1節 データの活用

1 度数分布

下のＡ表とＢ表は，ある中学校の１年生と２年生の女子 20 人ずつについて，反復横とびの記録を調べたものである。

とんだ回数のように，ある特性を数量的に表したものを **変量** といい，変量を集めたものを **データ** という。

［データ１］　反復横とび（K 中学校１年女子，２年女子）

Ａ表（１年女子）				Ｂ表（２年女子）			
番号	とんだ回数（回）	番号	とんだ回数（回）	番号	とんだ回数（回）	番号	とんだ回数（回）
1	34	11	35	1	32	11	38
2	39	12	46	2	46	12	34
3	41	13	37	3	37	13	47
4	39	14	29	4	42	14	42
5	30	15	36	5	33	15	28
6	52	16	33	6	48	16	43
7	28	17	24	7	41	17	35
8	44	18	38	8	54	18	44
9	31	19	43	9	40	19	39
10	34	20	48	10	50	20	42

度数分布表

右の表は，上のＡ表の記録を５回ごとの区間に区切り，その区間にはいる人数を調べたものである。

このように整理した１つ１つの区間を **階級**（かいきゅう） という。

右の表では，階級の幅は，５回
　　　　　階級の個数は，７個
である。

反復横とび（Ａ表）

階級（回）	度数（人）
（階級の幅） 以上　　未満	
20 ～ 25	1
25 ～ 30	2
30 ～ 35	5
35 ～ 40	6
40 ～ 45	3
45 ～ 50	2
50 ～ 55	1
計	20

（階級の個数）

各階級にはいるデータの値の個数(前ページの表では，人数)を，その階級の **度数**(どすう) といい，階級に応じて，度数を上のように整理した表を **度数分布表**(どすうぶんぷひょう) という。

また，最初の階級からその階級までの度数の合計を **累積度数** という。

反復横とび(A表)

階級(回)	度数(人)	累積度数(人)	
以上 未満			
20 〜 25	1	1	
25 〜 30	2	3	……1＋2＝3
30 〜 35	5	8	……3＋5＝8
35 〜 40	6	14	……8＋6＝14
40 〜 45	3	17	
45 〜 50	2	19	
50 〜 55	1	20	
計	20		

問1 前ページのB表を，右の度数分布表に整理しなさい。

反復横とび(B表)

階級(回)	度数(人)	累積度数(人)
以上 未満		
25 〜 30		
30 〜 35		
35 〜 40		
40 〜 45		
45 〜 50		
50 〜 55		
計		

問2 A表，B表の度数分布について，それぞれ，次のことを調べなさい。

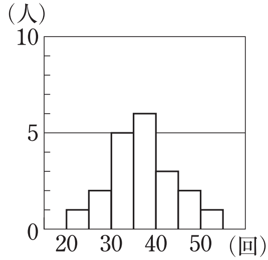
(1) 度数がもっとも多いのは，どの階級ですか。

(2) 記録が 45 回以上の生徒は何人ですか。

ヒストグラム・度数分布多角形

度数分布表は，グラフに表すと見やすくなることがある。

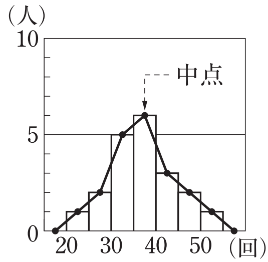
右のグラフは，前ページのA表の度数分布表を，横軸を回数，縦軸を人数としてグラフに表したものである。



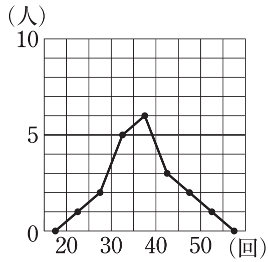
このグラフでは，階級の幅を横，度数を縦とする長方形を並べている。

この長方形の面積は，各階級の度数に比例している。このようなグラフを **ヒストグラム** という。

ヒストグラムで，1つ1つの長方形の上の辺の中点を，順に線分で結ぶと，右の図のような折れ線グラフができる。ただし，両端では，度数0の階級があるものと考える。このような折れ線グラフを **度数分布多角形** という。



