Measures of Dispersion:

① Variance

② Standard deviation

Variance ↴

· Population variance ⟶

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

Sample variance ⟶

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

x_i = Data points

μ = Population mean

N = population size

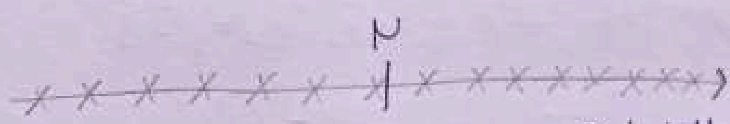
$\bar{x}$ = sample mean

n = Sample size

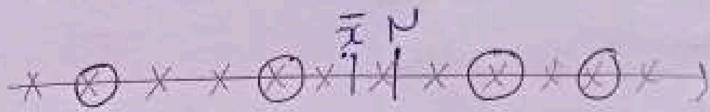
Why sample variance have $n-1$ in denominator? ☆ Interview question

Sample variance will be unbiased estimator of population variance.

ages

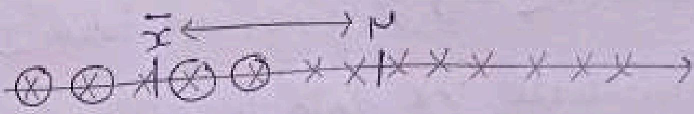


Case 1 ↴ taking sample data ~~with~~ containing all age groups



$\bar{x} \approx \mu \qquad s^2 \approx \sigma^2$

Case 2 ↴ taking sample data containing one agegroup.



This happens in case of skewed data

Researchers tried to calculate sample variance by $n-1$, $n-2$, $n-3$, . . . and by selecting different samples.

After many trials, they decided that $n-1$ is best since when we divide by $n-1$

$$\sigma^2 \approx s^2$$

Difference is getting reduced

Example:

Sample $\Rightarrow \{1, 2, 3, 4, 5\}$

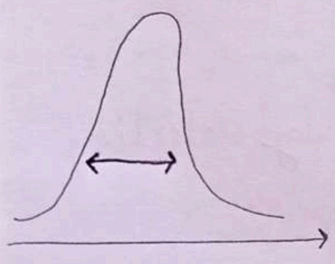
x_i	$\bar{x}$	$(x_i - \bar{x})^2$
1	3	4
2	3	1
3	3	0
4	3	1
5	3	4
		10

$$s^2 = \frac{10}{5-1} = 2.5$$

Variance $\rightarrow$ It gives an idea how well the data is spread.

Problem 1

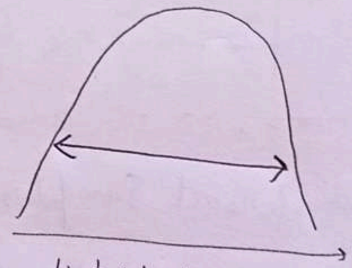
$s^2 = 2.5$



distribution

Problem 2

$s^2 = 6.5$



distribution

Standard deviation $\rightarrow$

Population STD

$$\sigma = \sqrt{\text{Population variance}}$$

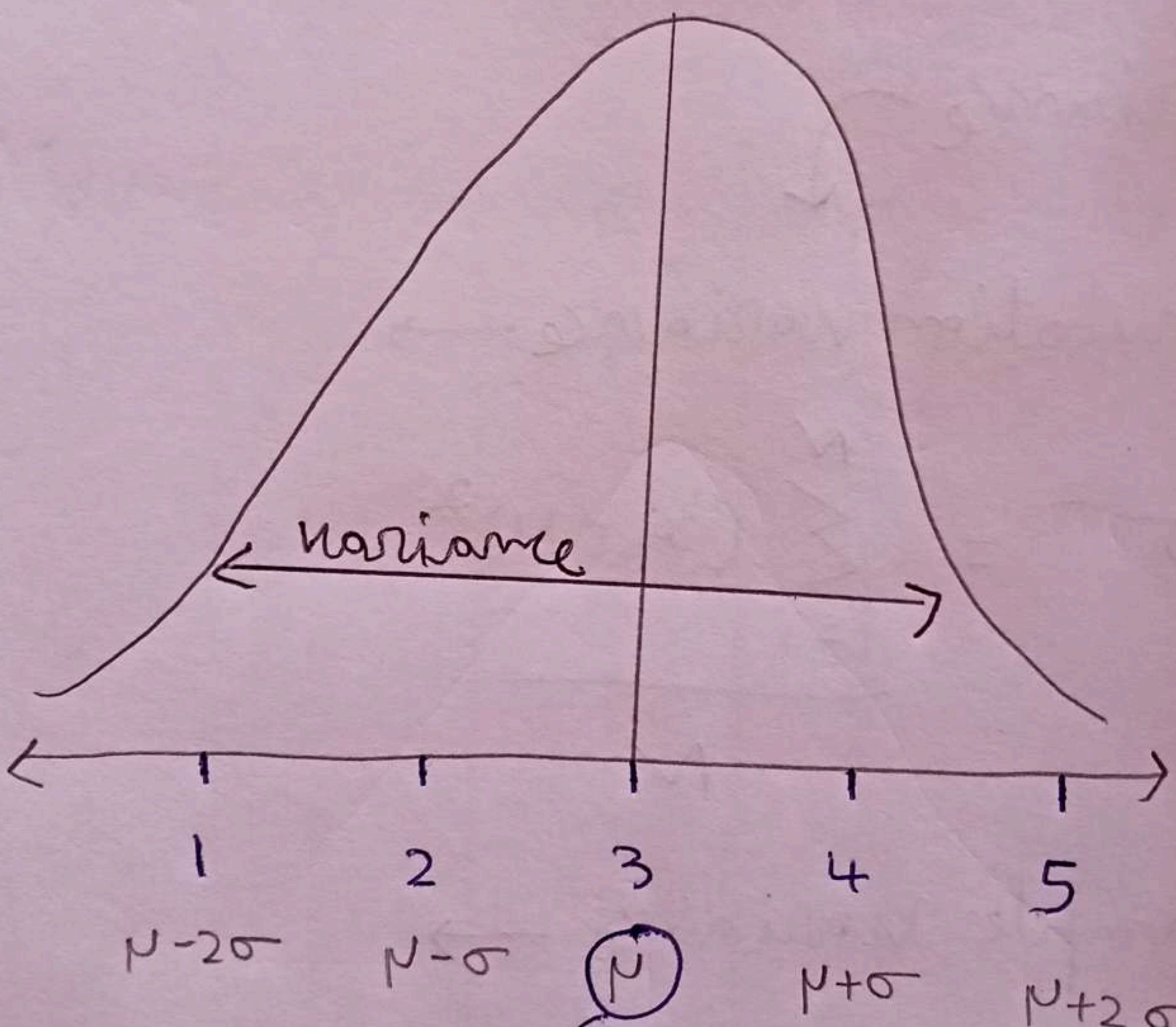
Sample STD

$$s = \sqrt{\text{Sample variance}}$$

Graphical understanding:

$\{1, 2, 3, 4, 5\}$

Lets say, $\mu = 3$ $\sigma = 1$



It can mean or median or mode

Skewness: