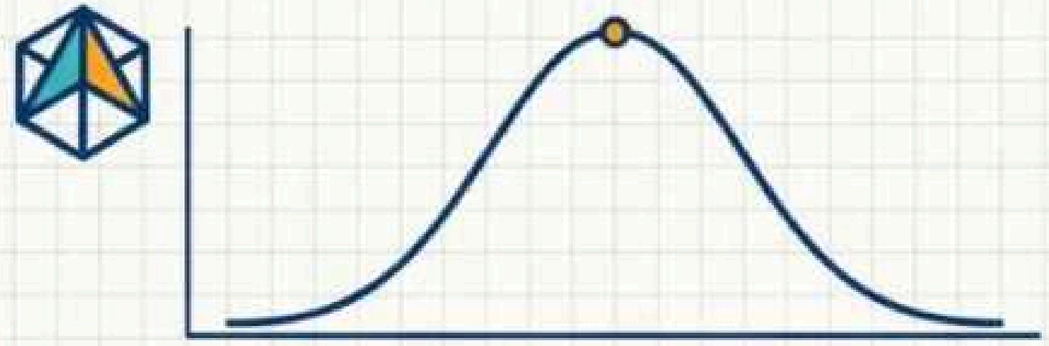
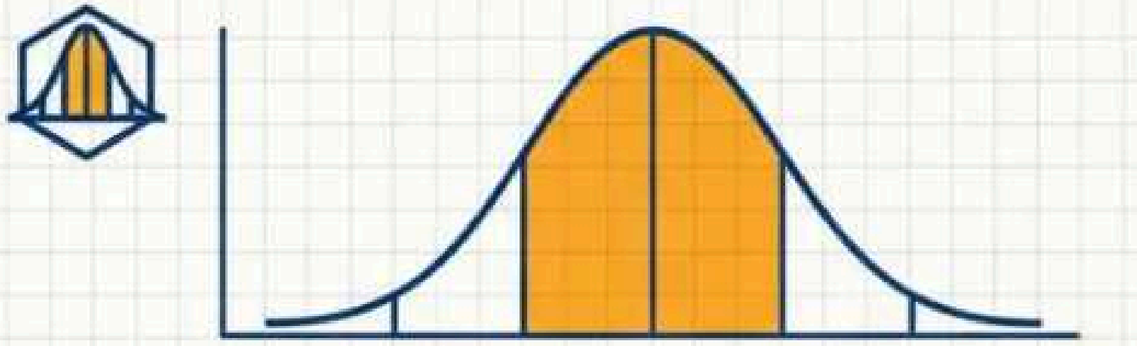
3.1 The Normal Distribution & Standard Scores

Using the famous bell-shaped curve and z-scores to compare values from different scales, find percentiles, and determine where the normal model fits (and where it fails).



