

1節 データの分析

1 データと度数分布表

(教科書 p.128)

電車で通学する生徒 20 人の A 班と、徒歩で通学する生徒 15 人の B 班に分けて、先月の読書時間の合計を調べたものである。

[A 班 (単位 時間)]

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

[B 班 (単位 時間)]

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

このように、何かを調べるために、調査して集めた資料や実験をして得られた資料のことを

(¹) という。

度数分布表とヒストグラム

(教科書 p.128)

下の表は、A 班のデータを、時間を 4 時間ずつの区間に分け、その区間に入る人数を調べてまとめたものである。

(²) : データを整理したときの各区間

(³) : 各階級に入っているデータの個数

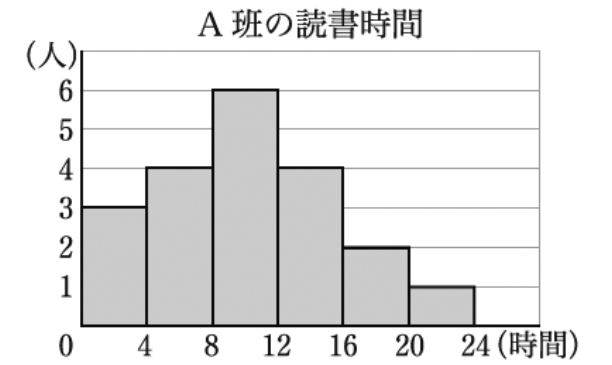
(⁴) : 階級の真ん中の値

(⁵) : 各階級に度数を対応させた表

A 班の読書時間

時間の (⁶) (時間)	(⁷) (時間)	(⁸) (人)
0以上 ~ 4 未満	2	3
4 ~ 8	6	4
8 ~ 12	10	6
12 ~ 16	14	4
16 ~ 20	18	2
20 ~ 24	22	1
計		20

右の図は、A 班の読書時間の度数分布表について、階級の幅を底辺、度数を高さとする長方形をすきまなく並べたグラフである。このようなグラフを (⁹) という。

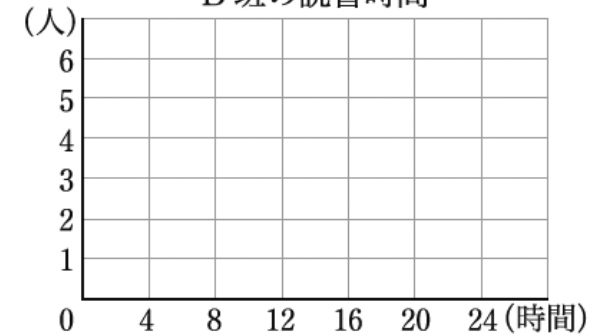


問1 B 班の読書時間を度数分布表にまとめ、ヒストグラムに表しなさい。また、度数の最も大きい階級を答えなさい。

B 班の読書時間

時間の階級 (時間)	階級値 (時間)	度数 (人)
0以上 ~ 4 未満		
4 ~ 8		
8 ~ 12		
12 ~ 16		
16 ~ 20		
20 ~ 24		
計		

B 班の読書時間



相対度数

(教科書 p.129)

A 班と B 班では、度数の合計が異なるので、各階級の度数をそのまま比べても違いがはっきりしない。度数の合計が異なるときは、各階級の度数が、度数の合計に対してどのような割合であるか調べてみるとよい。

(¹⁰) : 各階級の度数を度数の合計でわった値

(¹¹) : 各階級に相対度数を対応させた表

A 班, B 班の読書時間

時間の階級 (時間)	A 班		B 班	
	度数 (人)	相対 度数	度数 (人)	相対 度数
0 以上 ~ 4 未満	3	0.15		
4 ~ 8	4	0.20		
8 ~ 12	6	0.30		
12 ~ 16	4	0.20		
16 ~ 20	2	0.10		
20 ~ 24	1	0.05		
計	20	1.00		

問2 上の B 班の読書時間の相対度数分布表を完成しなさい。また、16 時間以上 20 時間未満の階級では、A 班と B 班で、どちらの相対度数が大きいか答えなさい。

2 代表値

(1) : データ全体の特徴を表す数値。

例) 平均値, 中央値, 最頻値 など

平均値と中央値

(教科書 p.130)

(2) : データの値の合計をデータの値の個数でわった値

例1 A 班の読書時間の平均値を求めてみよう。

問3 教科書 128 ページの B 班の読書時間の平均値を求めなさい。

(3) : データの値を小さい順に並べたときの中央の値

ただし、データの値の個数が偶数のときは、中央にある 2 つの値の平均値とする。

例2 A 班の読書時間の中央値を求めてみよう。

データの値を小さい順に並べると、次のようになる。

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

└──────────┬──────────┘

10人 10人

中央値は、() の平均値である。

よって、 $\frac{\quad}{2}$

問4 教科書 128 ページの B 班の読書時間を、値の小さい順に並べると次のようになる。B 班の読書時間の中央値を求めなさい。

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

データのなかに、ほかの値とかけ離れた値がある場合がある。このようなときは、平均値よりも中央値の方がその影響を受けにくいので、代表値としては、平均値より中央値が適当である。

最頻値

(教科書 p.131)

(4) : 度数分布表で、度数が最も大きい階級の階級値。

ヒストグラムにおいて、最も高い長方形に対応する階級値である。

例3 教科書 128 ページの A 班の読書時間の最頻値を求めてみよう。

右の度数分布表で、度数が最も大きい階級は

() 時間以上 () 時間未満

である。

したがって、最頻値はその階級値の () 時

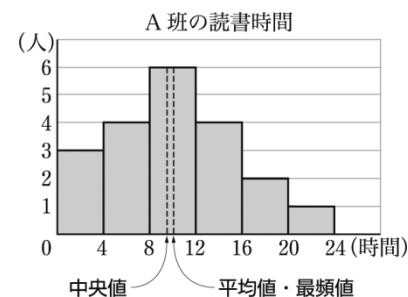
間である。

時間の階級 (時間)	階級値 (時間)	度数 (人)
0 以上 ~ 4 未満	2	3
4 ~ 8	6	4
8 ~ 12	10	6
12 ~ 16	14	4
16 ~ 20	18	2
20 ~ 24	22	1
計		20

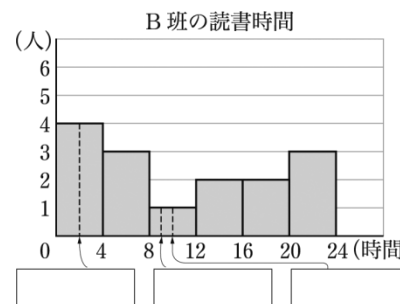
問5 P.1 (教科書 129 ページ) の問 1 でつくった度数分布表から、B 班の読書時間の最頻値を求めなさい。

例4 A 班と B 班について、読書時間の 3 つの代表値を比べてみよう。

A 班の読書時間の 3 つの代表値をヒストグラムに表すと、下の図のようになる。



問6 例 4 にならって、B 班の読書時間の 3 つの代表値を、ヒストグラムの に入れなさい。



3 四分位数と箱ひげ図

四分位数と四分位範囲

(教科書 p.132)

