

1. データの整理

◇ データ

人の身長、体重や運動の記録などのように、ある特性を表す数量を **変量** という。

数学では、ある変量の測定値や観測値の集まりを **データ** という。

また、データを構成する測定値や観測値の個数を、そのデータの **大きさ** という。

◇ 度数分布

度数分布表 データのとり値をいくつかの区間に区切って、各区間に含まれる値の個数(度数)を対応させたもの。
各区間を **階級**、区間の幅を **階級の幅**、各階級の中央の値を **階級値** という。

ヒストグラム 度数分布表を柱状のグラフで表したもの。

(例) 右の表は、ある40人の身長データを度数分布表で表したものである。階級の幅は4cmであり、階級値は上から順に160, 164, 168, 172, 176, 180(cm)である。

$$\frac{158+162}{2}$$

身長の度数分布表	
階級(cm)	度数
158 以上 162 未満	2
162 ~ 166	4
166 ~ 170	10
170 ~ 174	16
174 ~ 178	7
178 ~ 182	1
計	40

Q1 次のデータは、ある都市の1か月の毎日の最高気温を測定したものである。

13.4 11.1 12.4 10.1 13.1 8.3 9.1 9.6 14.8 16.7
15.2 16.8 15.5 12.8 13.7 15.4 15.4 16.2 17.1 15.5
18.2 15.3 18.7 19.7 20.7 17.6 18.4 19.7 20.8 20.0 (°C)

8°C以上10°C未満を階級の1つとして、どの階級の幅も2°Cである度数分布表を作れ。

[解答] データを整理すると、次のようになる。

8°C以上10°C未満 8.3, 9.1, 9.6
10°C以上12°C未満 11.1, 10.1
12°C以上14°C未満 13.4, 12.4, 13.1, 12.8, 13.7
14°C以上16°C未満 14.8, 15.2, 15.5, 15.4, 15.4, 15.5, 15.3
16°C以上18°C未満 16.7, 16.8, 16.2, 17.1, 17.6
18°C以上20°C未満 18.2, 18.7, 19.7, 18.4, 19.7
20°C以上22°C未満 20.7, 20.8, 20.0

よって、度数分布表は右のようになる。

階級(°C)	度数
8 以上 10 未満	3
10 ~ 12	2
12 ~ 14	5
14 ~ 16	7
16 ~ 18	5
18 ~ 20	5
20 ~ 22	3
計	30

1 次のデータは、ある高校の1年生女子20人の、ハンドボール投げの記録である。

16 12 17 14 16 19 13 14 11 14
14 17 11 15 16 15 18 16 17 15 (m)

このデータについて、11 m 以上 13 m 未満を階級の1つとして、どの階級の幅も2 mである度数分布表を作れ。

さらに、その度数分布表をもとにして、このデータのヒストグラムをかけ。

手順

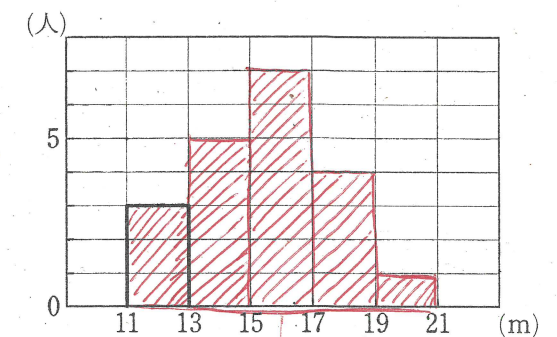
① データを整理 ... 小さい順に並べて
カウントする

② 度数分布表に記入 ... 階級の幅に注意。

※ データの数と、度数分布表の計の数が
一致し、それ等がないことをcheckする。

11, 11, 12,
13, 14, 14, 14, 14,
15, 15, 15, 16, 16, 16, 16,
17, 17, 17, 18,
19;

階級(m)	度数
11 以上 13 未満	3
13 ~ 15	5
15 ~ 17	7
17 ~ 19	4
19 ~ 21	1
計	20



ヒストグラム

階級の幅 2(m)

