

問 1

xy 平面上にデータ (x_k, y_k) を表す点 $P_k(x_k, y_k)$ をプロットするとき n 個の点 $P_1, P_2, \dots, P_n$ に最も隣接した直線を回帰直線という。点 P_k から y 軸に平行線を引き、直線 l との交点を Q_k とする。このとき線分の長さの 2 乗 $\overline{P_k Q_k}^2$ の和を最小にする直線 l を「 y の x への回帰直線」といい (このとき y を目的変数、 x を説明変数と呼ぶ)、次式で与えられる。

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

今 xy 平面上に $P_1(2, 1), P_2(4, 2), P_3(6, 3)$ の 3 点が存在する。このとき以下の問いに答えよ。

- (a) x の標準偏差 σ_x および y の標準偏差 σ_y を求めよ。
 (b) x と y の相関係数 r を求めよ。必要であれば以下の式を用いてよい。

$$r = \frac{s_{xy}}{\sigma_x \cdot \sigma_y}$$

$$s_{xy} = \frac{1}{n} \cdot \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}) : x \text{ と } y \text{ の共分散}$$

- (c) 式を用いて y の x への回帰直線を求めよ。
 (d) 回帰直線 $y = ax + b$ と置いた時、最小二乗法を用いて a と b を求めて回帰直線を求めよ。(この時、各点から y 軸に平行線を引き、直線 l との交点までの距離の二乗の和が最小になるようにせよ。) また、(c) で求めた式と比較してみよ。

ヒント：各点を式に代入して実際に距離を出し、偏微分を用いて距離が最小になるような a, b を求める。

解答

(a)

$$\bar{x} = \frac{2+4+6}{3} = 4, \bar{y} = \frac{1+2+3}{3} = 2, \text{ よって}$$

$$\sigma_x = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^n (x_k - \bar{x})^2} = 2\sqrt{\frac{2}{3}}$$

$$\sigma_y = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^n (y_k - \bar{y})^2} = \sqrt{\frac{2}{3}}$$

(b)

$$s_{xy} = \frac{1}{n} \cdot \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}) = \frac{4}{3}, \text{ よって}$$

$$r = \frac{s_{xy}}{\sigma_x \cdot \sigma_y} = 1$$

(c)

$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$ の公式に上記の値を代入し、回帰直線 $y = \frac{1}{2}x$ を得る。

(d)

各点と直線との距離の二乗の和を S と置くと

$$S = (2a+b-1)^2 + (4a+b-2)^2 + (6a+b-3)^2$$

$$= 56a^2 + 3b^2 + 14 + 24ab - 56a - 12b$$

S を a で偏微分すると

$$\frac{\partial S}{\partial a} = 112a + 24b - 56$$

よって、 $a = \frac{56-24b}{112}$ のとき S は最小になる。

これを S に代入して b で偏微分すると

$$\frac{\partial S}{\partial b} = 192b$$

よって、 $b=0$ のとき S は最小になる。

これを a の式に代入して $a = \frac{1}{2}$ となる。

以上まとめて回帰直線 $y = \frac{1}{2}x$ を得る。

この式は(c)で求めた回帰直線と一致していることが分かる。