1節　データの整理と分析

❶ データの整理

（教科書 p.162）

次の資料は，ある年の９の大阪の最高気温を日付順に横に並べたものである。

［ データ１　９月の大阪の最高気温(単位 ℃)］

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

（教科書 p.162）

データについて調べるとき，ただ個々の値を並べただけでは，全体の特徴がつかみにくい。このようなときには，表やグラフなどを用いてデータの特徴を把握するとよい。

（3　　　　　）：データを整理したときの区間

（4　　　　　）：各階級に入っているデータの値の個数

（5　　　　　　）：各階級の真ん中の値

（6　　　　　　　）：区間の幅

（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

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

（　　　　　　　）

階級の真ん中の値 31 は（　　　　　　）

差 2 は（　　　　　　）

度数分布のグラフ

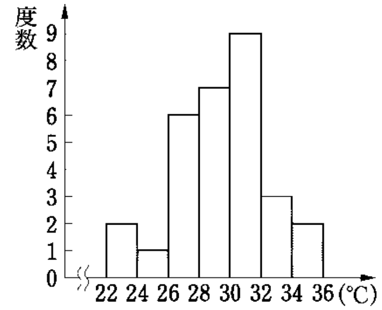
（教科書 p.163）

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。

データ１の度数分布のヒストグラムをかくと，右のようになる。

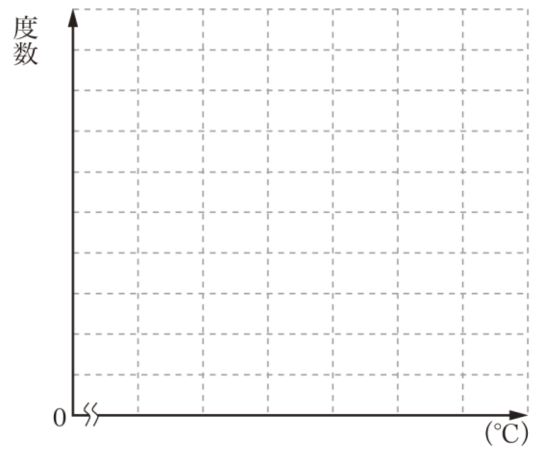


問1 次のデータは，ある年の９月の東京の最高気温を気温の低い順に横に並べたものである。22℃ 以上 24℃ 未満という階級から，幅を一定にとって順次階級をつくり，このデータの度数分布表を作成せよ。また，そのヒストグラムをかけ。

【データ２　9月の東京の最高気温（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

気温の階級 ℃以上　℃未満	度数
22～24	
24～26	
26～28	
28～30	
30～32	
計	



相対度数

（教科書 p.163）

（[10]　　　　　　）：各階級の度数を度数の合計で割った値

（[11]　　　　　　）＝ （[12]　　　　　　） / （[13]　　　　　　）

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

右の表で，たとえば，26℃以上 28℃未満の階級の相対度数は

になる。

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。

気温の階級 ℃以上　℃未満	度数	相対度数
22～24	6	
24～26	6	
26～28	8	
28～30	9	
30～32	1	
計	30	

❷ データの代表値

（教科書 p.164）

（[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.164）

変量 x の n 個の値 x_1，x_2，・・・，x_n からなるデータがあるとき，

（[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。

$$\overline{x} = \frac{x_1+x_2+\cdots+x_n}{n}$$

$$平均値 = \frac{データの値の総和}{データの値の個数}$$

例 1 教科書 162 ページのデータ 1 の平均値は

度数分布から求めた平均値

（教科書 p.164）

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 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

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。

階級 ℃以上　℃未満	階級値 x	度数 f	xf
22～24		6	
24～26		6	
26～28		8	
28～30		9	
30～32		1	
計		30	

中央値，最頻値

（教科書 p.165）

（3　　　　　　　）：データのすべての値を小さい順に並べたとき，中央の位置にある数値。

（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 $n+1$ 番目の数値の平均値を中央値とする。

例3 教科書 162 ページのデータ 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

中央値は

（5　　　　　　　）：度数分布表で，度数がもっとも多い階級の階級値。

（6　　　　　　　）ともいう。

例4 教科書 162 ページのデータ 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

「（　　　　　）以上（　　　　　）未満」なので

最頻値は（　　　　　）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　9月の東京の最高気温（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

❸ データの散らばり

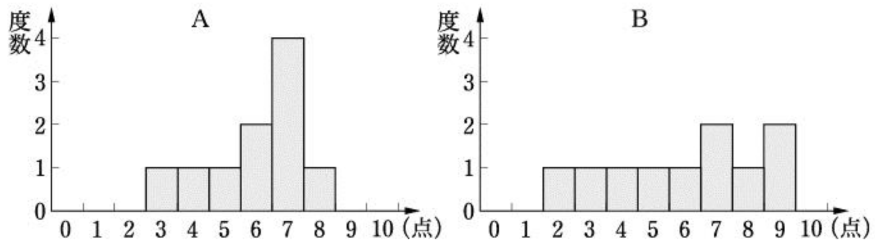
（教科書 p.166）

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

このとき，A, B ともに平均値は 6 点，中央値は 6.5 点である。

しかし，下のヒストグラムから，2 人の得点の分布には散らばり具合に違いが見られる。ここでは，データの分布の特徴について，散らばり具合を含めて考えてみよう。



四分位数

（教科書 p.166）

中央値をもとに，データの分布を表すための数値を考えてみよう。

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。

3　4　5　6　6　7　7　7　7　8

最小値　（1　　　　）　中央値 6.5　（2　　　　）　最大値

（3　　　　）

中央値を境にして 2 つの部分に分けたとき，

最小値を含む方の 5 個のデータの中央値 5 点を（4　　　　）という。

中央値 6.5 点を（5　　　　）という。

最大値を含む方の 5 個のデータの中央値 7 点を（6　　　　）という。

これらを合わせて（7　　　　）という。

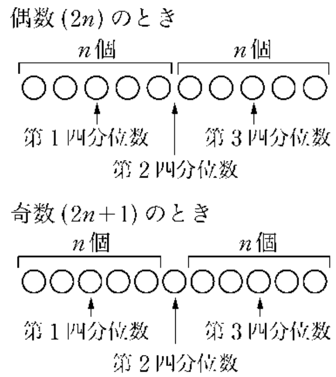
一般に，四分位数は，データの値の個数が偶数 $(2n)$，奇数 $(2n+1)$ のいずれのときも，次のようになる。

