here, $l_1 = 40$, $l_2 = 50$, $f = 10$, $q_3 = 33.75$ and $c = 30$

$$Q_3 = 40 + \frac{50 - 40}{10} (33.75 - 30)$$

$$= 40 + \frac{10}{10} \times 3.75$$

$$= 40 + \frac{37.5}{10} = 40 + 3.75 = 43.75$$

Quartile Deviation (Q. D) $= \frac{Q_3 - Q_1}{2} = \frac{43.75 - 26.25}{2} = \frac{17.50}{2} = 8.75$

Co- efficient of Q. D. $= \frac{Q_3 - Q_1}{Q_4 - Q_1} = \frac{43.75 - 26.25}{43.75 + 26.25} = \frac{17.50}{70} = 0.25$

Merits of Quartille Deviation

(1) Its calculation is very simple.

(2) Extreme value are less effective than in range.

(3) It is fit for the phenomena where middle portion in the distribution is to be studied.

Demerits of Quartile Deviation

(1) It is not based on all the items of the distribution.

(2) It is an unstable measure highly affected by fluctuation in sample.

(3) It is not capable of algebrical analysis. It is not fit for the cases where extreme items are more significant.

Model Questions

1. Define dispersion and give an account of different measures of dispersion.

2. Find out range and its co-efficient of the following :

145, 367, 268, 73, 70, 185, 670, 280, 870, 340.

3. Calculate range, inter Quartile Range and their co-efficients from the following :

x :	1	2	3	4	5	6	7	8	9
f :	6	8	6	9	8	7	5	6	3

4. Find out Percentile Range, Quartile Deviation and Co-efficient of quartile deviation from the following :

Class	:	10 - 12,	12 - 14	14 - 16,	16 - 18,	18 - 20,	20 - 22,	22 - 24
f	:	2	8	9	25	24	15	5

5. Calculate Inter-Quartile Range from the following :

Class	:	0 - 10,	10 - 20,	20 - 30,	30 - 40,	40 - 50.
f	:	4	10	15	10	4

♦

Lesson–19

DISPERSION II

In the previous lesson we have discussed various measures of dispersion based on range. Here, in this lesson we shall discuss average deviation and standard deviation.

Average Deviation

As the name sugests, average deviation is the arthmetic mean of the deviations of items from a measure of central tendency i.e. Mean, Median or Mode, Average deviation is also called Mean deviation. Related with a measure of central tendency, It is regarded a proper measure of dispersion, Although, average deviation can be calculated with the help of any of the three

averages mentioned above but in actual practice it is calculated right from mean or from median. Mode is usually not used as its value is indetermined, between mean and median the latter is supposed to be better than the former, because the sum of the deviations from the median is always less than the sum of the deviations from the mean. Median is also easily calculated. While taking deviations, plus and minus signs of deviations are ignored. Aggregate of such deviations is divided by the number of items in the series. The quotient obtained is the average deviation.

Average deviation is denoted by the Greek alphabet delta (δ). Alongwith delta the notation of the average used in getting deviations from different items of a series is also mentioned. Thus average deviation can be written symbollically in the following three ways :

1. Average deviation compuled with the use of arithmetic mean is denoted as δa or δx.
2. Average deviation computed with the use of median is denoted as δm or δM.
3. Average deviation computed with the use of mode is denoted as $-\delta z$ or δM.

Now we shall discuss procedure of computing average deviation in Individual, Discrete and Continuous series one by one.

(1) Individual Series — The following steps are taken in computation of average deviation in a series of individual observations :

(i) We decide the type of average to be used.

(ii) The value of the average selected is calculated.

(iii) Deviations are taken from each item of the series. In this process the value of the average calculated above is substracted fro each item of the series and balance is written in a column levelled (d) ignoring plus and minus sign. The sign of related average is also attacked with the notation of deviation (d) ignoring plus and minus sign. The sign of related average is also attached with the notation of deviation (d).

(iv) Aggregate of deviations is divided by the number of items in the series (N). The quotient so obtained is the average deviation.

