1節 データの整理と分析

1 データの整理

データ

（教科書 p.158）

次の資料は，あるクラスの「一か月の読書時間」について，電車・バス通学の A 班と徒歩・自転車通学の B 班に分けて調べた。

A 班 20 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

B 班 15 人（単位 時間）

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

このような資料を（[1]　　　　　）という。

また，読書時間のように，データの特性を表す数量を（[2]　　　　　）という。

データの整理

（教科書 p.158）

右の表は，A 班の読書時間の結果をもとに，0 時間から 24 時間までの間を 4 時間ずつの区間に分け，その区間に入っている人数を調べてまとめたものである。

読書時間（時間）	度数
以上～未満 0 ～ 4	3
4 ～ 8	4
8 ～ 12	6
12 ～ 16	4
16 ～ 20	2
20 ～ 24	1
計	20

（[3]　　　　　）：データを整理するために用いる区間

（[4]　　　　　）：区間の幅

（[5]　　　　　）：階級の真ん中の値

（[6]　　　　　）：階級に入っているデータの値の個数

（[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

差 4 は（[9]　　　　　）

真ん中の値 22 は（[10]　　　　　）

問1 B 班の読書時間の度数分布表を作成せよ。

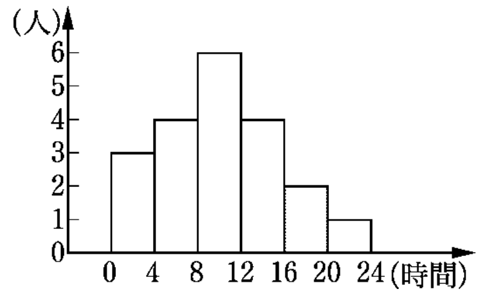
読書時間（時間）	度数
以上～未満 ～	
～	
～	
～	
～	
～	
計	

度数分布をグラフにした図を

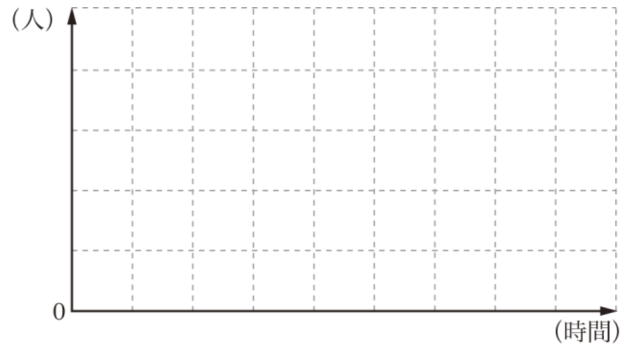
（[11]　　　　　）という。

A 班の読書時間のヒストグラムは右の図のようになる。

度数分布表やヒストグラムを用いるとデータの分布が見やすくなる。



問2 B 班の読書時間のヒストグラムをかけ。



相対度数

（教科書 p.159）

A 班とB 班は人数が異なるから，度数を見るだけではデータを比較しにくい。そこで，度数の代わりに，各階級の度数を度数の合計で割った値を用いるとよい。

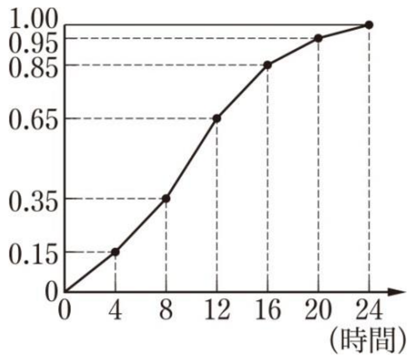
(12　　　　　　　)：各階級の度数を度数の合計で割った値

$$(^{13}\qquad\qquad) = \frac{(^{14}\qquad\qquad)}{(^{15}\qquad\qquad)}$$

(16　　　　　　　)：相対度数を小さい階級からその階級の値まで合計して得られる値

(17　　　　　　　)：各階級の累積相対度数を折れ線でつないだもの

読書時間（時間）	度数	相対度数	累積相対度数
以上～未満 0 ～ 4	3	0.15	0.15
4 ～ 8	4	0.20	0.35
8 ～ 12	6	0.30	0.65
12 ～ 16	4	0.20	0.85
16 ～ 20	2	0.10	0.95
20 ～ 24	1	0.05	1.00
計	20	1.00	



問3 問１で作成した度数分布表に B 班の読書時間の相対度数と累積相対度数を付け加えよ。

読書時間（時間）	度数	相対度数	累積相対度数
以上～未満 ～			
～			
～			
～			
～			
～			
計			

2 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

（教科書 p.160）

x を変量とし，データの n 個の値 x_1，x_2，…，x_n が与えられているとき

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

$$\frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

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

例1 教科書 158 ページの A 班の読書時間の平均値は

問4 教科書158 ページの B 班の読書時間の平均値を求めよ。

度数分布から平均値を求める方法を考えてみよう。

右の表は，変量 x の度数分布表である。

平均値 $\overline{x}$ は各階級に含まれるデータの値がすべてその階級値に等しいとみなして，次の式で計算する。

階級値	度数
x_1	f_1
x_2	f_2
⋮	⋮
x_r	f_r
計	n

$$\overline{x} = \frac{1}{n}(x_1 f_1 + x_2 f_2 + \cdots + x_r f_r)$$

例2 教科書 158 ページの A 班の読書時間の平均値を度数分布表から求めてみよう。

右の表より

階級値	度数
2	3
6	4
10	6
14	4
18	2
22	1
計	20

問5 教科書158 ページの問1で作成した度数分布表を用いて，B 班の読書時間の平均値を小数第 2 位を四捨五入して，小数第 1 位まで求めよ。

中央値

（教科書 p.161）

（[3]　　　　　）：データの値を小さい順に並べたとき，中央の順位にくる値。

（[4]　　　　　）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 $n+1$ 番目のデータの値の平均値を中央値とする。

例3 教科書 158 ページの A 班の読書時間の中央値を求めてみよう。

20 個のデータの値を小さい方から順に並べると

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

このデータの中央値は，10 番目の値 9 と 11 番目の値 10 の平均値

問6 教科書158 ページの B 班の読書時間の中央値を求めよ。

最頻値

（教科書 p.161）

（5　　　　　　　　）：度数分布表で，度数が最も多い階級の階級値。（　　　　　　　）ともいう。

例 4 教科書 158 ページの A 班の読書時間の最頻値を求めてみよう。

度数分布表で，読書時間の度数が最も大きい階級は

（　　）時間以上（　　　）時間未満なので

最頻値は（　　　）時間である。

読書時間（時間）	度数
以上〜未満 0 〜 4	3
4 〜 8	4
8 〜 12	6
12 〜 16	4
16 〜 20	2
20 〜 24	1
計	20

問7 教科書158 ページの問１で作成した度数分布表を用いて，B 班の読書時間の最頻値を求めよ。

3 箱ひげ図

四分位数

（教科書 p.162）

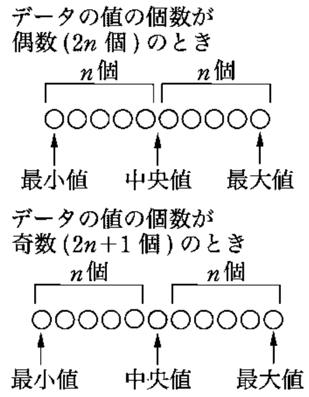
データの特徴をよりくわしく表すために，データの値を小さい方から順に並びかえて，次のような数値を考える。

