exercise 【分散と標準偏差】

●データ全体の特徴をあらわそう

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

偏差 $x-\overline{x}$ の 2 乗の平均値のこと。式で表すと $s^2=\frac{1}{n}\left\{\left(x_1-\overline{x}\right)^2+\left(x_2-\overline{x}\right)^2+\cdots\cdots+\left(x_n-\overline{x}\right)^2\right\}$

標準偏差

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

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)。

例）次のデータは 100 点満点のゲームでの男子 10 人の結果である。平均値、偏差、分散、標準偏差を求めよ。(計算は空きスペースで行うこと)

	x	偏差 $x-\overline{x}$	$\left(x-\overline{x}\right)^2$
A	67	9	81
B	42	−16	256
C	59	1	1
D	68	10	100
E	49	−9	81
F	53	−5	25
G	77	19	361
H	48	−10	100
I	77	19	361
J	40	−18	324
平均値	58	分散	169
		標準偏差	13

年（　）組（　　）番　氏名（　　　　解　答　　　　）

練習） 次のデータは 100 点満点のゲームでの女子 10 人の結果である。平均値、偏差、分散、標準偏差を求めよ。（計算は空きスペースで行うこと）

	x	偏差 $x-\bar{x}$	$(x-\bar{x})^2$
α	74	0	0
β	88	14	196
γ	68	−6	36
δ	70	−4	16
ε	60	−14	196
ζ	82	8	64
η	76	2	4
θ	66	−8	64
ι	82	8	64
κ	74	0	0
平均値	74	分散	64
		標準偏差	8

分散と平均値の関係式

分散は次のような求め方もできる。

$$(x\text{のデータの分散})$$
$$=(x^2\text{のデータの平均値})-(x\text{のデータの平均値})^2$$

ただしデータの値が小さくなければ大変なので注意。

（上のデータなどでは $74^2=5476$ などとなり大変になる）

練習 10） 次のデータについて，分散，標準偏差を求めよ。

5,　3,　6,　8,　5,　8,　5,　4,　6,　5

分散　s^2　は

$$s^2=\overline{x^2}-(\bar{x})^2$$
$$=32.5-30.25$$
$$=2.25$$

標準偏差は

$$s=\sqrt{2.25}=1.5$$

標準偏差と偏差値

四分位範囲（平均値 ± 四分位偏差）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 μ からのずれが　± 標準偏差 σ 以下の範囲には　68.27％，± 2 × 標準偏差以下だと 95.45%，さらに　± 3 × 標準偏差　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

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

「+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。