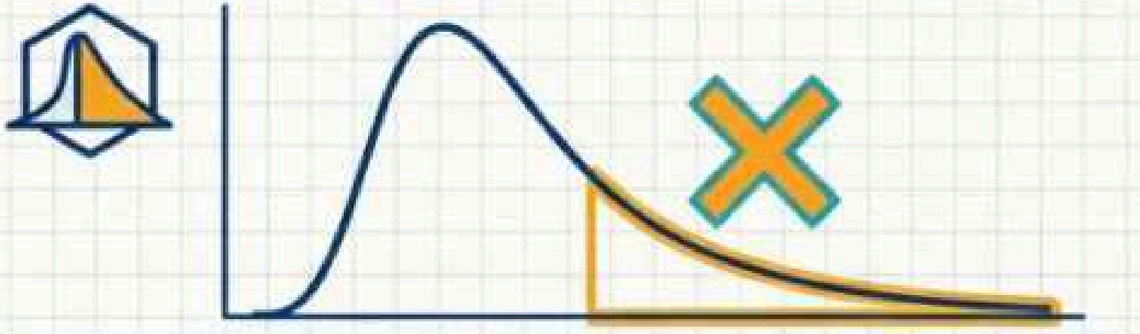
The Normal Distribution



The Empirical Rule

$$\div \frac{X - \mu}{\sigma} = Z\%$$

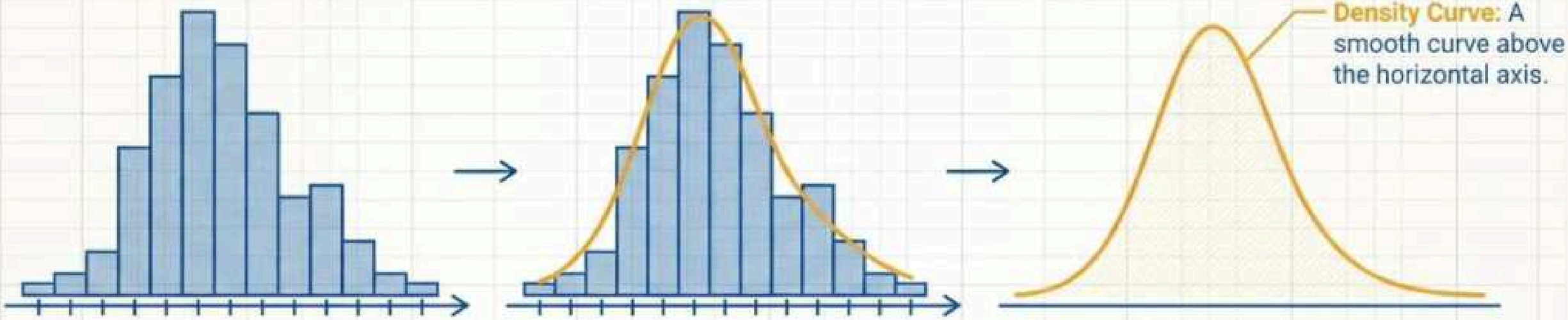
Z-Scores & Percentiles



Model Limits

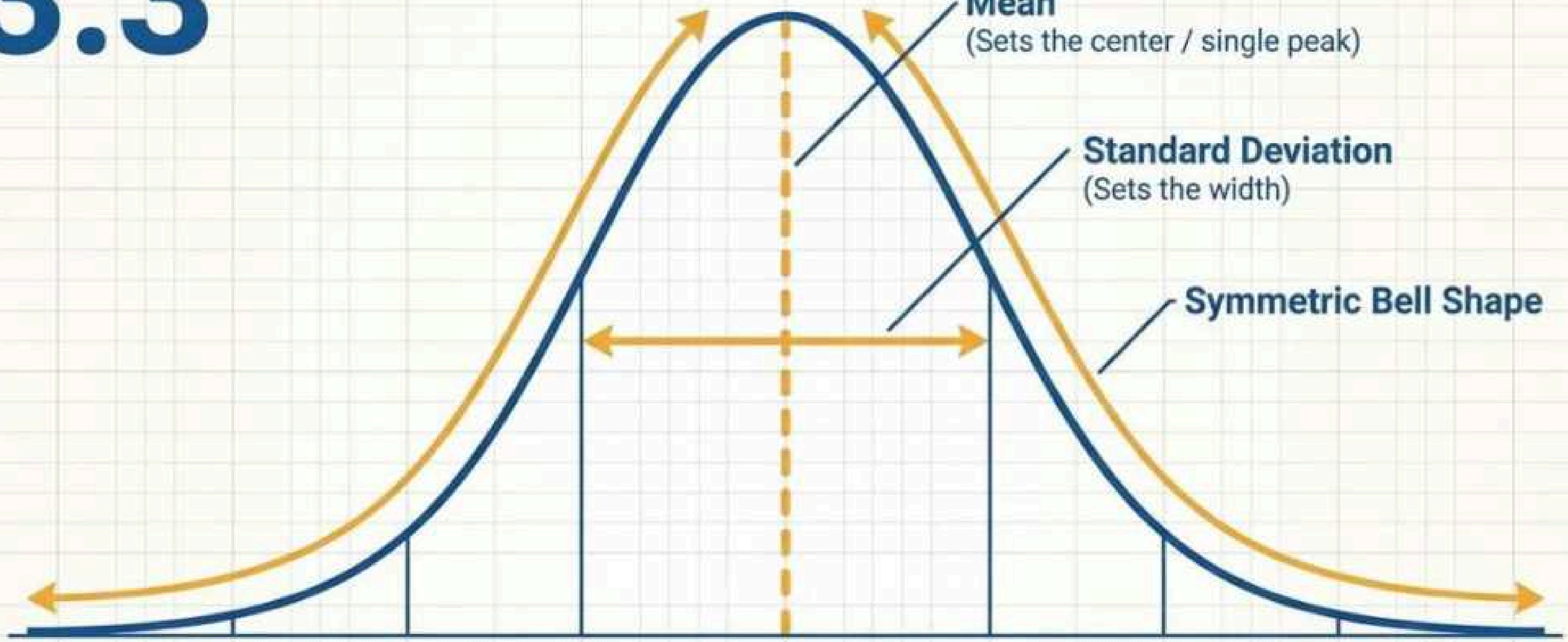
3.2

From Regular Histogram to Idealized Model



1) Total Area = Exactly 1
2) Area between two values = Proportion of observations falling between them

