

1 節 データの整理と分析

①データの整理

[データ1 平成 21 年 9 月の大阪の最高気温<日付順>(単位 °C)]

31.5	33.4	31.2	32.9	34.0	34.5	33.2	31.3	27.8	28.0	30.8
25.3	28.1	27.6	22.6	28.2	29.5	27.6	28.5	28.4	30.0	27.7
29.1	31.5	31.1	30.8	31.5	27.7	26.0	22.0			

このような資料を（¹ ）という。
また、気温、湿度、降水量のように、ある特性を数量的に表すものを（² ）という。

◆度数分布表

- (³)：データを整理したときの区間
- (⁴)：各階級に含まれるデータの値の個数
- (⁵)：各階級の端点の平均
- (⁶)：各階級の端点の差
- (⁷)：各階級に度数を対応させたもの
- (⁸)：度数分布を表にしたもの

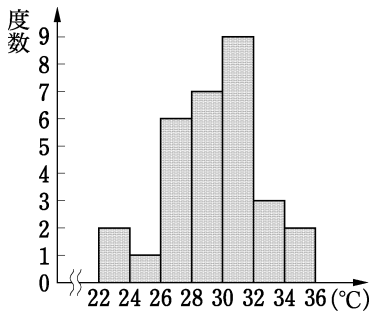
平均値 31 は (⁵)
差 2 は (⁶)

気温の(³) ℃以上 ℃未満	(⁴)
22～24	2
24～26	1
26～28	6
28～30	7
30～32	9
32～34	3
34～36	2
計	30

◆度数分布のグラフ

度数分布を表す右のようなグラフを
(⁹)という。

階級の幅が一定のとき、長方形の高さが
度数を表す。



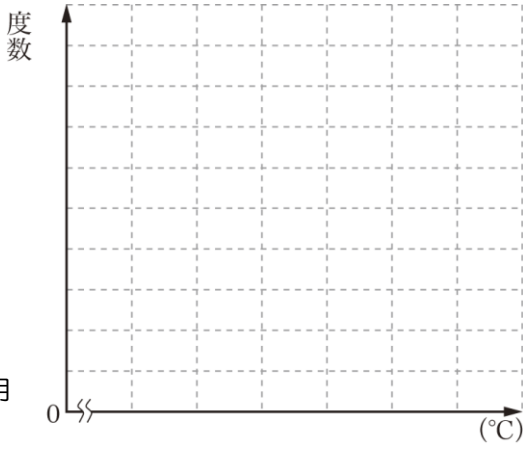
問 1 次のデータは、平成 21 年 9 月の東京の最高気温を気温の低い順に横に並べたものである。
22℃以上 24℃未満という階級から、幅を一定にとって順次階級をつくり、このデータの度
数分布表を作成せよ。また、そのヒストグラムをかけ。

【データ2 平成 21 年 9 月の東京の最高気温<気温の低い順> (単位 °C)】

23.2	23.3	23.4	23.7	23.8	23.9	24.0	24.4	24.5	25.1
25.7	25.8	26.2	26.4	26.5	26.5	26.9	26.9	27.6	27.8
28.1	28.3	28.5	28.5	28.8	28.8	28.9	29.0	29.3	31.5

気温の階級 ℃以上 ℃未満	階級値	度数		
22～24				
～				
～				
～				
～				
計				

問 3 で使用 問 2 で使用 問 3 で使用



◆相対度数

(¹⁰)：それぞれの度数を度数の
合計で割った値
 $(^{10}) = \frac{(^{11})}{(^{12})}$

右の表で、たとえば、26℃以上 28℃未満の階級の
相対度数は

$\frac{6}{30} = 0.2$

気温の階級 ℃以上 ℃未満	度数	相対 度数
22～24	2	0.07
24～26	1	0.03
26～28	6	0.20
28～30	7	0.23
30～32	9	0.30
32～34	3	0.10
34～36	2	0.07
計	30	1.00

問 2 問 1 で作成した度数分布表において、各階級の相対度数を求めよ。

②データの代表値

(¹) : データの特徴を表す数値。
例) 平均値, 中央値, 最頻値 など

◆平均値

n 個の値からなるデータがある。このデータの個々の値を次のように表す。

$x_1, x_2, \dots, x_n$

(²) : n 個の値からなるデータの総和を n で割った値。 $\bar{x}$ で表す。

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

平均値 = $\frac{\text{データの値の総和}}{\text{データの値の個数}}$

例1 教科書 140 ページのデータ 1 の平均値は

$$\frac{31.5 + 33.4 + \dots + 22.0}{30} = \frac{881.8}{30} \div 29.4 \text{ (}^\circ\text{C)}$$