問3 右の図は，A表の度数分布多角形です。これに **問1** でつくったB表の度数分布表をもとにして，度数分布多角形をかき入れなさい。



度数分布多角形を重ねると，2つのデータがくらべやすくなる。上の2つの度数分布多角形から，どのようなことがいえるだろうか。

相対度数

下の表は，ある 2 つの中学校の 1 年生の男子について，握力を調べ，その結果を度数分布表に整理したものである。

握　力（K，S 両中学校 1 年男子）

階級（kg）	K 中学校 度数（人）	S 中学校 度数（人）
以上　未満		
15 〜 20	1	5
20 〜 25	3	15
25 〜 30	6	28
30 〜 35	10	34
35 〜 40	8	25
40 〜 45	7	8
45 〜 50	2	4
50 〜 55	1	1
計	38	120

データの個数が異なる

2 校の握力の分布を比較するには，上のような度数分布表では全体の度数が違うので，比較しにくい。

このようなときは，各階級の度数の，全体に対する割合を求めて，その割合で比較することができる。

各階級の度数の，全体に対する割合を，その階級の **相対度数** という。

$$\textbf{相対度数}=\frac{\textbf{階級の度数}}{\textbf{度数の合計}}$$

例 1　相対度数の求め方

上の表の K 中学校で，25 kg 以上 30 kg 未満の階級の相対度数を小数第 2 位まで求める。

この階級の度数が 6，度数の合計が 38 だから，

$$\frac{6}{38}=0.1\overset{6}{\cancel{5}}\cancel{7}\cdots$$

したがって，　0.16

問 4 下の表は，前ページの表から，K中学校の相対度数を求めたものです。S中学校についても，相対度数を求めて，書き入れなさい。

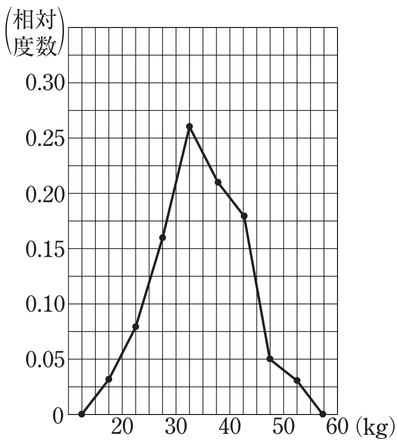
握　力 (K，S 両中学校 1 年男子)

階級 (kg)	K中学校		S中学校	
	度数 (人)	相対度数	度数 (人)	相対度数
以上 15 ～ 20 未満	1	0.03	5	
20 ～ 25	3	0.08	15	
25 ～ 30	6	0.16	28	
30 ～ 35	10	0.26	34	
35 ～ 40	8	0.21	25	
40 ～ 45	7	0.18	8	
45 ～ 50	2	0.05	4	
50 ～ 55	1	0.03	1	
計	38	1.00	120	

相対度数を求めると，個数の異なる 2 つのデータをくらべやすくなる。K中学校とS中学校の握力について，どのようなことがいえるだろうか。

注 相対度数の合計は，各度数を四捨五入しているから，**1** にならないこともある。

問 5 次の図は，上の表から，K中学校の相対度数を度数分布多角形に表したものです。この図に，S中学校の度数分布多角形をかき入れなさい。



累積度数と同じように，最初の階級から，その階級までの相対度数の合計を **累積相対度数** という。

握　力（K中学校1年男子）

階級（kg）	度数（人）	相対度数	累積相対度数
以上　未満			
15 ～ 20	1	0.03	0.03
20 ～ 25	3	0.08	0.11
25 ～ 30	6	0.16	0.27
30 ～ 35	10	0.26	
35 ～ 40	8	0.21	
40 ～ 45	7	0.18	
45 ～ 50	2	0.05	
50 ～ 55	1	0.03	
計	38	1.00	

