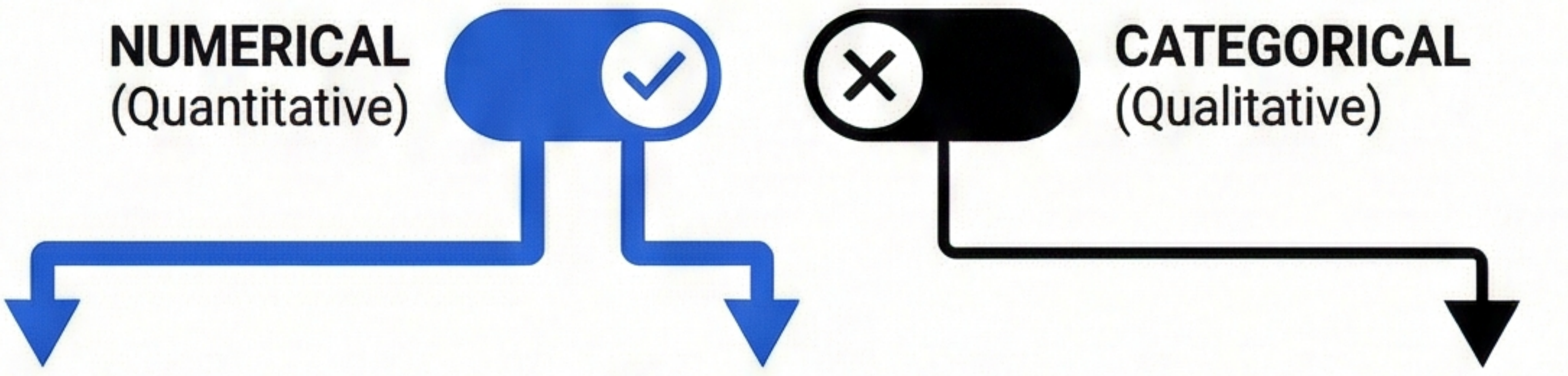
DESCRIPTIVE STATISTICS

Summarizing Data for Management Insight

DATA TYPE CONTEXT

NUMERICAL (Quantitative)

CATEGORICAL (Qualitative)

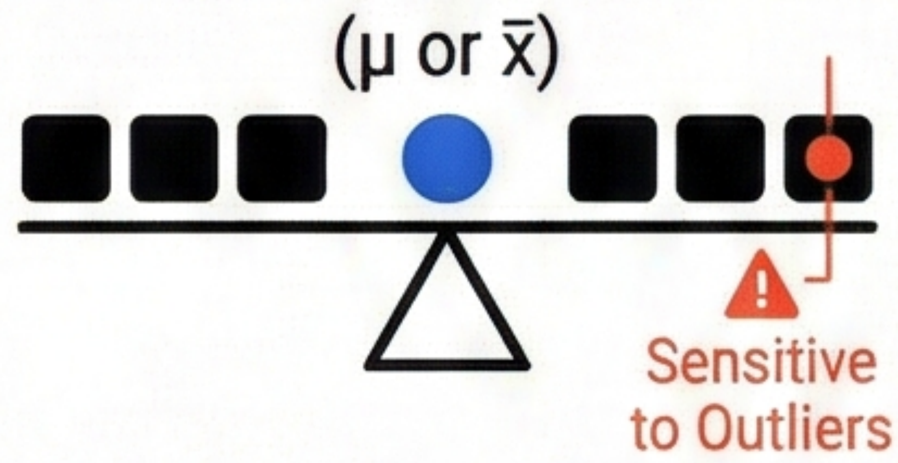


MEASURES OF CENTRAL TENDENCY (The "Center")

CENTRAL TENDENCY

Locating the typical or central value.

