

# Data analysis 1

A	B	C	D	E	F	G	H	I	J
14	10	13	13	15	18	6	10	11	10

表1 数学20点満点小テスト

3) 最大値

$$\underline{18}$$

4) 最小値

$$\underline{6}$$

5) 平均値

$$\frac{1}{10} \times (14 + 10 + 13 + 13 + 15 + 18 + 6 + 10 + 11 + 10)$$

$$= \frac{1}{10} \cdot 120 = \underline{12}$$

6) 中央値 (メジアン)

6 10 10 10 11 13 13 14 15 18

$$\text{中央値} = \frac{11 + 13}{2} = \underline{12}$$

7) 最頻値

$$\underline{10}$$

8) 範囲

$$18 - 6 = \underline{12}$$

9) 四分位範囲

6 10 10 10 11 | 13 13 14 15 18

$Q_1$                        $Q_2$                        $Q_3$

$$\text{四分位範囲} = Q_3 - Q_1$$

$$= 14 - 10 = \underline{4}$$

11) 外れ値

... 他のデータ群から極端に外れた値

ex.

0 6 10 10 10 11 13 13 14 15 18

$Q_1$                        $Q_2$                        $Q_3$

外れ値の基準例

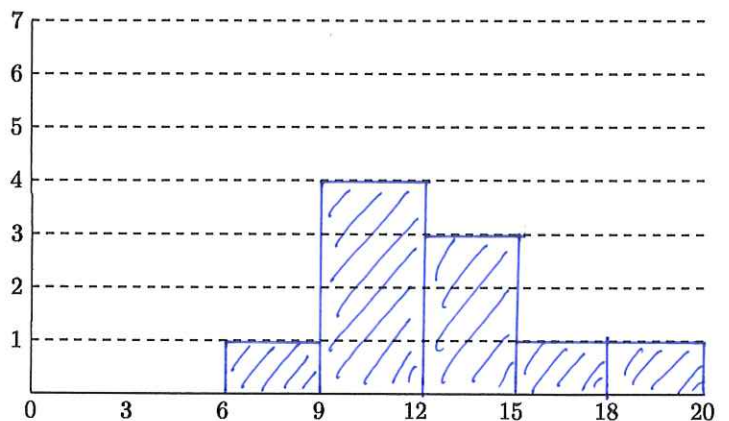
- $(Q_1 - 1.5 \times \text{四分位範囲}) \text{以下} \rightarrow 10 - 1.5 \times 4 = 6 \text{以下}$
- $(Q_3 + 1.5 \times \text{四分位範囲}) \text{以上} \rightarrow 14 + 1.5 \times 4 = 20 \text{以上}$

外れ値の背景を探ると、問題の発見や問題解決につながる

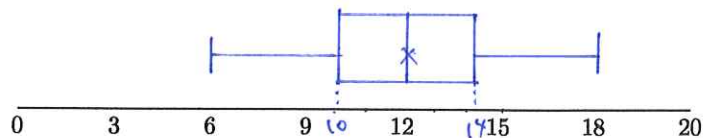
1) 度数分布表

階級	度数
18 以上 20 以下	1
15 以上 18 未満	1
12 以上 15 未満	3
9 以上 12 未満	4
6 以上 9 未満	1
3 以上 6 未満	0
0 以上 3 未満	0

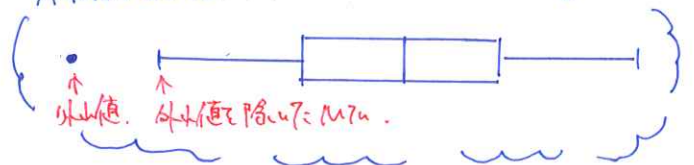
2) ヒストグラム



10) 箱ひげ図



外れ値の箱ひげ図



## Data analysis 2

A	B	C	D	E	F	G	H	I	J
17	9	12	10	12	15	5	11	9	10

表2 英語 20 点満点小テスト

3) 最大値

17

4) 最小値

5

5) 平均値

$$\frac{1}{10} (17 + 9 + 12 + 10 + 12 + 15 + 5 + 11 + 9 + 10)$$

$$= \frac{1}{10} \cdot 110 = \underline{11}$$

6) 中央値 (メジアン)

5 9 9 10 10 <sup>^</sup> 11 12 12 15 17

$$\text{中央値} = \frac{10 + 11}{2} = \underline{10.5}$$

7) 最頻値

9, 10, 12

8) 範囲

$$17 - 5 = \underline{12}$$

9) 四分位範囲

$$Q_1 = 9, \quad Q_3 = 12$$

$$\therefore Q_3 - Q_1 = 12 - 9 = \underline{3}$$

11) 外れ値の有無.

