

第 39 講 分散

学習目標

- データの値と平均値の差である偏差を求められる。
- 偏差を 2 乗する理由を理解できる。
- 偏差の 2 乗の平均として分散を求められる。
- 平均値が同じデータでも、分散によって散らばり方の違いを捉えられる。

平均値だけでは散らばり方は分らない

同じ平均値でもデータの様子は違う

次の 2 つのデータを比べます。

データ A :

4, 5, 6

データ B :

1, 5, 9

どちらも平均値は 5 です。

しかし、データ A は 5 の近くに集まり、データ B は 5 から大きく離れた値を含んでいます。

平均値だけでは、このようなデータの散らばり方の違いまでは表せません。

そこで、各データが平均値からどれくらい離れているかに注目します。

偏差とは平均との差

偏差

各データの値から平均値を引いた値を**偏差**といいます。

データの値を x 、平均値を $\bar{x}$ とすると、偏差は

$$x - \bar{x}$$

です。

偏差を実際に求める

データ

2, 4, 6

の平均値は 4 です。

したがって、各データの偏差は、

- $2 - 4 = -2$
- $4 - 4 = 0$
- $6 - 4 = 2$

です。

平均値より小さい値の偏差は負、平均値より大きい値の偏差は正になります。

偏差の符号

偏差 $x - \bar{x}$ は、

- データが平均値より小さいとき負
- データが平均値と等しいとき 0
- データが平均値より大きいとき正

になります。

偏差の絶対値が大きいほど、そのデータは平均値から遠く離れています。

偏差をそのまま平均してはいけない

偏差の和は 0 になる

先ほどのデータ

$$2, 4, 6$$

の偏差は、

$$-2, 0, 2$$

でした。

これらを足すと、

$$-2 + 0 + 2 = 0$$

です。

実は、どのようなデータでも偏差の和は 0 になります。

正の偏差と負の偏差が打ち消し合うため、偏差をそのまま平均しても散らばりの大きさを表すことができません。

偏差の平均は散らばりを表さない

「平均からどれだけ離れているか」を調べたいからといって、偏差をそのまま足して平均することとはできません。

偏差には正と負があり、互いに打ち消し合うからです。

そこで、偏差を 2 乗してから平均します。

偏差を 2 乗する

2 乗すると負の値も正になる

偏差が

$$-2, 0, 2$$

なら、その 2 乗は

$$4, 0, 4$$

です。

2 乗することで、平均値より上に離れていても下に離れていても、どちらも正の大きさとして扱えます。

また、平均値から大きく離れた値ほど、偏差の 2 乗も大きくなります。

この「偏差の 2 乗」を平均したものが分散です。

分散

データの各値について偏差を求め、その 2 乗を平均した値を**分散**といいます。

データが $x_1, x_2, \dots, x_n$ 、平均値が $\bar{x}$ のとき、分散は

$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

です。

分散を求める基本手順

分散は次の順に求めます。

1. 平均値を求める。
2. 各データから平均値を引いて偏差を求める。
3. 各偏差を 2 乗する。
4. 偏差の 2 乗をすべて足す。
5. データの個数で割る。

表にすると計算しやすい

平均値・偏差・偏差の 2 乗を整理する

データ

2, 4, 6

の分散を求めます。

平均値は 4 です。

データ x	偏差 $x - 4$	偏差の 2 乗 $(x - 4)^2$
2	-2	4
4	0	0
6	2	4

偏差の 2 乗の合計は 8 です。

データは 3 個なので、分散は

$$\frac{8}{3}$$

です。

最後に割るのはデータの個数

分散は、偏差の 2 乗の合計ではありません。

偏差の 2 乗をすべて足したあと、データの個数で割って平均します。

たとえば偏差の 2 乗の合計が 20、データが 5 個なら、分散は 20 ではなく 4 です。

分散の大きさが表すこと

平均値からの散らばりを比べる

次の 2 つのデータをもう一度比べます。

データ A :

4, 5, 6

データ B :

