

第5章「データの活用」 3. データの散らばりと四分位範囲

データの散らばり方を調べる方法を考えてみよう。
データの散らばりの程度を表す量を**散布度**という

<範囲>

右の表は、AさんとBさんの2人が、昨年1年間の各月に、図書館に行った回数のデータである。

2人とも1年間に48回図書館に行ったことから、1か月に図書館に行った回数の平均値はともに4回である。

一方、1か月に図書館に行った回数は

Aさん 最大：15回、最小：0回

Bさん 最大：6回、最小：2回

となり、回数の差が大きく異なる。

このような場合、平均値が等しくても、データの散らばり方は等しいとはいえない。

月	A (回数)	B (回数)
1	4	3
2	0	4
3	2	6
4	1	3
5	6	4
6	2	3
7	8	5
8	15	6
9	2	4
10	1	3
11	3	2
12	4	5
計	48	48

データのとる値のうち、最大のものから最小のものをひいた値を**範囲 (はんい)**という。
範囲は、データの散らばりの程度を表す。

(範囲)=(最大値)−(最小値)

範囲には、簡単に求められるという長所がある。しかし、データに例外的な値はずれ値というが含まれている場合にはその影響を大きく受けるという短所もある。

練習9 AさんとBさんについて、1か月に図書館に行った回数の範囲を、それぞれ求めなさい。

解答
Aさんについて、 $15 - 0 = 15$ より範囲は15回
Bさんについて、 $6 - 2 = 4$ より範囲は4回

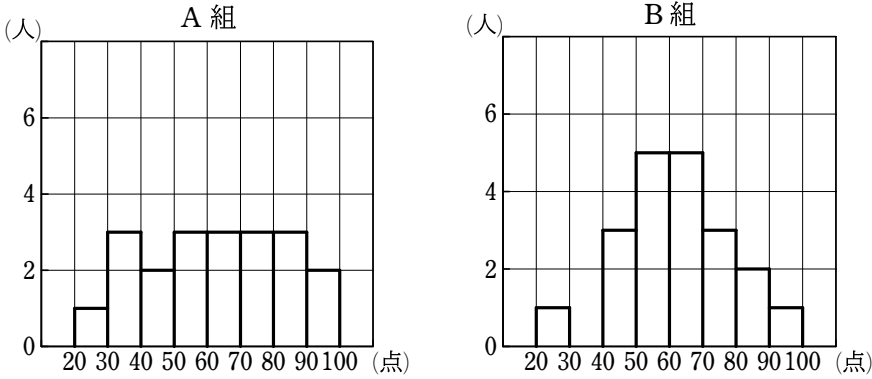
「範囲を求めよ」という問題に対して、なぜか不等号で表す者がいるので注意

<四分位数>

次のデータは、ある中学校のA組とB組のそれぞれ20人に行ったテストの結果を、得点の低い順に並べたものである。(単位は点)

A組	26	31	35	39	44	48	52	55	57	63
	67	68	74	75	78	82	85	87	92	93
B組	27	42	47	49	51	53	54	56	59	64
	66	68	69	69	72	76	79	82	85	94

それぞれのデータをヒストグラムで表すと、次の図のようになる。



A組のデータの範囲は $93 - 26 = 67$ (点),

B組のデータの範囲は $94 - 27 = 67$ (点)

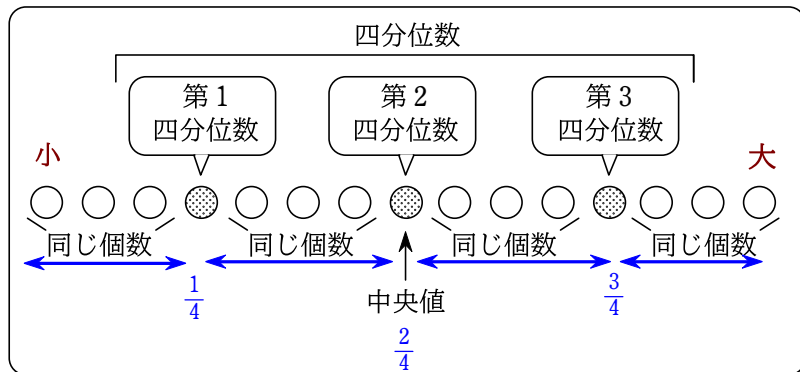
であり、ともにデータの範囲は等しいが、ヒストグラムの山の形や高さには違いが見られる。

