30.04.2020

B.COM PROG SEMESTER II-
BUSINESS MATHEMATICS AND STATISTICS

SIMPLE LINEAR CORRELATION

Correlation refers to a process for establishing whether or not relationships exist between two variables.

Correlation in Statistics

Methods of correlation summarize the relationship between two variables in a single number called the correlation coefficient. The correlation coefficient is usually given the symbol r and it ranges from -1 to +1.

A correlation coefficient quite close to 0, but either positive or negative implies little or no relationship between the two variables. A correlation coefficient close to plus 1 means a positive relationship between the two variables, with increases in one of the variables being associated with increases in the other variable.

A correlation coefficient close to -1 indicates a negative relationship between two variables, with an increase in one of the variables being associated with a decrease in the other variable. A correlation coefficient can be produced for ordinal, interval or ratio level variables, but has little meaning for variables which are measured on a scale which is no more than nominal.

For ordinal scales, the correlation coefficient which is usually calculated is Spearman's rho. For interval or ratio level scales, the most commonly used correlation coefficient is Pearson's r, ordinarily referred to as simply the correlation coefficient.

Correlation Coefficient

The correlation coefficient, r, is a summary measure that describes the extent of the statistical relationship between two interval or ratio level variables. The correlation coefficient is scaled so that it is always between

-1 and +1. When r is close to 0 this means that there is little relationship between the variables and the farther away from 0 r is, in either the positive or negative direction, the greater the relationship between the two variables.

Types of Correlation

1. Positive and Negative Correlation:

   Whether the correlation between the variables is positive or negative depends on its direction of change. The correlation is positive when both the variables move in the same direction, i.e. when one variable increases the other on an average also increases and if one variable decreases the other also decreases. The correlation is said to be negative when both the variables move in the opposite direction, i.e. when one variable increases the other decreases and vice versa.

2. Simple, Partial and Multiple Correlation:

   Whether the correlation is simple, partial or multiple depends on the number of variables studied. The correlation is said to be simple when only two variables are studied. The correlation is either multiple or partial when three or more variables are studied. The correlation is said to be Multiple when three variables are studied simultaneously. Such as, if we want to study the relationship between the yield of wheat per acre and the amount of fertilizers and rainfall used, then it is a problem of multiple correlations.

   Whereas, in the case of a partial correlation we study more than two variables, but consider only two among them that would be influencing each other such that the effect of the other influencing variable is kept constant. Such as, in the above example, if we study the relationship between the yield and fertilizers used during the periods when certain average temperature existed, then it is a problem of partial correlation.

3. Linear and Non-Linear (Curvilinear) Correlation:

Whether the correlation between the variables is linear or non-linear depends on the constancy of ratio of change between the variables. The correlation is said to be linear when the amount of change in one variable to the amount of change in another variable tends to bear a constant ratio. For example, from the values of two variables given below, it is clear that the ratio of change between the variables is the same:

X:	10	20	30	40	50
Y:	20	40	60	80	100

The correlation is called as non-linear or curvilinear when the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable. For example, if the amount of fertilizers is doubled the yield of wheat would not be necessarily be doubled.

Thus, these are three most important types of correlation classified on the basis of movement, number and the ratio of change between the variables. The researcher must study these carefully to determine the correlation methods to be used to identify the extent to which the variables are correlated.