Department of Sociology
Pre Ph. D course work
Paper-II
Unit-5

Topic-Measures of Relationship

Part-B

C) Karl Pearson's coefficient of correlation (or simple correlation) is the most widely used method of measuring the degree of relationship between two variables. This coefficient assumes the following:

(i) that there is linear relationship between the two variables;

(ii) that the two variables are casually related which means that one of the variables is independent and the other one is dependent; and

(iii) a large number of independent causes are operating in both variables so as to produce a normal distribution.

Karl Pearson's coefficient of correlation can be worked out thus.

$$\text{Karl Pearson's coefficient of correlation (or } r)^{*} = \frac{\Sigma\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n \cdot \sigma_X \cdot \sigma_Y}$$

* Alternatively, the formula can be written as:

$$r = \frac{\Sigma\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\Sigma\left(X_i - \bar{X}\right)^2 \cdot \Sigma\left(Y_i - \bar{Y}\right)^2}}$$

Or

$$r = \frac{\text{Covariance between } X \text{ and } Y}{\sigma_x \cdot \sigma_y} = \frac{\Sigma\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)/n}{\sigma_x \cdot \sigma_y}$$

Or

$$r = \frac{\Sigma X_i Y_i - n \cdot \bar{X} \cdot \bar{Y}}{\sqrt{\Sigma X_i^2 - n\bar{X}^2}\ \sqrt{\Sigma Y_i^2 - n\bar{Y}^2}}$$

(This applies when we take zero as the assumed mean for both variables, X and Y.)

where X_i = ith value of X variable

$\overline{X}$ = mean of X

Y_i = ith value of Y variable

$\overline{Y}$ = Mean of Y

n = number of pairs of observations of X and Y

σ_X = Standard deviation of X

σ_Y = Standard deviation of Y

In case we use assumed means (A_x and A_y for variables X and Y respectively) in place of true means, then Karl Person's formula is reduced to:

$$\frac{\frac{\sum dx_i \cdot dy_i}{n} - \left(\frac{\sum dx_i \cdot \sum dy_i}{n}\right)}{\sqrt{\frac{\sum dx_i^2}{n} - \left(\frac{\sum dx_i}{n}\right)^2} \sqrt{\frac{\sum dy_i^2}{n} - \left(\frac{\sum dy_i}{n}\right)^2}}$$

$$\frac{\frac{\sum dx_i \cdot dy_i}{n} - \left(\frac{\sum dx_i \cdot \sum dy_i}{n}\right)}{\sqrt{\frac{\sum dx_i^2}{n} - \left(\frac{\sum dx_i}{n}\right)^2} \sqrt{\frac{\sum dy_i^2}{n} - \left(\frac{\sum dy_i}{n}\right)^2}}$$

where
$$\sum dx_i = \sum(X_i - A_x)$$
$$\sum dy_i = \sum(Y_i - A_y)$$
$$\sum dx_i^2 = \sum(X_i - A_x)^2$$
$$\sum dy_i^2 = \sum(Y_i - A_y)^2$$
$$\sum dx_i \cdot dy_i = \sum(X_i - A_x)(Y_i - A_y)$$

n = number of pairs of observations of X and Y.

This is the short cut approach for finding 'r' in case of ungrouped data. If the data happen to be grouped data (i.e., the case of bivariate frequency distribution), we shall have to write Karl Pearson's coefficient of correlation as under:

$$\frac{\frac{\sum f_{ij} \cdot dx_i \cdot dy_j}{n} - \left(\frac{\sum f_i dx_i}{n} \cdot \frac{\sum f_j dy_j}{n}\right)}{\sqrt{\frac{\sum f_i dx_i^2}{n} - \left(\frac{\sum f_i dx_i}{n}\right)} \sqrt{\frac{\sum f_j dy_j^2}{n} - \left(\frac{\sum f_j dy_j}{n}\right)^2}}$$

where *fij* is the frequency of a particular cell in the correlation table and all other values are defined as earlier.

Karl Pearson's coefficient of correlation is also known as the product moment correlation coefficient. The value of '*r*' lies between ± 1. Positive values of *r* indicate positive correlationbetween the two variables (i.e., changes in both variables take place in the statement direction),whereas negative values of '*r*' indicate negative correlation i.e., changes in the two variables takingplace in the opposite directions. A zero value of '*r*' indicates that there is no association between thetwo variables. When $r = (+)\ 1$, it indicates perfect positive correlation and when it is (–)1, it indicates perfect negative correlation, meaning thereby that variations in independent variable (X) explain 100% of the variations in the dependent variable (Y). We can also say that for a unit change in independent variable, if there happens to be a constant change in the dependent variable in the same direction, then correlation will be termed as perfect positive. But if such change occurs in the opposite direction, the correlation will be termed as perfect negative. The value of '*r*' nearer to +1 or –1 indicates high degree of correlation between the two variables.

Email-artirani21nov@gmail.com
Contact-6200360965