また、範囲はデータの最大値と最小値だけで決まるため、データの中に極端に大きな値や小さな値があると、それによって範囲は大きく変わってしまう。そこで、データを値の大きさの順に並べて4等分し、中央値付近のデータについて考えることがある。

データを値の大きさの順に並べたとき、4等分する位置にくる値を

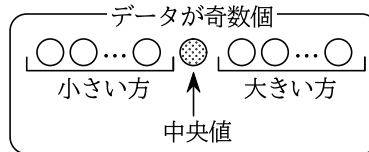
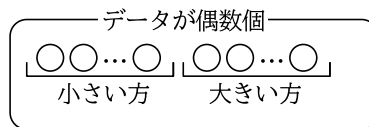
四分位数 (しぶんいすう) という。四分位数は、小さい方から順に

第1四分位数、**第2四分位数**、**第3四分位数** という。第2四分位数は中央値のことである。

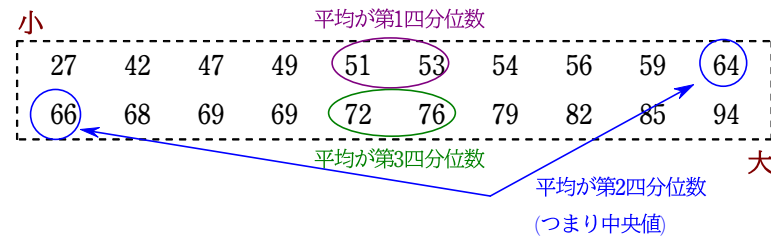


第1四分位数と第3四分位数は、次のように求める。

- [1] 値の大きさの順に並べたデータを、個数が同じになるように半分に分ける。ただし、データの個数が奇数のときは、中央値を除いて2つに分ける。
- [2] 半分にしたデータのうち、小さい方のデータの中央値が第1四分位数、大きい方のデータの中央値が第3四分位数となる。



練習 10 123 ページの B 組のデータの四分位数を求めなさい。



$$\text{第2四分位数は } \frac{64 + 66}{2} = 65$$

$$\text{第1四分位数は } \frac{51 + 53}{2} = 52$$

$$\text{第3四分位数は } \frac{72 + 76}{2} = 74$$

< 四分位範囲 >

122 ページで学んだ範囲よりも、中央値に近いところでのデータの散らばりの程度を調べる。

中央で調べれば、はずれ値の影響を受けにくい

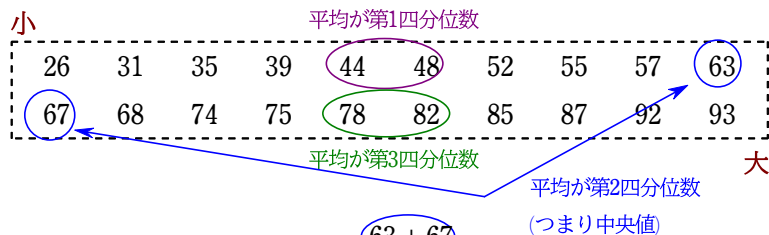
第3四分位数から第1四分位数をひいた差を **四分位範囲** という。

四分位範囲

$$(\text{四分位範囲}) = (\text{第3四分位数}) - (\text{第1四分位数})$$

四分位範囲内にはデータの約50%が含まれる

例 2 123 ページの A 組のデータの四分位数を求める。



$$\text{中央値すなわち第2四分位数は } \frac{63 + 67}{2} = 65$$

$$\text{第1四分位数は } \frac{44 + 48}{2} = 46$$

$$\text{第3四分位数は } \frac{78 + 82}{2} = 80$$

参考 四分位範囲を2でわった値を **四分位偏差 (へんさ)** という。

$$\begin{aligned} (\text{四分位偏差}) &= \frac{(\text{四分位範囲})}{2} \\ &= \frac{(\text{第3四分位数}) - (\text{第1四分位数})}{2} \end{aligned}$$