① データの中央値を求める。

② 右の図のように，中央値を境にしてデータの値の個数が等しくなるように２つの部分に分ける。

③ ２つに分けたうち，最小値を含む方のデータの中央値を求める。

④ ２つに分けたうち，最大値を含む方のデータの中央値を求める。



①の値を（1　　　　　　　　），③の値を（2　　　　　　　　），

④の値を（3　　　　　　　　），

これらを合わせて（4　　　　　　　　）という。

３つの四分位数の値は，データの値の小さい方から 25%，50%，75% に対応する数値である。

例 5 教科書 158 ページの A 班の読書時間の四分位数を求めてみよう。

20 個のデータの値を小さい方から順に並べると，

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

第１四分位数は

第２四分位数は

第３四分位数は

問8 教科書158 ページの B 班の読書時間の四分位数を求めよ。

【B 班 15 人（単位　時間）】

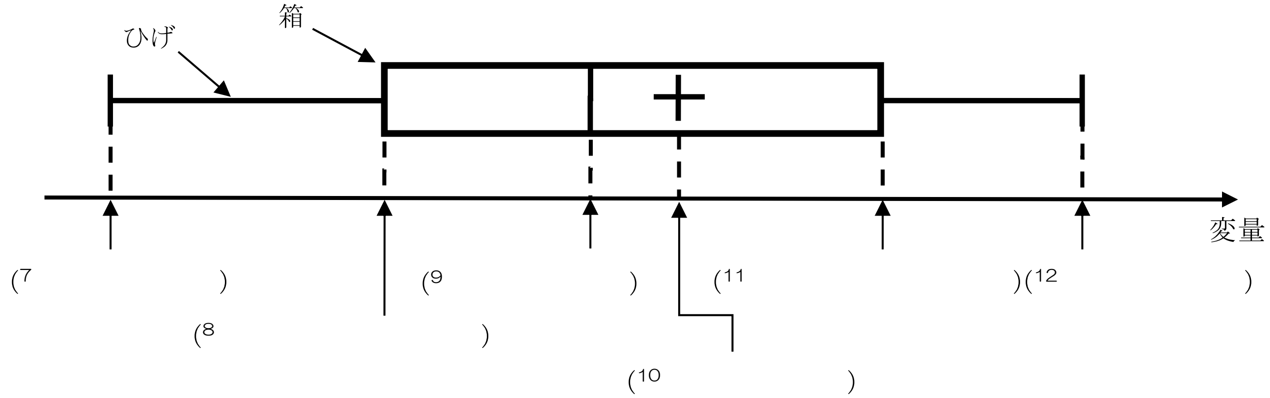
20	6	0	14	16	23	1	4	5	0
18	13	21	0	9					

箱ひげ図

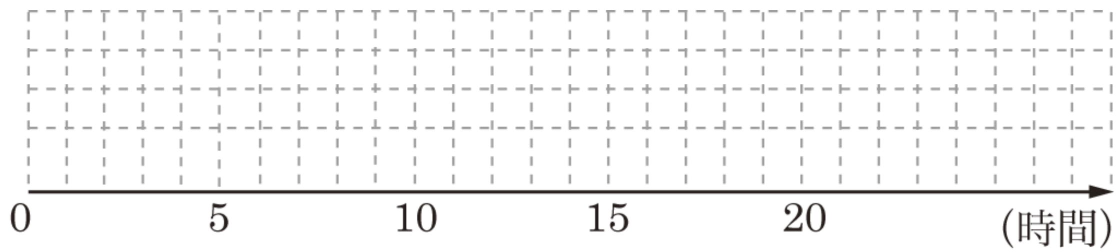
（教科書 p.163）

（5　　　　　　　　）：データの分布を最小値，第１四分位数，中央値（第2 四分位数)，第３四分位数，最大値の５つの数値を用いて要約する方法。

（6　　　　　　　　）：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。

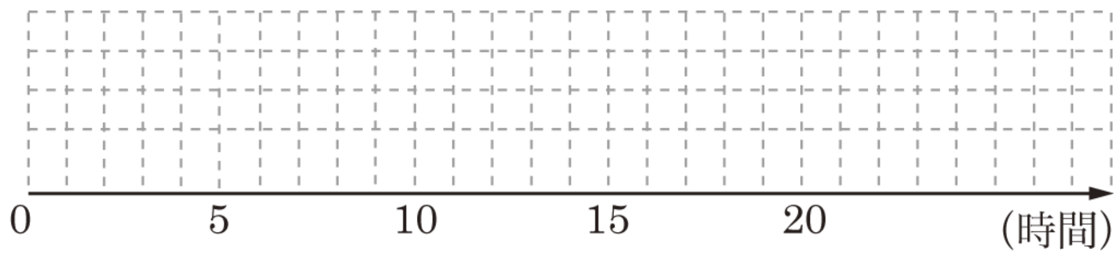


例 6



教科書 158 ページの A 班の読書時間の箱ひげ図をかいてみよう。

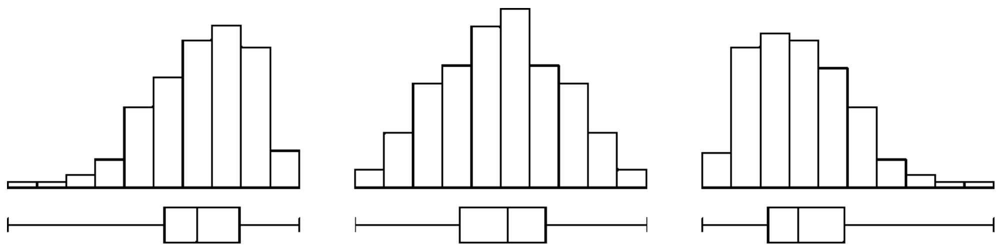
問9 教科書158 ページの B 班の読書時間の箱ひげ図をかけ。



箱ひげ図とヒストグラム

（教科書 p.163）

箱ひげ図はヒストグラムと同様に，データの分布を表現するのに適している。ヒストグラムでは，度数分布表のすべての階級の度数が必要であるのに対して，箱ひげ図は最小値，第 1 四分位数，中央値，第 3 四分位数，最大値の 5 つの数値がわかると，かくことができる。

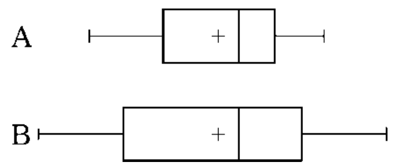


上の図のように，ヒストグラムの山の高い部分に箱ひげ図の箱が対応し，山のすその部分に箱ひげ図のひげが対応している。

4 箱ひげ図とデータの散らばり

（教科書 p.164）

右の図は，データの値の個数も平均値も中央値も等しい 2 つのデータ A，B の箱ひげ図である。



このように，平均値や中央値は等しくても，データの分布のようすが大きく異なることがある。箱ひげ図を用いることによって，データの分布のようすを視覚的に比較することを考える。

