

Understanding Distributions: Center, Spread, and Outliers – Beyond the Average



- **"Average" alone misleads:** salaries, delivery times, and customer spend hide extreme values



- **Three pillars:** center (typical value), spread (variability), and shape/outliers (the edges)



- **Core skill:** reading numbers to see the typical, the variable, and the extreme



- **Course roadmap:** central tendency → dispersion → outliers → shape → visuals → traps



CENTRAL
TENDENCY



DISPERSION



OUTLIERS



SHAPE



VISUALS

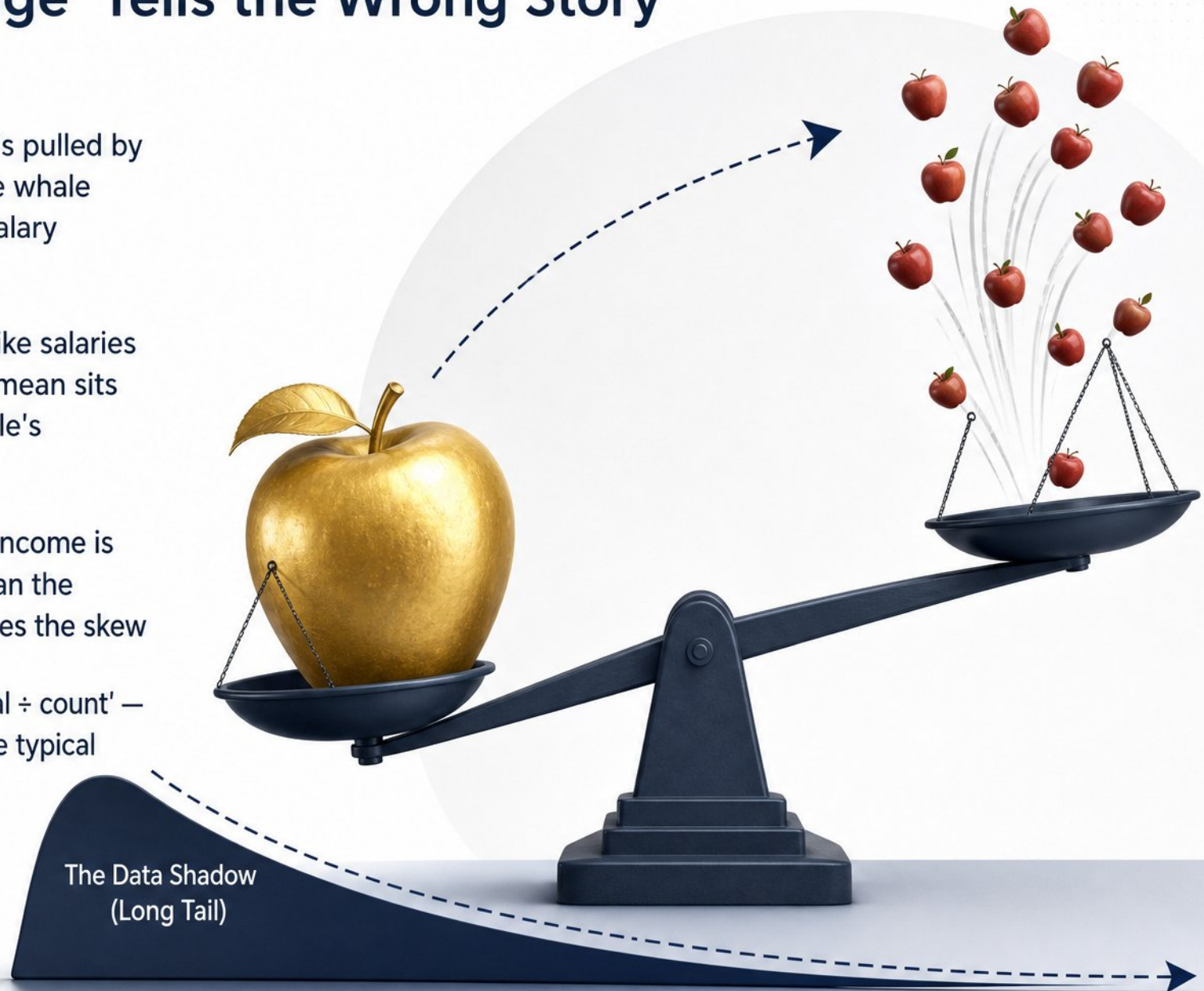


TRAPS

Why the Mean Lies:

