

データの必須知識②

問題 次の各問いに答えなさい。

2組の n 個のデータ $x_1, x_2, x_3, \dots, x_n$ と $y_1, y_2, y_3, \dots, y_n$ があります。この時、2つの変量 x と y の共分散 s_{xy} と相関係数 r の定義を Σ を使って言いなさい。

解答

$$x \text{ と } y \text{ の共分散の定義: } s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\text{相関係数 } r \text{ の定義: } r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

※ $\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ と $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$ は、それぞれ変量 x と y の標準偏差である。