第38回　第５章　データの分析

分散と標準偏差

監修・執筆
湯浅弘一

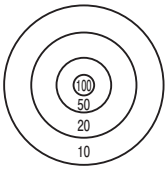
今回学ぶこと

前回学習した四分位数は、中央値をもとにしてデータの散らばりぐあいをとらえるものでした。今回学ぶ偏差、分散、標準偏差は平均値をもとにデータを見ていこうというものです。

学習のポイント

①偏差とは
②分散とは
③標準偏差を求める

ポイント1　偏差とは



得点が10点、20点、50点、100点と分かれた的に10人でダーツを1回ずつ投げました。結果は以下の通りでした。

人	A	B	C	D	E	F	G	H	I	J	合計
得点	10	10	10	20	100	20	50	20	10	20	270

10人の平均点は、270 ÷ 10 = 27点です。

平均点から10人の結果を見てみると…、

A、B、C、D、F、H、I、Jの8人は平均点を下回り、E、Gの2人は平均点を上回っています。

このように、平均値からデータを見る考え方を **偏差** といいます。

偏差＝(データの個々の値)－(平均値)　です。

ダーツのデータの場合、A、B、C、D、F、H、I、Jは、平均点を下回っているので、

偏差はマイナス。

E、Gは、平均点を上回っているので、

偏差はプラスです。

では、データを表にしてみましょう。x の平均値 $\overline{x}=27$ です。

人	得点 (x)	偏差 $(x-\overline{x})$
A	10	−17
B	10	−17
C	10	−17
D	20	−7
E	100	73
F	20	−7
G	50	23
H	20	−7
I	10	−17
J	20	−7
合計	270	0

※ $\overline{x}$ 平均を表す記号です。

偏差の和はいつも 0 です。

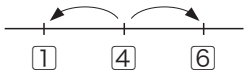
ポイント2　分散とは

サイコロを 8 回振ったとき出た目は、

1 → 2 → 6 → 4 → 4 → 4 → 5 → 6　でした。

平均は、$\dfrac{1+2+6+4+4+4+5+6}{8}=\dfrac{32}{8}=4$　となります。

平均値の4を中心に1から6までの範囲の目が出たことがわかります。



実際に出た目でも4が多くなっています。1〜6は、このデータの範囲（レンジ）です。

分散 とはデータの散らばりぐあいを表す数値です。範囲ではありません。

この場合、偏差（平均との差）を求めると − 3, − 2, 2, 0, 0, 0, 1, 2 となります。

これを平均すると 0 になってしまうので偏差を 2 乗した値の平均を求めます。

$\dfrac{(-3)^2+(-2)^2+2^2+0^2+0^2+0^2+1^2+2^2}{8}=\dfrac{22}{8}=\dfrac{11}{4}$　これが **分散** です。

ポイント3　標準偏差を求める

次に、標準偏差を求めてみましょう。まず、$\sqrt{分散}$＝標準偏差 と定めます。

前記の場合の標準偏差は$\sqrt{\dfrac{11}{4}}$ ですから、

$$\sqrt{\frac{11}{4}}=\frac{\sqrt{11}}{\sqrt{4}}=\frac{\sqrt{11}}{2}$$ です。

分散は、偏差の２乗の平均値なので、もとのデータの値と単位をそろえるために分散の正の平方根を求めたものが標準偏差です。分散も標準偏差もデータの散らばりぐあいを表す数値です。