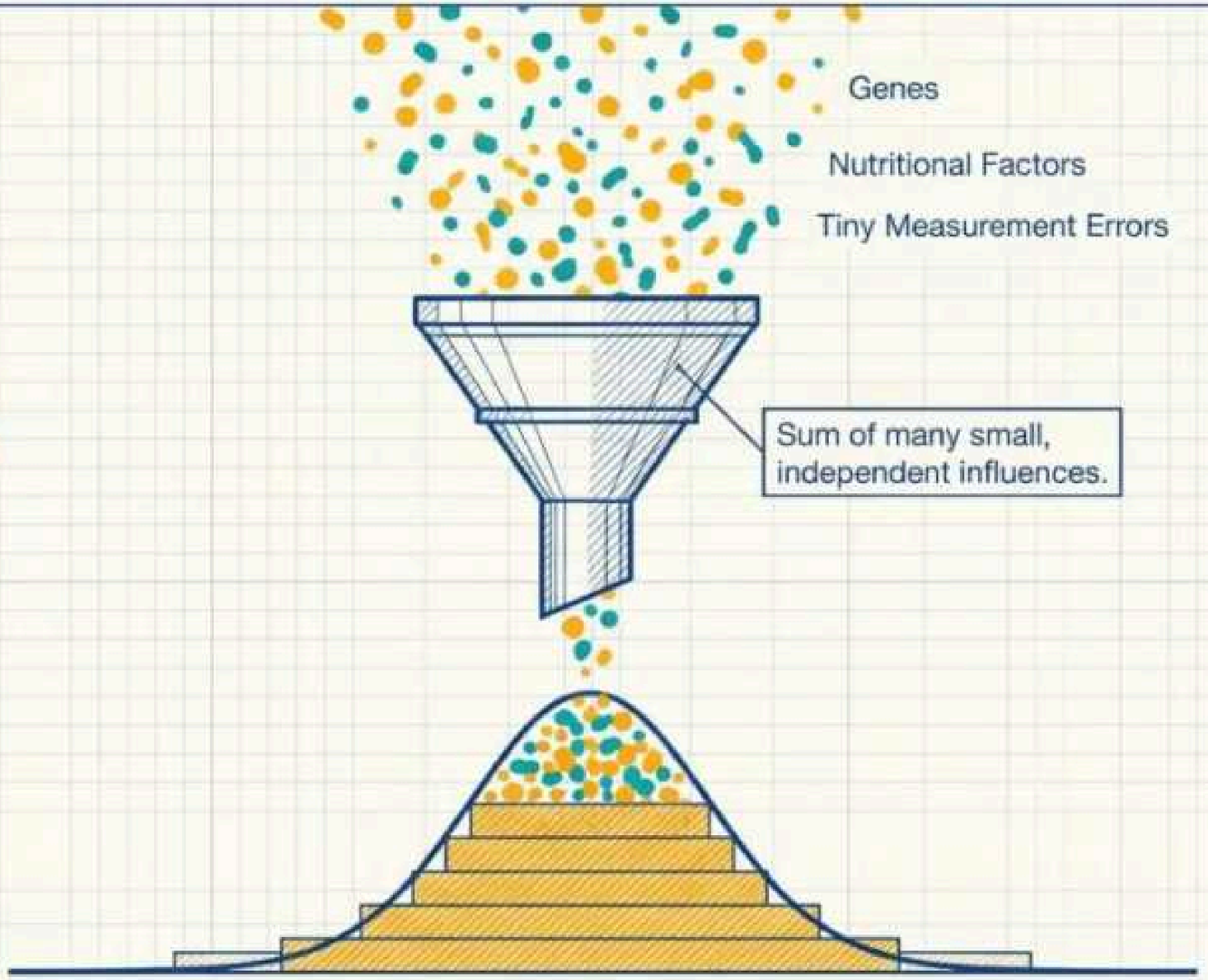
3.3



A **family of normal curves** exists for every pair of means and standard deviations—but once units are rescaled, they all share this exact shape.

