

# STATISTICS

## MEASURES OF CENTRAL TENDENCY :

An average value or a central value of a distribution is the value of variable which is representative of the entire distribution, this representative value are called the measures of central tendency.

Generally the following five measures of central tendency.

(a) Mathematical average

(i) Arithmetic mean

(ii) Geometric mean

(iii) Harmonic mean

(b) Positional average

(i) Median

(ii) Mode

### 1. ARITHMETIC MEAN :

(i) **For ungrouped dist. :** If  $x_1, x_2, \dots, x_n$  are  $n$  values of variate  $x_i$  then their A.M.  $\bar{x}$  is defined as

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\Rightarrow \sum x_i = n\bar{x}$$

(ii) **For ungrouped and grouped freq. dist. :** If  $x_1, x_2, \dots, x_n$  are values of variate with corresponding frequencies  $f_1, f_2, \dots, f_n$  then their A.M. is given by

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^n f_i x_i}{N}, \text{ where } N = \sum_{i=1}^n f_i$$

(iii) **By short method :** If the value of  $x_i$  are large, then the calculation of A.M. by using previous formula is quite tedious and time consuming. In such case we take deviation of variate from an arbitrary point  $a$ .

Let  $d_i = x_i - a$

$$\therefore \bar{x} = a + \frac{\sum f_i d_i}{N}, \text{ where } a \text{ is assumed mean}$$

(iv) **By step deviation method :** Sometime during the application of short method of finding the A.M. If each deviation  $d_i$  are divisible by a common number  $h$  (let)

$$\text{Let } u_i = \frac{d_i}{h} = \frac{x_i - a}{h}$$

$$\therefore \bar{x} = a + \left( \frac{\sum f_i u_i}{N} \right) h$$

### Illustration 1 :

If the mean of the series  $x_1, x_2, \dots, x_n$  is  $\bar{x}$ , then the mean of the series  $x_i + 2i, i = 1, 2, \dots, n$  will be-

(1)  $\bar{x} + n$

(2)  $\bar{x} + n + 1$

(3)  $\bar{x} + 2$

(4)  $\bar{x} + 2n$

**Solution :**

$$\text{As given } \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \dots(1)$$

If the mean of the series  $x_i + 2i, i = 1, 2, \dots, n$  be  $\bar{X}$ , then

$$\bar{X} = \frac{(x_1 + 2) + (x_2 + 2.2) + (x_3 + 2.3) + \dots + (x_n + 2.n)}{n}$$

$$= \frac{x_1 + x_2 + \dots + x_n}{n} + \frac{2(1 + 2 + 3 + \dots + n)}{n}$$

$$= \bar{x} + \frac{2n(n+1)}{2n} \quad \text{from (1)}$$

$$= \bar{x} + n + 1$$

Ans. (2)

**Illustration 2 :**

Find the A.M. of the following freq. dist.

$x_i$	5	8	11	14	17
$f_i$	4	5	6	10	20

**Solution :**

Here  $N = \sum f_i = 4 + 5 + 6 + 10 + 20 = 45$

$$\sum f_i x_i = (5 \cdot 4) + (8 \cdot 5) + (11 \cdot 6) + (14 \cdot 10) + (17 \cdot 20) = 606$$

$$\therefore \bar{x} = \frac{\sum f_i x_i}{N} = \frac{606}{45} = 13.47$$

**Illustration 3 :**

Find the mean of the following freq. dist.

$x_i$	5	15	25	35	45	55
$f_i$	12	18	27	20	17	6

**Solution :**

Let assumed mean  $a = 35$ ,  $h = 10$

$$\text{here } N = \sum f_i = 100, \quad u_i = \frac{(x_i - 35)}{10}$$

$$\therefore \sum f_i u_i = (12 \cdot -3) + (18 \cdot -2) + (27 \cdot -1) + (20 \cdot 0) + (17 \cdot 1) + (6 \cdot 2) = -70$$

$$\therefore \bar{x} = a + \left( \frac{\sum f_i u_i}{N} \right) h = 35 + \frac{(-70)}{100} \cdot 10 = 28$$