階級値

15~17であれば

15と17の中央の値

$$\frac{1}{2}(15+17)=16$$

階級値である。

2. データの代表値

中学の復習②

◇ 代表値

平均値 変数 x についてのデータが n 個の値 $x_1, x_2, \dots, x_n$ であるとき、このデータの平均値

$$\bar{x} \text{ は } \bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

中央値(メジアン) データを値の大きさに順に並べたとき、中央の位置にくる値。

データの大きさが偶数のとき、中央に2つの値が並ぶが、その場合は2つの値の平均値を中央値とする。

(例) データの値が 3, 4, 8, 9, 11, 12 のとき、中央値は $\frac{8+9}{2} = 8.5$

最頻値(モード) データにおいて、最も個数の多い値。

データが度数分布表に整理されているときは、

度数が最も大きい階級の階級値を最頻値とする。

2 次のデータの平均値、中央値を求めよ。

(1) 8, 15, 20, 15, 10

このデータを大きさの順に並べると

8, 10, 15, 15, 20

平均値は、 $\frac{1}{5}(8+10+15+15+20) = 13.6$

平均値 13.6
中央値 15

(2) 61, 40, 42, 45, 57, 49

このデータを値の大きさに順に並べると、

40, 42, 45, 49, 57, 61

データが偶数個なので、中央値は $\frac{1}{2}(45+49) = 47$

平均値は、 $\frac{1}{6}(40+42+45+49+57+61) = 47$

平均値 49
中央値 47

3 (1) 下の表は、男性 50 人の靴のサイズを調べた結果である。

このデータの最頻値を求めよ。

サイズ(cm)	24.0	24.5	25.0	25.5	26.0	26.5	27.0	計
人数	1	2	10	5	20	7	5	50

最も人数が多いもの

26.0cm が最頻値

答 26.0 (cm)

単位は
cm

(2) p. 2 の Q1 の度数分布表において、最頻値を求めよ。

度数が最も多い階級は

14~16 なので

最頻値は 14~16 の階級値 15

答 15

4 5 個の値 2, 9, 7, 3, a からなるデータの平均値が 5 であるとき、 a の値を求めよ。

データの平均値が 5 なので

$$\frac{1}{5}(2+9+7+3+a) = 5$$

$$21+a = 25$$

$$\therefore a = 4$$

答 $a = 4$

Q2 A, B, C の 3 つの家族の計 15 人が 5 問のクイズに答えた。

家族ごとの人数と正答数の平均値が右の表のようになるとき、15 人全員の正答数の平均値を求めよ。

	A	B	C
人数	4	5	6
正答数の平均値	2.5	3.4	3.5

解答 家族 A の正答数の合計は $4 \times 2.5 = 10$

家族 B の正答数の合計は $5 \times 3.4 = 17$

家族 C の正答数の合計は $6 \times 3.5 = 21$

よって、15 人全員の正答数の合計は $10 + 17 + 21 = 48$

ゆえに、15 人全員の正答数の平均値は $\frac{48}{15} = 3.2$

5 A, B, C の 3 つのグループの計 120 人が英語のテストを受けた。グループごとの人数と点数の平均値が右の表のようになるとき、120 人全員の点数の平均値を求めよ。

	A	B	C
人数	30	40	50
点数の平均値	70	65	50

グループ A の合計は、

$$30 \times 70 = 2100 \text{ (点)}$$

グループ B の合計は、 $40 \times 65 = 2600 \text{ (点)}$

グループ C の合計は、 $50 \times 50 = 2500 \text{ (点)}$ なので 120 人全員の合計は、

$$\text{合計で } 2100 + 2600 + 2500 = 7200 \text{ (点)}$$

$$\therefore \text{平均値} = \frac{7200}{120} = 60 \text{ (点)}$$

答 60 (点)

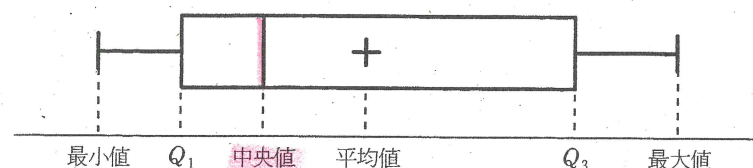
4. データの散らばりと四分位範囲 (2)