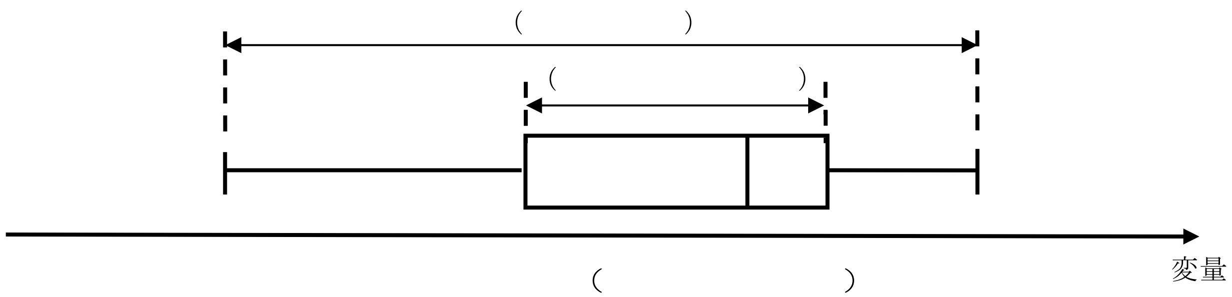
範囲と四分位数

（教科書 p.164）

（[1]　　　　　）：データの最大値から最小値を引いた値。（[2]　　　　　）ともいう。

（[3]　　　　　　　）：第 3 四分位数から第 1 四分位数を引いた値。

（[4]　　　　　　　）：四分位範囲を 2 で割った値。



$$（\qquad\qquad） = \frac{（\qquad\qquad）}{2}$$

例 7 教科書 158 ページの A 班の読書時間の範囲，四分位範囲，四分位偏差を求めよ。

範囲は　　　　　　　　　　(時間)

四分位範囲は　　　　　　　　(時間)

四分位偏差は　　　　　(時間)

問 10 教科書158 ページの B 班の読書時間の範囲，四分位範囲，四分位偏差をそれぞれ求めよ。

下の図 1 はある年の東京における各月の日ごとの平均気温の平均値を折れ線グラフで表したもの，図 2 は各月の日ごとの平均気温の分布を箱ひげ図で表したものである。

図 1

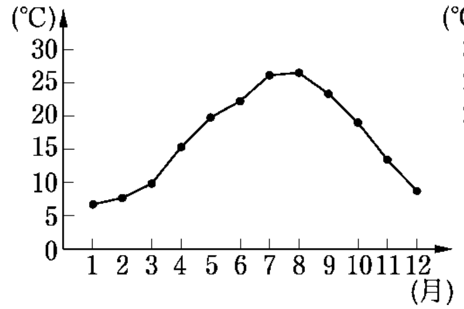
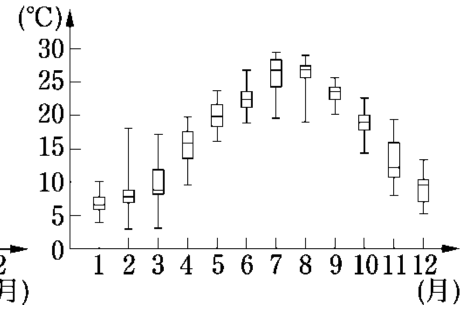
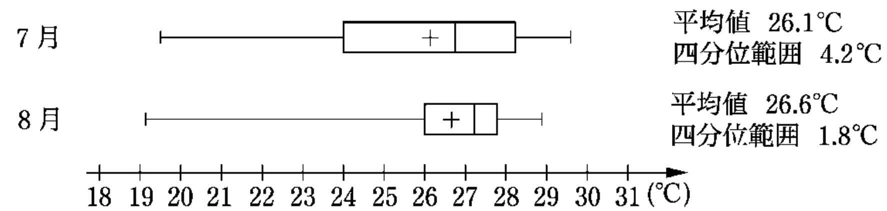


図 2



平均値のみでは，各月の日ごとの平均気温の分布のようすを比較することができないが，箱ひげ図を用いると，分布のようすを比較することが可能である。

たとえば，7 月と 8 月の箱ひげ図は次のようになる。



7 月と 8 月の平均値や範囲はあまり変わらないが，8 月の方が四分位範囲が小さい。

ゆえに，8 月の方が，平均気温の散らばりが小さいことがわかる。

問 11 次の文章は上の図 2 の 1 月と 2 月の日ごとの平均気温の分布について述べたものである。この文章が適切であるかどうかを答えよ。

「1 月と 2 月において，1 日の平均気温が 5℃ 以上 10℃ 以下であった日はそれぞれ月の半分以上である。」

5 分散と標準偏差

（教科書 p.166）

データの散らばり具合を数値で表すために，データの個々の値と平均値の差に着目する。
データの個々の値と平均値との差が大きいほど，その値は平均値から離れていることになる。
データの n 個の値を x_1，x_2，…，x_n とし，その平均値を $\overline{x}$ とする。

平均値からの（1　　　　　）：　各値の平均値との差 $x_1-\overline{x}$，$x_2-\overline{x}$，…，$x_n-\overline{x}$

偏差の平均値を計算すると，

$$\frac{1}{n}\{(x_1-\overline{x})+(x_2-\overline{x})+\cdots+(x_n-\overline{x})\}=\frac{1}{n}\{(x_1+x_2+\cdots+x_n)-n\overline{x}\}$$

$$=\frac{1}{n}(x_1+x_2+\cdots+x_n)-\overline{x}=\overline{x}-\overline{x}=0$$

となり，偏差の平均値ではデータの散らばり具合を表すことはできない。
そこで，偏差 2 乗した値 $(x_i-\overline{x})^2$ を考える。

（2　　　　　）：偏差の 2 乗の平均値

分散は，計算の過程で数値を 2 乗するため単位も変わるから，分散の正の平方根を考える。

（3　　　　　）：分散の正の平方根

分散と標準偏差	
分散	$s^2=\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}$
標準偏差	$\boldsymbol{s}=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}}$
	ただし，$\overline{x}$ は平均値

例 8 A 社と B 社の携帯音楽プレイヤーを充電してからの連続使用時間を 5 回調べた。

A 社　24　28　26　22　25　（単位　時間）

B 社　27　28　25　21　29　（単位　時間）

A 社の音楽プレーヤーの連続使用時間の標準偏差を求める。
連続使用時間を x として，平均値 $\overline{x}$ を求めると

連続使用時間の各値の偏差は

連続使用時間	24	28	26	22	25
連続使用時間の偏差					

分散 s^2 は

標準偏差 s は

問 12 例8における B 社の音楽プレイヤーの連続使用時間の標準偏差を小数第 3 位を四捨五入して小数第 2 位まで求めよ。ただし，$\sqrt{2}=1.414$ とする。

度数分布から標準偏差を求める方法を考える。
変量 x の平均値を $\overline{x}$ とすると，x の分散 s^2 は

$$s^2=\frac{1}{n}\{(x_1-\overline{x})^2f_1+(x_2-\overline{x})^2f_2+\cdots+(x_r-\overline{x})^2f_r\}$$

よって，x の標準偏差 s は

$$\boldsymbol{s}=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2f_1+(x_2-\overline{x})^2f_2+\cdots+(x_r-\overline{x})^2f_r\}}$$

階級値	度数
x_1	f_1
x_2	f_2
⋮	⋮
x_r	f_r
計	n

例題 1 右の度数分布表は，ある高校のバスケットボール部の部員 10 人でフリースローを 5 回ずつ行ったゲームの結果である。このゲームの得点の標準偏差を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。

得点 x	度数 f
0	1
1	0
2	3
3	1
4	4
5	1
計	10

解

点数 x	度数 f	xf	$x-\overline{x}$	$(x-\overline{x})^2$	$(x-\overline{x})^2f$
0	1				
1	0				
2	3				
3	1				
4	4				
5	1				
計	10				

平均値$\overline{x}$ は

分散 s^2 は

標準偏差 s は

問 13 下の表は，あるクラス 40 人に数学の小テストを行った結果である。このクラスの小テストの点数の標準偏差を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。ただし，$\sqrt{3}=1.732$ とする。

