3) Measure of Dispersion (Spread of the data)

1. Variance
2. Standard Deviation

If my data is spread, how can I use the Quantitative measure to measure the spreadness. So here we use Variance and SD.

1. Variance

Population Variance

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

$x_i \rightarrow$ Data points

$\mu \rightarrow$ Population mean

$N \rightarrow$ Population size

- This formula talks about the spread. How well your data is spread.
- we are squaring because distance can't be negative.

Sample Variance

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

$x_i \rightarrow$ Data points

$\bar{x} \rightarrow$ sample mean

$n \rightarrow$ sample size

Interview Question:

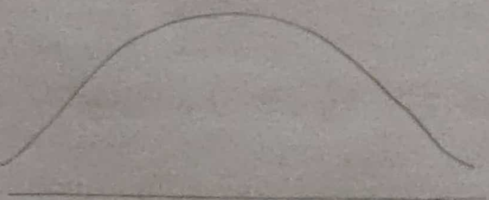
1) why we divide sample variance by $n-1$?

This is also called Bessel's Correction

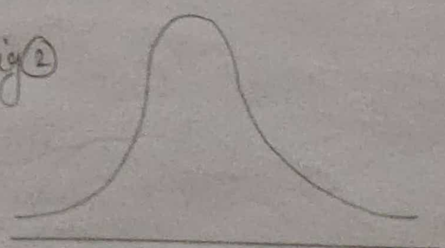
Ans: The sample variance is divided by $n-1$ so that we can create an unbiased estimator of the population variance.

* If Data is spread like below then, which is having the maximum variance?

fig①



fig②



Ans: fig① because, the more the spread, the more the variance.

2. Standard Deviation

Population S.D.

$\sigma = \sqrt{\text{Variance}}$

Sample S.D.

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

what this tells about:

How many S.D. (unit) x_i is away from Mean.

• example: $\{1, 2, 3, 4, 5\}$

$\mu = 3$

$$\text{Population variance } (\sigma^2) = \frac{[(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2]}{5}$$

$$= \frac{4+1+0+1+4}{5}$$

$$= \frac{10}{5}$$

$$= 2.$$

$\therefore \sigma = \sqrt{2} = 1.41$

* let's see how to use the S.D. to construct entire distribution.

$\mu = 3 \quad \sigma = 1.41$

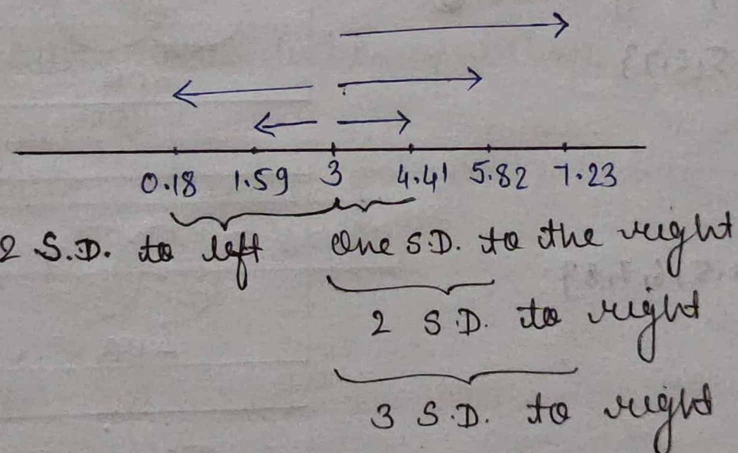
$3 - 1.41 = 1.59$

$1.59 - 1.41 = 0.18$

$3 + 1.41 = 4.41$

$4.4 + 1.41 = 5.82$

$5.8 + 1.41 = 7.23$



* **Qn**: where does 4.41 lies in the distribution?

Ans: It lies 1 S.D. ~~to the left~~ right to Mean.