06 Statistics

Measures of Central Tendency (OR Average Values)

An Average value of a distribution is the value of variable which is representative of the whole distribution. Generally Average values lies in the central part of the distribution and they are also called Measures of central tendency.

The following five measures of central tendency can be divided in two groups.

(I) Mathematical Average:
- (i) Arithmetic Mean or Mean
- (ii) Geometric Mean
- (iii) Harmonic Mean

(II) Positional Average:
- (i) Median
- (ii) Mode

Arithmetic Mean (A.M.)

(i) For Ungrouped (or Individual) Distribution:

If $x_1, x_2, ..., x_n$ are n values of variate then their arithmetic mean $(A.M.)$ is

$$= \frac{x_1 + x_2 + ... + x_n}{n}$$

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}} \quad \text{or} \quad \boxed{\sum_{i=1}^{n} x_i = n\bar{x}}$$

Illustration 1:

In a group of 30 observation mean of first 10 is 12 and last 20 is 9, then find mean of whole distribution.

Solution:

Sum of first 10 observation $= 10 \times 12 = 120$

Sum of last 20 observation $= 20 \times 9 = 180$

Sum of 30 observation $= 120 + 180 = 300$

Mean $= \frac{\sum x_i}{n} = \frac{300}{30} = 10$

Illustration 2:

Mean of n observation $1^2, 2^2, 3^2, ... n^2$ is $\frac{46n}{11}$. Then value of n is :

(A) 11 (B) 12 (C) 23 (D) 22

Ans. (A)

Solution:

$$\bar{x} = \frac{\sum x_i}{n} \Rightarrow \frac{46n}{11} = \frac{1^2 + 2^2 + ... n^2}{n}$$

$$\Rightarrow \frac{46n}{11} = \frac{n(n+1)(2n+1)}{6n}$$

Solving

$n = 11$

Illustration 3:

Find the $A.M.$ of the series $1,2,4,8,16,....,2^n$

(A) $\frac{2^{n+1}-1}{n+1}$ (B) $\frac{2^{n+2}-1}{n}$ (C) $\frac{2^n-1}{n+1}$ (D) $\frac{2^n-1}{n}$

Solution:

$$\frac{1+2+4......2^n}{n+1}=\frac{2^{n+1}-1}{2-1}\cdot\frac{1}{n+1}$$

Arithmetic Mean (A.M.)

For Ungrouped Frequency Distribution:

If $x_1, x_2, ..., x_n$ are n values of variate x_i and their frequencies are $f_1, f_2, ..., f_n$ respectively, then AM of this frequency distribution is -

$$\bar{x}=\frac{f_1x_1+f_2x_2+......+f_nx_n}{(f_1+f_2+......+f_n)}$$

$$\bar{x}=\frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \quad \text{or} \quad \boxed{\bar{x}=\frac{\sum_{i=1}^{n} f_i x_i}{N}} \qquad \text{Where } N=\sum_{i=1}^{n} f_i$$

Note: For AM of grouped frequency distribution, we find the middle values of each group (or class) and changed it into ungrouped frequency distribution, then used above formula.

Illustration 4:

Find the A.M. of the following frequency distribution.

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

Solution:

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

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

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

Short Method/Assumed Mean Method (By Change of Origin)

In this method we take deviation (d_i) of each variate (x_i) from any central value (a).

Let $x_i-a=d_i \Rightarrow x_i=a+d_i$

Then $\bar{x}=\frac{\sum f_i(a+d_i)}{N}=\frac{a\sum f_i}{N}+\frac{\sum f_i d_i}{N}$

$\boxed{\bar{x}=a+\frac{\sum f_i d_i}{N}}$ Where a is assumed mean (any central value of variate x_i)

Step Deviation Method (By Change of Origin and Scale)

In this method, if each value of deviation (d_i) is divisible by any common number h (let)

Let $\frac{d_i}{h}=u_i \Rightarrow d_i=hu_i$

Then $\boxed{\bar{x}=a+\frac{\sum f_i u_i}{N}\times h}$

Illustration 5:

Find the $A.M$ of the following frequency distribution

Wage(in Rs.)	800	820	860	900	920	980	1000
No. of Worker	7	14	19	24	20	11	5

Solution:

Let $a = 900, h = 20$

x_i	f_i	$d_i = x_i$	$u_i = d_i / h$	$f_i u_i$
800	7	–100	–5	–35
820	14	–80	–4	–56
860	19	–40	–2	–38
900	24	0	0	0
920	20	20	1	20
980	11	80	4	44
1000	5	100	5	25
	N = 100			$\sum f_i u_i = -40$

$\because \ \bar{x} = a + \dfrac{\sum f_i u_i}{N} \times h$

$= 900 + \dfrac{(-40)}{100} \times 20$

= 900 – 8

= 892

(v) Weighted Mean (W.M.)

