04.05.2020

B.COM PROG SEMESTER II-
BUSINESS MATHEMATICS AND STATISTICS

Karl Pearson's Coefficient of Correlation

Definition: Karl Pearson's Coefficient of Correlation is widely used mathematical method wherein the numerical expression is used to calculate the degree and direction of the relationship between linear related variables.

Pearson's method, popularly known as a **Pearsonian Coefficient of Correlation,** is the most extensively used quantitative methods in practice. The coefficient of correlation is denoted by **"r".**

If the relationship between two variables X and Y is to be ascertained, then the following formula is used:

$$r = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2}\sqrt{(Y-\overline{Y})^2}}$$

Where, $\overline{X}$ = mean of X variable
$\overline{Y}$ = mean of Y variable

Properties of Coefficient of Correlation

- The value of the coefficient of correlation (r) always lies between ±1**.** Such as:

r=+1, perfect positive correlation
r=-1, perfect negative correlation
r=0, no correlation

- The coefficient of correlation is independent of the origin and scale. By origin, it means subtracting any non-zero constant from the given value of X and Y the value of "r" remains unchanged. By scale it means, there is no effect on the value of "r" if the value of X and Y is divided or multiplied by any constant.
- The coefficient of correlation is a geometric mean of two regression coefficient. Symbolically it is represented as:

$$r = \sqrt{b_{xy} + b_{yx}}$$

- The coefficient of correlation is " zero" when the variables X and Y are independent. But, however, the converse is not true.

Assumptions of Karl Pearson's Coefficient of Correlation

1. The relationship between the variables is "Linear", which means when the two variables are plotted, a straight line is formed by the points plotted.
2. There are a large number of independent causes that affect the variables under study so as to form a Normal Distribution. Such as, variables like price, demand, supply, etc. are affected by such factors that the normal distribution is formed.
3. The variables are independent of each other.

Note: The coefficient of correlation measures not only the magnitude of correlation but also tells the direction. Such as, r = -0.67, which shows correlation is negative because the sign is **"-"** and the magnitude is 0.67**.**

Pearson Correlation Coefficient Formula

The most common formula is the Pearson Correlation coefficient used for linear dependency between the data set. The value of the coefficient lies between -1 to +1. When the coefficient comes down to zero, then the data is considered as not related. While, if we get the value of +1, then the data are positively correlated and -1 has a negative correlation.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where, n = Quantity of Information
Σx = Total of the First Variable Value
Σy = Total of the Second Variable Value
Σxy = Sum of the Product of & Second Value
Σx^2 = Sum of the Squares of the First Value
Σy^2 =Sum of the Squares of the Second Value