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。



例5 的当てゲームにおける B の得点の四分位数を求めてみよう。B の得点を小さい順に並べかえると

2　3　4　5　6　7　7　8　9　9

第 1 四分位数は（　　　　）点

第 2 四分位数は（　　　　）点

第 3 四分位数は（　　　　）点

問5 次のデータはあるクラスの 1 班，2 班の 1 か月の読書時間である。1 班，2 班の四分位数をそれぞれ求めよ。

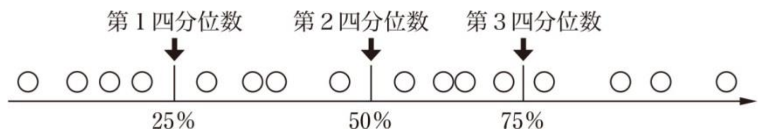
【1 班 20 人（単位　時間）】

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

【2 班 15 人（単位　時間）】

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

第 1 四分位数は，データの値を小さい順に並べたとき，値の小さい方から 25% の位置を示す値である。同様に，第 2 四分位数，第 3 四分位数は，値の小さい方から 50%，75% の位置を示す値である。

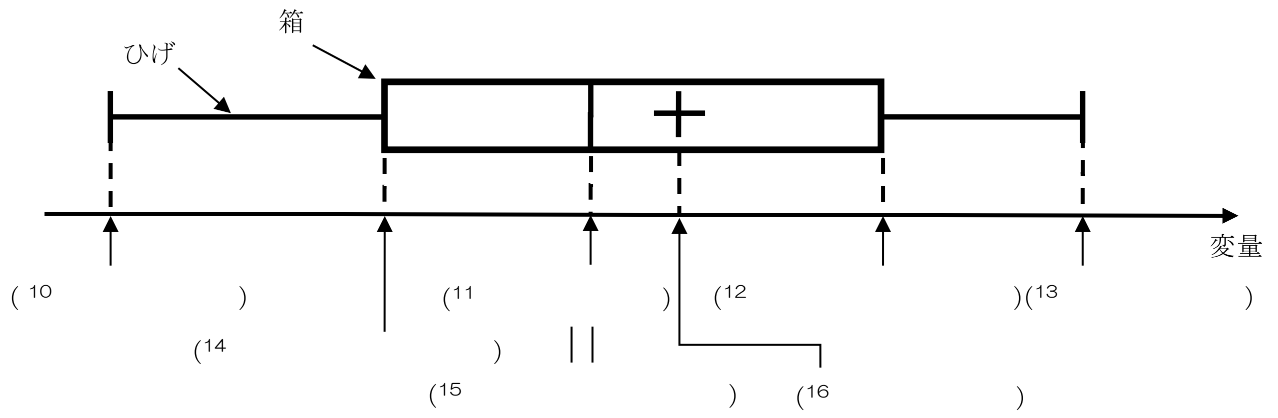


箱ひげ図

（教科書 p.168）

（[8]　　　　　）：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

（[9]　　　　　）：5 数要約を用いてデータの分布を表す図

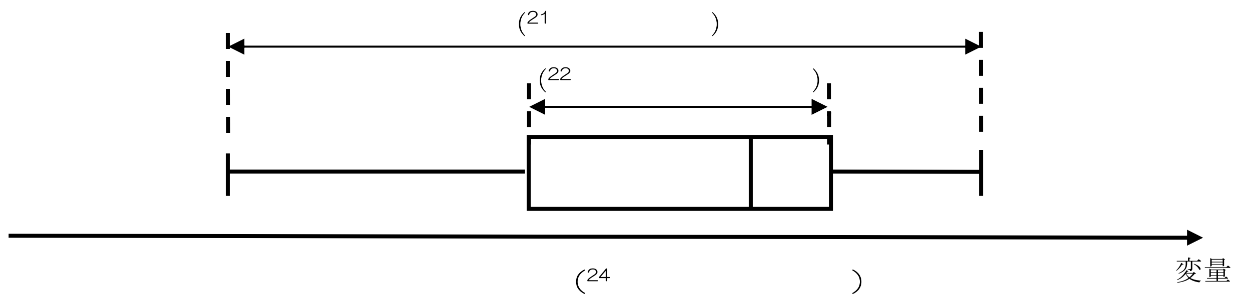


（[17]　　　　　）：データの最大値から最小値を引いた値。（[18]　　　　　）ともいう。

（[19]　　　　　）：第 3 四分位数から第 1 四分位数を引いた値。

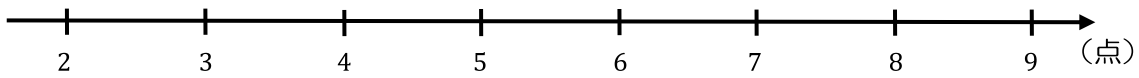
（[20]　　　　　）：四分位範囲を 2 で割った値。

箱ひげ図では，ひげを含めた全長が範囲を表し，箱の横の長さが四分位範囲を表す。



（[23]　　　　　）＝ $\dfrac{（[24]\qquad）}{2}$

例6 教科書 166 ページの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（　　　　　　　　　）(点)

四分位範囲は（　　　　　　　　　）(点)

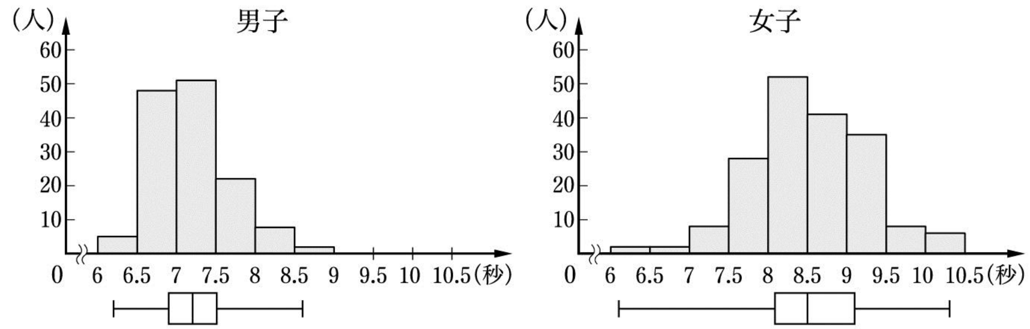
四分位偏差は（　　　　　　　）(点)