◇ 箱ひげ図

箱ひげ図 データの分布を表す図。

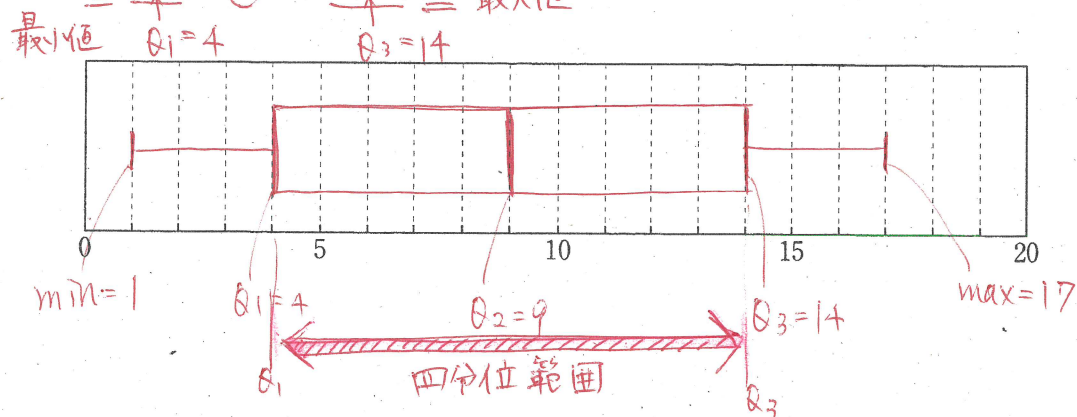
データの最小値、第1四分位数、中央値、第3四分位数、最大値を用いて、データの分布の特徴を表す。箱の長さは四分位範囲を表す。また、平均値を記入することもある。

複数のデータの分布を比較したいとき、箱ひげ図は有効である。

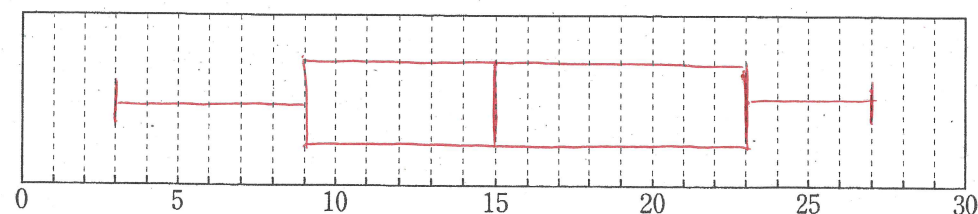


8 p.6の⑥(1), (2)のデータについて、箱ひげ図をかけ。ただし、平均値は記入しなくてよい。

(1) 1, 3, 5, 7, 9, 11, 13, 15, 17

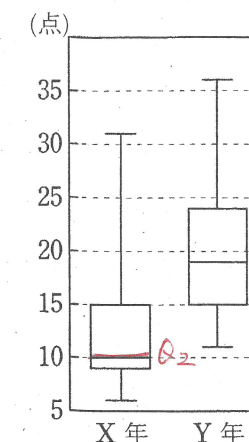


(2) 3, 6, 9, 12, 13, 15, 18, 21, 23, 24, 27



Q4 右の図は、スポーツ選手50人のある年X年、Y年の得点の記録の箱ひげ図である。この箱ひげ図から読み取れることとして、正しいと断定できるものを、次の①～③の中からすべて選べ。

- ① X年、Y年ともに、12点以上とった選手は25人よりも多い。✗
- ② X年からY年にかけて得点が減った選手は1人もいない。?
- ③ 四分位範囲に基づくと、X年の方がデータの散らばりの度合いが小さい。OK



50人の中央値は25、26番目

解答 ① X年の上から26番目の選手の得点は、中央値である10点以下である。

よって、X年に12点以上とった選手は25人以下であるから、正しくない。

② 箱ひげ図では、各個人の選手の得点の変化を調べることはできないから、正しいと断定できない。

③ X年の四分位範囲は $15-9=6$ (点)、Y年の四分位範囲は $24-15=9$ (点)

