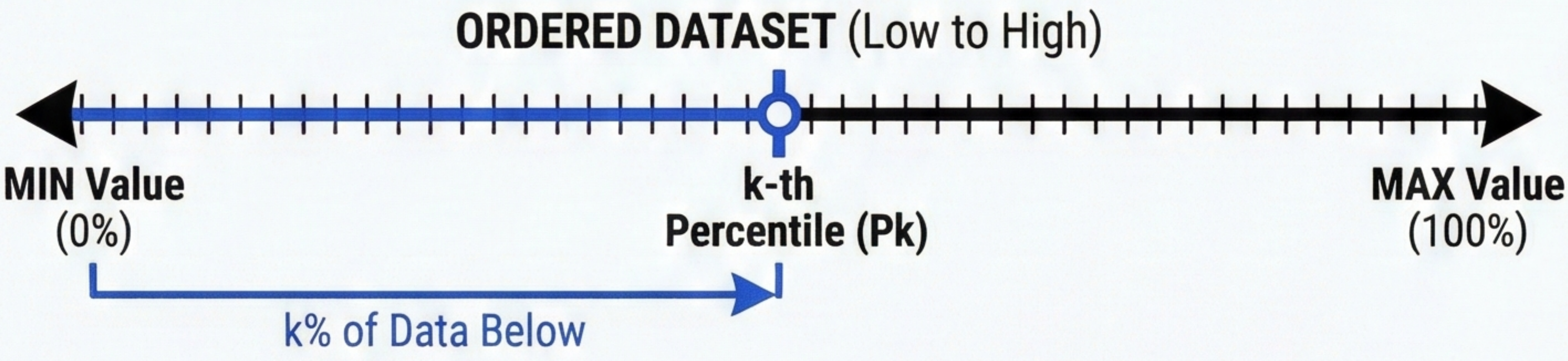
QUARTILES & PERCENTILES: DIVIDING ORDERED DATA

Understanding Data Position and Distribution for Management Analysis

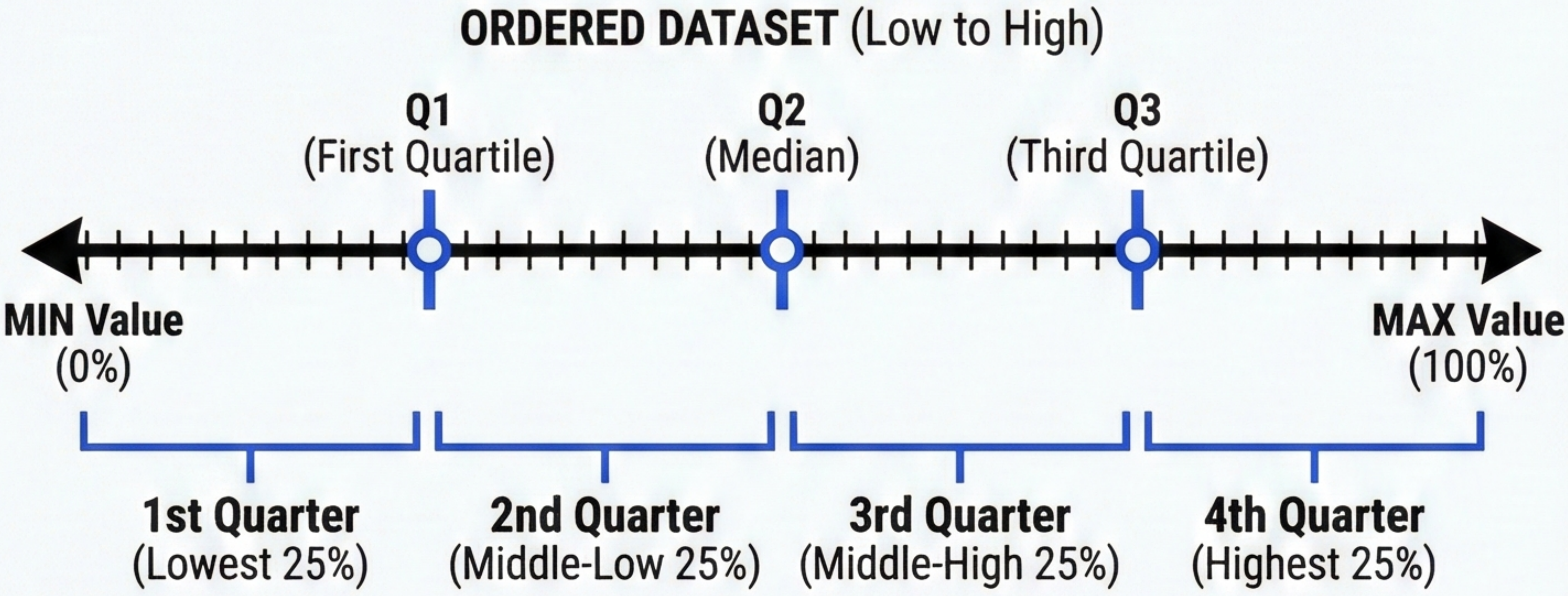
1. FOUNDATION: PERCENTILES (Pk)

Values that divide an ordered dataset into 100 equal parts.
Pk is the value below which k% of observations fall.