If $x_1, x_2, ..., x_n$ are n values of variate x_i and their weightage are $w_1, w_2, ..., w_n$ respectively, then their weighted mean is -

$$W.M. = \frac{w_1x_1 + w_2x_2 + ... + w_nx_n}{w_1 + w_2 + ... + w_n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Illustration 6:

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

Solution:

$$\text{Weighted Mean} = \frac{1.1^2 + 2.2^2 + ... + n.n^2}{1^2 + 2^2 + ... + n^2} = \frac{1^3 + 2^3 + ... + n^3}{1^2 + 2^2 + ... + n^2} = \frac{[n(n+1)/2]^2}{[n(n+1)(2n+1)/6]} = \frac{3n(n+1)}{2(2n+1)}$$

(vi) Combined Mean (C.M.)

If $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ are mean of observations of k series which have $n_1, n_2, ..., n_k$ terms respectively, then their combined mean is-

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

If $k = 2$ then $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Illustration 7:

Average salary of male employee in a firm was Rs. 5200 and of female employee was Rs. 4200. The mean salary of all employee was Rs. 5000, then percentage of male employee in a firm is:

Solution:

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

$$5000 = \frac{5200n_1 + 4200n_2}{n_1 + n_2} \Rightarrow n_1 = 4n_2$$

Percentage of male = $\frac{n_1}{n_1 + n_2} \times 100$

$= \frac{4n_2}{5n_2} \times 100 = 80\%$

Properties of AM

(1) The sum of deviation of variate from their AM is always zero

i.e. $\sum(x_i - \bar{x}) = 0$, $\sum f_i(x_i - \bar{x}) = 0$

(2) The sum of square of deviation of variate from their AM is minimum

i.e. $\sum(x_i - \bar{x})^2 \le \sum(x_i - a)^2$ (where a is any central value)

(3) If the AM of the observations $x_1, x_2, ..., x_n$ is $\bar{x}$ then

AM of $x_1 \pm \lambda, x_2 \pm \lambda, ..., x_n \pm \lambda$ is $\bar{x} \pm \lambda$

AM of $\lambda x_1, \lambda x_2, ..., \lambda x_n$ is $\lambda\bar{x}$

AM of $ax_1 \pm b, ax_2 \pm b, ..., ax_n \pm b$ is a $\bar{x} \pm b$(Where λ, a, b are constant)

(4) AM is not affected by any change in assumed mean.

Illustration 8:

Mean of 16 observations is 16. If one observation of value 16 is deleted and 3 observations 3, 4, 5 are added to this set, then find mean of new set of observations.

Solution:

Sum of 16 observation $= 16 \times 16 = 256$

New sum = $256 - 16 + (3 + 4 + 5) = 252$

New means = $\frac{\sum x_i}{n} = \frac{252}{18} = 14$

Illustration 9:

Mean age of 25 teacher in a school is 40 years. A teacher is retire at age of 60 years and a new teacher is appointed in his place. If new mean age of teacher in school is 39 year. Then age of newly appointed teacher is

(A) 35 (B) 30 (C) 40 (D) 25

Solution:

Sum of age of 25 teachers $= 25 \times 40 = 1000$ years.

Let age of newly appointed teacher is x year

New sum = $1000 - 60 + x$

New means = $\frac{1000 - 60 + x}{25}$

$$39 = \frac{940 + x}{25}$$

$x = 35$ years

Median:

Median is the middle value of a series when the values of the series are arranged in ascending or descending order therefore median divides an arranged series into two equal parts.

