

データ解析演習問題

2006.10.16

提出期限: 2006年10月19日(木) 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, 2変数 x, u, y に関する n 個のデータ与えられている. 以下の文

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

章中の空欄 ~ の中に最も適切な 数式 を記入せよ.

最小2乗法は, データ (x_i, u_i, y_i) ($i = 1, \dots, n$) に最も良くあてはまる平面 $y = a + bx + cu$ を求める方法の一つである. ここで, 平面 $y = a + bx + cu$ のあてはまりの度合を測る尺度として, 残差平方和

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

が用いられる. 即ち, Q を最小にする平面 $y = a + bx + cu$ がデータ (x_i, u_i, y_i) ($i = 1, \dots, n$) に最も良くあてはまる直線であると考えるのである.

Q を3変数 a, b, c に関する関数とみて, $Q(a, b, c)$ を最小にする a, b, c を求めるには, 次の連立方程式の解を求めればよい.

$$\frac{\partial Q}{\partial a} = \text{イ} -2 \sum_{i=1}^n \{y_i - (a + bx_i + cu_i)\} = 0,$$

$$\frac{\partial Q}{\partial b} = \boxed{\text{ウ } -2 \sum_{i=1}^n x_i \{y_i - (a + bx_i + cu_i)\}} = 0,$$

$$\frac{\partial Q}{\partial c} = \boxed{\text{エ } -2 \sum_{i=1}^n u_i \{y_i - (a + bx_i + cu_i)\}} = 0.$$

a, b, c について整理すると,

$$\begin{aligned} a + \bar{x}b + \bar{u}c &= \bar{y}, \\ s_x^2 b + s_{xu}c &= s_{xy}, \\ s_{xu}b + s_u^2 c &= s_{uy} \end{aligned} \tag{A.1}$$

が得られる. 求める a, b, c は, この連立方程式 (A.1) の解である.

単回帰分析のときと同様にして, 回帰方程式が定める平面 $y = a + bx + cu$ は, $a + \bar{x}b + \bar{u}c = \bar{y}$ によって, 必ず点 $(\bar{x}, \bar{u}, \bar{y})$ を通ることがわかる.

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.

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

以下の文章中の空欄 オ ~ タ の中に適切な数値を記入せよ.

連立方程式 (A.1) は,

$$\begin{array}{l} a + \text{オ } 74.9 \\ \text{ク } 213.29 \\ \text{サ } -10.66 \end{array} \quad b + \begin{array}{l} \text{カ } 44.4 \\ \text{ケ } -10.66 \\ \text{シ } 86.04 \end{array} \quad c = \begin{array}{l} \text{キ } 25.6 \\ \text{コ } -388.74 \\ \text{ス } 296.96 \end{array},$$

となる. この連立方程式を解けば, 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る.

$$\begin{aligned} a &= \boxed{\text{セ } 5.853}, \\ b &= \boxed{\text{ソ } -1.660}, \\ c &= \boxed{\text{タ } 3.246}. \end{aligned}$$

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.