Correlation Coefficient

Karl Pearson's Method

Karl Pearson's Coefficient of Correlation

- $r = \frac{Covariance}{\sigma_x \sigma_y}$
- $r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})/N}{\sqrt{\frac{\Sigma(X-\bar{X})^2}{N} \times \frac{\Sigma(Y-\bar{Y})^2}{N}}} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \times \sum(Y-\bar{Y})^2}}$

Where,

r = coefficient of correlation.

$\bar{X}$ and $\bar{Y}$ are mean of X and Y series.

N refers to number of pairs of observations.

$$r = \sum(X\text{-}X\ bar)(Y\text{-}Ybar) \ / \ \sqrt{(X\text{-}Xbar)sq * (Y\text{-}Ybar)sq}$$

$$= 100/\sqrt{50*242} = 100/\sqrt{12100} = 100/110 = 0.909$$

X	Y	x-xbar	x-xbar. Sq	y-ybar	x-xbar*y-ybar	Y-Ybar	Y-Ybar.sq
5	31	0	0	-19	0	1	1
11	40	6	36	-10	-60	10	100
4	30	-1	1	-20	20	0	0
5	34	0	0	-16	0	4	16
3	25	-2	4	-25	50	-5	25
2	20	-3	9	-30	90	-10	100
X bar=5	Y bar= 30		50		100		242

Interpretation of Correlation Coefficient (r)

- The value of correlation coefficient ‘r’ ranges from -1 to +1
- If r = +1, then the correlation between the two variables is said to be perfect and positive
- If r = -1, then the correlation between the two variables is said to be perfect and negative
- If r = 0, then there exists no correlation between the variables

Properties of Correlation coefficient

- The correlation coefficient lies between -1 & +1

  symbolically: $(-1 \leq r \geq 1)$

- The correlation coefficient is independent of the change of origin & scale.
- The coefficient of correlation is the geometric mean of two regression coefficient.

  $r = \sqrt{bxy * byx}$

- The sign of correlation coefficient and both the regression coefficients are same. If one regression coefficient is (+ve), the other regression coefficient will also be (+ve) and the correlation coefficient will also be (+ve). Similarly, if one is negative (-ve) the other two will also be negative (-ve).

Advantages of Pearson's Coefficient

- It summarizes in one value, the degree of correlation & direction of correlation also.