When 'Average' Tells the Wrong Story

- The arithmetic mean is pulled by extreme values — one whale customer or C-suite salary dramatically shifts it
- In right-skewed data like salaries or delivery times, the mean sits well above most people's experience
- U.S. mean household income is roughly 40% higher than the median; the gap exposes the skew
- The mean answers 'total ÷ count' — it is a poor proxy for the typical individual experience

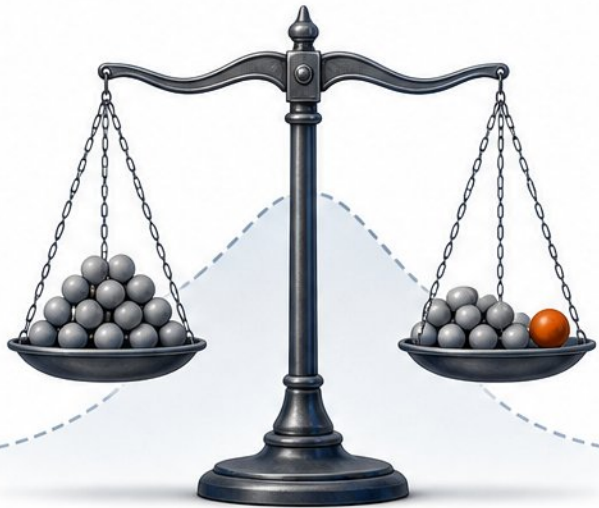


Measures of Central Tendency:

Mean, Median, and Mode — Choosing the Right One

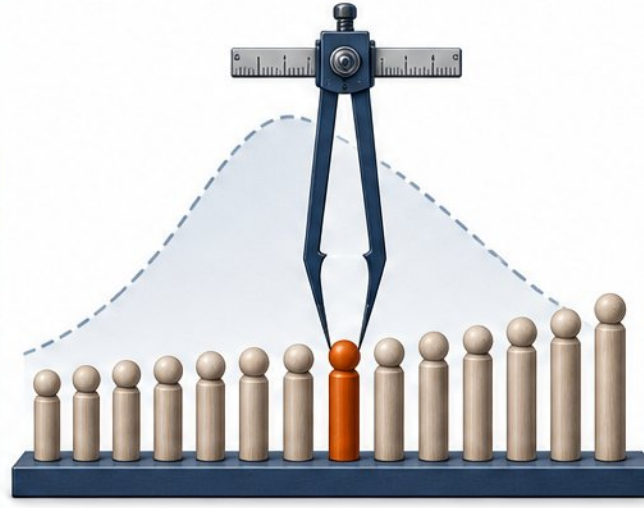
MEAN

Balances all values



MEDIAN

Finds the middle ground

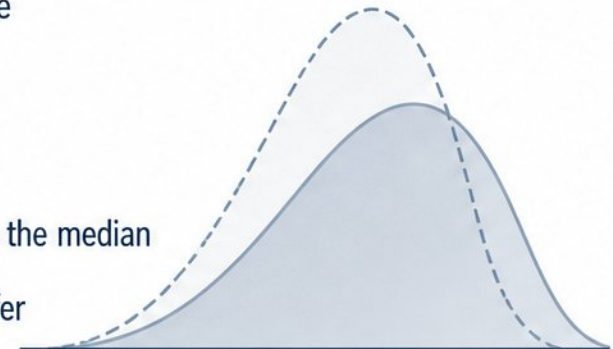


MODE

Highlights what happens most



- **Mean** = $\text{sum} \div \text{count}$; pulled by extremes — a \$1M salary can inflate the company average
- **Median** = middle value when sorted; ignores outliers — the honest 'typical' salary
- **Mode** = most frequent value; shows majority — best for categorical data or spotting clusters
- **Mean-Median gap** as a **diagnostic**: if mean is >20% higher than median, report both and lead with the median
- **In right-skewed data** (salaries, delivery times), mean overstates typical experience — median is safer