……$0.03+0.08=0.11$

……$0.11+0.16=0.27$

問6 上の表で，各階級の累積相対度数を求め，空欄に書き入れなさい。

問　題

1 次の表は，あるクラスの生徒のテストの点数の度数分布表です。これをもとにして，次の問いに答えなさい。

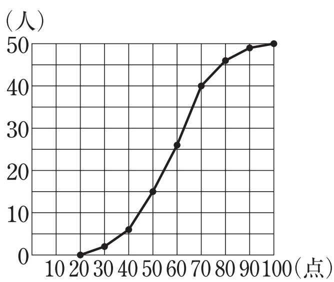
階級（点）	度数（人）	相対度数	累積度数	累積相対度数
以上　未満				
20 ～ 30	2			
30 ～ 40	4			
40 ～ 50	9			
50 ～ 60	11			
60 ～ 70	14			
70 ～ 80	6			
80 ～ 90	3			
90 ～ 100	1			
計				

(1)　上の表を完成させなさい。

(2)　点数が60点未満の生徒は，全体の何％ですか。

(3)　累積度数と累積相対度数の関係について，気づいたことを答えなさい。

(4)　右のグラフは，上の表において，何を表しているか答えなさい。



2　代表値

代表値

データの値全体を1つの値で代表させ，これを基準にして，ものごとを考えたり判断したりすることがある。

このようなとき，データの値全体を代表する値を **代表値** という。

代表値には，**平均値**，**中央値**（**メジアン**），**最頻値**（さいひんち）（**モード**）などがある。

それぞれの代表値には，
どのような特徴があるだろうか。

平均値

5人の身長が次のようなとき，その平均値を求めてみましょう。
173，169，163，176，170　（cm）

平均値は，次の式で求める。

$$\text{平均値}=\frac{\text{データの個々の値の合計}}{\text{データの個数}}$$

で，5人の身長の平均値は，次のようになる。

$$\frac{173+169+163+176+170}{5}=170.2\,(\text{cm})$$

問7　ある中学校の生徒20人について，先月読んだ本の冊数を調べたところ，下のような結果になりました。この20人の読んだ本の冊数の平均値を求めなさい。

3，6，14，5，7，2，1，18，5，4，
8，5，9，13，10，11，3，2，8，11　（冊）

で，仮に平均値を 170 cm として考えてみよう。

この 5 つの値と仮の平均値 170 cm との差は，それぞれ

$3,\ -1,\ -7,\ 6,\ 0$

で，この差の平均値は，

$$\frac{1}{5}\times\{3+(-1)+(-7)+6+0\}=\frac{1}{5}=0.2$$

となるから，5 人の身長の平均値は，

$$170+0.2=170.2\ (\mathrm{cm})$$

上のような仮の平均値のことを仮平均といい，データの平均値は，仮平均を用いると次のように表すことができる。

（平均値）＝（仮平均）＋（個々の値と仮平均との差の平均値）

度数分布表だけが与えられた場合は，データの個々の値がわからないので，前ページのような計算はできない。そのようなときには，どのような値を平均値として考えればよいだろうか。

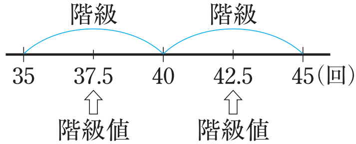
度数分布表では，1 つの階級にはいっているデータの個々の値はいろいろだが，「どの値もすべて，その階級のまん中の値である」とみなして，その平均値を計算する。

度数分布表で，各階級のまん中の値を **階級値** という。

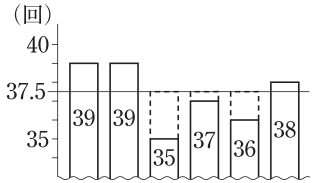
例えば，3 ページのA表で，35 回以上 40 回未満の階級の階級値は，

$$\frac{35+40}{2}=37.5\ (回)$$

である。



35 回以上 40 回未満の階級にはいっている 6 人の生徒の記録は，すべて 37.5 回とみなすと，この 6 人の記録の合計は，



階級値×度数

で，

$$37.5\times6=225\ (\text{回})$$

となる。

ほかの階級でも同じように考えて，度数分布表から全体の回数の合計を出し，平均値を求めることができる。

例 2 度数分布表から求めた平均値

3 ページのA表を用いると，平均値は，

$$\frac{22.5\times1+27.5\times2+32.5\times5+37.5\times6+42.5\times3+47.5\times2+52.5\times1}{20}$$

$$=\frac{740}{20}=37\ (\text{回})$$

実際の平均値は何回だろうか。

問 8 下の表は，あるクラスの通学時間についての度数分布表です。表の空欄をうめて，このクラスの通学時間の平均値を求めなさい。

通　学　時　間

階 級 (分)	階級値 (分)	度数 (人)	階級値×度数
以上 0 ～ 10 未満		5	
10 ～ 20		9	
20 ～ 30		10	
30 ～ 40	35	3	105
40 ～ 50		2	
50 ～ 60		1	
計		30	

