$\boxed{1}$ 2つの変量 x, y の n 個の組 $(x_1,\ y_1), (x_2,\ y_2), \cdots, (x_n,\ y_n)$ について，相関係数 r_{xy} の値域が $-1 \leqq r_{xy} \leqq 1$ であることを証明したい。途中までは次のように書いた。

［証明］

新しい変量 z を，$z_k = \dfrac{x_k - \overline{x}}{s_x} + \dfrac{y_k - \overline{y}}{s_y}$ $(k = 1,\ 2,\ \cdots,\ n)$ で定める。

両辺を 2 乗すると，

$$z_k^2 = \frac{(x_k - \overline{x})^2}{s_x^2} + 2 \cdot \frac{(x_k - \overline{x})(y_k - \overline{y})}{s_x s_y} + \frac{(y_k - \overline{y})^2}{s_y^2} \geqq 0$$

$k = 1, 2, \cdots, n$ について平均をとると，

$$\frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2}{s_x^2} + 2 \cdot \frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})}{s_x s_y} + \frac{\frac{1}{n}\sum_{k=1}^{n}(y_k - \overline{y})^2}{s_y^2} \geqq 0$$

(1) 上の証明の続きを書き，$r_{xy} \geqq -1$ を証明しなさい。

解答 $s_x^2 = \dfrac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2$，$s_y^2 = \dfrac{1}{n}\sum_{k=1}^{n}(y_k - \overline{y})^2$，$s_{xy} = \dfrac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})$ より，

$$1 + 2 \cdot \frac{s_{xy}}{s_x s_y} + 1 \geqq 0$$

$r_{xy} = \dfrac{s_{xy}}{s_x s_y}$ より，

$$1 + 2r_{xy} + 1 \geqq 0$$

$$r_{xy} \geqq -1$$

等号は，すべての k について $z_k = 0$ のとき成立する。 … 終

(2) $w_k = \dfrac{x_k - \overline{x}}{s_x} - \dfrac{y_k - \overline{y}}{s_y}$ $(k = 1,\ 2,\ \cdots,\ n)$ とおくことで，$r_{xy} \leqq 1$ を証明しなさい。

解答 両辺を 2 乗すると，

$$w_k^2 = \frac{(x_k - \overline{x})^2}{s_x^2} - 2 \cdot \frac{(x_k - \overline{x})(y_k - \overline{y})}{s_x s_y} + \frac{(y_k - \overline{y})^2}{s_y^2} \geqq 0$$

$k = 1, 2, \cdots, n$ について平均をとると，

$$\frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2}{s_x^2} - 2 \cdot \frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})}{s_x s_y} + \frac{\frac{1}{n}\sum_{k=1}^{n}(y_k - \overline{y})^2}{s_y^2} \geqq 0$$

$$\frac{s_x^2}{s_x^2} - 2 \cdot \frac{s_{xy}}{s_x s_y} + \frac{s_y^2}{s_y^2} \geqq 0$$

$$1 - 2r_{xy} + 1 \geqq 0$$

$$r_{xy} \leqq 1$$

等号は，すべての k について $w_k = 0$ のとき成立する。 … 終

参考 $r_{xy} = 1$ の成立条件「すべての k について $w_k = 0$」とはどのような場合か考えてみよう。

$$w_k = 0 \iff \frac{x_k - \overline{x}}{s_x} = \frac{y_k - \overline{y}}{s_y} \iff y_k = \frac{s_y}{s_x}(x_k - \overline{x}) + \overline{y}$$

これは，散布図ですべての点 $(x_k,\ y_k)$ が，1 つの直線 $y = \dfrac{s_y}{s_x}(x - \overline{x}) + \overline{y}$ 上に並ぶことを表している。