Mean
(μ or $\bar{x}$)

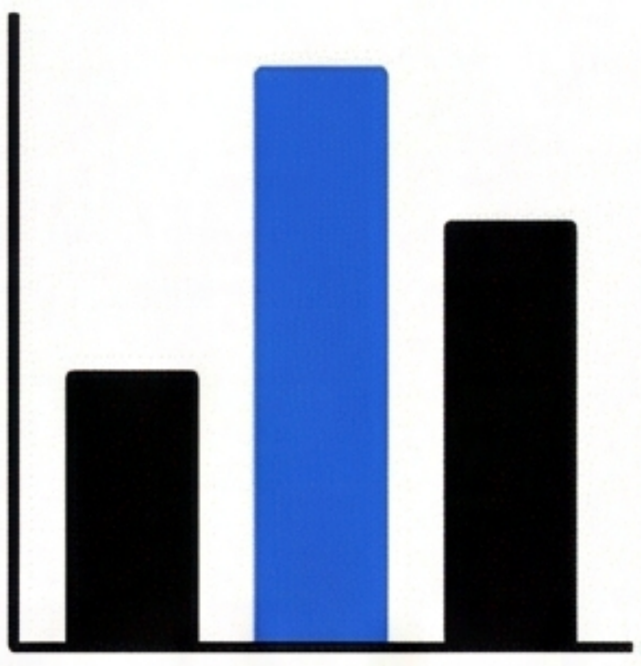


Arithmetic Average
(Balance Point)

Ordered

Median
(M)

Middle Value
(Positional Center)
Robust to Outliers



Mode

Most Frequent Value
(Peak)
For Categorical & Numerical

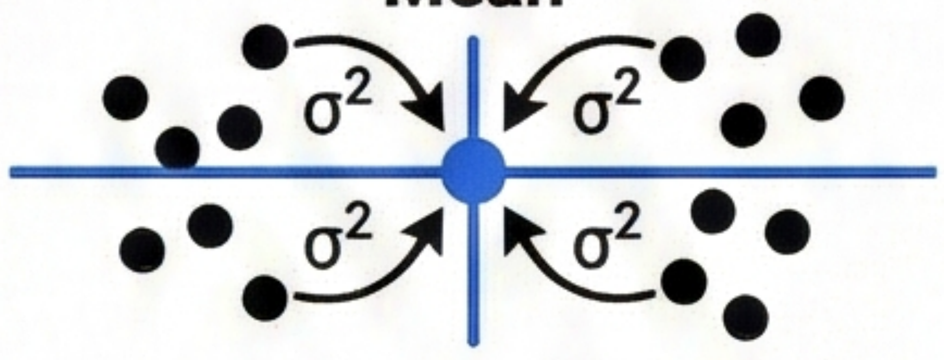
MEASURES OF DISPERSION (The "Spread")

DISPERSION

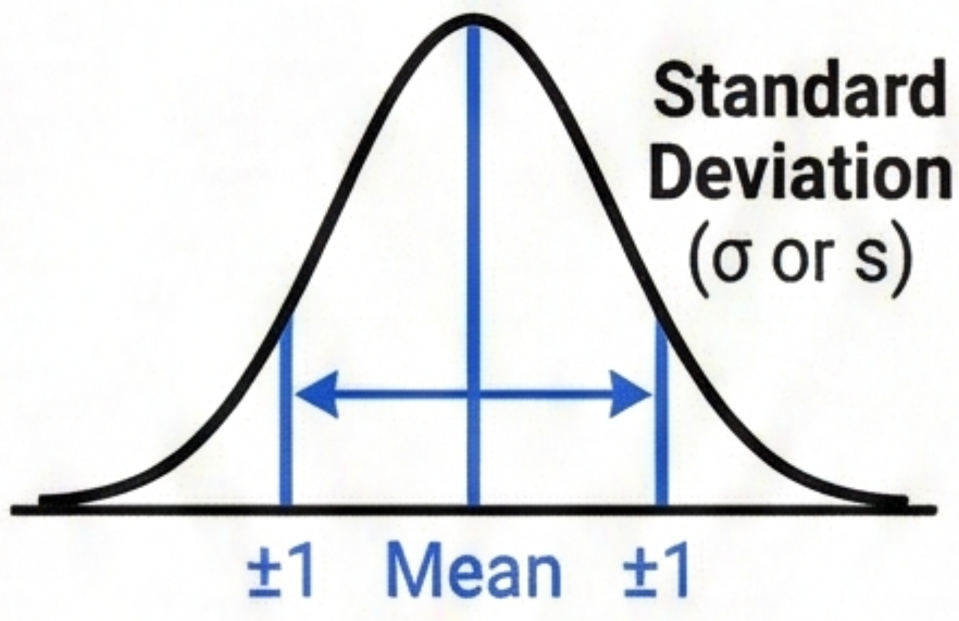
Quantifying the variability or spread.

