# Other Terms

## Bias :

In simple terms bias is the bias in the data. i.e., data is skewed. For example if I want to measure average sugar level from population of 1000 people using sample of 100, then I must pick sample randomly. If I pick sample of young and healthy people only then the average sugar level will be low, so this is not the correct representation of population. So this is called bias.

## Covariance :

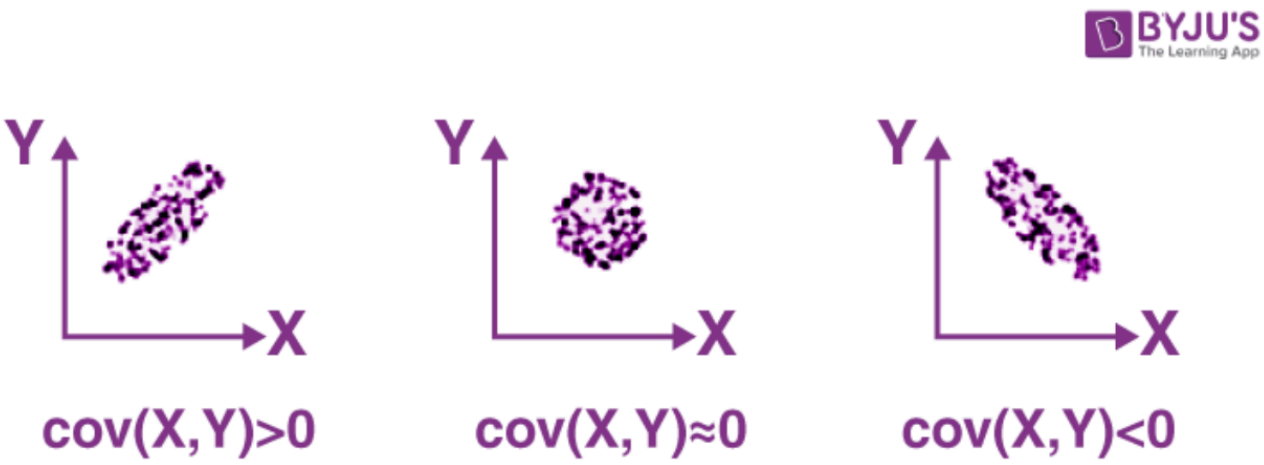
It is a measure of the relationship between two random variables and to what extent they are changed together. Or in simple words it defines the direction of change of one variable if other changes and to which extent the change will be. It is measured in units, the unit of covariance is the multiplication of units of two variables. There are three types of covariance - positive, negative and zero.

1. Positive : Variables move in same direction.
2. Negative : Variables move in opposite direction.
3. Zero : There is no relationship between variables.

It is given by formula,

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

Where $\bar{x}$ is mean of x and $\bar{y}$ is mean of y.

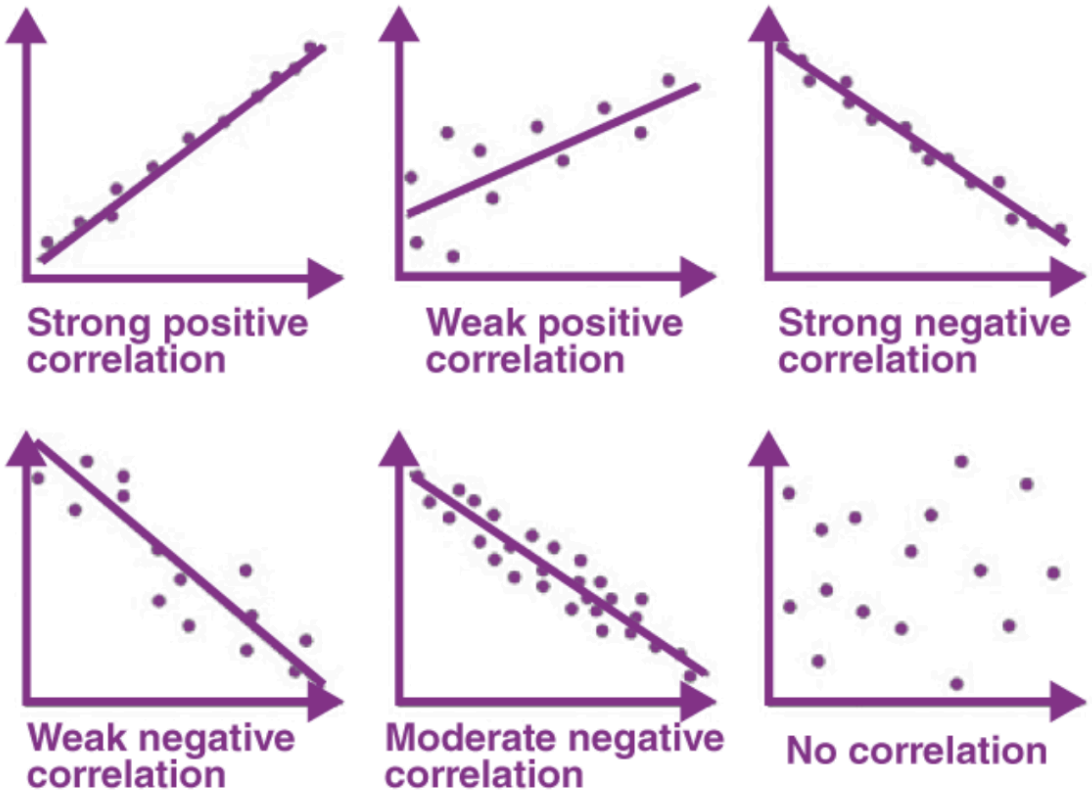


# Correlation :

We know that covariance is the evaluation of changes between any two variables. Where as correlation estimates the depth of relationship of variables. It tells how strong is the relationship between two variables.

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \cdot \sigma_y}$$

Where $\rho$ is correlation and $\sigma$ is standard deviation.




# Correlation VS Covariance :

| Correlation | Covariance |
|---|---|
| It shows how strongly one variable is related with other variable. | It shows to which extent variables change with each other. |
| It is between -1 and 1. | It is between -∞ and ∞. |
| It is unit less. | It has unit of multiplication of variable. |
| It measures direction and strength of the **linear** relationship between two variables. | It indicates the direction of the **linear** relationship between the given two variables. |

# Causation :

Causation is the statistical measure of occurrence of once event as the cause of occurrence of another event. It is different from correlation, correlation is the relation between two variables, where as causation is the cause of occurrence of one event due to another event. Causation indicates causal relationship where as correlation indicates correlation relationship. Lets take an example.

Smoking is injurious to health, it **causes** lung cancer. - Causation
Smoking is **correlated** with drinking alcohol. - Correlation

In first statement smoking is cause of lung cancer where as in second statement smoking is correlation of alcoholism. Smoking doesn't cause alcoholism, but it correlates alcoholism.

# Percentile :

Percentile is the value between 0 to 100. It defines how many percent values are less than or equal to given value. Lets say salary at 75 percentile is 80000, that means that are 75% salaries which are less than or equal to 80000.

Quantile is the number of percentile. I.e., value at 25 percentile is called 1st quintile, 50 percentile is called 2nd quantile and so on.