Symbolically :

(a) Average deviation from arithmetic mean (δa) = $\frac{\Sigma da}{N}$

Where da = $|x - \bar{x}|$

(b) Average deviation from median (δM) = $\frac{\Sigma dm}{N}$

Where dM = $|x - Median|$

(c) Average deviation from mode (z) = $\frac{\Sigma dz}{N}$

Where dz = $|x - Mode|$

Average deviation, as calculated above, would be an absolute measure of dispersion, expressed in the same units in which the original data are. In order to transform it into a relative measure, it is divided by the average from which it has been calculated. It is known as co-efficient

of dispersion. Thus co-efficient of dispersion from mean, median and mode respectively are represented as below :

Co-efficient of mean deviation $= \frac{a}{\dot{a}} or \frac{\delta \bar{x}}{x}$

Co-efficient of median deviation $= \frac{\delta M}{M}$

Co-efficient of mode deviation $= \frac{\delta z}{z}$

The following example will explain the process of calculation of average deviation and its co-efficient in a series of individual observations :

Example 1.

Calculate average deviation from mean, median and mode and their co-efficient from the following :

x : 20, 20.25, 35, 40, 40, 40, 50, and 64

Solution :

Calculation of δ a, δ m and δ z

x	+ ignored da/a = 38	+ ignored dm/m = 40	ignored dz/z = 40
20	18	20	20
20	18	20	20
25	13	15	15
35	3	5	5
40	2	0	0
40	2	0	0
40	2	0	0
58	20	18	18
64	26	24	24
342	104	102	102

(1) Mean $= \frac{\Sigma x}{N} = \frac{342}{9} = 3$

(2) Median :

M = Value of $\frac{N+1}{2}$ th unit.

= Value of $\frac{9+1}{2}$ th unit.

= Value of 5th unit. = 40

As 40 occurs most frequenctly in the distribution it is the mode value in the distribution.

$\therefore$ z = 40

(1) Deviation from mean (δ a) = $\frac{\Sigma da}{N} = \frac{104}{9} = 11.56$

(2) Deviatoin from median (δ M) = $\frac{\Sigma dM}{N} = \frac{102}{9} = 11.33$

(3) Deviation from mode (δ z) = $\frac{\Sigma dz}{N} = \frac{102}{9} = 11.33$

Calculation of co-efficient of δ a, δ M and δ z.

(i) Co-efficient of δ a = $\frac{\delta a}{a} = \frac{11.56}{38} = 0.304$

(ii) Co-efficient of δ M = $\frac{\delta M}{M} = \frac{11.33}{40} = 0.283$

(iii) Co-efficient of δ z = $\frac{\delta z}{z} = \frac{11.33}{40} = 0.283$

Discrete Series :

The following steps are taken in the calculation of average deviation in a discrete series :

(i) Thevalue of the average (mean, median, mode) is calculated on the basis of which the average deviation has to be calculated.

(ii) Deviations from the average are obtained by substracting average value from the value of each item in the series. (d)

(iii) Deviations are multiplied to the corresponding frequency and aggregates of the product (fd) is obtained.

(iv) In the last, average deviation is calculated with the help of the following formula :

Deviation from mean (δ a) = $\frac{\Sigma fda}{N}$

Deviation from median (δ M) = $\frac{\Sigma fdM}{N}$

Deviation from mode (δ z) = $\frac{\Sigma fdz}{N}$

The following example can explain the procedure :

Example 2.

Calculate mean deviation and its co-efficient from the following :

x	: 2	3	4	5	6	7	8	9	10	11
f	: 9	6	2	2	2	4	3	3	2	3

Solution :

Any specific average has not been mentioned here. In such cases median is generally used in calculation of dispersion. Let us calculate average deviation from median (δ M) in this case.

Calculation of Average Deviation from Median (m)

x	f	Cumulative Frequency (c.f)	(x – M) m/m = 5	f. dm
2	9	9	3	27
3	6	15	2	12
4	2	17	1	2
5	2	19	0	0
6	2	21	1	2
7	4	25	2	8
8	3	28	3	9
9	3	31	4	12
10	2	33	5	10
11	3	36	6	18
	$\Sigma f = 36$			$\Sigma f.\ d\,M = 100$

Median :

M = Value of $\frac{N+1}{2}$ th unit

= Value of $\frac{36+1}{2}$ th unit

Value of 18.5th unit

Which lies in a cumulative frequency 19. Thus 5 is the median Value.