Understanding Spread: Range, IQR, Variance, and Standard Deviation



- Range = Max – Min:
simple but shattered by
a single extreme value



- IQR = Q3 – Q1:
spread of the middle 50%,
robust partner for the median



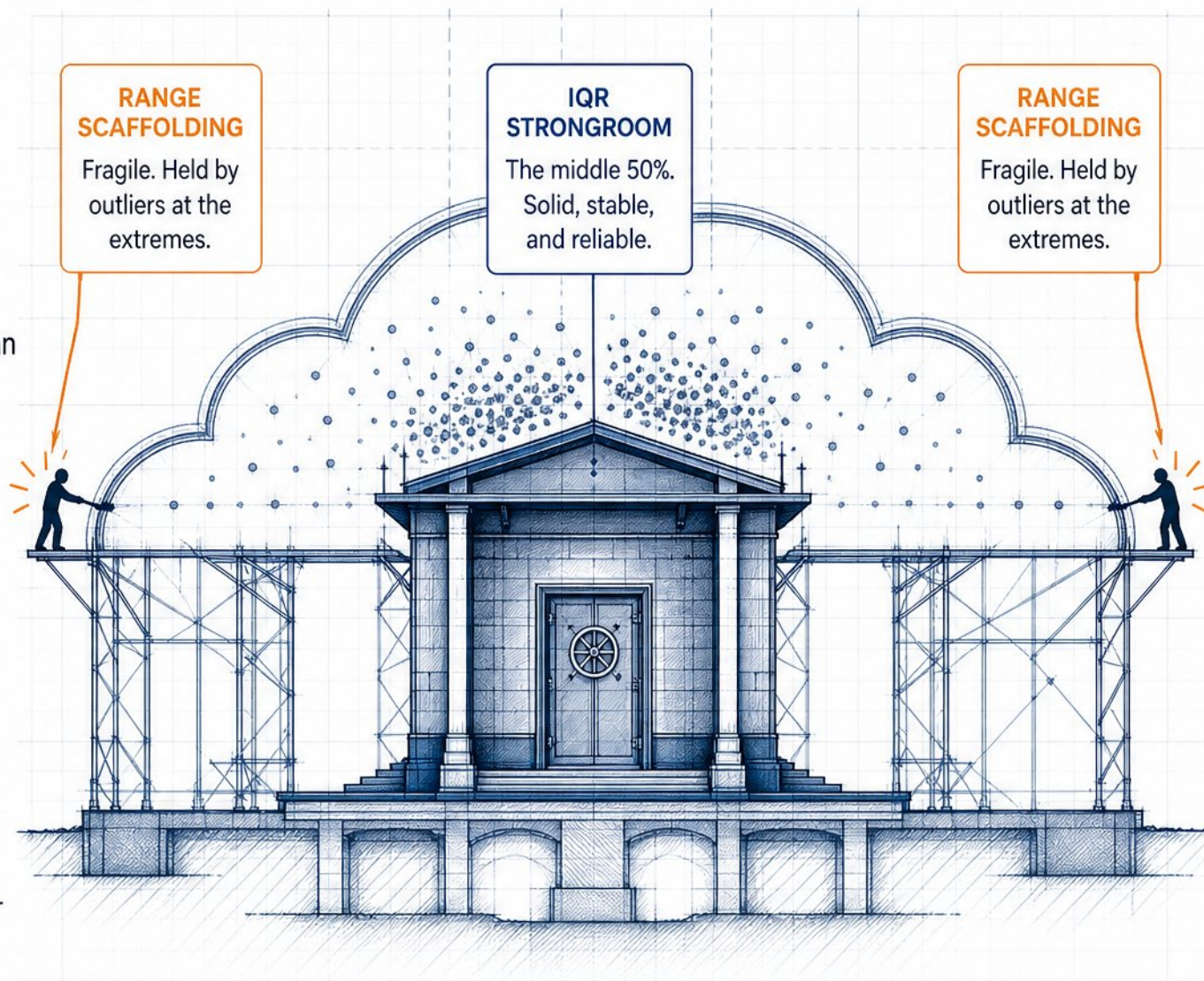
- Variance: average
squared distance from
mean; units are squared
(dollars²)