**Illustration 4 :**

If a variable takes the value 0, 1, 2, ..., n with frequencies proportional to the binomial coefficients  ${}^nC_0, {}^nC_1, \dots, {}^nC_n$  then the mean of the distribution is-

$$(1) \frac{n(n+1)}{4}$$

$$(2) \frac{n}{2}$$

$$(3) \frac{n(n-1)}{2}$$

$$(4) \frac{n(n+1)}{2}$$

**Solution :**

$$N = \sum f_i = k [{}^nC_0 + {}^nC_1 + \dots + {}^nC_n] = k2^n$$

$$\sum f_i x_i = k [1 \cdot {}^nC_1 + 2 \cdot {}^nC_2 + \dots + n \cdot {}^nC_n] = k \sum_{r=1}^n r \cdot {}^nC_r = kn \sum_{r=1}^n {}^{n-1}C_{r-1} = kn2^{n-1}$$

$$\text{Thus } \bar{x} = \frac{1}{2^n} (n \cdot 2^{n-1}) = \frac{n}{2}.$$

**Ans. (2)**

(v) **Weighted mean :** If  $w_1, w_2, \dots, w_n$  are the weights assigned to the values  $x_1, x_2, \dots, x_n$  respectively then their weighted mean is defined as

$$\text{Weighted mean} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + \dots + w_n} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

**Illustration 5 :**

Find the weighted mean of first n natural numbers when their weights are equal to their squares respectively

**Solution :**

$$\text{Weighted Mean} = \frac{1 \cdot 1^2 + 2 \cdot 2^2 + \dots + n \cdot n^2}{1^2 + 2^2 + \dots + n^2} = \frac{1^3 + 2^3 + \dots + n^3}{1^2 + 2^2 + \dots + n^2} = \frac{[n(n+1)/2]^2}{[n(n+1)(2n+1)/6]} = \frac{3n(n+1)}{2(2n+1)}$$

(vi) **Combined mean** : If  $\bar{x}_1$  and  $\bar{x}_2$  be the means of two groups having  $n_1$  and  $n_2$  terms respectively then the mean (combined mean) of their composite group is given by

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

$$\text{If there are more than two groups then, combined mean} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + \dots}{n_1 + n_2 + n_3 + \dots}$$

**Illustration 6 :**

The mean income of a group of persons is Rs. 400 and another group of persons is Rs. 480. If the mean income of all the persons of these two groups is Rs. 430 then find the ratio of the number of persons in the groups.

**Solution :**

$$\text{Here } \bar{x}_1 = 400, \bar{x}_2 = 480, \bar{x} = 430$$

$$\therefore \bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \Rightarrow 430 = \frac{400n_1 + 480n_2}{n_1 + n_2}$$

$$\Rightarrow \frac{n_1}{n_2} = \frac{5}{3}$$

**(vii) Properties of Arithmetic mean :**

- Sum of deviations of variate from their A.M. is always zero i.e.  $\sum(x_i - \bar{x}) = 0$ ,  $\sum f_i(x_i - \bar{x}) = 0$
- Sum of square of deviations of variate from their A.M. is minimum i.e.  $\sum(x_i - \bar{x})^2$  is minimum
- If  $\bar{x}$  is the mean of variate  $x_i$  then
  - A.M. of  $(x_i + \lambda) = \bar{x} + \lambda$
  - A.M. of  $(\lambda x_i) = \lambda \bar{x}$
  - A.M. of  $(ax_i + b) = a\bar{x} + b$  (where  $\lambda, a, b$  are constant)
- A.M. is independent of change of assumed mean i.e. it is not effected by any change in assumed mean.

