

相関係数 (correlation coefficient)

n 組のデータ (x_i, y_i) の相関係数 R ($-1 \leq R \leq 1$) は

$$R \equiv \frac{S_{xy}}{S_x S_y} \quad (1)$$

で定義されます。ここで、 x の分散 (variance) S_x^2 と y の分散 S_y^2 は

$$S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2, \quad (2)$$

$$S_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \quad (3)$$

と表され、 x の平均 (average) $\bar{x}$ と y の平均 $\bar{y}$ は

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (4)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (5)$$

と表されます。さらに、共分散 (covariance) S_{xy} は

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \quad (6)$$

です。