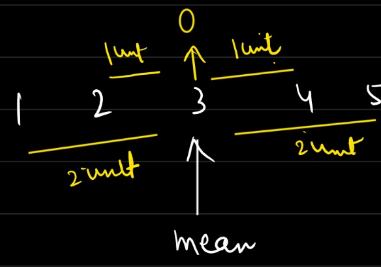
* Measures of Spread / dispersion
  1. Mean deviation
  2. Variance
  3. Standard deviation

1. Mean deviation



$$\frac{\sum_{i=1}^{n} |x_i - \mu|}{N}$$

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

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

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

2. Variance

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

N → no. of elements

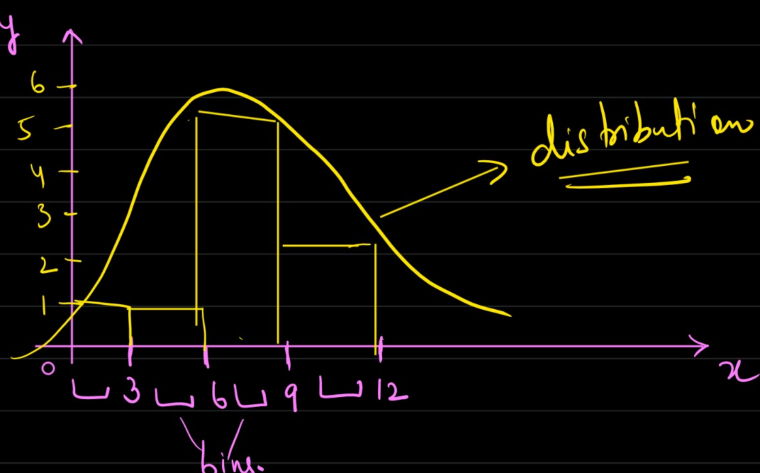
Sample Variance

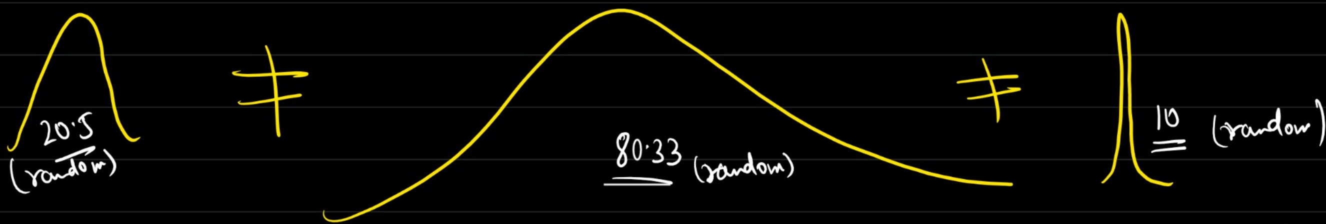
$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$\bar{x}$ → sample mean

$n-1$ → no. of elements − 1

* 2, 4, 6, 6, 6, 8, 8, 11, 10





* Variance talks about spread at an overall level.

* Spread ↑ Variance ↑

How to Calculate variance

→ Calculate mean
→ For each no in data, subtract the mean & the no
→ Square of difference
→ Calculate the avg of square of difference

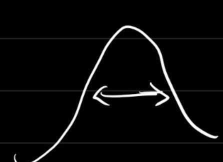
data = {1, 2, 3, 3, 4, 4}

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
			6.82

$\bar{x} = 2.83$

$$S^2 = \frac{6.82}{n-1} = \frac{6.82}{6-1} = 1.37$$

* Variance ↑ Spread ↑



→ Spread ↓ Var ↓

③ Standard deviation

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

↳ square root of variance

$$s = \sqrt{var}$$
$$= \sqrt{1.37} = \underline{\underline{1.17}}$$

Population

$$\sigma = \sqrt{var_p}$$

Sample

$$S = \sqrt{var_s}$$

1, 2, 3, 3, 4, 4

↳ var = 1.37 (can be any no)

↳ std dev = 1.17

Why std dev?

→ variance can be huge no because it talks about spread at an overall level. Comparison of each no wrt variance becomes difficult.

→ $$\frac{\sum_{i=1}^{n} (x-\mu)^2}{N} \Rightarrow \frac{(\text{dis} - \text{Avg dis})^2}{N} = \frac{(3m - 5m)}{N} = \underline{\underline{m^2}}$$

