IT Skill for Chemist

Standard Deviation

Standard deviation: Standard deviation is the measure of how spread out or dispersed the value in a data set are.

Definition: Standard deviation (denoted as σ or population for sample) shows how much the values deviates from the mean (average of the data set).

Formula: For a population

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

For sample :

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

Note: We divide by n−1 (not n) — this is called **Bessel's correction**, and it corrects bias in the estimation.

Where:

- σ = population standard deviation
- s = sample standard deviation
- N = number of values in the population
- x_i = each value in the population
- μ = population mean
- n = number of values in the sample
- $\bar{x}$ = sample mean

Key points:

i. Low standard deviation: Data points are closed to the mean (less spread).

ii. High standard deviation: Data points are most spread out from the mean.

iii. It is always non-negative.

iv. Unit of the standard deviation is the same as data.

Population vs Sample: Key Differences

Definition	Entire group of individuals or data points	Subset of the population
Size	Denoted by **N** (usually larger)	Denoted by **n** (usually smaller)
Mean	Population mean $\mu = \frac{1}{N}\sum x_i$	Sample mean $\bar{x} = \frac{1}{n}\sum x_i$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N}\sum(x_i - \mu)^2}$	$s = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$
Denominator in SD	Divide by **N**	Divide by **n − 1** (Bessel's correction)
Purpose	Describes the actual data spread in a full set	Estimates the spread of the **entire population** based on a sample
Accuracy	Precise (if all data is known)	Approximate (used for inference)
Use Case	When data for the entire population is available	When data is available only for a part of the population

Q1. A student performs an acid-base titration 5 times to determine the volume of NaOH (titrate) requires to neutralized a fixed volume of HCl(analyte). The volume reading (ml) are :

25.1, 25.0, 24.9, 25.2, 25.0

Find the standard deviation.

Solution:

Step1: Calculate the mean (or average volume)

$\bar{x}=\frac{25.1+25.0+24.9+25.2+25.0}{5}\ ml$

$=\frac{125.2}{5}ml$= 25.04 ml

Step 2: Find the deviation from the mean and their squares

Volume (ml)	Deviation $(x_i - \bar{x})$	Square deviation $(x_i - \bar{x})^2$
25.1	(25.1-25.04)=0.06	$(0.06)^2$=0.0036
25.0	(25.0-25.04)= -0.04	$(-0.04)^2$ = 0.016
24.9	(24.9-25.04)= - 0.14	$(-0.14)^2$ = 0.0196
25.2	(25.2-25.04)=0.16	$(0.16)^2$ =0.0256
25.0	(25.0-25.04)=- 0.04	$(-0.04)^2$ =0.0016

Step 3:

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

$$S=\sqrt{\frac{0.0036+0.0016+0.0196+0.0256+0.0016}{5-1}}$$

$=\sqrt{\frac{0.052}{4}} = \sqrt{0.013}$= 0.114 ml

Interpretation:

Mean value used =25.04 ml

Standard deviation =0.114ml

Small standard deviation indicates the titration result is precise, meaning the student perform the titration consistently.

Mean, Median and Mode

1. **Mean (average):** Mean is the sum of all values divided by number of values.

   Formula: Mean= $\frac{\text{Sum of all values}}{\text{number of values}}$

Example: Data: 2,4,6,8,10

Mean= $\frac{2+4+6+8+10}{5} = \frac{30}{5}$ =6

2. **Median (Middle value):** The median is the middle value when the data values are arranged in ascending order.

Steps: i. Arrange data in order.

ii. If **odd number of values**, middle value is the median.

iii. If **even number**, average of the two-middle value is the median.

Example: In odd case: data: 3,5,7 Median =5

In even case: data 4,6,8,10, Median = $\frac{6+8}{2} = 7$

3. **Mode (**most frequent value**):** The mode is the number that appears most frequently.

Example : data : 2, 3, 3, 5,7

Mode= 3 (Because it occurs repeatedly).

N.B.: No mode → if all numbers occur once.

(‘bimodal’ or multimodal)More than one mode → if two or more value occurs equally.