

第 38 講 データと代表値

学習目標

- ・ データと代表値の意味を説明できる。
- ・ データから平均値・中央値・最頻値を求められる。
- ・ データの個数が奇数の場合と偶数の場合の中央値を正しく求められる。
- ・ データの特徴や目的に応じて、平均値・中央値・最頻値のどれを見たとよいか判断できる。

データを 1 つの値で表す

たくさんの値を見やすくする

あるクラスの小テストの点数や、1 週間の気温のように、調べて得られた複数の値を**データ**といいます。

たとえば、5 人の小テストの点数が

6, 7, 7, 8, 12

だったとします。

データをそのまま並べれば詳しい情報は分かりますが、人数が増えると全体の特徴をつかみにくくなります。

そこで、データ全体を代表する値を使って、データの中心的な様子を表します。

このような値を**代表値**といいます。

代表値

データ全体の特徴を、1 つの数値で代表して表す値を**代表値**といいます。

代表的なものに、

- ・ 平均値
- ・ 中央値
- ・ 最頻値

があります。

平均値

すべての値を使って求める

平均値は、すべてのデータの値を足し、その個数で割って求めます。

データが

6, 7, 7, 8, 12

なら、合計は 40、個数は 5 です。

したがって、平均値は

$$\frac{40}{5} = 8$$

です。

平均値

データの値の合計を、データの個数で割った値を**平均値**といいます。
データが $x_1, x_2, \dots, x_n$ の n 個なら、平均値は

$$\frac{x_1 + x_2 + \dots + x_n}{n}$$

です。

平均値の特徴

平均値は、すべてのデータの値を使って計算します。
そのため、データ全体を反映しやすい一方、他の値から大きく離れた値があると、その影響を受けやすいという特徴があります。

中央値

小さい順に並べて中央を見る

中央値を求めるときは、まずデータを小さい順に並べます。
たとえば、

3, 8, 4, 10, 5

を小さい順にすると、

3, 4, 5, 8, 10

です。
データは 5 個なので、中央にある 3 番目の値 5 が中央値です。

中央値

データを大きさの順に並べたとき、中央に位置する値を**中央値**といいます。
データの個数が奇数なら、ちょうど中央の値を使います。
データの個数が偶数なら、中央に並ぶ 2 つの値の平均を中央値とします。

データが偶数個の場合

データ

2, 5, 7, 8, 11, 15

は 6 個あります。
中央の 2 つは 7 と 8 なので、中央値は

$$\frac{7 + 8}{2} = \frac{15}{2}$$

です。

中央値は並べ替えてから求める

中央値は、与えられた順番の中央を見るものではありません。
必ずデータを大きさの順に並べてから中央を確認します。

最頻値

最も多く現れる値を見る

データ

2, 3, 3, 4, 4, 4, 7

では、4 が 3 回現れ、最も多くなっています。
したがって、このデータの最頻値は 4 です。

最頻値

データの中で、最も多く現れる値を**最頻値**といいます。

最頻値の特徴

最頻値は、データの中で最もよく現れる値を表します。
たとえば、最もよく売れるサイズや、最も多い回答を知りたいときに役立ちます。

最頻値が 1 つとは限らない

最も多く現れる回数が同じ値が複数ある場合、最頻値が複数になることがあります。
たとえば、

1, 1, 2, 2, 3

では、1 と 2 がともに 2 回現れます。
この場合、最頻値は 1 と 2 です。

平均値・中央値・最頻値を比べる

代表値は目的によって使い分ける

次のデータを考えます。

4, 5, 5, 6, 30

平均値は、

$$\frac{4 + 5 + 5 + 6 + 30}{5} = 10$$

です。

中央値は 5、最頻値も 5 です。

このデータでは、30 が他の値から大きく離れています。そのため、平均値は 30 の影響を受けて 10 まで大きくなっています。

一方、中央値は並べたときの中央を見るので、離れた値の影響を平均値ほど強く受けません。

このように、どの代表値がデータの特徴を表しやすいかは、データの様子や知りたいことによって変わります。

代表値の使い分け

- ・ **平均値**：すべての値を使って全体をまとめたいときに役立つ。
- ・ **中央値**：極端に大きい値や小さい値があり、中央付近の様子を見たいときに役立つ。
- ・ **最頻値**：最もよく現れる値を知りたいときに役立つ。

