Properties of Karl Pearson's Co-efficient of correlation (r) :

5. Prove that $-1 \leq r \leq 1$ i.e., r lies between -1 and 1. **(Important) (Property No. 1)**

Solution : Suppose, X and Y are two variables which takes values (x_i, y_i) ; $i = 1,2, \dots, n$ with means $\bar{x}, \bar{y}$ and standard deviation σ_x, σ_y respectively.

Let us consider,

$$\sum \left[\frac{x-\bar{x}}{\sigma_x} \pm \frac{y-\bar{y}}{\sigma_y}\right]^2 \geq 0$$

⇨ $\sum \left[(\frac{x-\bar{x}}{\sigma_x})^2 + (\frac{y-\bar{y}}{\sigma_y})^2 \pm 2\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x \sigma_y}\right] \geq 0$

⇨ $\frac{1}{{\sigma_x}^2}\sum(x-\bar{x})^2 + \frac{1}{{\sigma_y}^2}\sum(y-\bar{y})^2 \pm \frac{2}{\sigma_x \sigma_y}\sum(x-\bar{x})(y-\bar{y}) \geq 0$

Dividing both sides by n, we get

$$\frac{1}{{\sigma_x}^2}\frac{\sum(x-\bar{x})^2}{n} + \frac{1}{{\sigma_y}^2}\frac{\sum(y-\bar{y})^2}{n} \pm \frac{2}{\sigma_x \sigma_y}\frac{\sum(x-\bar{x})(y-\bar{y})}{n} \geq 0$$

⇨ $\frac{1}{{\sigma_x}^2}. {\sigma_x}^2 + \frac{1}{{\sigma_y}^2}. {\sigma_y}^2 \pm \frac{2}{\sigma_x \sigma_y} cov(x, y) \geq 0$

⇨ $1 + 1 \pm 2r \geq 0$

⇨ $2 \pm 2r \geq 0$

⇨ $2\ (\ 1 \pm r\ ) \geq 0$

⇨ $1 \pm r\ \geq 0$

⇨ Either $1 + r\ \geq 0$ or $1 - r\ \geq 0$

⇨ $r\ \geq\ -1$ or $1 \geq r$

$\therefore -1 \leq r \leq 1$

Special Conditions (Interpretation)

i) The least value of r is –1 and the most is +1. If $r = \pm 1$, there is a perfect positive correlation between two variables. If $r = -1$, there is a perfect negative correlation.

ii) If $r = 0$, then there is no linear relation between the variables. However, there may be non-linear relationship between the variables.

iii) If r is positive but close to zero, then there will be weak positive correlation and if r is close to +1, then there will be strong positive correlation.

6. Prove that Correlation co-efficient r is independent of change in origin and scale. **(Property No. 2)**

Solution : Suppose, X and Y are the original variables and after changing origin and scale, we have

$$U = \frac{X-a}{h} \quad \text{and} \quad V = \frac{Y-b}{k} \quad \text{where } a, b, h, k \text{ are all constants.}$$

⇨ $X - a = hU$ and $Y - b = kV$

⇨ $X = a + hU$ and $Y = b + kV$

⇨ $\bar{X} = a + h\bar{U}$ and $\bar{Y} = b + k\bar{V}$

$$\therefore X - \bar{X} = h(U - \bar{U})$$

$$and\ Y - \bar{Y} = K(V - \bar{V})$$

Now, $r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$

$$= \frac{hk\sum(u-\bar{u}).(v-\bar{v})}{\sqrt{\sum h^2(u-\bar{u})^2 \sum k^2(v-\bar{v})^2}}$$

$$= \frac{\sum h(u-\bar{u}).k(v-\bar{v})}{hk\sqrt{\sum(u-\bar{u})^2 \sum(v-\bar{v})^2}}$$

$$= r_{uv}$$

$$\therefore \quad r_{xy} = r_{uv}$$

$Hence, Proved.$

7. Prove that Two independent variables are uncorrelated but the converse is not true. **(Property No. 3)**

Solution : If two variables are independent then their covariance is zero, i.e., $cov(X, Y) = 0$.

$$\therefore \quad r_{xy} = \frac{cov\,(X,Y)}{\sigma_x \sigma_y} = \frac{0}{\sigma_x \sigma_y} = 0$$

Thus, if two variables are independent their co-efficient of correlation is zero, i.e., independent variables are uncorrelated.

But, the converse is not true. If $r_{xy} = 0$, then there does not exist any linear correlation between the variables as Karl Pearson's coefficient of correlation r_{xy} is a measure of only linear relationship. However, there may be strong non-linear or curvilinear relationship even through $r_{xy} = 0$.

Let us consider, an illustration for bivariate distribution

x :	–3	–2	–1	0	1	2	3
y :	9	4	1	0	1	4	9

Applying the formula of r_{xy}, we get $r_{xy} = 0$. But, X and Y are not independent and they are related by the non-linear relation $y = x^2$. Hence Proved.

8. Write another Properties of Karl Pearson's Co-efficient of correlation (r).

Solution : (i) Correlation co-efficient r is a pure number independent of unit of measurement. **(Property No. 4)**

(ii) Correlation co-efficient is symmetric. **(Property No. 5)**

9. Write the assumptions of Karl Pearson's co-efficient of correlation.

Solution : There are three assumptions :

(i) The variables x and y are linearly related.
(ii) There is a cause and effect relationship between factors affecting the values of the variables x and y.
(iii) The random variables x and y are normally distributed.