(i) For Ungrouped Distribution:

Let n be the number of terms of series, first we arranged the series in the ascending or descending order then in this arranged series.

$$\text{Median} = \begin{cases} \left(\frac{n+1}{2}\right)^{th} \text{term} & (\text{when } n \text{ is odd}) \\ \dfrac{\left(\frac{n}{2}\right)^{th} \text{term} + \left(\frac{n}{2}+1\right)^{th} \text{term}}{2} & (\text{when } n \text{ is even}) \end{cases}$$

(ii) For Ungrouped Frequency Distribution:

(a) First, we arranged the value of variate x_i, in ascending order with their frequency.

(b) Prepare cumulative frequency $(c.f.)$ column and find the value of N.

$$\text{Median} = \begin{cases} \left(\frac{N+1}{2}\right)^{th} \text{term} & (\text{when } N \text{ is odd}) \\ \dfrac{\left(\frac{N}{2}\right)^{th} \text{term} + \left(\frac{N}{2}+1\right)^{th} \text{term}}{2} & (\text{when } N \text{ is even}) \end{cases}$$

Illustration 10:

Find the median of following frequency distribution.

x_i	42	52	62	72	47	57	67
f_i	3	6	11	9	8	8	5

Solution:

x_i	f_i	$C.f$
42	3	3
47	8	11
52	6	17
57	8	25
62	11	36
67	5	41
72	9	50

Here $N = 50$ (even)

$$\text{Median} = \frac{\left(\frac{N}{2}\right)^{th} \text{term} + \left(\frac{N}{2}+1\right)^{th} \text{term}}{2}$$

$$= \frac{25^{th} \text{ term} + 26^{th} \text{ term}}{2}$$

$$= \frac{57+62}{2} = \frac{119}{2} = 59.5$$

(iii) For Grouped Frequency Distribution:

(a) Prepare the cumulative frequency $(c.f.)$ column and find the value of $\frac{N}{2}$

(b) Find the class which contain the value of $c.f.$ is equal or just greater to $\frac{N}{2}$. This is median class.

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

Where ℓ — lower limit of median class.

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

f — frequency of median class.

h — Class interval of median class.

Illustration 11:

Find the median of following frequency distribution.

Class	0–5	5–10	10–15	15–20	20–25
f_i	6	10	8	9	7

Solution:

Class	f_i	$C.f.$
0–5	6	6
5–10	10	16
10–15	8	24
15–20	9	33
20–25	7	40

Here $\frac{N}{2} = \frac{40}{2} = 20$

Median Class is 10– 15

$\therefore$ Median $= 10 + \frac{(20-16)}{8} \times 5$

$= 10 + 2.5 = 12.5$

Illustration 12:

Find the median of the following frequency dist.

Class	1–10	11–20	21–30	31–40	41–50
f_i	5	6	8	7	4

Solution:

Class	f_i	$C.f.$
0.5–10.5	5	5
10.5–20.5	6	11
20.5–30.5	8	19
30.5–40.5	7	26
40.5–50.5	4	30

Here $\frac{N}{2} = \frac{30}{2} = 15$

Median class is 20.5 – 30.5

$\therefore$ Median $= 20.5 + \frac{(15-11)}{8} \times 10$

$= 20.5 + 5 = 25.5$

Illustration 13:

Set of 100 observation is divided into equal class interval whose median is 25. If median class is 20-30 and 45 observation is below from 20 then $'f'$ is:

(A) 10 (B) 20 (C) 30 (D) 15

Ans. (A)

Solution:

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

$$25 = 20 + \frac{\frac{100}{2} - 45}{f} \times 10$$

Solving, $f = 10$

Mode:

Mode is the value of variate which have maximum frequency.

(i) for ungrouped distribution

The value of variate which repeated maximum number of times.

(ii) for ungrouped frequency distribution.

The value of variate which have maximum frequency.

(iii) for grouped frequency distribution.

First, we find the class which have maximum frequency. This is modal class, then

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

Where ℓ — lower limit of modal class.

f_0 — frequency of modal class.

f_1 — frequency of the class preceding modal class.

f_2 — frequency of the class succeeding modal class.

h — class interval of modal class.

Illustration 14:

Find the mode of the following frequency distribution.

x_i	4	5	6	7	8	9
f_i	3	4	5	8	7	6

Solution:

Here the variate 7 have maximum frequency.

$\therefore$ Mode = 7

Illustration 15:

Find the mode of following frequency distribution.