代表値は、1つだけを機械的に使うのではなく、その特徴を理解して選びます。

代表値だけでは元のデータすべては分からない

代表値が同じでも、元のデータの並び方や特徴が同じとは限りません。

代表値はデータを簡潔に見るための値であり、データのすべての情報を表すものではありません。

例題

例題 1

次のデータの平均値、中央値、最頻値を求めなさい。

4, 6, 6, 7, 12

例題 1 の解答・解説

合計は 35、データは 5 個なので、平均値は

$$\frac{35}{5} = 7$$

です。

すでに小さい順に並んでおり、中央の 3 番目は 6 なので、中央値は 6 です。

最も多く現れる値は 6 なので、最頻値も 6 です。

例題 2

次のデータの平均値と中央値を求めなさい。

3, 5, 8, 8, 10, 14

例題 2 の解答・解説

合計は 48、データは 6 個なので、平均値は

$$\frac{48}{6} = 8$$

です。

データは 6 個なので、中央の 2 つである 8 と 8 の平均をとります。

したがって、中央値は 8 です。

例題 3

5 人の待ち時間が次のようになりました。

2, 3, 3, 4, 28

平均値と中央値を求めなさい。また、このデータの「ふつうの待ち時間」を表す値として平均値と中央値のどちらを見るとよいか、その理由を説明しなさい。

例題 3 の解答・解説

合計は 40 なので、平均値は

$$\frac{40}{5} = 8$$

です。

中央の 3 番目の値は 3 なので、中央値は 3 です。

28 は他の値から大きく離れており、平均値を大きくしています。

このデータの中央付近の様子を見る目的なら、中央値 3 の方が特徴を捉えやすいと考えられます。

演習問題**練習問題 1**

次のデータの平均値を求めなさい。

5, 7, 8, 9, 11

練習問題 2

次のデータの中央値を求めなさい。

9, 3, 6, 5, 12

練習問題 3

次のデータの中央値を求めなさい。

2, 4, 5, 7, 9, 13

練習問題 4

次のデータの最頻値を求めなさい。

1, 2, 2, 3, 3, 3, 5, 6

練習問題 5

次のデータの平均値、中央値、最頻値をそれぞれ求めなさい。

4, 5, 5, 6, 10

練習問題 6

次のデータの平均値、中央値、最頻値をそれぞれ求めなさい。

2, 2, 4, 4, 6, 6

練習問題 7

ある 6 人の 1 日の読書時間は、次のようになりました。単位は分です。

10, 12, 13, 15, 16, 90

- (1) 平均値を求めなさい。
- (2) 中央値を求めなさい。
- (3) この 6 人の読書時間の中央付近の様子を表すには、平均値と中央値のどちらが適していると考えられるか、理由とともに答えなさい。

練習問題 8

ある店で販売された T シャツのサイズは、次のようになりました。

M, L, M, S, M, L, M

最も多く売れたサイズを知りたいとき、平均値・中央値・最頻値のうち、どの代表値の考え方が適しているか答えなさい。また、実際にその値を答えなさい。

演習問題の解答・解説

練習問題 1 の解答・解説

合計は 40、データは 5 個です。

したがって、平均値は

$$\frac{40}{5} = 8$$

です。

練習問題 2 の解答・解説

小さい順に並べると、

3, 5, 6, 9, 12

です。

中央の 3 番目の値は 6 なので、中央値は 6 です。

練習問題 3 の解答・解説

データは 6 個なので、中央の 2 つは 5 と 7 です。
したがって、中央値は

$$\frac{5+7}{2} = 6$$

です。

練習問題 4 の解答・解説

3 が 3 回現れ、最も多くなっています。
したがって、最頻値は 3 です。

練習問題 5 の解答・解説

合計は 30、データは 5 個なので、平均値は 6 です。
中央の 3 番目の値は 5 なので、中央値は 5 です。
最も多く現れる値も 5 なので、最頻値は 5 です。

練習問題 6 の解答・解説

合計は 24、データは 6 個なので、平均値は 4 です。
中央の 2 つは 4 と 4 なので、中央値は 4 です。
2, 4, 6 はそれぞれ 2 回ずつ現れます。
したがって、最頻値は 2, 4, 6 です。

練習問題 7 の解答・解説

(1) 合計は 156 です。
したがって、平均値は

$$\frac{156}{6} = 26$$

分です。

(2) 中央の 2 つは 13 と 15 なので、中央値は

$$\frac{13+15}{2} = 14$$

分です。

(3) 90 分という他の値から大きく離れた値があり、平均値はその影響を受けています。
中央付近の様子を表す目的なら、中央値 14 分の方が適していると考えられます。

練習問題 8 の解答・解説

最も多く売れたサイズを知りたいので、最も多く現れる値を見る最頻値の考え方が適しています。
 M が 4 回で最も多く現れています。
したがって、最頻値は M です。