点数 x	0	1	2	3	4	5	6	7	8	9	10	計
人数	2	4	7	8	4	6	2	1	2	3	1	40

分散と平均値の関係式

（教科書 p.169）

分散 s^2 を表す式は，次のように変形できる。

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

$$=\frac{1}{n}\{(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2\overline{x}(x_1+x_2+\cdots+x_n)+n(\overline{x})^2\}$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2\overline{x}\cdot\frac{1}{n}(x_1+x_2+\cdots+x_n)+(\overline{x})^2$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2(\overline{x})^2+(\overline{x})^2$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-(\overline{x})^2$$

一般に，変量 x の分散について，次のことが成り立つ。

（4　　　　　　　）＝（5　　　　　　　）－（6　　　　　　　）

例 9 下の表は，高校 1 年生の A さんが毎朝通学に利用しているバスの乗車時間を 6 日間調べた結果である。

バスの乗車時間 x（分）	9	13	10	9	10	12

x の平均値 $\overline{x}$ と x^2 の平均値 $\overline{x^2}$ は，次のようになる。

よって分散 s^2 は

問 14 下の表は，A さんの 1 日の学習時間を 5 日間記録したものである。

1 日の学習時間 x（時間）	1	2	3	1	3

このとき，1 日の学習時間の分散を求めよ。

問 題

（教科書 p.170）

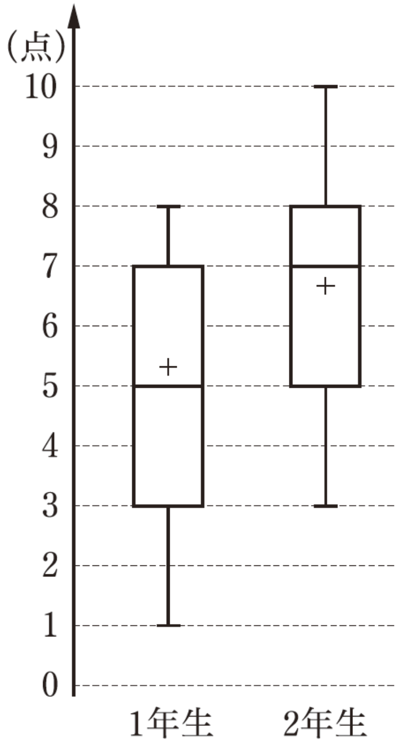
1 右の図は，ある学校の 1 年生と 2 年生それぞれ 20 人に対して，10 点満点の漢字テストを実施したときの得点の箱ひげ図である。このとき，次の①～④のうち，右の図から必ずしも正しいとは限らないものをすべて選べ。

① 1 年生の得点のデータの方が 2 年生の得点のデータより四分位範囲が大きい。

② どちらの学年にも，平均点以下の生徒が 10 人以上いる。

③ どちらの学年にも，得点が 5 点以上の生徒が 10 人以上いる。

④ どちらの学年にも，得点がちょうど 8 点の生徒がいる。



2 くじを 20 回引いて，当たった回数だけ得点できるゲームがある。右の表は，ある学校の生徒 5 人がこのゲームを行ったときの得点を記録したものである。ただし，生徒 4 の得点は 5 人の得点の平均値以下であった。このとき，次の問に答えよ。

	得点
生徒 1	8
生徒 2	14
生徒 3	10
生徒 4	a
生徒 5	$18-a$
平均値	m
分散	6

(1) 5 人の得点の平均値 m を求めよ。

(2) 表中の a の値を求めよ。

(3) 右の表の得点に対して，全員に 10 点を加えたとき，分散の値はどのように変化するか。①～③のうちから適するものを選べ。

① 大きくなる　　② 変わらない　　③ 小さくなる

1 節 データの整理と分析

1 データの整理

データ

（教科書 p.158）

次の資料は，あるクラスの「一か月の読書時間」について，電車・バス通学の A 班と徒歩・自転車通学の B 班に分けて調べた。

A 班 20 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

B 班 15 人（単位 時間）

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

このような資料を（[1] データ ）という。

また，読書時間のように，データの特性を表す数量を（[2] 変量 ）という。

データの整理

（教科書 p.158）

右の表は，A 班の読書時間の結果をもとに，0 時間から 24 時間までの間を 4 時間ずつの区間に分け，その区間に入っている人数を調べてまとめたものである。

読書時間（時間）	度数
以上～未満 0 ～ 4	3
4 ～ 8	4
8 ～ 12	6
12 ～ 16	4
16 ～ 20	2
20 ～ 24	1
計	20

（[3] 階級 ）：データを整理するために用いる区間

（[4] 階級の幅 ）：区間の幅

（[5] 階級値 ）：階級の真ん中の値

（[6] 度数 ）：階級に入っているデータの値の個数

（[7] 度数分布 ）：各階級に度数を対応させたもの

（[8] 度数分布表 ）：度数分布を表にしたもの

差 4 は（[9] 階級の幅 ）

真ん中の値 22 は（[10] 階級値 ）

問1 B 班の読書時間の度数分布表を作成せよ。

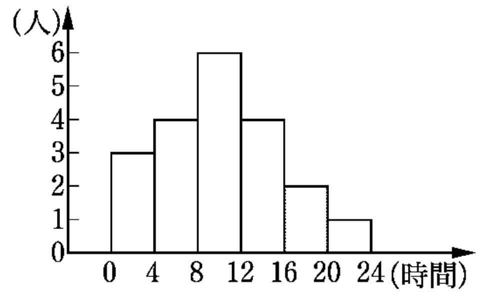
読書時間（時間）	度数
以上～未満 0～4	4
4～8	3
8～12	1
12～16	2
16～20	2
20～24	3
計	15

度数分布をグラフにした図を

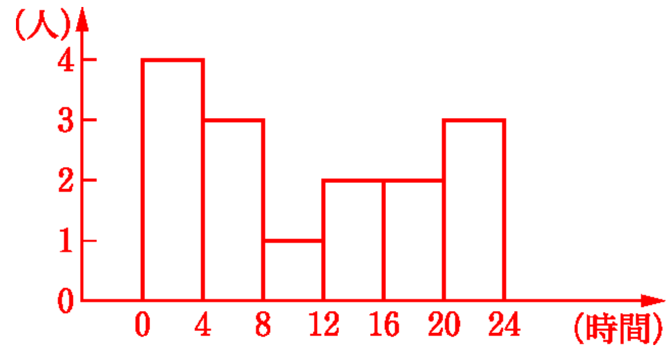
（[11] ヒストグラム ）という。

A 班の読書時間のヒストグラムは右の図のようになる。

度数分布表やヒストグラムを用いるとデータの分布が見やすくなる。



問2 B 班の読書時間のヒストグラムをかけ。



相対度数

（教科書 p.159）

A 班とB 班は人数が異なるから，度数を見るだけではデータを比較しにくい。そこで，度数の代わりに，各階級の度数を度数の合計で割った値を用いるとよい。

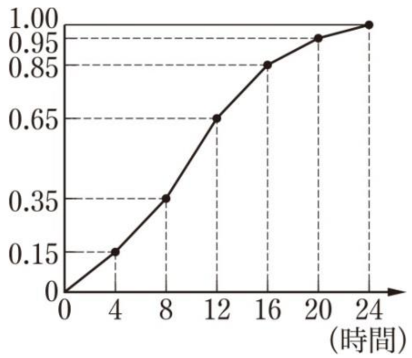
（[12] 相対度数 ）：各階級の度数を度数の合計で割った値

（[13] 相対度数 ）＝ （[14] その階級の度数 ）／（[15] 度数の合計 ）

（[16] 累積相対度数 ）：相対度数を小さい階級からその階級の値まで合計して得られる値

（[17] 累積相対度数折れ線 ）：各階級の累積相対度数を折れ線でつないだもの

読書時間（時間）	度数	相対度数	累積相対度数
以上～未満 0 ～ 4	3	0.15	0.15
4 ～ 8	4	0.20	0.35
8 ～ 12	6	0.30	0.65
12 ～ 16	4	0.20	0.85
16 ～ 20	2	0.10	0.95
20 ～ 24	1	0.05	1.00
計	20	1.00	



