L-3 Data Analytics

Variance & Standard Deviation

$\bar{X}$ (Mean)

(Data points) | (Data Points)

$$\text{Variance} = \frac{1}{n}\left[\sum_{i=1}^{N} (x_i - \mu)^2\right]$$

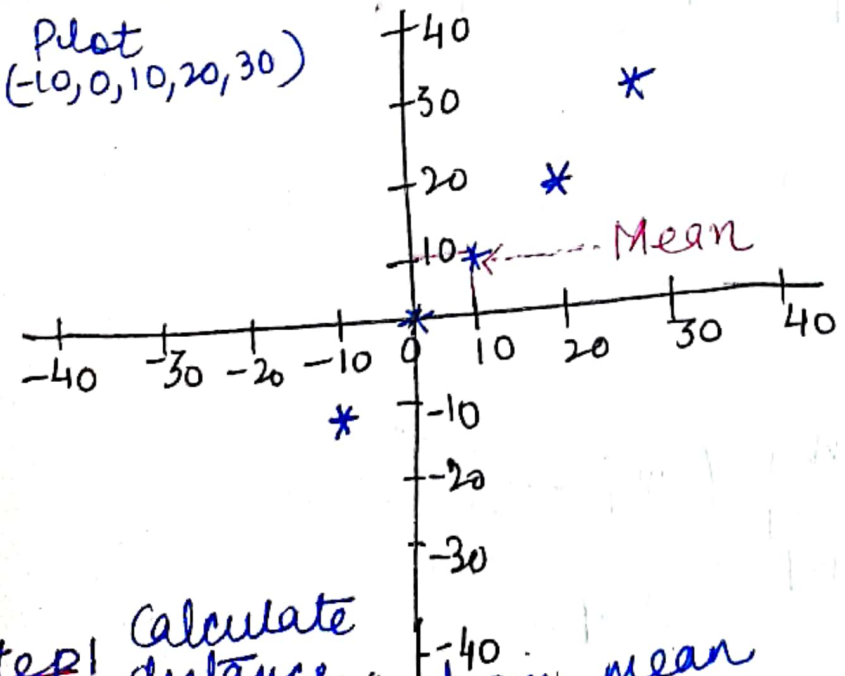
where $x_i \rightarrow$ data points ($i = 0$ to n)

$\mu \rightarrow$ Mean

$n \rightarrow$ total no. of data points

Variance tells the spread ie distance of the data points from Mean Value.

Plot (-10, 0, 10, 20, 30)



distance of all data points from Mean.

$(-10 - 10) = -20$ (x_i, $\bar{X}$)

$(0 - 10) = -10$

$10 - 10 = 0$

$20 - 10 = 10$

Step 1 Calculate distance of data points from mean

$x_i - \bar{X}$

Since it is seen some values are negative.

Step2 Squaring

So Squaring to remove the impact of neg. values.

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

Step3. Averaging

$$\frac{(x_i - \bar{X})^2}{N}$$

Ily distance from each point is measured so. sum is done represented.