研究

偏差値って知ってますか？

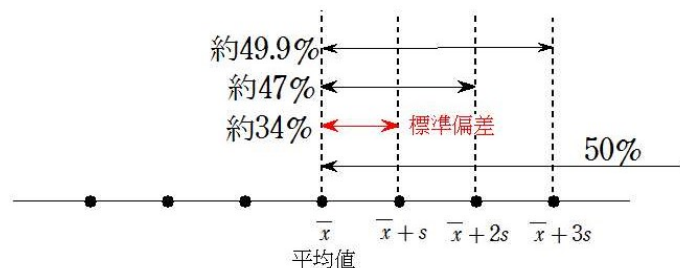
偏差値が50だと自分は集団の中のどこにいるの？

偏差値が60だと自分は集団の中のどこにいるの？

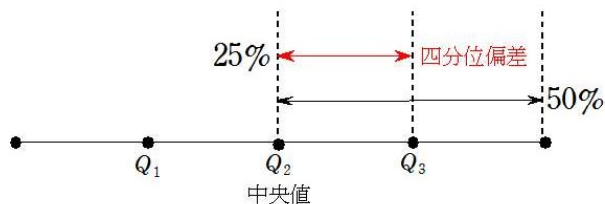
標準偏差と四分位偏差

● 標準偏差 (standard deviation)

記号で s や SD , σ などで表されます。[平均- s]から[平均+ s]の間には、全体の約 68%が含まれることが分かります。



● 四分位偏差 (Quartile deviation)



第1四分位数と第3四分位数の間の区間には、データ全体のほぼ半分が入っており、データの中に極端に大きな値や小さな値があっても、影響を受けにくい。

一般に、データが中央値付近に集中しているほど、四分位範囲は小さくなり、データの散らばりの程度は小さいといえる。

例 3 123 ページの A 組のデータの四分位範囲を求める。

第1四分位数は 46, 第3四分位数は 80 であるから 四分位範囲は $80 - 46 = 34$

小										大									
26	31	35	39	44	48	52	55	57	63										
67	68	74	75	78	82	85	87	92	93										
平均が第1四分位数46										平均が第3四分位数80									

練習 11 123 ページの A 組と B 組のデータについて、次の問いに答えなさい。

- (1) B 組のデータの四分位範囲を求めなさい。
- (2) A 組と B 組の四分位範囲から、データの散らばりの程度が大きいのはどちらの組であると考えられるか答えなさい。

解答

小

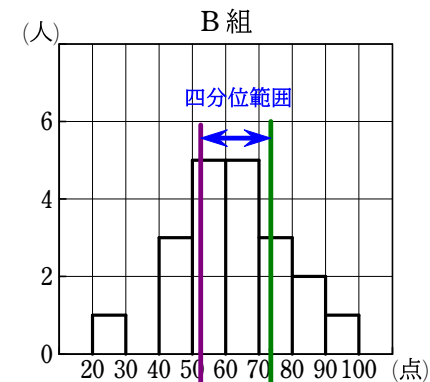
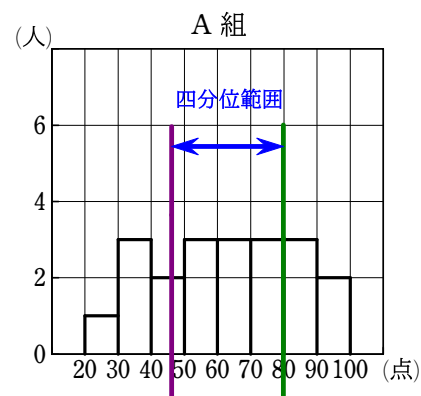
平均が第1四分位数52