問3 問１で作成した度数分布表にＢ班の読書時間の相対度数と累積相対度数を付け加えよ。

読書時間（時間）	度数	相対度数	累積相対度数
以上～未満 0～4	4	0.27	0.27
4～8	3	0.20	0.47
8～12	1	0.07	0.54
12～16	2	0.13	0.67
16～20	2	0.13	0.80
20～24	3	0.20	1.00
計	15	1.00	

2 代表値

（1　代表値　）：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

（教科書 p.160）

x を変量とし，データの n 個の値 x_1，x_2，…，x_n が与えられているとき

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

$$\frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

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

例1 教科書 158 ページの A 班の読書時間の平均値は

$$\frac{1}{20}(3 + 10 + 7 + \cdots + 23 + 9 + 5) = \frac{1}{20} \cdot 200 = 10\text{(時間)}$$

問4 教科書158 ページの B 班の読書時間の平均値を求めよ。

B 班の読書時間の平均値は

$\frac{1}{15}(20 + 6 + 0 + 14 + \cdots + 9) = \frac{1}{15} \cdot 150$

$= 10$ (時間)

度数分布から平均値を求める方法を考えてみよう。

右の表は，変量 x の度数分布表である。

平均値 $\overline{x}$ は各階級に含まれるデータの値がすべてその階級値に等しいとみなして，次の式で計算する。

階級値	度数
x_1	f_1
x_2	f_2
⋮	⋮
x_r	f_r
計	n

$$\overline{x} = \frac{1}{n}(x_1 f_1 + x_2 f_2 + \cdots + x_r f_r)$$

例2 教科書 158 ページの A 班の読書時間の平均値を度数分布表から求めてみよう。

右の表より

$$\frac{1}{20}(2 \cdot 3 + 6 \cdot 4 + 10 \cdot 6 + 14 \cdot 4 + 18 \cdot 2 + 22 \cdot 1) = \frac{1}{20} \cdot 204 = 10.2\ \text{(時間)}$$

階級値	度数
2	3
6	4
10	6
14	4
18	2
22	1
計	20

問5 教科書158 ページの問1で作成した度数分布表を用いて，B 班の読書時間の平均値を小数第 2 位を四捨五入して，小数第 1 位まで求めよ。

表より，B 班の読書時間の平均値は

$\frac{1}{15}(2 \cdot 4 + 6 \cdot 3 + 10 \cdot 1 + 14 \cdot 2 + 18 \cdot 2 + 22 \cdot 3)$

$= \frac{1}{15} \cdot 166$

$= 11.06\cdots$

$\fallingdotseq 11.1$ (時間)

階級値	度数
2	4
6	3
10	1
14	2
18	2
22	3
計	15

中央値

（教科書 p.161）

（3　中央値　）：データの値を小さい順に並べたとき，中央の順位にくる値。

（4　メジアン　）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 $n+1$ 番目のデータの値の平均値を中央値とする。

例3 教科書 158 ページの A 班の読書時間の中央値を求めてみよう。

20 個のデータの値を小さい方から順に並べると

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

このデータの中央値は，10 番目の値 9 と 11 番目の値 10 の平均値

$\frac{1}{2}(9 + 10) = 9.5$ (時間)

問6 教科書158 ページの B 班の読書時間の中央値を求めよ。

B 班の読書時間の 15 個のデータの値を小さい方から順に並べると

0　0　0　1　4　5　6　9　13　14　16　18　20　21　23

となる。このデータの中央値は

8 番目の値 9 時間である。

最頻値

（教科書 p.161）

（5　最頻値　）：度数分布表で，度数が最も多い階級の階級値。（　モード　）ともいう。

例4 教科書 158 ページの A 班の読書時間の最頻値を求めてみよう。

度数分布表で，読書時間の度数が最も大きい階級は

（　8　）時間以上（　12　）時間未満なので

最頻値は（　10　）時間である。

読書時間（時間）	度数
以上～未満 0 ～ 4	3
4 ～ 8	4
8 ～ 12	6
12 ～ 16	4
16 ～ 20	2
20 ～ 24	1
計	20

問7 教科書158 ページの問１で作成した度数分布表を用いて，B 班の読書時間の最頻値を求めよ。

問１の度数分布表で，度数が最も多い階級は 0 時間以上 4 時間未満であるから，B 班の読書時間の最頻値は，この階級の階級値 2 時間である。

3 箱ひげ図

四分位数

（教科書 p.162）

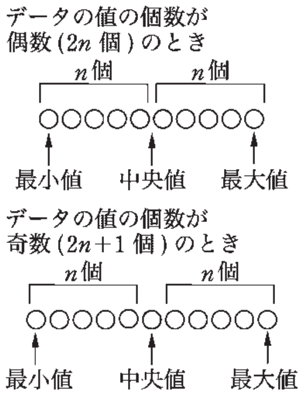
データの特徴をよりくわしく表すために，データの値を小さい方から順に並びかえて，次のような数値を考える。

① データの中央値を求める。

② 右の図のように，中央値を境にしてデータの値の個数が等しくなるように２つの部分に分ける。

③ ２つに分けたうち，最小値を含む方のデータの中央値を求める。

④ ２つに分けたうち，最大値を含む方のデータの中央値を求める。



①の値を（[1] 第２四分位数 ），③の値を（[2] 第１四分位数 ），

④の値を（[3] 第３四分位数 ），

これらを合わせて（[4] 四分位数 ）という。

３つの四分位数の値は，データの値の小さい方から 25%，50%，75% に対応する数値である。

例 5 教科書 158 ページの A 班の読書時間の四分位数を求めてみよう。

20 個のデータの値を小さい方から順に並べると，

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

第１四分位数は $Q_1 = \frac{1}{2}(5+6) = 5.5$（時間）

第２四分位数は $Q_2 = 9.5$（時間）

第３四分位数は $Q_3 = \frac{1}{2}(14+15) = 14.5$（時間）

問8 教科書158 ページの B 班の読書時間の四分位数を求めよ。

【B 班 15 人（単位　時間）】

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

B 班の読書時間の 15 個のデータの値を小さい方から順に並べると

0　0　0　1　4　5　6　9　13　14　16　18　20　21　23

となる。このデータの第２四分位数は９時間である。

第１四分位数は 0　0　0　1　4　5　6 の中央値より１時間である。

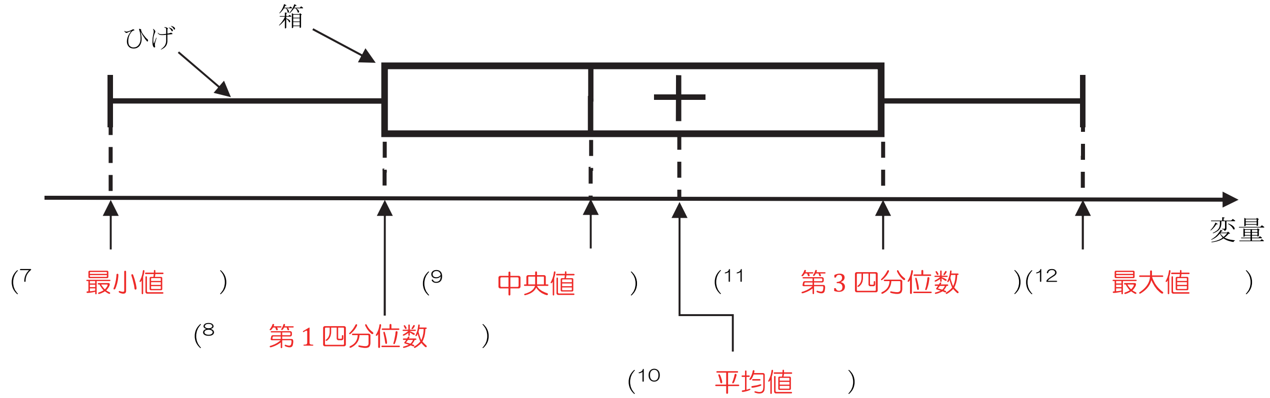
第３四分位数は 13　14　16　18　20　21　23 の中央値より 18 時間である。