中央値

データの値を大きさの順に並べたとき，その中央の値を **中央値** または，**メジアン** という。

データの個数が奇数の場合は，まん中の値が中央値であり，データの個数が偶数の場合は，中央に並ぶ 2 つの値の平均をとって中央値とする。

例 3 中央値

(1) 5, 2, 9, 2, 8, 1, 7 の中央値

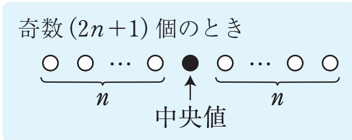
値を大きさの順に並べると，

1, 2, 2, $\boxed{5}$, 7, 8, 9

① ② ③ ④ ⑤ ⑥ ⑦

4 番目がまん中であるから，

中央値は 5 である。



(2) 6, 3, 8, 7, 0, 2, 5, 7 の中央値

値を大きさの順に並べると，

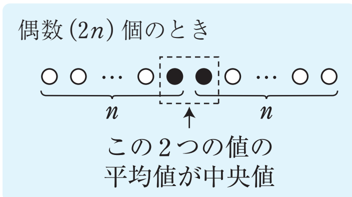
0, 2, 3, $\boxed{5,\ 6}$, 7, 7, 8

① ② ③ ④ ⑤ ⑥ ⑦ ⑧

4, 5 番目の値 5, 6 が中央に並ぶので，中央値は，

$$\frac{5+6}{2}=5.5$$

である。



問 9 問 7 の生徒 20 人が先月読んだ本の冊数の中央値を求めなさい。

最頻値

データの中で，もっとも多く現れる値を **最頻値**（さいひんち） または，**モード** という。度数分布表では，度数のもっとも多い階級の階級値を最頻値とする。

例 4 最頻値

ある生徒の小テスト（10 点満点）の12 回分の結果が，

7，6，6，8，10，8，9，7，6，8，5，8

となったとき，点数の最頻値は，8 点

問10 問 7 の生徒 20 人が先月読んだ本の冊数の最頻値を求めなさい。

範囲

2 つの容器 A，B に，卵が 10 個ずつはいっています。それぞれの容器にはいった卵の重さを 1 個ずつはかると，右の表のようになりました。
いずれの容器も，平均値は 50.5 g，中央値は 50.6 g となっています。
2 つの容器の卵の重さの分布は，ほぼ同じといってよいでしょうか。

平均値，中央値の他にどんな値を調べるとよいだろうか。

［データ 2］ 卵の重さ（g）

容器A	容器B
50.1	43.2
48.7	50.3
50.5	57.1
52.1	53.7
47.8	50.2
48.4	44.9
52.2	50.9
50.7	55.3
53.3	45.8
51.2	53.6

の容器 A と B では，最大の値と最小の値に違いがある。平均値や中央値に加え，最大値と最小値の差を考えると，より詳しく分布を調べることができる。

データの最大の値と最小の値の差を，分布の **範囲**（はんい） または，**レンジ** という。

範囲＝最大値－最小値

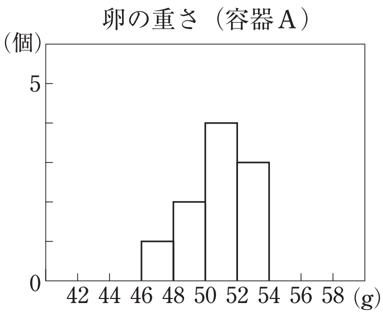
問11 容器 A，B の卵の重さの範囲を求めなさい。

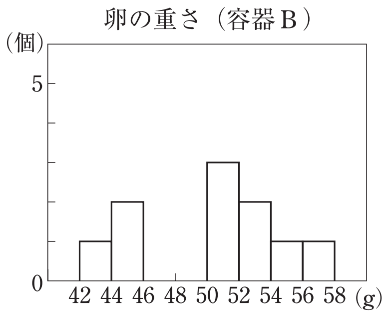
データ 2 における容器 A，B の度数分布表は右のようになる。

度数分布表から，2 つの容器の卵の重さについて，どんなことがいえるだろうか。

卵の重さ

階級 (g)	容器A 度数 (個)	容器B 度数 (個)
以上 42 ～ 44 未満	0	1
44 ～ 46	0	2
46 ～ 48	1	0
48 ～ 50	2	0
50 ～ 52	4	3
52 ～ 54	3	2
54 ～ 56	0	1
56 ～ 58	0	1
計	10	10





