Covariance:

Covariance is a statistical measure that describes the degree to which two variables are linearly related. It measures how much two variables change together, such that when one variable increases, does the other variable also increase, or does it decrease?

If the covariance between two variables is positive, it means that the variables tend to move together in the same direction. If the covariance is negative, it means that the variables tend to move in opposite directions. A covariance of zero indicates that the variables are not linearly related.

Covariance Formula	
Population	**Sample**
$\sigma_{xy} = \frac{\sum(X - \mu_x)(Y - \mu_y)}{N}$	$s_{xy} = \frac{\sum(X - \bar{x})(Y - \bar{y})}{n - 1}$
X, Y – The Value of X and Y in the Population μ_x, μ_y – The population Mean of X and Y N – Total Number of Observations	X, Y – The Value of X and Y in the Sample Data $\bar{x}, \bar{y}$ – The Sample Mean of X and Y n - Total Number of Observations

Example:

Experience (X)	Salary (Y)	X - $\overline{X}$	Y - $\overline{Y}$	(X - $\overline{X}$) $\times$ (Y - $\overline{Y}$)
2	1			
5	2			
8	5			
12	12			
13	10			

Disadvantage of using Covariance

One limitation of covariance is that it does not tell us about the strength of the relationship between two variables, since the magnitude of covariance is affected by the scale of the variables.

Correlation:

Correlation refers to a statistical relationship between two or more variables. Specifically, it measures the degree to which two variables are related and how they tend to change together.

Correlation is often measured using a statistical tool called the correlation coefficient, which ranges from -1 to 1. A correlation coefficient of -1 indicates a perfect negative correlation, a correlation coefficient of 0 indicates no correlation, and a correlation coefficient of 1 indicates a perfect positive Correlation.

$$r = \frac{cov(x, y)}{\sigma_x \cdot \sigma_y}$$

Derived :

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

r = Pearson correlation coefficient

x_i = x variable sample

y_i = y variable sample

$\bar{x}$ = mean of values in x variable

$\bar{y}$ = mean of values in y variable