箱ひげ図

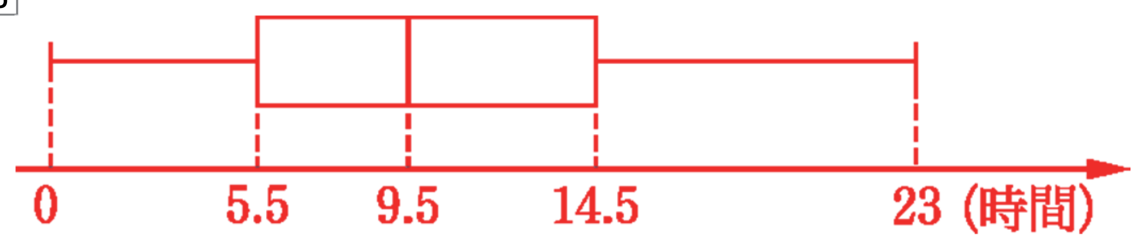
（教科書 p.163）

（[5] ５数要約 ）：データの分布を最小値，第１四分位数，中央値（第2 四分位数），第３四分位数，最大値の５つの数値を用いて要約する方法。

（[6] 箱ひげ図 ）：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。

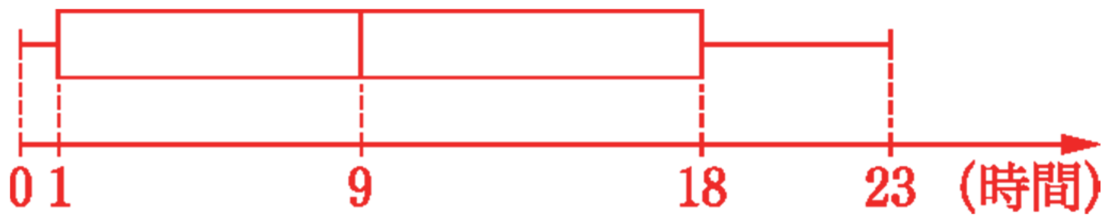


例 6



教科書 158 ページの A 班の読書時間の箱ひげ図をかいてみよう。

問9 教科書158 ページの B 班の読書時間の箱ひげ図をかけ。

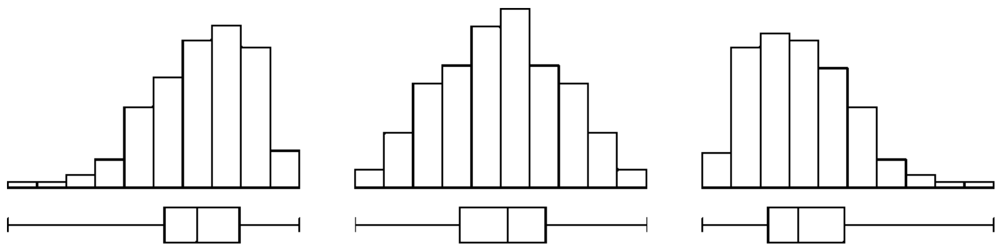


最小値０，第１四分位数１，第２四分位数９，第３四分位数 18，最大値 23 であるから，箱ひげ図は図のようになる。

箱ひげ図とヒストグラム

（教科書 p.163）

箱ひげ図はヒストグラムと同様に，データの分布を表現するのに適している。ヒストグラムでは，度数分布表のすべての階級の度数が必要であるのに対して，箱ひげ図は最小値，第 1 四分位数，中央値，第 3 四分位数，最大値の 5 つの数値がわかると，かくことができる。

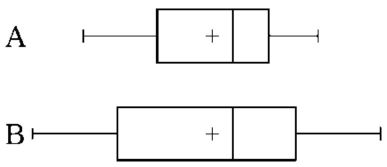


上の図のように，ヒストグラムの山の高い部分に箱ひげ図の箱が対応し，山のすその部分に箱ひげ図のひげが対応している。

4 箱ひげ図とデータの散らばり

（教科書 p.164）

右の図は，データの値の個数も平均値も中央値も等しい２つのデータ A，B の箱ひげ図である。



このように，平均値や中央値は等しくても，データの分布のようすが大きく異なることがある。箱ひげ図を用いることによって，データの分布のようすを視覚的に比較することを考える。

