Pre Ph. D Course work
Department of sociology
Paper- II
Unit-5

Topic- MEASURES OF RELATIONSHIP

So far we have dealt with those statistical measures that we use in context of univariate population i.e., the population consisting of measurement of only one variable. But if we have the data on two variables, we are said to have a bivariate population and if the data happen to be on more than two variables, the population is known as multivariate population. If for every measurement of a variable, *X*, we have corresponding value of a second variable, *Y*, the resulting pairs of values are called a bivariate population. In addition, we may also have a corresponding value of the third variable, *Z*, or the forth variable, *W*, and so on, the resulting pairs of values are called a multivariate population. In case of bivariate or multivariate populations, we often wish to know the relation of the two and/or more variables in the data to one another. We may like to know, for example, whether the number of hours students devote for studies is somehow related to their family income, to age, to sex or to similar other factor. There are several methods of determining the relationship between variables, but no method can tell us for certain that a correlation is indicative of causal relationship. Thus we have to answer two types of questions in bivariate or multivariate populations viz.,

(i) Does there exist association or correlation between the two (or more) variables? If yes, of what degree?

(ii) Is there any cause and effect relationship between the two variables in case of the bivariate population or between one variable on one side and two or more variables on the other side in case of multivariate population? If yes, of what degree and in which direction?

The first question is answered by the use of correlation technique and the second question by the technique of regression. There are several methods of applying the two techniques, but the important
ones are as under:

In case of bivariate population: Correlation can be studied through (a) cross tabulation; (b) Charles Spearman's coefficient of correlation; (c) Karl Pearson's coefficient of correlation;

whereas cause and effect relationship can be studied through simple regression equations.

In case of multivariate population: Correlation can be studied through (a) coefficient of multiple correlation; (b) coefficient of partial correlation; whereas cause and effect relationship can be studied through multiple regression equations. We can now briefly take up the above methods one by one.

Cross tabulation approach is specially useful when the data are in nominal form. Under it we classify each variable into two or more categories and then cross classify the variables in these subcategories.Then we look for interactions between them which may be symmetrical, reciprocal or asymmetrical.

A symmetrical relationship is one in which the two variables vary together, but we assume that neither variable is due to the other. A reciprocal relationship exists when the two variables mutually influence or reinforce each other. Asymmetrical relationship is said to exist if one variable (the independent variable) is responsible for another variable (the dependent variable). The cross classification procedure begins with a two-way table which indicates whether there is or there is not an interrelationship between the variables. This sort of analysis can be further elaborated in which case a third factor is introduced into the association through cross-classifying the three variables. By doing so we find conditional relationship in which factor *X* appears to affect factor *Y* only when factor *Z* is held constant. The correlation, if any found through this approach is not considered a very powerful form of statistical correlation and accordingly we use some other methods when data happen to be either ordinal or interval or ratio data.

Charles Spearman's coefficient of correlation (*or rank correlation*) is the technique of determining the degree of correlation between two variables in case of ordinal data where ranks are given to the different values of the variables. The main objective of this coefficient is to determine the extent to which the two sets of ranking are similar or dissimilar. This coefficient is determined as under

$$\text{Spearman's coefficient of correlation (or } r_s) = 1 - \left[\frac{6\sum d_i^2}{n\left(n^2-1\right)}\right]$$

where *di* = difference between ranks of *i*th pair of the two variables;

n = number of pairs of observations.

As rank correlation is a non-parametric technique for measuring relationship between paired observations of two variables when data are in the ranked form.

Email-artirani21nov@gmail.com
Contact-6200360965