Range
Max – Min (Total Spread)

Mean

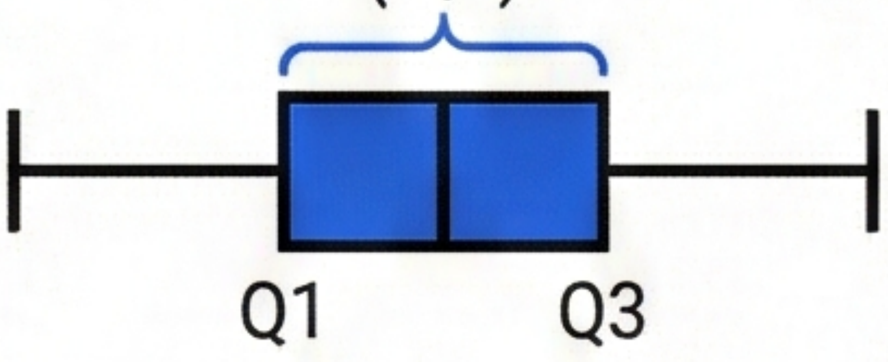


Variance (σ^2 or s^2)
Average Squared Deviation
Units are squared



√Variance
(Typical Distance from Mean)
Original Units, Most Common

Interquartile Range
(IQR)