27	42	47	49	51	53	54	56	59	64
66	68	69	69	72	76	79	82	85	94

平均が第3四分位数74

大

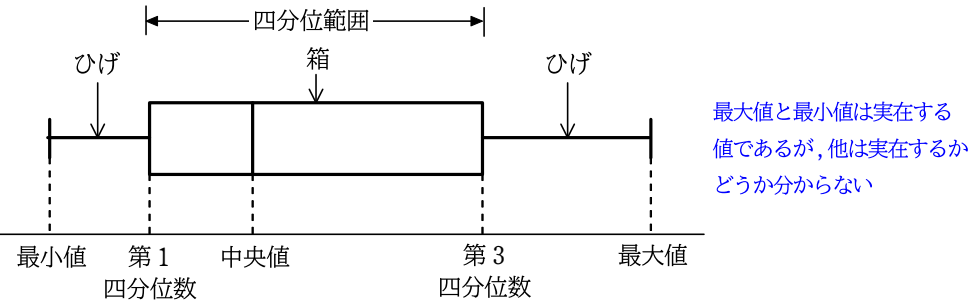
- (1) 第1四分位数は 52, 第3四分位数は 74 であるから 四分位範囲は $74 - 52 = 22$
- (2) A組のデータの四分位範囲は34, B組のデータの四分位範囲は22であるから A組のデータの四分位範囲の方が大きい。
よって、データの散らばりの程度が大きいのはA組



<箱ひげ図>

四分位数や四分位範囲を使って、データの分布を図で表してみよう。

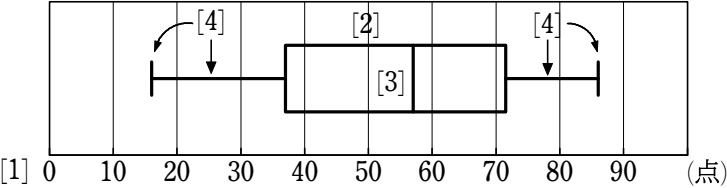
データの散らばりのようすを図で表すと、次の図のようになる。



上の図を **箱ひげ図** という。箱ひげ図は、データの最小値、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、箱とひげで表している。箱の横の長さは、四分位範囲を表す。

箱ひげ図は、次の手順でかくとよい。

- [1] 横軸にデータの目もりをとる。
- [2] 第1四分位数を左端、第3四分位数を右端とする長方形(箱)をかく。
- [3] 箱の中に中央値を示す縦線をひく。
- [4] 最小値、最大値を表す縦線をひき、箱の左端から最小値までと、箱の右端から最大値まで、線分(ひげ)をひく。



注意 箱ひげ図は、縦向きにかくこともある。

例 4 次のデータは、2019 年の東京と那覇における月ごとの平均気温を、気温の低い順に並べたものである。(単位は℃)

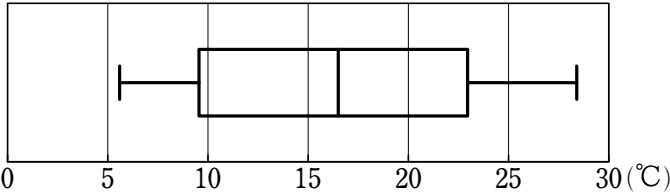
東京	5.6	7.2	8.5	10.6	13.1	13.6	19.4	20.0	21.8	24.1	25.1	28.4
那覇	18.1	19.9	20.0	20.0	22.3	23.1	24.2	26.0	26.5	28.0	28.9	29.2

(気象庁のホームページより)

2 つのデータについて、最大値、最小値、四分位数を表にまとめると、次のようになる。

	最小値	第1四分位数	中央値	第3四分位数	最大値
東京	5.6	9.6	16.5	23.0	28.4
那覇	18.1	20.0	23.7	27.3	29.2

よって、東京の箱ひげ図は、次のようになる。

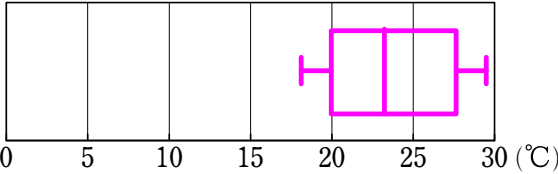


練習 12 例 4 について、次の問いに答えなさい。

- (1) 那覇の箱ひげ図を下の図にかき入れなさい。
- (2) 東京と那覇の箱ひげ図から、寒暖の差が大きいのはどちらであると考えられるか答えなさい。

解答

(1)



