

箱ひげ図

講師

湯浅 弘一



身近にあることは？

前回は中央値について学習しました。

例えば、5人の所持金が300円、300円、300円、300円、900円だとします。

この場合の中央値は、金額を小さい順に並べたときの真ん中の値ですから、300円です。

また、1人目から4人目までは300円となっていて、300円の人が一番多いことがわかります。

5人目の900円は、データ全体で見ると外れ値であることもわかりますね。

中央値の方が、平均値よりデータのばらつきがわかります。

ちなみにこの5人の平均金額は420円。

1人目から4人目までの300円が平均以下ということになります。

平均以上が1人だけ。変な感じがしますよね。

簡単にいえば、中央値を求めることでデータ全体の散らばり具合がわかり、平均値を求めることで、すべてのデータをたいらにしたときの値がわかるのです。



確認しましょう

問題1

次の5個のデータの中央値を求めなさい。

5, 6, 8, 10, 13

【考え方】

5個のデータですから、

真ん中のデータは小さい方から3番目であり、大きい方から3番目にあたります。

よって、中央値は8です。

問題2

次の6個のデータの中央値を求めなさい。

5, 6, 8, 10, 13, 19

【考え方】

6個のデータですから、真ん中のデータは存在しません。

5, 6, 8, | 10, 13, 19

上のように、真ん中にデータがないのです。

このような場合には・・・

5, 6, 8, | 10, 13, 19

中央を挟む2つの数の平均をとります。

この場合は

$$\frac{8+10}{2} = 9$$

となります。

よって、中央値は9です。

問題3

次の11個のデータの四分位数を求めなさい。

2, 3, 5, 8, 10, 11, 12, 14, 17, 20, 34

【考え方】

データを小さい順に並べたとき4等分する位置にくる値を四分位数といいます。

求め方は以下の通り！

- (1) 全体のデータの中央値を求め、これを第2四分位数とする。
- (2) (1)を除いて、小さい方を下位、大きい方を上位とする。
- (3) (2)の下位の中央値を第1四分位数、上位の中央値を第3四分位数とする。

この問題の場合は・・・

- (1) 第2四分位数（中央値）は11

2, 3, 5, 8, 10, 11, 12, 14, 17, 20, 34

- (2) 下位と上位に分ける

2, 3, 5, 8, 10, 11, 12, 14, 17, 20, 34
 下位 上位

- (3) 下位の中央値が第1四分位数なので5、上位の中央値が第3四分位数なので17

2, 3, 5, 8, 10, 11, 12, 14, 17, 20, 34
 下位 上位

問題 4

次の 11 個のデータの四分位数を求めなさい。

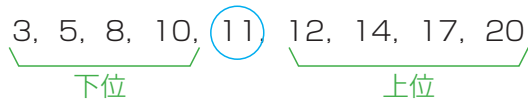
3, 5, 8, 10, 11, 12, 14, 17, 20

【考え方】

まず、3, 5, 8, 10, 11, 12, 14, 17, 20 の中央値を求めます。

9 個のデータの真ん中は小さい方から 5 番目、大きい方からも 5 番目ですから、11。

次に、上位と下位を考えます。



下位の中央値は、下位の真ん中にデータがないので

$$\frac{5+8}{2} = 6.5$$

6.5 が第 1 四分位数です。

上位の中央値は、上位の真ん中にデータがないので

$$\frac{14+17}{2} = 15.5$$

15.5 が第 3 四分位数です。

問題5

以下の9個のデータを箱ひげ図で表しなさい。

1, 4, 7, 8, 9, 13, 15, 20, 23

【考え方】

第2四分位数は9、

第1四分位数は $\frac{4+7}{2} = 5.5$

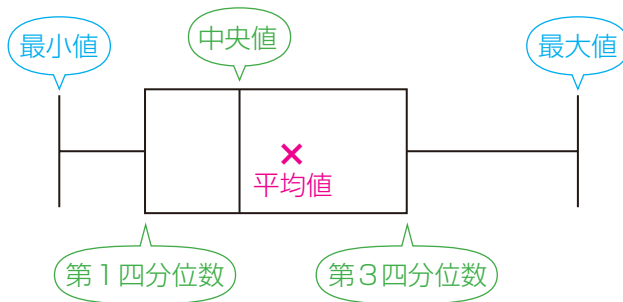
第3四分位数は $\frac{15+20}{2} = 17.5$

これを以下のようにして箱ひげ図で表しましょう。

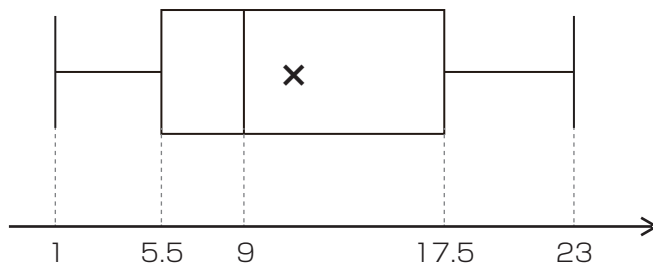
ひげの左端が最小値、右端を最大値にします。

そして、箱の左端が第1四分位数。箱の右端が第3四分位数。

さらに、中央値(第2四分位数)を入れ、最後に平均値を×で入れます。



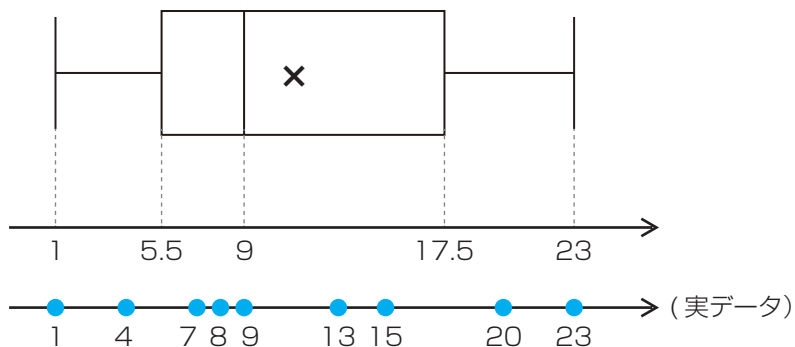
この問題の箱ひげ図は以下ようになります。



さて、箱ひげ図から何がわかるのでしょうか？

幅が広いほど1つ1つのデータの間が開いていて、幅が狭いほどデータが集まっているということがわかります。

先ほどの問題で、箱ひげ図と実際のデータを比較してみましょう。



(参考)

以下は箱ひげ図とヒストグラムの対応を表しています。

ヒストグラムのようにデータの分布を詳しく表せませんが、箱ひげ図からもおよその分布を知ることができます。

