Measures of Data Dispersion

It describes the spread or variability of data in a dataset They provide insights about how much the data values differ from each other.

Dispersion can't be negative.

1. **Variance**

Variance measures the dispersion of a set of data points around their mean values.

- **Variance for Population data:**
  Variance for population data is denoted as - σ^2 (Sigma square) =
$$\sigma^2 = \frac{I}{N}\sum_{i=1}^{N}(xi - \mu)^2$$
  It is sum of squared differences between observed values and the population mean divided by the total number of observations.
- **Variance for Sample Data:**
  Variance for sample data is denoted by - s^2
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$
  It is sum of squared differences between observed sample values and sample mean divided by number of sample observations minus 1.

2. **Standard Deviation**

This is the most common measure of variability. The square root of variance is known as Standard deviation.

- Sample Standard Deviation (s^2)= $\sqrt{s^2}$
- Population Standard Deviation (σ^2)= $\sqrt{\sigma^2}$

3. **Coefficient of variation**

Expresses as the standard deviation relative to mean. It allows for meaningful comparison of variability across datasets with different units or scales, which would not be possible with standard deviation.

- Population coefficient of variation $= \frac{\boldsymbol{\sigma}}{\boldsymbol{\mu}}$
- Sample coefficient of variation $= \frac{\boldsymbol{s}}{\bar{\boldsymbol{x}}}$

B. Measures of working with more than one variable

1. **Covariance –**
   Gives the sense of direction.
   - Sample-Covariance =

$$\boldsymbol{S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}}$$

   - Population Covariance

$$\boldsymbol{\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}}$$

   - If covariance is having +ve sign (>0) then it represents that two variables move together.
   - If covariance is having _ve sign (<0) then it represents that two variables move opposite direction.
   - If covariance is equal to 0 then it represents that two variables are independent.

2. **Correlation**
   Adjusts covariance so that the relationship between the two variables becomes easy and intuitive to interpret.

   Sample Correlation:

$$r = \frac{cov(X,Y)}{s_x s_y}$$

   Population Correlation:

$$p = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

   - Correlation of 1 is known as Perfect Positive Correlation entire variability of one variable is explained by the other variable.
   - Correlation of 0 between two variable represents Absolutely Independent Variable.
   - Negative Correlation also known as Perfect negative Correlation.