# データ解析演習問題

2006.10.19

提出期限: 2006 年 10 月 23 日 (月) 12:00
提出場所: システム棟 5F レポート提出 BOX

学籍番号: ____________________

氏名: ____________________

## A.

以下の表 A.1 のように, 3 変数 $x, u, y$ に関する $n$ 個のデータ与えられている. データ

表 A.1: 変数 $x, u, y$ のデータ

| | $x$ | $u$ | $y$ |
|---|---|---|---|
| 1 | $x_1$ | $u_1$ | $y_1$ |
| 2 | $x_2$ | $u_2$ | $y_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $i$ | $x_i$ | $u_i$ | $y_i$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $n$ | $x_n$ | $u_n$ | $y_n$ |

$(x_i, u_i, y_i)$ $(i = 1, \cdots, n)$ が与えられているときに, 平面 $y = a + bx + cu$ に対する残差平方和 $Q$ は,

$$Q = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \{y_i - (a + bx_i + cu_i)\}^2 \tag{A.1}$$

で定義される. $Q(a, b, c)$ を最小にする $a, b, c$ は, 次の連立方程式の解であることを前回 (2006 年 10 月 16 日) 見た.

$$\sum_{i=1}^{n} \{y_i - (a + bx_i + cu_i)\} = 0, \tag{A.2}$$

$$\sum_{i=1}^{n} x_i\{y_i - (a + bx_i + cu_i)\} = 0, \tag{A.3}$$

$$\sum_{i=1}^{n} u_i\{y_i - (a + bx_i + cu_i)\} = 0. \tag{A.4}$$

式 (A.2) によって, $\sum_{i=1}^{n} \epsilon_i = 0$ であるから, 両辺を $n$ で割って, $\bar{\epsilon} = 0$ を得る. したがって,

$$s_\epsilon^2 = \frac{1}{n}\sum_{i=1}^{n} (\epsilon_i - \bar{\epsilon})^2 = \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 = Q/n \tag{A.5}$$

である.

以下の小問に答えよ. 式の証明では, どの式をどこで用いたのかを記入すること.

## A.1.　式 (A.2) を用いて, 以下の式を証明せよ.

$$\bar{\hat{y}} = \bar{y}. \tag{A.6}$$

$$\begin{aligned}
\bar{\hat{y}} &= \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\{a + bx_i + cu_i\} \\
&= \frac{1}{n}\sum_{i=1}^{n}y_i \quad \text{((A.2) を用いた. )} \\
&= \bar{y}.
\end{aligned}$$

## A.2.　式 (A.2)〜式 (A.4) を用いて, 以下の式を証明せよ.

$$\sum_{i=1}^{n}\epsilon_i\{(a + bx_i + cu_i) - \bar{y}\} = 0. \tag{A.7}$$

$$\begin{aligned}
\sum_{i=1}^{n}\epsilon_i\{(a + bx_i + cu_i) - \bar{y}\} &= \sum_{i=1}^{n}\{\epsilon_i(a - \bar{y}) + b\epsilon_i x_i + c\epsilon_i u_i\} \\
&= (a - \bar{y})\sum_{i=1}^{n}\epsilon_i + b\sum_{i=1}^{n}\epsilon_i x_i + c\sum_{i=1}^{n}\epsilon_i u_i \\
&= (a - \bar{y}) \cdot 0 + b \cdot 0 + c \cdot 0 \quad \text{((A.2),(A.3,(A.4) を用いた. )} \\
&= 0.
\end{aligned}$$

**A.3. 式 (A.5), 式 (A.6), 及び, 式 (A.7) を用いて, 以下の式を証明せよ.**

$$s_y^2 = s_\epsilon^2 + s_{\hat{y}}^2. \tag{A.8}$$

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-\bar{y})^2 &= \textstyle\sum_{i=1}^{n}\{(y_i-\hat{y}_i)+(\hat{y}_i-\bar{y})\}^2 \\
&= \sum_{i=1}^{n}\{\epsilon_i+(\hat{y}_i-\bar{y})\}^2 \\
&= \sum_{i=1}^{n}\{\epsilon_i^2+2\epsilon_i(\hat{y}_i-\bar{y})+(\hat{y}_i-\bar{y})^2\} \\
&= \sum_{i=1}^{n}\epsilon_i^2+2\sum_{i=1}^{n}\epsilon_i(\hat{y}_i-\bar{y})+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \\
&= \sum_{i=1}^{n}\epsilon_i^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \quad \text{((A.7) を用いた. )} \\
&= \sum_{i=1}^{n}\epsilon_i^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2 \quad \text{((A.6) を用いた. )}.
\end{aligned}$$

ゆえに, $\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}\epsilon_i^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2$. この式の両辺を $n$ で割って,

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2+\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2 \tag{A.9}$$

であるが, 式 (A.5) より (A.8) を得る.

## B.

表 B.1 に示すように, 3 変数 $x, u, y$ についての $n = 10$ 個のデータが得られているとする.
前回 (2006 年 10 月 16 日) の演習問題で, 回帰方程式 $y = a + bx + cu$ の切片 $a$ 及び偏回帰係数 $b, c$ が以下のように求まった. (小数点以下 4 桁で四捨五入した. )

$$a = 5.853, b = -1.660, c = 3.246.$$

以下の文章中の空欄 $\boxed{\text{ア}}$ 〜 $\boxed{\text{オ}}$ の中に適切な <u>数値</u> を, $\boxed{\text{カ}}$ の中に適切な文を, 記入せよ.

$y$ の分散 $s_y^2$, 残差 $\epsilon$ の分散 $s_\epsilon^2$, 予測値の分散 $s_{\hat{y}}^2$ を求めよ. (小数点以下第 3 桁を四捨五入して, 小数点以下第 2 桁まで記入せよ. )

$s_y^2 =$ $\boxed{\text{ア } 1624.84}$ , $s_\epsilon^2 =$ $\boxed{\text{イ } 15.54}$ , $s_{\hat{y}}^2 =$ $\boxed{\text{ウ } 1609.30}$
.

寄与率 $R^2$, 及び, 自由度調整済み寄与率 $R^{*2}$ は, 以下のようになる. (小数点以下第 5 桁を四捨五入して, 小数点以下第 4 桁まで記入せよ. )

表 B.1: $x, u, y$ に関するデータ

| No. | $x$ | $u$ | $y$ |
|---|---|---|---|
| 1 | 94 | 58 | 38 |
| 2 | 57 | 43 | 44 |
| 3 | 78 | 47 | 32 |
| 4 | 61 | 45 | 56 |
| 5 | 66 | 43 | 32 |
| 6 | 93 | 32 | -42 |
| 7 | 71 | 36 | 2 |
| 8 | 53 | 53 | 95 |
| 9 | 88 | 30 | -42 |
| 10 | 88 | 57 | 41 |

$R^2 =$ エ 0.9904 , $R^{*2} =$ オ 0.9877

したがって, カ 回帰方程式はデータを良く表現できている と言える.

## C.

本講義「データ解析」についての感想, 要望, 質問等があれば記せ.