データの散らばりぐあいは、代表値ではとらえられない。

データの値で(最大値) - (最小値)を、そのデータの分布の()という。

範囲を調べて、データの散らばりぐあいを比べてみよう。

下の表はバスケットボール部のAさんとBさんの、最近10試合で成功したシュートの本数である。

Aさん	(単位 本)	Bさん	(単位 本)
5 3 6 5 6 7 8 7 8 10		4 5 13 5 9 6 6 7 6 5	

2人の成功したシュートの本数の範囲は、それぞれ

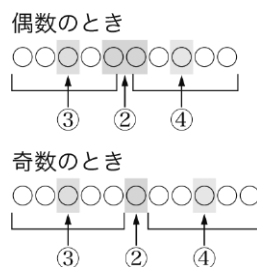
Aさん() (本)

Bさん() (本)

範囲の値を比べると、散らばりぐあいはAさんよりBさんの方が大きいといえる。

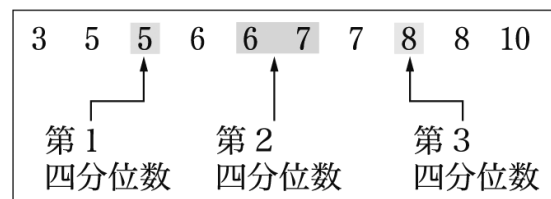
範囲は、データのなかにほかとかけ離れた値がある場合、その値の影響を受けやすい。その影響を少なくしたものに、()がある。次のようにして求める。

- ① データの値を小さい順に並べ、中央値を境にして2つに分ける。
- ② 中央値を、()という。
- ③ 最小値を含む方のデータの中央値を()という。
- ④ 最大値を含む方のデータの中央値を()という。
- ⑤ (第3四分位数) - (第1四分位数)が、四分位範囲である。



また、四分位範囲を2でわった値を()という。

例5 Aさんの成功したシュートの本数を、小さい順に並べかえると右のようになるから



第2四分位数()本

第1四分位数()本

第3四分位数()本

四分位範囲()本

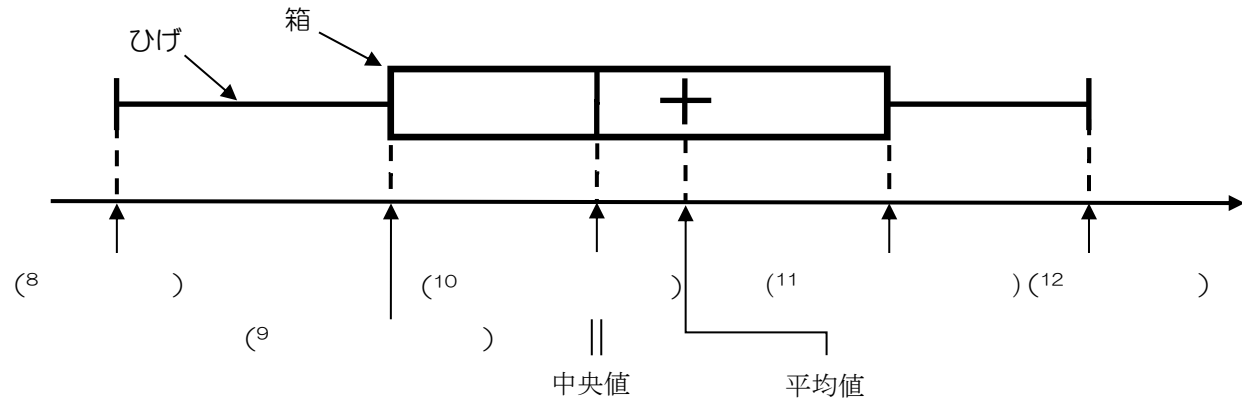
四分位偏差()本

問7 Bさんの成功したシュートの本数の、四分位範囲、四分位偏差を求めなさい。また、AさんとBさんの四分位範囲を比べて、どちらの散らばりぐあいが大きいか答えなさい。

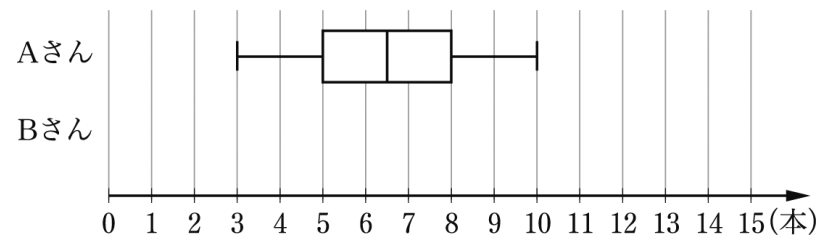
箱ひげ図

(教科書 p.133)

(7) : 最小値, 第1四分位数, 第2四分位数(中央値), 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。



例6 例5の結果から, 箱ひげ図は次のようになる。



問8 Bさんの成功したシュートの本数の箱ひげ図を, 上の図に表しなさい。

4 分散と標準偏差

(教科書 p.134)

データの散らばりぐあいを、データの個々の値と平均値との差を用いて考えてみよう。

(¹) : (データの個々の値) - (平均値)

下の表は、ある野球チームのA投手と、B投手が、最近5試合で奪った三振の個数である。

(単位 個)

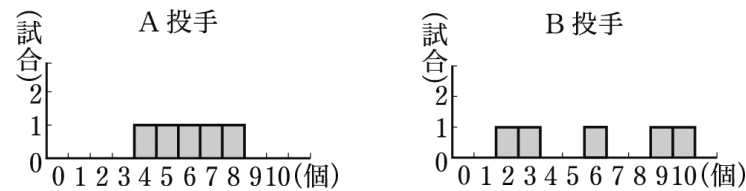
	1 試合目	2 試合目	3 試合目	4 試合目	5 試合目
A 投手	4	8	7	5	6
B 投手	10	2	9	6	3

平均値は、それぞれ

$$\text{A 投手} \quad \frac{4+8+7+5+6}{5} = \frac{30}{5} = 6 \text{ (個)}$$

$$\text{B 投手} \quad \frac{10+2+9+6+3}{5} = \frac{30}{5} = 6 \text{ (個)}$$

となり、等しくなる。しかし、三振の個数をヒストグラムに表すと、散らばりぐあいは異なる。



次に、A投手、B投手の三振の個数の偏差とその合計を求めると下の表のようになる。

	1 試合目	2 試合目	3 試合目	4 試合目	5 試合目	計
A 投手の偏差	-2	2	1	-1	0	0
B 投手の偏差	4	-4	3	0	-3	0

偏差の合計はどちらも0になるため、偏差の平均値でデータ全体の散らばりぐあいを表すことはできない。

そこで、偏差を2乗した値を考える。

偏差の2乗は0以上の値になるので、合計が0に近いほど平均値からの散らばりぐあいが小さい。

(²) : 偏差の2乗の平均値。 s^2 で表す。

(³) : 分散の正の平方根。 s で表す。

分散と標準偏差

偏差 (データの個々の値) - (平均値)

分散 $s^2 = (\text{偏差})^2$ の平均値 = $\frac{(\text{偏差})^2 \text{ の合計}}{\text{データの値の個数}}$