問題

1 あるクラスの生徒 35 人のテストの結果の平均値を求めると，65 点でした。この結果からかならずいえることを，次の(ア)～(ウ)から選びなさい。 p.8～10

(ア) 点数が 65 点だった生徒がいちばん多い。

(イ) 点数が 18 番目に高かった生徒の点数は 65 点である。

(ウ) 全員の点数を合計すると 2275 点である。

3　データの散らばりと四分位数

次のデータは，A組 (9 人) とB組 (10 人) で漢字の小テスト (10 点満点) を行ったときの点数である。

データ 3　漢字テストの点数

A組	5,	3,	9,	7,	10,	4,	8,	8,	9	
B組	9,	6,	4,	8,	8 ,	5,	9,	6,	7,	8

この 2 組のデータの平均点はいずれも 7 点で同じであるが，点数の分布の様子は異なっているように見える。

このようなデータの散らばり具合を数値で表すことを考えてみよう。

四分位数

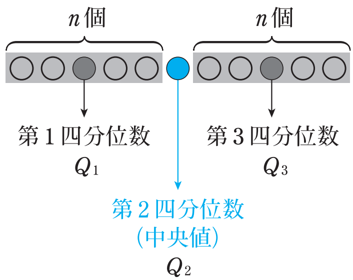
データの値を小さい順に並べたとき，そのデータを 4 等分する位置の値を **四分位数** という。

小さい方から順に，**第 1 四分位数，第 2 四分位数，第 3 四分位数** といい，それぞれ Q_1，Q_2，Q_3 で表す。

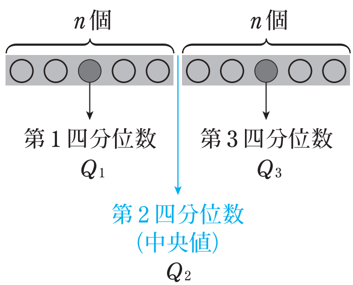
とくに，第 2 四分位数 Q_2 は，データの中央値である。

四分位数を求めるには，まず，データの値を小さい順に並べて，中央値 Q_2 を求める。次に，Q_2 を境にして左と右に分け，それぞれの中央値 Q_1，Q_3 を求めればよい。

奇数$(2n+1)$のとき



偶数$(2n)$のとき



例 5 **データ 3 における，A組とB組の点数の四分位数**

(1)　A組のデータの値を小さい順に並べると，

3，4，5，7，8，8，9，9，10

Q_1　中央値　Q_3

よって，$Q_1=\dfrac{4+5}{2}=4.5$ (点)，$Q_2=8$ (点)，

$Q_3=\dfrac{9+9}{2}=9$ (点)

(2)　B組のデータの値を小さい順に並べると，

4，5，6，6，7，8，8，8，9，9

Q_1　中央値　Q_3

よって，$Q_1=6$ (点)，$Q_2=\dfrac{7+8}{2}=7.5$ (点)，$Q_3=8$ (点)

問12 次のデータは，C組の生徒 8 人の漢字の小テスト (10 点満点) の点数です。このデータの四分位数 Q_1，Q_2，Q_3 を求めなさい。

4，8，5，9，5，7，8，6

データの範囲が大きいほど，散らばり具合が大きいと考えられる。

第 3 四分位数 Q_3 と第 1 四分位数 Q_1 との差 Q_3-Q_1 を **四分位範囲** といい，四分位範囲の半分を **四分位偏差** という。

たとえば，

データ 3 のA組の四分位範囲は，$Q_3-Q_1=9-4.5=4.5$ (点)

