

■データ分析■（大問 6）：20 点分

ここの（1）（2）（4）の 3 問、15 点を確実に押さえておきましょう。

計算力に余裕のある人は（3）も。偏差とかで計算のややこしさはありますが、理解していると楽だし公式が書かれていることが多いので、地道に解くならさらに+5 点の可能性もあります。

そもそも「データ分析」とは。

集団の傾向を見るためのもの。身近なところではある試験の成績データから順位とか平均点とか偏差値とか合格判定とか出すとか、全小中学生の身長とか体重とかのデータを取っておけば全体の傾向とか変化とかもわかります（そして最近では栄養が足りてないから平均身長が低くなってきているなんてことも言われたりしています）。

その他にも、全世帯の年収とかのデータを取って分析すれば貧困状態にある家庭がどれくらいの数あるとか、国全体が豊かになっているのか貧しくなっているのか分かります（平均値と中央値の違いなどが分かっていたらデータを読み間違えてしまいます）。2020 年にもあった 5 年に 1 度の国勢調査とかも大きなデータ収集の動きですね。

Step 1：基本用語を抑えよう 第 6 問 (1)

● 1-1：言葉通りの意味で分かる 4 語

「**最大値**」：書かれたデータの中で一番数字が大きい値。漢字通りの意味

「**最小値**」：書かれたデータの中で一番数字が小さい値。漢字通りの意味

「**範囲**」：書かれたデータの数値の幅。「最大値」－「最小値」で求める。

「**最頻値**」：書かれたデータの数値で、一番多く登場している数値。漢字通りの意味

<例で確かめ> 次の 8 つのデータで確認

19 23 3 44 5 18 41 23

- 最小値は _____ ,
- 最大値は _____ ,
- 範囲は _____ 。
- 最頻値は _____

手順1：わかりやすいように小さいもの順に並び替える。

3 5 18 19 23 23 41 44

- ・最小値は一番左の 3 , ・最大値は一番右の 44 ,
- ・範囲は最大値(44)から最小値(3)を引く $44 - 3 =$ 41
- ・最頻値は登場回数2回が最高なので 23 (他は1回ずつの登場)

● 1-2：少し整理や計算が要る2語

「平均」：書かれたデータの数値を均一に平らにした時の数値。書かれたデータの数値をすべて足して、足したデータの数で割って求める。

「中央値」：書かれたデータを順番に並べた時に、ちょうど真ん中になる数値。

<例で確かめ> 引き続き、さっきの小さいもの順に並べ替えたデータで。

3 5 18 19 23 23 41 44

平均：平均はデータをすべて足して ($3+5+18+19+23+23+41+44=176$)
データの数(8つ)で割ると ($176\div 8=22$) 22

中央値：並べ替えたデータのちょうど真ん中。外側から順番に探っていく(※がど真ん中)

3 **5** **18** **19** ※ **23** **23** **41** **44**

19と23のちょうど間になるので、19と23の平均を出す

$$19+23=42 \quad 42\div 2=21$$

なので、中央値は 21

※ 中央値は、データの数(奇数個か偶数個か)によって出し方に少し注意が必要です。
(あとに出てくる「第1四分位数」「第3四分位数」の出し方も同じ考え方になります)

【高認 類題 1-1】基本用語がわかれば解ける問題

次のデータは、小学生10人に、筆箱に入っている鉛筆の本数を聞いたものである。

4, 6, 4, 8, 1, 5, 4, 2, 5, 3 (本)

このデータの

最小値は (本), 最大値は (本), 範囲は , 最頻値は (本)

中央値は (本), 平均値は (本)

—————〔解き方と解答（高認 類題 1-1）〕—————

データを小さい順に並び替える。

1 2 3 4 4 4 5 5 6 8

最小値は 1（本），最大値は 8（本），範囲は 7：最大値8－最小値1，

最頻値は 4（本）：4が3回、5が2回、他は1回，

中央値は 4（本）：4と4の間なので4，

平均値は 4.2（本）：10つの合計が42。42÷10=4.2

【高認 類題 1-2】グラフからデータを読まないといけない問題

次の図は、10人のバスケットボール選手が5本のフリースローを入れた本数をヒストグラムに表したものである。

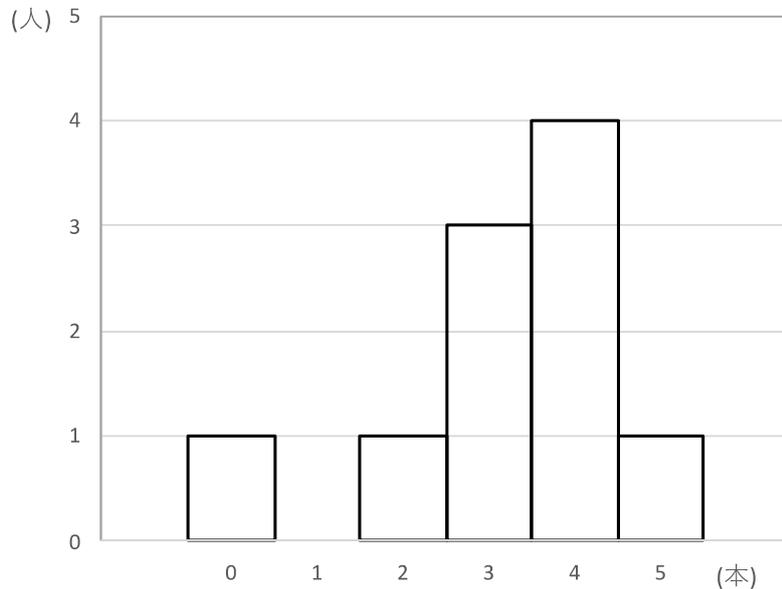


図 1. フリースローの本数と人数

このデータの 最小値は _____（本），最大値は _____（本），範囲は _____，最頻値は _____（本）
中央値は _____（本），平均値は _____（人）