◆度数分布から求めた平均値

度数分布が与えられたとき、階級値 x と度数 f の積 xf を求めて、その総和を度数の合計で割ったものを、度数分布から求めた平均値という。

例2 例1で求めたデータ1の平均値を
度数分布表から求めてみよう。
度数分布表に積 xf の欄をつけ加えると
右のようになるから、
求める値は

$$\frac{(\text{³})}{(\text{⁴})} \div (\text{⁵}) \text{ }^\circ\text{C}$$

階級 ℃以上 ℃未満	階級値 x	度数 f	xf
22～24	23	2	46
24～26	25	1	25
26～28	27	6	162
28～30	29	7	203
30～32	31	9	279
32～34	33	3	99
34～36	35	2	70
計		30	884

問3 問1で作成した度数分布表から、平均値を小数第2位を四捨五入して求めよ。

◆中央値、最頻値

(⁶) : データの値を小さい順に並べたとき、中央の順位にある数値。
(⁷) ともいう。

※ただし、データの値の個数が偶数、すなわち $2n$ 個のときは、第 n 番目と第 $n+1$ 番目の数値の
平均値を中央値とする。

例3 データ1の中央値を求めてみよう。
データの値を小さい順に並べかえると

22.0	22.6	25.3	26.0	27.6	27.6	27.7	27.7	27.8	28.0
28.1	28.2	28.4	28.5	29.1	29.5	30.0	30.8	30.8	31.1
31.2	31.3	31.5	31.5	31.5	32.9	33.2	33.4	34.0	34.5

$$\text{中央値は } \frac{(\text{⁸}) + (\text{⁹})}{2} = (\text{¹⁰}) \text{ (}^\circ\text{C)}$$

(¹¹) : 度数分布表で、度数がもっとも多い階級の階級値。
(¹²) ともいう。

例4 データ1の最頻値を求めてみよう。
右の度数分布表で、度数がもっとも多い階級は
「(¹³) 以上 (¹⁴) 未満」なので
最頻値は (¹⁵) である。

気温の階級 ℃以上 ℃未満	度数
22～24	2
24～26	1
26～28	6
28～30	7
30～32	9
32～34	3
34～36	2
計	30

問4 問1のデータ2の中央値を求めよ。また、問1で作成した度数分布表から最頻値を求めよ。

【データ2 平成21年9月の東京の最高気温<気温の低い順> (単位 ℃)】

23.2	23.3	23.4	23.7	23.8	23.9	24.0	24.4	24.5	25.1
25.7	25.8	26.2	26.4	26.5	26.5	26.9	26.9	27.6	27.8
28.1	28.3	28.5	28.5	28.8	28.8	28.9	29.0	29.3	31.5

中央値は (¹⁶)

最頻値は (¹⁷)