$\therefore$ $M = 5$

Average Deviation from Median (δM)

$$\delta M = \frac{\Sigma fdM}{N} = \frac{100}{36} = 2.78$$

Co-efficient of $\delta M = \frac{\delta M}{M} = \frac{2.78}{5} = 0.556$

Continuous Series :

The procedure of calculation of average deviation in a continous series is almost similar to the procedure used in a discrete series. The only difference between the two is that in a continous series deviations are taken from the mid values of the different class. The following example can explain the procedure :

Example 3.

Calculate average deviation and its co-efficient from the following data :

Class :	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
Frequency	10	20	30	20	10

Solution :

Median will be proper to be used for calculation of average deviation here.

Calculation of δM

Class	f	C.f	Mid Points	x – M = dm (M = 35)	f.dm
10-20	10	10	15	20	200
20-30	20	30	25	10	200
30-40	30	60	35	0	0
40-50	20	80	45	10	200
50-60	10	90	55	20	200
	$\Sigma f = 90$				$\Sigma f.dM = 800$

Calculation of Median :

Size of m $= \frac{N}{2}$ th unit

$= \frac{90}{2}$ th unit. = 45 th unit.

This size of m. i.e. 45 th unit has been included in cumulative frequency 60. Thus 30-40 is the median class.

Now :

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

Here, $l_1 = 30$, $l_2 = 40$, $f = 30$, $m = 45$ and $c = 30$.

By putting value in the above formula :

$$M = 30 + \frac{40-30}{30}(45 - 40)$$

$$= 30 + \frac{10}{30} \times 15 = 30 + 5 = 35$$

$$\delta M = \frac{\Sigma fdM}{N} = \frac{800}{90} = 8.89$$

Co-efficient of $\delta M = \frac{\delta M}{M} = \frac{8.89}{35} = 0.254$

Merits of Average or Mean Deviation :

(1) Based on all the items of distribution, it is superior meansure of dispersion.

(2) It is simple and easy.

(3) It is less affected by the extreme values of the distribution.

(4) Either of the three average (Mean, Median and Mode) can be used in its calculation.

Demerits of Average or Means Deviation :

(1) Mean deviation is not capable of further mathematical treatment because it ignores the algebraic signs of the deviations.

(2) It is not a well defined measure of dispersion, as it can be computed by taking deviations from any average. Mean deviation calculated from various averages will not e the same.

(3) It assigns equal importance to mean, median and mode whereas arithmetic mean is generally accepted as the best average. Sometimes mode is ill-defined. In such cases calculation of mode deviation is not advised.

Standard Deviation

Standard deviation is regarded as the most important measure of dispersion. It is denoted by Greek letter (σ) small sigma. In order to overcome the drawbacks of average deviation Karl Pearson introduced the concept of standard deviation in 1893. According while computing standard deviation of a set of data, the deviation of the different values are taken from the arithmetic mean. Unlike average deviation, algebaric signs are mentioned in course of taking deviations in the calculation of standard deviation.

Incorporting all the above features standard deviation has been defined as the square root of the average of the squared deviation from arithmetic mean. Standard deviation is used in advance mathematical applications.

$$\text{Standard Deviation } (\sigma) = \frac{\sqrt{\Sigma d^2_a}}{N}$$

The relative measure of standard deviation is called 'co-efficient of Variatoin." It is computed by dividing standard deviation (σ) by the arithmetic mean ($\bar{X}$ or a) of the distribution.

Symbolically :

$$\text{Co-efficient of Variation} = \frac{\sigma}{x}$$

Calculation of Variation

(1) Individual Series :

There are two methods of calculating standard deviation in an individual series–(i) Direct Method and (ii) Short-cut Method.