よって、データの散らばりの度合いが小さいのはX年であるから、正しい。

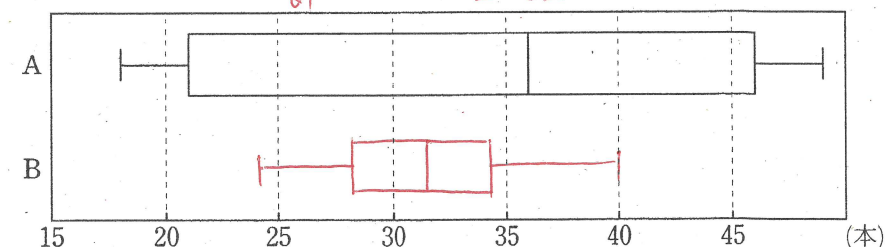
以上から、正しいと断定できるのは③

9 2人の野球選手A、Bの最近10年間のホームランの数は次のようであった。

A 35, 18, 27, 46, 48, 21, 49, 38, 37, 20

B 24, 28, 33, 33, 31, 39, 40, 30, 28, 34 (単位は本)

(1) Aのデータの箱ひげ図は下ようになる。Bのデータの箱ひげ図を、Aのデータの箱ひげ図と並べてかけ。



(2) データの散らばりの度合いが大きいのはA、Bのどちらと考えられるか。(1)で得られた箱ひげ図を用いて比較せよ。

答 A

5. 分散と標準偏差

◇ 分散, 標準偏差

変量 x についてのデータが n 個の値 $x_1, x_2, \dots, x_n$ であるとする。

偏差 データの各値と平均値 $\bar{x}$ の差 $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$

分散 偏差の2乗の平均値。 s^2 で表す。

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

また、分散は次のようにして求めることもできる。

$$s^2 = (x^2 \text{ のデータの平均値}) - (x \text{ のデータの平均値})^2$$

標準偏差 $s = \sqrt{\text{分散}}$

参考 変量 x の測定単位が、例えば cm であるとき、分散 s^2 の単位は cm^2 となり変量 x の測定単位と異なる。そこで、変量 x の測定単位と同じ単位である s を用いることも多い。

データの値が平均値の周りに集中しているほど、それぞれの偏差の絶対値は小さくなり、分散、標準偏差も小さくなる傾向にある。

(偏差の2乗)の平均

Q5 次の変量 x のデータの、分散 s^2 と標準偏差 s を求めよ。

5, 8, 9, 11, 12

解答 $\bar{x} = \frac{1}{5}(5+8+9+11+12) = \frac{45}{5} = 9$

よって、右の表から

$$s^2 = \frac{1}{5} \times 30 = 6, \quad s = \sqrt{6} \approx 2.4$$

x	5	8	9	11	12	計
$x - \bar{x}$	-4	-1	0	2	3	0
$(x - \bar{x})^2$	16	1	0	4	9	30

標準偏差 = $\sqrt{\text{分散}}$

$\bar{x} = 9$ のため

$$x - \bar{x} = 5 - 9 = -4$$

← 偏差の和は必ず0

偏差の2乗

$$(-4)^2 = 16$$

10 変量 x のデータが 10, 14, 14, 16, 17, 19 であるとする。

(1) x のデータの平均値 $\bar{x}$ を求めよ。

$$\bar{x} = \frac{1}{6}(10+14+14+16+17+19) = \frac{90}{6} = 15$$

答 $\bar{x} = 15$

(2) x のデータの偏差と偏差の2乗について、右の表を完成させよ。

x	10	14	14	16	17	19	計
$x - \bar{x}$	-5	-1	-1	1	2	4	0
$(x - \bar{x})^2$	25	1	1	1	4	16	48