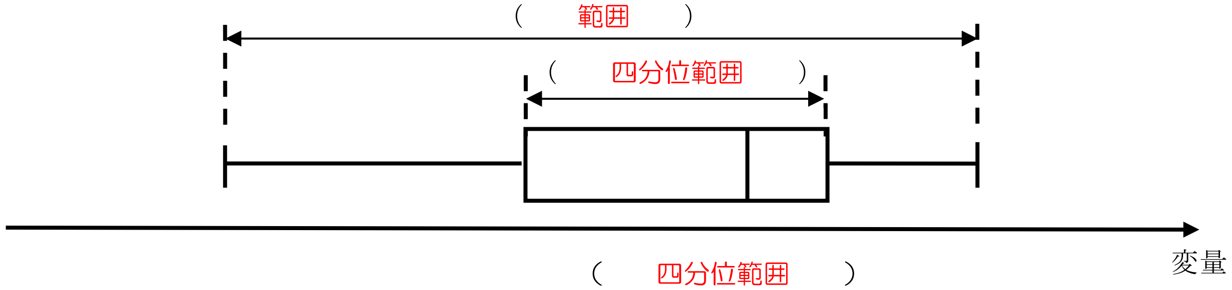
範囲と四分位数

（教科書 p.164）

（1 範囲 ）：データの最大値から最小値を引いた値。（2 レンジ ）ともいう。

（3 四分位範囲 ）：第３四分位数から第１四分位数を引いた値。

（4 四分位偏差 ）：四分位範囲を２で割った値。



$$（\ 四分位偏差\ ）=\frac{（\ 四分位範囲\ ）}{2}$$

例 7 教科書 158 ページの A 班の読書時間の範囲，四分位範囲，四分位偏差を求めよ。

範囲は　$23-0=23$（時間）

四分位範囲は　$14.5-5.5=9$（時間）

四分位偏差は　$\frac{9}{2}=4.5$（時間）

問 10 教科書158 ページの B 班の読書時間の範囲，四分位範囲，四分位偏差をそれぞれ求めよ。

範囲は　$23-0=23$（時間）

四分位範囲は　$18-1=17$（時間）

四分位偏差は　$\frac{17}{2}=8.5$（時間）

下の図１はある年の東京における各月の日ごとの平均気温の平均値を折れ線グラフで表したもの，図２は各月の日ごとの平均気温の分布を箱ひげ図で表したものである。

図１

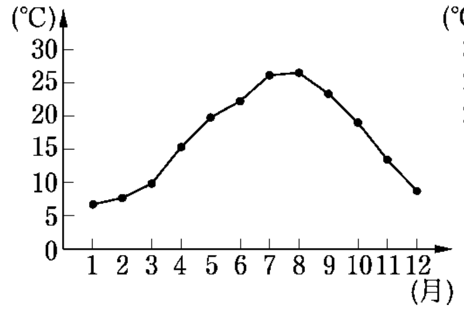
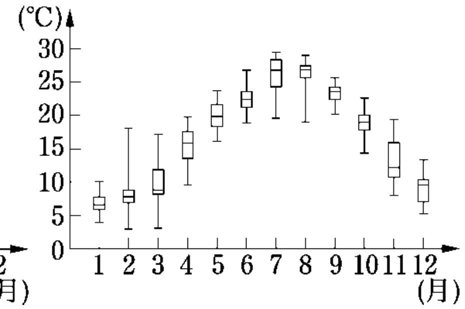
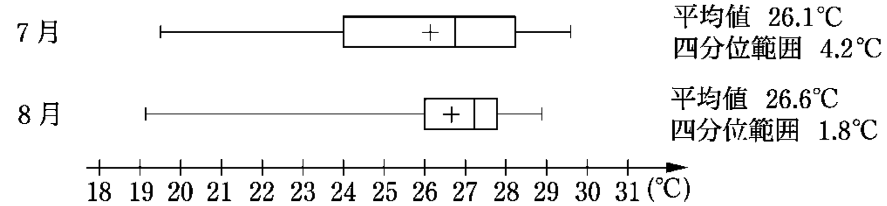


図２



平均値のみでは，各月の日ごとの平均気温の分布のようすを比較することができないが，箱ひげ図を用いると，分布のようすを比較することが可能である。

たとえば，７月と８月の箱ひげ図は次のようになる。



７月と８月の平均値や範囲はあまり変わらないが，８月の方が四分位範囲が小さい。

ゆえに，８月の方が，平均気温の散らばりが小さいことがわかる。

問 11 次の文章は上の図２の１月と２月の日ごとの平均気温の分布について述べたものである。この文章が適切であるかどうかを答えよ。

「１月と２月において，１日の平均気温が 5℃ 以上 10℃ 以下であった日はそれぞれ月の半分以上である。」

図２より，１月，２月ともに日ごとの平均気温の第１四分位数は 5℃ 以上であり，第３四分位数は 10℃ 以下である。このことから両月とも，月の半分以上は１日の平均気温が 5℃ 以上 10℃ 以下であることがわかる。

したがって，この文章は適切である。

5 分散と標準偏差

（教科書 p.166）

データの散らばり具合を数値で表すために，データの個々の値と平均値の差に着目する。
データの個々の値と平均値との差が大きいほど，その値は平均値から離れていることになる。
データの n 個の値を x_1，x_2，⋯，x_n とし，その平均値を $\overline{x}$ とする。

平均値からの（[1] 偏差 ）： 各値の平均値との差 $x_1-\overline{x}$，$x_2-\overline{x}$，⋯，$x_n-\overline{x}$

偏差の平均値を計算すると，

$$\frac{1}{n}\{(x_1-\overline{x})+(x_2-\overline{x})+\cdots+(x_n-\overline{x})\}=\frac{1}{n}\{(x_1+x_2+\cdots+x_n)-n\overline{x}\}$$

$$=\frac{1}{n}(x_1+x_2+\cdots+x_n)-\overline{x}=\overline{x}-\overline{x}=0$$

となり，偏差の平均値ではデータの散らばり具合を表すことはできない。
そこで，偏差２乗した値 $(x_i-\overline{x})^2$ を考える。

（[2] 分散 ）：偏差の２乗の平均値

分散は，計算の過程で数値を２乗するため単位も変わるから，分散の正の平方根を考える。

（[3] 標準偏差 ）：分散の正の平方根

分散と標準偏差

分散 $s^2=\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}$

標準偏差 $\boldsymbol{s}=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}}$

ただし，$\overline{x}$ は平均値

例 8 A 社と B 社の携帯音楽プレイヤーを充電してからの連続使用時間を５回調べた。

A 社　24　28　26　22　25　（単位　時間）

B 社　27　28　25　21　29　（単位　時間）

A 社の音楽プレーヤーの連続使用時間の標準偏差を求める。
連続使用時間を x として，平均値 $\overline{x}$ を求めると

$\overline{x}=\frac{1}{5}(24+28+26+22+25)=\frac{1}{5}\cdot125=25$（時間）

連続使用時間の各値の偏差は

連続使用時間	24	28	26	22	25
連続使用時間の偏差	−1	3	1	−3	0

分散 s^2 は

$s^2=\frac{1}{5}\{(-1)^2+3^2+1^2+(-3)^2+0^2\}=\frac{1}{5}\cdot20=4$

標準偏差 s は

$s=\sqrt{4}=2$（時間）

問 12 例8における B 社の音楽プレイヤーの連続使用時間の標準偏差を小数第３位を四捨五入して小数第２位まで求めよ。ただし，$\sqrt{2}=1.414$ とする。

B 社の音楽プレイヤーの連続使用時間を x，その平均値を $\overline{x}$ とする。

$\overline{x}=\frac{1}{5}(27+28+25+21+29)$

$=\frac{1}{5}\cdot130=26$（時間）

連続使用時間の各値の偏差は

連続使用時間 x	27	28	25	21	29
連続使用時間の偏差 $x-\overline{x}$	1	2	−1	−5	3

よって，分散 s^2 は

$s^2=\frac{1}{5}\{1^2+2^2+(-1)^2+(-5)^2+3^2\}$

$=\frac{1}{5}\cdot40=8$

したがって，標準偏差 s は

$s=\sqrt{8}=2\sqrt{2}$

$=2\cdot1.414=2.828\fallingdotseq2.83$（時間）

度数分布から標準偏差を求める方法を考える。
変量 x の平均値を $\overline{x}$ とすると，x の分散 s^2 は

階級値	度数
x_1	f_1
x_2	f_2
⋮	⋮
x_r	f_r
計	n

$$s^2=\frac{1}{n}\{(x_1-\overline{x})^2f_1+(x_2-\overline{x})^2f_2+\cdots+(x_r-\overline{x})^2f_r\}$$

よって，x の標準偏差 s は

$$s=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2f_1+(x_2-\overline{x})^2f_2+\cdots+(x_r-\overline{x})^2f_r\}}$$

例題 1 右の度数分布表は，ある高校のバスケットボール部の部員 10 人でフリースローを 5 回ずつ行ったゲームの結果である。このゲームの得点の標準偏差を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。

得点 x	度数 f
0	1
1	0
2	3
3	1
4	4
5	1
計	10

解

点数 x	度数 f	xf	$x-\overline{x}$	$(x-\overline{x})^2$	$(x-\overline{x})^2f$
0	1	0	-3	9	9
1	0	0	-2	4	0
2	3	6	-1	1	3
3	1	3	0	0	0
4	4	16	1	1	4
5	1	5	2	4	4
計	10	30			20

平均値 $\overline{x}$ は　$\overline{x}=\frac{1}{10}\cdot 30=3$（点）

分散 s^2 は　$s^2=\frac{1}{10}\cdot 20=2$

標準偏差 s は　$s=\sqrt{2}=1.414\cdots\fallingdotseq 1.41$（点）

問 13 下の表は，あるクラス 40 人に数学の小テストを行った結果である。このクラスの小テストの点数の標準偏差を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。ただし，$\sqrt{3}=1.732$ とする。