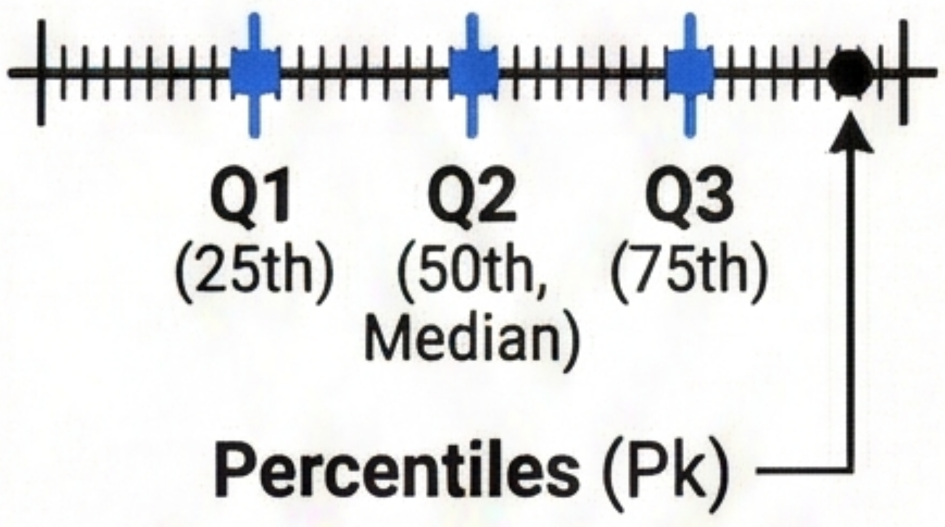


Q3 - Q1
(Middle 50% Spread)
Robust to Outliers

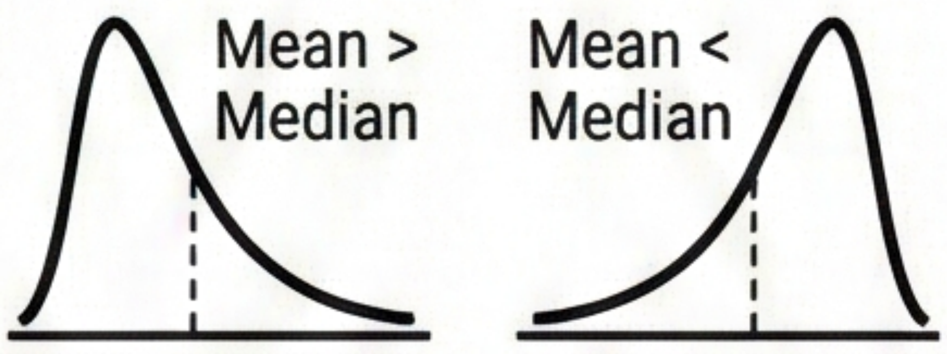
MEASURES OF POSITION & SHAPE (The "Form")

POSITION & SHAPE

Describing location and distribution form.

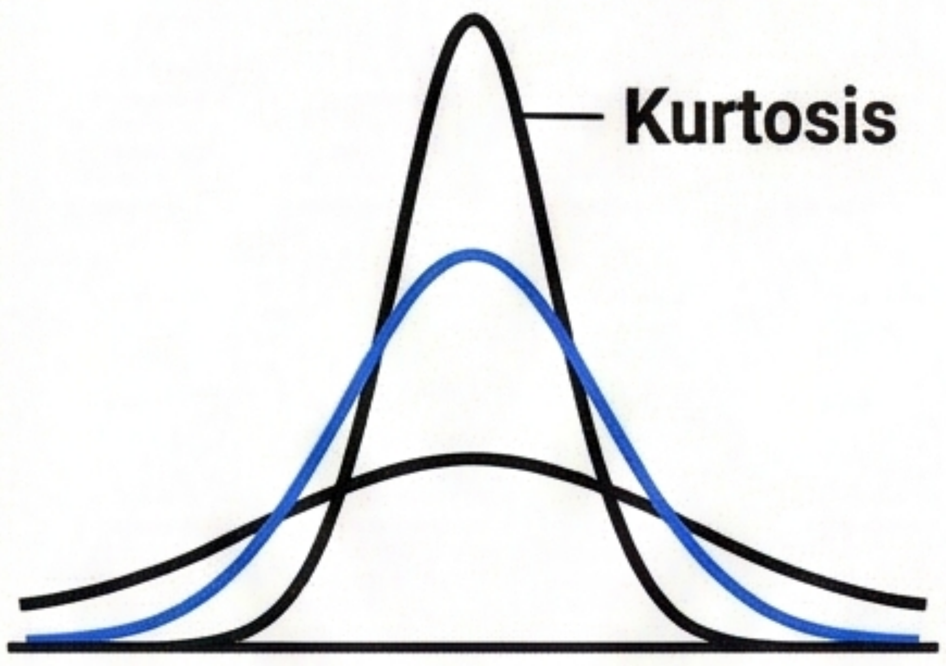


Percentiles (Pk)
Value below which k% of data falls.



Positively Skewed
(Right Tail)

Negatively Skewed
(Left Tail)



Peakedness/Tails
relative to Normal.

Descriptive statistics provide a foundational snapshot of data, distinct from inferential statistics which make predictions.

Zbigniew Galar

MEASURES OF CENTRAL TENDENCY

Describing the Center of a Dataset with a Single Value for Management Decision-Making.

EXAMPLE DATASET (N=6)

OUTLIER

2	3	3	4	8	25

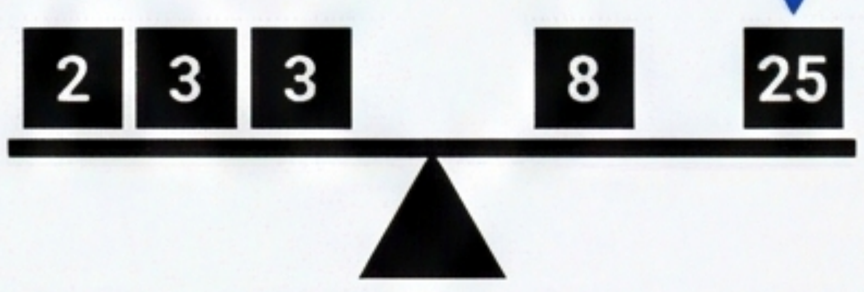
Ordered Numerical Data

1. MEAN (ARITHMETIC AVERAGE)

The mathematical balance point.