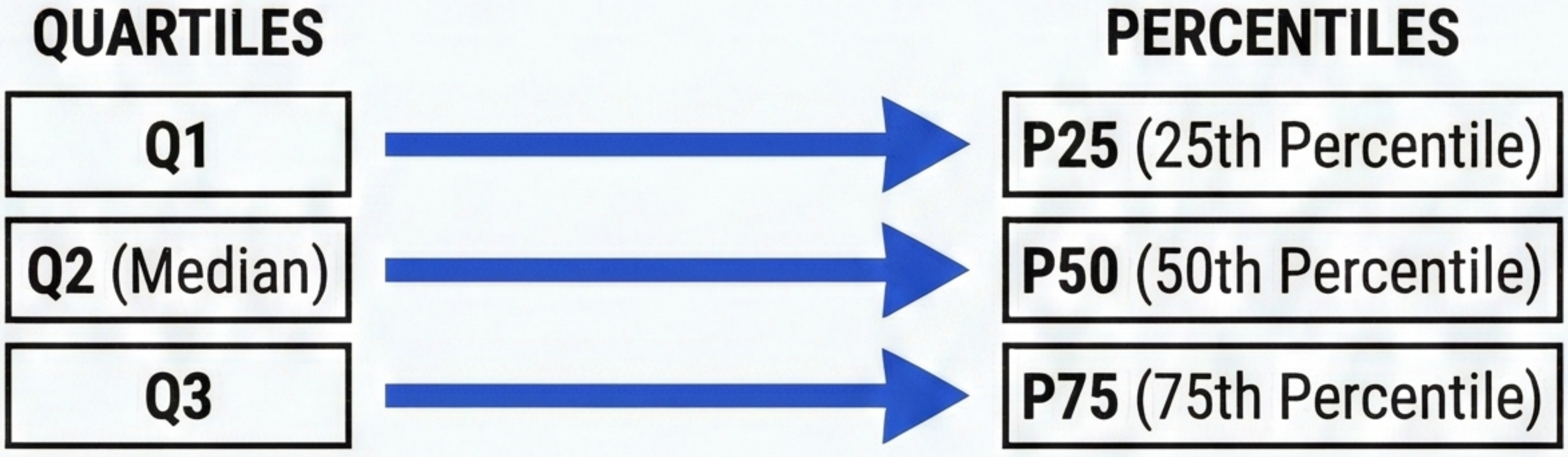


2. CORE CONCEPT: QUARTILES (Q)

Specific percentiles that divide the data into four equal quarters.



3. SYNTHESIS: RELATIONSHIP & KEY METRICS



IQR = Q3 - Q1 (Spread of the Middle 50%)

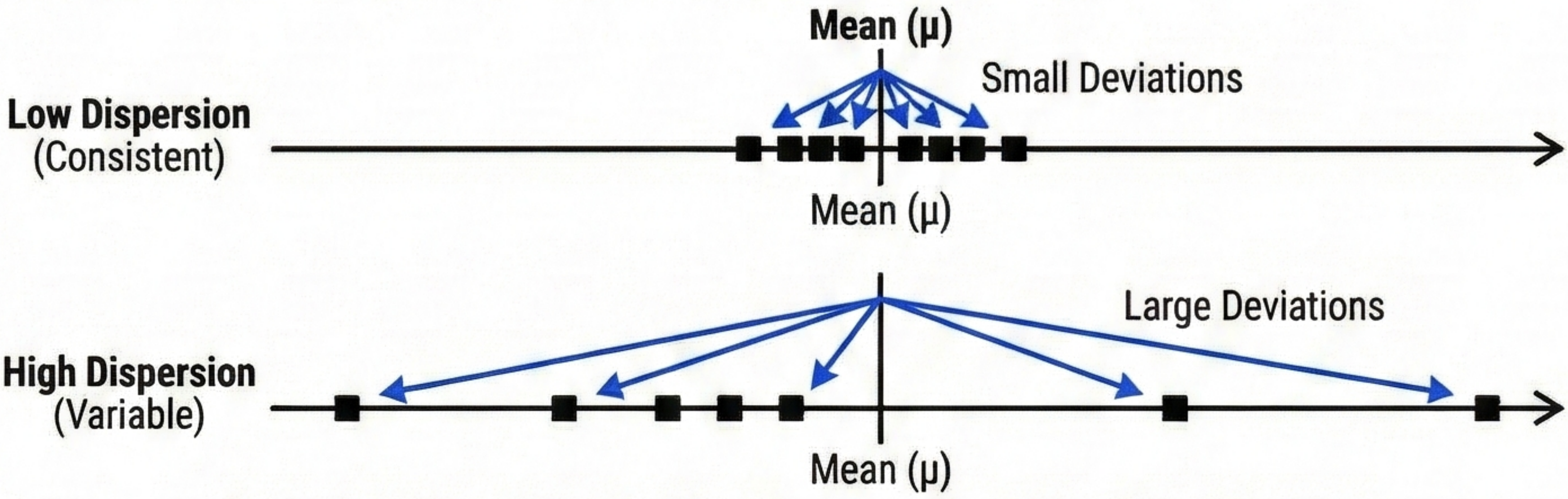
⚠ **NOTE:** Values outside Q1 - 1.5*IQR or Q3 + 1.5*IQR are often flagged as POTENTIAL OUTLIERS.

