

## 問5. 以下の問いに答えよ。(5-1~5-3: 計4点、5-4: 各2点)

(i)-(iv)の4セットのデータ  $(x, y)$  に対して、Pearson相関係数の計算と線形単回帰分析を行ったところ、回帰直線  $y = a_0 + a_1 \cdot x$  のパラメータ  $a_0, a_1$ 、決定係数  $r^2$ 、 $X$  と  $Y$  の間のPearson相関係数  $\text{corr}$  が、それぞれ以下のように得られた。

(i)  $a_0: 0.04, a_1: -0.26, r^2: 0.77, \text{corr}: -0.87$

(ii)  $a_0: 1.73, a_1: 0.74, r^2: 0.49, \text{corr}: 0.70$

(iii)  $a_0: -10.44, a_1: -0.49, r^2: 0.15, \text{corr}: -0.39$

(iv)  $a_0: 6.85, a_1: 0.00, r^2: 0.01, \text{corr}: 0.03$

5-1.  $X$  と  $Y$  の間の関係がもっとも強いといえるデータセットはどれか。

☐ (i) ☐ (ii) ☐ (iii) ☐ (iv)

5-2.  $X$  の値が増加すると  $Y$  の値が減少する傾向にあるデータセットはどれか。該当するものをすべて挙げる

☐ (i) ☐ (ii) ☐ (iii) ☐ (iv)

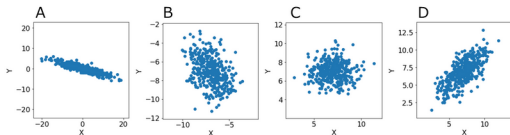
5-3. 次の散布図 A-D は、(i)-(iv) のいずれかのデータ点を可視化したものである。それぞれ、(i)-(iv) のどれに対応するか答えよ。

A  
(未選択)

B  
(未選択)

C  
(未選択)

D  
(未選択)



5-4. 以下の操作を行い、問いに答えよ。

- cs3-mid-p5.csv の全データを読み込む。
- $x$  列を説明変数、 $y$  列を目的変数にする。
- 線形単回帰分析を行い、回帰式  $y = a_0 + a_1 \cdot x$  を求める。 $a_0$  と  $a_1$  の値を小数第3位まで求めよ。
- $x$  の値が 4.12 であるとき、 $y$  の予測値、および実際の  $y$  との差 (予測値が大きいときは正、予測値が小さいときは負の値とせよ) をそれぞれ 5-4. (1), (2) に小数第3位まで答えよ。 ※「実際の  $y$ 」は、1の入力データの中で説明変数  $x$  が 4.12 であるデータの  $y$  とすること。

5-3.  $a_0$

5-3.  $a_1$

5-4. (1)

5-4. (2)

ipynbファイルのアップロード:

未提出

htmlファイルのアップロード:

未提出