基準.

$$\cdot Q_1 - 1.5(Q_3 - Q_1) = (9 - 4.5) \times 1.5$$

i.e. 4.5  $\times 1.5 \rightarrow 6.75$

$$\cdot Q_3 + 1.5(Q_3 - Q_1) = (12 + 4.5) \times 1.5$$

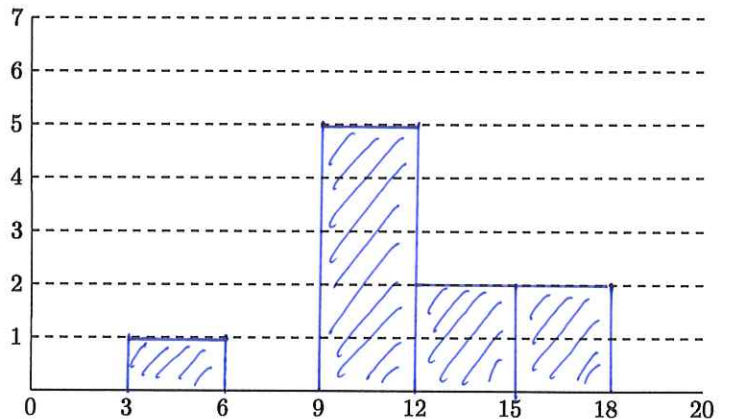
i.e. 16.5  $\times 1.5 \rightarrow 24.75$

17 は 外れ値

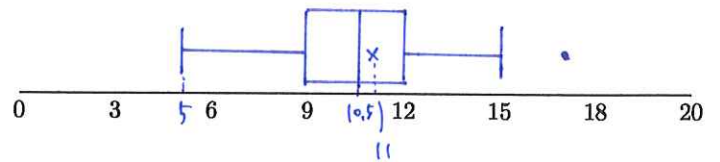
1) 度数分布表

階級	度数
18 以上 20 以下	
15 以上 18 未満	2
12 以上 15 未満	2
9 以上 12 未満	5
6 以上 9 未満	0
3 以上 6 未満	1
0 以上 3 未満	0

2) ヒストグラム



10) 箱ひげ図



# Data analysis 3

平均をもとに、散らばり具合を調べたい!

	$x$	$x - \bar{x}$	$(x - \bar{x})^2$
A	14	2	4
B	10	-2	4
C	13	1	1
D	13	1	1
E	15	3	9
F	18	6	36
G	6	-6	36
H	10	-2	4
I	11	-1	1
J	10	-2	4
計	120	0	100

$\bar{x}$ : 平均値  $\bar{x} = \frac{1}{10} \times 120 = 12$

偏差・分散・標準偏差

偏差... 各値から平均を引いたもの  
 $x_i - \bar{x}$

分散... 偏差の2乗和の平均

$$V = \frac{1}{n} \{ (x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}$$

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

標準偏差...  $\sigma = \sqrt{10}$

計算してみよう。

1) 分散

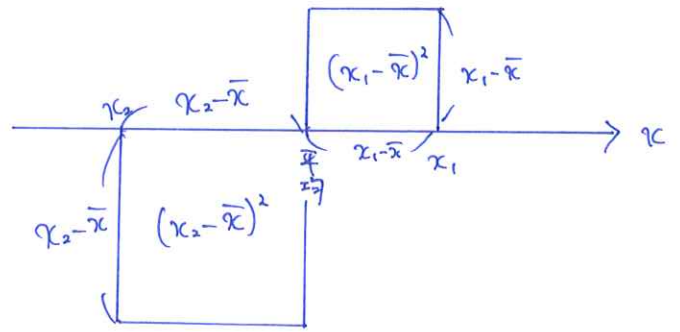
$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 100 \text{ だと}$$