Zbigniew Galar

DISPERSION MEASURES: QUANTIFYING DATA SPREAD

Understanding variability, risk, and consistency in datasets for management decision-making.

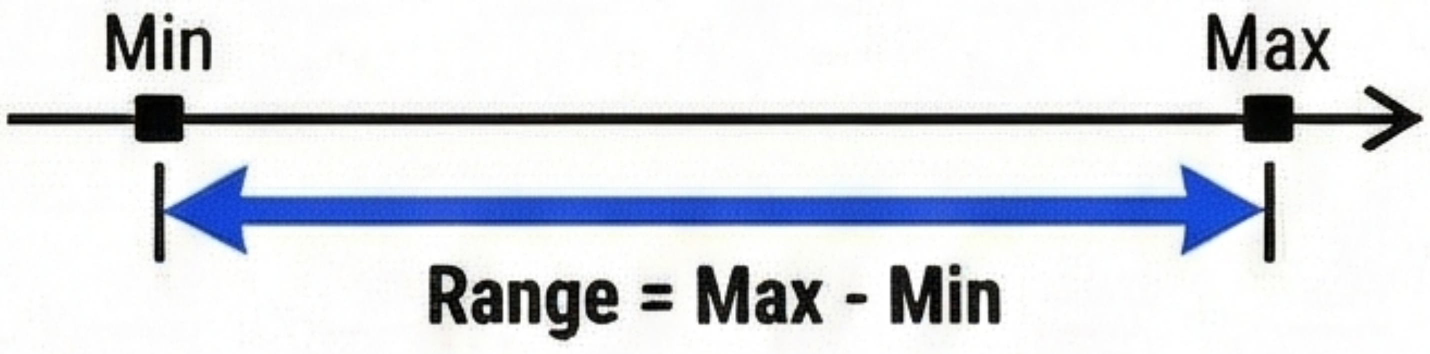
1. THE CONCEPT: LOW vs. HIGH DISPERSION



Dispersion measures quantify how far data points spread from the center.

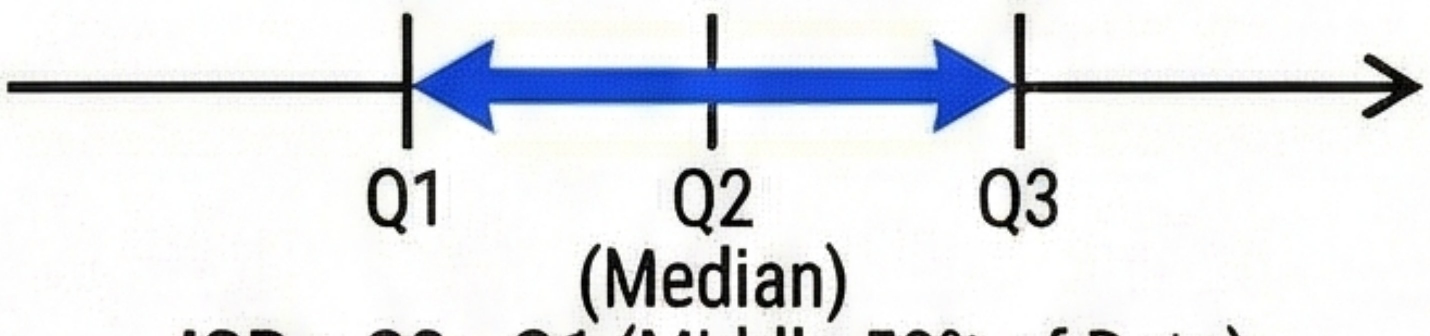
2. POSITIONAL MEASURES (ABSOLUTE DISTANCE)

RANGE



⚠ Highly sensitive to outliers.

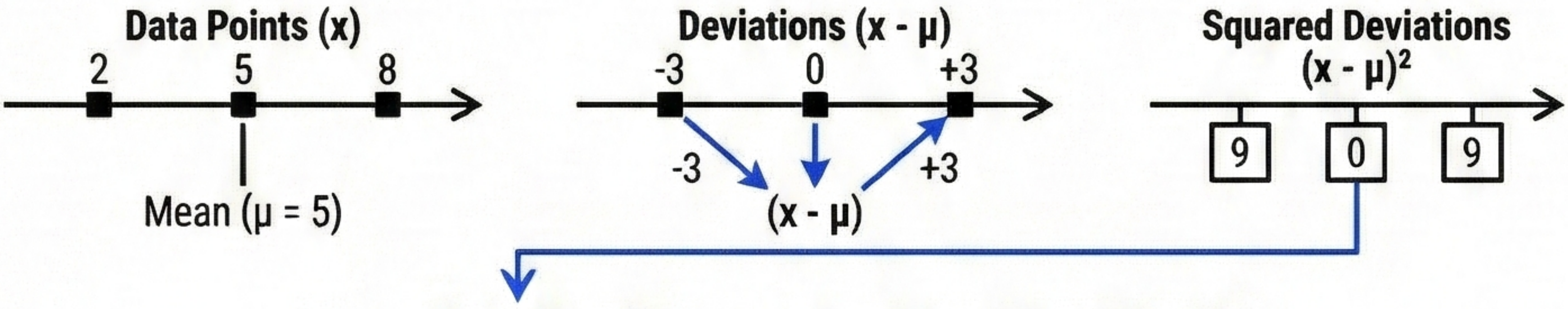
INTERQUARTILE RANGE (IQR)



IQR = Q3 - Q1 (Middle 50% of Data)

Robust to outliers. Focuses on the core.