ALERT: Highly Sensitive to Outliers. The '25' pulls the mean upwards significantly.

$$\frac{\text{Sum of Values } (\sum x)}{\text{Count of Values } (N)} = \text{Mean } (\mu \text{ or } \bar{x}) \rightarrow \frac{(2+3+3+4+8+25)}{6} = \frac{45}{6} = \boxed{7.5}$$



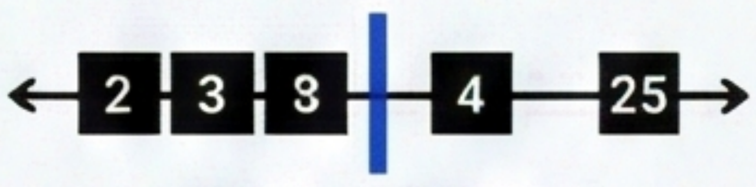
2. MEDIAN (MIDDLE VALUE)

The positional center, splitting data in half.

Average of Middle Values:

$$2 \quad 3 \quad 3 \;|\; 4 \quad 8 \quad 25 \rightarrow \frac{3+4}{2} = \boxed{3.5}$$

Two Middle Values (Even N)



MEDIAN POINT
(Robust to Outliers)

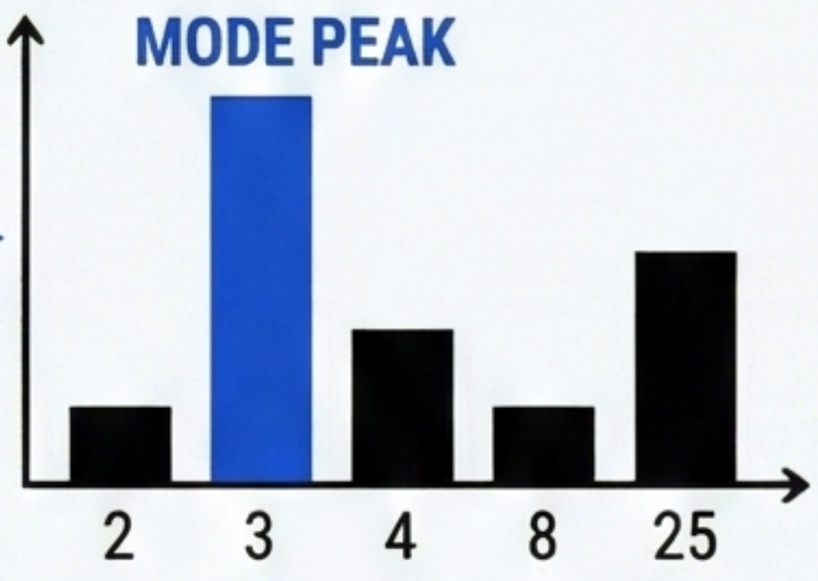
Changing '25' to '100' would NOT change the Median.

3. MODE (MOST FREQUENT)

The value with the highest occurrence peak.

$$2 \quad 3 \quad \boxed{3 \; 3} \quad 4 \quad 8 \quad 25 \rightarrow \boxed{3}$$

Highest Frequency (Count = 2)



Useful for Categorical Data. Can be Non-unique or None.

