02.05.2020

B.COM PROG SEMESTER II-
BUSINESS MATHEMATICS AND STATISTICS

SIMPLE LINEAR CORRELATION

COVARIANCE

Covariance is a measure of the relationship between two random variables. The metric evaluates how much – to what extent – the variables change together. In other words, it is essentially a measure of the variance between two variables. However, the metric does not assess the dependency between variables.

Formula for Covariance

The covariance formula is similar to the formula for correlation and deals with the calculation of data points from the average value in a dataset. For example, the covariance between two random variables X and Y can be calculated using the following formula (for population)

$$\text{Cov}(X, Y) = \frac{\sum(X_i - \bar{X})(Y_j - \bar{Y})}{n}$$

Where:
Xi – the values of the X-variable
Yj – the values of the Y-variable
$\bar{X}$ – the mean (average) of the X-variable
$\bar{Y}$ – the mean (average) of the Y-variable
n – the number of data points

Covariance vs. Correlation

Covariance and correlation both primarily assess the relationship between variables. The closest analogy to the relationship between them is the relationship between the variance and standard deviation.

Covariance measures the total variation of two random variables from their expected values. Using covariance, we can only gauge the direction of the relationship (whether the variables tend to move in tandem or show an inverse relationship). However, it does not indicate the strength of the relationship, nor the dependency between the variables.

On the other hand, correlation measures the strength of the relationship between variables. Correlation is the scaled measure of covariance. It is dimensionless. In other words, the correlation coefficient is always a pure value and not measured in any units.

The relationship between the two concepts can be expressed using the formula below:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Where:
$\rho(X,Y)$ – the correlation between the variables X and Y
$\mathrm{Cov}(X,Y)$ – the covariance between the variables X and Y
σX – the standard deviation of the X-variable
σY – the standard deviation of the Y-variable

Correlation Formula

Correlation shows the relation between two variables. Correlation coefficient shows the measure of correlation. To compare two datasets we use the correlation formulas.