3. DEVIATION-BASED MEASURES (FROM MEAN)



4. VARIANCE (σ^2)

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

$(9+0+9)/3 = 6$

(Squared Units)

STANDARD DEVIATION (σ)

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

$\sigma = \sqrt{6} \approx 2.45$

(Original Units)

The average distance of points from the mean.

⚠ Sensitive to outliers due to squaring.

4. SUMMARY & APPLICATION GUIDE

MEASURE	KEY CHARACTERISTIC	OUTLIER SENSITIVITY	BEST USE CASE
Range	Simplest, max vs. min spread.	High ⚠	Quick snapshot, error checking.
IQR	Spread of the middle 50%.	Low ✔	Skewed data, presence of outliers.
Variance	Average squared deviation.	High ⚠	Mathematical modeling, advanced stats.
Std. Deviation	Average deviation in original units.	High ⚠	Normal distribution, risk assessment.

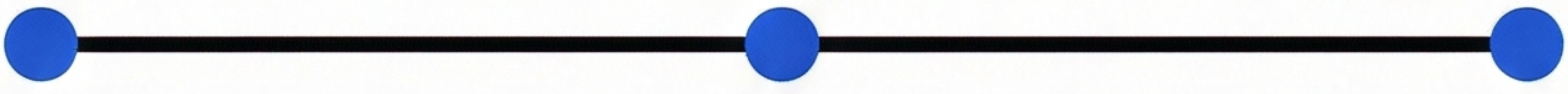
Zbigniew Galar

PEARSON CORRELATION COEFFICIENT (r)

Quantifying Linear Association Between Two Continuous Variables
(A Measure of Relationship Strength & Direction)

MODULE 1: CORE CONCEPT - THE RANGE OF 'r'

THE SCALE: FROM PERFECT NEGATIVE TO PERFECT POSITIVE



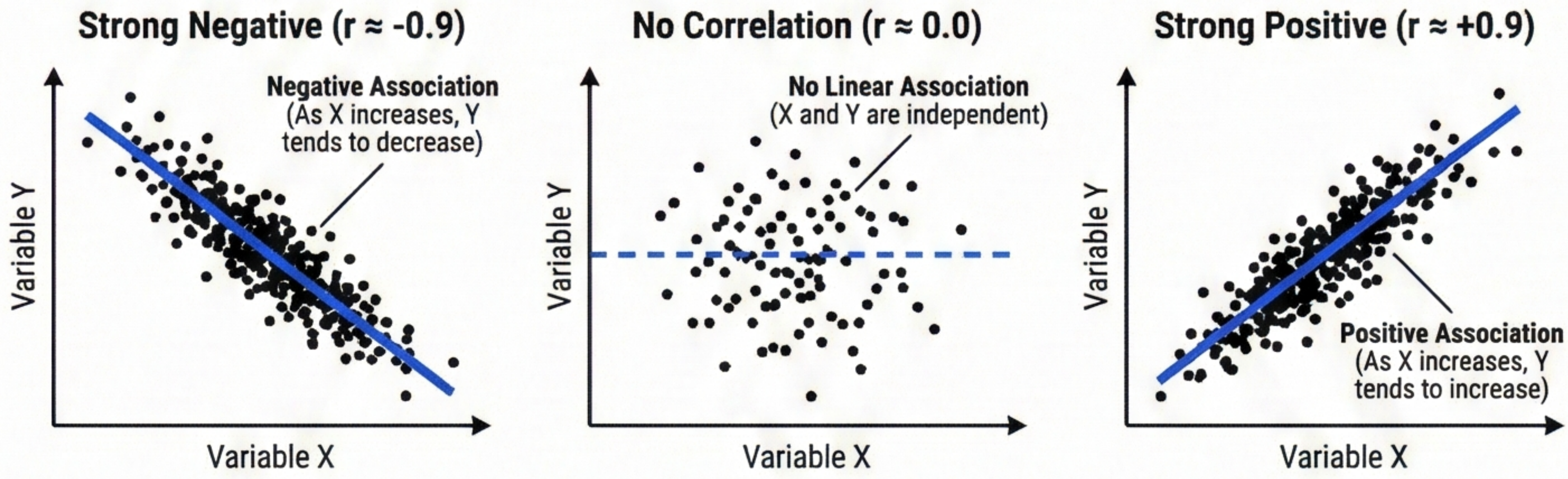
-1.0 (Perfect Negative) | **0.0** (No Correlation) | **+1.0** (Perfect Positive)

The coefficient 'r' is a normalized value between -1 and +1, indicating both the strength and direction of the linear relationship.

