Covariance

Variance: Variance helps us understand how far our random variable is spread out from the mean, for example, the income of the people may have a high variance as some people may have high income levels.

The formula for variance for a sample is given by:

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

where n is the number of samples (e.g. the number of people) and $\bar{x}$ is the mean of the random variable x (mean of the income).

Covariance: It measures how much two random variables vary together. e.g. The income of a person and the expenses of that person in a population. More precisely, covariance refers to the measure of how two random variables in a data set will change together. A positive covariance means that the two variables at hand are positively related, and they move in the same direction. A negative covariance means that the variables are inversely related, or that they move in opposite directions.

The formula for covariance is given by:

$$\sigma(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

where n is the number of samples (e.g. the number of people) and $\bar{x}$ is the mean of the random variable x (represented as a vector).

The variance σ^2x of a random variable x can be also expressed as the covariance with itself by σ(x,x).

Covariance Matrix: Following from the previous equations, the covariance matrix for the two dimensions is given by:

$$C = \begin{pmatrix} \sigma(x,x) & \sigma(x,y) \\ \sigma(y,x) & \sigma(y,y) \end{pmatrix}$$

In this matrix, the variances appear along the diagonal and covariances appear in the off-diagonal elements.

Note: You can use the function **numpy.cov** to get the covariance matrix in Python.