

# データ解析演習問題

2006.10.12

提出期限: 2006年10月16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, 2変数  $x, y$  に関する  $n$  個のデータ与えられている. 以下の文章

表 A.1: 変数  $x, y$  のデータ

	$x$	$y$
1	$x_1$	$y_1$
2	$x_2$	$y_2$
$\vdots$	$\vdots$	$\vdots$
$i$	$x_i$	$y_i$
$\vdots$	$\vdots$	$\vdots$
$n$	$x_n$	$y_n$

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

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

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

$Q$  を2変数  $a, b$  に関する関数とみて,  $Q(a, b)$  を最小にする  $a$  と  $b$  を求めるには, 次の連立方程式の解を求めればよい.

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

$a, b$  について整理すると,

$$\begin{aligned} a + \bar{x}b &= \bar{y}, \\ s_x^2 b &= s_{xy} \end{aligned}$$

が得られる. この連立方程式を解いて,

$$\begin{aligned} a &= \boxed{\text{エ } \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x}}, \\ b &= \boxed{\text{オ } \frac{s_{xy}}{s_x^2}} \end{aligned}$$

を得る.

したがって, データ  $(x_i, y_i)$  ( $i = 1, \dots, n$ ) に最も良くあてはまる直線 (= 回帰方程式) は,

$$\boxed{\text{カ } y = \bar{y} + \frac{s_{xy}}{s_x^2} (x - \bar{x})}$$

である. この直線は必ず  $(\bar{x}, \bar{y})$  を通ることがわかる.

## B.

表 B.1 に示すような 2 変数  $x, y$  についての  $n = 10$  個のデータが得られているとする. 以下の文章中の空欄  ~  の中に適切な 数値 を記入せよ.

表 B.1:

No	$x$	$y$
1	69	-147
2	43	-76
3	32	-53
4	40	-72
5	68	-141
6	43	-79
7	38	-58
8	49	-95
9	34	-53
10	62	-124

変数  $x$  の平均  $\bar{x}$  と変数  $y$  の平均  $\bar{y}$  は,

$$\bar{x} = \boxed{\text{キ } 47.8},$$

$$\bar{y} = \boxed{\text{ク } -89.8},$$

となり, 変数  $x$  の分散  $s_x^2$  と変数  $y$  の分散  $s_y^2$  は,

$$s_x^2 = \text{ケ } 170.36,$$

$$s_y^2 = \text{コ } 1143.36$$

となる. また,  $x$  と  $y$  の共分散  $s_{xy}$  は,

$$s_{xy} = \text{サ } -439.66$$

である. ゆえに, 回帰方程式  $y = a + bx$  の  $y$  切片  $a$  及び傾き  $b$  は以下の通りである.

$$a = \text{シ } 33.56,$$

$$b = \text{ス } -2.58.$$

### 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.