Range of Possible Values

MODULE 2: VISUAL INTERPRETATION (Scatter Plots)

VISUALIZING RELATIONSHIPS: STRENGTH & DIRECTION



MODULE 3: THE FORMULA (A Deconstructed View)

UNDERSTANDING THE MECHANISM: COVARIANCE vs. VARIANCE

Covariance (X, Y)

$$r = \frac{\boxed{\dfrac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{n-1}}}{\boxed{\sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n-1}}} * \boxed{\sqrt{\dfrac{\sum(y_i - \bar{y})^2}{n-1}}}}$$

Measures Joint Variability
(How X and Y move together)

Measures Spread of X
(Standard Deviation of X)

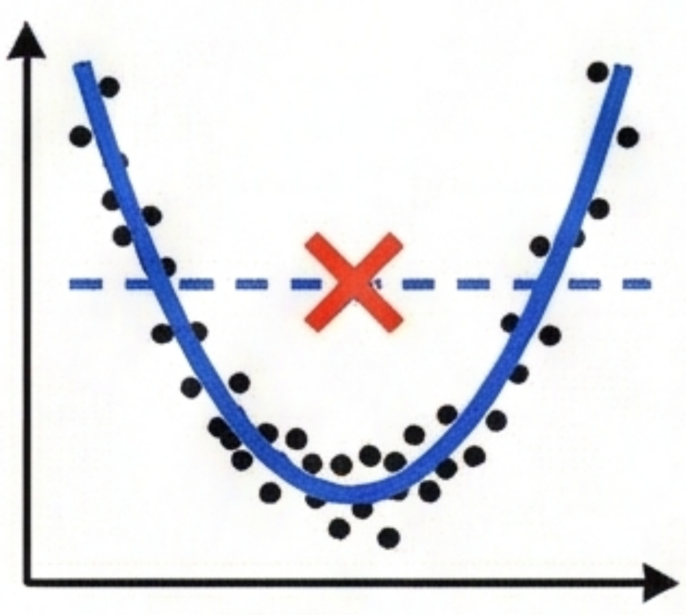
Measures Spread of Y
(Standard Deviation of Y)

The formula standardizes the joint variability by the individual spreads, making 'r' unitless and comparable.

MODULE 4: CRITICAL LIMITATIONS (Management Insight)

LIMITATIONS: WHAT 'r' DOES NOT CAPTURE

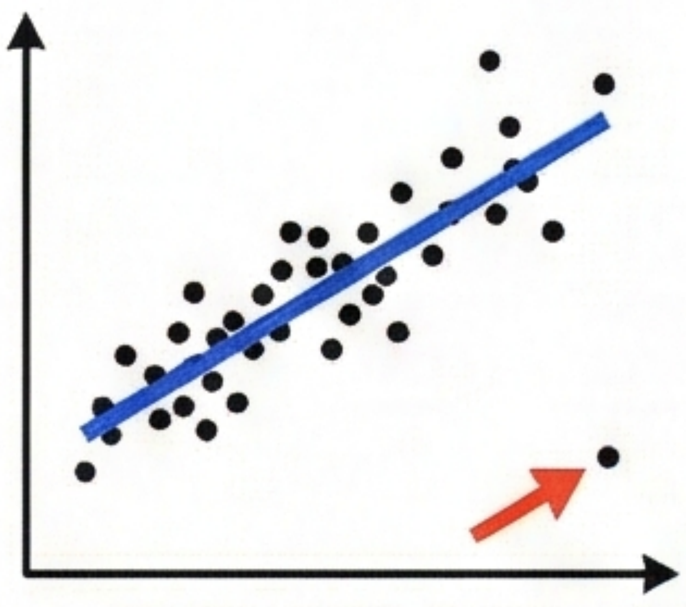
NON-LINEAR RELATIONSHIPS



Pearson 'r' only measures **linear** patterns. It will fail to detect strong non-linear associations (like this curved relationship).

⚠ **Alert:** Always plot the data first to check for linearity.

OUTLIER SENSITIVITY



Pearson 'r' is highly sensitive to outliers. A single extreme value can significantly distort the coefficient and mislead the interpretation.

⚠ **Alert:** Outliers can inflate or deflate 'r' substantially.