四分位偏差は，$4.5\div 2=2.25$ (点)

である。

箱ひげ図

データの散らばりを示す指標として考えた5つの数値

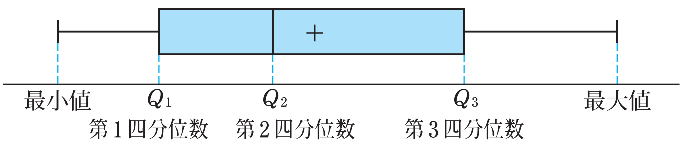
最小値，第1四分位数 Q_1，中央値（第2四分位数 Q_2），

第3四分位数 Q_3，最大値

を **5数要約** という。

この5数要約を1つの図に示したものが **箱ひげ図** である。

箱ひげ図は，Q_1 と Q_3 を両端とする長方形（箱）をかき，中央値 Q_2 で箱の内部に線を引き，最小値と Q_1，Q_3 と最大値を線分（ひげ）で結んだ図である。



箱の右端から左端までの長さが四分位範囲を示している。

箱ひげ図に平均値を記入するときは，上の図のように＋記号で表すこともある。

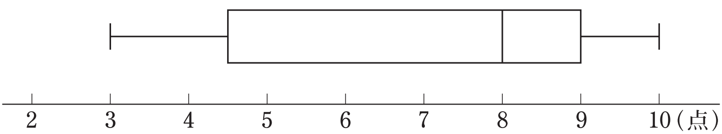
箱ひげ図は，データの分布のおおまかな様子を見るのに便利である。

例6 **14ページのデータ3における，A組の点数の箱ひげ図**

例5 から，

最小値3，$Q_1=4.5$，$Q_2=8$，$Q_3=9$，最大値10

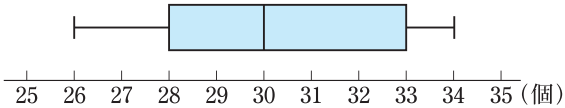
よって，下のような箱ひげ図が得られる。



問13 データ3において，B組の点数の箱ひげ図をかきなさい。

例 7 箱ひげ図を用いて，データの傾向を読みとる。

みかんが入った段ボール箱が 40 箱ある。次の箱ひげ図は，1 箱の中に入っているみかんの個数を数えた結果を表している。

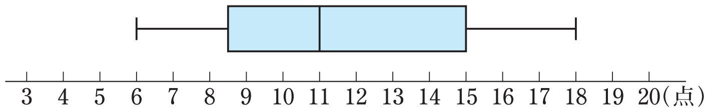


箱ひげ図から，次のことがわかる。

・1 箱に入っているみかんの個数は 26 個から 34 個である。

・28 個から 33 個のみかんが入った箱は，約 20 箱ある。

問14 下の箱ひげ図は，45 人の生徒が受験したテストの結果を表したものです。次の □ にあてはまる数を答えなさい。



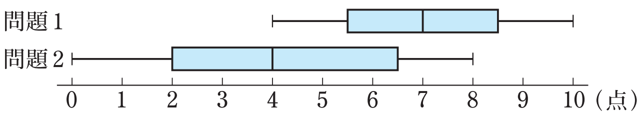
(1) テストの結果は，□ 点から □ 点までである。

(2) 約半分の生徒の点数が，9 点以上 □ 点以下である。

(3) □ 点以上の生徒は，約 11 人いる。

複数のデータの分布の様子を比較するときにも，箱ひげ図は有効である。

たとえば，次の箱ひげ図は，45 人が受験したテストにおける問題 1 と問題 2 の結果を表している。



箱ひげ図から，次のことなどがわかる。

7 点以上をとった人数を比較すると，問題 1 の人数は問題 2 の人数の 2 倍以上である。

例題 1 次のデータは，ある人の最高血圧をＡ機器とＢ機器で 10 回ずつ測定した結果です。

