CHAPTER # 08

CORRELATION , MULTIPLE AND PARTIAL CORRELATION

- **Correlation**

  The interdependence of two or more variables is called correlation.

  Or

  The liner relationship b/w two or more variables is called correlation. For example, an increase in the amount of rainfall will increase the sales of raincoats. Ages and weights of children are correlated with each other.

- **Positive Correlation**

  The correlation in the same direction is called positive correlation. If one variable increase other is also increase, and one variable is decrease other is also decrease. For example, an increase in heights of children is usually accompanied by an increase in their weights. The length of an iron bar will increase as the temperature increase.

- **Negative Correlation**

  The correlation in opposite (different) direction is called negative correlation. If one variable increase other is decrease, and one variable is decrease other is increase. For example, the volume gas will decrease as the pressure increase.

- **No Correlation Or Zero Correlation**

  If there are no relationship b/w two variables then it is called no correlation or zero correlation.

- **Coefficient of Correlation**

  It is a measurement of the degree of interdependence b/w the variable. It is a pure number and lies b/w -1 to +1 and intermediate value of zero indicates the absence of correlation. it denoted by r.

- **Properties of Correlation Coefficient**

  1. The correlation coefficient is symmetrical with respect to X and Y i.e. $r_{xy} = r_{yx}$
  2. The correlation co-efficient is the geometric mean of the two regression coefficients.

     $$r = \sqrt{b \times d} \text{ Or } r = \sqrt{b_{xy} \times b_{yx}}$$

  3. The correlation coefficient is independent of origin and unit of measurement i.e. $r_{xy} = r_{uv}$
  4. The correlation coefficient lies b/w -1 and +1.i.e. $-1 \le r \le +1$
  5. It is a pure number.

- **Formulas of Correlation Coefficient**

  For ungrouped Data

  (1). $$r = r_{xy} = r_{yx} = \frac{\sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}\right)}}$$

(2). $r = r_{xy} = r_{yx} = \dfrac{\sum (X-\bar{X}) \sum (Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2 \sum (Y-\bar{Y})^2}}$

(3). $r = r_{xy} = r_{yx} = \dfrac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{\left(\sum X^2 - n\bar{X}^2\right)\left(\sum Y^2 - n\bar{Y}^2\right)}}$

(4). $r = r_{xy} = r_{yx} = \dfrac{\sum (X-\bar{X}) \sum (Y-\bar{Y})}{nS_x S_y}$

(5). $r = r_{xy} = r_{yx} = \dfrac{\sum XY - n\bar{X}\bar{Y}}{nS_x S_y}$

(6). $r = r_{xy} = r_{yx} = \dfrac{\sum D_X D_Y - \dfrac{\left(\sum D_X\right)\left(\sum D_Y\right)}{n}}{\sqrt{\left(\sum D_X{}^2 - \dfrac{\left(\sum D_X\right)^2}{n}\right)\left(\sum D_y{}^2 - \dfrac{\left(\sum D_y\right)^2}{n}\right)}}$

Where $D_X = X - A$, $D_Y = Y - B$

(7). $r = r_{uv} = r_{vu} = \dfrac{\sum UV - \dfrac{\left(\sum U\right)\left(\sum V\right)}{n}}{\sqrt{\left(\sum U^2 - \dfrac{\left(\sum U\right)^2}{n}\right)\left(\sum V^2 - \dfrac{\left(\sum V\right)^2}{n}\right)}}$

Where $U = \dfrac{X-A}{h} = \dfrac{D_X}{h}$, $V = \dfrac{Y-B}{k} = \dfrac{D_y}{k}$

(8). $r = r_{xy} = r_{yx} = \sqrt{b_{xy} \times b_{yx}}$ **Or** $r = r_{xy} = r_{yx} = \sqrt{b \times d}$

Where $b = b_{yx} = r\dfrac{S_y}{S_x}$, $d = b_{xy} = r\dfrac{S_x}{S_y}$

For Grouped Data

(1). $r = r_{xy} = r_{yx} = \dfrac{\sum fXY - \dfrac{\left(\sum fX\right)\left(\sum fY\right)}{\sum f}}{\sqrt{\left(\sum fX^2 - \dfrac{\left(\sum fX\right)^2}{\sum f}\right)\left(\sum Y^2 - \dfrac{\left(\sum fY\right)^2}{\sum f}\right)}}$

(2). $$r = r_{xy} = r_{yx} = \frac{\sum fD_X D_Y - \frac{\left(\sum fD_X\right)\left(\sum fD_Y\right)}{\sum f}}{\sqrt{\left(\sum fD_X{}^2 - \frac{\left(\sum fD_X\right)^2}{\sum f}\right)\left(\sum fD_y{}^2 - \frac{\left(\sum fD_y\right)^2}{\sum f}\right)}}$$