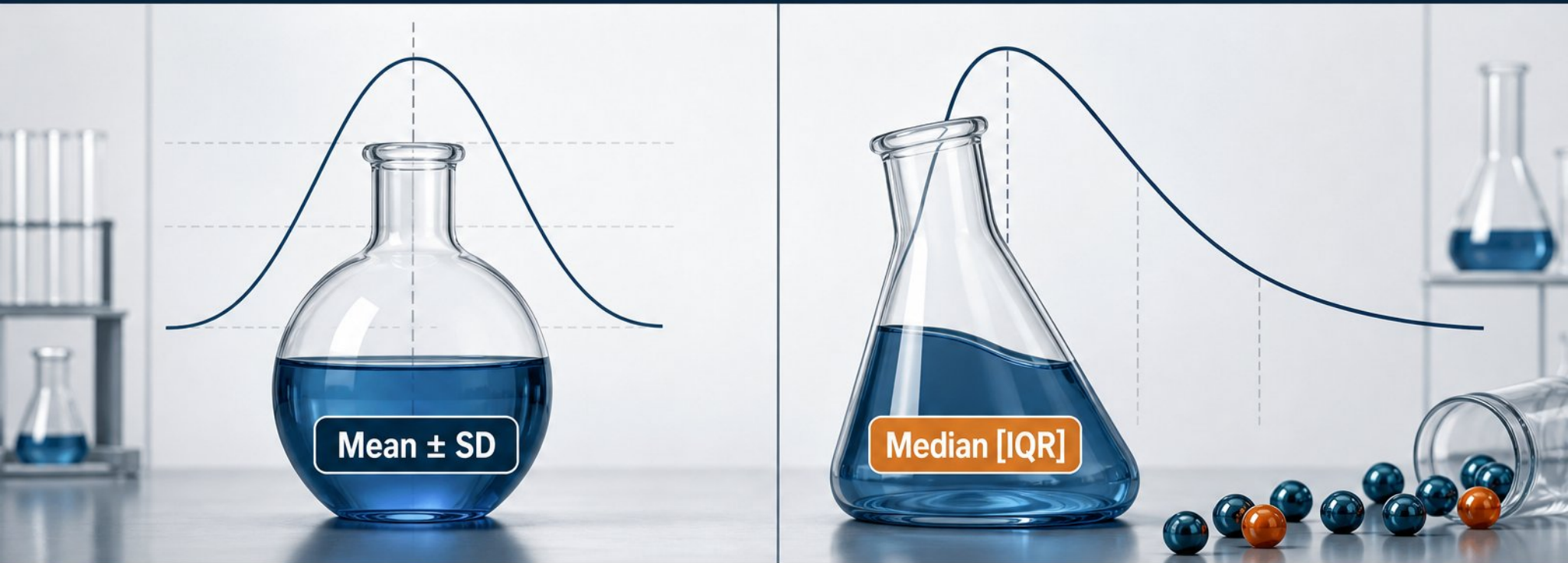
- Standard Deviation (SD):
square root of variance,
in original units (dollars,
minutes)



- Business use: 'typical
delivery is 20 min ±5 min' —
enables buffer sizing and
risk communication



Reporting Rules: Mean + SD or Median + IQR?