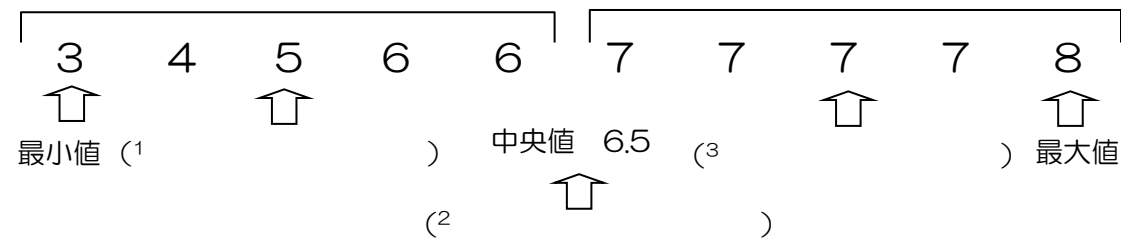
③データの散らばり

A, Bの2人で10点満点の的当てゲームを10回行い、成績は右の表のようになった。

回	1	2	3	4	5	6	7	8	9	10
Aの得点	7	3	6	4	8	7	7	5	7	6
Bの得点	9	7	2	8	6	3	4	7	9	5

◆四分位数

上の的当てゲームにおけるAの得点を小さい順に並べかえる。



中央値を境にして2つに分けたとき,

最小値を含む方の 5 個のデータの中央値 5 点を (1) という。

中央値 6.5 点を (2) という。

最大値を含む方の5個のデータの中央値7点を(3))という。

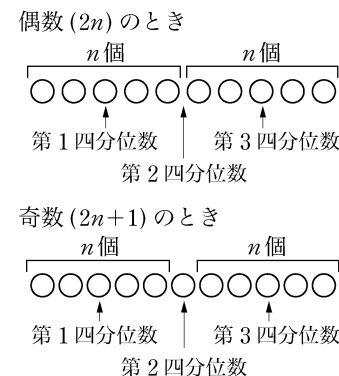
これらを合わせて (4) という。

※ 四分位数はデータを小さい順に並べて、25%ずつ4つの部分に分けた値といえる。

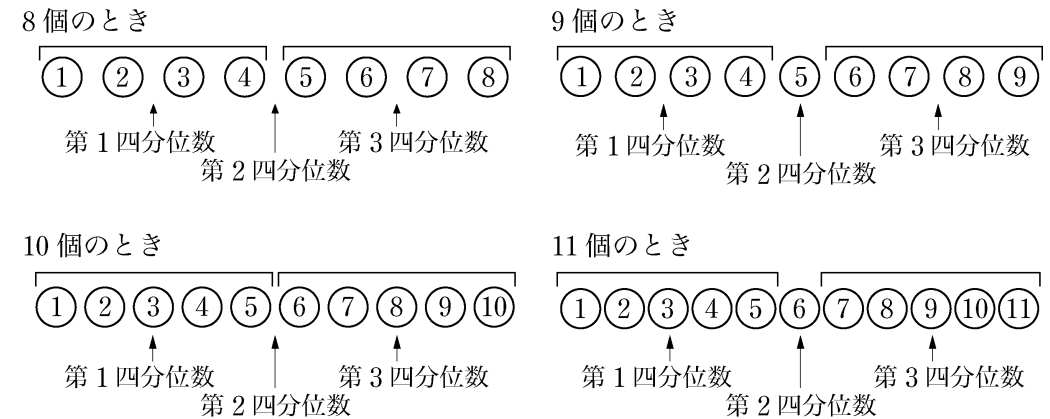
一般に、四分位数は、データの値の個数が偶数($2n$)、奇数($2n+1$)のいずれのときも、次のようになる。

データの値を小さい順に並べかえて、右の図のように、中央値を境にして2つの部分に分ける。

- ① 最小値を含む方の n 個のデータの中央値が第1四分位数である。
- ② 中央値が第2四分位数である。
- ③ 最大値を含む方の n 個のデータの中央値が第3四分位数である。



※四分位数の求め方は、データの値の個数によって次の4通りがある。



例5 的当てゲームにおけるBの得点を小さい順に並べかえると



第1四分位数は（⁵）点

第2四分位数は（⁶）点

第3四分位数は（7 ）点

問5 次のデータはあるクラスのA班、B班の1か月の読書時間である。A班、B班の四分位数をそれぞれ求めよ。

【A班 20人（単位 時間）】

【B班 15人（単位 時間）】

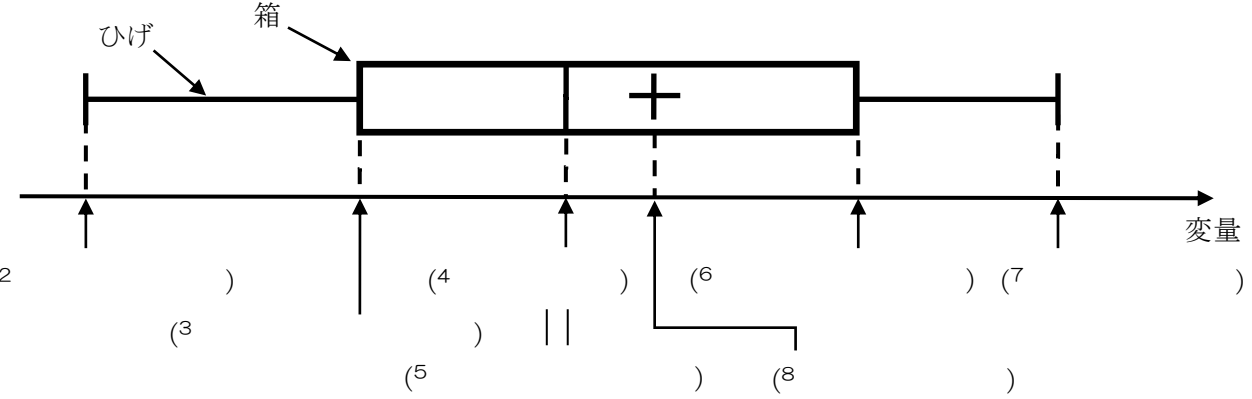
3	10	7	14	5	9	15	0	13	18
0	8	11	10	15	19	6	23	9	5