標準偏差 $s = \sqrt{s^2} = (\text{分散の正の平方根})$

例7 教科書 134 ページのA投手が奪った三振の個数から偏差の2乗の合計を計算すると、右の表のようになる。したがって、分散 s^2 は

$$s^2 =$$

標準偏差 s は

$$s =$$

A 投手	三振の個数	偏差	(偏差) ²
1 試合目	4	-2	4
2 試合目	8	2	4
3 試合目	7	1	1
4 試合目	5	-1	1
5 試合目	6	0	0
計	30	0	10

問9 右の表を完成して、B投手が奪った三振の個数の分散を求めなさい。また、標準偏差を、四捨五入して小数第2位まで求めなさい。

B 投手	三振の個数	偏差	(偏差) ²
1 試合目	10		
2 試合目	2		
3 試合目	9		
4 試合目	6		
5 試合目	3		
計	30		

問10 A投手とB投手が奪った三振の個数の標準偏差を比べて、どちらの散らばりぐあいが大きいか答えなさい。

5 相関関係

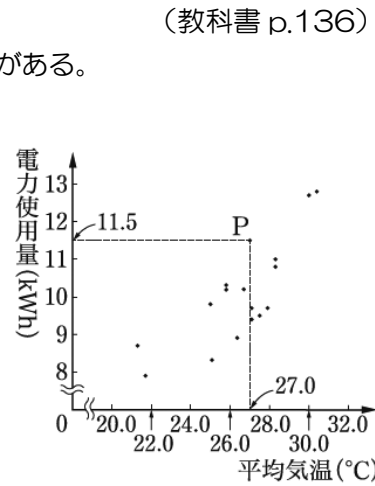
散布図

平均気温と電力使用量のように、互いに関係があると考えられる数量がある。

このような数量を組にしたデータについて考えてみよう。

(¹) : 右の図のように、2つのデータの値の組を x 座標、 y 座標とする点として平面上に表した図。

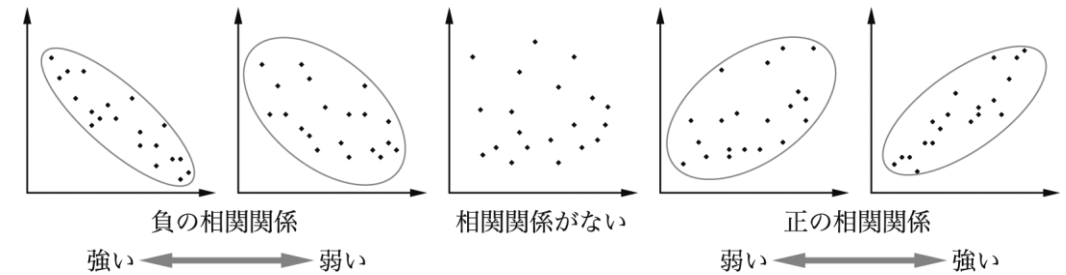
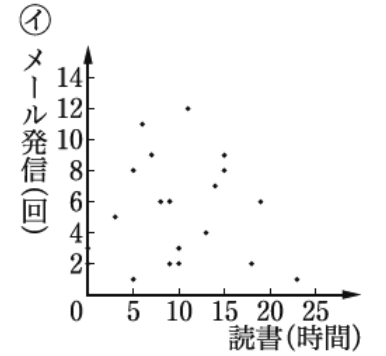
右の図は、夏の1日ごとの平均気温と世帯あたりの電力使用量を散布図に表したものである。図中の点Pは、平均気温が27°Cで、電力使用量が11.5kWhであることを表している。



読書時間とメール発信回数を散布図に表すと、右の図①のようになる。

読書時間とメール発信回数には正の相関関係も負の相関関係もみられない。

(⁴) : 正の相関も負の相関もみられない。



相関関係

(教科書 p.136)

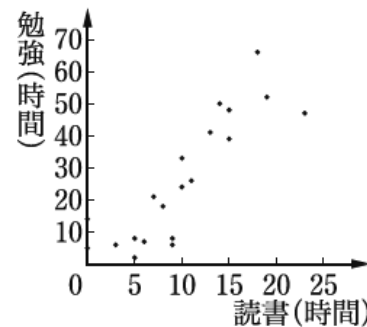
下の表は、20人の生徒について、先月の読書時間と、勉強時間、テレビ視聴時間、1日のメール発信回数の平均を調べたデータである。

読書(時間)	3	10	7	14	5	9	15	0	9	18	0	8	11	10	15	19	6	23	13	5
勉強(時間)	6	33	21	50	2	8	39	5	6	66	14	18	26	24	48	52	7	47	41	8
テレビ視聴(時間)	58	37	47	26	55	52	48	38	56	4	48	25	36	31	24	19	53	18	20	62
メール発信(回)	5	2	9	7	8	6	8	3	2	2	2	6	12	3	9	6	11	1	4	1

読書時間と勉強時間を散布図に表すと、右の図のようになる。

この散布図から、読書時間と勉強時間は、一方が増加すれば他方も増加する傾向があることがわかる。

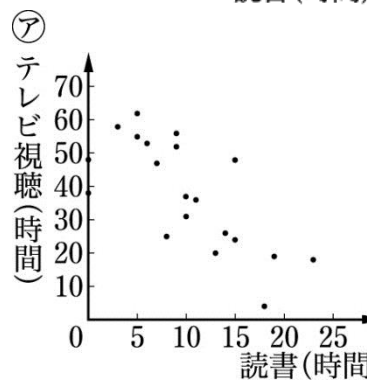
(²) : 一方が増加すれば他方も増加する傾向。



読書時間とテレビ視聴時間を散布図に表すと、右の図㉗のようになる。

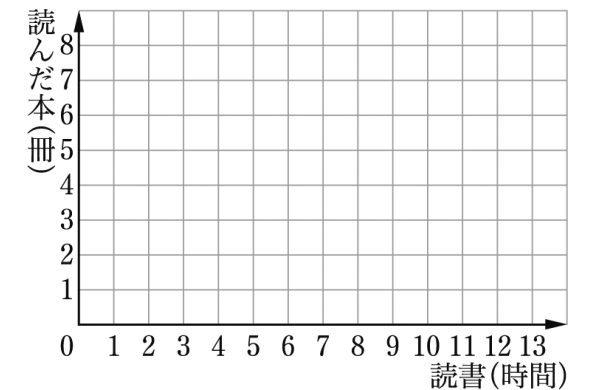
この散布図から、読書時間とテレビ視聴時間は、一方が増加すれば他方が減少する傾向があることがわかる。

(³) : 一方が増加すれば他方が減少する傾向



問11 次の表は、5人の生徒の、先月の読書時間と読んだ本の冊数を示したものである。散布図をつくり、読書時間と読んだ本の冊数には、どのような相関関係があるか答えなさい。

生徒	読書(時間)	読んだ本(冊)
a	5	2
b	8	6
c	9	4
d	12	8
e	11	5



6 相関係数

相関係数の意味

(教科書 p.138)

相関関係を調べたい2つの数量を x, y とする。 x の偏差と y の偏差の積の平均値を⁽¹⁾) という。また、共分散を x の標準偏差と y の標準偏差の積でわった値を⁽²⁾) という。相関係数は記号 r で表す。

共分散と相関係数	
共分散	x と y の偏差の積の平均値
相関係数	$r = \frac{\text{共分散}}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})}$