問6 教科書166ページの的当てゲームにおける B の得点の箱ひげ図をかけ。また，B の得点の範囲，四分位範囲，四分位偏差を求めよ。

ヒストグラムと箱ひげ図

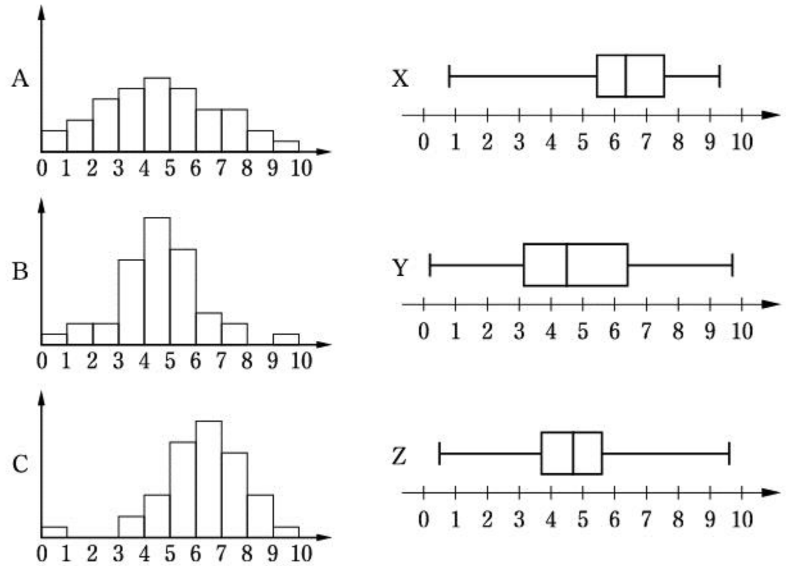
（教科書 p.169）

下のヒストグラムと箱ひげ図は，ある高校の 1 年生の体力テストにおける 50m 走の結果を，男子と女子に分けて表したものである。



上の図では，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。一般に，分布が 1 つの山の形をしたヒストグラムとなる場合，箱ひげ図をみることによってヒストグラムのおおよその形を知ることができる。

問7 次の A，B，C のヒストグラムについて，それぞれに対応する箱ひげ図として適切なものを X，Y，Z の中から選べ。



箱ひげ図とデータの散らばり

（教科書 p.170）

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で表し，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

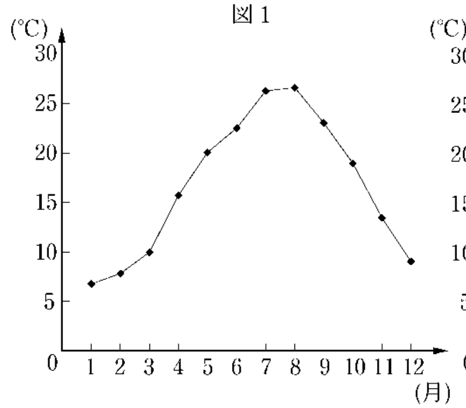


図２

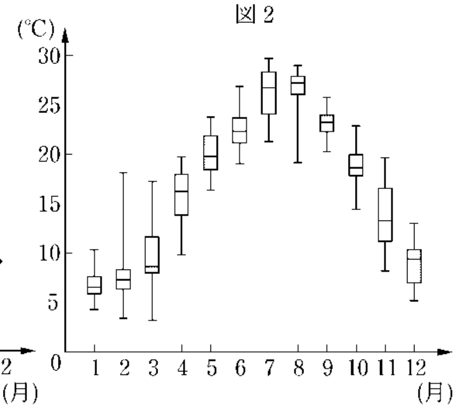
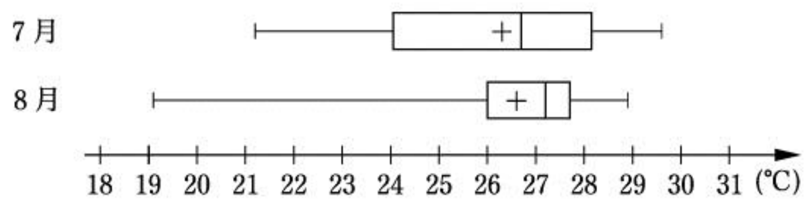


図１のように，平均値のみでは各月の日ごとの平均気温の分布のようすを比較することはできないが，図２のように箱ひげ図を用いることで，分布のようすを，散らばり具合いを含めて比較することができる。

たとえば，７月と８月の箱ひげ図は下の図のようになる。



この箱ひげ図より，７月と８月の平均気温の平均値はあまり変わらないが，箱の横の長さやひげの長さを比較することにより，データの散らばり具合の違いがわかる。

このように箱ひげ図は複数のデータの分布を比較するのに適している。

問8 167 ページの問５の１班，２班について，それぞれ箱ひげ図をかき，分布を比較せよ。

分散と標準偏差

（教科書 p.171）

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，…，x_n であり，その平均値を $\overline{x}$ とするとき