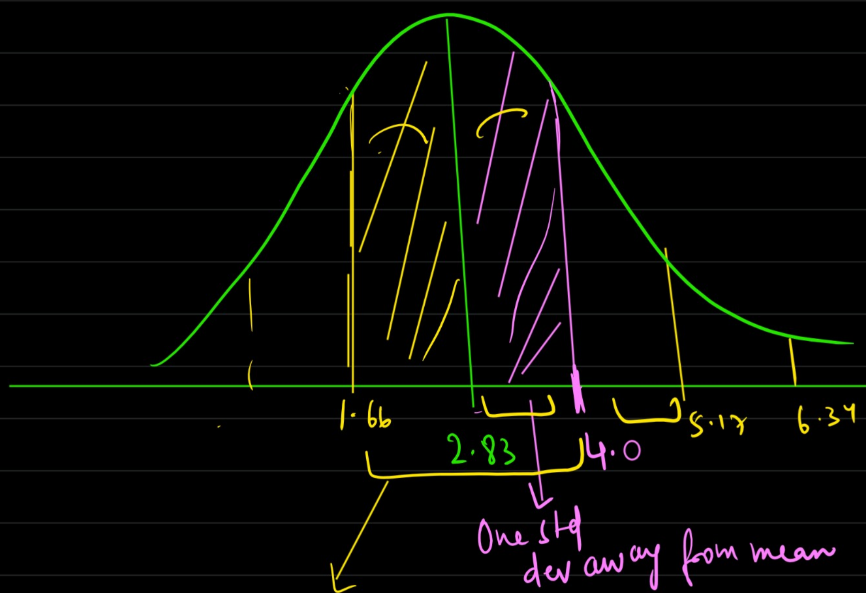
* In Variance dimension change.

in std dev $= \sqrt{var} = \sqrt{m^2} = m$

1, 2, 3, 3, 4, 4.

↳ var = 1.37

↳ std dev = 1.17 → Standard way of knowing where your data lies.



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

−1 sd — 1 s.d

68% — 1 std

95% — 2 std

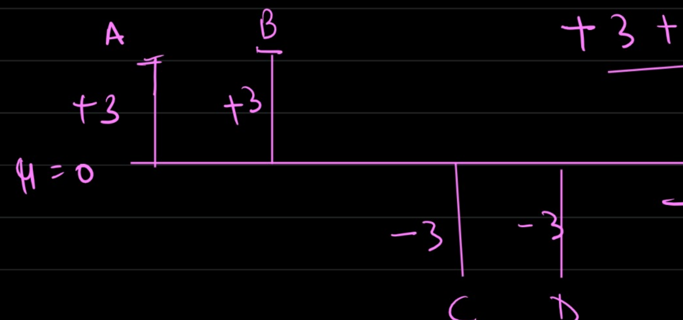
99.7% — 3 std.

* Variance

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

Why square?

*

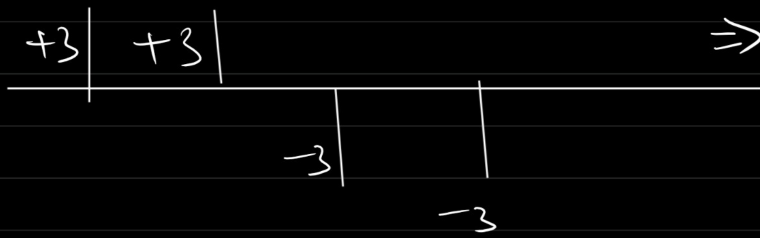


$$\frac{+3 +3 -3 -3}{4} = 0$$

→ +ve & −ve values are negating each other.

if Absolute value

Sc-1



$\Rightarrow \dfrac{|3|+|3|+|-3|+|-3|}{4}$

$= \dfrac{12}{4} = 3$

mean deviation

Sc-2

$\mu = 0$

$\Rightarrow \dfrac{|+8|+|+1|+|-2|+|-1|}{4} = \dfrac{12}{4} = 3$

Is the spread same? $\rightarrow$ NO

if Squared difference

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

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

* Variance Sample = $\frac{\sum (x_i - \bar{x})^2}{n-1}$ → Bessel Correction — Why ?

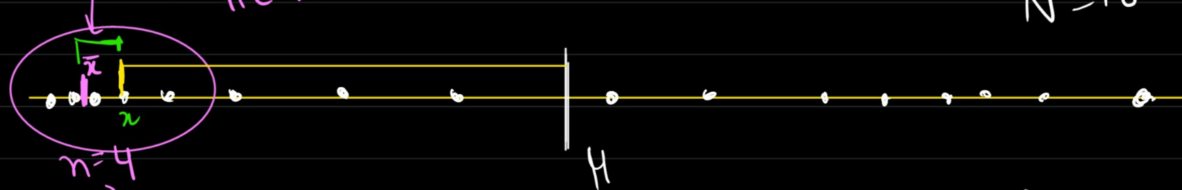
Why you calculate any statistics of sample?

→ No access to complete population

* We use $n-1$ rather the N is because Sample variance will be unbiased estimator.

→ Since no access to population, therefore we are estimating the variance of population using sample

Sample mean will lie here.



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

$(x-\mu)^2 > (x-\bar{x})^2 = \text{Var} \quad \frac{(x_i-\mu)^2}{}$

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

$\Rightarrow \frac{(x-\mu)^2}{N}$

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

$n \longrightarrow n-1 \longrightarrow$ In most of of the case you are reducing numerator.

$$S^2 = \frac{\sum (x-\bar{x})^2}{n-1} \longrightarrow \text{Smaller}$$