20	6	0	14	16	23	1	4	5	0
18	13	21	0	9					

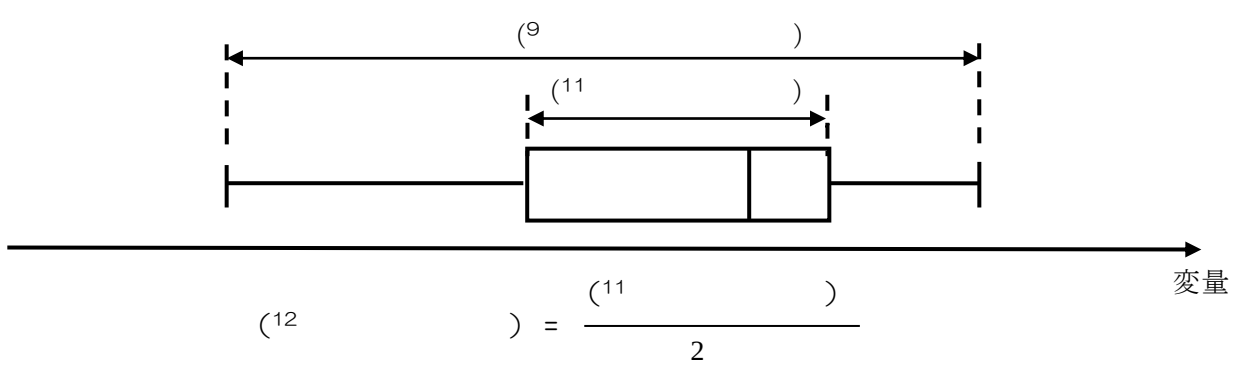
	第 1 四分位数	第 2 四分位数	第 3 四分位数
A 班			
B 班			

◆箱ひげ図

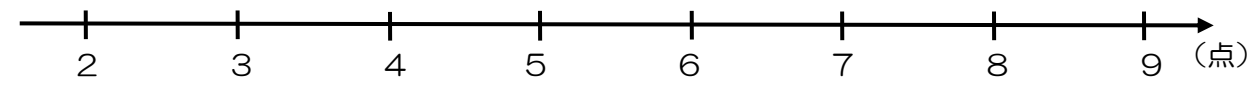
(1) : 最小値, 第1四分位数, 中央値(第2四分位数), 第3四分位数, 最大値を箱と線(ひげ)を用いて表した図。データの分布を表す。



(9) : データの最大値から最小値を引いた値。(10) ともいう。
(11) : 第3四分位数から第1四分位数を引いた値。
(12) : 四分位範囲を2で割った値。



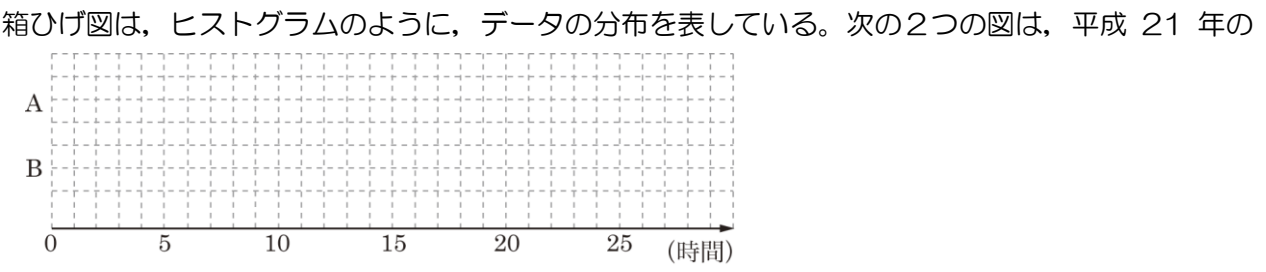
例6 的当てゲームにおけるAの得点の箱ひげ図をかいてみよう。



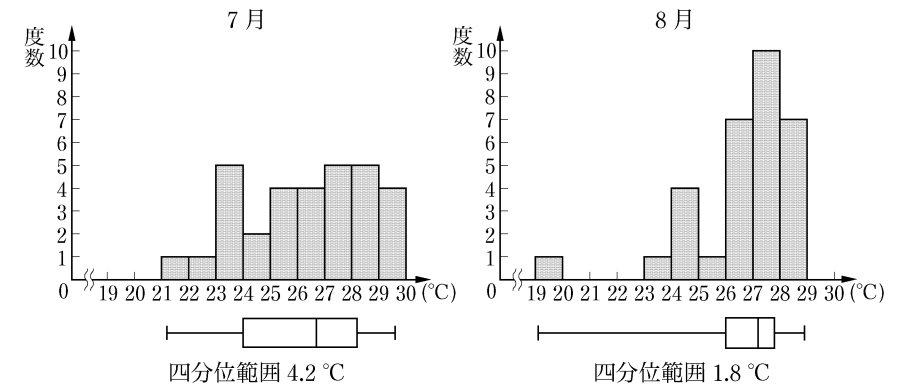
このとき, 範囲は (13), 四分位範囲は (14), 四分位偏差は (15)

的当てゲームにおけるBの得点の箱ひげ図は A 右の通りである。Aの得点の箱ひげ図と比べることによって, データの分布の違いがわかる。 B

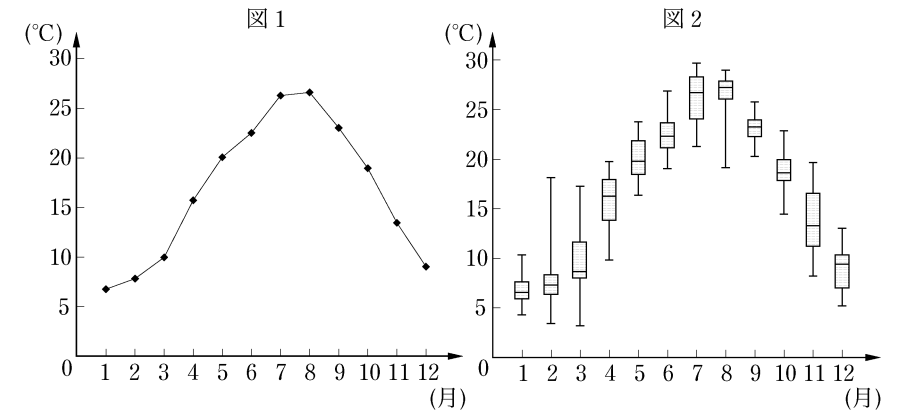
問6 問5のA班, B班について, それぞれの四分位偏差を求めよ。また, 箱ひげ図をかき, 並べよ。