(2) 箱ひげ図が横に長いのは東京なので、寒暖の差が大きいのは東京

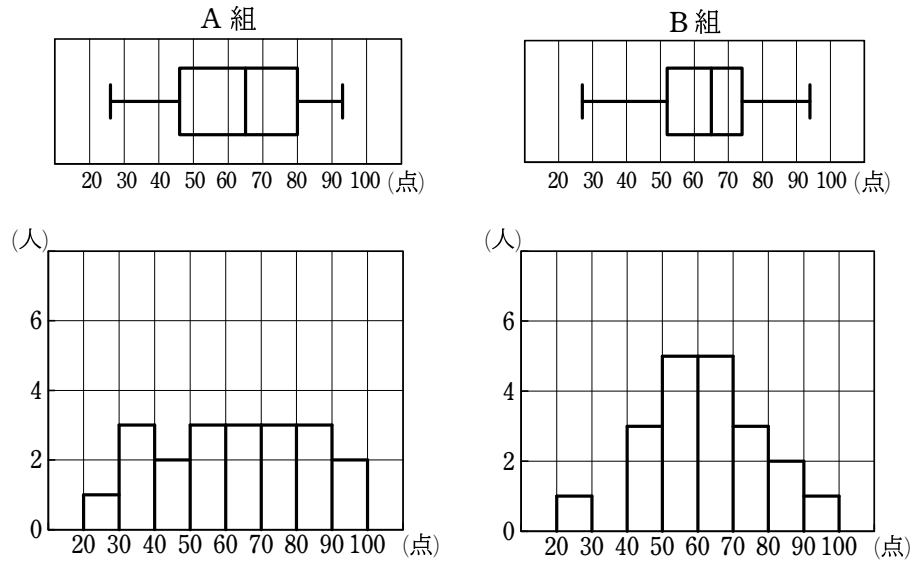
複数のデータの散らばりのようすを比べる場合は、箱ひげ図を利用するとよい。

箱ひげ図の「ひげ」は、最小値や最大値が他のデータの値と大きく離れている場合に、影響を受けて長くなる。

一方、「箱」はその影響を受けにくい。

箱ひげ図とヒストグラムの関係について、考えてみよう。

123 ページの A 組と B 組のデータについて、箱ひげ図とヒストグラムを比べると、次の図のようになる。



上の図から、ヒストグラムの山の位置と、箱ひげ図の箱の位置がだいたい対応し、ヒストグラムのすそにあたる部分が、箱ひげ図のひげに対応していることがわかる。ヒストグラムのすそが左に伸びていれば、箱ひげ図のひげも左に伸びる。

箱ひげ図では、ヒストグラムほどにはデータの散らばりのようすが表現されないが、大まかなようすを知ることができる。

探究

データの傾向と調査

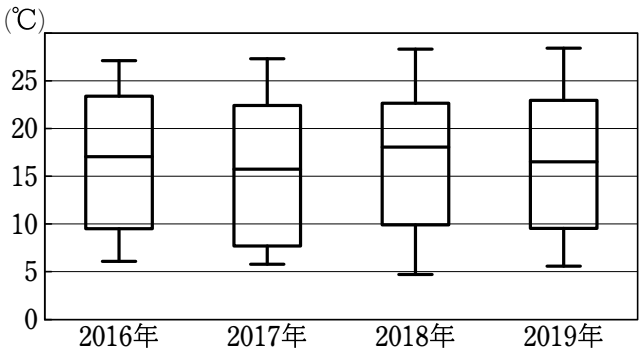
127 ページの例 4 では、2019 年の東京における月ごとの平均気温を調べて、箱ひげ図に表すことを学びました。

そこで、同じように、2016 年、2017 年、2018 年の月ごとの平均気温を調べて、東京の気温が以前と比べてどのように変化しているかを調査することになりました。

それぞれの年の月ごとの平均気温について、最大値、最小値、四分位数をまとめた表と、箱ひげ図をつくると次のようになりました。

	最小値	第 1 四分位数	中央値	第 2 四分位数	最大値
2016 年	6.1	9.5	17.1	23.4	27.1
2017 年	5.8	7.7	15.8	22.4	27.3
2018 年	4.7	9.9	18.1	22.7	28.3
2019 年	5.6	9.6	16.5	23.0	28.4

(気象庁のホームページより)



先生

今までに学んだことを思い出しながら、データの傾向について、気づいたことを話し合ってみましょう。また、そのように考えた理由も説明してみましょう。