Measures of Dispersion

Measures of Dispersion describe the spread or variability of a dataset.

While **measures of central tendency** (mean, median, mode) tell us where the center is, **measures of dispersion** tell us how far the data values are from that center.

High dispersion means data points are spread out widely, while low dispersion means they are clustered close to the center.

Common measures of dispersion:

- Range
- Variance
- Standard Deviation
- Interquartile Range (IQR)

1. Range

The simplest measure of dispersion, calculated as the difference between the largest and smallest values in the dataset.

- Formula

Range= Maximum Value − Minimum Value

- Example

Dataset: 5, 8, 12, 20, 25

Range = 25 − 5 = 20

- Advantages
  - Easy to calculate and understand.

- Limitations
  - Only depends on extreme values (max and min), ignoring the rest of the data.
  - Sensitive to outliers.

2. Variance

Variance measures the average squared deviation from the mean.

It shows how far data points are from the mean, on average.

- Formula (**Population Variance**) **VS** Formula (**Sample Variance**)

Population Variance	Sample Variance
$\sigma^2 = \dfrac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$	$s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
$\sigma^2 =$ population variance	$s^2 =$ sample variance
$x_i =$ value of i^{th} element	$x_i =$ value of i^{th} element
$\mu =$ population mean	$\bar{x} =$ sample mean
$N =$ population size	$n =$ sample size

- Example

Dataset: 4, 8, 6

Mean = 6

Variance = $((4-6)^2 + (8-6)^2 + (6-6)^2) / 3 =$ 2.67

- Advantages
  - Considers all data points.

- Limitations
  - Units are squared, making interpretation less intuitive.

3. Standard Deviation

The square root of variance, bringing the unit back to the same as the original data.

- Formula (**Population Standard Deviation**) **VS** Formula (**Sample Standard Deviation**)

Standard Deviation Formula	
Population	**Sample**
$\sigma = \sqrt{\dfrac{\sum(X-\mu)^2}{N}}$	$s = \sqrt{\dfrac{\sum(X-\bar{x})^2}{n-1}}$
X – The Value in the data distribution *μ – The population Mean* *N – Total Number of Observations*	*X – The Value in the data distribution* *$\bar{x}$ – The Sample Mean* *n - Total Number of Observations*

- Example

Using the variance example above:

Standard Deviation = $(2.67)^{1/2} \approx 1.63$

- Advantages
  - Easier to interpret than variance because it uses the same units as the data.

- Limitations
  - Sensitive to outliers.

4. Interquartile Range (IQR)

Measures the range of the middle 50% of the data by subtracting the 25th percentile (Q1) from the 75th percentile (Q3).

- Formula

IQR = Q3 – Q1

- Example

Dataset: 2, 4, 6, 8, 10, 12, 14

Q1 = 4, **Q3** = 12

IQR = 12 − 4 = 8

- Advantages
    - Not affected by extreme values (robust measure).
- Limitations
    - Ignores data outside the middle 50%.