Zbigniew Galar

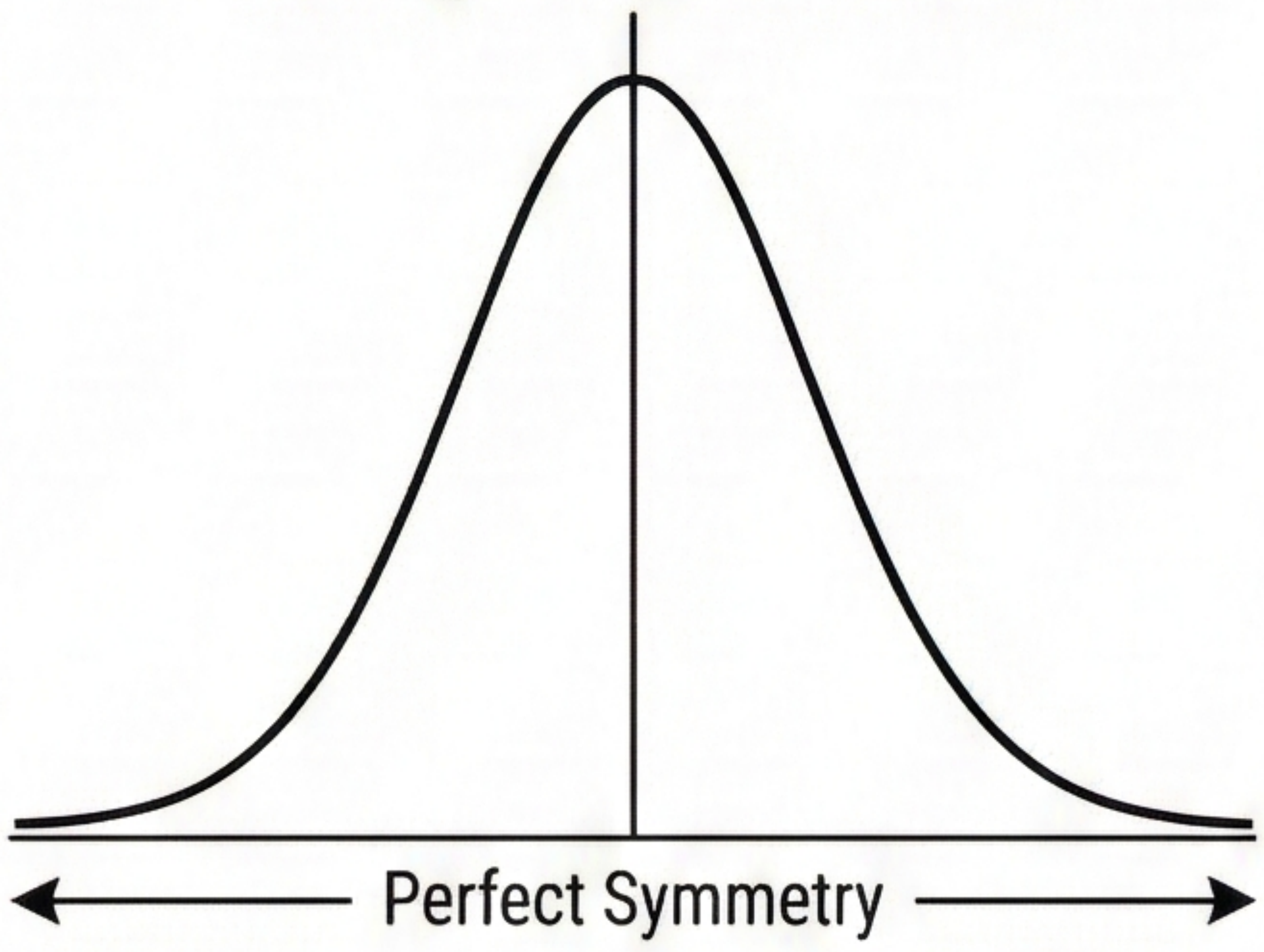
NORMAL DISTRIBUTION (GAUSSIAN): THE BELL CURVE MODEL

A Symmetrical Probability Distribution Defined by Mean (μ) and Standard Deviation (σ)

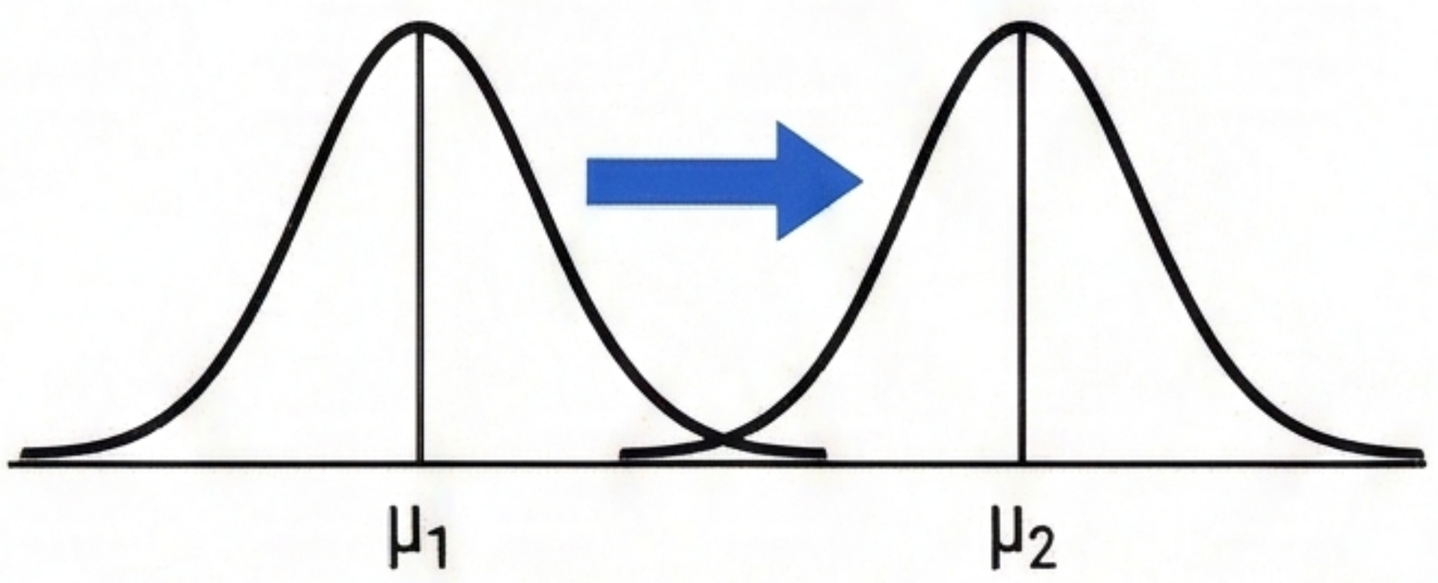
MODULE 1: THE FOUNDATION (Shape & Symmetry)

