確率統計学　解答例

2016.06.15

■　データ (x_1, y_1)，(x_2, y_2)，$\cdots$，(x_n, y_n) が与えられているとし，データ x_1，x_2，$\cdots$，x_n の平均を $\bar{x}$，標準偏差を s_x，データ y_1，y_2，$\cdots$，y_n の平均を $\bar{y}$，標準偏差を s_y，データ x_1，x_2，$\cdots$，x_n と y_1，y_2，$\cdots$，y_n の共分散を s_{xy} と表す．このとき，$s_x > 0$ かつ $s_y > 0$ の場合に対して

$$f(a,b) = \sum_{k=1}^{n} \{y_k - (a\,x_k + b)\}^2$$

を最小にする (a, b) を求めよ．

（解） 定義により

$$\sum_{k=1}^{n}(x_k - \bar{x}) = 0, \qquad \sum_{k=1}^{n}(y_k - \bar{y}) = 0, \qquad s_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2,$$
$$s_y^2 = \frac{1}{n}\sum_{k=1}^{n}(y_k - \bar{y})^2, \qquad s_{xy} = \frac{1}{n}\sum_{k=1}^{n}\{(x_k - \bar{x})\,(y_k - \bar{y})\}$$

であることに注意すると，$f(a,b)$ は

$$\begin{aligned}
\frac{f(a,b)}{n} =& \frac{1}{n}\sum_{k=1}^{n}\{(y_k - \bar{y}) - a\,(x_k - \bar{x}) + (\bar{y} - a\,\bar{x} - b)\}^2 \\
=& \frac{1}{n}\sum_{k=1}^{n}(y_k - \bar{y})^2 - 2\,a\,\frac{1}{n}\sum_{k=1}^{n}\{(y_k - \bar{y})\,(x_k - \bar{x})\} + a^2\,\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2 \\
&+ \frac{2\,(\bar{y} - a\,\bar{x} - b)}{n}\left\{\sum_{k=1}^{n}(y_k - \bar{y}) - a\sum_{k=1}^{n}(x_k - \bar{x})\right\} + (\bar{y} - a\,\bar{x} - b)^2 \\
=& s_y^2 - 2\,a\,s_{xy} + a^2\,s_x^2 + (\bar{y} - a\,\bar{x} - b)^2 = s_y^2 - \frac{s_{xy}^2}{s_x^2} + s_x^2\left(a - \frac{s_{xy}}{s_x^2}\right)^2 + (\bar{y} - a\,\bar{x} - b)^2
\end{aligned}$$

と表されるので，$f(a,b)$ が最小値をとる (a, b) は

$$a = \frac{s_{xy}}{s_x^2}, \qquad b = \bar{y} - a\,\bar{x} = \bar{y} - \frac{s_{xy}}{s_x^2}\cdot\bar{x}$$

である．■