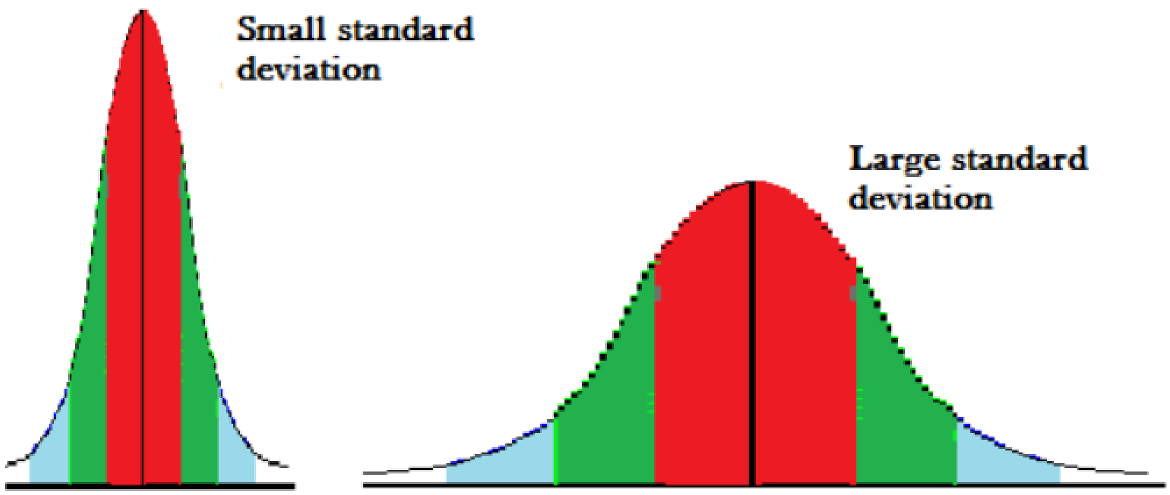
5. Comparison

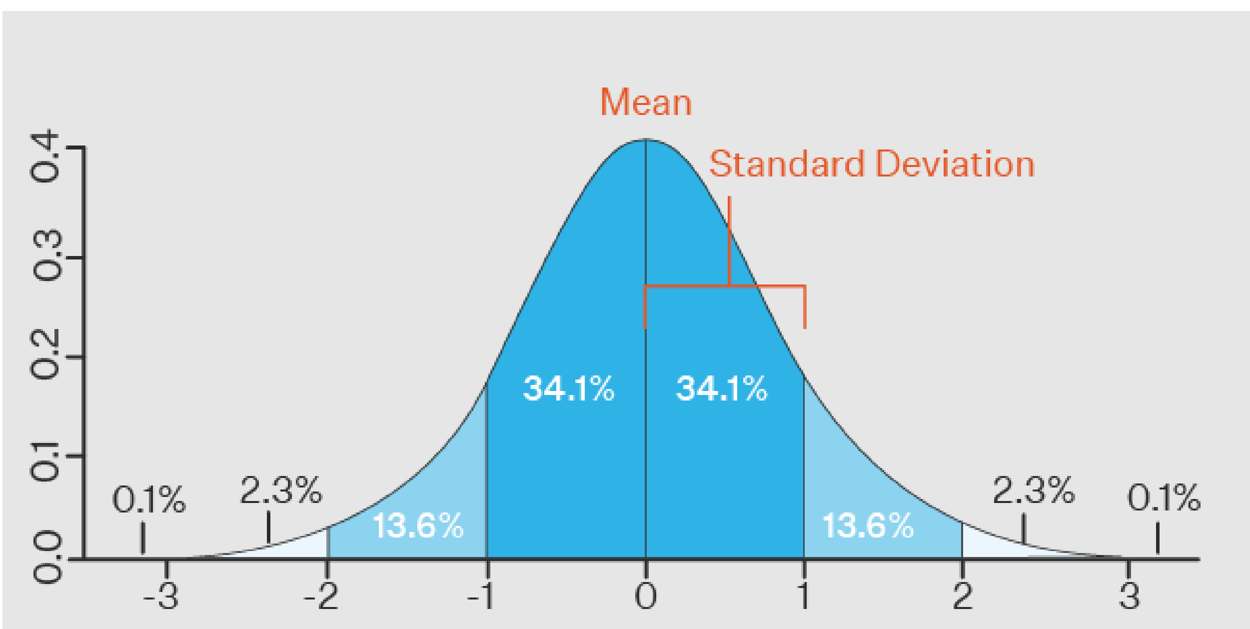
Measure	**Uses All Data?**	**Sensitive to Outliers?**	**Units**
Range	✗ No	☑ Yes	Same as data
Variance	☑ Yes	☑ Yes	Squared units
Standard Deviation	☑ Yes	☑ Yes	Same as data
IQR	✗ No	✗ No	Same as data

