Rank Correlation.

Let us suppose that a group of n individuals is arranged in order of merit or proficiency in possession of two characteristics A and B.

These rank in two characteristics will, in general, be different.

For example, if we consider the relation between intelligence and beauty, it is not necessary that a beautiful individual is intelligent also.

Rank Correlation: $\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$, which is the Spearman's formulae for rank correlation coefficient.

Where $\sum d^2$ = sum of the squares of the difference of two ranks and n is the number of pairs of observations.

Note: We always have, $\sum d_i = \sum (x_i - y_i) = \sum x_i - \sum y_i = n(\bar{x}) - n(\bar{y}) = 0$, $(\because \bar{x} = \bar{y})$

If all d's are zero, then $r = 1$, which shows that there is perfect rank correlation between the variable and which is maximum value of r.

If however some values of x_i are equal, then the coefficient of rank correlation is given by

$$r = 1 - \frac{6\left[\sum d^2 + \left(\frac{1}{12}\right)(m^3 - m)\right]}{n(n^2 - 1)}$$

Where m is the number of times a particular x_i is repeated.

Positive and Negative rank correlation coefficients

Let r be the rank correlation coefficient then, if

- $r > 0$, it means that if the rank of one characteristic is high, then that of the other is also high or if the rank of one characteristic is low, then that of the other is also low. e.g., if the two characteristics be height and weight of persons, then $r > 0$ means that the tall persons are also heavy in weight.

- $r = 1$, it means that there is perfect correlation in the two characteristics i.e., every individual is getting the same ranks in the two characteristics. Here the ranks are of the type (1, 1), (2, 2),....., (n, n).

- $r < 1$, it means that if the rank of one characteristics is high, then that of the other is low or if the rank of one characteristics is low, then that of the other is high. e.g., if the two characteristics be richness and slimness in person, then $r < 0$ means that the rich persons are not slim.

- $r = -1$, it means that there is perfect negative correlation in the two characteristics i.e, an individual getting highest rank in one characteristic is getting the lowest rank in the second characteristic. Here the rank, in the two characteristics in a group of n individuals are of the type (1, n), $(2, n-1), \ldots, (n, 1)$.
- $r = 0$, it means that no relation can be established between the two characteristics.

Important Tips

- If $r = 0$, the variable x and y are said to be uncorrelated or independent.
- If $r = -1$, the correlation is said to be negative and perfect.
- If $r = +1,$ the correlation is said to be positive and perfect.
- Correlation is a pure number and hence unitless.
- Correlation coefficient is not affected by change of origin and scale.
- If two variate are connected by the linear relation $x + y = K$, then x, y are in perfect indirect correlation. Here $r = -1$.
- If x, y are two independent variables, then $\rho(x+y, x-y) = \frac{\sigma_x^2 - \sigma_y^2}{\sigma_x^2 + \sigma_y^2}$.
- $r(x,y) = \frac{\sum u_i v_i - \frac{1}{n}\sum u_i . \sum v_i}{\sqrt{\sum u_i^2 - \frac{1}{n}\left(\sum u_i\right)^2}\sqrt{\sum v_i^2 - \frac{1}{n}\left(\sum v_i\right)^2}}$, where $u_i = x_i - A,\ v_i = y_i - B$.