# データの分析【分散と標準偏差】 p.187～189

青チャートⅠ：例題182～184

**【態度目標】しゃべる、質問する、説明する、動く、協力する、貢献する**

**【内容目標】散らばりの度合いを表す値を求められるようになろう**

## □分散と標準偏差

データの中の**いくつかの代表的な値**を用いて散らばりの度合いを表す値 ⇒ 四分位数

データの中の **すべての値** を用いて散らばりの度合いを表す値 ⇒ ？

◎ 次のことを用いて、平均値の周りに各値がどのように分布しているかを考えてみよう

**偏差**　データの各値と平均値 $\overline{x}$ との差のこと。$x-\overline{x}$ で表す。

偏差の総和は 0 であるので、偏差の平均ももちろん 0 。

**分散** (variance)　**偏差 $x-\overline{x}$ の 2 乗の平均値**のこと。

$V$，$\sigma^2$ などを用いることもある

変量 $x$ のデータの値が $x_1$，$x_2$，……，$x_n$ で，その平均値が $\overline{x}$ のとき

式で表すと　$s^2=\dfrac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}$

変量 $x$ の測量単位の２乗になってしまう

↓

平方根を取ることで変量 $x$ の測量単位と同じになる

**標準偏差** (standard deviation)

分散の正の平方根のこと。$s$ で表す。要するに $s=\sqrt{\text{分散}}$ 。

$$s=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}}$$

標準偏差が小さくなるほどデータは平均値の周りに集中しており、散らばりの度合いが小さくなる。逆に標準偏差が大きくなれば散らばりの度合いが大きいといえる（分散も同様である）。

＜注意＞ 標準偏差の単位は，変量 $x$ の単位と同じである。

**例８）** 10 人の生徒の漢字テストの得点 $x$ (点) が，下の表で与えられている。ただし，平均値 $\overline{x}$ は，

$\overline{x}=\dfrac{1}{10}\times 70=7$ (点) である。

| | A | B | C | D | E | F | G | H | I | J | 合計 | 平均 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 9 | 3 | 4 | 10 | 10 | 5 | 7 | 9 | 10 | 3 | 70 | 7 |
| $x-\overline{x}$ | 2 | $-4$ | $-3$ | 3 | 3 | $-2$ | 0 | 2 | 3 | $-4$ | 0 | |
| $(x-\overline{x})^2$ | 4 | 16 | 9 | 9 | 9 | 4 | 0 | 4 | 9 | 16 | 80 | 8 |

偏差は平均との差を計算

偏差を２乗

「偏差の２乗の平均」が分散

分散は　$s^2=\dfrac{1}{10}\times 80=8$，標準偏差は　$s=\sqrt{8}\fallingdotseq 2.8$ (点) 終

「$\sqrt{\text{分散}}$」が標準偏差

「元データが整数値なので小数第１位までにした

平均を求めるときに仮平均を使ってもよい

数式で書くと（問題集などでは下のように書かれることが多い）

$$s^2=\frac{1}{10}\{(9-7)^2+(3-7)^2+(4-7)^2+(10-7)^2+(10-7)^2+(5-7)^2+(7-7)^2+(9-7)^2+(10-7)^2+(3-7)^2\}$$

$$=\frac{1}{10}\times 80=8$$

参考 平均偏差 ～ 偏差の２乗ではなく，偏差の絶対値を用いて平均値を計算して散らばりを考える方法もある。

$$\frac{1}{n}\{|x_1-\overline{x}|+|x_2-\overline{x}|+\cdots\cdots+|x_n-\overline{x}|\}\quad\text{ただし}\quad \overline{x}=\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)$$