必ず0になること確認

(3) x のデータの分散 s^2 と標準偏差 s を求めよ。

ただし、 $\sqrt{2} = 1.41$ とし、 s は小数第2位を四捨五入して小数第1位まで答えよ。

$$\text{分散} = \frac{1}{6} \times 48 = 8$$

$$\text{標準偏差} = \sqrt{\text{分散}} = \sqrt{8} = 2\sqrt{2}$$

$$= 2 \times 1.41 = 2.82$$

$$s^2 = 8$$

$$s \approx 2.8$$

11 次のデータは、8人の生徒の小テストの得点 x (点) である。

3, 3, 5, 5, 7, 9, 11, 13

(1) このデータの平均値 $\bar{x}$ を求めよ。

$$\bar{x} = \frac{1}{8}(3+3+5+5+7+9+11+13) = \frac{56}{8} = 7 \text{ (点)}$$

答 7 (点)

(2) x^2 のデータについて、次の表を完成させよ。

x	3	3	5	5	7	9	11	13	計
x^2	9	9	25	25	49	81	121	169	488

(3) 次のことを利用して、 x のデータの分散 s^2 と標準偏差 s を求めよ。

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

ただし、 $\sqrt{3} = 1.73$ とし、 s は小数第2位を四捨五入して小数第1位まで答えよ。

$$x^2 \text{ のデータの平均値 } \bar{x^2} = \frac{1}{8} \times 488 = 61$$

$$x \text{ のデータの平均値 } \bar{x} = 7$$

$$(x \text{ のデータの分散}) = 61 - 7^2 = 61 - 49 = 12$$

$$\text{標準偏差} = \sqrt{\text{分散}} = \sqrt{12} = 2\sqrt{3}$$

$$\sqrt{3} = 1.73 \text{ とき } s = 2 \times 1.73 = 3.46$$

$$s^2 = 12$$

$$s \approx 3.5 \text{ (点)}$$

(4) 小テストの日に欠席していた1人の生徒についても後日同じ小テストを行ったところ、得点は7点であった。この生徒を加えた9人の小テストの得点の分散は、加える前の分散と比較してどうなるか。正しいものを、次の①～③のうちから1つ選べ。

① 小さくなる ② 変わらない ③ 大きくなる

平均点の7(点)と同じ ⇒ 平均値は変化しない

→ 分散は (偏差)² の平均値

↑
変化する
① 前 8 人 だけ (大) 答 ①
② 後 9 人 だけ (小)

よって分散は小さくなる。

7. データの相関 (1)

◇ 散布図, 相関関係

散布図 2つの変量からなるデータを平面上に図示したもの。

散布図を用いると、2つの変量の間の関係が見やすくなる。

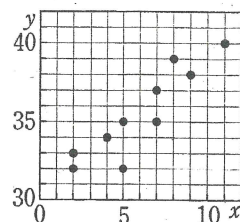
相関関係 2つの変量のデータにおいて、一方が増えると他方も増える傾向^①が認められるとき、2つの変量の間に **正の相関関係** があるという。
逆に、一方が増えると他方が減る傾向^②が認められるとき、2つの変量の間に **負の相関関係** があるという。

どちらの傾向も認められないときは、相関関係がないという。

また、①の傾向が著しいとき **強い正の相関関係** があるといい、

②の傾向が著しいとき **強い負の相関関係** があるという。

(例) 右上の散布図において、変量 x , y の間には正の相関関係があると考えられる。



◇ 相関表

2つの変量の間の関係を調べるとき、2つの度数分布表を組み合わせた **相関表** を用いることがある。

データの数が非常に多いとき、散布図を使うと同じ場所に点がいくつも重なりデータの分布がとらえにくいことがある。そのようなときは相関表を使うとよい。

$y \backslash x$	40 以上 50 未満	50 ~ 60	60 ~ 70	計
90 以上 100 未満		1	20	21
80 ~ 90	5	11	18	34
70 ~ 80	2	3		5
計	7	15	38	60

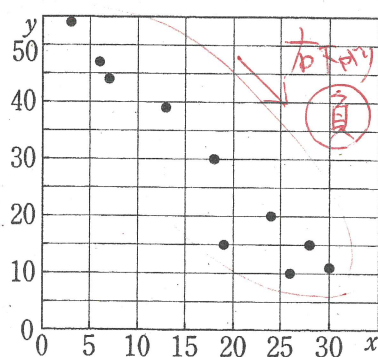
Q7 右の表は、京都市のある年の
1 ~ 10 月における平均気温 $x(^{\circ}\text{C})$ と
1 世帯当たりのガスの平均消費量
 $y(\text{m}^3)$ を調べた結果である。