箱ひげ図は, ヒストグラムのように, データの分布を表している。次の2つの図は, 平成 21 年の東京における, 7月と8月の1日の平均気温の分布をヒストグラムと箱ひげ図で表している。ヒストグラムと箱ひげ図を比較すると, データの分布を同じように表していることがわかる。



下の図1は各月の1日の平均気温の平均, 図2は各月の1日の平均気温の分布の箱ひげ図のグラフである。箱ひげ図を用いると, 最高値, 最低値, 中央値, 四分位範囲などを比較することができるようになる。



◆分散と標準偏差

データの散らばり具合を表す数値として以下の値を用いる。
データの個々の値が $x_1, x_2, \dots, x_n$ であり、その平均値を $\bar{x}$ とする。

平均値からの (1) あるいは単に (1) : 個々の値と平均値との差 $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$

(2) : 偏差の2乗の平均値 $s^2 = \frac{1}{n} \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}$

(3) : 分散の正の平方根 $s = \sqrt{\frac{1}{n} \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}}$

Q&A

Q：散らばり具合を調べるのに、なぜ偏差の平均値を用いないのか？
A：偏差の平均値を実際に計算すると
$$\frac{1}{n} \{ (x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) \} = \frac{1}{n} \{ (x_1 + x_2 + \dots + x_n) - n\bar{x} \}$$
$$= \frac{1}{n} (x_1 + x_2 + \dots + x_n) - \bar{x} = \bar{x} - \bar{x} = 0$$
となり、常に0になってしまう。そのため、偏差の2乗の平均値を用いている。

Q：偏差の平均値が0になるのであれば、偏差の絶対値の平均値を用いればよいのではないか？
A：個々の値を計算するとき、絶対値記号を外すのは場合分けが必要で統一して計算できない。平均との差の2乗は統一して計算することができるため、一般的に偏差の2乗を用いている。

例7 的当てゲームにおけるA, B2人の得点の分散, 標準偏差を求めてみよう。

回	1	2	3	4	5	6	7	8	9	10
Aの得点	7	3	6	4	8	7	7	5	7	6
Bの得点	9	7	2	8	6	3	4	7	9	5

A, Bともに平均値は6点であるので,

回	1	2	3	4	5	6	7	8	9	10
Aの得点の偏差										
Bの得点の偏差										

A, Bそれぞれの得点の分散を s_A^2, s_B^2 とすると,

$s_A^2 =$

$s_B^2 =$

A, Bそれぞれの得点の標準偏差を s_A, s_B とすると

$s_A =$

$s_B =$

したがって、分散, 標準偏差の値は、ともに (4) の方が大きいので、得点の散らばり具合は (5) の方が大きい。

問7 次の表はある生徒の5回のテストの得点である。得点の分散と標準偏差を求めよ。ただし、標準偏差については教科書170ページの表を用いて、小数第3位を四捨五入して答えよ。

回	1	2	3	4	5
得点	72	91	79	88	75

回	1	2	3	4	5
得点の偏差					

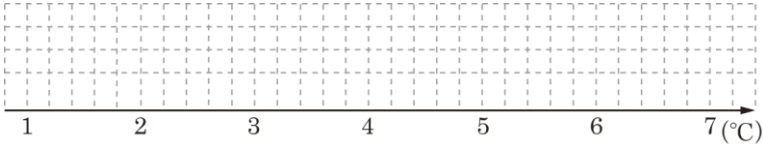
問題

- 1
- 右の表は、50 人の生徒が受けたあるテストの結果を，度数分布表にまとめたものである。それぞれの階級の点数をとった生徒の数を度数とする。このとき，次の間に答えよ。
- (1) 表の空所をうめて度数分布表を完成させ，最頻値を求めよ。

点数の階級 点以上 点未満	度数	相対度数
0 ～ 10		0.12
10 ～ 20	8	0.16
20 ～ 30		
30 ～ 40		0.28
40 ～ 50	12	0.24
計	50	

- 2
- 平成21 年の仙台における 1 日の平均気温を調べた。次の間に答えよ。
- (1) 次のデータは 3 月 1 日から 10 日までの 1 日の平均気温を順に並べたものである。このデータの箱ひげ図をかけ。