$$V = \frac{1}{10} \times 100 = 10$$

2) 標準偏差

$$\sigma = \sqrt{10}$$

## 偏差・分散・標準偏差のイメージ



平均化

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

分散

標準偏差

- 分散... 面積の平均
- 標準偏差... 平均の正偏差の辺の長さ

分散と平均の関係を調べてみよう。

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

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

$$= \frac{1}{n} \left( \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \right)$$

$$= \frac{1}{n} \left( \sum_{i=1}^n x_i^2 \right) - 2\bar{x} \cdot \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} n \bar{x}^2$$

$$= \frac{1}{n} \left( \sum_{i=1}^n x_i^2 \right) - (\bar{x})^2$$

分散と平均の関係式

$$(\text{分散}) = \left( \frac{\text{各々の}^2}{2\text{乗の平均}} \right) - (\text{平均})^2$$

# Data analysis 4

復習として、授業で行ったことを再現してみよう。

	$y$	$y - \bar{y}$	$(y - \bar{y})^2$
A	17	6	36
B	9	-2	4
C	12	1	1
D	10	-1	1
E	12	1	1
F	15	4	16
G	5	-6	36
H	11	0	0
I	9	-2	4
J	10	-1	1
計	110	0	100

$$\bar{x} = 11$$

偏差・分散・標準偏差

計算してみよう。

1) 分散

$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 100$$

$$\begin{aligned} (\text{分散}) &= \frac{1}{10} \cdot 100 \\ &= 10 \end{aligned}$$

2) 標準偏差

$$\sigma = \sqrt{10}$$

偏差・分散・標準偏差のイメージ

分散と平均の関係を調べてみよう。

$$s^2 =$$

=

=

=

=

分散と平均の関係式



# Data analysis 5

2つの変量の間の関係を調べたい!

	数(x)	英(y)	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x}) \times (y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
A	14	17	2	6	12	4	36
B	10	9	-2	-2	4	4	4
C	13	12	1	1	1	1	1
D	13	10	1	-1	-1	1	1
E	15	12	3	1	3	9	1
F	18	15	6	4	24	36	16
G	6	5	-6	-6	36	36	36
H	10	11	-2	0	0	4	0
I	11	9	-1	-2	2	1	4
J	10	10	-2	-1	2	4	1
計	120	110	0	0	83	100	100