- Symmetric & outlier-free data → report **mean $\pm$ standard deviation**



- Skewed or heavy-tailed data → report **median (IQR)** or **median [Q1-Q3]**



- Never give a center measure alone — **always pair it with a spread measure**



- Template: “**Median is ___, with half falling between ___ and ___ ;**
the mean of ___ reflects the influence of the tail”

Detecting and Interpreting Outliers



- **Outliers:** values inconsistent with the rest — can be legitimate signals or errors



- **IQR method:** robust, uses medians; works for any distribution shape



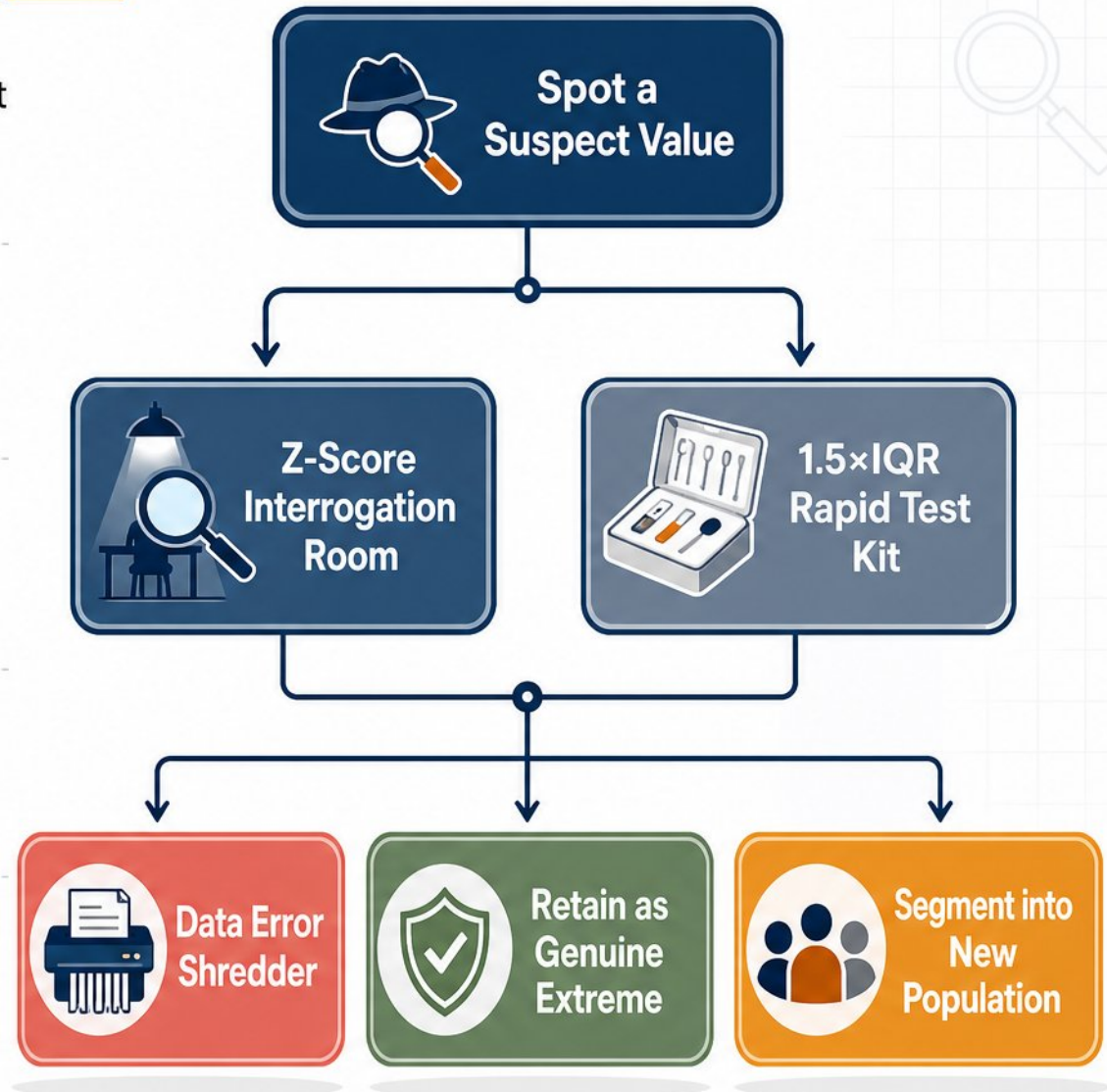
- **Z-score:** needs normal data; mean & SD affected by outliers themselves