—————〔解き方と解答（高認 類題 1-2）〕—————

グラフの縦のメモリが（人）、横のメモリが（本）となっているので、

横のメモリの0(本)のところは1(人)、1(本)のところは0(人)、2(本)のところは1(人)、3(本)が3(人)、4(本)が4(人)、5(本)が1(人)ということが読み取れます。

なので、それを 10 人分のシュート成功本数のデータを並べると、↓ になります。

0 2 3 3 3 4 4 4 4 5

※ ちなみに、それを表にするとこんな感じ（参考）表 1.フリースローの本数と人数

本数	0	1	2	3	4	5
人数	1	0	1	3	4	1

表のパターンも慣れておくといいかもです。

ここまでできると【高認 類題 1-1】と同じように解けますね。

なので、答えだけ記しときますね。

最小値は 0（本），最大値は 5（本），範囲は 5，最頻値は 4（本）

中央値は 3.5（本），平均値は 3.2（人）

※ ヒストグラムの方は、見た目で「平均」のイメージが掴みやすいと思います。

Step2：箱ひげ図 第6問 (2)

● 2-1：箱ひげ図

箱ひげ図とは、「最大値・最小値・中央値・四分位数」がひと目でわかる ↓ のようなグラフ。

Step1 で確認した用語から、最大値、中央値、最小値をグラフで示すところら。

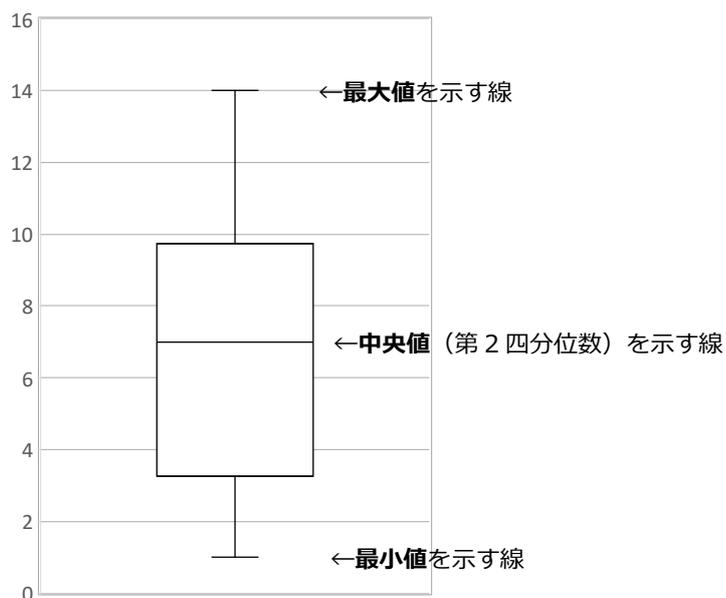
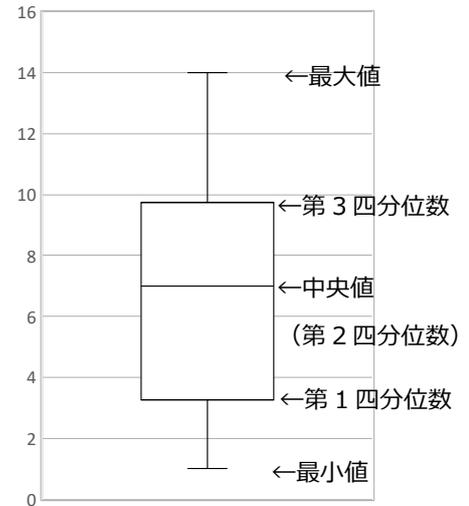


図 2. 箱ひげ図の例

● 2-2：箱ひげ図の見方（箱の両端からひげのようなものが出ている図）

- ① ひげの両端は「**最小値**」と「**最大値**」
- ② 箱の真ん中の線は「**中央値**」。またの名を「**第2四分位数**」
- ③ 箱の両端は、下が「**第1四分位数**」、上が「**第3四分位数**」。
「四分位数」：全体数を4つに分けるための区切り。
4つに分けるので、区切りは3つ必要ということになる。
(それぞれ「**第1四分位数**」「**中央値**」「**第3四分位数**」)



「中央値」は全体を2つに分けるため（半分）の区切りなので、「二分位数」と言ってもいいかもしれません。

ともあれ、「中央値」で2分の1で区切って、
下の2分の1をさらに半分で区切ったものが「第1四分位数」、
上の2分の1をさらに半分で区切ったものが「第3四分位数」

なので、第1四分位数も第3四分位数も「中央値の中央値」といえるので、中央値の出し方をマスターしていれば四分位数も出せます。

<例で確かめ> さっきの小さいもの順に並べ替えたデータで四分位数を出してみよう。

6 1 4 5 8 4 2 3 5 4

手順1：わかりやすいように小さいもの順に並び替え

1 2 3 4 4 4 5 5 6 8

- 最小値は 1 , • 最大値は 8 ,
- 中央値は ※ の位置なので 4
1 2 3 4 4 ※ 4 5 5 6 8
- 第1四分位数は、下半分のさらに中央値なので 3
1 2 3 4 4
- 第3四分位数は、上半分のさらに中央値なので 5
4 5 **5** 6 8

※ 高認の1ポイント：箱ひげ図は、全体を並べて区切っているだけなので、平均値は読み取れない。