CORE CONCEPT: SYMMETRY & CENTER

Mean (μ) = Median = Mode



MEAN (μ) - Location

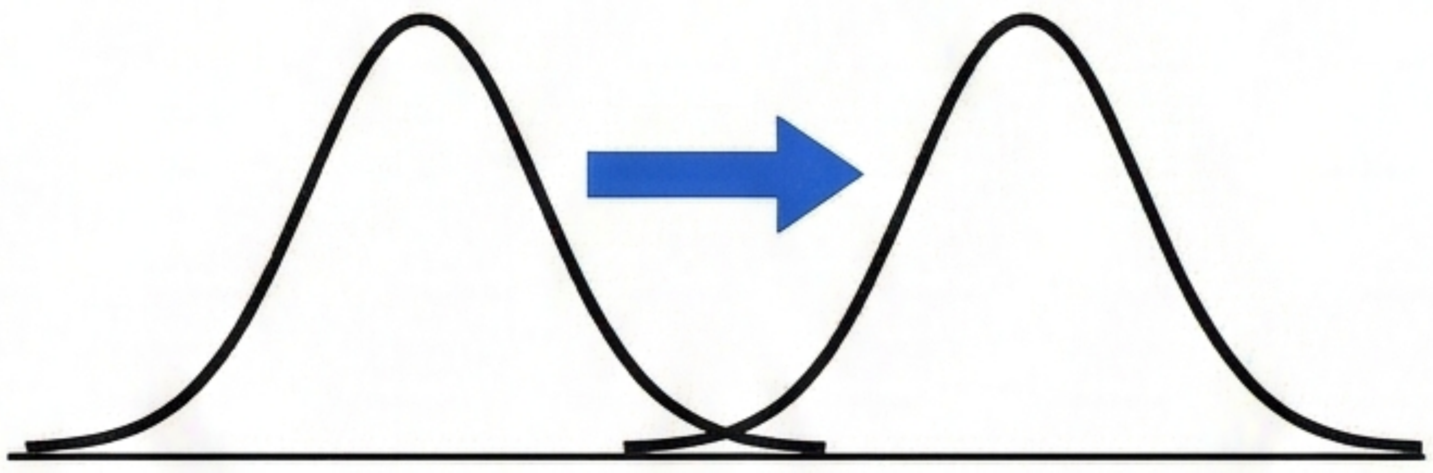


Determines Center/Position on the Axis

MODULE 2: THE PARAMETERS (Controlling the Form)