**Do yourself - 1 :**

(i) If in an examination different weights are assigned to different subjects Physics (2), Chemistry (1), English (1), Mathematics (2) A student scores 60 in Physics, 70 in Chemistry, 70 in English and 80 in Mathematics, then weighted mean is -

- (1) 60                      (2) 70                      (3) 80                      (4) 85

(ii) The mean of the following freq. table is 50 and  $\Sigma f = 120$

class	0-20	20-40	40-60	60-80	80-100
f	17	$f_1$	32	$f_2$	19

the missing frequencies are-

- (1) 28, 24                      (2) 24, 36                      (3) 36, 28                      (4) None of these

(iii) A student obtained 75%, 80%, 85% marks in three subjects. If the marks of another subject are added then his average marks can not be less than-

- (1) 60%                      (2) 65%                      (3) 80%                      (4) 90%

**2. GEOMETRIC MEAN :**

(i) **For ungrouped dist. :** If  $x_1, x_2, \dots, x_n$  are  $n$  positive values of variate then their geometric mean  $G$  is given by

$$G = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{1/n}$$

$$\Rightarrow G = \text{antilog} \left[ \frac{1}{n} \sum_{i=1}^n \log x_i \right]$$

- (ii) For freq. dist. : If  $x_1, x_2, \dots, x_n$  are  $n$  positive values with corresponding frequencies  $f_1, f_2, \dots, f_n$  resp. then their G.M.

$$G = (x_1^{f_1} \times x_2^{f_2} \times \dots \times x_n^{f_n})^{1/N}$$

$$\Rightarrow G = \text{antilog} \left[ \frac{1}{N} \sum_{i=1}^n f_i \log x_i \right]$$

**Note :-** If  $G_1$  and  $G_2$  are geometric means of two series which containing  $n_1$  and  $n_2$  positive values resp. and  $G$  is geometric mean of their combined series then

$$G = (G_1^{n_1} \times G_2^{n_2})^{\frac{1}{n_1+n_2}}$$

$$\Rightarrow G = \text{antilog} \left[ \frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2} \right]$$

**Illustration 7 :**

Find the G.M. of 1, 2,  $2^2$ , ...,  $2^n$

**Solution :**

$$\begin{aligned} \text{G.M.} &= (1 \cdot 2 \cdot 2^2 \cdot \dots \cdot 2^n)^{\frac{1}{n+1}} \\ &= \left[ 2^{\frac{n(n+1)}{2}} \right]^{\frac{1}{n+1}} = 2^{n/2} \end{aligned}$$

**3. HARMONIC MEAN :**

- (i) For ungrouped dist. : If  $x_1, x_2, \dots, x_n$  are  $n$  non-zero values of variate then their harmonic mean  $H$  is defined as

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

- (ii) For freq. dist. : If  $x_1, x_2, \dots, x_n$  are  $n$  non-zero values of variate with corresponding frequencies  $f_1, f_2, \dots, f_n$  respectively then their H.M.

$$H = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}} = \frac{N}{\sum_{i=1}^n \frac{f_i}{x_i}}$$

**Illustration 8 :**

Find the H.M. of  $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{17}$

**Solution :**

$$\text{H.M.} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{16}{2+3+\dots+17} = \frac{2}{19}$$

**Note :-** If  $A, G, H$  are A.M. G.M. H.M. of a series respectively then

$$A \geq G \geq H$$

#### 4. MEDIAN :

The median of a series is the value of middle term of the series when the values are written in ascending order. Therefore median, divided an arranged series into two equal parts.

**Formulae of median :**

(i) **For ungrouped distribution :** Let  $n$  be the number of variate in a series then

$$\text{Median} = \begin{cases} \left(\frac{n+1}{2}\right)^{\text{th}} \text{ term, (when } n \text{ is odd)} \\ \text{Mean of } \left(\frac{n}{2}\right)^{\text{th}} \text{ and } \left(\frac{n}{2}+1\right)^{\text{th}} \text{ terms, (when } n \text{ is even)} \end{cases}$$

(ii) **For ungrouped freq. dist. :** First we prepare the cumulative frequency (c.f.) column and Find value of  $N$  then

$$\text{Median} = \begin{cases} \left(\frac{N+1}{2}\right)^{\text{th}} \text{ term, (when } N \text{ is odd)} \\ \text{Mean of } \left(\frac{N}{2}\right)^{\text{th}} \text{ and } \left(\frac{N}{2}+1\right)^{\text{th}} \text{ terms, (when } N \text{ is even)} \end{cases}$$

