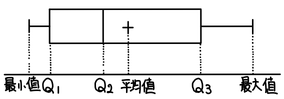
6

・分散と標準偏差

四分位範囲：中央値の周りにおけるデータの散らばり

　?　：平均値の周りにおけるデータの散らばり



変量 x についてのデータの値が $x_1, x_2, x_3, \cdots, x_n$ とn個あるとする。

$x_1-\bar{x},\ x_2-\bar{x},\ \cdots,\ x_n-\bar{x}$ ← 個々のデータが平均値からどのくらいズレているか

を平均値からの偏差という。

偏差の平均値を考えてみると

$\frac{1}{n}\{(x_1-\bar{x})+(x_2-\bar{x})+\cdots+(x_n-\bar{x})\}$

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

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

$=\bar{x}-\bar{x}$

$=0$ ← どんなデータであっても0

これは、散らばりの度合いを表す量として不適切である。

そこで、偏差の2乗の平均値を考えてみる。

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

この値を分散といい、平均値の周りの散らばりの度合いを表す量である。

また、変量 x の測定単位をそろえるため、分散の正の平方根 $\sqrt{s^2}$ を散らばりの度合いを表す量として用いることもある。

この $\sqrt{s^2}$ を s で表し標準偏差という。

> 分散　$s^2=\frac{1}{n}\{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2\}$
>
> 標準偏差　$s=\sqrt{s^2}\ (s>0)$

(例1) 3.5.5.6.8.9 における分散と標準偏差を求めよ。

$\bar{x}=\frac{1}{6}(3+5+5+6+8+9)=6$

x	3	5	5	6	8	9
$x-\bar{x}$	-3	-1	-1	0	2	3
$(x-\bar{x})^2$	9	1	1	0	4	9

← 偏差の和0

$s^2=\frac{1}{6}(9+1+1+0+4+9)=4$ //

$s=\sqrt{4}=2$ //

> 分散　$s^2=\overline{x^2}-(\bar{x})^2$
>
> ※ $\overline{x^2}$ は $x_1^2, x_2^2, \cdots, x_n^2$ の平均値を表す

(証明)

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

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

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

$=\overline{x^2}-2\bar{x}\,\bar{x}+(\bar{x})^2$ ← $\overline{x^2}$ と $(\bar{x})^2$ は意味が異なる

$=\overline{x^2}-(\bar{x})^2$ □

(例2) 2.4.6.8.9 における分散を求めよ。

$\bar{x}=\frac{1}{5}(2+4+6+8+9)=\frac{29}{5}$

x	2	4	6	8	9
$x-\bar{x}$	$-\frac{19}{5}$	$-\frac{9}{5}$	$\frac{1}{5}$	$\frac{11}{5}$	$\frac{16}{5}$
$(x-\bar{x})^2$	$\frac{361}{25}$	$\frac{81}{25}$	$\frac{1}{25}$	$\frac{121}{25}$	$\frac{256}{25}$

← 偏差の和0

$s^2=\frac{1}{5}\left(\frac{361}{25}+\frac{81}{25}+\frac{1}{25}+\frac{121}{25}+\frac{256}{25}\right)=6.56$

別解

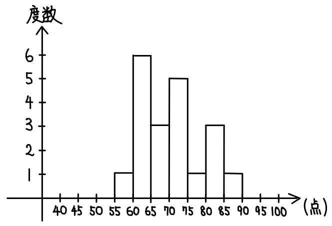
x	2	4	6	8	9
x^2	4	16	36	64	81

$\overline{x^2}=\frac{1}{5}(4+16+36+64+81)=\frac{201}{5}$

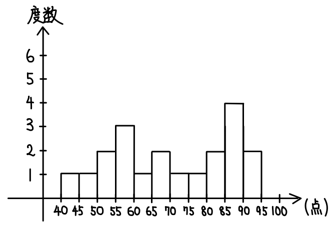
$s^2=\overline{x^2}-(\bar{x})^2=\frac{201}{5}-\left(\frac{29}{5}\right)^2=6.56$

(例3)

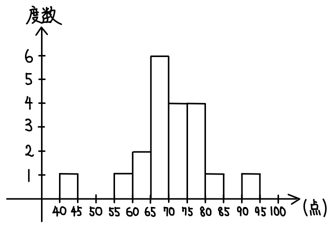
(データI)　平均値：70　分散：64



(データII)　平均値：70　分散：260



(データIII)　平均値：70　分散：111.1



散らばりの度合い

I < III < II