Measures of spread / dispersion

1. Variance
2. Standard deviation

* Mean- deviation

0

1 unit ↑ 1 unit 2 unit

1, 2, 3, 4, 5

2 unit ↑

mean

$$\Rightarrow \frac{2+1+0+1+2}{5}$$

$$\Rightarrow \frac{6}{5} = 1.2$$

→ On an avg each of the data is 1.2 units away from mean value

* Variance — The average of the squared differences from the mean.

Population variance

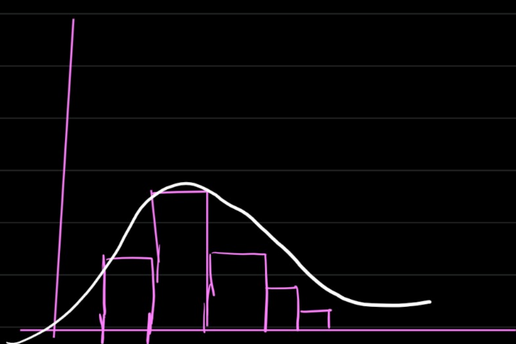
$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

μ → population mean

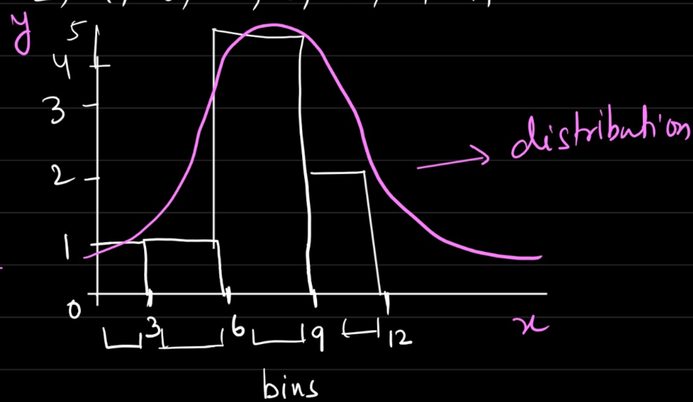
Sample variance

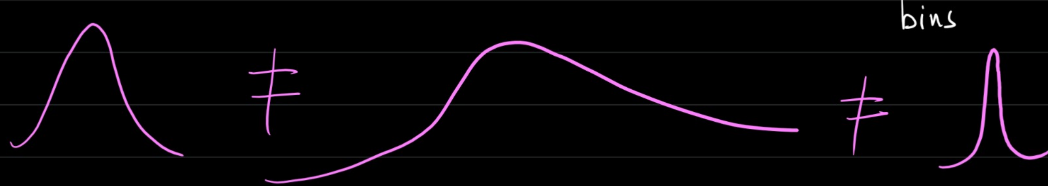
$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\boxed{n-1}}$$

$\bar{x}$ → sample mean



✓ ✓
2, 4, 6, 6, 6, 8, 8, 11, 10





data = {1, 2, 3, 3, 4, 4}

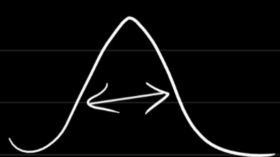
How to Calculate variance

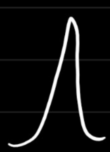
- → Calculate mean
- → for each no in data, Subtract the mean and the no
- → Square of difference
- → Calculate the avg of square of difference.

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
1	2.83	-1.83	3.34
2	2.83	-0.83	0.68
3	2.83	0.17	0.03
3	2.83	0.17	0.03
4	2.83	1.17	1.37
4	2.83	1.17	1.37
2.83			6.82

$$s^2 = \frac{6.82}{n-1} = \frac{6.82}{5} = 1.37$$

Variance ↑ Spread ↑





→ spread ↓ Var ↓

* Standard deviation

Standard deviation is a measure of how spread out numbers are.

