相関係数の定義

　2 変量のデータ $\{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$ に対し、2 変量 x と y の**相関係数** r を次のように定義する。

$$r = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_N - \bar{x})(y_N - \bar{y})}{\sqrt{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}\sqrt{(y_1 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2}}$$

また、2 変量のデータ $\{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$ に対し、次の統計量 S_{xy} を 2 変量 x と y の**共分散**と呼び、

$$S_{xy} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_N - \bar{x})(y_N - \bar{y})}{N - 1}$$

データ $\{x_1, x_2, \cdots, x_N\}$ に対し、次の統計量 s^2 を x の**分散**（分母が $N-1$ のものを不偏分散、N のものを標本分散）と呼ぶので、

$$s^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{N - 1}$$

相関係数の定義式は次のように変形できる。

$$r = \frac{x \text{ と } y \text{ の共分散}}{\sqrt{x \text{ の分散}}\sqrt{y \text{ の分散}}}$$

　さらに、相関係数 r は $-1 \leq r \leq 1$ の値をとります。1 に近いほど正の相関が、-1 に近いほど負の相関が強く、0 に近いほど相関が弱いことになります。