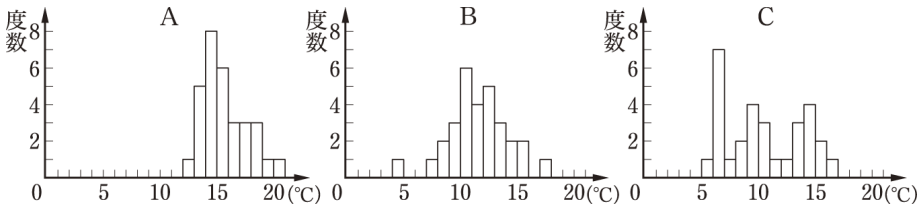
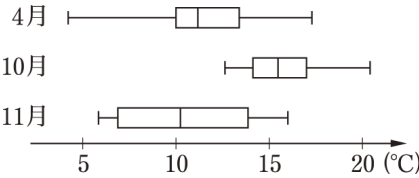
2.6 1.8 1.3 2.5 5.8 5.8 5.3 6.4 6.2 6.9



- (2) テストの点数の平均値を度数分布表から求めよ。

点数の階級 点以上 点未満	階級値 x	度数 f	xf
0 ～ 10			
10 ～ 20		8	
20 ～ 30			
30 ～ 40			
40 ～ 50		12	
計		50	

- (2) 右の図は 4 月，10 月，11 月の 1 日の平均気温の箱ひげ図である。4 月，10 月，11 月のヒストグラムを A～C の中からそれぞれ選べ。



3 A, B の 2 人が10点満点の小テストを 5 回受け、下の表のデータを得た。A, B それぞれの得点の範囲、分散を求めよ。

回	1	2	3	4	5
A の得点	7	5	8	6	4
B の得点	2	7	10	5	6

回	1	2	3	4	5
A の得点の偏差					
B の得点の偏差					

参考 分散の計算

変量 x の分散の求め方は、ほかにも以下の式で求めることができる。

$$(x \text{ の分散}) = (x^2 \text{ の平均値}) - (x \text{ の平均値})^2$$

$$(x \text{ の標準偏差}) = \sqrt{(x^2 \text{ の平均値}) - (x \text{ の平均値})^2}$$

例 1 上の分散の計算式を用いて、的当てゲームにおける A の得点の分散を求めてみよう。

$$\begin{aligned} s_A^2 &= \frac{1}{10} \left(7^2 + 3^2 + 6^2 + 4^2 + 8^2 + 7^2 + 7^2 + 5^2 + 7^2 + 6^2 \right) - 6^2 \\ &= \frac{1}{10} \times 382 - 36 = 2.2 \end{aligned}$$

問 1 上の計算式を用いて、的当てゲームにおける B の得点の分散を求めよ。

【的当てゲームにおける B の得点】

回	1	2	3	4	5	6	7	8	9	10
B の得点	9	7	2	8	6	3	4	7	9	5

2 節 データの相関

①相関

◆ 散布図

(¹) : 右の図のように、縦軸、横軸に 2 つのデータを示した図。

(²) : 一方が増加すれば他方も増加する傾向

(³) : 一方が増加すれば他方が減少する傾向

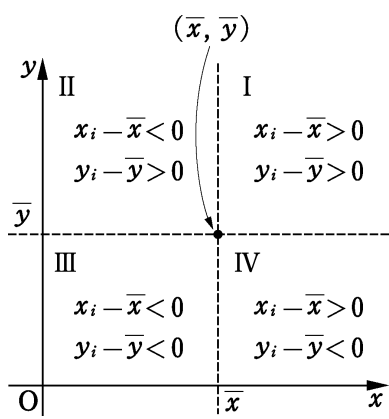
(⁴) : 正の相関も負の相関もみられない

(²) (³) (⁴)

②相関係数

◆ 相関係数

散布図の n 個のデータを $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ とする。
右の図のように、 $(\bar{x}, \bar{y})$ を中心に平面を 4 分割し、各部分を I, II, III, IV とする。
 x と y に正の相関があれば、散布図の点は I と III に多く集まり、負の相関があれば II と IV に多く集まる傾向がある。
散布図の各点 (x_i, y_i) ($i = 1, 2, 3, \dots, n$) が
I または III に属するときは $(x_i - \bar{x})(y_i - \bar{y}) > 0$
II または IV に属するときは $(x_i - \bar{x})(y_i - \bar{y}) < 0$
である。よって、



$(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})$
の値を考える。

$(x_i - \bar{x})(y_i - \bar{y})$ の平均値は

$$s_{xy} = \frac{1}{n} \{ (x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}) \}$$

この値を (⁵) といい、正の相関があるなら正に、負の相関があるなら負になる傾向がある。 x, y に相関がないときは、0 に近い値になる。

共分散はデータの単位が異なると相関の度合いを比較することができない。そこで、共分散を x, y の標準偏差の積で割った値を考える。

x, y の (⁶) : x, y の共分散を、 x, y の標準偏差の積で割った値。 r で表す。

$$x, y \text{ の } (^6) = \frac{(x, y \text{ の共分散})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})}$$

$$r = \frac{s_{xy}}{s_x s_y}$$

$$s_{xy} = \frac{1}{n} \{ (x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}) \}$$