(i) Direct Method :

In the direct method, deviations are taken from the actual mean of the series. This method is appropriate when the actual mean is a whole number. The following steps are taken in the computation of standard deviation in an individual series by direct method :

(A) The actual arithmetic mean ($\bar{X}$) of the series is found out.

(B) The deviations of the data from the actual mean by subtracting the mean from each item of the data are found out. ($d = X - \bar{X}$)

(C) Each deviation is squared up (d^2) and their sum total (Σd^2) is found out.

(D) Sum of the squared deviations is divided by the number of items (N). The result is the mean of the squared deviations. ($\Sigma d^2 a \div N$)

(E) Square root of the mean of the spuared deviations is obtained. The result is the standard deviation.

Symbolically :

$$(\sigma) = \frac{\sqrt{\Sigma d^2 a}}{N} \text{ or } \sqrt{\frac{\Sigma (x - \bar{x})^2}{N}}$$

Here, σ = Standard deviation

Σ d^2a = Sum of the squared deviatons from arithmetic mean.

N = Number of items in the series.

The following example can explain the procedure.

Example 4.

Calculate standard deviation from the following data.

x : 20, 20, 25, 35, 40, 40, 40, 58 and 64

Solution :

x	da/a = 38	d^2a
20	–18	324
20	–18	324
25	–13	169
35	–3	9
40	2	4
40	2	4
40	2	4
58	20	400
64	26	676
$= \bar{x} \Sigma x = 342$	Σ d2a = 1914	$\frac{\Sigma x}{N} = \frac{342}{9} = 38$

$$\sigma = \frac{\sqrt{\Sigma d^2 a}}{N} = \sqrt{\frac{1914}{9}} = \sqrt{212.66} = 14.58$$

σ = 14.58

(ii) Short - cut Method :

As the name shows, this method simplifies the calculation work. It saves time. In direct method. if actual mean comes in decimal the calculation of squares of deviations becomes very difficult. In such circumstances we badly feel the desirability of short-cut method. For calculating standard deviation in an individual series by short-cut method the following procedures are adopted :

(A)

(1) Any number is as

(1) Any number is asumed as mean of the distribution. This assumed mean is denoted by

'A', For the ease of calculation, the Value in the middle of the distribution iis generally taken as assumed mean (A).

(2) Deviations are found out from the assumed mean (dA = x – A) and their sum total is obtained (ΣdA).

(3) The deviation are squared up and totally (Σd^2A).

(4) The standard deviation is obtained by the use of the following formula.

$$\sigma = \sqrt{\frac{\Sigma d^2 A}{N} - (\frac{\Sigma dA}{N})^2}$$

Here : d^2A = Sum of the square of the deviations from assumed mean.

N = Number of items.

ΣdA = Sum of the deviations from assumed mean.

A = Assumed mean.

Example 5.

Calculate standard deviation and the co-efficient of Variation of data given in example no. 4 by short-cut method.

Solution :

x	dA/A = 40	d^2A
20	–20	400
20	–20	400
25	–15	225
35	–5	25
40	0	0
40	0	0
40	0	0
58	18	324
64	24	576
	$\Sigma dA = -18$	$\Sigma d^2A = 1950$

$$\overline{X} = A + \frac{\Sigma dA}{N} = 40 + \frac{-18}{9} = 40 - 2 = 38$$

Assumed mean (A) = 40.

$$\sigma = \sqrt{\frac{\Sigma d^2 A}{N} - (\frac{\Sigma dA}{N})^2} = \sqrt{\frac{1950}{9} - (\frac{-18}{9})} = \sqrt{216.666 - 4}$$

$$= \sqrt{212.666} = 14.58.$$

Co-efficient of Standard Deviation.

Co-efficient of $\sigma = \frac{\sigma}{\bar{x}} = \frac{14.58}{38} = 0.3837$

(2) Discrete Series :

There are again two methods of calculating standard deviation in a discrete series—

(i) Direct Method and (ii) Short-cut Method.

(i) Direct Method :

For caculating standard deviation in a discrete series by direct method the following steps are taken.

(A) Actual arithmetic mean ($\bar{x}$ or a) of the series is computed.