- **Modified Z-score (MAD):** robust for small or skewed datasets



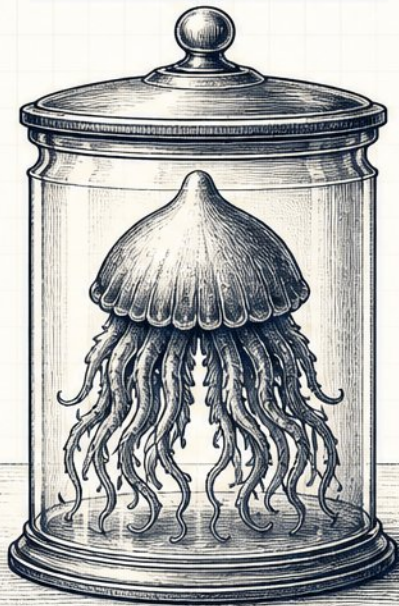
- **Decision:** investigate first → correct (error), retain (genuine), or segment (different population)



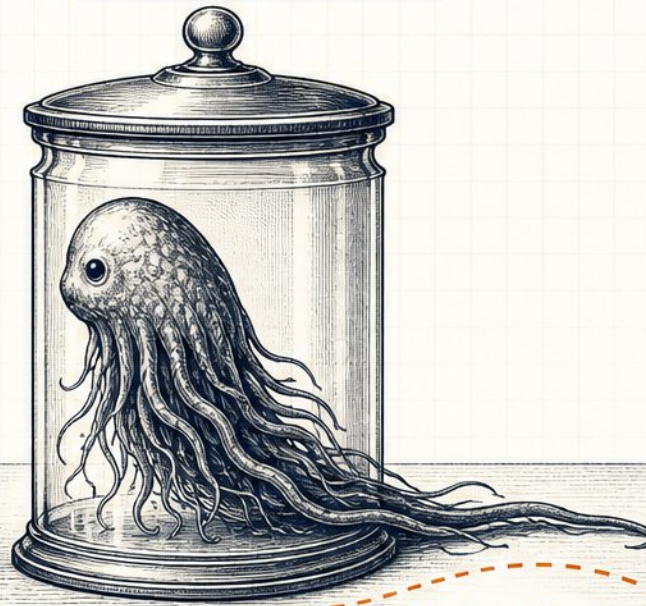
Skew and Shape: Reading the Whole Distribution

- Symmetry vs. right-skew (mean > median) vs. left-skew (mean < median)
- Modality: unimodal, bimodal (hidden by box plots), multimodal — split before summarizing
- Shape decides the summary: symmetric → mean/SD; skewed → median/IQR; bimodal → segment first
- Example: right-skewed wait times — the average hides the painful experience of the worst-hit customers

