

Data analysis 3

平均をもとに、散らばり具合を調べたい！

	x	$x - \bar{x}$	$(x - \bar{x})^2$
A	14	2	4
B	10	-2	4
C	13	1	1
D	13	1	1
E	15	3	9
F	18	6	36
G	6	-6	36
H	10	-2	4
I	11	-1	1
J	10	-2	4
計	120	0	100

\bar{x} : 平均値 $\bar{x} = \frac{1}{10} \times 120 = 12$

偏差・分散・標準偏差

偏差... 各値から平均を引いたもの
 $x_i - \bar{x}$

分散... 偏差の2乗和の平均

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

$$= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

標準偏差... $\sigma = \sqrt{10}$

計算してみよう。

1) 分散

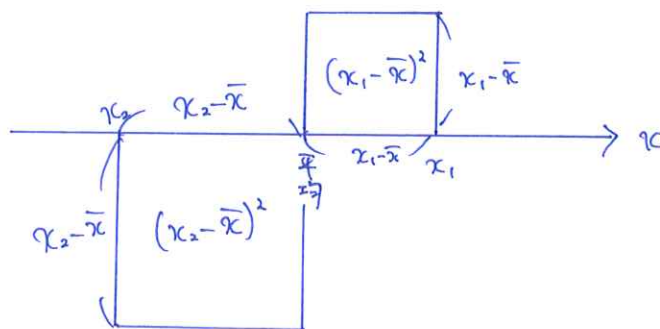
$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 100 \text{ 7kade}$$

$$V = \frac{1}{10} \times 100 = 10$$

2) 標準偏差

$$\sigma = \sqrt{10}$$

偏差・分散・標準偏差のイメージ



平均化

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

分散

標準偏差

- 分散... 面積の平均
- 標準偏差... 平均の正偏差の辺の長さ

分散と平均の関係を調べてみよう。

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2)$$

$$= \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \right)$$

$$= \frac{1}{n} \left(\sum_{i=1}^n x_i^2 \right) - 2\bar{x} \cdot \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \times n\bar{x}^2$$

$$= \frac{1}{n} \left(\sum_{i=1}^n x_i^2 \right) - (\bar{x})^2$$

分散と平均の関係式

$$(\text{分散}) = \left(\frac{\text{各値の2乗の平均}}{\text{2乗の平均}} \right) - (\text{平均})^2$$