相関係数の話

2 つの変量 x, y の組 $(x_1,\ y_1), (x_2,\ y_2), \cdots, (x_n,\ y_n)$ について，x, y をそれぞれ標準化した値を成分として，次のベクトルを定める。

$$\vec{X} = \left(\frac{x_1 - \overline{x}}{s_x},\ \frac{x_2 - \overline{x}}{s_x},\ \cdots,\ \frac{x_n - \overline{x}}{s_x}\right)$$
$$\vec{Y} = \left(\frac{y_1 - \overline{y}}{s_y},\ \frac{y_2 - \overline{y}}{s_y},\ \cdots,\ \frac{y_n - \overline{y}}{s_y}\right)$$

x, y の相関係数 r_{xy} は，$\dfrac{\vec{X} \cdot \vec{Y}}{n}$ と表される。

$$\begin{aligned}\frac{\vec{X} \cdot \vec{Y}}{n} &= \frac{1}{n}\sum_{k=1}^{n} \frac{x_k - \overline{x}}{s_x} \cdot \frac{y_k - \overline{y}}{s_y} \\ &= \frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})}{s_x s_y} \\ &= \frac{s_{xy}}{s_x s_y} \\ &= r_{xy}\end{aligned}$$

ところで，

$$\begin{aligned}\left|\vec{X}\right| &= \sqrt{\sum_{k=1}^{n}\left(\frac{x_k - \overline{x}}{s_x}\right)^2} \\ &= \sqrt{\frac{n}{s_x^2} \cdot \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2} \\ &= \sqrt{\frac{n}{s_x^2} \cdot s_x^2} \\ &= \sqrt{n}\end{aligned}$$

であるから，$\vec{X}$ と $\vec{Y}$ のなす角を θ とすると，

$$\begin{aligned}r_{xy} = \frac{\vec{X} \cdot \vec{Y}}{n} &= \frac{\left|\vec{X}\right|\left|\vec{Y}\right| \cos\theta}{n} \\ &= \frac{\sqrt{n}\sqrt{n}\cos\theta}{n} \\ &= \cos\theta\end{aligned}$$

$-1 \leqq r_{xy} \leqq 1$ であることが確認できる。