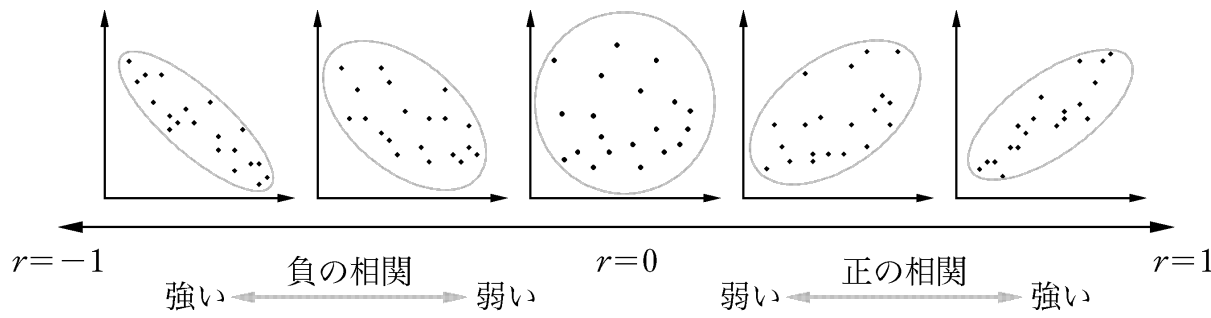
$$s_x = \sqrt{\frac{1}{n} \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}}$$

$$s_y = \sqrt{\frac{1}{n} \{ (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2 \}}$$

一般に、相関係数 r の値については、次の不等式が成り立つ。

$$-1 \leq r \leq 1$$

(⁷) の相関が強いほど r の値は 1 に近づき、(⁸) の相関が強いほど r の値は -1 に近づく。



例題 1 下の表は、5人の生徒 a, b, c, d, e が受けた、数学と国語のテストの点数である。数学（ x 点）と国語（ y 点）の点数の相関係数 r を、小数第3位を四捨五入して答えよ。

	x (点)	y (点)
a	72	69
b	76	76
c	93	88
d	84	75
e	85	87

数学の平均点 $\bar{x}$ ， 国語の平均点 $\bar{y}$

$$\bar{x} =$$

$$\bar{y} =$$

となり、下の表のようにして計算すると、

	x	y	$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
a	72	69					
b	76	76					
c	93	88					
d	84	75					
e	85	87					
合計							

$$r =$$

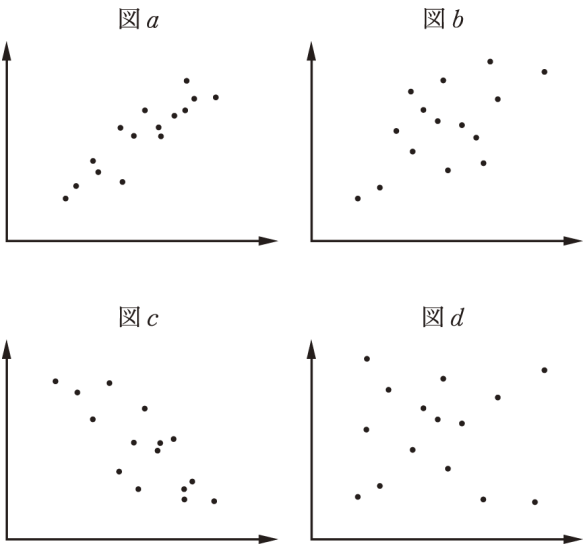
問 1 下の表は4人の生徒a, b, c, d の数学と英語の小テストの点数である。数学と英語の点数の相関係数を求めよ。

	a	b	c	d
数学	4	7	5	4
英語	6	9	8	9

	x	y	$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
a	4	6					
b	7	9					
c	5	8					
d	4	9					
計							

問題

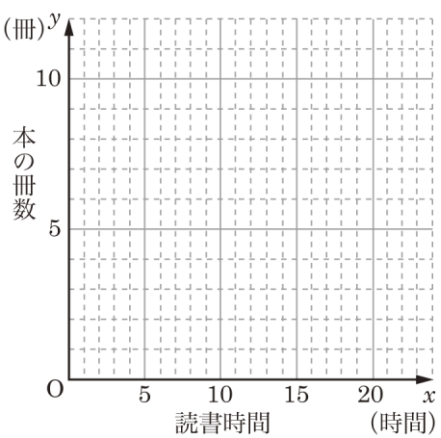
4 次の a , b , c , d の散布図に対応する相関係数を下の 4 つから選べ。



$r = 0.6$, $r = 0$, $r = -0.8$, $r = 0.9$

5 あるクラスの生徒10人に、10日間の合計読書時間 (x 時間) と実際に読んだ本の冊数 (y 冊) を調査したところ、右の表のようになった。次の問に答えよ。

(1) 右のデータの散布図をかけ。



番号	x (時間)	y (冊)
1	11	4
2	10	7
3	8	3
4	14	7
5	4	2
6	6	4
7	15	5
8	16	7
9	5	3
10	11	8