相関係数 r の値については、次の不等式が成り立つ。

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

⁽³⁾) の相関関係が強いほど r の値は 1 に近づき、⁽⁴⁾) の相関関係が強いほど -1 に近づく。

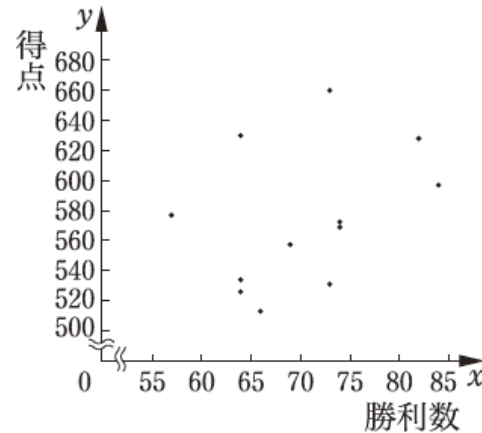
問 12 プロ野球 12 チームについて、1 年間の勝利数と失点数の相関係数は -0.73 、勝利数と得失点差の相関係数は 0.90 である。

次の①、②から、正しいものを選びなさい。

- (1) 勝利数と失点数には
- ① 正の相関関係がある
 - ② 負の相関関係がある

(2) 勝利数と得点より、勝利数と得失点差の方が

- ① 正の相関関係が強い
- ② 正の相関関係が弱い



相関係数

(教科書 p.139)

例8 読書時間と読んだ本の冊数の相関係数を求めてみよう。教科書 137 ページの問 11 について、読書時間を x とし、読んだ本の冊数を y とする。

(x の平均値) =

(y の平均値) =

下の表のようにして計算すると

生徒	x	y	x の偏差	y の偏差	(x の偏差) ²	(y の偏差) ²	偏差の積
a	5	2	-4	-3	16	9	12
b	8	6	-1	1	1	1	-1
c	9	4	0	-1	0	1	0
d	12	8	3	3	9	9	9
e	11	5	2	0	4	0	0
計	45	25	0	0	30	20	20

x, y の共分散は

x の標準偏差は

y の標準偏差は

相関係数 r は

$r =$

問 13 次の表は、4人の生徒の数学と英語の小テストの得点を示したものである。表を完成することにより、数学の得点 x と英語の得点 y の相関係数を求めなさい。

生徒	x	y	x の偏差	y の偏差	$(x \text{ の偏差})^2$	$(y \text{ の偏差})^2$	偏差の積
a	6	5					
b	7	5					
c	7	8					
d	8	10					
計							

復習問題

(教科書 p.140)

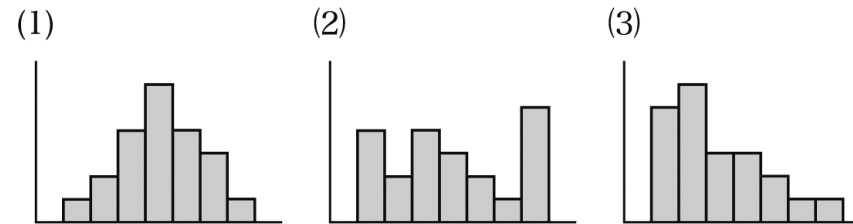
1 次の資料は、1年生の図書委員8人が1年間に学校の図書室から借りた本の冊数を調べたものである。

7	12	8	40	9	4	8	8
---	----	---	----	---	---	---	---

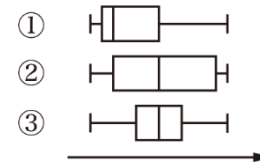
(1) 借りた本の冊数の平均値、中央値を求めなさい。

(2) 40冊借りた生徒を除いて、残り7人が借りた本の冊数の平均値、中央値を求めなさい。

2 次のヒストグラムについて、対応する箱ひげ図を選びなさい。

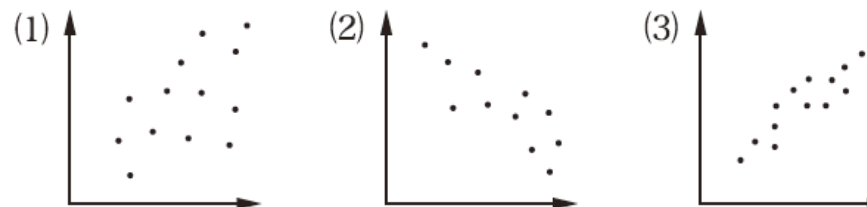


箱ひげ図



3 次の散布図について、対応する相関係数を選びなさい。

散布図



相関係数

- ① 0.9 ② 0.5 ③ -0.8

1節 データの分析

1 データと度数分布表

(教科書 p.128)

電車で通学する生徒 20 人の A 班と、徒歩で通学する生徒 15 人の B 班に分けて、先月の読書時間の合計を調べたものである。

[A 班 (単位 時間)]

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

[B 班 (単位 時間)]

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

このように、何かを調べるために、調査して集めた資料や実験をして得られた資料のことを

(¹ **データ**) という。

度数分布表とヒストグラム

(教科書 p.128)

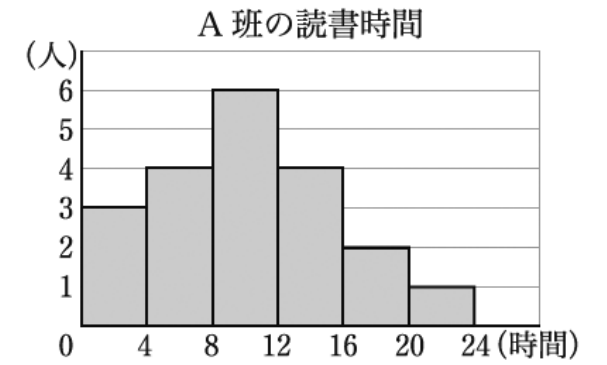
下の表は、A 班のデータを、時間を 4 時間ずつの区間に分け、その区間に入る人数を調べてまとめたものである。

- (² **階級**) : データを整理したときの各区間
- (³ **度数**) : 各階級に入っているデータの個数
- (⁴ **階級値**) : 階級の真ん中の値
- (⁵ **度数分布表**) : 各階級に度数を対応させた表

A 班の読書時間

時間の (⁶ 階級) (時間)	(⁷ 階級値) (時間)	(⁸ 度数) (人)
0以上 ~ 4 未満	2	3
4 ~ 8	6	4
8 ~ 12	10	6
12 ~ 16	14	4
16 ~ 20	18	2
20 ~ 24	22	1
計		20

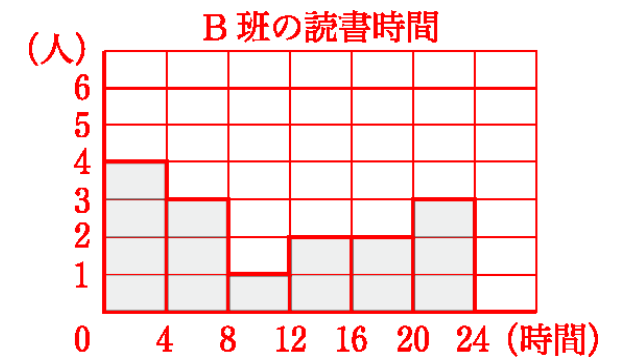
右の図は、A 班の読書時間の度数分布表について、階級の幅を底辺、度数を高さとする長方形をすきまなく並べたグラフである。このようなグラフを (⁹ **ヒストグラム**) という。