(iii) **For grouped freq. dist. :** Prepare c.f. column and find value of  $\frac{N}{2}$  then find the class which contain value of c.f. is equal or just greater to  $N/2$ , this is median class

$$\therefore \text{Median} = \ell + \frac{\left(\frac{N}{2} - F\right)}{f} \quad h$$

where

$\ell$  — lower limit of median class

$f$  — freq. of median class

$F$  — c.f. of the class preceeding median class

$h$  — Class interval of median class

#### Illustration 9 :

Find the median of following freq. dist.

class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
f	8	30	40	12	10

**Solution :**

class	$f_i$	c.f.
0 – 10	8	8
10 – 20	30	38
20 – 30	40	78
30 – 40	12	90
40 – 50	10	100

Here  $\frac{N}{2} = \frac{100}{2} = 50$  which lies in the value 78 of c.f. hence corresponding class of this c.f. is 20-30 is the median class, so

$$\ell = 20, f = 40, F = 38, h = 10$$

$$\therefore \text{Median} = \ell + \frac{\left(\frac{N}{2} - F\right)}{f} \quad h = 20 + \frac{(50 - 38)}{40} \times 10 = 23$$

## 5. MODE :

In a frequency distribution the mode is the value of that variate which have the maximum frequency

**Method for determining mode :**

(i) **For ungrouped dist. :** The value of that variate which is repeated maximum number of times

(ii) **For ungrouped freq. dist. :** The value of that variate which have maximum frequency.

(iii) **For grouped freq. dist. :** First we find the class which have maximum frequency, this is model class

$$\therefore \text{Mode} = \ell + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} h$$

where  $\ell$  — lower limit of model class

$f_0$  — freq. of the model class

$f_1$  — freq. of the class preceeding model class

$f_2$  — freq. of the class succeeding model class

$h$  — class interval of model class

### Illustration 10 :

Find the mode of the following frequency dist

class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
$f_i$	2	18	30	45	35	20	6	3

### Solution :

Here the class 30–40 has maximum freq. so this is the model class

$$\ell = 30, f_0 = 45, f_1 = 30, f_2 = 35, h = 10$$

$$\therefore \text{Mode} = \ell + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} h = 30 + \frac{45 - 30}{2 \times 45 - 30 - 35} \times 10 = 36$$

## 6. RELATION BETWEEN MEAN, MEDIAN AND MODE :

In a moderately asymmetric distribution following relation between mean, median and mode of a distribution.

It is known as imprical formula.

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

**Note** (i) Median always lies between mean and mode

(ii) For a symmetric distribution the mean, median and mode are coincide.

### Do yourself - 2 :

(i) Median of the distribution :

$${}^{20}C_0, {}^{20}C_{19}, {}^{20}C_2, {}^{20}C_{17}, {}^{20}C_4, {}^{20}C_{15}, {}^{20}C_9$$

$${}^{20}C_{10}, {}^{20}C_6, {}^{20}C_{13}, {}^{20}C_{12} \text{ will be -}$$

$$(1) {}^{20}C_6$$

$$(2) {}^{20}C_{15}$$

$$(3) {}^{20}C_9$$

$$(4) \text{None}$$

(ii) Let a, b, c and d are real numbers ( $d > a > b > c$ ). If mean and median of the distribution a, b, c, d are 5 and 6 respectively then the value of  $-a + 3d + 3c - b$  is :

$$(1) 8$$

$$(2) 10$$

$$(3) 12$$

$$(4) \text{None}$$

(iii) Let median of 23 observations is 50 if smallest 13 observations are increased by 2 then median will become :-

$$(1) 50$$

$$(2) 52$$

$$(3) \text{Can't say anything} \quad (4) \text{None of these}$$

## 7. MEASURES OF DISPERSION :

The dispersion of a statistical distribution is the measure of deviation of its values about the their average (central) value.

It gives an idea of scatteredness of different values from the average value.

Generally the following measures of dispersion are commonly used.

- (i) Range (ii) Mean deviation (iii) Variance and standard deviation

**(i) Range :** The difference between the greatest and least values of variate of a distribution, are called the range of that distribution.

