Standard deviation

(i) For the ungrouped data:

$$Variance = \sigma^2 = \frac{\sum(x_i-\bar{x})^2}{N} = \frac{1}{N}\sum x_i^2 - \left(\frac{1}{N}\sum x_i\right)^2 = \frac{1}{N}\sum x_i^2 - \bar{x}^2$$

(ii) Discrete or Continuous frequency distribution

$$\sigma^2 = \frac{1}{N}\sum_i f_i(x_i-\bar{x})^2$$

$$= \frac{1}{N}\sum f_i x_i^2 - \left(\frac{1}{N}\sum f_i x_i\right)^2 \; or \; \frac{1}{N}\sum f_i d_i^2 - \left(\frac{1}{N}\sum f_i d_i\right)^2 \quad h^2\left[\frac{1}{N}\sum f_i\, d_i{}^2 - \left(\frac{1}{N}\sum f_i d_i\right)^2\right]$$

(Discrete case) (Continuous case)

$$Standard\ deviation = \sqrt{variance} \quad \sigma = \sqrt{\frac{1}{N}\sum (x_i-\bar{x})^2} \quad or \quad \sqrt{\frac{1}{N}\sum f_i(x_i-\bar{x})^2}$$

$\bar{x}$ – Arithmetic mean of the distribution

Recall:

$\bar{x} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i$ (mean for discrete), $\bar{x} = A + \frac{h}{N}\sum_{i=1}^{n} f_i d_i$, (mean for continuous)

(2) Relations between measures of dispersion

(i) *Quartile deviation = 2 / 3 (standard deviation)*

(ii) *Mean deviation = 4 / 5 (standard deviation)*

Coefficient of Variation

$$\text{Coefficient of Variation : C.V.} = \frac{\sigma}{\bar{x}} \times 100 \text{ (Relative Measure)}$$

For comparing the variability of two series, we calculate the co-efficient of variations for each series. The series having greater C.V. is said to be more variable than the other and the series having lesser C.V. is said to be more consistent (or homogendus) than the other.

3) The score of two players A and B in ten innings during a certain season are:

A	32	28	47	63	71	39	10	60	96	14
B	19	31	48	53	67	90	10	62	40	80

Find which of the two players A, B is more consistent in scoring.

Solution: Calculation of Coefficient of Variation

X	$(X-\bar{X})$	$(X-\bar{X})^2$
32	-14	196
28	-18	324
47	+1	1
63	+17	289
71	+25	625
39	-7	49
10	-36	1296
60	+14	196
96	+50	2500
14	-32	1024
$\sum X = 460$	0	6500

Y	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$
19	-31	961
31	-19	361
48	-2	4
53	+3	9
67	+17	289
90	+40	1600
10	-40	1600
62	+12	144
40	-10	100
80	+30	900
$\sum Y = 500$	0	5968

$$\bar{X} = \frac{460}{10} = 46$$

$$\sigma_A^2 = \frac{\sum(x_i-\bar{x})^2}{N} = \frac{6500}{10} = 650$$

$$\bar{Y} = \frac{500}{10} = 50$$

$$\sigma_B^2 = \frac{\sum(y_i-\bar{y})^2}{N} = \frac{5968}{10} = 596.8$$

$$\sigma_A = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}} = 25.5 \qquad \sigma_B = \sqrt{\frac{\sum(y_i - \bar{y})^2}{N}} = 24.43$$

$$\mathrm{C.V.}_{(A)} = \frac{\sigma_A}{\bar{x}} \text{ x } 100 = 55.43 \qquad \mathrm{C.V.}_{(B)} = \frac{\sigma_B}{\bar{y}} \text{ x } 100 = 48.86$$

Example:

Calculate the mean and standard deviation for the following table giving the age distribution of 542 members.

Age in years :	20—30	30—40	40—50	50—60	60—70	70—80	80—90
No. of members :	3	61	132	153	140	51	2

Solution: Also find CV.

Here we take $d = \frac{x - A}{h} = \frac{x - 55}{10}$

Age group	*Mid–value* (x)	*Frequency* (f)	$d = \frac{x-55}{10}$	fd	fd^2
20 — 30	25	3	–3	–9	27
30 — 40	35	61	–2	–122	244
40 — 50	45	132	–1	–132	132
50 — 60	55	153	0	0	0
60 — 70	65	140	1	140	140
70 — 80	75	51	2	102	204
80 — 90	85	2	3	6	18
		$N = \Sigma f = 542$		$\Sigma fd = -15$	$\Sigma fd^2 = 765$

$$\bar{x} = A + h\frac{\Sigma fd}{N} = 55 + \frac{10 \times (-15)}{542} = 55 - 0{\cdot}28 = 54{\cdot}72 \text{ years.}$$

$$\sigma^2 = h^2\left[\frac{1}{N}\Sigma fd^2 - \left(\frac{1}{N}\Sigma fd\right)^2\right] = 100\left[\frac{765}{542} - (0{\cdot}28)^2\right]$$

$= 100 \times 1{\cdot}333 = 133{\cdot}3$

σ (standard deviation) $= 11{\cdot}55$ years

$$CV = 100 \times \frac{\sigma}{\bar{x}}$$