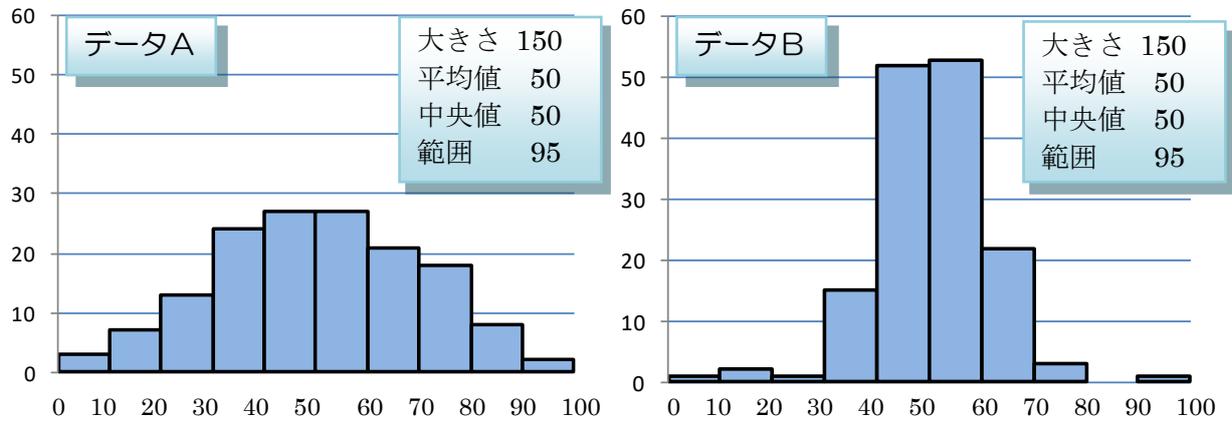


●データの特徴をあらわそう

例) 下の図はデータの大きさ、平均値、中央値、範囲がすべて等しい2つのデータA、Bのヒストグラムである。



問1) この表から読み取れることは何か。

- データAの方が低く、データBの方が高くデータが集まっている。
- データAの方が幅広くデータがあり、データBの方が狭くデータが集まっている。

など

◎ 次のことを用いてデータの散らばりの度合いを比較しよう

しぶんいすう
四分位数とは

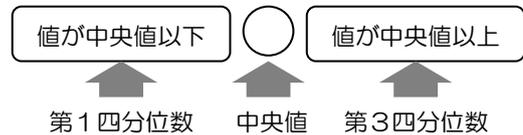
データの値を大きさの順に並べたとき、4等分する位置の値を四分位数という。四分位数は、小さい方から順に第1四分位数、第2四分位数(中央値と同じ値)、第3四分位数といい、順に Q_1 、 Q_2 、 Q_3 で表す。

四分位数を意味する英語は quartile

四分位数の求め方

- ① データを値の大きさの順に並べ、中央値(第2四分位数)を求める。
- ② ①の中央値を境界としてデータの個数を2等分し、値が中央値以下の下組と値が中央値以上の上組に分ける。ただし、データの大きさが奇数のとき、①の中央値はどちらの組にも含めないものとする。
- ③ 下組の中央値(第1四分位数)、上組の中央値(第3四分位数)を求める。

◇データの大きさが奇数



◇データの大きさが偶数



四分位範囲とは

第3四分位数 Q_3 から第1四分位数 Q_1 を引いた値 $Q_3 - Q_1$ のこと。この中に中央値周辺に並ぶ約50%のデータが含まれる。よって、四分位範囲は、データの中に極端に飛び離れた値がある場合でも、その影響を受けにくい。また、データの値が中央値の周りに集中しているほど、四分位範囲は小さくなる傾向にある。逆に四分位範囲が大きいほど、データの散らばりが大きいと言える。

四分位偏差とは

四分位範囲を2で割った値。つまり四分位範囲の半分のこと。

問2) データAの中央値、第1四分位数、第3四分位数、四分位範囲、四分位偏差を求めよう。

1, 2, 3, 11, 13, 14, 14, 14, 15, 19, 21, 23, 23, 23, 23, 26, 27, 27, 27, 27, 27, 27, 28, 28, 29, 30, 30, 30, 31, 31, 32, 32, 32, 33, 34, 35, 35, 35, 36, 36, 36, 36, 37, 37, 37, 38, 39, 39, 39, 40, 40, 42, 42, 42, 43, 43, 44, 44, 45, 45, 45, 45, 45, 45, 46, 46, 46, 46, 46, 48, 48, 48, 49, 49, 49, 50, 50, 50, 50, 51, 51, 52, 52, 53, 53, 53, 53, 53, 53, 54, 55, 55, 56, 56, 56, 57, 57, 57, 57, 58, 59, 60, 60, 61, 61, 61, 62, 63, 63, 64, 64, 64, 65, 65, 66, 66, 67, 68, 68, 68, 69, 69, 70, 70, 71, 72, 72, 72, 72, 73, 73, 74, 74, 75, 75, 76, 77, 78, 78, 78, 82, 82, 83, 85, 85, 87, 89, 89, 96, 96

問3) データBの各値は右の表のようになっている。
ヒストグラムと合わせて確認し、データA, Bの違いを説明しよう。

第1四分位数	36
中央値	50
第3四分位数	65
四分位範囲	29
四分位偏差	14.5

データB

第1四分位数	45
中央値	50
第3四分位数	56
四分位範囲	11
四分位偏差	5.5

• データAの方が四分位範囲が大きいので
データAの方が散らばりの度合いが大きい。

練習7) 次のデータA, Bのそれぞれについて、四分位範囲と四分位偏差を求めよ。
また、データの散らばりの度合いが大きいのはA, Bのどちらか。

A 21, 29, 32, 36, 38, 40, 49, 53, 55, 68, 80
B 25, 31, 39, 42, 45, 46, 50, 53, 54, 65, 80

第1四分位数 Q_1 中央値 第3四分位数 Q_3

- ① 小さい方から順に並べる
- ② 中央値→第1四分位数、第3四分位数を調べる
- ③ $Q_3 - Q_1$ を計算して四分位範囲、範囲を2で割って四分位偏差を求める

※^{みきはす}幹葉図を用いてもよい
例えばデータAであれば
一の位のみを書き並べていく

20	1	9
30	2	6 8
40	0	9
50	3	5
60	8	
70		
80	0	

A. データAの四分位範囲は $55 - 32 = 23$ 四分位偏差は 11.5
データBの四分位範囲は $54 - 39 = 15$ 四分位偏差は 7.5
よって散らばりの度合いが大きいと考えられるのは データA