＜２次元正規分布に従うデータと回帰直線１＞

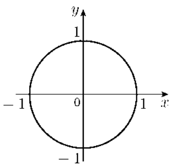
$(U_1,\ V_1),\ (U_2,\ V_2),\ \cdots,\ (U_n, V_n)$ は独立な 2 次元標準正規分布に従う確率変数列とし、2 次元データ

$$(*)\qquad \begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} U_i \\ V_i \end{pmatrix} + \begin{pmatrix} m_X \\ m_Y \end{pmatrix} \qquad (i = 1,\ 2,\ \cdots,\ n)$$

を考える。この一次変換が、定数 $\lambda_1 > \lambda_2 > 0,\ \ 0 < \theta < \frac{\pi}{2}$ に対して

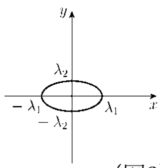
$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} \lambda_1\cos\theta & -\lambda_2\sin\theta \\ \lambda_1\sin\theta & \lambda_2\cos\theta \end{pmatrix}$$

と表されている場合、この変換 $(*)$ は図 1→ 図 2→ 図 3→ 図 4 のようになる。



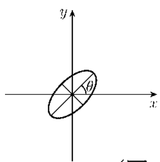
(図1)

$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \Longrightarrow$ $\begin{pmatrix} x \text{ 方向に } \lambda_1 \text{ 倍} \\ y \text{ 方向に } \lambda_2 \text{ 倍} \end{pmatrix}$



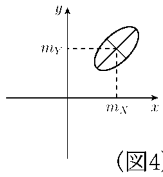
(図2)

$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \Longrightarrow$ (θ 回転)



(図3)

$+\begin{pmatrix} m_X \\ m_Y \end{pmatrix} \Longrightarrow$ (平行移動)



(図4)

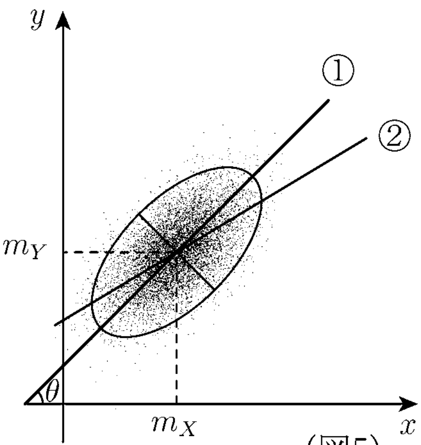
このとき 2 次元データ $(X_i,\ Y_i)\ (i = 1,\ 2,\ \cdots,\ n)$ の散布図は図 5 のような $(m_X,\ m_Y)$ を中心として、中心軸が直線

$$y = (\tan\theta)(x - m_X) + m_Y \ \cdots\ ①$$

である楕円の形になる。

図 5 の直線②は前ページで導いた 2 次元正規分布の場合の「$X = x$ のときの Y の条件付平均値」が表す直線

$$y = \rho\frac{\sigma_2}{\sigma_1}(x - m_X) + m_Y \ \cdots\ ②$$



(図5)

である。ただし $\sigma_1 = \sqrt{a^2 + c^2},\ \sigma_2 = \sqrt{b^2 + d^2},\ \rho = \dfrac{ab + cd}{\sigma_1\sigma_2}$ である。このとき、次が成り立つ。

$\boxed{1}$ n が十分大きいとき、直線①は 2 次元データ $(X_i,\ Y_i)\ (i = 1,\ 2,\ \cdots,\ n)$ の直交回帰直線とほぼ一致する。

$\boxed{2}$ n が十分大きいとき、直線②は 2 次元データ $(X_i,\ Y_i)\ (i = 1,\ 2,\ \cdots,\ n)$ の回帰直線とほぼ一致する。