（[25]　　　　　　　　　　　　）あるいは単に（[26]　　　　　　）

：個々の値と平均値との差　$x_1-\overline{x}$，$x_2-\overline{x}$，…，$x_n-\overline{x}$

偏差の総和を計算すると

$(x_1-\overline{x})+(x_2-\overline{x})+\cdots+(x_n-\overline{x})$

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

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

偏差の平均値ではデータの散らばり具合いを表すことができない。

偏差を２乗した値

$(x_1-\overline{x})^2$，$(x_1-\overline{x})^2$，…，$(x_1-\overline{x})^2$

を考えると，これらの値はすべて０以上で，データの値が平均値 $\overline{x}$ から離れているほど大きくなる。

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

例7 教科書 166 ページの的当てゲームにおける A の得点の分散 ${s_A}^2$，標準偏差 s_A を求めてみよう。

A の得点の平均値は 6 点であるから，偏差は次の表のようになる。

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
偏差	1	−3	0	−2	2	1	1	−1	1	0

よって

166 ページの的当てゲームにおける B の得点の分散 ${s_B}^2$，標準偏差 s_B を，例 7 と同様に求めると，

となる。したがって，分散，標準偏差の値は（　　　　　）の方が大きいので，得点の散らばり具合は（　　　　　）の方が大きいことがわかる。

問9 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。

回	1	2	3	4	5
得点	82	76	63	80	64

Training

（教科書 p.173）

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，次の問に答えよ。

(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	

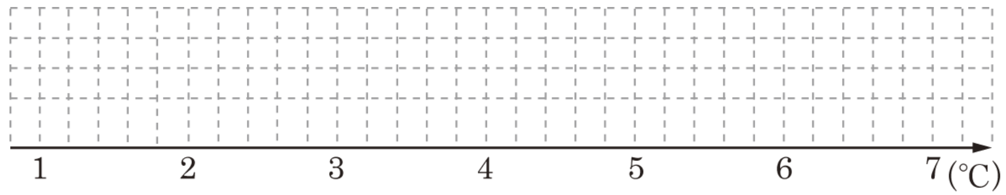
(2)　テストの点数の平均値を度数分布表から求めよ。

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

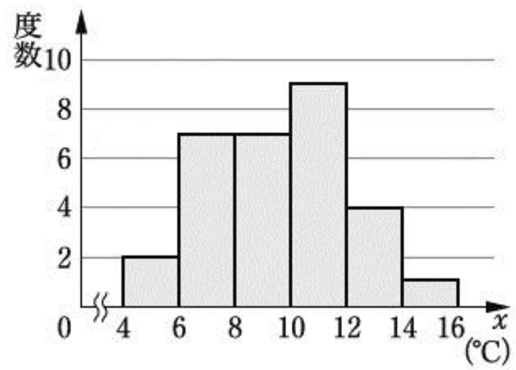
2　ある都市における 1 日の平均気温を調べた。次の問に答えよ。

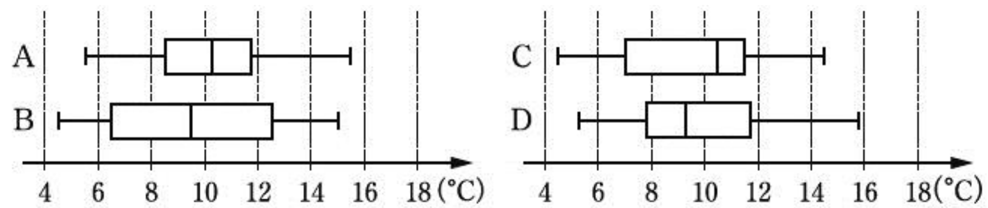
(1)　次のデータは 3 月 1 日から 10 日までの 1 日の平均気温を順に並べたものである。このデータの箱ひげ図をかけ。

2.6　1.8　1.3　2.5　5.8　5.8　5.3　6.4　6.2　6.9



(2)　右の図は 11 月の 30 日間における 1 日の平均気温のデータのヒストグラムである。このデータを箱ひげ図にまとめると，次の A ～ D のいずれかになった。このデータの箱ひげ図は A ～ D のどれか。





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

参考 分散の計算

（教科書 p.174）

5 個の値 x_1，x_2，x_3，x_4，x_5 からなるデータがある。このデータの平均値を $\overline{x}$ とすると

$$\frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5) = \overline{x}$$

が成り立つ。

データの分散を s^2 とすると

$$\begin{aligned} s^2 &= \frac{1}{5}\{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 + (x_4 - \overline{x})^2 + (x_5 - \overline{x})^2\} \\ &= \frac{1}{5}\{({x_1}^2 + {x_2}^2 + {x_3}^2 + {x_4}^2 + {x_5}^2) - 2\overline{x}(x_1 + x_2 + x_3 + x_4 + x_5) + 5(\overline{x})^2\} \\ &= \frac{1}{5}({x_1}^2 + {x_2}^2 + {x_3}^2 + {x_4}^2 + {x_5}^2) - 2\overline{x} \cdot \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5) + (\overline{x})^2 \\ &= \frac{1}{5}({x_1}^2 + {x_2}^2 + {x_3}^2 + {x_4}^2 + {x_5}^2) - 2(\overline{x})^2 + (\overline{x})^2 \\ &= \frac{1}{5}({x_1}^2 + {x_2}^2 + {x_3}^2 + {x_4}^2 + {x_5}^2) - (\overline{x})^2 \end{aligned}$$

となり，データの値を 2 乗した値の平均値から，データの平均値の 2 乗を引いた値が分散に等しいことがわかる。

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

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

例1 上の分散の計算式を用いて，教科書 166 ページの的当てゲームにおける A の得点の分散を求めてみよう。

問1 上の計算式を用いて，教科書166ページの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

１節　データの整理と分析

❶ データの整理

（教科書 p.162）

次の資料は，ある年の９の大阪の最高気温を日付順に横に並べたものである。

［ データ１　９月の大阪の最高気温(単位 ℃)］

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]　データ　）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]　変量　）という。

度数分布表

（教科書 p.162）

データについて調べるとき，ただ個々の値を並べただけでは，全体の特徴がつかみにくい。このようなときには，表やグラフなどを用いてデータの特徴を把握するとよい。

（[3]　階級　）：データを整理したときの区間
（[4]　度数　）：各階級に入っているデータの値の個数
（[5]　階級値　）：各階級の真ん中の値
（[6]　階級の幅　）：区間の幅
（[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

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

階級の真ん中の値 31 は（　階級値　）

差２は（　階級の幅　）

（　度数分布表　）

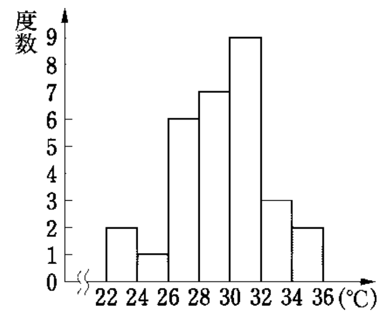
度数分布のグラフ

（教科書 p.163）

度数分布を表す右のようなグラフを（[9]　ヒストグラム　）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。

データ１の度数分布のヒストグラムをかくと，右のようになる。

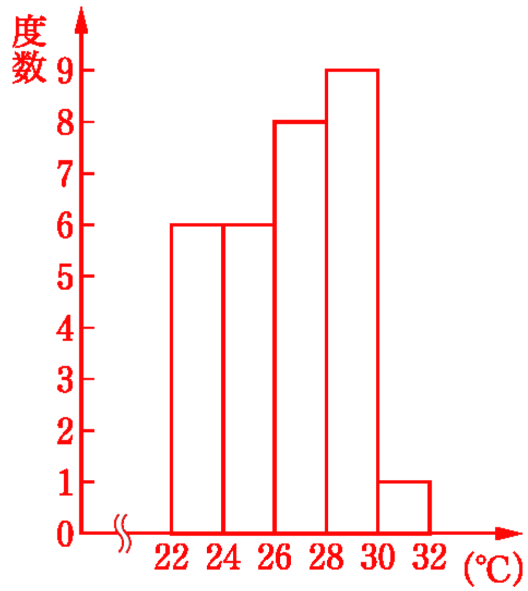


問1 次のデータは，ある年の９月の東京の最高気温を気温の低い順に横に並べたものである。22℃以上24℃未満という階級から，幅を一定にとって順次階級をつくり，このデータの度数分布表を作成せよ。また，そのヒストグラムをかけ。

【データ２　9月の東京の最高気温（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

気温の階級 ℃以上　℃未満	度数
22～24	6
24～26	6
26～28	8
28～30	9
30～32	1
計	30



相対度数

（教科書 p.163）

（[10]　相対度数　）：各階級の度数を度数の合計で割った値

（[11]　相対度数　）＝ $\dfrac{（^{12}\ 度数\ ）}{（^{13}\ 度数の合計\ ）}$

右の表で，たとえば，26℃以上 28℃未満の階級の相対度数は

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

になる。

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。

相対度数を度数分布表に並べて記入すると，表のようになる。

気温の階級 ℃以上　℃未満	度数	相対度数
22～24	6	0.20
24～26	6	0.20
26～28	8	0.27
28～30	9	0.30
30～32	1	0.03
計	30	1.00