(3). $$r = r_{uv} = r_{vu} = \frac{\sum fUV - \frac{\left(\sum fU\right)\left(\sum fV\right)}{\sum f}}{\sqrt{\left(\sum fU^2 - \frac{\left(\sum fU\right)^2}{\sum f}\right)\left(\sum V^2 - \frac{\left(\sum fV\right)^2}{\sum f}\right)}}$$

(4). $r = r_{xy} = r_{yx} = \sqrt{b_{xy} \times b_{yx}}$ **Or** $r = r_{xy} = r_{yx} = \sqrt{b \times d}$

Where $$b_{vu} = \frac{\sum fUV - \frac{\left(\sum fU\right)\left(\sum fV\right)}{\sum f}}{\sum fU^2 - \frac{\left(\sum fU\right)^2}{\sum f}} \quad , \quad b_{yx} = \frac{k}{h} b_{vu}$$

$$b_{uv} = \frac{\sum fUV - \frac{\left(\sum fU\right)\left(\sum fV\right)}{\sum f}}{\sum fV^2 - \frac{\left(\sum fV\right)^2}{\sum f}} \quad , \quad b_{xy} = \frac{h}{k} b_{uv}$$

- **Rank Correlation**

  Sometimes, the actual measurement or counts of individuals or objects are either not available or accurate assessment is not possible. They are then arranged in order according to some characteristic of interest. Such an ordered arrangement is called a ranking and the order given to an individual or object is called its rank. The correlation b/w two such sets of rankings are known as Rank Correlation.

  Rank Correlation = $r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$ (Spearman's Formula)

  Where d=difference b/w ranks of corresponding values of X and Y
  n= number of pairs of values (X, Y) in the data.

- **Rank Correlation for Tied Ranks**

  The spearman's coefficient or rank correlation applies only when no ties are present. In case there are ties in ranks, the ranks are adjusted by assigning the mean of the ranks which the tied objects or observations would have if they were ordered.

  Rank Correlation for Tied = $r_s = 1 - \frac{6\left(\sum d^2 + a\right)}{n(n^2 - 1)}$

$$a = \frac{1}{12}\left(t_1^3 - t_1\right) + \frac{1}{12}\left(t_2^3 - t_2\right) + ...$$

Where t= tied values

- **Multiple Correlation**

  Multiple correlation coefficient measures the degree of relationship b/w a variable and a group of variables and variable is not included in that group e.g. $R_{y.12}$, $R_{1.23}$

(1). $R_{1.23} = R_{1.32} = \sqrt{\dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$

Or

$$R_{1.23} = R_{1.32} = \sqrt{1 - \frac{\Delta}{\Delta_{11}}}$$

(2). $R_{2.13} = R_{2.31} = \sqrt{\dfrac{r_{21}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}}$

Or

$$R_{2.13} = R_{2.31} = \sqrt{1 - \frac{\Delta}{\Delta_{22}}}$$

(3). $R_{3.12} = R_{3.21} = \sqrt{\dfrac{r_{31}^2 + r_{32}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}}$

Or

$$R_{3.12} = R_{3.12} = \sqrt{1 - \frac{\Delta}{\Delta_{33}}}$$

Where $\Delta = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix} = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix}$

$$\Delta_{11} = 1 - r_{23}^2$$
$$\Delta = 1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}$$

$$\because r_{12} = r_{21} \ , \ r_{23} = r_{32} \ , \ r_{13} = r_{31}$$

Hence $R_{1.23}^2, R_{2.13}^2, R_{3.12}^2$ are known as coefficient of multiple determination

- **Partial Correlation**

  Correlation b/w two variable keeping the effects of all other variables as constant is called partial correlation for example $r_{12.3}, r_{13.2}, r_{23.1}$

(1). $r_{12.3} = r_{21.3} = \dfrac{r_{12} - r_{13}r_{23}}{\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}}$

Or

$$r_{12.3} = r_{21.3} = \sqrt{b_{12.3} \times b_{21.3}}$$

(2). $r_{13.2} = r_{31.2} = \dfrac{r_{13} - r_{12}r_{32}}{\sqrt{\left(1-r_{12}^2\right)\left(1-r_{32}^2\right)}}$

Or

$$r_{13.2} = r_{31.2} = \sqrt{b_{13.2} \times b_{31.2}}$$

(3). $r_{23.1} = r_{32.1} = \dfrac{r_{23} - r_{21}r_{31}}{\sqrt{\left(1-r_{21}^2\right)\left(1-r_{31}^2\right)}}$

Or

$$r_{23.1} = r_{32.1} = \sqrt{b_{23.1} \times b_{32.1}}$$

$$\because r_{12} = r_{21} \ , \ r_{23} = r_{32} \ , \ r_{13} = r_{31}$$