問1 B 班の読書時間を度数分布表にまとめ、ヒストグラムに表しなさい。また、度数の最も大きい階級を答えなさい。

B 班の読書時間

時間の階級 (時間)	階級値 (時間)	度数 (人)
0 以上 ~ 4 未満	2	4
4 ~ 8	6	3
8 ~ 12	10	1
12 ~ 16	14	2
16 ~ 20	18	2
20 ~ 24	22	3
計		15



度数の最も大きい階級は 0 時間以上 4 時間未満である。

相対度数

(教科書 p.129)

A 班と B 班では、度数の合計が異なるので、各階級の度数をそのまま比べても違いがはっきりしない。度数の合計が異なるときは、各階級の度数が、度数の合計に対してどのような割合であるか調べてみるとよい。

(¹⁰ **相対度数**) : 各階級の度数を度数の合計でわった値

(¹¹ **相対度数分布表**) : 各階級に相対度数を対応させた表

A 班, B 班の読書時間

時間の階級 (時間)	A 班		B 班	
	度数 (人)	相対 度数	度数 (人)	相対 度数
0 以上 ~ 4 未満	3	0.15	4	0.27
4 ~ 8	4	0.20	3	0.20
8 ~ 12	6	0.30	1	0.07
12 ~ 16	4	0.20	2	0.13
16 ~ 20	2	0.10	2	0.13
20 ~ 24	1	0.05	3	0.20
計	20	1.00	15	1.00

問2 上の B 班の読書時間の相対度数分布表を完成しなさい。また、16 時間以上 20 時間未満の階級では、A 班と B 班で、どちらの相対度数が大きいか答えなさい。

16 時間以上 20 時間未満の階級では、相対度数は B 班の方が大きい。

2 代表値

(¹ **代表値**) : データ全体の特徴を表す数値。

例) 平均値, 中央値, 最頻値 など

平均値と中央値

(教科書 p.130)

(² **平均値**) : データの値の合計をデータの値の個数でわった値

例1 A 班の読書時間の平均値を求めてみよう。

$$\frac{3+10+7+\dots+13+5}{20} = \frac{200}{20} = 10 \text{ (時間)}$$

問3 教科書 128 ページの B 班の読書時間の平均値を求めなさい。

$$\frac{6+20+0+\dots+0+9}{15} = \frac{150}{15} = 10 \text{ (時間)}$$

(³ **中央値**) : データの値を小さい順に並べたときの中央の値

ただし, データの値の個数が偶数のときは, 中央にある 2 つの値の平均値とする。

例2 A 班の読書時間の中央値を求めてみよう。

データの値を小さい順に並べると, 次のようになる。

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

└──────────────────┘ └──────────────────┘

10人 10人

中央値は, (**10 番目の値と 11 番目**) の平均値である。

よって, $\frac{9+10}{2} = 9.5 \text{ (時間)}$

問4 教科書128ページのB班の読書時間を, 値の小さい順に並べると次のようになる。B班の読書時間の中央値を求めなさい。

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

**中央値は, 8 番目の値 9 であるから
9 時間**

データのなかに, ほかの値とかけ離れた値がある場合がある。このようなときは, 平均値よりも中央値の方がその影響を受けにくいので, 代表値としては, 平均値より中央値が適当である。

最頻値

(教科書 p.131)

(⁴ **最頻値**) : 度数分布表で, 度数が最も大きい階級の階級値。

ヒストグラムにおいて, 最も高い長方形に対応する階級値である。

例3 教科書 128 ページの A 班の読書時間の最頻値を求めてみよう。

右の度数分布表で, 度数が最も大きい階級は (**8**) 時間以上 (**12**) 時間未満である。
したがって, 最頻値はその階級値の (**10**) 時間である。

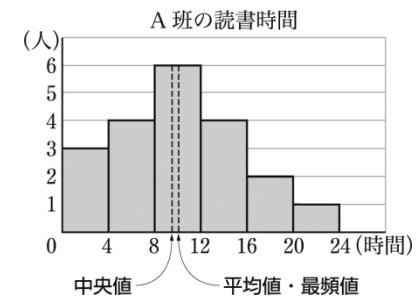
時間の階級 (時間)	階級値 (時間)	度数 (人)
0 以上 ~ 4 未満	2	3
4 ~ 8	6	4
8 ~ 12	10	6
12 ~ 16	14	4
16 ~ 20	18	2
20 ~ 24	22	1
計		20

問5 P.1 (教科書 129 ページ) の問 1 でつくった度数分布表から, B 班の読書時間の最頻値を求めなさい。

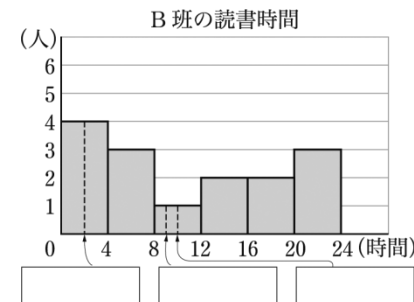
**度数が最も大きい階級は 0 時間以上 4 時間未満である。
したがって, 最頻値はその階級値であるから 2 時間**

例4 A 班と B 班について, 読書時間の 3 つの代表値を比べてみよう。

A 班の読書時間の 3 つの代表値をヒストグラムに表すと, 下の図のようになる。



問6 例 4 にならって, B 班の読書時間の 3 つの代表値を, ヒストグラムの に入れなさい。



(左から順に) 最頻値, 中央値, 平均値

3 四分位数と箱ひげ図

四分位数と四分位範囲

(教科書 p.132)

データの散らばりぐあいは、代表値ではとらえられない。

データの値で(最大値) - (最小値)を、そのデータの分布の(1 範囲)という。

範囲を調べて、データの散らばりぐあいを比べてみよう。

下の表はバスケットボール部のAさんとBさんの、最近10試合で成功したシュートの本数である。

Aさん	(単位 本)	Bさん	(単位 本)
5 3 6 5 6 7 8 7 8 10		4 5 13 5 9 6 6 7 6 5	

2人の成功したシュートの本数の範囲は、それぞれ

$$A \text{さん} (10 - 3 = 7) \text{ (本)}$$

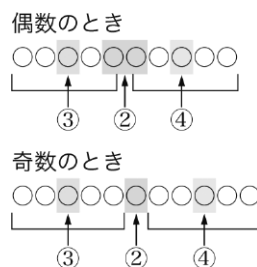
$$B \text{さん} (13 - 4 = 9) \text{ (本)}$$

範囲の値を比べると、散らばりぐあいはAさんよりBさんの方が大きいといえる。

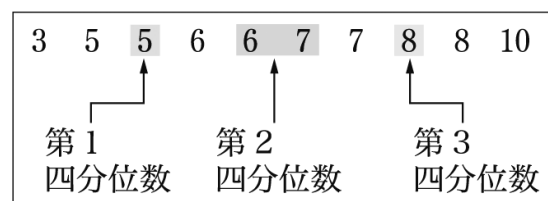
範囲は、データのなかにほかとかけ離れた値がある場合、その値の影響を受けやすい。その影響を少なくしたものに、(2 四分位範囲)がある。次のようにして求める。