Class	0–10	10–20	20–30	30–40	40–50
f_i	8	30	40	10	12

Solution:

Here model class is 20– 30 (which have maximum frequency)

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

$= 20 + \frac{(40-30)}{(2\times 40-30-10)} \times 10$

= 20 + 2.5

= 22.5

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 empirical formula.

Mode = 3 Median – 2 Mean

Measures of Dispersion:

What is Dispersion?

Dispersion refers to how "spread out" a group of scores is. To see what we mean by spread out, consider table. These table represent the scores on two quizzes. The mean score for each quiz is 7.0. Despite the equality of means, you can see that the distributions are quite different. Specifically, the scores on Quiz 1 are more densely packed and those on Quiz 2 are more spread out. The differences among students were much greater on Quiz 2 than on Quiz 1.

Quiz 1	2	6	5	4	3		
Quiz 2	2	4	3	3	2	3	3

The terms dispersion, spread, and variability are synonyms, and refer to how spread out a distribution is. Just as in the section on central tendency where we discussed measures of the centre of a distribution of scores, in this chapter we will discuss measures of the dispersion of a distribution. There are four frequently used measures of dispersion: range, interquartile range, variance, and standard deviation.

The dispersion of a distribution is, the measure of deviation of variates from their average value.

Generally, the following measures of dispersion are used.

(1) Range

(2) Mean Deviation

(3) Variance and Standard Deviation

Range:

The range of a distribution is the difference of highest value (H) and least value (L) of the distribution.

The Range of the grouped distribution is the difference of upper limit of maximum class (H) and lower limit of minimum class (L). Therefore

Range = $H - L$

Coefficient of Range = $\frac{\text{Difference of extreme values}}{\text{Sum of extreme values}} = \frac{H-L}{H+L}$

Illustration 16:

Range of data 7, 8, 2, 1, 3, 13, 18 is

(A) 10 (B) 15 (C) 17 (D) 11

Ans. (C)

Solution:

Range of data = difference of extreme values = 18 − 1 = 17

Illustration 17:

Coefficient of range 5, 2, 3, 4, 6, 8, 10 is

(A) $\frac{2}{3}$ (B) $\frac{1}{3}$ (C) $\frac{3}{5}$ (D) $\frac{1}{2}$

Ans. (A)

Solution:

Coefficient of range = $\frac{\text{Difference of extreme values}}{\text{Sum of extreme values}} = \frac{10-2}{10+2} = \frac{2}{3}$

Mean Deviation (M.D.):

The mean of absolute value of deviation of each variate from any central value are called mean deviation about that central value.

Let A be any central value (Mean, Median, Mode) of a dist. then M.D. about A are defined as

$M.D. = \frac{\sum_{i=1}^{n}|x_i - A|}{n}$ (For ungrouped dist.)

$M.D. = \frac{\sum_{i=1}^{n} f_i|x_i - A|}{N}$ (For frequency dist.)

Illustration 18:

Find the $M.D.$ of the following numbers about their median

34, 38, 42, 44, 46, 48, 54, 55, 63, 70

Solution:

Here $n = 10$ (even)

Median (M) = $\frac{\left(\frac{n}{2}\right)\text{th term} + \left(\frac{n}{2}+1\right)\text{th term}}{2} = \frac{5^{th}\text{ term} + 6^{th}\text{ term}}{2}$

$= \frac{46+48}{2} = 47$

$\because \sum|x_i - M|$ = 13 + 9 + 5 + 3 + 1 + 1 + 7 + 8 + 16 + 23 = 86

$\therefore M.D. = \frac{\sum|x_i - M|}{n} = \frac{86}{10} = 8.6$

Illustration 19:

Find the $M.D.$ of the following frequency dist. about Median.