❷ データの代表値

（教科書 p.164）

（[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.164）

変量 x の n 個の値 x_1, x_2, ・・・, x_n からなるデータがあるとき，

（[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。

$$\overline{x} = \frac{x_1+x_2+\cdots+x_n}{n}$$

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

例 1 教科書 162 ページのデータ 1 の平均値は

$$\frac{31.5+33.4+\cdots+22.0}{30} = \frac{881.8}{30} \fallingdotseq 29.4\ (℃)$$

度数分布から求めた平均値

（教科書 p.164）

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 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

$$\frac{884}{30} \fallingdotseq 29.5\ (℃)$$

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。

度数分布表に階級値 x と度数 f の積 xf の欄をつけ加えると，表のようになる。

階級 ℃以上　℃未満	階級値 x	度数 f	xf
22～24	23	6	138
24～26	25	6	150
26～28	27	8	216
28～30	29	9	261
30～32	31	1	31
計		30	796

表より，平均値は

$$\frac{796}{30} = 26.53\cdots \fallingdotseq 26.5\ (℃)$$

中央値，最頻値

（教科書 p.165）

（3　中央値　）：データのすべての値を小さい順に並べたとき，中央の位置にある数値。

（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 $n+1$ 番目の数値の平均値を中央値とする。

例3 教科書 162 ページのデータ 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

中央値は　$\frac{29.1+29.5}{2} = 29.3\ (℃)$

（5　最頻値　）：度数分布表で，度数がもっとも多い階級の階級値。

（6　モード　）ともいう。

例4 教科書 162 ページのデータ 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

「（　30℃　）以上（　32℃　）未満」なので

最頻値は（　31℃　）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　9月の東京の最高気温（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

$\frac{26.5+26.5}{2} = 26.5\ (℃)$

度数がもっとも多い階級は，「28℃以上 30℃未満」なので，最頻値はその階級値 29℃

❸ データの散らばり

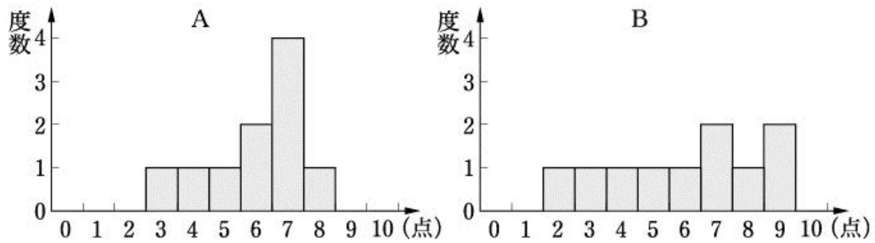
（教科書 p.166）

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

このとき，A，B ともに平均値は 6 点，中央値は 6.5 点である。

しかし，下のヒストグラムから，2 人の得点の分布には散らばり具合に違いが見られる。ここでは，データの分布の特徴について，散らばり具合を含めて考えてみよう。



四分位数

（教科書 p.166）

中央値をもとに，データの分布を表すための数値を考えてみよう。

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。

3　4　5　6　6　7　7　7　7　8

最小値（1 第 1 四分位数 ）　中央値　6.5　（2 第 3 四分位数 ）最大値

（3 第 2 四分位数 ）

中央値を境にして 2 つの部分に分けたとき，

最小値を含む方の 5 個のデータの中央値 5 点を（4 第 1 四分位数 ）という。

中央値 6.5 点を（5 第 2 四分位数 ）という。

最大値を含む方の 5 個のデータの中央値 7 点を（6 第 3 四分位数 ）という。

これらを合わせて（7 四分位数 ）という。

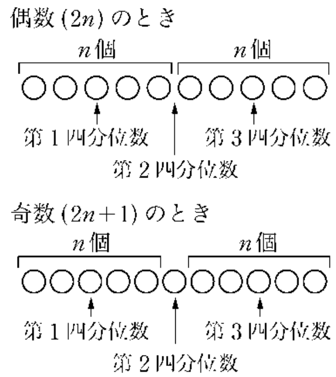
一般に，四分位数は，データの値の個数が偶数 $(2n)$，奇数 $(2n+1)$ のいずれのときも，次のようになる。

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。



例5 的当てゲームにおける B の得点の四分位数を求めてみよう。B の得点を小さい順に並べかえると

2　3　4　5　6　7　7　8　9　9

第 1 四分位数は（ 4 ）点

第 2 四分位数は（ $\frac{6+7}{2}=6.5$ ）点

第 3 四分位数は（ 8 ）点

問5 次のデータはあるクラスの１班，２班の１か月の読書時間である。１班，２班の四分位数をそれぞれ求めよ。

【1 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

【2 班 15 人（単位　時間)】

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

1 班のデータの値を小さい順に並べかえると

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

0　0　3　5　5　6　7　8　9　9

の中央値より

$\frac{5+6}{2} = 5.5$ (時間)

第 2 四分位数は，データの中央値より

$\frac{9+10}{2} = 9.5$ (時間)

第 3 四分位数は

10　10　11　13　14　15　15　18　19　23

の中央値より

$\frac{14+15}{2} = 14.5$ (時間)

2 班のデータの値を小さい順に並べかえると

0　0　0　1　4　5　6　9　13　14　16　18　20　21　23

第 1 四分位数は

0　0　0　1　4　5　6

の中央値より 1 時間

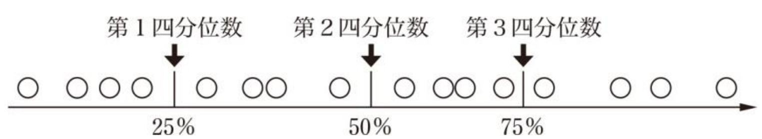
第 2 四分位数は，データの中央値より 9 時間

第 3 四分位数は

13　14　16　18　20　21　23

の中央値より 18 時間

第 1 四分位数は，データの値を小さい順に並べたとき，値の小さい方から 25% の位置を示す値である。同様に，第 2 四分位数，第 3 四分位数は，値の小さい方から 50%，75% の位置を示す値である。

