

# データ解析演習問題

2006.10.05

提出期限: 2006年10月10日(火)17:00

提出場所: システム棟5F レポート提出BOX

A.

A.1.

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.

表 A.1: 変数  $x, y$  のデータ

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

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

$$\bar{x} = (1) \frac{1}{n} \sum_{i=1}^n x_i,$$

$$\bar{y} = (2) \frac{1}{n} \sum_{i=1}^n y_i$$

と定義され, 変数  $x$  の分散  $s_x^2$  と変数  $y$  の分散  $s_y^2$  は,

$$s_x^2 = (3) \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$s_y^2 = (4) \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

と定義される. また,  $x$  と  $y$  の共分散  $s_{xy}$ , 及び, 相関係数  $r_{xy}$  は

$$s_{xy} = (5) \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

$$r_{xy} = \boxed{(6) \frac{s_{xy}}{s_x s_y}}$$

と定義される.

## A.2.

$x_i, y_i$  ( $i = 1, \dots, n$ ) が以下の表 A.2 のように与えられているとする. 問題 A(1) の答えに

表 A.2: 変数  $x, y$  のデータ

$i$	$x_i$	$y_i$
1	53	71
2	56	75
3	47	63
4	15	22
5	53	71
6	55	74
7	76	101
8	12	18
9	39	53
10	41	56

基づいて,  $\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy}, r_{xy}$  を計算すると

$$\begin{aligned} \bar{x} &= \boxed{(7) 44.7}, \\ \bar{y} &= 60.4, \\ s_x^2 &= \boxed{(8) 335.41}, \\ s_y^2 &= 562.44, \\ s_{xy} &= \boxed{(9) 434.32}, \\ r_{xy} &= \boxed{(10) 0.999963} \end{aligned}$$

となる.

したがって, 変数  $x$  と変数  $y$  の間には,  $\boxed{(11) \text{正の強い相関がある}}$  といえる.

## B.

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.