- ① データの値を小さい順に並べ、中央値を境にして2つに分ける。
- ② 中央値を、(3 第2四分位数)という。
- ③ 最小値を含む方のデータの中央値を(4 第1四分位数)という。
- ④ 最大値を含む方のデータの中央値を(5 第3四分位数)という。
- ⑤ (第3四分位数) - (第1四分位数)が、四分位範囲である。

また、四分位範囲を2でわった値を(6 四分位偏差)という。



例5 Aさんの成功したシュートの本数を、小さい順に並べかえると右のようになるから



$$\text{第2四分位数} (\frac{6+7}{2} = 6.5) \text{ 本}$$

$$\text{第1四分位数} (5) \text{ 本}$$

$$\text{第3四分位数} (8) \text{ 本}$$

$$\text{四分位範囲} (8 - 5 = 3) \text{ 本}$$

$$\text{四分位偏差} (3 \div 2 = 1.5) \text{ 本}$$

問7 Bさんの成功したシュートの本数の、四分位範囲、四分位偏差を求めなさい。また、AさんとBさんの四分位範囲を比べて、どちらの散らばりぐあいが大きいか答えなさい。

Bさんの成功したシュートの本数を小さい順に並べかえると

4 5 5 5 6 6 6 7 9 13

第2四分位数は中央値であるから

$$\frac{6+6}{2} = 6 \text{ (本)}$$

第1四分位数は、最小値を含む方のデータの中央値であるから 5本

第3四分位数は、最大値を含む方のデータの中央値であるから 7本

したがって、四分位範囲は

$$7 - 5 = 2 \text{ (本)}$$

四分位偏差は

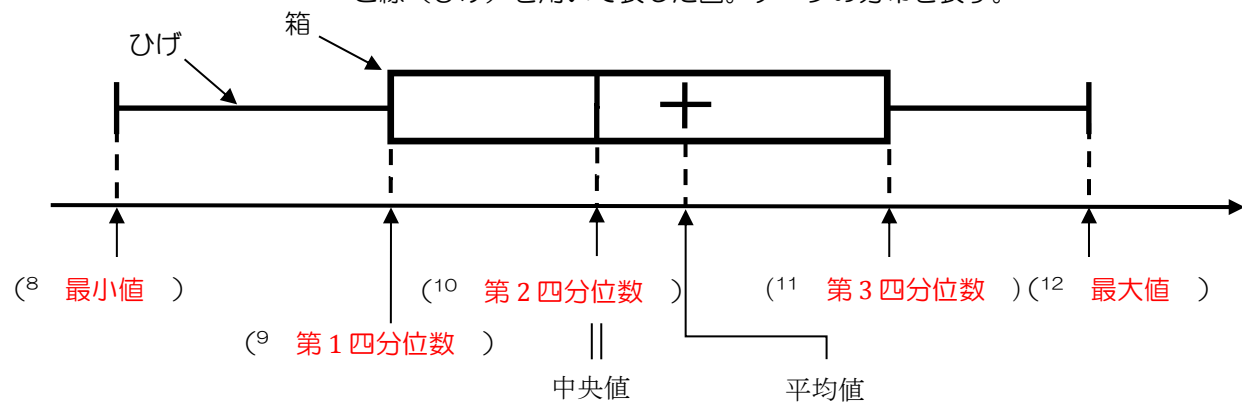
$$2 \div 2 = 1 \text{ (本)}$$

また、AさんとBさんの四分位範囲から、散らばりぐあいはAさんの方が大きい。

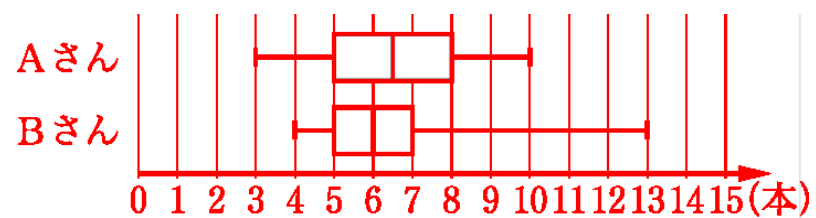
箱ひげ図

(教科書 p.133)

(7 箱ひげ図) : 最小値, 第1四分位数, 第2四分位数(中央値), 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。



例6 例5の結果から, 箱ひげ図は次のようになる。



問8 Bさんの成功したシュートの本数の箱ひげ図を, 上の図に表しなさい。

4 分散と標準偏差

(教科書 p.134)

データの散らばりぐあいを、データの個々の値と平均値との差を用いて考えてみよう。

(¹ **偏差**) : (データの個々の値) - (平均値)

下の表は、ある野球チームのA投手と、B投手が、最近5試合で奪った三振の個数である。

(単位 個)

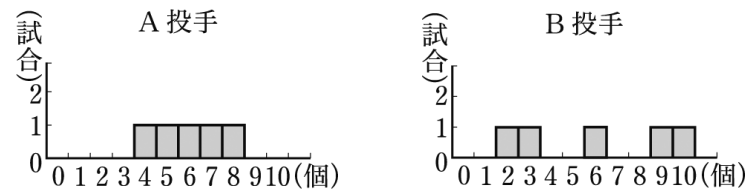
	1 試合目	2 試合目	3 試合目	4 試合目	5 試合目
A 投手	4	8	7	5	6
B 投手	10	2	9	6	3

平均値は、それぞれ

$$\text{A 投手} \quad \frac{4+8+7+5+6}{5} = \frac{30}{5} = 6 \text{ (個)}$$

$$\text{B 投手} \quad \frac{10+2+9+6+3}{5} = \frac{30}{5} = 6 \text{ (個)}$$

となり、等しくなる。しかし、三振の個数をヒストグラムに表すと、散らばりぐあいは異なる。



次に、A投手、B投手の三振の個数の偏差とその合計を求めると下の表のようになる。

	1 試合目	2 試合目	3 試合目	4 試合目	5 試合目	計
A 投手の偏差	-2	2	1	-1	0	0
B 投手の偏差	4	-4	3	0	-3	0

偏差の合計はどちらも0になるため、偏差の平均値でデータ全体の散らばりぐあいを表すことはできない。

そこで、偏差を2乗した値を考える。

偏差の2乗は0以上の値になるので、合計が0に近いほど平均値からの散らばりぐあいが小さい。

(² **分散**) : 偏差の2乗の平均値。 s^2 で表す。

(³ **標準偏差**) : 分散の正の平方根。 s で表す。

分散と標準偏差	
偏差	(データの個々の値) - (平均値)
分散	$s^2 = (\text{偏差})^2$ の平均値 = $\frac{(\text{偏差})^2 \text{の合計}}{\text{データの値の個数}}$
標準偏差	$s = \sqrt{s^2} = (\text{分散の正の平方根})$

例7 教科書 134 ページのA投手が奪った三振の個数から偏差の2乗の合計を計算すると、右の表のようになる。したがって、分散 s^2 は

$$s^2 = \frac{10}{5} = 2$$

標準偏差 s は

$$s = \sqrt{2} = 1.4142 \dots \approx 1.41 \text{ (個)}$$

A 投手	三振の個数	偏差	(偏差) ²
1 試合目	4	-2	4
2 試合目	8	2	4
3 試合目	7	1	1
4 試合目	5	-1	1
5 試合目	6	0	0
計	30	0	10