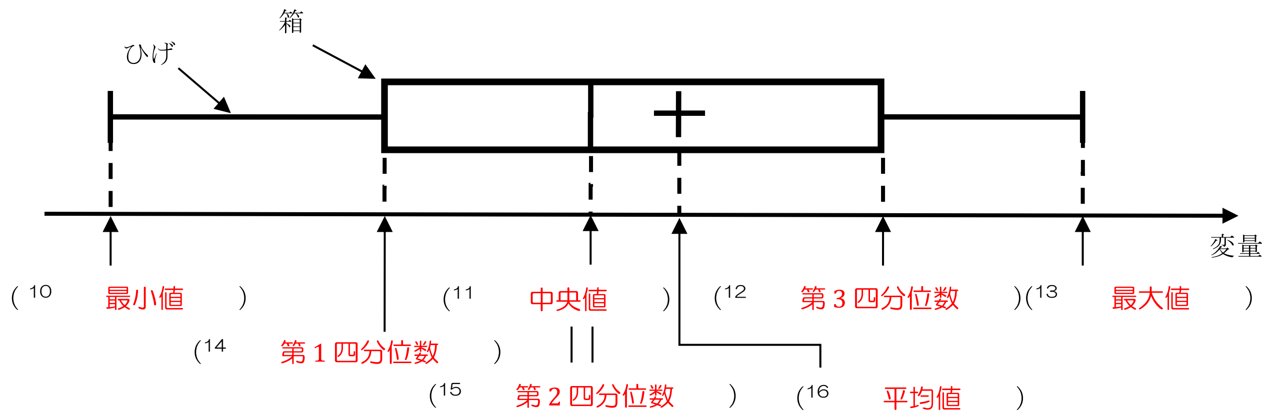


箱ひげ図

（教科書 p.168）

(8　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

(9　箱ひげ図　)：5 数要約を用いてデータの分布を表す図

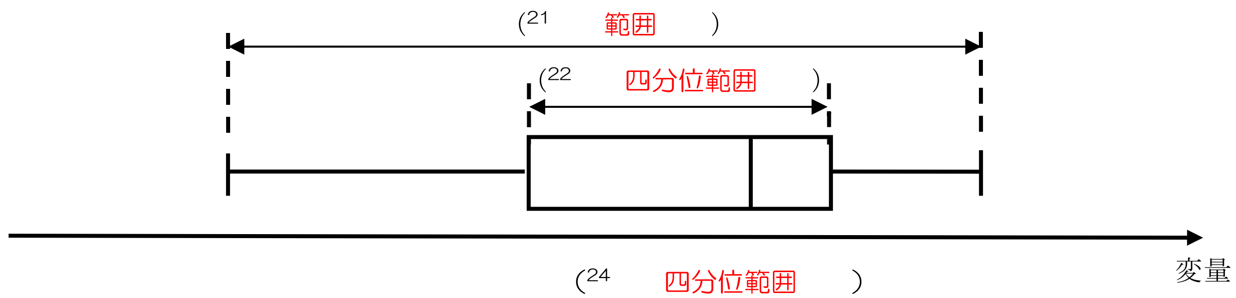


(17　範囲　)：データの最大値から最小値を引いた値。(18　レンジ　) ともいう。

(19　四分位範囲　)：第 3 四分位数から第 1 四分位数を引いた値。

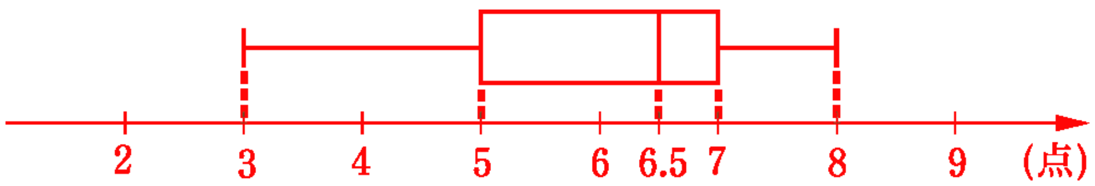
(20　四分位偏差　)：四分位範囲を 2 で割った値。

箱ひげ図では，ひげを含めた全長が範囲を表し，箱の横の長さが四分位範囲を表す。



(23　四分位偏差　) = $\dfrac{(24\ \text{四分位範囲}\ )}{2}$

例6 教科書 166 ページの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,

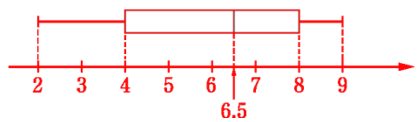
範囲は（　8 − 3 = 5　）(点)

四分位範囲は（　7 − 5 = 2　）(点)

四分位偏差は（　$\frac{2}{2} = 1$　）(点)

問6 教科書166ページの的当てゲームにおける B の得点の箱ひげ図をかけ。また，B の得点の範囲，四分位範囲，四分位偏差を求めよ。

B の得点の箱ひげ図は次のとおりである。



範囲は　9 − 2 = 7 (点)

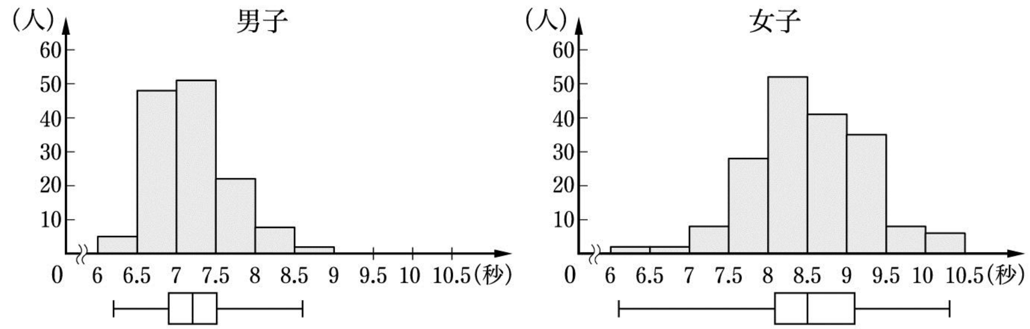
四分位範囲は　8 − 4 = 4 (点)

四分位偏差は　$\frac{4}{2} = 2$ (点)

ヒストグラムと箱ひげ図

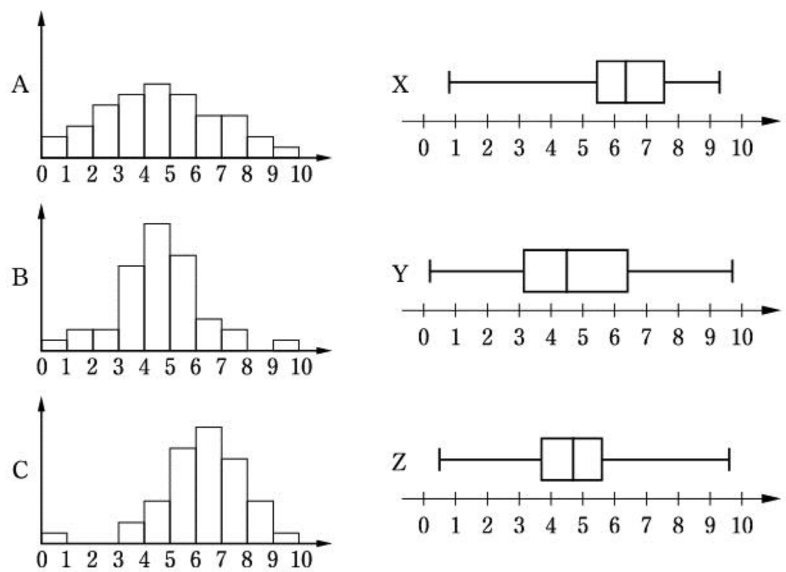
（教科書 p.169）

下のヒストグラムと箱ひげ図は，ある高校の 1 年生の体力テストにおける 50m 走の結果を，男子と女子に分けて表したものである。



上の図では，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。一般に，分布が 1 つの山の形をしたヒストグラムとなる場合，箱ひげ図をみることによってヒストグラムのおおよその形を知ることができる。

