応用確率統計（演習 14）　　　　　　　　　　　　　　平成 26 年 01 月 30 日

問 1

xy 平面上にデータ$\left(x_k y_k\right)$を表す点$P_k\left(x_k y_k\right)$をプロットするとき n 個の点$P_1, P_2, \cdots, P_n$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)、次式で表される。

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

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。このとき以下の問に答えよ。

(a) x の標準偏差 δ x および y の標準偏差 δ 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 と y の共分散

(c) 公式を用いて 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。