$\bar{x} = 12, \bar{y} = 11$

① 偏差の計算

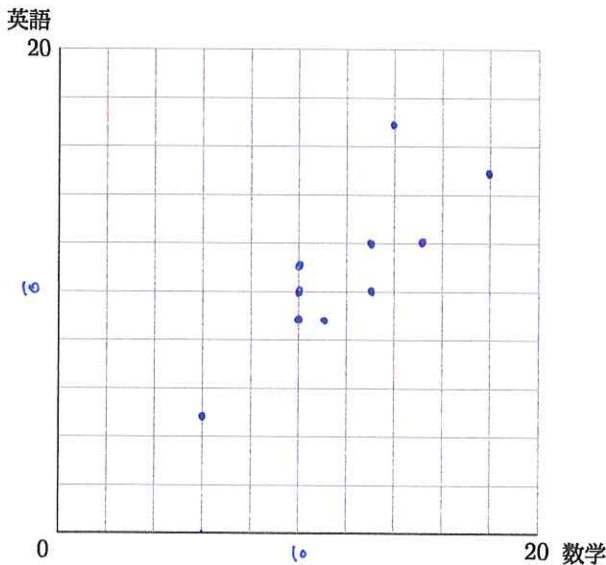
② 分散  $\sigma_x^2 = 10$

$\sigma_y^2 = 10$

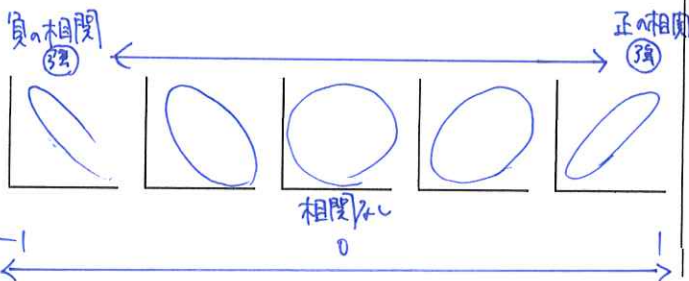
③ 標準偏差

$\sigma_x = \sqrt{10}, \sigma_y = \sqrt{10}$

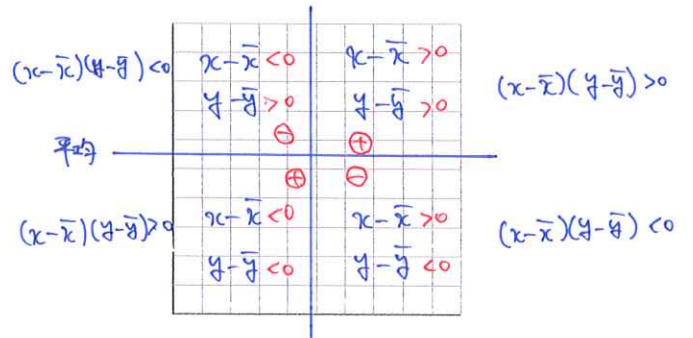
1) 散布図 を描いてみる。



2) 散布図 からみる 相関



2つの変量間の関係を数値で評価したい!



右上、左下 傾きが正のとき → 正の相関  
右下、左上 " " → 負の相関

### 共分散と相関係数

共分散 → xとyの偏差積の平均

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

相関係数

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$r < -1$  近い → 負の相関 強い  
 $r < 1$  近い → 正の相関 強い

計算してみよう。

3) 共分散

$$\sigma_{xy} = \frac{1}{10} \times 83 = 8.3$$

4) 相関係数

$$r = \frac{8.3}{\sqrt{10} \cdot \sqrt{10}} = 0.83$$

因果関係に??

... 原因が他を起すという関係

ex.

気温が高いからアイスが売れた。  
等等。

※) 気をつけて  
「相関」 → 「因果」 + 逆はなし!!



$$\text{共分散} = \frac{1}{8} (0 + (-6) + 10 + 0 + (-1) + (-9) + 0)$$

$$= -2$$

令和4年度第1学年(3組)2学期期末考査 数学② 表

R4. 12.6

解答は全て、解答欄に記入すること(欄外は採点しない)。

1 変数  $x$  のデータが、次のように与えられている。

24, 8, 11, 15, 26, 13, 12, 15, 11, 2, 17

以下の問いに答えよ。(30点)

- 解答欄の表に合わせて、度数分布表をかけ。
- 解答欄のグラフに合わせて、ヒストグラムをかけ。
- (1) で作成した 度数分布表 において、最頻値を求めよ。

以下の問題は、元データをもとに答えよ。

- 最大値を求めよ。
- 平均値を求めよ。

$$154 \div 11 = 14$$

(6) 中央値を求めよ。

(7) 範囲を求めよ。  
26 - 2 = 24

(8) 四分位範囲を求めよ。

$$17 - 11 = 6$$

(9) 箱ひげ図を描け。ただし、外れ値がある場合には、 $\circ$  で表すこと。  
(外れ値の基準は下の通り)

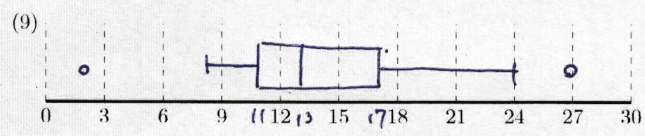
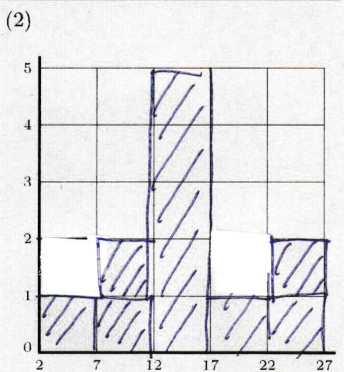
$$Q_1 - 1.5 \times (\text{四分位範囲}) \text{ 以下の値}$$

