This rule is applied to any distribution (Population or Sample) and guarantees the inclusion of minimum fraction of the data in the constructed interval whereas the actual fraction of the included (especially in bell-shaped distributions) will exceed $\left(1-\frac{1}{k^2}\right)$.

4.5.3 Co-efficient of Variation. The variability of two or more than two sets of data cannot be compared unless we have a relative measure of dispersion. For this purpose, Karl Pearson (1857–1936) introduced a relative measure of variation, known as the *co-efficient of variation*, abbreviated C.V. which expresses the standard deviation as a percentage of the arithmetic mean of a data set. Symbolically, it is defined as

$$\text{C.V.} = \frac{S}{\bar{x}} \times 100, \text{ for sample data,}$$

$$= \frac{\sigma}{\mu} \times 100, \text{ for population data.}$$

As the coefficient of variation is a pure number without units, it is therefore used to compare the variation in two or more data sets or distributions that are measured in different units, e.g. one may be measured in hours and the other in kilograms or rupees. A large value of C.V. indicates that the variability is great and a small value of C.V. indicates less variability.

The coefficient of variation is also used to compare the performance of two candidates or of two players given their scores in various papers or games, the smaller the coefficient of variation the more consistent is the performance of the candidates or players. Thus it is used as a criterion for the consistent performance of the candidates or the players. It should be noted that this co-efficient is unreliable when the arithmetic mean is very small.

Example 4.9 Using the co-efficient of variation, determine whether or not there is greater variation among the prices of certain similar commodities given, than among the life in hours under test.

Prince in Rupees: 8, 13, 18, 23, 30

Life in hours: 130, 150, 180, 250, 345

We have to compute the mean and the standard deviation for each set so that the corresponding coefficient of variation can be obtained. The necessary arithmetic is shown below:

Price in Rupees (X)		Life in hours (Y)	
X	X^2	Y	Y^2
8	64	130	16900
13	169	150	22500
18	324	180	32400
23	529	250	62500
30	900	345	119025
92	1986	1055	253325

Price of Commodities

$\bar{X} = \text{Rs.}\dfrac{92}{5} = \text{Rs.}18.4$

$s_X = \sqrt{\dfrac{1986}{5} - \left(\dfrac{92}{5}\right)^2}$

$= \sqrt{397.2 - 338.56}$

$= \sqrt{58.44} = \text{Rs. } 7.66$

$\therefore \ \text{C.V.} = \dfrac{7.66}{18.4} \text{x} 100 = 41.63\%$

Life in Hours

$\bar{Y} = \dfrac{1055}{5} = 211 \text{ hours}$

$s_Y = \sqrt{\dfrac{253325}{5} - \left(\dfrac{1055}{5}\right)^2}$

$= \sqrt{50665 - 44521}$

$= \sqrt{6144} = 78.38 \text{ hours}$

$\therefore \ \text{C.V.} = \dfrac{78.38}{211} \text{x} 100 = 37.15\%$

We see that the co-efficient of variation for the prices of commodities (X) is larger than that for the life in hours (Y). Hence the prices of certain similar commodities are showing greater variation than that among the life in hours under test.

Example 4.10 Goals scored by two teams A and B in a football season were as follows:

No. of goals scored in a match (x_i)	Number of matches	
	A	B
0	27	17
1	9	9
2	8	6
3	5	5
4	4	3

By calculating the co-efficient of variation in each case, find which team may be considered more consistent.

The necessary arithmetic is shown below:

No. of goals	Team A			Team B		
(x_i)	f_i	$f_i x_i$	$f_i x_i^2$	f_j	$f_j x_i$	$f_j x_i^2$
0	27	0	0	17	0	0
1	9	9	9	9	9	9
2	8	16	32	6	12	24
3	5	15	45	5	15	45
4	4	16	64	3	12	48
Total	53	56	150	40	68	126

Team A:

$$\text{Mean} = \frac{\Sigma f_i x_i}{n} = \frac{56}{53} = 1.06, \text{ and}$$

$$s = \sqrt{\frac{\Sigma f_i x_i^2}{n} - \left(\frac{\Sigma f_i x_i}{n}\right)^2}$$

$$= \sqrt{\frac{150}{53} - \left(\frac{56}{53}\right)^2} = \sqrt{1.7138} = 1.308$$

$$\therefore \quad \text{C.V.} = \frac{s}{\bar{x}} \text{x} 100 = \frac{1.308}{1.06} \text{x} 100 = 123.4\%$$

Team B:

$$\text{Mean} = \frac{\Sigma f_j x_i}{n} = \frac{48}{40} = 1.20, \text{ and}$$

$$s = \sqrt{\frac{\Sigma f_j x_i^2}{n} - \left(\frac{\Sigma f_j x_i}{n}\right)^2}$$

$$= \sqrt{\frac{126}{40} - \left(\frac{48}{40}\right)^2} = \sqrt{1.71} = 1.308$$

Thus $\text{C.V.} = \frac{s}{\bar{x}} \text{x} 100 = \frac{1.308}{1.20} \text{x} 100 = 109.0\%$

We see that the co-efficient of variation for the team B is smaller than that for the team A. Hence team B is more consistent than team A.

4.5.4 Properties of Variance and Standard Deviation. The variance and standard deviation have the following useful and interesting properties:

i) The variance of a constant is equal to zero. If a is any constant, then

$$\text{Var}(a) = \frac{1}{N}\Sigma[a-a]^2 \quad (\because \text{mean of a constant is constant itself})$$

$$= 0$$

ii) The variance is independent of the origin, *i.e.* it remains unchanged when a constant is added to or subtracted from each observation of the variable X. Symbolically,

$$\text{Var}(X + a) = \text{Var}(X)$$

Now $\text{Var}(X+a) = \frac{1}{N}\Sigma[(x_i + a) - (\mu + a)]^2 \quad (\because \frac{\Sigma(x_i + a)}{N} = \mu + a)$