直感的ではあるが，絶対値を外す際に場合分けが必要になるなど扱いにくさもあるのであまり用いられない。

# データの分析【分散と標準偏差】 p.187～189

青チャートⅠ：例題182～184

分散 $s^2$ の式は，次のように変形される。

$$s^2=\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}$$

$$=\frac{1}{n}\{(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)-2\overline{x}(x_1+x_2+\cdots\cdots+x_n)+n(\overline{x})^2\}$$

$$=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)-2\overline{x}\cdot\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)+(\overline{x})^2$$

$$=\overline{x^2}-2\overline{x}\cdot\overline{x}+(\overline{x})^2$$

$$=\overline{x^2}-(\overline{x})^2$$

ただし，$\overline{x^2}$ は，$x^2$ のデータ $x_1{}^2$，$x_2{}^2$，……，$x_n{}^2$ の平均値を表す。

**（$x$ のデータの分散）＝（$x^2$ のデータの平均値）－（$x$のデータの平均値）$^2$**

つまり，データの値の『**２乗の平均値**』**から，**『**平均値の２乗**』**を引いた値が分散に等しい。**

ただし，データの値が小さくなければ大変なので注意。

<注意> 変量 $x^2$ のデータの平均値を $\overline{x^2}$ と書くと，上の式は $s^2=\overline{x^2}-(\overline{x})^2$ と書ける。

**例題１）** テスト A の得点を $x$（点）とする。 変量 $x$ のデータの標準偏差 $s_x$ を求めよ。

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| テスト A | 1 | 1 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 9 | 9 | 10 |
| テスト B | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 7 | 7 |
| テスト C | 1 | 2 | 3 | 3 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 7 | 7 | 8 | 10 |

解答 $x$ のデータの分散を $s_x{}^2$ とする。

$$\overline{x}=\frac{1}{20}(1+1+2+\cdots\cdots+10)=\frac{100}{20}=5$$

平均

$$\overline{x^2}=\frac{1}{20}(1^2+1^2+2^2+\cdots\cdots+10^2)=\frac{640}{20}=32$$

２乗の平均

よって $s_x{}^2=\overline{x^2}-(\overline{x})^2=32-5^2=7$

２乗の平均－平均の２乗

ゆえに $s_x=\sqrt{s_x{}^2}=\sqrt{7}\fallingdotseq 2.6$（点）

補足 $s^2=\dfrac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}$

を用いて，$x$ のデータの分散 $s_x{}^2$ を求めると，次のようになる。

$$s_x{}^2=\frac{1}{20}\{(1-5)^2+(1-5)^2+(2-5)^2+\cdots\cdots+(10-5)^2\}=\frac{140}{20}=7$$

深める データの散らばりの度合いを表す四分位範囲，分散の値のうち，外れ値の影響を受けやすいのはどちらか，説明してみよう。

# データの分析【分散と標準偏差】 p.187〜189

青チャートⅠ：例題182〜184

## □分散と平均値の関係式

**例題）** 15 個の値からなるデータがあり，そのうちの 10 個の値の平均値は 9，分散は 3，
残り 5 個の値の平均値は 6，分散は 9 である。

(1) このデータの平均値を求めよ。

(2) このデータの分散を求めよ。

| | 10 個のデータ | 5 個のデータ |
|---|---|---|
| 平均値 | 9 | 6 |
| 総和 | $9\times10$ | $6\times5$ |
| 15個のデータの総和 | $9\times10+6\times5=120$ | |
| 15個のデータの平均値 | $\frac{120}{15}=8$ | |
| 2乗の平均値 | $a$ | $b$ |
| 2乗の総和 | $a\times10$ | $b\times5$ |
| 15個のデータでの総和 | $a\times10+b\times5$ | |