月	1	2	3	4	5	6	7	8	9	10
x	3	6	7	13	18	24	28	30	26	19
y	54	47	44	39	30	20	15	11	10	15

(1) 2つの変量 x , y の散布図と相関表をかけ。

(2) x と y の間には、正、負どちらの相関関係があると考えられるか。

解答 (1) 散布図, 相関表はそれぞれ次のようになる。



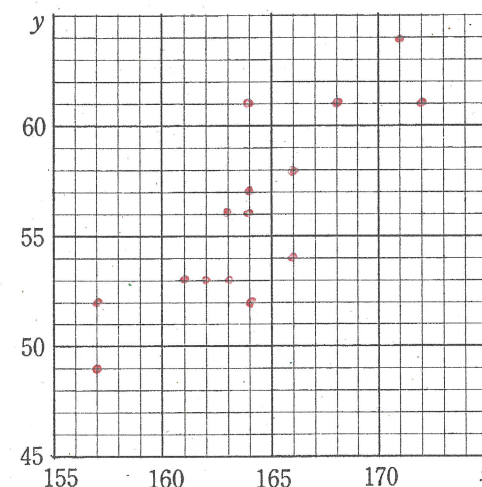
$y \backslash x$	0 以上 10 未満	10 ~ 20	20 ~ 30	30 ~ 40	計
以上 ~ 未満					
50 ~ 60	1				1
40 ~ 50	2				2
30 ~ 40		2			2
20 ~ 30			1		1
10 ~ 20		1	2	1	4
計	3	3	3	1	10

(2) x が増えると y が減る傾向にあるから、**負** の相関関係があると考えられる。

15 下の表は、ある女子サッカーチームの選手 15 人について、身長を $x(\text{cm})$ 、体重を $y(\text{kg})$ として、 x , y を調べた結果である。

番号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	164	171	161	164	162	163	172	166	157	157	166	164	164	163	168
y	61	64	53	56	53	53	61	54	52	49	58	52	57	56	61

(1) x , y の散布図をかけ。



(2) x , y の相関表を完成させよ。

$y \backslash x$	155 以上 160 未満	160 ~ 165	165 ~ 170	170 ~ 175	計
60 以上 65 未満		1	1	2	4
55 ~ 60		3	1		4
50 ~ 55	1	4	1		6
45 ~ 50	1				1
計	2	8	3	2	15

散布図より 右上がり → 正の相関

(3) x と y の関係として最も適当なものを、次の① ~ ③のうちから1つ選べ。

- ① x と y の間には正の相関関係がある。
- ② x と y の間には負の相関関係がある。
- ③ x と y には相関関係がない。

8. データの相関 (2)

◇ 共分散

2つの変量 x, y のデータが、 n 個の値の組として $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ のように与えられるとき、 x の偏差と y の偏差の積 $(x_k - \bar{x})(y_k - \bar{y})$ の平均値を x と y の **共分散** といい、 s_{xy} で表す。

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

◇ 相関係数

2つの変量 x, y の共分散 s_{xy} を、 x の標準偏差 s_x と y の標準偏差 s_y の積 $s_x s_y$ で割った量を **相関係数** といい、 r で表す。

$$\begin{aligned} r &= \frac{s_{xy}}{s_x s_y} \leftarrow \frac{(x \text{ と } y \text{ の共分散})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})} \\ &= \frac{\frac{1}{n} \{ (x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}) \}}{\sqrt{\frac{1}{n} \{ (x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}} \sqrt{\frac{1}{n} \{ (y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2 \}}} \\ &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\{ (x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \} \{ (y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2 \}}} \end{aligned}$$