COMPARISON & APPLICATION SUMMARY

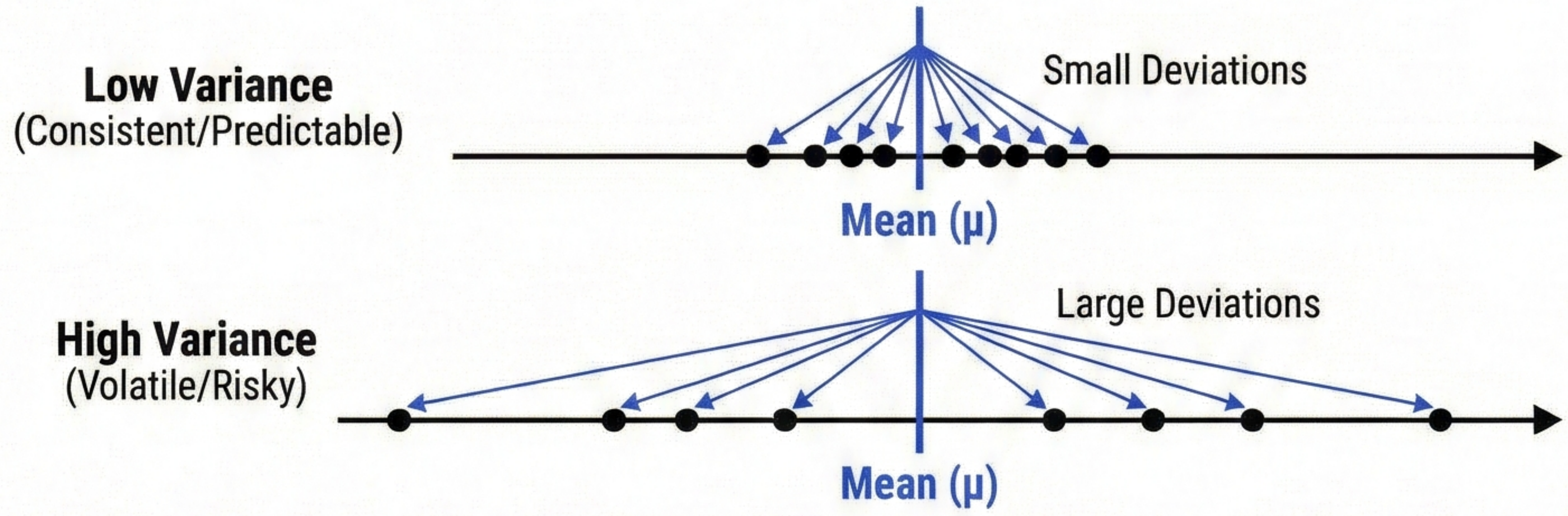
MEASURE	BEST USE CASE (Cobalt Blue Focus)	WATCH OUT FOR (Vermilion Alert)
MEAN	Symmetric Data, No Outliers.	Outliers, Skewed Data.
MEDIAN	Skewed Data, Presence of Outliers.	Ignores Data Magnitudes.
MODE	Categorical Data, Finding Peaks.	Not Always Central, Multimodal.

Zbigniew Galar

VARIANCE (σ^2 or s^2)

Quantifying Data Dispersion & Risk by Averaging Squared Deviations from the Mean

1. THE CORE CONCEPT: SPREAD & CONSISTENCY



Vermilion Alert: High variance indicates greater uncertainty or risk in management decisions.

2. THE CALCULATION FLOW (STEP-BY-STEP)

1. Find the Mean (μ or $\bar{x}$)

$$\mu = \frac{\sum x}{N}$$

↓

2. Calculate Deviations ($x - \mu$)

-3	0	+4

Distances from Mean

↓

3. Square Deviations $(x - \mu)^2$

9	0	16

Focus: Squaring eliminates negative values & emphasizes larger deviations.

↓

4. Average the Squares (Variance, σ^2)

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N}$$

$\sigma^2 = 12.5$ (Squared Units)

Vermilion Alert: Result is in squared units, not original units.

3. THE FORMULAS (POPULATION vs. SAMPLE)

Population Variance (σ^2)

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N}$$

N = Total Population Size

μ = Population Mean

Sample Variance (s^2)

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

n = Sample Size

$\bar{x}$ = Sample Mean

Vermilion Alert: (n-1) corrects bias for estimation.