(2) 読書時間と本の冊数の相関係数 r を求めよ。

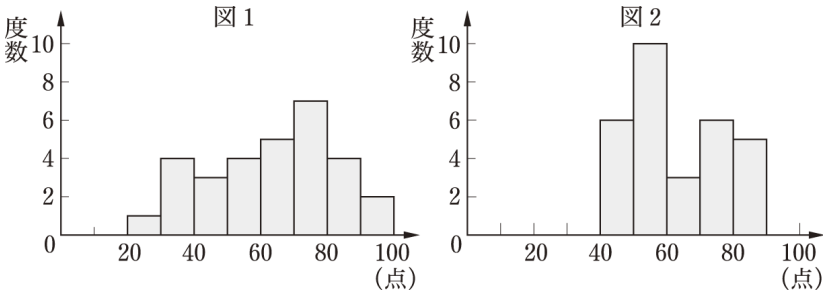
	x	y	$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	11	4					
2	10	7					
3	8	3					
4	14	7					
5	4	2					
6	6	4					
7	15	5					
8	16	7					
9	5	3					
10	11	8					
計							

練習問題

1 A, B, C 3 つの組で、100 点満点のテストを実施したところ、各組の成績は次のようになった。

組	人数	平均値	中央値	標準偏差
A	30	62.5	66.0	18.8
B	30	62.5	59.0	14.7
C	30	61.3	61.5	19.6

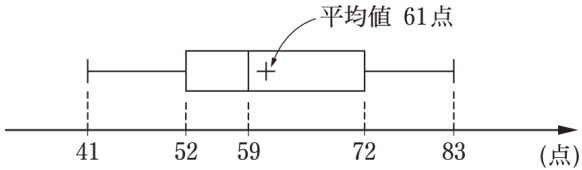
(1) 各組の点数にもとづいて、0 点以上 10 点未満というように、階級の幅が10 点のヒストグラムをつくった。次の図1, 2 のヒストグラムは、A組, B組のいずれかのものである。どちらがA組のものか答えよ。



(2) C組には、40 点未満の生徒が 5 人いて、その点数は、25, 30, 35, 36, 38であった。この 5 人について、再試験を実施したところ 5 人の平均点は40 点になった。他の生徒の点数は変わらないとして、C組の平均点は何点になるか求めよ。

(3) 次の表はB組の出席番号 1 番から10 番までの生徒のテストの点数で、下はその箱ひげ図である。表の a , b , c , d をうめよ。ただし、 a , b , c , d はこの順に点数が高いものから並んでいるとする。

出席番号	1	2	3	4	5	6	7	8	9	10
点数	56	59	a	b	76	c	52	83	46	d



2 右の表は、ある高校の女子生徒 10 人の握力を調べたものである。
次の問に答えよ。

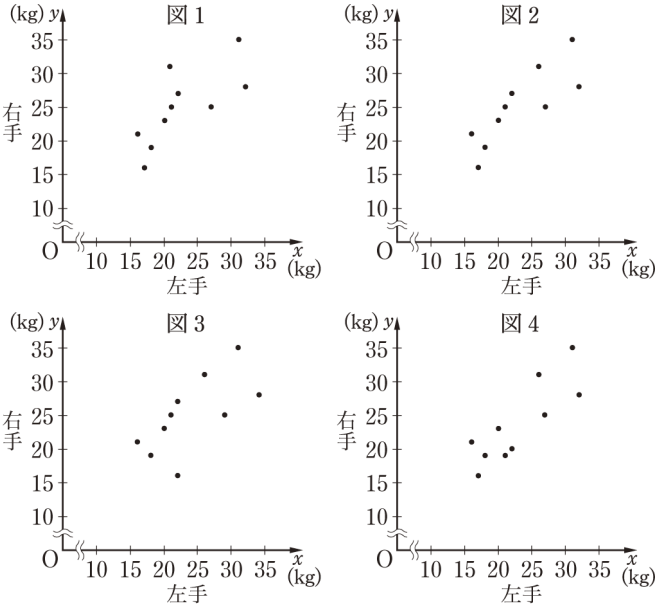
(1) A の値を求めよ。

(2) 右手の握力について、I 班の平均値は II 班の平均値より 3.6 kg 高い。このとき、I 班 2 番の生徒の握力から II 班 6 番の生徒の握力を引いた値 $B - C$ を求めよ。

(3) B , C の値を求めよ。

班	番号	左手の握力 $x\text{ (kg)}$	右手の握力 $y\text{ (kg)}$
I	1	31	35
	2	22	B
	3	18	19
	4	27	25
	5	32	28
II	6	A	C
	7	16	21
	8	26	31
	9	17	16
	10	20	23
平均値		23	25
標準偏差		5.4	5.3

(4) 左手の握力と右手の握力の散布図を下の図 1 ～ 4 の中から選べ。



(5) 左手の握力と右手の握力の相関係数 r を、小数第 3 位を四捨五入して求めよ。ただし、表中の標準偏差の値を用いてよい。

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
計					