↳ Square root of Variance

$$s = \sqrt{var}$$

$$std \Rightarrow \sqrt{1.37} = 1.17$$

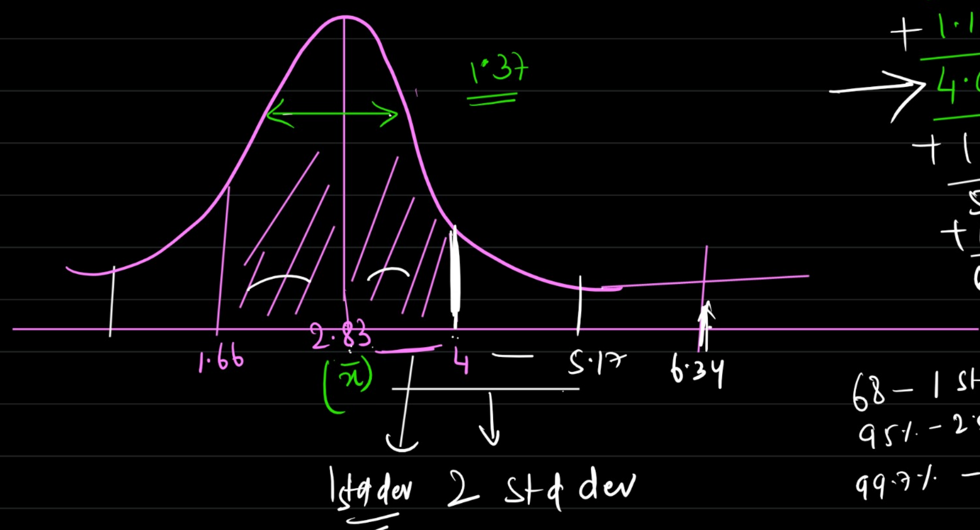
Std dev of Population $\Rightarrow \sigma = \sqrt{Var_p}$

Std dev of Sample $\Rightarrow s = \sqrt{var_s}$

1, 2, 3, 3, 4, 4

$\to$ var $= 1.37$

$\to$ std dev $= 1.17$ $\longrightarrow$ standard way of knowing where your data lies



$$\begin{array}{r} 2.83 \\ -1.17 \\ \hline 1.66 \end{array}$$

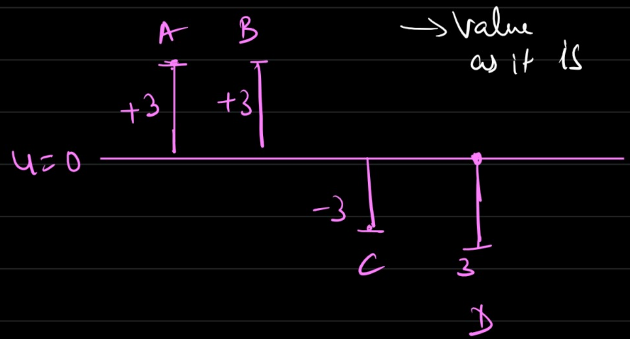
$$\begin{array}{r} 2.83 \\ +1.17 \\ \hline 4.00 \\ +1.17 \\ \hline 5.17 \\ +1.17 \\ \hline 6.34 \end{array}$$

68 – 1 std
95% – 2 std
99.7% – 3 std

* Variance $\longrightarrow$

$$Var_p = \sigma^2 = \frac{\sum_{i=1}^{n} (x-\mu)^2}{n}$$

why square?



$\to$ value as it is

$x-\mu$

$$\frac{+3+3+(-3)+(-3)}{4} = 0$$

$\downarrow$

+ve and -ve are negating each other

absolute value

$$\to \frac{|3|+|3|+|-3|+|-3|}{4} = \frac{12}{4} = 3$$

mean deviation

$\to$

$$\to \frac{|+8|+|+1|+|-2|+|-1|}{4} = \frac{12}{4} = 3$$

$$\sqrt{\frac{3^2+3^2+(-3)^2+(-3)^2}{4}} = \sqrt{\frac{36}{4}} = 3$$

$$\sqrt{\frac{8^2+1^2+(-2)^2+(-1)^2}{4}} = \sqrt{\frac{64+1+4+1}{4}} = \sqrt{\frac{70}{4}} = 4.184$$

Bessel correction

* Variance$_{\text{sample}}$ $= \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$ $\longrightarrow$ Why?

We use $n-1$ rather than n is because sample variance will be unbiased estimator

↓

We are estimating variance of population using variance of sample

↓

$(x-\bar{x})^2$

Population $(x-\mu)^2$

$(x-\mu)^2 > (x-\bar{x})^2$ = var.

$2 = \dfrac{10}{5} = \dfrac{8}{4} = 2$

$n \longrightarrow n-1$. $\longrightarrow$ In most of the cases you are reducing numeration

$\Rightarrow S^2 = \dfrac{\sum(x-\bar{x})^2}{n-1}$ $\rightarrow$ smaller