Class	0–10	10–20	20–30	30–40	40–50
f_i	2	8	10	4	6

Solution:

Class	f_i	$c.f.$	x_i	$f_i\|x_i - M\|$
$0-10$	2	2	5	$2\cdot 20 = 40$
$10-20$	8	10	15	$8\cdot 10 = 80$
$20-30$	10	20	25	$10\cdot 0 = 0$
$30-40$	4	24	35	$4\cdot 10 = 40$
$40-50$	6	30	45	$6\cdot 20 = 120$
	$N = 30$			$\sum f_i\|x_i - M\| = 280$

Here $\frac{N}{2} = 15$, Median class is $20-30$

$$\therefore \text{ Median } = \ell + \frac{\left(\frac{N}{2} - F\right)}{f} \times h = 20 + \frac{(15-10)}{10} \times 10 = 25$$

$$\therefore M.D.\text{ about Median} = \frac{\sum f_i |x_i - M|}{N} = \frac{280}{30} = 9.33$$

Illustration 20:

If mean deviation of number $1, 1+d, 1+2d, \ldots.. 1+50\,d$ from their mean is 325, then value of $'d'$ is :

Solution:

$x_i \to 1, 1+d, 1+2d, \ldots.. 1+50\,d$

Mean = $1 + 25\,d$

$$M.D. = \frac{\sum f_i |x_i - (1+25d)|}{n}$$

$$325 = \frac{25d + 24d + \ldots\ldots + d + 0 + d + \ldots.25d}{51}$$

$$325 = \frac{2d[1+2+\ldots\ldots+25]}{51}$$

$$\frac{325 \times 51}{2} = d \cdot \frac{25(26)}{2}$$

$$d = \frac{51}{2}$$

$d = 25.5$

Illustration 21:

If mean deviation of number $a, 2a, 3a, \ldots\ldots, 50a$ from median is 50, then $|a|$ equals:

Solution:

$x_i \to a, 2a, 3a, \ldots\ldots.50a$

Median = $25.5a$

$$\text{M.D.} = \frac{\sum f_i |x_i - 25.5a|}{n}$$

$$50 = \frac{24.5a + 23.5a + \ldots\ldots + 0.5a + \ldots..24.5a}{50}$$

$$2500=\frac{2a}{10}(5+15+25+.....+245)$$

$$2500=\frac{2a}{10}\times\frac{25}{2}(5+245)$$

$a = 4$

$\therefore\ |a| = 4$

Variance and Standard Deviation:

Variance:

The Mean of square of deviation of each variate from their mean are called variance, it is denoted by σ^2.

Standard Deviation (S.D.):

The positive square root of the variance are called standard deviation. It is denoted by σ.

i.e., = $\boxed{\text{S.D.}=+\sqrt{\text{variance}}}$

Formulae for Variance:

(i) For ungrouped distribution

$$\sigma^2=\frac{\sum(x_i-\bar{x})^2}{n}$$

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

$$\sigma^2=\frac{\sum d_i^2}{n}-\left(\frac{\sum d_i}{n}\right)^2 \qquad \text{Where } d_i=x_i-a$$

$$\sigma^2=h^2\left[\frac{\sum u_i^2}{n}-\left(\frac{\sum u_i}{n}\right)^2\right] \qquad \text{Where } u_i=\frac{d_i}{h}$$

Illustration 22:

Find variance of first $'n'$ even natural number is:

Solution:

$x_i \rightarrow 2,4,6,.....,2n$

$$=\frac{\sum x_i}{n}=\frac{n(n+1)}{n}=n+1$$

$$\text{Variance }(x_i)=\frac{\sum x_i^2}{n}-(\bar{x})^2$$

$$=\frac{2^2+4^2+....(2n)^2}{n}-(n+1)^2$$

$$=4\times\frac{n(n+1)(2n+1)}{6n}-(n+1)^2$$

$$=\frac{4}{6}(n+1)(2n+1)-(n+1)^2=\frac{n^2-1}{3}$$

Illustration 23:

The sum of 100 observation and sum of there square are 400 and 2475 respectively. Later on three observation 3, 4, 5 are found incorrect. If these incorrect data is omitted then variance of remaining observation is :

(A) 8.00 (B) 8.25 (C) 9.00 (D) 9.25

Ans. (C)

Solution:

New $\sum x_i = 400 - (3 + 4 + 5)$

$= 400 - 12 = 388$

New $\sum x_i^2 = 2475 - (3^2 + 4^2 + 5^2)$

$= 2475 - 50 = 2425$

So var $(x_i) = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2$

$= \frac{2425}{97} - \left(\frac{388}{97}\right)^2 = 25 - 16 = 9$

Illustration 24:

If mean and variance of data $a, b, 8, 5, 10$ is 6 and 6.8 respectively. Then value of and b is :

Solution:

Mean = 6

$\frac{a+b+8+5+10}{5} = 6$

$a + b = 7$ ...(i)

and var $(x_i) = 6.8$

$\frac{a^2+b^2+8^2+5^2+10^2}{5} - (6)^2 = 6.8$

$a^2 + b^2 = 25$ ...(ii)

from (i) and (ii)

$\begin{cases} a = 3, b = 4 \\ a = 4, b = 3 \end{cases}$

Illustration 25:

If $S.D.$ of number 2, 3, a and 11 is 3.5 then which of them is true:

(A) $3a^2 - 34a + 91 = 0$ (B) $3a^2 - 23a + 44 = 0$

(C) $3a^2 - 26a + 55 = 0$ (D) $3a^2 - 32a + 84 = 0$

Ans. (D)

Solution:

$\text{Var}(x_i) = \frac{\sum x_i^2}{n} - (\bar{x})^2$

$(3.5)^2 = \frac{2^2+3^2+a^2+11^2}{4} - \left(\frac{2+3+a+11}{4}\right)^2$

Solving $3a^2 - 32a + 84 = 0$

Variance (For Frequency Distribution)

Formulae for Variance:

(ii) For frequency distribution:

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

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

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

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

Note: (i) $\sigma_x^2 = \sigma_d^2 = h^2\sigma_u^2 = \sigma^2$

(ii) The variance of first n natural numbers is $\boxed{\sigma^2 = \frac{n^2-1}{12}}$

(iii) The variance of A.P. $a, a+d, a+2d, \ldots\ldots\ldots\ldots, a+2nd$ is $\boxed{\sigma^2 = \frac{n(n+1)}{3}d^2}$

Properties of Variance:

(i) If the variance of $x_1, x_2, \ldots\ldots\ldots, x_n$ is σ^2 then
var. of $x_1 \pm \lambda, x_2 \pm \lambda, \ldots\ldots\ldots\ldots, x_n\lambda$ is σ^2
var. of $\lambda x_1, \lambda x_2, \ldots\ldots\ldots\ldots, \lambda x_n$ is $\lambda^2\sigma^2$
var. of $ax_1 + b, ax_2 b, \ldots\ldots\ldots\ldots, ax_n b$ is $a^2\sigma^2$ (where $\lambda, a, b,$ are constant)
i.e. the variance and $S.D.$ are not depend on change of origin (add & subtract) but they are depend on change of scale (Multiplication & Division).

(ii) If are mean and σ_1^2, σ_2^2 are variance of two series which have n_1, n_2 terms resp. and is their combined mean then their combined variance σ^2 is

$$\sigma^2 = \frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}$$

where $d_1 = \bar{x}_1 - \bar{x} = \frac{n_2(\bar{x}_1 - x_2)}{n_1 + n_2}, d_2 = \bar{x}_2 - \bar{x} = \frac{n_1(\bar{x}_2 - \bar{x}_1)}{n_1 + n_2}$

$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2} + \frac{n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}$$

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

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

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

Illustration 26:

$S.D.$ of distribution $41, 42, 43, \ldots.55$ is

Solution:

Variance $(41, 42, 43, \ldots.55)$ = Variance $(1, 2, \ldots.15)$

= $\frac{15^2 - 1}{12}$

$= \frac{225-1}{12} = \frac{224}{12} = \frac{112}{6} = \frac{56}{3}$

$\therefore\ S.D. = \sqrt{\frac{56}{3}}$

Illustration 27:

If $\sum_{i=1}^{10}(x_1 - 10) = 12$ and $\sum_{i=1}^{10}(x_1 - 10)^2 = 18$, then $S.D.$ is

(A) $\frac{2}{5}$ (B) $\frac{3}{5}$ (C) $\frac{4}{5}$ (D) $\frac{6}{5}$

Ans. (B)

Solution:

$\Sigma(x_i - 10) = 12,\ \Sigma(x_i - 10)^2 = 18$

$\Sigma d_i = 12 \qquad \Sigma d_i^2 = 18$

Variance $(d_i) = \frac{\Sigma d_i^2}{n} - \left(\frac{\Sigma d_i}{n}\right)^2 = \frac{18}{10} - \left(\frac{12}{10}\right)^2$

$= \frac{180-144}{100} = \frac{36}{100} = \frac{9}{25}$ $\qquad \therefore\ S.D. = \frac{3}{5}$