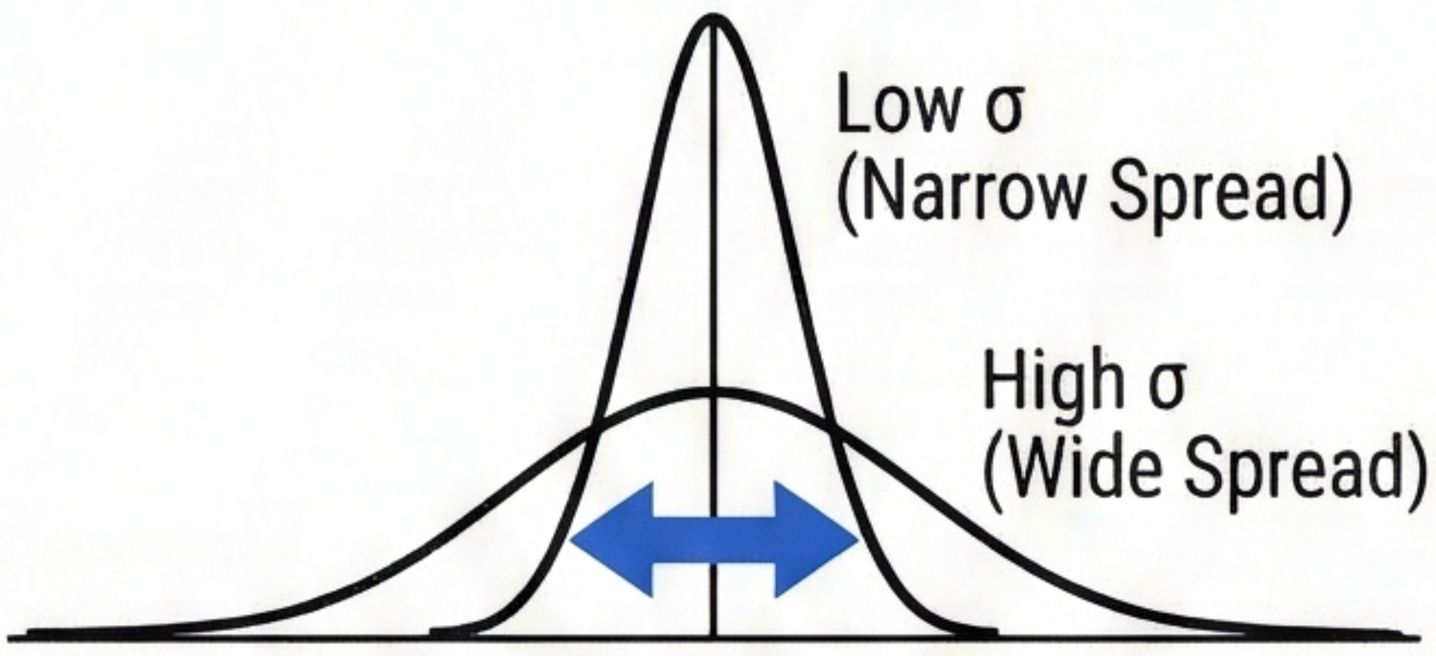
TWO DRIVERS: LOCATION & SPREAD

MEAN (μ) - Location



Data clusters around the central mean, with tails tapering equally in both directions.

STANDARD DEVIATION (σ) - Spread

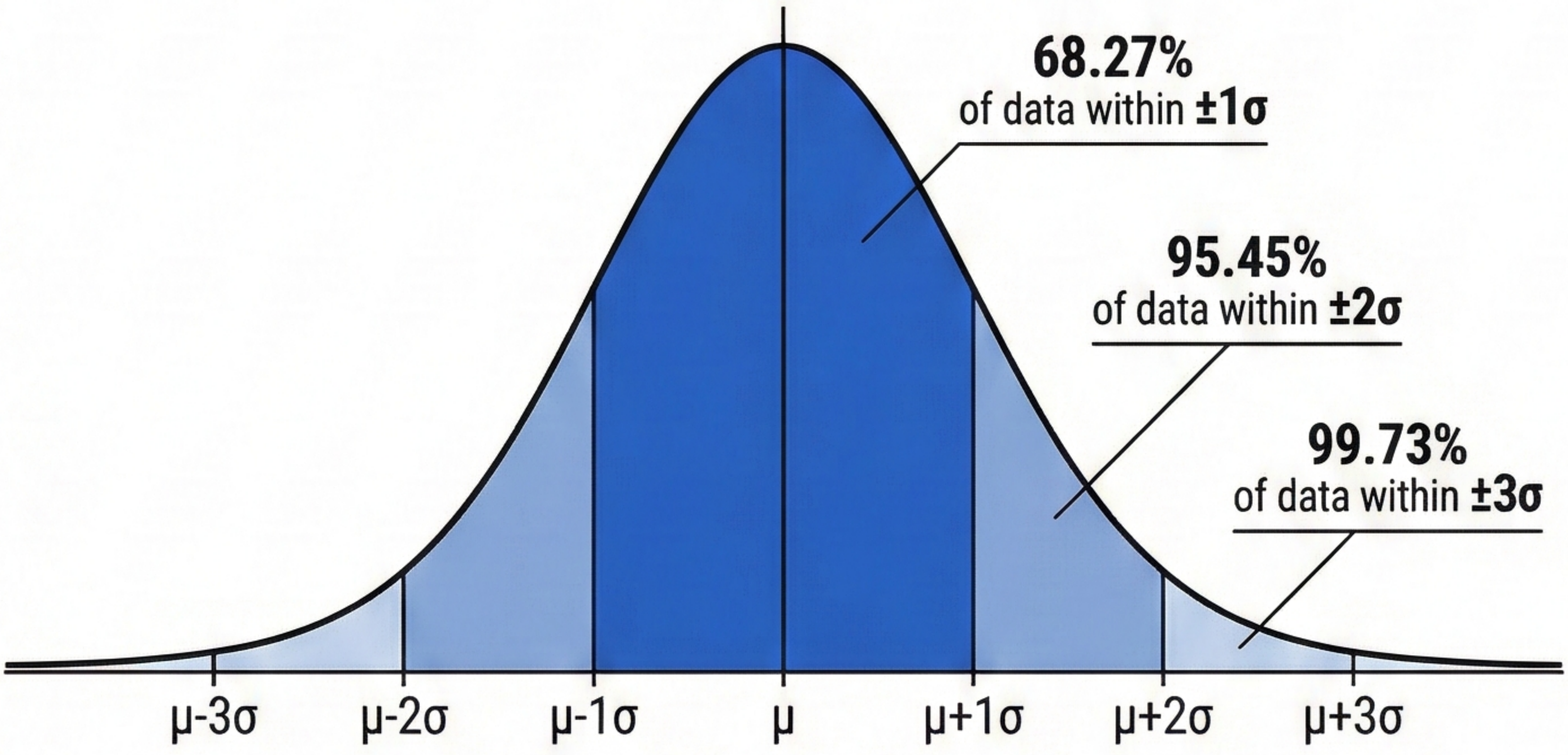


Quantifies Variability/Dispersion

Determines Width & Flatness

MODULE 3: THE EMPIRICAL RULE (68-95-99.7% Principle)

THE 68-95-99.7 RULE: PREDICTABLE PROBABILITIES



⚠ **TAIL ALERT:** Asymptotic tails never touch the axis.
Extreme outliers (beyond ±3σ) are rare but possible.

Zbigniew Galar