$$Q_3 + 1.5 \times (\text{四分位範囲}) \text{ 以上の値}$$

【解答欄】

1	(1)	下図に記入	(2)	下図に記入	(3)	14.5
	(4)	26	(5)	14	(6)	13
	(7)	24	(8)	6	(9)	下図に記入

階級	度数
22以上27未満	2
17以上22未満	1
12以上17未満	5
7以上12未満	2
2以上7未満	1
計	11



1年 組 番

氏名 NO.1

2 次の表は、8人の休日の携帯の使用時間と勉強時間を調査した結果である。

	A	B	C	D	E	F	G	H
携帯(時間)	3	7	0	5	6	7	2	10
勉強(時間)	6	3	8	5	5	4	9	8

計 平均  
40 5  
48 6

以下の問いに答えよ。(30点)

(1) 携帯使用時間について、分散を求めよ。

$$s^2 = \frac{1}{8} (4 + 4 + 25 + 0 + 1 + 4 + 9 + 25) = \frac{1}{8} \cdot 72 = 9$$

(2) 携帯使用時間について、標準偏差を求めよ。

$$s = \sqrt{9} = 3$$

(3) 相関係数を求めよ。ただし、小数第3位を四捨五入せよ。

$$r = \frac{-2}{3 \cdot 2} = -\frac{1}{3} = -0.3333$$

(4) (3) の計算結果から、携帯使用時間と勉強時間の間には、どのような相関があると考えられるか。

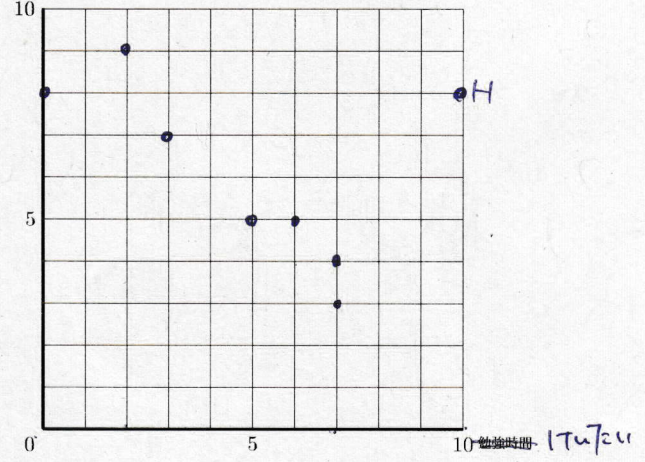
(5) 散布図を描け。

(6) 計算した相関係数と散布図から読み取れることを記述せよ。

【解答欄】

2	(1)	9	(2)	3	(3)	-0.33
	(4)	弱い負相関			(5)	下図に記入
	(6)	Hを除く。強い負相関が見て取れる。				

(5) 勉強時間



X:知理	+	Y:思	=	合計
------	---	-----	---	----



- 3 A, Bの2グループの計50人全員が、1問1点の10問のクイズに答えた。下の表は、その正答数の結果である。

グループ	人数	平均得点	分散
A	30	4	8
B	20	9	18

以下の問いに答えよ。(12点)

- (1) AとB全体の得点について、平均値を求めよ。

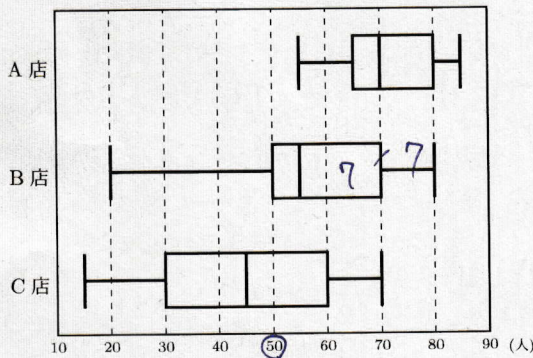
$$(30 \times 4 + 20 \times 9) \div 50 = 300 + 180 = 480 \div 50 = 9.6$$