(B) Deviations from mean is found out ($da = x - \bar{x}$)

(C) Deviations are squared up and multipled by their respective frequencey and the total of he product (Σ f. d2a) is found out.

(D) Standard deviation is computed by the use of the following formula :

$$\sigma = \sqrt{\frac{\Sigma fd^2a}{N}}$$

(ii) Short-cut Method :

The following steps are taken while computing standard deviation in a discrete series by short-cut method :

(A) Any number is assumed as mean of the series. The mean is denoted by 'A'.

(B) Deviations are obtained from the assumed mean. This is denoted BY 'dA'/

(C) Deviations are multiplied by their respective frequencies and total of the product (Σ fda) is obtained.

(D) Each of f-da is multipied by thier respective dA and the product $f.d^2A$ is totalled to get Σfd^2A.

(E) The standard deviation is obtained by the use of the following formula.

$$\sigma = \sqrt{\frac{\Sigma fd^2A}{N} - \left(\frac{\Sigma f.dA}{N}\right)^2}$$

The following example can explain the procedures involved in calculation of standard deviation by direct as well as short-cut method.

Example 6.

Calculate standard deviation of the following series by direct and short-cut method.

x	:	2	3	4	5	6	7	8
f	:	5	9	11	15	10	3	2

Solution :

Calculation of standard deviation by Direct Method

x	f	f.x	(x–a) da/o = 4.6	d^2a	f. d^2a
2	5	10	–2.6	6.76	33.80
3	9	27	–1.6	2.56	23.04
4	11	44	–0.6	0.36	3.26
5	15	75	+0.6	0.16	2.40
6	10	60	+1.4	1.96	19.60
7	3	21	+2.4	5,76	17.28
8	2	16	+3.4	11.56	23.12
	$\Sigma f = 55$	$\Sigma f.x. = 253$			$\Sigma fd^2a = 123.20$

$$\bar{X} = \frac{\Sigma fx}{N} = \frac{253}{55} = 4.6$$

$$\sigma = \sqrt{\frac{\Sigma fd^2A}{N}} = \sqrt{\frac{123}{55}}$$

$$= \sqrt{2.24} = 1.4966 \text{ or } 1.5$$

Calculation of standard deviation by short-cut method

x	f	(x–A) dA/A = 4	f.dA	(dA × fdA) $f.d^2A$
2	5	–2	–10	20
3	9	–1	–9	9
4	11	0	0	0
5	15	1	15	15
6	10	2	20	40
7	3	3	9	27
8	2	3	8	32
	55		33	143

Assumed mean (A) = 4

$$\sigma = \sqrt{\frac{\Sigma fd^2A}{N} - \left(\frac{\Sigma fdA}{N}\right)^2}$$

$$= \sqrt{\frac{143}{55} - \left(\frac{33}{55}\right)^2} = \sqrt{2.6 - 0.36}$$

$$= \sqrt{2.4} = 1.49066 \text{ or } 1.5$$

(3) Continous Series :

There are three methods of calculating standard deviation in a continous series–

(i) Direct Method (ii) Short-cut Method and (iii) Std-deviation Method.

In a continuous series, the mid values of the class intervals are found out. Then the procedure is the same as followed in discrete series.

The following example will explain the procedure clearly :

Example 7.

Calculate the standard deviation of the following by direct method :

Class :	10-20	20-30	30-40	40-50	50-60
f :	10	20	30	20	10

Solution :

Class	**f**	**M.V. (x)**	**f.x**	**da/a=35**	**d2a**	**f.d²a**
10-20	10	15	150	–20	400	4000
20-30	20	25	500	–10	100	2000
30-40	30	35	1050	0	0	0
40-50	20	45	900	10	100	2000
50-60	10	55	550	20	400	4000
	90		3150	0	1000	12000

$$\bar{x} = \frac{\Sigma f.x}{N} = \frac{3150}{90} = 35$$

$$\sigma = \sqrt{\frac{\Sigma fd^2{}_a}{N}} = \sqrt{\frac{1200}{90}}$$

$$= \sqrt{133.33} = 11.547 \text{ or } 11.55$$

Example 8.

