Statistics

1. **Measures of Dispersion**
The dispersion is the measure of variations in the values of the variable. It measures the degree of scatterdness of the observation in a distribution around the central value.

2. **Range**
Range is defined as the difference between two extreme observations of the distribution.
Range of distribution = Largest observation − Smallest observation

3. **Mean Deviation**
Mean deviation is defined as the arithmetic mean of the absolute deviations of all the values taken about any central value 'a' (mean or median).

$$\therefore \text{MD} = \frac{\text{Sum of absolute values of deviations from 'a'}}{\text{Number of observations}}$$

(i) **Mean deviation for ungrouped data** Let n observations be $x_1, x_2, x_3, \dots, x_n$, then mean deviation about their mean or median is given by

$$\text{MD} = \frac{\Sigma |x_i - A|}{n}$$

where, A = mean or median

(ii) **Mean deviation for discrete frequency distribution** Let the given data consist of discrete observations $x_1, x_2, x_3, \dots, x_n$ occurring with frequencies $f_1, f_2, f_3, \dots, f_n$ respectively, then

$$\text{MD} = \frac{\Sigma f_i |x_i - A|}{\Sigma f_i} = \frac{\Sigma f_i |x_i - A|}{N}$$

where, A = mean or median

(iii) **Mean deviation for continuous frequency distribution** Here, the procedure is same as for a discrete frequency distribution. The only difference is that here we have to obtain the mid-points of the various classes and take the deviations of these mid-points from the given central value.

Note Median $= l + \dfrac{\frac{N}{2} - cf}{f} \times h$

where, l = lower limit, f = frequency

h = width of median class

and cf = cumulative frequency of class just preceding the median class.

4. Variance

Variance is the arithmetic mean of the square of the deviations about mean $\bar{x}$. Let $x_1, x_2, \ldots, x_n$ be n observations with $\bar{x}$ as the mean, then the variance denoted by σ^2, is given by $\sigma^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n}$.

5. Standard Deviation

If σ^2 is the variance, then σ is called the standard deviation which is given by $\sigma = \sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n}}$.

Thus, standard deviation (SD) $= \sqrt{\text{Variance}}$

(i) **Standard deviation for ungrouped data** SD of n observations $x_1, x_2, \ldots, x_n$ is given by

$$\sigma = \sqrt{\frac{\Sigma(x_1 - \bar{x})^2}{n}}.$$

(ii) **Standard deviation of a discrete frequency distribution** Let the discrete frequency distribution be $x_i : x_1, x_2, \ldots, x_n$ and $f_i : f_1, f_2, \ldots, f_n$, then

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

or $$\sigma = \frac{1}{N}\sqrt{N\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2}$$

where, f_i's are the frequency of x_i's and $N = \sum_{i=1}^{n} f_i$.

Also, by shortcut method,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{\Sigma f_i d_i}{N}\right)^2}$$

where, $d_i = x_i - a$, a = assumed mean

Standard Deviation of a continuous frequency distribution

If there is a frequency distribution of n classes and each class defined by its mid-point x_i, with corresponding frequency f_i, then $\sigma = \dfrac{1}{N}\sqrt{N\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2}$

6. Coefficient of Variation

The measures of variability which is independent of units, is called coefficient of variation denoted as CV and defined as

$$\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Mean}} \times 100$$

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

Note *If two series have equal means, then the series with greater SD or variance is said to be more variable than the other and the series with lesser value of SD is said to be more consistent than the other.*