If the distribution is grouped distribution, then its range is the difference between upper limit of the maximum class and lower limit of the minimum class.

$$\text{Also, coefficient of range} = \frac{\text{difference of extreme values}}{\text{sum of extreme values}}$$

### Illustration 11 :

Find the range of following numbers 10, 8, 12, 11, 14, 9, 6

### Solution :

Here greatest value and least value of the distribution are 14 and 6 resp. therefore

$$\text{Range} = 14 - 6 = 8$$

**(ii) Mean deviation (M.D.) :** The mean deviation of a distribution is, the mean of absolute value of deviations of variate from their statistical average (Mean, Median, Mode).

If A is any statistical average of a distribution then mean deviation about A is defined as

$$\text{Mean deviation} = \frac{\sum_{i=1}^n |x_i - A|}{n} \quad (\text{for ungrouped dist.})$$

$$\text{Mean deviation} = \frac{\sum_{i=1}^n f_i |x_i - A|}{N} \quad (\text{for freq. dist.})$$

**Note :-** Mean deviation is minimum when it taken about the median

### Illustration 12 :

Find the mean deviation of number 3, 4, 5, 6, 7

### Solution :

Here  $n = 5$ ,  $\bar{x} = 5$

$$\begin{aligned} \therefore \text{Mean deviation} &= \frac{\sum |x_i - \bar{x}|}{n} \\ &= \frac{1}{5} [|3 - 5| + |4 - 5| + |5 - 5| + |6 - 5| + |7 - 5|] \\ &= \frac{1}{5} [2 + 1 + 0 + 1 + 2] = \frac{6}{5} = 1.2 \end{aligned}$$

### Illustration 13 :

Find the mean deviation about mean from the following data

$x_i$	3	9	17	23	27
$f_i$	8	10	12	9	5

**Solution :**

$x_i$	$f_i$	$f_i x_i$	$ x_i - \bar{x} $	$f_i  x_i - \bar{x} $
3	8	24	12	96
9	10	90	6	60
17	12	204	2	24
23	9	207	8	72
27	5	135	12	60
	$N = 44$	$\Sigma f_i x_i = 660$		$\Sigma f_i  x_i - \bar{x}  = 312$

$$\text{Mean } (\bar{x}) = \frac{\Sigma f_i x_i}{N} = \frac{660}{44} = 15$$

$$\text{Mean deviation} = \frac{\Sigma f_i |x_i - \bar{x}|}{N} = \frac{312}{44} = 7.09$$

**Do yourself - 3 :**

- (i) The mean deviation about median from the following data 340, 150, 210, 240, 300, 310, 320, is-  
 (1) 52.4 (2) 52.5 (3) 52.8 (4) none of these

- (ii) The mean deviation of the series  $a, a + d, a + 2d, \dots, a + 2nd$  from its mean is-

- (1)  $\frac{n+1}{2n+1} |d|$  (2)  $\frac{n(n+1)}{2n+1} |d|$  (3)  $\frac{n(n-1)}{2n+1} |d|$  (4) none of these

**(iii) Variance and standard deviation :** The variance of a distribution is, the mean of squares of deviation of variate from their mean. It is denoted by  $\sigma^2$  or  $\text{var}(x)$ .

The positive square root of the variance are called the standard deviation. It is denoted by  $\sigma$  or S.D.

Hence standard deviation =  $+\sqrt{\text{variance}}$

**Formulae for variance :**

**(i) for ungrouped dist. :**

$$\sigma_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$\sigma_x^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2 = \frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2$$

$$\sigma_d^2 = \frac{\Sigma d_i^2}{n} - \left(\frac{\Sigma d_i}{n}\right)^2, \text{ where } d_i = x_i - a$$