Calculate the standard deviation of the data given in example no. 7 by short cut method.

Solution :

Class	f	M.V.(x)	dA/A	fdA	f.d²A
10-20	10	15	–30	–300	9000
20-30	20	25	–20	–400	8000
30-40	30	35	–10	–300	3000
40-50	20	45	0	0	0
50-60	10	55	10	100	1000
	90			–900	21000

Assumed mean (A) = 45

$$\sigma = \sqrt{\frac{\sum fd^2A}{N} - \frac{\sum fdA}{N}} = \sqrt{\frac{2100}{90} - \frac{-900}{90}}$$

$$= \sqrt{23.333 - 10} = \sqrt{133.33} = 11.5466 or 11.55$$

(iii) Step-deviation Method :

This method is applied to simplify the calculation process in continuous series of equal class intervals only. This can not be applied in a series of individual observations and in a discrete series having irregular gap between different items. The steps taken to calculate standard deviations by step deviation method are as follows :

(A) Deviation are taken from the assumed mean.

(B) Deviations are divided by a common factor which is generally equal to the magnitude of the class interval (i) and di A is found out.

(C) Each diA is multiplied by its corresponding frequency and sum total of the product ($\sum$ f.d'tA) is found out.

(D) Each of fd'A is multiplied by its corresponding diA and sum total of the product ($\sum$ f.d'²A) is obtained.

(E) In the last, standard deviation iscomputed with the use of the following formula.

$$\sigma = i \times \sqrt{\frac{\sum fdi^2A}{N} - \left(\frac{\sum fdiA}{N}\right)^2}$$

Here : i = Magnitude of the class interval or the common factor used or the equal gap between different discrete variables.

The following example can explain the process :

Example 9.

Calculate the standard deviation of the data given in example no. 7 using step deviation method.

Solution :

Standard deviation through step-deviation method

Class	f	M.V	dA /45	dt /i = 10	fd'A	f.dt²
10-20	10	15	–30	–3	–30	90
20-30	20	25	–20	–2	–40	80
30-40	30	35	–10	–1	–30	30
40-50	10	55	10	1	10	10
50-60	10	55	10	1	1	10
Total	90			–5	–90	210

$$\sigma = i \times \sqrt{\frac{\Sigma fdi^2}{N} - \left(\frac{\Sigma fdi}{N}\right)^2}$$

$$= 10 \times \sqrt{\frac{210}{90} - \frac{90^2}{90}} = 10 \times \sqrt{2.33 - 1}$$

$$= 10 \times \sqrt{1.3333} = 10 \times 1.1546 = 11.546 \text{ or } 11.56$$

Co-effiecient of Variation :

Like all other measures of dispersion, co-efficient of variation is calculated by dividing standard deviation by arithmetic mean of the series.

Symbolically :

$$\text{Co-efficient of Variation (C.V.)} = \frac{\sigma}{\bar{x}}$$

Co-effiecient of Variation is the relative measure of dispersion used in comperative study of dispersion between two or more series.

Symbolically :

$$\text{Co-efficient of variation (C.V)} = \frac{\sigma}{\bar{x}} \times 100$$

Co-efficient of variation is used to compare variability, homogeneity, stability, uniformity and consistency in two sets of data. The series having higher co-efficeient of variation has higher degree of variability.

Variance — Variance is the second moment of dispersion. The square of the standard deviation is called variance.

Symbolically :

$$\text{Variance} = \sigma^2 \text{ or } \sigma = \sqrt{\text{variance}}$$

Variance can also be calculated by removing root from the formula used in the calculation of standard deviation.

Merits of Standard Deviation

(1) It is considered to be the best measure of dispersion. Based on all the values in a distribution, it is popularly used in higher mathematical applications owing to its mathematical characteristics.

(2) It is less affected by fluctuation in sample.

(3) The standard deviation is a well defined and definite measure of dispersion. This measure can be obtained in every condition. Combined standard deviation can also be calculated for two or more series.

(4) The standard deviation is of great importance for sampling theory. It facilitates statistics study like skewness, correlation, test of significance, area under normal curve, regression etc.