問9 右の表を完成して、B投手が奪った三振の個数の分散を求めなさい。また、標準偏差を、四捨五入して小数第2位まで求めなさい。

教科書 134 ページより、B投手が奪った三振の個数の平均値は6個である。このことから、偏差、偏差の2乗を計算すると、表のようになる。

分散を s^2 とすると

$$s^2 = \frac{50}{5} = 10$$

標準偏差を s とすると

$$\begin{aligned} s &= \sqrt{10} \\ &= 3.1622 \dots \\ &\approx 3.16 \text{ (個)} \end{aligned}$$

B 投手	三振の個数	偏差	(偏差) ²
1 試合目	10	4	16
2 試合目	2	-4	16
3 試合目	9	3	9
4 試合目	6	0	0
5 試合目	3	-3	9
計	30	0	50

問10 A投手とB投手が奪った三振の個数の標準偏差を比べて、どちらの散らばりぐあいが大きいか答えなさい。

A投手とB投手が奪った三振の個数の標準偏差から、散らばりぐあいはB投手の方が大きい。

5 相関関係

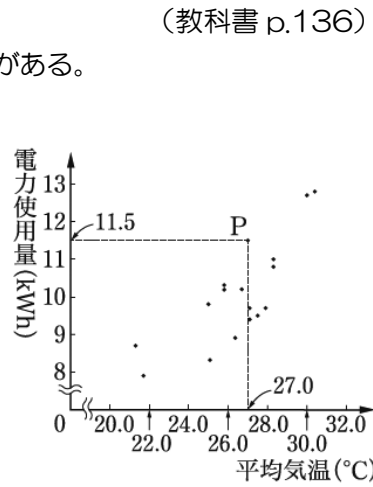
散布図

平均気温と電力使用量のように、互いに関係があると考えられる数量がある。

このような数量を組にしたデータについて考えてみよう。

(¹ **散布図**) : 右の図のように、2つのデータの値の組を x 座標、 y 座標とする点として平面上に表した図。

右の図は、夏の1日ごとの平均気温と世帯あたりの電力使用量を散布図に表したものである。図中の点Pは、平均気温が27°Cで、電力使用量が11.5kWhであることを表している。



相関関係

(教科書 p.136)

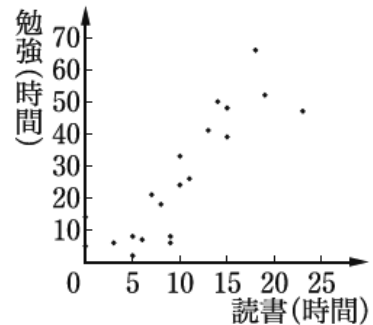
下の表は、20人の生徒について、先月の読書時間と、勉強時間、テレビ視聴時間、1日のメール発信回数の平均を調べたデータである。

読書(時間)	3	10	7	14	5	9	15	0	9	18	0	8	11	10	15	19	6	23	13	5
勉強(時間)	6	33	21	50	2	8	39	5	6	66	14	18	26	24	48	52	7	47	41	8
テレビ視聴(時間)	58	37	47	26	55	52	48	38	56	4	48	25	36	31	24	19	53	18	20	62
メール発信(回)	5	2	9	7	8	6	8	3	2	2	2	6	12	3	9	6	11	1	4	1

読書時間と勉強時間を散布図に表すと、右の図のようになる。

この散布図から、読書時間と勉強時間は、一方が増加すれば他方も増加する傾向があることがわかる。

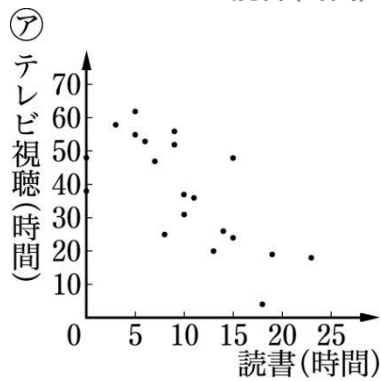
(² **正の相関関係**) : 一方が増加すれば他方も増加する傾向。



読書時間とテレビ視聴時間を散布図に表すと、右の図②のようになる。

この散布図から、読書時間とテレビ視聴時間は、一方が増加すれば他方が減少する傾向があることがわかる。

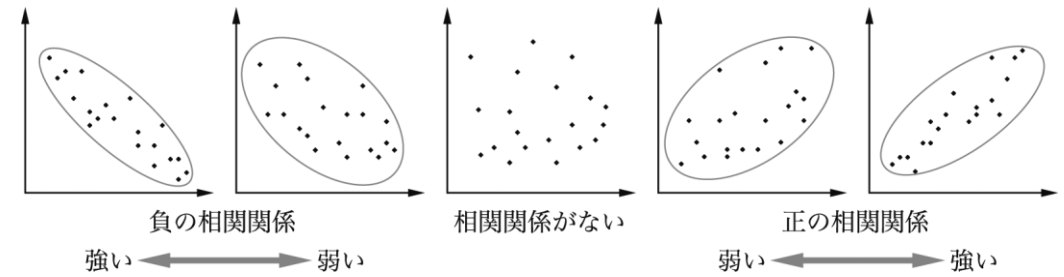
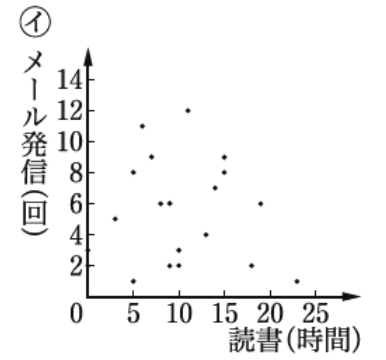
(³ **負の相関関係**) : 一方が増加すれば他方が減少する傾向



読書時間とメール発信回数を散布図に表すと、右の図①のようになる。

読書時間とメール発信回数には正の相関関係も負の相関関係もみられない。

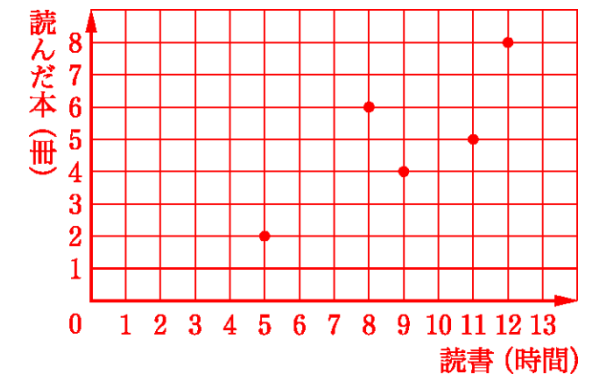
(⁴ **相関関係がない**) : 正の相関も負の相関もみられない。



問 11 次の表は、5人の生徒の、先月の読書時間と読んだ本の冊数を示したものである。散布図をつくり、読書時間と読んだ本の冊数には、どのような相関関係があるか答えなさい。

生徒	読書(時間)	読んだ本(冊)
a	5	2
b	8	6
c	9	4
d	12	8
e	11	5

正の相関関係がある。



6 相関係数

相関係数の意味

(教科書 p.138)

相関関係を調べたい2つの数量を x, y とする。 x の偏差と y の偏差の積の平均値を⁽¹⁾ **共分散** という。また、共分散を x の標準偏差と y の標準偏差の積でわった値を⁽²⁾ **相関係数** という。相関係数は記号 r で表す。