Ａ機器	107	105	102	103	105	102	100	106	98	95
Ｂ機器	99	102	101	97	97	99	102	100	97	103

（単位は mmHg）

(1) 各機器のデータについて，箱ひげ図を並べてかきなさい。

(2) 2 機器のデータを比較して，どちらの散らばりが大きいか答えなさい。

考え方 (2) 範囲や四分位範囲に注目して判断する。

解答：(1) 略　(2) Ａ機器

問15 次のデータは，ある人の最低血圧をＡ機器とＢ機器で 9 回ずつ測定した結果です。

Ａ機器	71	72	69	81	78	75	67	78	62
Ｂ機器	57	55	58	51	50	51	60	55	58

（単位は mmHg）

(1) 各機器のデータについて，箱ひげ図を並べてかきなさい。

(2) 2 機器のデータを比較して，どちらの散らばりが大きいか答えなさい。

ヒストグラムと箱ひげ図

同じデータを使って表示されたヒストグラムと箱ひげ図を比較して，その分布の特徴を考えてみよう。

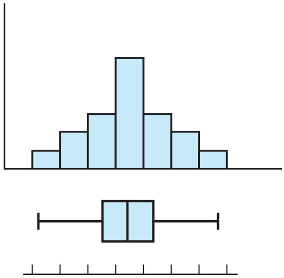
箱ひげ図では，ヒストグラムほどにデータの分布が詳しく表現できないが，おおよその様子はわかる。

1．ヒストグラムの形状がほぼ対称である場合

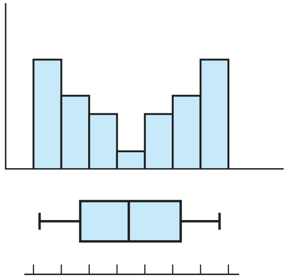
① 箱ひげ図は中央値を中心にほぼ対称になる。

② データが中央に集まるほど箱の長さは短い。

A



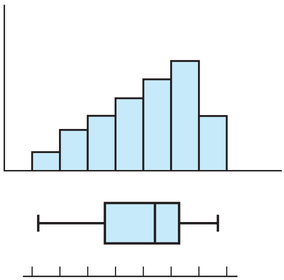
B



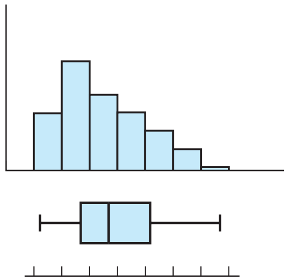
2．ヒストグラムの形状が偏っている場合

次のC，Dのようなヒストグラムでは，箱は最頻値の方に寄り，反対側のひげは長くなる。

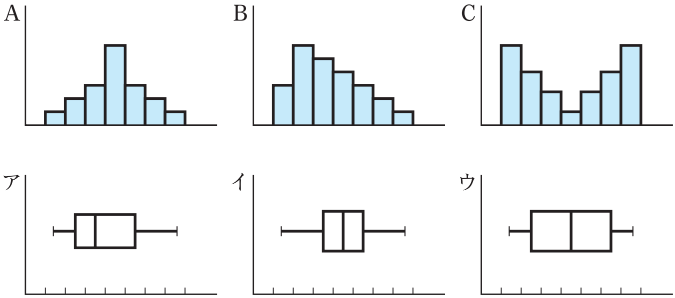
C



D



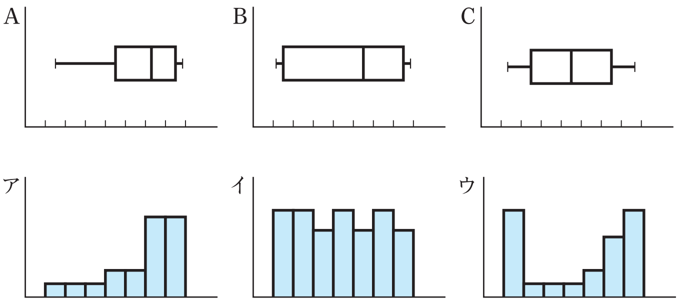
例題 2　次のＡ～Ｃのヒストグラムについて，対応する箱ひげ図をア～ウの中から選び，その理由を説明しなさい。



考え方　ヒストグラムの形状に着目する。

解答：Ａはイ，Ｂはア，Ｃはウ

問16　次のＡ～Ｃの箱ひげ図について，対応するヒストグラムをア～ウの中から選び，その理由を説明しなさい。



4　近似値

近似値

右の長方形の縦の長さと横の長さを，それぞれ，ものさしを使って，mm の位まで測ってみましょう。

長さなどを測って得られた値を **測定値** という。

上の長方形で，横の長さはほぼ 42 mm と読みとれるが，きっちり 42 mm ではない。このように，測定値は，どんなに精密に測ろうとしても，測定機器の精度や測定の方法に一定の限界があるため，正確には真の値になっていない。