(1) $\dfrac{9\times10+6\times5}{15}=8$

(2) 10 個の値の 2 乗の平均値を $a$ とすると

$a-9^2=3$

すなわち　$a=84$

残りの 5 個の値の 2 乗の平均値を $b$ とすると

$b-6^2=9$

すなわち　$b=45$

よって，15 個の値の 2 乗の和は

$a\times10+b\times5=84\times10+45\times5=1065$

したがって，15 個の値の分散は

『２乗の平均値』−『平均値の２乗』

$\dfrac{1065}{15}-8^2=71-64=7$

**度数分布から標準偏差を求める**

度数分布から標準偏差を求める場合は，各偏差に対応する度数を掛けたものの平均値を用いれば良い。

**練習）** 下の表は，あるクラス 40 人に数学の小テストを行った結果である。このクラスの小テストの得点の標準偏差を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。

ただし，$\sqrt{3}=1.732$ とする。

| 得点 $x$ | 度数 $f$ | $xf$ | 偏差 $x-\bar{x}$ | $(x-\bar{x})^2$ | $(x-\bar{x})^2f$ |
|---|---|---|---|---|---|
| 0 | 2 | 0 | −4 | 16 | 32 |
| 1 | 4 | 4 | −3 | 9 | 36 |
| 2 | 7 | 14 | −2 | 4 | 28 |
| 3 | 8 | 24 | −1 | 1 | 8 |
| 4 | 4 | 16 | 0 | 0 | 0 |
| 5 | 6 | 30 | 1 | 1 | 6 |
| 6 | 2 | 12 | 2 | 4 | 8 |
| 7 | 1 | 7 | 3 | 9 | 9 |
| 8 | 2 | 16 | 4 | 16 | 32 |
| 9 | 3 | 27 | 5 | 25 | 75 |
| 10 | 1 | 10 | 6 | 36 | 36 |
| 計 | 40 (人) | 160 | | | 270 |
| 平均 | | $\bar{x}=4$ | | 分散 $s^2$ | 6.75 |
| | | | | 標準偏差 $s$ | 2.60 |

平均や分散の計算は
度数を掛けたものの
平均値を計算すること

$$\sqrt{\frac{270}{40}}=\sqrt{\frac{27}{4}}$$
$$=\frac{3\sqrt{3}}{2}$$
$$=2.598$$
$$\fallingdotseq 2.60$$

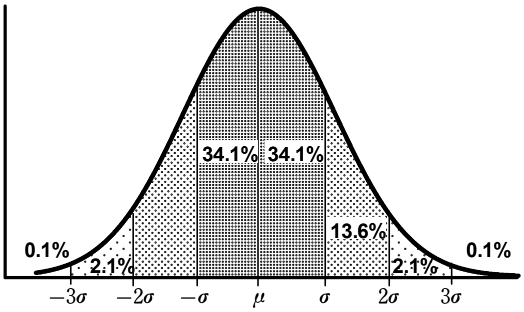
# データの分析【分散と標準偏差】 p.187～189

青チャートⅠ：例題182～184

## 標準偏差と偏差値

四分位範囲にはデータの約50%が含まれることになります。一方、平均値 $\mu$ からのずれが $\pm$標準偏差 $\sigma$ 以下の範囲には 68.27%, $\pm 2\times$標準偏差 $\sigma$ 以下だと 95.45%, さらに $\pm 3\times$標準偏差 $\sigma$ だと 99.73% となります。このことを用いて模擬試験などでは偏差値として数値化し，全体（母集団）との位置関係を示しています。公式は次のようになり

$$(\text{偏差値}) = \frac{10\times\{(\text{得点})-(\text{平均点})\}}{(\text{標準偏差})}+50$$




| | 得点 | 平均値 | 標準偏差 | 偏差値 |
|---|---|---|---|---|
| 第１回 | 81 | 60 | 14 | 65 |
| 第２回 | 76 | 60 | 8 | 70 |

「 ＋50 」とするので偏差値 50 が集団の中央(平均点)となるのです。ただしあくまでも分布内の数値であり，確率に関わる数値なので目安として捉えましょう。たとえば， ２回のテストの点数がそれぞれ，81 点，76 点であっても平均点や標準偏差によって表のように偏差値が求められます。第１回より第２回のテストの点数は下がっていますが，偏差値は上がっています。偏差値は全体の中で自分がどの位置にいるかを示していますから，この場合は、がっかりしなくてもよいのかもしれません。