- (2) AとB全体の得点について、分散を求めよ。

$$(30 \times 2 + 20 \times 18) \div 50 = 240 + 360 = 600 \div 50 = 12$$

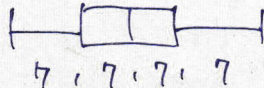
- (3) 全員の得点を10倍して、100点満点で評価する。このとき、AとB全体の得点の分散を求めよ。

- 4 次の図は、A店、B店、C店の1日の入店者数を31日間調べたデータを、箱ひげ図に表したものである。



以下の問いに答えよ。(12点)

- (1) 1日の入店者数が50人を超えた日が16日以上あったのはどの店か。全て答えよ。



- (2) 1日の入店者数が50人以下となる日が8日以上あったのはどの店か。全て答えよ。

- (3) B店において、1日の入店者数が60人を超えたのは、最大で何日あった可能性があるか。

【解答欄】

3	(1)	6	(2)	A	(3)	200
4	(1)	A, B	(2)	B, C	(3)	15人

- 5 以下の問いに当てはまるものを、記号で答えよ。(16点)

- (1) 分散と平均値の関係について、aからdの中から1つ選べ。

( $s^2$ : 分散,  $\bar{x}$ : 平均とする.)

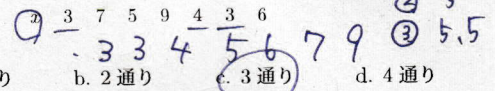
a.  $s^2 = \frac{1}{n}(x_1 + \dots + x_n)^2 - (\bar{x})^2$

b.  $s^2 = \frac{1}{n}\{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\} - (\bar{x})^2$

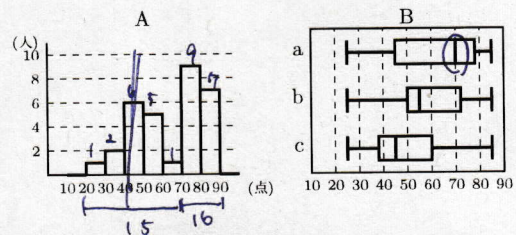
c.  $s^2 = \frac{1}{n}(x_1^2 + \dots + x_n^2) - (\bar{x})^2$

d.  $s^2 = (\bar{x})^2 - \frac{1}{n}\{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\}$

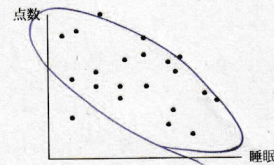
- (2)  $x$ を0以上の整数とする。次のデータに対して、中央値として何通りの値がありえるか。aからdの中から1つ選べ。



- (3) 下の図Aは、31人の生徒の数学のテストの得点をヒストグラムにしたものである。ただし、各階級は0点以上10点未満のように区切っている。このデータを箱ひげ図にまとめたとき、最も当てはまるものを、図Bのaからcのうち1つ選べ。



- (4) 下の図は、数学の点数と睡眠時間のデータについての散布図である。ただし、重なっている点はないものとする。



相関係数の値として最も適切な値をaからfの中から1つ選べ。

- a. 1.5 b. 0.95 c. 0.33 d. -0.33 e. -0.95 f. -2

- (5) ある会社では、既に販売しているペンAを改良したペンBを開発した。書きやすさの評価のために、無作為に選んだ20人に、AとBのどちらが書きやすいかアンケートを行ったところ、15人がBと回答した。この結果からの消費者からの評価として最も適するものをaからeから選べ。ただし、基準となる確率を0.05とし、以下の公正なコイン投げ200セット(1セット20回)の結果を利用して考察せよ。

表の枚数	5	6	7	8	9	10	11	12	13	14	15	16	計
度数	4	10	15	19	27	33	29	26	21	12	3	1	200

- a. 「Aの方が書きやすいと評価されている」と判断してよい。  
 b. 「Bの方が書きやすいと評価されている」と判断してよい。  
 c. 「どちらも書きやすいと評価されている」と判断してよい。  
 d. 「どちらも書きにくいと評価されている」と判断してよい。  
 e. 評価できない。

【解答欄】

5	(1)	C	(2)	c	(3)	a	(4)	d	(5)	/
---	-----	---	-----	---	-----	---	-----	---	-----	---