1, 5, 9

どちらも平均値は 5 です。

データ A の偏差は $-1, 0, 1$ なので、偏差の 2 乗は $1, 0, 1$ です。

したがって、分散は

$$\frac{1 + 0 + 1}{3} = \frac{2}{3}$$

です。

データ B の偏差は $-4, 0, 4$ なので、偏差の 2 乗は $16, 0, 16$ です。

したがって、分散は

$$\frac{16 + 0 + 16}{3} = \frac{32}{3}$$

です。

データ B の方が分散が大きく、平均値から大きく散らばっていることが数値にも表れています。

分散と散らばり

分散は 0 以上の値になります。

一般に、

- ・ 分散が小さいほど、データは平均値の近くに集まっている。
 - ・ 分散が大きいほど、データは平均値から広く散らばっている。
- と考えられます。

すべて同じ値なら分散は 0

データが

7, 7, 7, 7

なら、平均値は 7 です。

すべての偏差が 0 なので、偏差の 2 乗もすべて 0 です。

したがって、分散は 0 です。

データがまったく散らばっていないことを表しています。

例題

例題 1

次のデータの偏差と分散を求めなさい。

3, 5, 7

例題 1 の解答・解説

平均値は

$$\frac{3 + 5 + 7}{3} = 5$$

です。

偏差は順に $-2, 0, 2$ です。

偏差の 2 乗は $4, 0, 4$ なので、分散は

$$\frac{4 + 0 + 4}{3} = \frac{8}{3}$$

です。

例題 2

次のデータの分散を求めなさい。

2, 4, 4, 6

例題 2 の解答・解説

平均値は

$$\frac{2 + 4 + 4 + 6}{4} = 4$$

です。

偏差は $-2, 0, 0, 2$ です。

偏差の 2 乗は $4, 0, 0, 4$ なので、分散は

$$\frac{4 + 0 + 0 + 4}{4} = 2$$

です。

例題 3

次の 2 つのデータについて、それぞれ分散を求め、どちらの方が平均値のまわりに大きく散らばっているか答えなさい。

データ A:

8, 10, 12

データ B:

4, 10, 16

例題 3 の解答・解説

どちらも平均値は 10 です。

データ A の偏差は $-2, 0, 2$ なので、分散は

$$\frac{4 + 0 + 4}{3} = \frac{8}{3}$$

です。

データ B の偏差は $-6, 0, 6$ なので、分散は

$$\frac{36 + 0 + 36}{3} = 24$$

です。

したがって、分散が大きいデータ B の方が、平均値のまわりに大きく散らばっています。

演習問題

練習問題 1

次のデータについて、平均値と各データの偏差を求めなさい。

1, 3, 5

練習問題 2

次のデータの分散を求めなさい。

1, 3, 5

練習問題 3

次のデータの分散を求めなさい。

4, 4, 6, 6

練習問題 4

次のデータの分散を求めなさい。

2, 2, 2, 2, 2

練習問題 5

次のデータの分散を求めなさい。

1, 2, 3, 4, 5

練習問題 6

次の 2 つのデータは、どちらも平均値が 6 です。

データ A :

$$5, 6, 7$$

データ B :

$$2, 6, 10$$

それぞれの分散を求め、どちらの方が散らばりが大きいかわか答えなさい。

練習問題 7

ある 4 個のデータの平均値は 5 で、各データの偏差は

$$-3, -1, 1, 3$$

です。

(1) 偏差の和を求めなさい。

(2) 分散を求めなさい。

練習問題 8

ある生徒は、データ

$$2, 4, 6$$

について、平均値が 4 なので偏差を $-2, 0, 2$ と求め、その平均

$$\frac{-2 + 0 + 2}{3} = 0$$

を「分散」としました。

この考えの誤りを説明し、正しい分散を求めなさい。

演習問題の解答・解説

練習問題 1 の解答・解説

平均値は

$$\frac{1 + 3 + 5}{3} = 3$$

です。

したがって、偏差は順に

$$-2, 0, 2$$

です。

練習問題 2 の解答・解説

平均値は 3 なので、偏差は $-2, 0, 2$ です。

偏差の 2 乗は $4, 0, 4$ です。

したがって、分散は

$$\frac{4 + 0 + 4}{3} = \frac{8}{3}$$

です。

練習問題 3 の解答・解説

平均値は

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

です。

偏差は $-1, -1, 1, 1$ なので、偏差の 2 乗はすべて 1 です。

したがって、分散は

$$\frac{1 + 1 + 1 + 1}{4} = 1$$

です。

練習問題 4 の解答・解説

平均値は 2 です。

すべてのデータが平均値と等しいので、偏差はすべて 0 です。

したがって、分散は 0 です。

練習問題 5 の解答・解説

平均値は

$$\frac{1 + 2 + 3 + 4 + 5}{5} = 3$$

です。

偏差は $-2, -1, 0, 1, 2$ です。

偏差の 2 乗は $4, 1, 0, 1, 4$ なので、分散は

$$\frac{4 + 1 + 0 + 1 + 4}{5} = 2$$

です。

練習問題 6 の解答・解説

データ A の偏差は $-1, 0, 1$ です。

したがって、分散は

$$\frac{1 + 0 + 1}{3} = \frac{2}{3}$$

です。

データ B の偏差は $-4, 0, 4$ です。

したがって、分散は

$$\frac{16 + 0 + 16}{3} = \frac{32}{3}$$

です。

よって、データ B の方が分散が大きく、散らばりも大きいと分かります。

練習問題 7 の解答・解説

(1) 偏差の和は

$$-3 - 1 + 1 + 3 = 0$$

です。

(2) 偏差の 2 乗は $9, 1, 1, 9$ です。

したがって、分散は

$$\frac{9 + 1 + 1 + 9}{4} = 5$$

です。

練習問題 8 の解答・解説

分散は偏差そのものの平均ではなく、**偏差の 2 乗の平均**です。

偏差をそのまま足すと、正の偏差と負の偏差が打ち消し合ってしまいます。

このデータの偏差は $-2, 0, 2$ なので、偏差の 2 乗は $4, 0, 4$ です。

したがって、正しい分散は

$$\frac{4 + 0 + 4}{3} = \frac{8}{3}$$

です。