Standard Deviation and Variance

↳ Measures of dispersion

Dataset 1

$\{1, 2, 3, 9, 5\}$

$\text{Mean} = \frac{1+2+3+9+5}{5}$

$= \frac{15}{5} = 3$

Dataset 2

$\{0, 1, 2, 5, 7\}$

$\text{Mean} = \frac{0+1+2+5+7}{5}$

$= \frac{15}{5} = 3$

Range = Max - Min

DI: Range = 5 - 1 = 4 → 4

Range = 7 - 0 = 7 → 7

Variance

Definition :- Variance is basically the spread of the data from the mean

2 types :-

Population Variance

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

Sample variance

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

D1: $\{1, 2, 3, 9, 5\}$

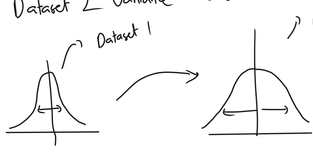
x	μ	x − μ	(x − μ)²
1	3	(1-3)	
2	3	(2-3)	
3	3	(3-3)	
9	3	(9-3)	
5	3	(5-3)	

D2: $\{0, 1, 2, 5, 7\}$

$(x-\mu)^2$ → $\frac{(x-\mu)^2}{N} = 2$ → $= 6.8$

Dataset 1 Variance = 2

Dataset 2 Variance = 6.8 → spread is more than Dataset 1 (from the mean)



Variance is giving the overall spread

D1 variance = 2

D2 variance = 6.8

$\{1, 2, 3, 9, 5\}$ → mean, square → m^2

$\{0, 1, 2, 5, 7\}$ → mean, square → m^2

Standard Deviation :-

↳ just the square root of variance → $\sqrt{9}$

↳ gives us the spread of a certain data with respect to the mean (on average) → 3

D1 variance = 2 → std = $\sqrt{2} = 1.414$

D2 variance = 6.8 → std = $\sqrt{6.8} = 2.608$

D1 = $\{1, 2, 3, 9, 5\}$

D2 = $\{0, 1, 2, 5, 7\}$

Variance = Overall spread

Standard deviation = average spread or range of values around the mean

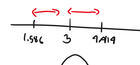
Dataset 1

mean	std
3	1.414

$\{0, 1, 2, 3, 9, 5\}$

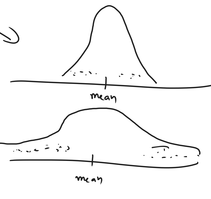
3 + 1.414 = 4.414

3 − 1.414 =



std = 10

std = 50



Why did we use n-1 in sample variance?

↳ instead of N

based on degrees of freedom

Sample variance → we use sample mean to find the sample variance → there is uncertainty → to prevent this we used (n-1) → gives us unbiased estimate