【準備】

変量 x の n 個の値 x_1，x_2，……，x_n の平均値を $\overline{x}$ とするとき，$x_1-\overline{x}$，$x_2-\overline{x}$，……，$x_n-\overline{x}$ をそれぞれの値の **偏差** という。偏差の 2 乗の平均値を，変量 x の **分散** といい，s^2 で表す。

$$s^2=\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}$$

分散の正の平方根を **標準偏差** といい， s で表す。

$$s=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}}$$

偏差の積 $(x-\overline{x})(y-\overline{y})$ の平均値を，x と y の **共分散** といい， s_{xy} で表す。

$$s_{xy}=\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$$

x の標準偏差 s_x と y の標準偏差 s_y の積 s_xs_y で，共分散 s_{xy} を割った値を **相関係数** といい， r で表す。

$$r=\frac{s_{xy}}{s_xs_y}$$

【定理】

相関係数 r のとり得る値の範囲は $-1\leqq r\leqq 1$ である。

〈注意〉証明には，数学 B「数列」で学習する Σ（シグマ）を用いる。

証明

$x_i-\overline{x}=x_i'$ ，$y_i-\overline{y}=y_i'$ とおくと　$r=\dfrac{\sum_{i=1}^{n}x_i'\,y_i'}{n\,s_xs_y}$

ここで　$\left(\dfrac{x_i'}{s_x}+\dfrac{y_i'}{s_y}\right)^2=\dfrac{x_i'^2}{s_x^2}+2\dfrac{x_i'\,y_i'}{s_xs_y}+\dfrac{y_i'^2}{s_y^2}\geqq 0$ であるから

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i'}{s_x}+\frac{y_i'}{s_y}\right)^2=\frac{1}{s_x^2}\left(\frac{1}{n}\sum_{i=1}^{n}x_i'^2\right)+2\frac{\sum_{i=1}^{n}x_i'\,y_i'}{n\,s_xs_y}+\frac{1}{s_y^2}\left(\frac{1}{n}\sum_{i=1}^{n}y_i'^2\right)=\frac{s_x^2}{s_x^2}+2r+\frac{s_y^2}{s_y^2}=2+2r\geqq 0$$

よって　$r\geqq -1$

$\left(\dfrac{x_i'}{s_x}-\dfrac{y_i'}{s_y}\right)^2=\dfrac{x_i'^2}{s_x^2}-2\dfrac{x_i'\,y_i'}{s_xs_y}+\dfrac{y_i'^2}{s_y^2}\geqq 0$ であるから，同様の計算により　$\dfrac{1}{n}\sum_{i=1}^{n}\left(\dfrac{x_i'}{s_x}-\dfrac{y_i'}{s_y}\right)^2=2-2r\geqq 0$

よって　$r\leqq 1$　　したがって　$-1\leqq r\leqq 1$

ポイント　$\left(\dfrac{x_i'}{s_x}+\dfrac{y_i'}{s_y}\right)^2$ の 1 から n までの和の中に，標準偏差，相関係数が含まれることを利用する。