4. KEY IMPLICATION: INTERPRETATION

INTERPRETING THE VALUE

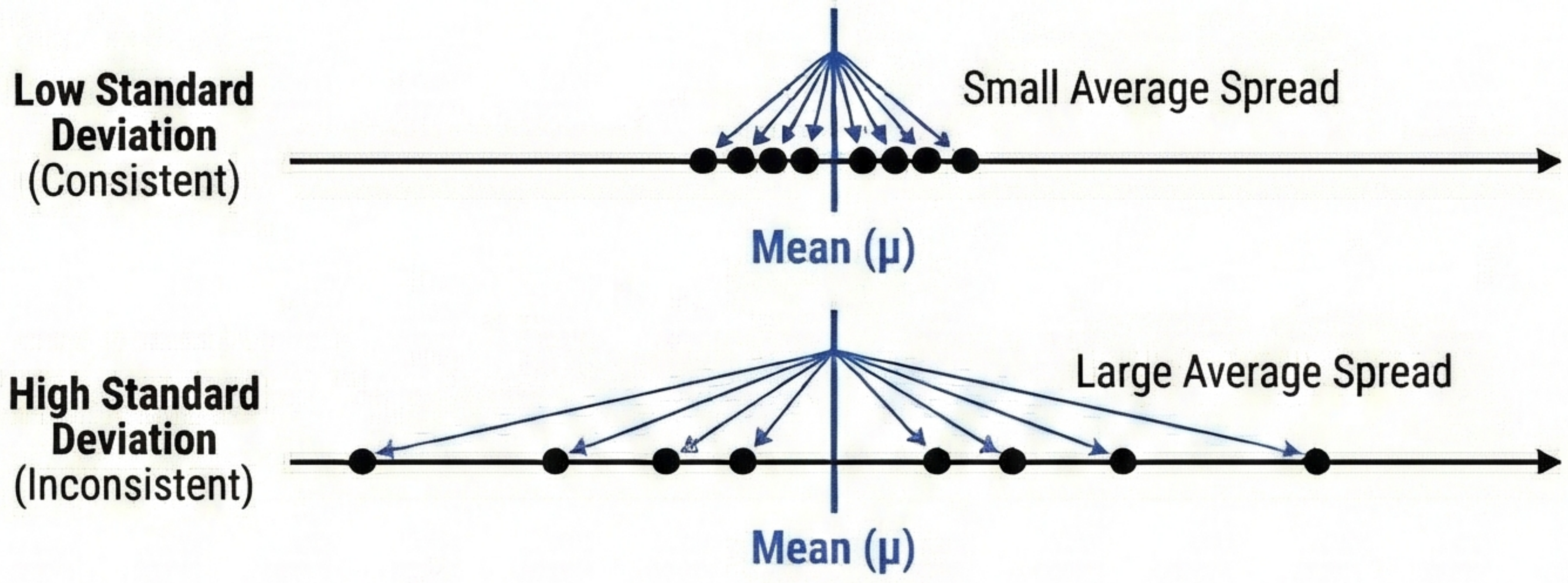
- **Zero Variance ($\sigma^2 = 0$):** All data points are identical to the mean (No Spread).
- **Positive Variance ($\sigma^2 > 0$):** Data points are spread out. Larger values indicate greater dispersion.

Zbigniew Galar

STANDARD DEVIATION (σ)

Quantifying Data Spread from the Mean

1. CORE CONCEPT: SPREAD VS. CONSISTENCY



2. CALCULATION FLOW (STEP-BY-STEP)

1. Find the Mean (μ)

$$\mu = \frac{\sum x}{N}$$

2. Calculate Deviations (x − μ)

-2	0	+3

Distances from Mean

3. Square Deviations $(x - \mu)^2$

4	0	9

⚠ **Alert:** Squaring handles negative values (Positive Only)

4. Average the Squares (Variance, σ^2)

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

$\sigma^2 = 6.5$ (Squared Units)

5. Take Square Root (Standard Deviation, σ)

$$\sigma = \sqrt{\sigma^2}$$

$\sigma \approx 2.55$ (Original Units)

3. FORMULAS (POPULATION vs. SAMPLE)

Population (σ)

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

N = Total Population Size

μ = Population Mean

Sample (s)