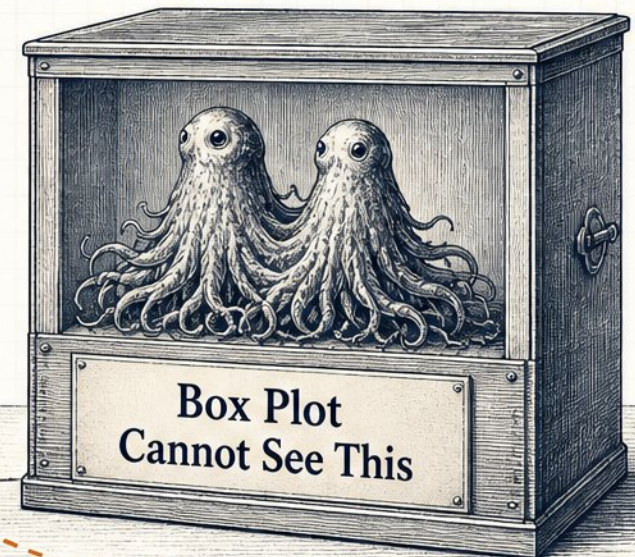
SYMMETRIC
(BELL-SHAPED)



RIGHT-SKEWED
(LONG TAIL TO THE RIGHT)



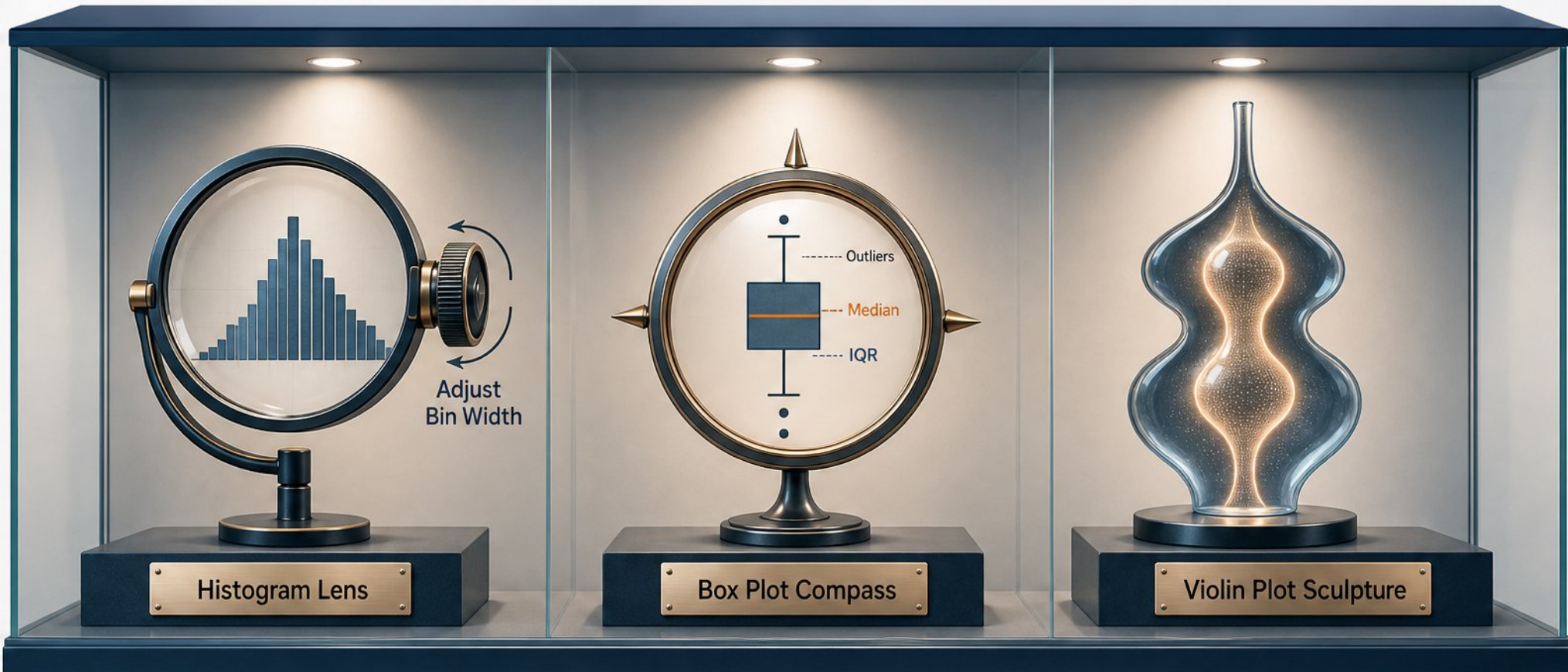
BIMODAL
(HIDDEN BY BOX PLOTS)



