

Ch= 17 Correlation and Regression

* Correlation

→ Karl Pearson's Method :

Correlation coefficient = r

$$(i) \quad r_{xy} = \frac{\text{Covariance}}{(\text{sd of } x)(\text{sd of } y)} = \frac{\text{Cov}(x, y)}{s_x \cdot s_y}$$

where, $\text{Cov}(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$

$$= \frac{\sum xy - n\bar{x}\bar{y}}{n}$$

$$\therefore s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \quad \text{and} \quad s_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}$$

$$(ii) \quad r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \cdot \sqrt{\sum (y - \bar{y})^2}}$$

$$(iii) \quad r_{xy} = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}$$

mean of $x = \bar{x} = \frac{\sum x_i}{n}$

mean of $y = \bar{y} = \frac{\sum y_i}{n}$

$$(iv) r = \frac{n \sum uv - (\sum u)(\sum v)}{\sqrt{n \sum u^2 - (\sum u)^2} \cdot \sqrt{n \sum v^2 - (\sum v)^2}}$$

where, $u = X - A$ or $\frac{x - A}{c_x}$,

$v = y - B$ or $\frac{y - B}{c_y}$

$$\left. \begin{aligned} (v) r_{xy} &= \frac{\sum (x - \bar{x})(y - \bar{y})}{n \cdot s_x \cdot s_y} \\ (vi) r_{xy} &= \frac{\sum xy - n\bar{x}\bar{y}}{n \cdot s_x \cdot s_y} \end{aligned} \right\} \text{(Especially for short sums)}$$

→ Spearman's Rank Correlation Method.

(i) $r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$ (when the obs. are not repeated)

(ii) $r = 1 - \frac{6 [\sum d^2 + CF]}{n(n^2 - 1)}$ (when some of the obs. repeated)

where, $d = \text{rank of } x - \text{Rank of } y$

$$d = R_x - R_y$$

CF = Correction factor = $\sum \left(\frac{m^3 - m}{12} \right)$

$m = \text{No. of times obs. is repeated}$

* Regression

→ Equation of regression line

$$\hat{y} = a + bx$$

where, $b = b_{yx}$ = Regression coefficient

$$(i) b_{yx} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

$$(ii) b_{yx} = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

$$(iii) b = b_{yx} = \frac{n \sum uv - (\sum u)(\sum v)}{n \sum u^2 - (\sum u)^2}$$

Here, $u = X - A$ and $v = Y - B$

$$(iv) b = b_{yx} = \frac{n \sum uv - (\sum u)(\sum v)}{n \sum u^2 - (\sum u)^2} \times \frac{C_y}{C_x}$$

Here, $u = \frac{X - A}{C_x}$ & $v = \frac{Y - B}{C_y}$

$$(v) b_{yx} = r \cdot \frac{S_y}{S_x}$$

$$(vi) b_{yx} = \frac{\text{Cov}(x, y)}{S_x^2} = \frac{r \cdot S_x \cdot S_y}{S_x^2}$$

$$(vii) a = \bar{y} - b\bar{x}$$

(viii) Coefficient of Determination (R^2)

$$R^2 = [r(y, \hat{y})]^2 = [r(x, y)]^2 = r^2$$

=> in Illustration

Coefficient of Determination

$$R^2 = \left[\frac{\text{Cov}(x, y)}{s_x \cdot s_y} \right]^2$$

error find karwa mate

$$e = y - \hat{y}$$

→ Coefficient of concurrent Deviations

refer = Example 17.15 (Pg-17, 24)

$$\text{formula} = r_c = \pm \sqrt{\frac{\pm (2c - m)}{m}}$$

where m = no. of pairs of deviation

c = no. of positive signs in product of Deviation column

= no. of concurrent deviation.

(ex) Coefficient of non-determination
 $= (1 - r^2)$