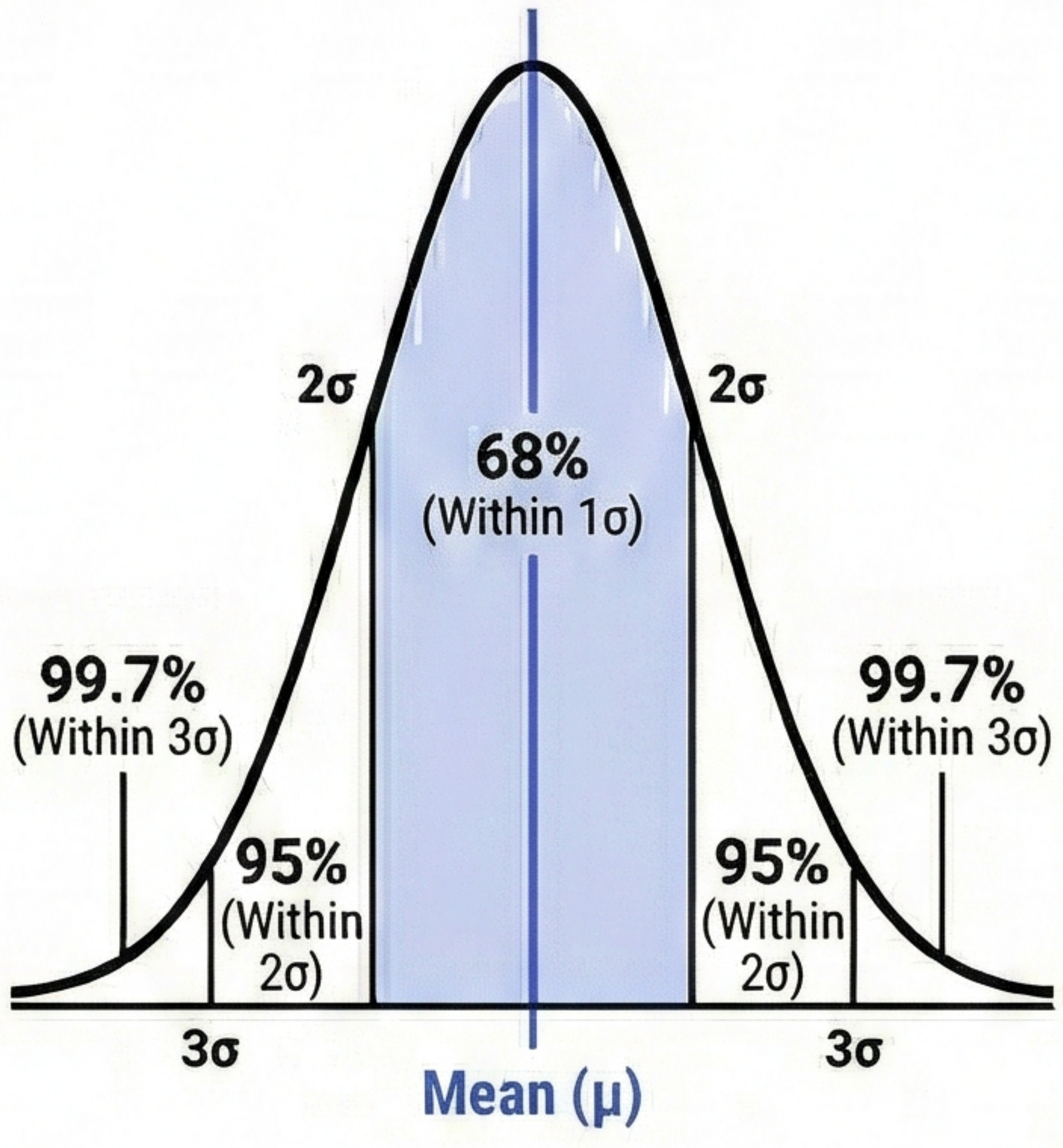
$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{(n - 1)}}$$

n = Sample Size

$\bar{x}$ = Sample Mean

Vermilion Alert: (n-1) corrects bias

4. KEY IMPLICATION: THE NORMAL CURVE



Zbigniew Gaiar