**(ii) For freq. dist. :**

$$\sigma_x^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N}$$

$$\sigma_x^2 = \frac{\Sigma f_i x_i^2}{N} - (\bar{x})^2 = \frac{\Sigma f_i x_i^2}{N} - \left(\frac{\Sigma f_i x_i}{N}\right)^2$$

$$\sigma_d^2 = \frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2$$

$$\sigma_u^2 = h^2 \left[ \frac{\Sigma f_i u_i^2}{N} - \left(\frac{\Sigma f_i u_i}{N}\right)^2 \right] \text{ where } u_i = \frac{d_i}{h}$$



(iii) Coefficient of S.D. =  $\frac{\sigma}{\bar{x}}$

Coefficient of variation =  $\frac{\sigma}{\bar{x}} \times 100$  (in percentage)

Note :-  $\sigma^2 = \sigma_x^2 = \sigma_d^2 = h^2 \sigma_u^2$

**Illustration 14 :**

Find the variance of first n natural numbers

**Solution :**

$$\sigma^2 = \frac{\sum x_i^2}{n} - \left( \frac{\sum x_i}{n} \right)^2 = \frac{\sum n^2}{n} - \left( \frac{\sum n}{n} \right)^2 = \frac{n(n+1)(2n+1)}{6n} - \left( \frac{n(n+1)}{2n} \right)^2 = \frac{n^2-1}{12}$$

**Illustration 15 :**

If  $\sum_{i=1}^{18} (x_i - 8) = 9$  and  $\sum_{i=1}^{18} (x_i - 8)^2 = 45$ , then find the standard deviation of  $x_1, x_2, \dots, x_{18}$

**Solution :**

Let  $(x_i - 8) = d_i$

$$\therefore \sigma_x = \sigma_d = \sqrt{\frac{\sum d_i^2}{n} - \left( \frac{\sum d_i}{n} \right)^2} = \sqrt{\frac{45}{18} - \left( \frac{9}{18} \right)^2} = \sqrt{\frac{5}{2} - \frac{1}{4}} = \frac{3}{2}$$

**Illustration 16 :**

Find the coefficient of variation of first n natural numbers

**Solution :**

For first n natural numbers.

$$\text{Mean } (\bar{x}) = \frac{n+1}{2}, \text{ S.D.}(\sigma) = \sqrt{\frac{n^2-1}{12}}$$

$$\therefore \text{coefficient of variance} = \frac{\sigma}{\bar{x}} \times 100 = \sqrt{\frac{n^2-1}{12}} \times \frac{1}{\left(\frac{n+1}{2}\right)} \times 100 = \sqrt{\frac{(n-1)}{3(n+1)}} \times 100$$

**8. MEAN SQUARE DEVIATION :**

The mean square deviation of a distribution is the mean of the square of deviations of variate from assumed mean. It is denoted by  $S^2$

Hence 
$$S^2 = \frac{\sum (x_i - a)^2}{n} = \frac{\sum d_i^2}{n} \quad (\text{for ungrouped dist.})$$

$$S^2 = \frac{\sum f_i (x_i - a)^2}{N} = \frac{\sum f_i d_i^2}{N} \quad (\text{for freq. dist.}), \quad \text{where } d_i = (x_i - a)$$

**Illustration 17 :**

The mean square deviation of a set of n observations  $x_1, x_2, \dots, x_n$  about a point c is defined as  $\frac{1}{n} \sum_{i=1}^n (x_i - c)^2$

The mean square deviation about -2 and 2 are 18 and 10 respectively, then standard deviation of this set of observations is-

- (1) 3 (2) 2 (3) 1 (4) None of these

**Solution :**

$$\therefore \frac{1}{n} \sum (x_i + 2)^2 = 18 \text{ and } \frac{1}{n} \sum (x_i - 2)^2 = 10$$

$$\Rightarrow \sum (x_i + 2)^2 = 18n \text{ and } \sum (x_i - 2)^2 = 10n$$

$$\Rightarrow \sum (x_i + 2)^2 + \sum (x_i - 2)^2 = 28n \text{ and } \sum (x_i + 2)^2 - \sum (x_i - 2)^2 = 8n$$

$$\Rightarrow 2\sum x_i^2 + 8n = 28n \text{ and } 8\sum x_i = 8n$$

$$\Rightarrow \sum x_i^2 = 10n \text{ and } \sum x_i = n$$

$$\Rightarrow \frac{\sum x_i^2}{n} = 10 \text{ and } \frac{\sum x_i}{n} = 1$$

$$\therefore \sigma = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{10 - (1)^2} = 3$$