点数 x	0	1	2	3	4	5	6	7	8	9	10	計
人数	2	4	7	8	4	6	2	1	2	3	1	40

得点 x	度数 f	xf	$x-\overline{x}$	$(x-\overline{x})^2$	$(x-\overline{x})^2f$
0	2	0	-4	16	32
1	4	4	-3	9	36
2	7	14	-2	4	28
3	8	24	-1	1	8
4	4	16	0	0	0
5	6	30	1	1	6
6	2	12	2	4	8
7	1	7	3	9	9
8	2	16	4	16	32
9	3	27	5	25	75
10	1	10	6	36	36
計	40	160			270

xf の値は上の表のようになるから，テストの点数の平均値 $\overline{x}$ は

$$\overline{x}=\frac{1}{40}\cdot 160=4\text{（点）}$$

さらに，$x-\overline{x}$，$(x-\overline{x})^2$，$(x-\overline{x})^2f$ の値は上の表のようになるから，点数の分散 s^2 は

$$s^2=\frac{1}{40}\cdot 270=\frac{27}{4}$$

よって，標準偏差 s は

$$s=\sqrt{\frac{27}{4}}=\frac{3\sqrt{2}}{2}$$

$$=\frac{3\cdot 1.732}{2}$$

$$=2.598\fallingdotseq 2.60\text{（点）}$$

分散と平均値の関係式

（教科書 p.169）

分散 s^2 を表す式は，次のように変形できる。

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

$$=\frac{1}{n}\{(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2\overline{x}(x_1+x_2+\cdots+x_n)+n(\overline{x})^2\}$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2\overline{x}\cdot\frac{1}{n}(x_1+x_2+\cdots+x_n)+(\overline{x})^2$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2(\overline{x})^2+(\overline{x})^2$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-(\overline{x})^2$$

一般に，変量 x の分散について，次のことが成り立つ。

（4　x の分散　）＝（5　x^2 の平均値　）－（6　（x の平均値）^{2}　）

例 9 下の表は，高校 1 年生の A さんが毎朝通学に利用しているバスの乗車時間を 6 日間調べた結果である。

バスの乗車時間 x（分）	9	13	10	9	10	12

x の平均値 $\overline{x}$ と x^2 の平均値 $\overline{x^2}$ は，次のようになる。

$$\overline{x}=\frac{1}{6}(9+13+10+9+10+12)=\frac{1}{6}\cdot 63=\frac{21}{2}$$

$$\overline{x^2}=\frac{1}{6}(9^2+13^2+10^2+9^2+10^2+12^2)=\frac{1}{6}\cdot 675=\frac{225}{2}$$

よって分散 s^2 は

$$s^2=\frac{225}{2}-\left(\frac{21}{2}\right)^2=\frac{9}{4}=2.25$$

問 14 下の表は，A さんの 1 日の学習時間を 5 日間記録したものである。

1 日の学習時間 x（時間）	1	2	3	1	3

このとき，1 日の学習時間の分散を求めよ。

1 日の学習時間 x の平均値 $\overline{x}$ は

$$\overline{x}=\frac{1}{5}(1+2+3+1+3)=\frac{1}{5}\cdot 10$$

$$=2$$

x^2 の平均値 $\overline{x^2}$ は

$$\overline{x^2}=\frac{1}{5}(1^2+2^2+3^2+1^2+3^2)=\frac{1}{5}\cdot 24=\frac{24}{5}$$

x の分散 s^2 は

$$s^2=\overline{x^2}-(\overline{x})^2=\frac{24}{5}-2^2=\frac{4}{5}=0.8$$

問 題

（教科書 p.170）

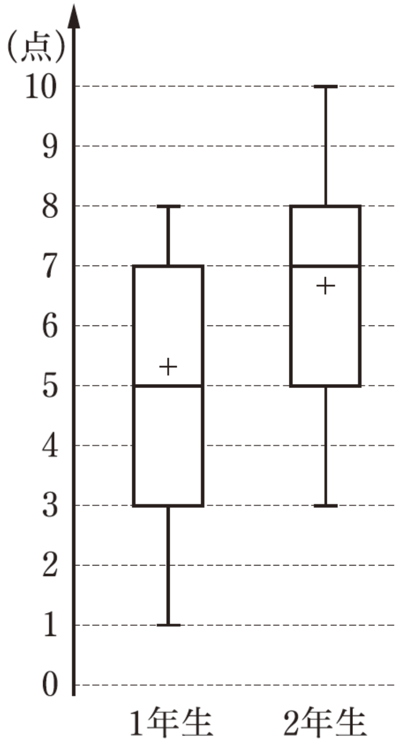
1 右の図は，ある学校の 1 年生と 2 年生それぞれ 20 人に対して，10 点満点の漢字テストを実施したときの得点の箱ひげ図である。このとき，次の①～④のうち，右の図から必ずしも正しいとは限らないものをすべて選べ。

① 1 年生の得点のデータの方が 2 年生の得点のデータより四分位範囲が大きい。

② どちらの学年にも，平均点以下の生徒が 10 人以上いる。

③ どちらの学年にも，得点が 5 点以上の生徒が 10 人以上いる。

④ どちらの学年にも，得点がちょうど 8 点の生徒がいる。



1 年生の四分位範囲は $7-3=4$（点）

2 年生の四分位範囲は $8-5=3$（点）

よって，①は正しい。

②は 2 年生において，平均点以下の生徒が 10 人以上いるとは限らない。

（例）2 年生 20 人の得点が次のような場合，

3，4，4，4，5，5，6，6，6，7，7，7，7，7，7，9，10，10，10，10

平均点は 6.7 となり，平均点以下の生徒は 9 人となる。

1 年生，2 年生ともに中央値は 5 点以上である。よって，③は正しい。

④は 2 年生において，得点がちょうど 8 点の生徒がいるとは限らない。

（例） ②の（例）では 8 点の生徒はいない。

2 くじを 20 回引いて，当たった回数だけ得点できるゲームがある。右の表は，ある学校の生徒 5 人がこのゲームを行ったときの得点を記録したものである。ただし，生徒 4 の得点は 5 人の得点の平均値以下であった。このとき，次の問に答えよ。

	得点
生徒 1	8
生徒 2	14
生徒 3	10
生徒 4	a
生徒 5	$18-a$
平均値	m
分散	6

(1) 5 人の得点の平均値 m を求めよ。

5 人の得点の平均値 m は

$$m=\frac{1}{5}\{8+14+10+a+(18-a)\}=\frac{1}{5}\cdot 50=10$$

(2) 表中の a の値を求めよ。

各生徒の得点の偏差は，表のようになる。

-2	4	0	$a-10$	$8-a$

…表(*)

よって

$$\frac{1}{5}\{(-2)^2+4^2+0^2+(a-10)^2+(8-a)^2\}=6$$

が成り立つ。

この式を計算し，整理すると

$$a^2-18a+77=0$$

$$(a-7)(a-11)=0$$

条件より $a \leqq 10$ であるから $a=7$

(3) 右の表の得点に対して，全員に 10 点を加えたとき，分散の値はどのように変化するか。①～③のうちから適するものを選べ。

① 大きくなる ② 変わらない ③ 小さくなる

全員に 10 点加えた得点を x として，その平均値を $\overline{x}$ とする。

$$\overline{x}=\frac{1}{5}(5m+50)=m+10=20$$

10 点加えた得点 x	18	24	20	$a+10$	$28-a$
$x-\overline{x}$	-2	4	0	$a-10$	$8-a$

10 点加えた得点の各値の偏差は表(*)と一致する。よって，全員に 10 点加えた得点の分散と加える前の分散は変わらない。

したがって，正解は②