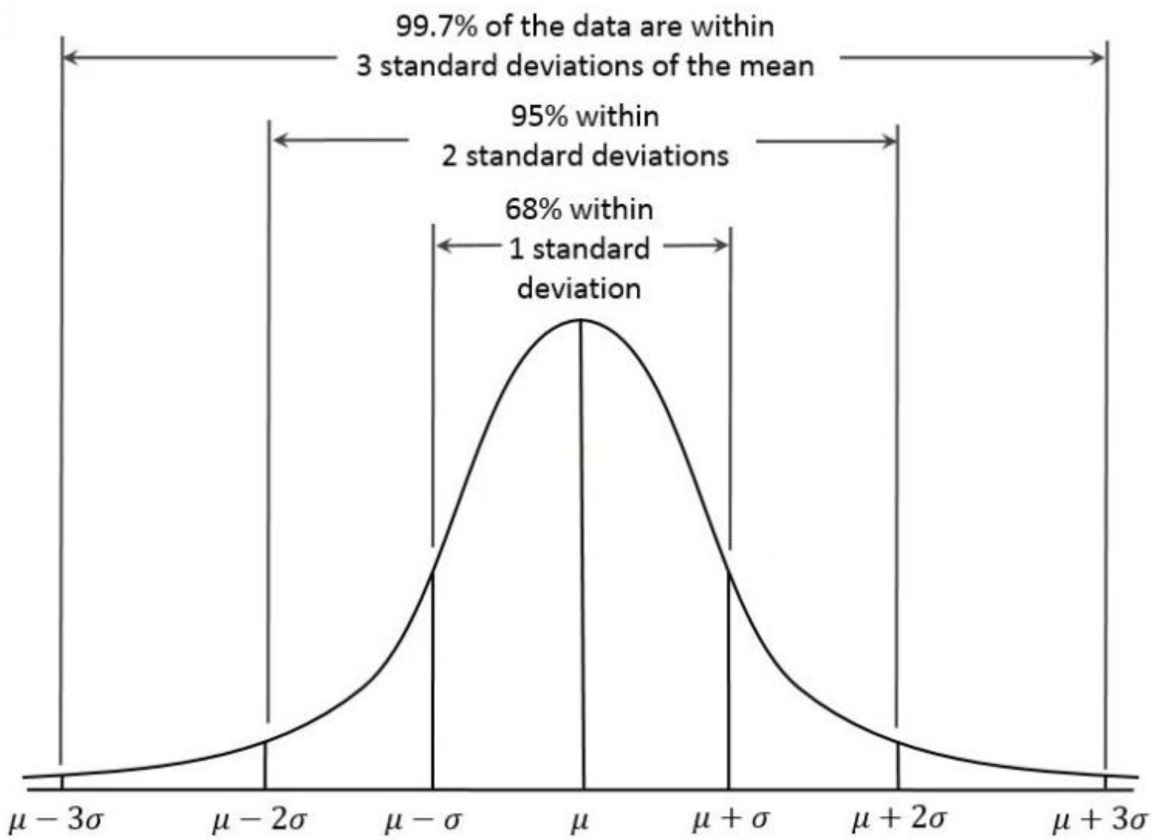
6. Common Mistakes

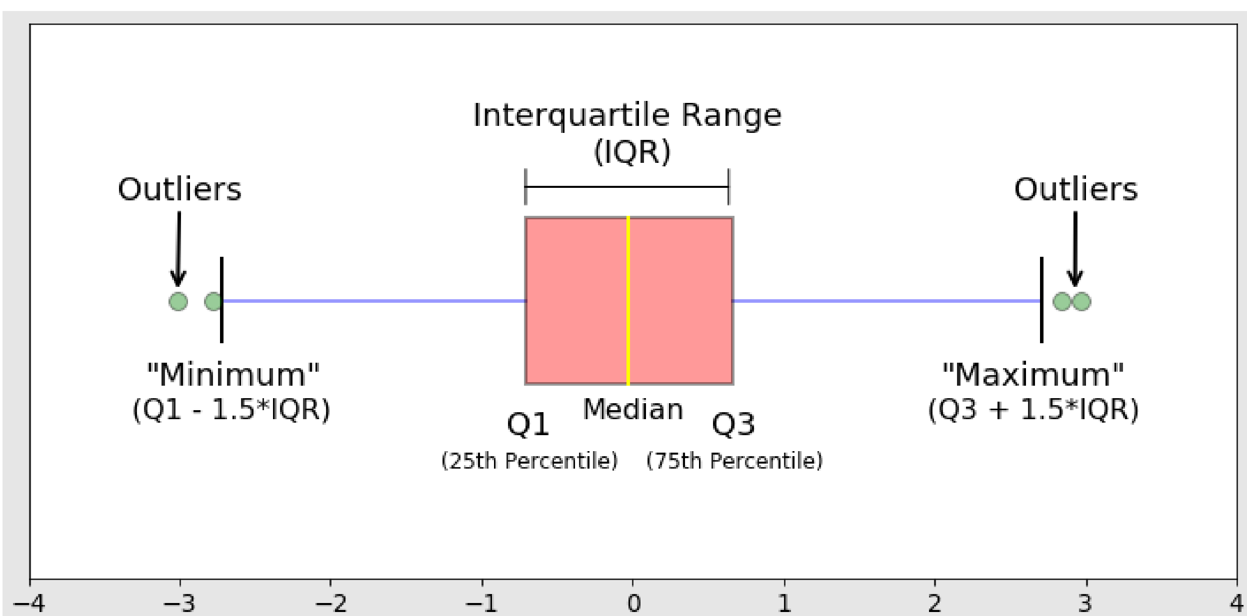
- Using range to judge spread in datasets with extreme outliers.
- Forgetting to take the square root of variance when interpreting standard deviation.
- Mixing up sample and population formulas.

7. Visual Example






99.7% of the data are within
3 standard deviations of the mean
95% within
2 standard deviations
68% within
1 standard
deviation
$\mu - 3\sigma$
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$
$\mu + 3\sigma$


Interquartile Range
(IQR)
Outliers
Outliers
"Minimum"
(Q1 - 1.5*IQR)
Median
Q1
(25th Percentile)
Q3
(75th Percentile)
"Maximum"
(Q3 + 1.5*IQR)
−4
−3
−2
−1
0
1
2
3
4