Spearman's rank correlation coefficient :-

Let a group of n individuals is arranged in order of merit of two characteristics A and B. The ranks in the two characteristics will in general be different.

eg: If we consider the relation between intelligence and beauty, it is not necessary that a beautiful individuals is intelligent also.

Let (x_i, y_i) ; $i = 1, 2, \ldots n$ be the ranks of the i^{th} individual in two characteristic A & B respectively.

The coefficient of correlation between A & B is given by ~~Spearman's~~ is given by.

$$\rho = 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)}$$

and known as Spearman's rank correlation coefficient.

where $\quad d_i = x_i - y_i \quad$ (difference betn two ranks

or $\quad d_i = (x_i - \bar{x}) - (y_i - \bar{y})$

as both the characteristic A & B is assigning the values $1, 2, - -, n$.

$$\therefore \bar{x} = \bar{y} = \frac{1}{n}(1+2+ \cdots +n) = \frac{1}{n}\left(\frac{n(n+1)}{2}\right)$$

$$= \left(\frac{n+1}{2}\right)$$

also $\sigma_x^2 = \frac{1}{n} \Sigma x_i^2 - \bar{x}^2 = \frac{1}{n}(1^2+2^2+ \cdots +n^2) - \left(\frac{n+1}{2}\right)^2$

$$= \frac{n(n+1)(2n+1)}{n \times 6} - \frac{(n+1)^2}{2} = \frac{n^2-1}{12}$$

$$\sigma_x^2 = \frac{n^2-1}{12} = \sigma_y^2$$

So if $d_i = (x_i - \bar{x}) - (y_i - \bar{y})$

then squaring and summing over i from 1 to n we get

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} [(x_i - \bar{x}) - (y_i - \bar{y})]^2$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$\Sigma d_i^2 = n\sigma_x^2 + n\sigma_y^2 - 2\rho\sigma_x\sigma_y$$

where ρ is the rank corr. coff. betn. A and B

$$\frac{1}{n}\Sigma d_i^2 = 2\sigma_x^2 - 2\rho\sigma_x^2$$

$$\Rightarrow 1 - \rho = \frac{\Sigma d_i^2}{2n\sigma_x^2}$$

$$\rho = 1 - \frac{\sum_{i=1}^{n} d_i^2}{2n\sigma_x^2} = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} \quad \left[\text{as } \sigma_x^2 = \frac{n^2-1}{12} \right.$$

which is the Spearman's formula for the rank correlation coefficient.

Tied Ranks :— If some of the individuals receive the same rank in a ranking of merit they are said to be tied. Let m individuals are tied say $(K+1)^{th}$, $(K+2)^{th}$ .... $(K+m)^{th}$. then each of these m individuals is assigned a common rank which is the arithmatic mean of the ranks $K+1, K+2, \ldots, K+m$

then in calculation we reduce sum of squares by $\frac{m(m^2-1)}{12}$.

Let $T_x = \frac{m_i(m_i^2-1)}{12}$ & $T_y = \frac{m_j^3 - m_j}{12}$

then

$$\rho(X,Y) = 1 - \frac{6\left[\Sigma d^2 + T_x + T_y\right]}{n(n^2-1)}$$

limits of rank corr. coeff. is also $-1 \le \rho \le 1$

Regression :- The term "regression" means "stepping back towards the average"

It was first used by a British biometrician "Sir Francis Galton" (1822 – 1911).

Defn :- Regression analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data.

or

Defn :- In Statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).

The case when we have one explanatory variable (indept variable) is called simple linear regression.

The case when we have more than one explanatory variable (indept. variable) the process is called multiple linear regression.

Lines or regression :- If we draw variables in a bivariate distribution and if these variables are related i.e. the points in the scatter diagram will cluster round some curve is called the "curve of regression". If the obtained curve is a straight line it is called the line of regression and we say there is linear regression between the variables. otherwise regression is said to be curvilinear.

The line of regression is the line which gives the best estimate to the value of one variable for any specific value of the other variables.

The line of regression is the line of "best fit" obtained by principle of least squares.

Let us suppose that in the bivariate distn. (x_i, y_i); $i = 1, 2, \ldots, n$; Y is a dependent variable and X is indept. variable.