Illustration 28:

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
		$N = 50$		$\Sigma f_i u_i = -21$	$\Sigma f_i u_i^2 = 71$

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

Miscellaneous Examples:

Illustration 29:

The mean of 2 samples of sizes 50 & 40 were found to be 63 and 54. Their variance were 81 & 36. Find the variance of combined sample of size 90

(A) 9 (B) 81 (C) 3 (D) 243

Ans. (B)

Solution:

$\sum x_i = 63\times 50 = 3150\,;\ \sum y_i = 40\times 54 = 2160$

var(x_i) = 81 = $\frac{\sum x_i^{\ 2}}{50} - (63)^2$, var(y_i) = 36 = $\frac{\sum y_i^{\ 2}}{40} - (54)^2$

$\sum x_i^{\ 2} = 202500$

$\sum y_i^{\ 2} = 118080$

Combined variance = $\frac{\left(\sum x_i^{\ 2} + \sum y_i^{\ 2}\right)}{90} - \left(\frac{\sum x_i + \sum y_i}{90}\right)^2$

= $\frac{320580}{90} - (59)^2$

= 3562 – 3481

= 81

Illustration 30:

The mean and variance of 7 observation are 7 and $\frac{100}{7}$. If 5 of the observation are 2, 4, 7, 11, 10, find the remaining 2 observations.

(A) 3, 6 (B) 3, 12 (C) 4, 11 (D) 5, 10

Ans. (B)

Solution:

$\frac{\sum x_i}{7} = 7$

$2 + 4 + 7 + 11 + 10 + a + b = 49 \Rightarrow a + b = 15$

$\frac{\Sigma(x_i - \bar{x})^2}{7} = \frac{100}{7}$

$\Rightarrow 25 + 9 + 0 + 16 + 9 + (7 - a)^2 + (7 - b)^2 = 100$

$\Rightarrow (7 - a)^2 + (7 - b)^2 = 41 \Rightarrow a = 3$

$b = 12$

Illustration 31:

If $\sum_{i=1}^{11}(x_i - 4) = 11$ and $\sum_{i=1}^{11}(x_i - 4)^2 = 44$ then find variance of $x_1, x_2, x_3 \ldots\ldots x_{11}$.

(A) 4 (B) 3 (C) 7 (D) 11

Ans. (B)

Solution:

$var(x_i - 4) = \frac{44}{11} - \left(\frac{11}{11}\right)^2 = 3$

var(x_i) = var$(x_i - 4)$ = 3

Illustration 32:

If difference between mean and mode is 3, the difference between mean and median is

(A) 3 (B) 1 (C) 4 (D) 2

Ans. (B)

Solution:

Mode + 2 mean = 3 median

(mean –3) + 2 mean = 3 median

3(mean – median) = 3

Illustration 33:

The mean and variance of 10 numbers were calculated as 11.3 and 3.3 respectively. It was subsequently found that one of the a number was misread as 10 instead of 12. How does the variance change?

(A) variance decreases (B) variance increases

(C) nothing can be said about variance (D) variance remains unchanged

Ans. (A)

Solution:

Let x_n misread value $(x_n) = 10 (x_n)_{\text{actual}} = 12$

$\sigma^2 = 3.3\ \bar{x} = 11.3 \Rightarrow \sum_{i=1}^{n-1} x_i = 113 - 10 = 103 = 10.\left(\bar{x}\right) - 10$

$$\sigma^2 = \frac{\sum_{i=1}^{n-1} x_i^2 + x_n^2}{10} - \left(\bar{x}\right)^2$$

$$\sum_{i=1}^{n-1} x_i^2 = -67 + 10\left(\bar{x}\right)^2 \qquad \text{...(i)}$$

$$\Rightarrow (\sigma^2)_{\text{actual}} = \frac{\sum_{i=1}^{n-1} x_i^2 + \left(x_n\right)_{\text{actual}}^2}{10} - \left(\bar{x}\right)_{\text{actual}}^2 = \frac{-67 + 10\left(\bar{x}\right)^2 + 144}{10} - \left(\frac{10\left(\bar{x}\right) - 10 + 12}{10}\right)^2$$

$= (\sigma^2)_{\text{actual}} = 3.14$