問7 次の A，B，C のヒストグラムについて，それぞれに対応する箱ひげ図として適切なものを X， Y，Z の中から選べ。



ヒストグラムの山の高さと位置，および箱ひげ図の箱の横の長さを比較すると

A に対応する箱ひげ図は Y

B に対応する箱ひげ図は Z

C に対応する箱ひげ図は X

箱ひげ図とデータの散らばり

（教科書 p.170）

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で表し，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

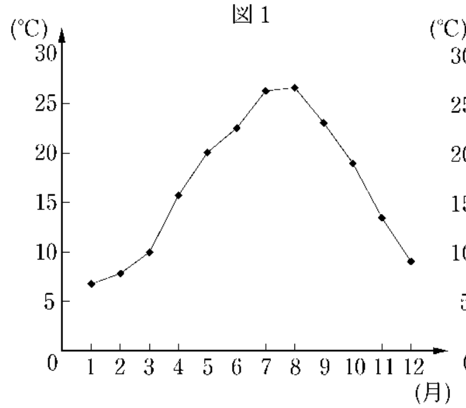


図 2

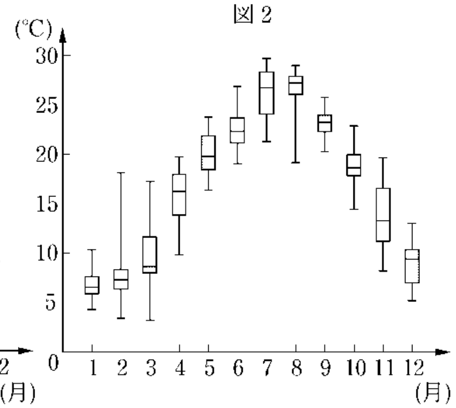
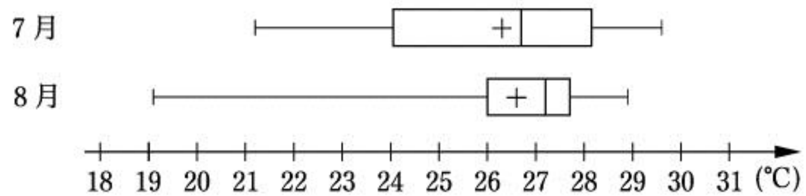


図 1 のように，平均値のみでは各月の日ごとの平均気温の分布のようすを比較することはできないが，図 2 のように箱ひげ図を用いることで，分布のようすを，散らばり具合を含めて比較することができる。

たとえば，7 月と 8 月の箱ひげ図は下の図のようになる。

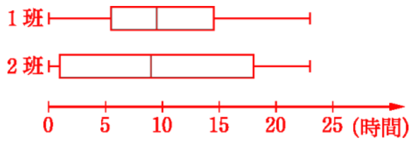


この箱ひげ図より，7 月と 8 月の平均気温の平均値はあまり変わらないが，箱の横の長さやひげの長さを比較することにより，データの散らばり具合の違いがわかる。

このように箱ひげ図は複数のデータの分布を比較するのに適している。

問8 167 ページの問 5 の 1 班，2 班について，それぞれ箱ひげ図をかき，分布を比較せよ。

1 班，2 班の読書時間の箱ひげ図は次のようになる。



箱の横の長さを比較すると，2 班の方が散らばり具合が大きい。

分散と標準偏差

（教科書 p.171）

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，…，x_n であり，その平均値を $\overline{x}$ とするとき

（[25]　平均値からの偏差　）あるいは単に（[26]　偏差　）

：　個々の値と平均値との差　$x_1-\overline{x}$，$x_2-\overline{x}$，…，$x_n-\overline{x}$

偏差の総和を計算すると

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

偏差の平均値ではデータの散らばり具合いを表すことができない。

偏差を 2 乗した値

$$(x_1-\overline{x})^2,\ (x_1-\overline{x})^2,\ \cdots,\ (x_1-\overline{x})^2$$

を考えると，これらの値はすべて 0 以上で，データの値が平均値 $\overline{x}$ から離れているほど大きくなる。

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

例7 教科書 166 ページの的当てゲームにおける A の得点の分散 $s_A{}^2$，標準偏差 s_A を求めてみよう。

A の得点の平均値は 6 点であるから，偏差は次の表のようになる。

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
偏差	1	−3	0	−2	2	1	1	−1	1	0

よって

$$s_A{}^2 = \frac{1}{10}\{1^2 + (-3)^2 + 0^2 + (-2)^2 + 2^2 + 1^2 + 1^2 + (-1)^2 + 1^2 + 0^2\}$$
$$= 2.2$$

$s_A = \sqrt{2.2} \fallingdotseq 1.48$（点）

166 ページの的当てゲームにおける B の得点の分散 $s_B{}^2$，標準偏差 s_B を，例 7 と同様に求めると，

$s_B{}^2 = 5.4$

$s_B = \sqrt{5.4} \fallingdotseq 2.32$（点）

となる。したがって，分散，標準偏差の値は（ B ）の方が大きいので，得点の散らばり具合は（ B ）の方が大きいことがわかる。

問9 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。

回	1	2	3	4	5
得点	82	76	63	80	64

得点の平均値は

$\frac{1}{5}(82 + 76 + 63 + 80 + 64) = 73$

よって偏差は次の表のようになる。

回	1	2	3	4	5
得点の偏差	9	3	−10	7	−9

したがって，分散は

$\frac{1}{5}\{9^2 + 3^2 + (-10)^2 + 7^2 + (-9)^2\} = 64$

標準偏差は

$\sqrt{64} = 8$（点）

Training

（教科書 p.173）

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，次の問に答えよ。

点数の階級 点以上　点未満	度数	相対度数
0 ～ 10	6	0.12
10 ～ 20	8	0.16
20 ～ 30	10	0.20
30 ～ 40	14	0.28
40 ～ 50	12	0.24
計	50	1.00