相関係数 r は相関関係の強弱をみるための量であり、次のことが成り立つ。

[1] $-1 \leq r \leq 1$

[2] r の値が **1** に近いことは正の相関関係が強いことの見安であり、 r の値が **-1** に近いことは負の相関関係が強いことの見安である。

Q8 右の表は、生徒 5 人が受けた物理のテストの点数 x (点) と、化学のテストの点数 y (点) のデータである。

x と y の相関係数 r を求めよ。また、 x と y の間にはどのような相関関係があると考えられるか答えよ。

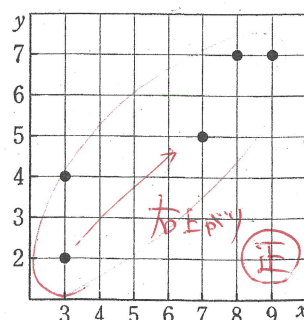
生徒	A	B	C	D	E
x	7	3	8	9	3
y	5	2	7	7	4

【解答】 $\bar{x} = \frac{1}{5}(7+3+8+9+3) = 6, \quad \bar{y} = \frac{1}{5}(5+2+7+7+4) = 5$

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
A	7	5	1	0	0	1	0
B	3	2	-3	-3	9	9	9
C	8	7	2	2	4	4	4
D	9	7	3	2	6	9	4
E	3	4	-3	-1	3	9	1
計	30	25			22	32	18

よって $r = \frac{22}{\sqrt{32} \sqrt{18}} = \frac{22}{4\sqrt{2} \times 3\sqrt{2}} = \frac{11}{12} \approx 0.916 \dots \approx 0.92$

r は 1 に近いから、 x と y の間には **正** の相関関係があると考えられる。



16 ある 2 つの変量 x, y のデータについて、 x の標準偏差が 2.70、 y の標準偏差が 3.60、 x と y の共分散が 3.24 であった。 x と y の相関係数 r を求めよ。ただし、小数第 3 位を四捨五入せよ。

$$\text{相関係数 } r = \frac{\text{共分散}}{\sqrt{x} \sqrt{y}} = \frac{3.24}{2.70 \times 3.60} = \frac{3.24}{9.72} = \frac{1}{3}$$

答 $r \approx 0.33$

$$r = 0.333 \dots$$

17 右の表は、ある店で月曜日から金曜日に売れたアイスクリームの個数 x (個) と缶コーヒーの本数 y (本) のデータである。

曜日	月	火	水	木	金
x	8	10	9	7	6
y	11	7	9	13	10

(1) x, y のデータの平均値 $\bar{x}, \bar{y}$ をそれぞれ求めよ。

$$\bar{x} = \frac{1}{5}(8+10+9+7+6) = \frac{40}{5} = 8$$

$$\bar{y} = \frac{1}{5}(11+7+9+13+10) = \frac{50}{5} = 10$$

$$\bar{x} = 8 \text{ (個)}$$

$$\bar{y} = 10 \text{ (本)}$$

(2) 下の表を完成させよ。

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
月	8	11	0	1	0	0	1
火	10	7	2	-3	-6	4	9
水	9	9	1	-1	-1	1	1
木	7	13	-1	3	-3	1	9
金	6	10	-2	0	0	4	0
計	40	50			-10	10	20

(3) (2) の表を利用して、 x と y の相関係数 r を求めよ。ただし $\sqrt{2} = 1.41$ とし、計算結果は小数第 3 位を四捨五入せよ。

$$\begin{aligned} &\text{ } x \text{ の標準偏差 } \sqrt{10} \\ &\text{ } y \text{ の標準偏差 } \sqrt{20} \\ &\text{ } x \text{ と } y \text{ の共分散 } -10 \end{aligned} \quad r = \frac{-10}{\sqrt{10} \sqrt{20}} = -\frac{1}{\sqrt{2}} = -\frac{\sqrt{2}}{2}$$

$$\therefore r = -\frac{1.41}{2} = -0.705 \quad \text{答 } -0.71$$

(4) x と y の関係として正しいものを、次の ①～③のうちから 1 つ選べ。

- ① アイスクリームが多く売れる日は、缶コーヒーも売れやすい傾向がある。
- ② アイスクリームが多く売れる日は、缶コーヒーは売れにくい傾向がある。
- ③ ①、② のような傾向は認められない。

(3) より $r \approx -0.71$ と負の強い相関がある。

よって ②

答 ②