The Data Shadow
(Long Tail)



Visual Tools for Distribution Literacy

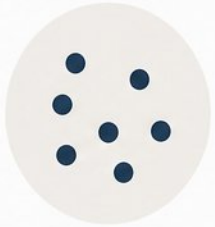


- Histograms reveal shape, but bin width changes the story — always try multiple sizes
- Box plots: compact, compare many groups, show median/IQR/outliers, but hide bimodality
- Violin plots: box plot + density curve, expose multi-peaks and skew; need $n > 30$ per group
- Audience-aware choice: box plots for executive dashboards; violin plots for analytical deep dives

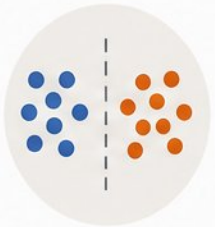
Common Traps When Interpreting Summary Statistics



- **Simpson's Paradox:** An aggregate trend can reverse when split by subgroup — overall conversion rises while every segment falls.



- **Small sample deception:** With few data points, the mean appears stable but is highly volatile. Always report sample size.



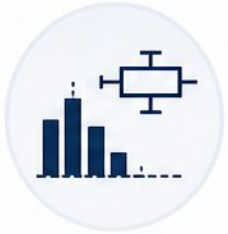
- **Bimodality hidden by averages:** A mean or median can sit between two real clusters, describing a customer who doesn't exist.



- **The single-number habit:** Dashboards showing only averages create false confidence. Include a spread measure and check segments.



Applying Distribution Thinking to Everyday Business Decisions



- Replace average sales with median + IQR to see true rep/store variability



- Set targets at P75/P90, not on an average half already exceed



- Use P95 delivery ≤ 2 days — mean-based SLAs hide tail pain



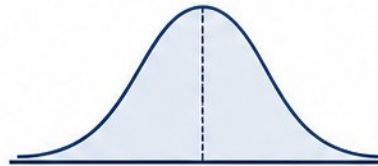
- Translate skew: "Most spend \$35, but whales pull the average to \$58"

THE HONEST DASHBOARD GADGET



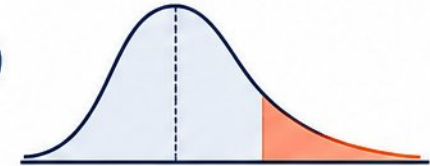
REALITY CHECK

MEAN (AVERAGE) SALES



VS.

MEDIAN SALES



MEANS CAN MISLEAD. MEDIANS REVEAL REALITY.



WORST-CASE PROTECTOR

p95 DELIVERY TIME

SLA



≤ 2 DAYS

Focus on the tail to protect the worst experience.



The Distribution Checklist:

What to Ask Before You Trust a Number



- **Ask 6 questions:** mean or median? Gap size? IQR or SD? Outliers checked? Shape (symmetric/skewed/bimodal)? Does segmentation change the story?



- **Audit dashboards:** add median and p90 for any KPI where the mean–median gap exceeds 20%.



- **Mental habit:** "What is this number hiding?" – every summary discards information; skill is knowing what was discarded.



Case Practice and Self-Assessment

- Mean salary \$112K but right-skewed → report median + IQR
- Mean delivery 24h but p95 is 72h → SLA on p95, not mean
- Overall conversion up 5%, every segment down → Simpson's Paradox
- Flag mean–median gap >20%; always segment before trusting aggregates



PersonWise.

PersonWise • AI digital-human course platform



Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/258bf581/public