TOPIC 2.6- Correlation

Correlation

If the change in one variable affects change in the other variable, then the riables are said to be correlated.

Karl - Pearson's Coefficient of correlation

Correlation coefficient between two random variables x and y, usually denoted by $r(x,y)$ is a numerical measure of linear relationship between them and is defined as

$$r(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$$

where $\text{cov}(x,y) = \frac{1}{n}\sum xy - \bar{x}\,\bar{y}$

$$\sigma_x = \sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2}, \quad \bar{x} = \frac{\sum x}{n}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2}, \quad \bar{y} = \frac{\sum y}{n}$$

Note

(a) Limits of correlation coefficient is $-1 \le r \le 1$.

(b) when $r = 1$, the correlation is perfect and positive

(c) when $r = 0$, two independent variables are uncorrelated.

(d) Correlation coefficient may also be denoted by $\rho(x,y)$ or ρ_{xy}.

① Find the coefficient of correlation between industrial production and export using the following data.

Production (x)	55	56	58	59	60	60	62
Export (y)	35	38	37	39	44	43	44

Correlation coefficient $r(x,y) = \dfrac{cov(x,y)}{\sigma_x \sigma_y}$

where $cov(x,y) = \dfrac{\sum xy}{n} - \bar{x}\,\bar{y}$

$$\sigma_x = \sqrt{\frac{1}{n}\sum x^2 - (\bar{x})^2} \quad \text{and} \quad \sigma_y = \sqrt{\frac{1}{n}\sum y^2 - (\bar{y})^2}$$

X	Y	$U = X-58$	$V = Y-40$	UV	U^2	V^2
55	35	−3	−5	15	9	25
56	38	−2	−2	4	4	4
58	37	0	−3	0	0	9
59	39	1	−1	−1	1	1
60	44	2	4	8	4	16
60	43	2	3	6	4	9
62	44	4	4	16	16	16
		4	0	48	38	80

$$\bar{U} = \frac{\Sigma U}{n} = 0.5714$$

$$\bar{V} = \frac{\Sigma V}{n} = 0$$

$$\text{COV}(U,V) = \frac{\Sigma UV}{n} - \bar{U}\,\bar{V} = \frac{48}{7} = 6.8571$$

$$\sigma_u = \sqrt{\frac{\sum U^2}{n} - \bar{U}^2} = \sqrt{\frac{38}{7} - (0.5714)^2}$$

$$= 2.2588$$

$$\sigma_v = \sqrt{\frac{\sum V^2}{n} - \bar{V}^2} = \sqrt{\frac{80}{7}} = 3.3806$$

$$r(x,y) = r(u,v) = \frac{6.8571}{2.2588 \times 3.3806}$$

$$= 0.898$$