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

Population Variance

Sample variance

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

Standard Deviation $\sqrt{\sigma^2}$

Q1

Population 1

-10, 0, 10, 20, 30

Mean

$$\frac{-10+0+10+20+30}{5}$$

$$\Rightarrow \frac{50}{5} \Rightarrow 10.$$

Population 2

8, 9, 10, 11, 12

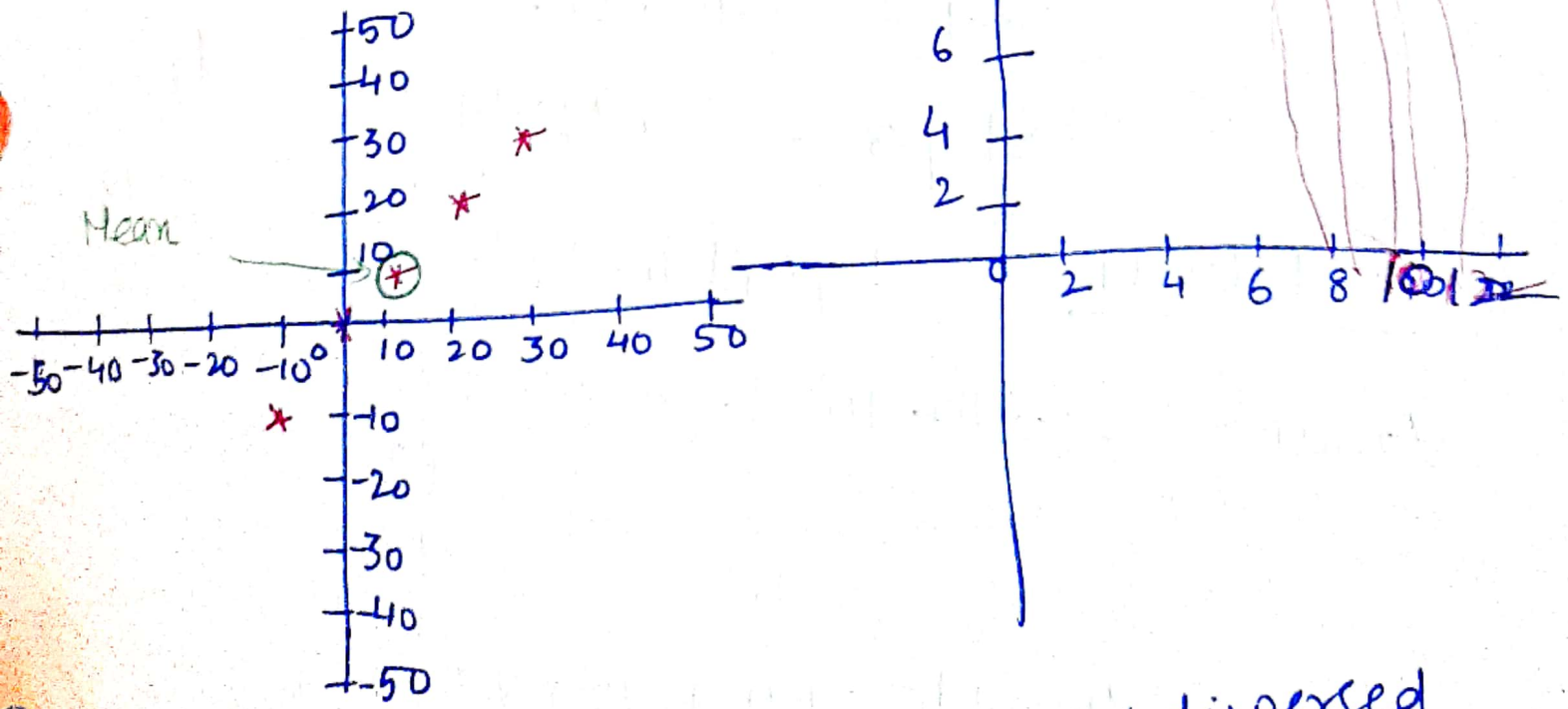
Mean

$$= \frac{8+9+10+11+12}{5}$$

$$= \frac{50}{5} = 10$$

Both have same means.

But it can be clearly seen both have different Dispersion as compared to Mean.



> dispersed as compared to

< dispersed as compared to mean

to calculate the dispersion or spread.

RANGE $-10, 0, 10, 20, 30$

Population 1 → largest − smallest

Range $\Rightarrow 30-(-10)$

$\Rightarrow 40.$

Population 2 $8, 9, 10, 11, 12$

Range = largest − smallest

$= 12-8$

$= 4.$

VARIANCE Step 1 − Distance of data Points from mean

Step 2 − sum all.

Step 3 − square.

Step 3 − Averaging.

Population 1. $-10, 0, 10, 20, 30$

$$\text{Variance} = \frac{(-10-10)^2 + (0-10)^2 + (10-10)^2 + (20-10)^2 + (30-10)^2}{5}$$

$$\Rightarrow \frac{400+100+0+100+400}{5} \Rightarrow \frac{1000}{5}$$

Population 2.

8, 9, (10), 11, 12.

$$\text{Variance} = \frac{(8-10)^2 + (9-10)^2 + (10-10)^2 + (11-10)^2 + (12-10)^2}{5}$$

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

$$\Rightarrow \frac{10}{5} \Rightarrow 2.$$

Standard Deviation

for population 1 $= \sqrt{\sigma^2}$

$= \sqrt{200}$

$= 10\sqrt{2}$

Population 2 $= \sqrt{2}$.