(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。

0 点以上 10 点未満の階級の度数は

$50 \times 0.12 = 6$

30 点以上 40 点未満の階級の度数は

$50 \times 0.28 = 14$

20 点以上 30 点未満の階級の度数は

$50 - (6 + 8 + 14 + 12) = 10$

20 点以上 30 点未満の階級の相対度数は

$\frac{10}{50} = 0.20$

度数がもっとも多い階級は「30 点以上 40 点未満」なので，最頻値はその階級値 35 点

(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	5	6	30
10 ～ 20	15	8	120
20 ～ 30	25	10	250
30 ～ 40	35	14	490
40 ～ 50	45	12	540
計		50	1430

度数分布表に階級値 x と度数 f の積 xf の欄をつけ加えると，表のようになる。

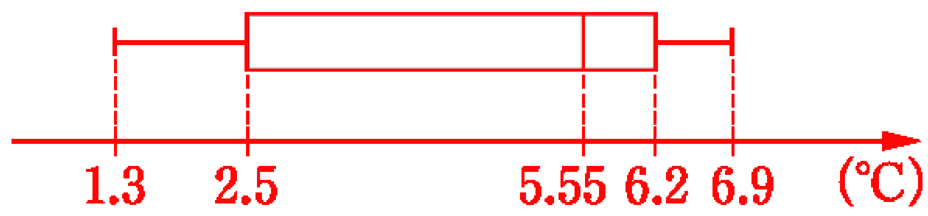
表より，平均値は

$\frac{1430}{50} = 28.6$（点）

2　ある都市における 1 日の平均気温を調べた。次の問に答えよ。

(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



データの値を小さい順に並べかえると

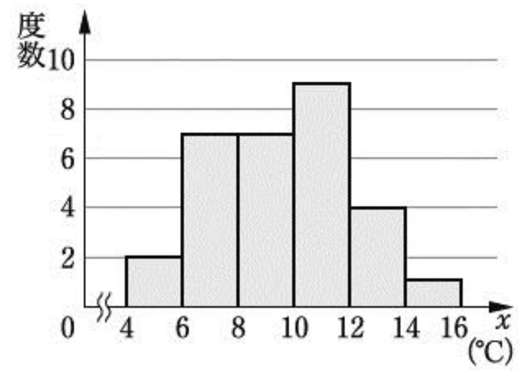
1.3　1.8　2.5　2.6　5.3　5.8　5.8　6.2　6.4　6.9　（単位 ℃）

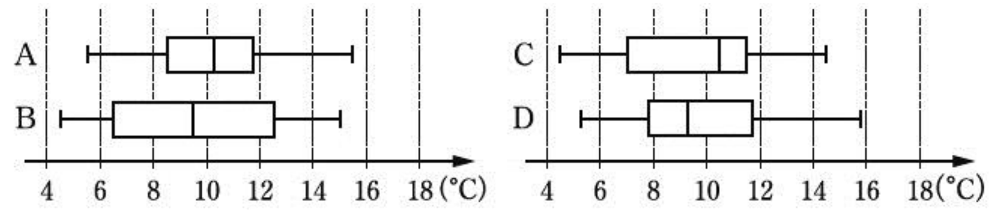
中央値　$\frac{5.3+5.8}{2} = 5.55$ (℃)

第 1 四分位数　2.5 ℃

第 3 四分位数　6.2 ℃

(2)　右の図は 11 月の 30 日間における 1 日の平均気温のデータのヒストグラムである。このデータを箱ひげ図にまとめると，次の A ～ D のいずれかになった。このデータの箱ひげ図は A ～ D のどれか。





ヒストグラムより

第 1 四分位数は 6℃ から 8℃ の間

中央値は 8℃ から10℃ の間

第 3 四分位数は 10℃ から 12℃ の間

これらを満たす箱ひげ図は D だけである。

よって，このデータの箱ひげ図は D

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

A の平均値

$\frac{1}{5}(7+5+8+6+4)=6$（点）

B の平均値

$\frac{1}{5}(2+7+10+5+6)=6$（点）

よって，偏差は表のようになる。

回	1	2	3	4	5
A の得点の偏差	1	−1	2	0	−2
B の得点の偏差	−4	1	4	−1	0

A の分散

$\frac{1}{5}\{1^2+(-1)^2+2^2+0^2+(-2)^2\}=\frac{1}{5}\times 10=2$

B の分散

$\frac{1}{5}\{(-4)^2+1^2+4^2+(-1)^2+0^2\}=\frac{1}{5}\times 34=6.8$

参考 分散の計算

（教科書 p.174）

5 個の値 x_1，x_2，x_3，x_4，x_5 からなるデータがある。このデータの平均値を $\overline{x}$ とすると

$$\frac{1}{5}(x_1+x_2+x_3+x_4+x_5)=\overline{x}$$

が成り立つ。

データの分散を s^2 とすると

$$\begin{aligned}s^2&=\frac{1}{5}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+(x_3-\overline{x})^2+(x_4-\overline{x})^2+(x_5-\overline{x})^2\}\\&=\frac{1}{5}\{(x_1{}^2+x_2{}^2+x_3{}^2+x_4{}^2+x_5{}^2)-2\overline{x}(x_1+x_2+x_3+x_4+x_5)+5(\overline{x})^2\}\\&=\frac{1}{5}(x_1{}^2+x_2{}^2+x_3{}^2+x_4{}^2+x_5{}^2)-2\overline{x}\cdot\frac{1}{5}(x_1+x_2+x_3+x_4+x_5)+(\overline{x})^2\\&=\frac{1}{5}(x_1{}^2+x_2{}^2+x_3{}^2+x_4{}^2+x_5{}^2)-2(\overline{x})^2+(\overline{x})^2\\&=\frac{1}{5}(x_1{}^2+x_2{}^2+x_3{}^2+x_4{}^2+x_5{}^2)-(\overline{x})^2\end{aligned}$$

となり，データの値を 2 乗した値の平均値から，データの平均値の 2 乗を引いた値が分散に等しいことがわかる。

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

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

例1 上の分散の計算式を用いて，教科書 166 ページの的当てゲームにおける A の得点の分散を求めてみよう。

$$\begin{aligned}s_A{}^2&=\frac{1}{10}(7^2+3^2+6^2+4^2+8^2+7^2+7^2+5^2+7^2+6^2)-6^2\\&=\frac{1}{10}\times382-36=2.2\end{aligned}$$

問1 上の計算式を用いて，教科書166ページの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

B の得点の分散 $s_B{}^2$ は

$$\begin{aligned}s_B{}^2&=\frac{1}{10}(9^2+7^2+2^2+8^2+6^2+3^2+4^2+7^2+9^2+5^2)-6^2\\&=\frac{1}{10}\times414-36=5.4\end{aligned}$$