共分散と相関係数	
共分散	x と y の偏差の積の平均値
相関係数	$r = \frac{\text{共分散}}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})}$

相関係数 r の値については、次の不等式が成り立つ。

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

⁽³⁾ **正**) の相関関係が強いほど r の値は 1 に近づき、⁽⁴⁾ **負**) の相関関係が強いほど -1 に近づく。

問 12 プロ野球 12 チームについて、1 年間の勝利数と失点数の相関係数は -0.73 、勝利数と得失点差の相関係数は 0.90 である。

次の①、②から、正しいものを選びなさい。

(1) 勝利数と失点数には

- ① 正の相関関係がある
- ② 負の相関関係がある

勝利数と失点数の相関係数は -0.73 であるから

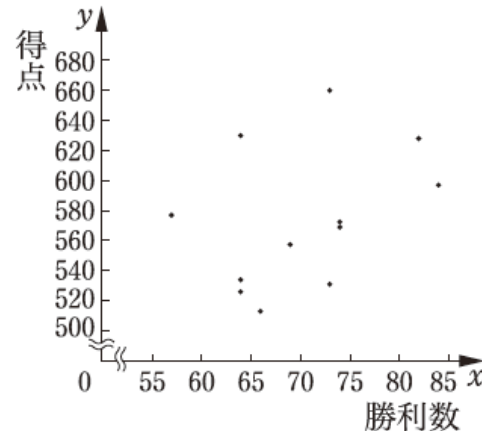
② 負の相関関係がある

(2) 勝利数と得点より、勝利数と得失点差の方が

- ① 正の相関関係が強い
- ② 正の相関関係が弱い

教科書138ページより、勝利数と得点の相関係数は 0.37 であり、勝利数と得失点差の相関係数は 0.90 であるから

① 正の相関関係が強い



相関係数

(教科書 p.139)

例8 読書時間と読んだ本の冊数の相関係数を求めてみよう。教科書 137 ページの問 11 について、読書時間を x とし、読んだ本の冊数を y とする。

$$(x \text{ の平均値}) = \frac{5+8+9+12+11}{5} = 9 \text{ (時間)}$$

$$(y \text{ の平均値}) = \frac{2+6+4+8+5}{5} = 5 \text{ (冊)}$$

下の表のようにして計算すると

生徒	x	y	x の偏差	y の偏差	$(x \text{ の偏差})^2$	$(y \text{ の偏差})^2$	偏差の積
a	5	2	-4	-3	16	9	12
b	8	6	-1	1	1	1	-1
c	9	4	0	-1	0	1	0
d	12	8	3	3	9	9	9
e	11	5	2	0	4	0	0
計	45	25	0	0	30	20	20

$$x, y \text{ の共分散は } \frac{20}{5} = 4$$

$$x \text{ の標準偏差は } \sqrt{\frac{30}{5}} = \sqrt{6}$$

$$y \text{ の標準偏差は } \sqrt{\frac{20}{5}} = \sqrt{4} = 2$$

相関係数 r は

$$r = \frac{4}{\sqrt{6} \times 2} = \frac{\sqrt{6}}{3} = 0.816 \dots \approx 0.82$$

問 13 次の表は、4人の生徒の数学と英語の小テストの得点を示したものである。表を完成することにより、数学の得点 x と英語の得点 y の相関係数を求めなさい。

生徒	x	y	x の偏差	y の偏差	$(x \text{ の偏差})^2$	$(y \text{ の偏差})^2$	偏差の積
a	6	5	-1	-2	1	4	2
b	7	5	0	-2	0	4	0
c	7	8	0	1	0	1	0
d	8	10	1	3	1	9	3
計	28	28	0	0	2	18	5

$$(x \text{ の平均値}) = \frac{6+7+7+8}{4} = 7 \text{ (点)}$$

$$(y \text{ の平均値}) = \frac{5+5+8+10}{4} = 7 \text{ (点)}$$

x , y の共分散は

$$\frac{5}{4}$$

x の標準偏差は

$$\sqrt{\frac{2}{4}} = \frac{\sqrt{2}}{2}$$

y の標準偏差は

$$\sqrt{\frac{18}{4}} = \frac{3\sqrt{2}}{2}$$

相関係数を r とすると

$$r = \frac{5}{4} \div \left(\frac{\sqrt{2}}{2} \times \frac{3\sqrt{2}}{2} \right)$$

$$= \frac{5}{6}$$

$$= 0.833 \dots$$

$$\approx 0.83$$

復習問題

(教科書 p.140)

1 次の資料は、1年生の図書委員8人が1年間に学校の図書室から借りた本の冊数を調べたものである。

7 12 8 40 9 4 8 8

(1) 借りた本の冊数の平均値、中央値を求めなさい。

平均値は

$$\begin{aligned} & \frac{7+12+8+40+9+4+8+8}{8} \\ &= \frac{96}{8} \\ &= 12 \text{ (冊)} \end{aligned}$$

データの値を小さい順に並べると

4 7 8 8 8 9 12 40

中央値は、4番目と5番目の平均値であるから

$$\frac{8+8}{2} = 8 \text{ (冊)}$$

(2) 40冊借りた生徒を除いて、残り7人が借りた本の冊数の平均値、中央値を求めなさい。

平均値は

$$\begin{aligned} & \frac{7+12+8+9+4+8+8}{7} \\ &= \frac{56}{7} \\ &= 8 \text{ (冊)} \end{aligned}$$

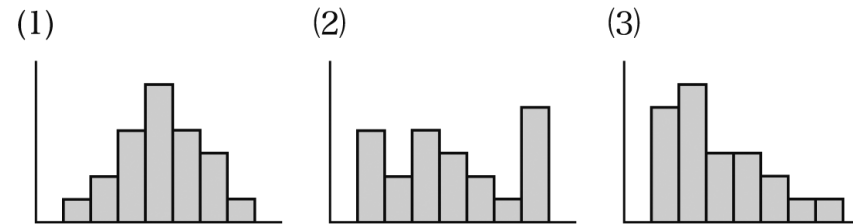
データの値を小さい順に並べると

4 7 8 8 8 9 12

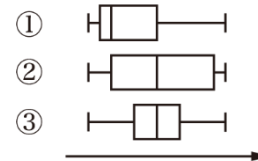
中央値は、4番目の値8であるから

$$8 \text{ (冊)}$$

2 次のヒストグラムについて、対応する箱ひげ図を選びなさい。



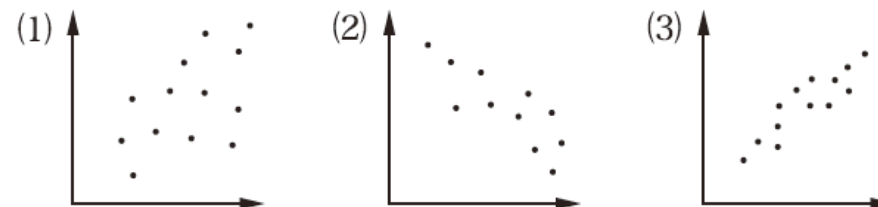
箱ひげ図



(1) ③ (2) ② (3) ①

3 次の散布図について、対応する相関係数を選びなさい。

散布図



相関係数

① 0.9 ② 0.5 ③ -0.8

(1) ② (2) ③ (3) ①