

確率統計と情報処理のレポート問題 (その 3)

問題 1

- 身長と体重のデータを行列配置のオブジェクトに入力。ただし, 1 行目は身長で 2 行目は体重
- 身長と体重のデータについて, 以下ことを求めよ。
 - (1) ヒストグラム
 - (2) 平均と中央値
 - (3) 箱ひげ図 (学籍番号-boxplot.pdf 例: 20916***-boxplot.pdf)
 - (4) 範囲, 四分位範囲, 平均偏差, 分散, 標準偏差
- 締め切りは 2011 年 11 月 11 日 (金) 13 時
- 21016***-目白花子-101111.txt

問題 2

(a) $Q(a, b)$ を a と b について偏微分した式

$$\begin{cases} \frac{\partial Q}{\partial a} = -2 \sum_{i=1}^n \{y_i - (a + bx_i)\} = -2n\bar{y}_n + 2an + 2nb\bar{x}_n = 0 \\ \frac{\partial Q}{\partial b} = -2 \sum_{i=1}^n x_i \{y_i - (a + bx_i)\} = -2 \sum_{i=1}^n x_i y_i + 2na\bar{x}_n + 2b \sum_{i=1}^n x_i^2 = 0 \end{cases}$$

を確認せよ。また,

$$(1) \quad \begin{cases} \frac{\partial Q}{\partial a} = 0 \\ \frac{\partial Q}{\partial b} = 0 \end{cases}$$

の解が Q を最小にする a と b の値になるのはなぜか?

(b) 連立方程式 (1) の解が

$$b = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}_n \bar{y}_n}{\sum_{i=1}^n x_i^2 - n\bar{x}_n^2}, \quad a = \bar{y}_n - \frac{s_{xy}}{s_x^2} \bar{x}_n$$

(c)

$$\begin{aligned} -2 \sum_{i=1}^n \{y_i - (a + bx_i)\} &= -2n\bar{y}_n + 2an + 2nb\bar{x}_n \\ -2 \sum_{i=1}^n x_i \{y_i - (a + bx_i)\} &= -2 \sum_{i=1}^n x_i y_i + 2na\bar{x}_n + 2b \sum_{i=1}^n x_i^2 \end{aligned}$$

等号を確認せよ。

- 締め切りは 2011 年 11 月 11 日 (金) 13 時
- このレポートは A4 のレポート用紙にかき, 数研前のレポート入れに提出すること。

問題 3

- (a) 50m 走と走り幅跳びのデータを 10 に選び出し, それぞれをオブジェクト x と y に入力せよ。
- (b) オブジェクト x と y の散布図を作成せよ。
- (c) 回帰直線 $y = \hat{a} + \hat{b}x$ を求め, 散布図 (21016***-scatter.pdf) に描きいれよ。
- (d) 相関係数を計算せよ。
 - 締め切りは 2011 年 11 月 11 日 (金) 13 時
 - 21016***-目白花子-101111.txt