**Ans. (1)**

### 9. RELATION BETWEEN VARIANCE AND MEAN SQUARE DEVIATION :

$$\therefore \sigma^2 = \frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2$$

$$\Rightarrow \sigma^2 = s^2 - d^2, \quad \text{where } d = \bar{x} - a = \frac{\sum f_i d_i}{N}$$

$$\Rightarrow s^2 = \sigma^2 + d^2 \Rightarrow s^2 \geq \sigma^2$$

Hence the variance is the minimum value of mean square deviation of a distribution

### Illustration 18 :

Determine the variance of the following frequency dist.

class	0-2	2-4	4-6	6-8	8-10	10-12
$f_i$	2	7	12	19	9	1

**Solution :**

Let  $a = 7, h = 2$

class	$x_i$	$f_i$	$u_i = \frac{x_i - a}{h}$	$f_i u_i$	$f_i u_i^2$
0-2	1	2	-3	-6	18
2-4	3	7	-2	-14	28
4-6	5	12	-1	-12	12
6-8	7	19	0	0	0
8-10	9	9	1	9	9
10-12	11	1	2	2	4
		$N = 50$		$\sum f_i u_i = -21$	$\sum f_i u_i^2 = 71$

$$\therefore \sigma^2 = h^2 \left[ \frac{\sum f_i u_i^2}{N} - \left(\frac{\sum f_i u_i}{N}\right)^2 \right] = 4 \left[ \frac{71}{50} - \left(\frac{-21}{50}\right)^2 \right] = 4[1.42 - 0.1764] = 4.97$$

10. MATHEMATICAL PROPERTIES OF VARIANCE :

- $\text{Var}(x_i + \lambda) = \text{Var}(x_i)$   
 $\text{Var}(\lambda x_i) = \lambda^2 \cdot \text{Var}(x_i)$   
 $\text{Var}(ax_i + b) = a^2 \cdot \text{Var}(x_i)$   
 where  $\lambda, a, b$ , are constant
- If means of two series containing  $n_1, n_2$  terms are  $\bar{x}_1, \bar{x}_2$  and their variance's are  $\sigma_1^2, \sigma_2^2$  respectively and their combined mean is  $\bar{x}$  then the variance  $\sigma^2$  of their combined series is given by following formula

$$\sigma^2 = \frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{(n_1 + n_2)} \quad \text{where } d_1 = \bar{x}_1 - \bar{x}, d_2 = \bar{x}_2 - \bar{x}$$

i.e. 
$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2} + \frac{n_1n_2}{(n_1 + n_2)^2} (\bar{x}_1 - \bar{x}_2)^2$$

Do yourself - 4 :

- (i) The variance of first 20-natural numbers is-

(1)  $\frac{133}{4}$                       (2)  $\frac{379}{12}$                       (3)  $\frac{133}{2}$                       (4)  $\frac{399}{4}$

- (ii) The mean and variance of a series containing 5 terms are 8 and 24 respectively. The mean and variance of another series containing 3 terms are also 8 and 24 respectively. The variance of their combined series will be-

(1) 20                      (2) 24                      (3) 25                      (4) 42

- (iii) Variance of the data given below is

Size of item	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Frequency	3	7	22	60	85	32	8

(1) 1.29                      (2) 2.19                      (3) 1.32                      (4) none of these

- (iv) The mean and variance of 5 observations of an experiment are 4 and 5.2 respectively. If from these observations three are 1, 2 and 6, then the remaining will be-

(1) 2, 9                      (2) 5, 6                      (3) 4, 7                      (4) 3, 8

ANSWERS FOR DO YOURSELF

1. (i) 2                      (ii) 1                      (iii) 1

2. (i) 2                      (ii) 3                      (iii) 3

3. (i) 3                      (ii) 2

4. (i) 1                      (ii) 2                      (iii) 3                      (iv) 3