測定値などのように，真の値に近い値のことを **近似値**（きんじち） という。

測定値以外の例として，円周率として用いる 3.14 のようなものもある。

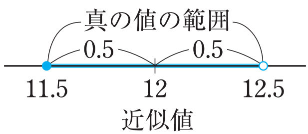
また，近似値から真の値をひいた差を **誤差**（ごさ） という。

誤差＝近似値－真の値

例 8　真の値の範囲

ある数 a の小数第 1 位を四捨五入した近似値が 12 であるとすると，a の値の範囲は，

$$11.5 \leqq a < 12.5$$



となる。このとき，誤差の絶対値は 0.5 以下であるといえる。

問17 ある数 a の小数第 2 位を四捨五入した近似値が 1.6 であるとき，a の値の範囲を，不等号を使って表しなさい。

有効数字

前ページの長方形で，縦の長さの測定値は 20 mm である。これを cm の単位で表すと 2 cm であるが，mm の位まで測ったことをはっきりさせるために，2.0 cm と表すことがある。このように，測定などによって得られた数のうち，意味のある数字を **有効数字** といい，その数字の個数を有効数字のけた数という。

例えば，2.0 cm での有効数字は 2 と 0 で，これは，有効数字 2 けたの近似値である。

大きい数字や小さい数字の場合，有効数字をはっきりさせるためには，整数部分が 1 けたの小数と，10 の累乗の積の形で表す。

例 9 **有効数字をはっきりさせた表し方**

地球の直径 12750 km を，

有効数字 3 けたで表すと，1.28×10^4 (km)

有効数字 2 けたで表すと，1.3×10^4 (km)

整数部分が 1 けたの小数

問18 次の近似値で，有効数字が 3 けたであるとき，整数部分が 1 けたの小数と，10 の累乗の積の形で表しなさい。

(1) 木星の直径 143000 km

(2) ある体育館の広さ 1210 m^2

補充問題

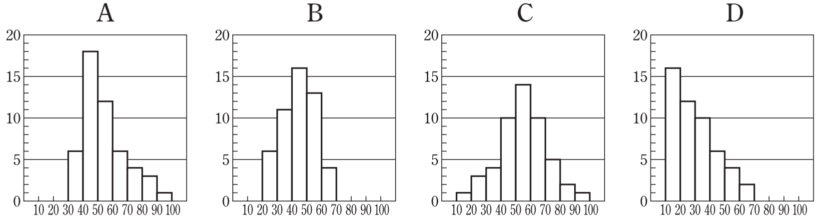
1 次の⑴～⑷のそれぞれにあてはまるものを，Ａ～Ｄのヒストグラムからすべて選びなさい。

⑴　平均値がもっとも小さいもの

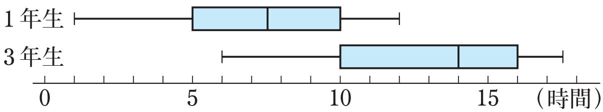
⑵　範囲がもっとも大きいもの

⑶　中央値が，40 以上 50 未満の階級にふくまれているもの

⑷　平均値と中央値と最頻値がほとんど同じになるもの



2 次の 2 つの箱ひげ図は，ある高校の 1 年生と 3 年生が，月曜日から金曜日までに行った数学の家庭学習の時間の分布を表しています。



次のＡ～Ｄについて，正しいものを選びなさい。

Ａ．１年生の学習時間で，最も長いのは 12 時間である。

Ｂ．３年生で，学習時間が 10 時間から 14 時間までの区間にいる人数は，14 時間から 16 時間までの区間にいる人数の 2 倍である。

Ｃ．学年全体の人数に対して，学習時間が 10 時間から 14 時間までの区間にいる人数の割合は，３年生の方が 1 年生よりも約 2 倍多い。

Ｄ．３年生の 50％ 以上は，すべての 1 年生よりも学習時間が長い。