第33回

分散・標準偏差

講師
湯浅 弘一

まず，次の表を見てください。

最近５試合で，バスケットの A 選手と B 選手がシュートを決めた本数です。

	試合①	試合②	試合③	試合④	試合⑤	合 計	平均値
A 選手	7	6	4	5	8	30	6
B 選手	4	3	10	9	4	30	6

（単位：本）

平均値は同じですが，１つ１つの値を見ると，A 選手と B 選手は違う特徴がありそうです。

A 選手に比べて，B 選手はシュート数が少ないときと多いときの差が大きそうですね。

今回は，このようなデータ全体のばらつきを数値にする方法を学びます。

1 偏 差

偏差は，次のような式で表されます。

（偏差）＝（データの個々の値）－（平均値）

	試合①	試合②	試合③	試合④	試合⑤	合 計	平均値
A 選手	7	6	4	5	8	30	6
平均値との差	1	0	−2	−1	2		
B 選手	4	3	10	9	4	30	6
平均値との差	−2	−3	4	3	−2		

（単位：本）

偏差がマイナス…平均値未満

偏差がプラス……平均値を超えている

偏差＝０…………平均値と等しい

つまり，「偏差」は個々のデータが平均値からどのくらい離れているかを表しています。

偏差には以下の特徴があります。

・偏差の合計は０である。

・偏差の平均値は０である。

2 分散を求める

「偏差」はプラスとマイナスが混在していて，平均が0なので，“データ全体が”平均から見てどの程度離れているか，いわゆる“散らばり”がわかりません。
そこで，偏差の2乗を考えてみましょう。
すると・・・

（単位：本）

	試合①	試合②	試合③	試合④	試合⑤	合 計	平均値
A選手	7	6	4	5	8	30	6
Aの偏差	1	0	−2	−1	2	0	0
(Aの偏差)^2	1	0	4	1	4	10	2
B選手	4	3	10	9	4	30	6
Bの偏差	−2	−3	4	3	−2	0	0
(Bの偏差)^2	4	9	16	9	4	42	8.4

すべて0以上の値になります。これらの平均を「分散」と定めます。
「分散」は個々のデータの散らばり具合を表します。
この場合，
A選手の分散＝2，B選手の分散＝8.4
となり，A選手よりもB選手のほうが，シュート数の散らばり具合が大きいことがわかります。

やってみよう！

8人の10点満点の小テスト成績が以下の表のようになったとします。
この8人のデータの分散を求めなさい。

名 前	A	B	C	D	E	F	G	H
得 点	8	6	4	10	2	3	9	10

（単位：点）

【答 え】
まず，平均値を求めます。
$(8+6+4+10+2+3+9+10)\div 8=52\div 8=6.5$
偏差と偏差の2乗を求めて表にすると

名 前	A	B	C	D	E	F	G	H
得 点	8	6	4	10	2	3	9	10
偏 差	1.5	−0.5	−2.5	3.5	−4.5	−3.5	2.5	3.5
(偏 差)^2	2.25	0.25	6.25	12.25	20.25	12.25	6.25	12.25

$$\begin{aligned}\text{分散} &= (2.25+0.25+6.25+12.25+20.25+12.25+6.25+12.25)\div 8\\ &= 72\div 8\\ &= 9\end{aligned}$$

3 標準偏差

ポイント２で扱ったＡ選手・Ｂ選手のデータを見直してみましょう。

	試合①	試合②	試合③	試合④	試合⑤	合計	平均値
A 選手（本）	7	6	4	5	8	30	6
A の偏差（本）	1	0	−2	−1	2	0	0
（A の偏差）^{2}（本）^{2}	1	0	4	1	4	10	2
B 選手（本）	4	3	10	9	4	30	6
B の偏差（本）	−2	−3	4	3	−2	0	0
（B の偏差）^{2}（本）^{2}	4	9	16	9	4	42	8.4

分散

分散

「偏差」は単位が（本）ですが，平均値が０になってしまうため，全体としての散らばり具合を比べることができませんでした。
「分散」は，個々の値が０以上なので，全体としての散らばり具合を比べることができますが，「偏差」の２乗の平均値なので，単位は（本）^{2} になってしまいます。

そこで，「分散」にルートをつけて，単位を（本）に戻したものが「標準偏差」です。
「標準偏差」もデータ全体の散らばり具合を表します。
A 選手と B 選手の標準偏差を求めると次のようになります。

A 選手　$\sqrt{2} = 1.4142\cdots$　となるので，約 1.4（本）

B 選手　$\sqrt{8.4} = 2.8982\cdots$　となるので，約 2.9（本）

4 分散・標準偏差からわかること

ここまでの公式を実際に使ってみましょう！
ばんびさんとあすみさんが，５００g の粘土のかたまりから，１０個の粘土玉を作りました。
１つ１つの重さを測ったところ，次のようになりました。

ばんび

	1	2	3	4	5	6	7	8	9	10	合計
重さ（g）	46.5	50.0	48.2	51.5	49.6	48.5	50.7	54.2	47.3	53.5	500

あすみ

	1	2	3	4	5	6	7	8	9	10	合計
重さ（g）	55.1	53.2	50.2	49.8	46.2	49.2	43.7	49.6	52.6	50.4	500

「偏差」と「(偏差)^2」を計算したところ，以下のような結果になりました。

ばんび

	1	2	3	4	5	6	7	8	9	10	合計
重さ (g)	46.5	50.0	48.2	51.5	49.6	48.5	50.7	54.2	47.3	53.5	500
偏 差	−3.5	0	−1.8	1.5	−0.4	−1.5	0.7	4.2	−2.7	3.5	0
(偏 差)^2	12.25	0	3.24	2.25	0.16	2.25	0.49	17.64	7.29	12.25	57.82

あすみ

	1	2	3	4	5	6	7	8	9	10	合計
重さ (g)	55.1	53.2	50.2	49.8	46.2	49.2	43.7	49.6	52.6	50.4	500
偏 差	5.1	3.2	0.2	−0.2	−3.8	−0.8	−6.3	−0.4	2.6	0.4	0
(偏 差)^2	26.01	10.24	0.04	0.04	14.44	0.64	39.69	0.16	6.76	0.16	98.18

例 題

ばんびさんとあすみさん，それぞれの分散と標準偏差を求めなさい。

【解 説】

分散は偏差２乗の平均なので、

ばんびさんの分散

$= (12.25 + 0 + 3.24 + 2.25 + 0.16 + 2.25 + 0.49 + 17.64 + 7.29 + 12.25) \div 10$

$= 57.82 \div 10$

$= 5.782$

あすみさんの分散

$= (26.01 + 10.24 + 0.04 + 0.04 + 14.44 + 0.64 + 39.69 + 0.16 + 6.76 + 0.16) \div 10$

$= 98.18 \div 10$

$= 9.818$

標準偏差は，分散にルートをつけた値なので，

ばんびさんの標準偏差 $= \sqrt{5.782} = 2.404\cdots$

あすみさんの標準偏差 $= \sqrt{9.818} = 3.1333\cdots$

※ ばんびさんのほうが，あすみさんよりも標準偏差が小さい・・・ということは，
均等な重さで粘土玉を作ることができた，ということですね！