3.4

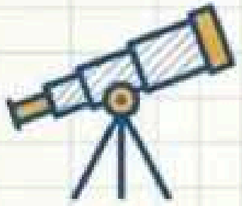
The Architecture of the Bell Shape



Sums of small, independent influences tend **naturally toward the bell shape.**
This tendency is made precise by the **Central Limit Theorem** (explored in Part 2).

3.5

Carl Friedrich Gauss



Used the curve to describe errors in astronomical measurements (The 'Gaussian' Curve).

Adolphe Quetelet



Belgian scientist who applied the curve to human traits.

Francis Galton

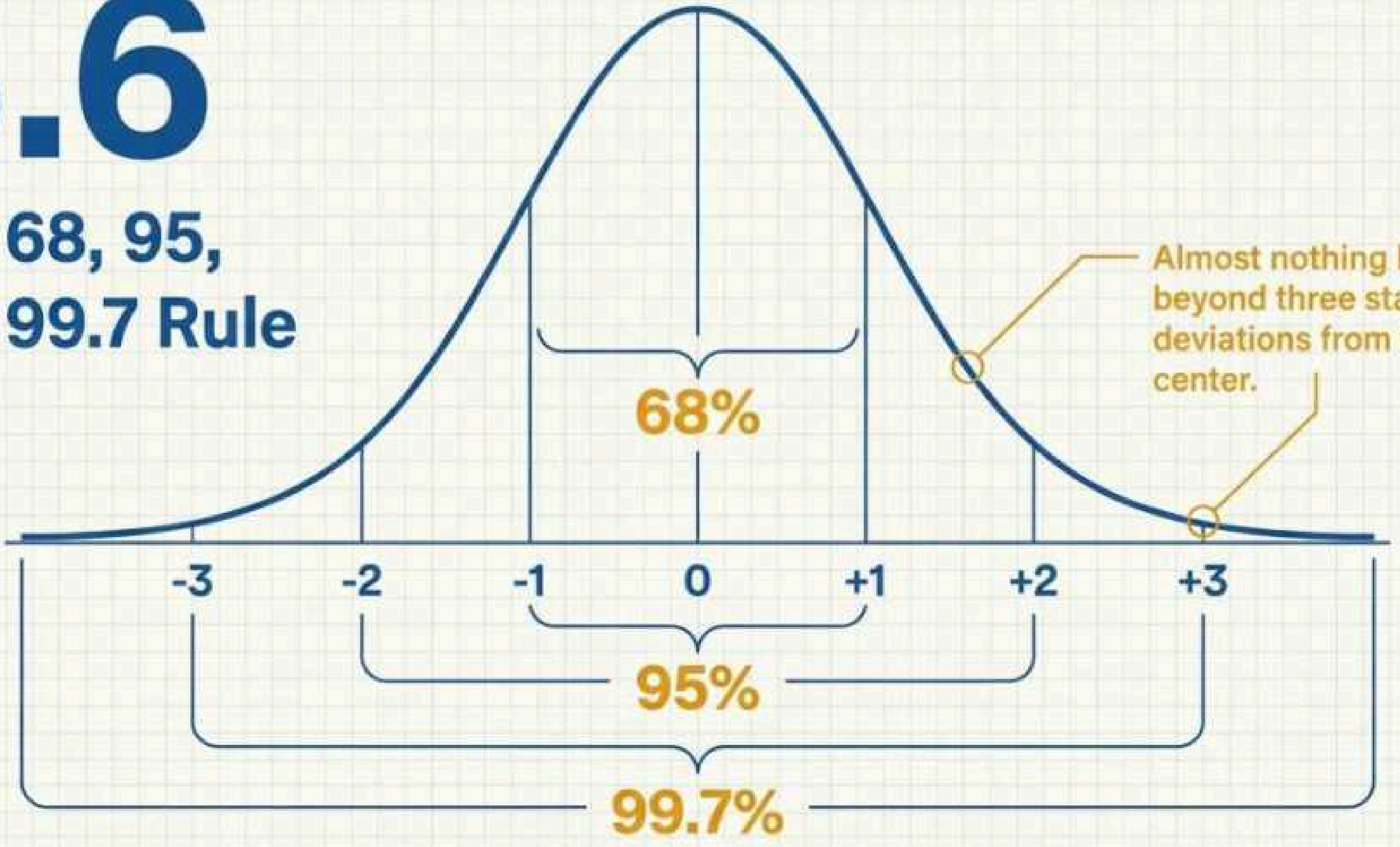


Applied the curve to the study of heredity.

Today, it is the workhorse of statistics.

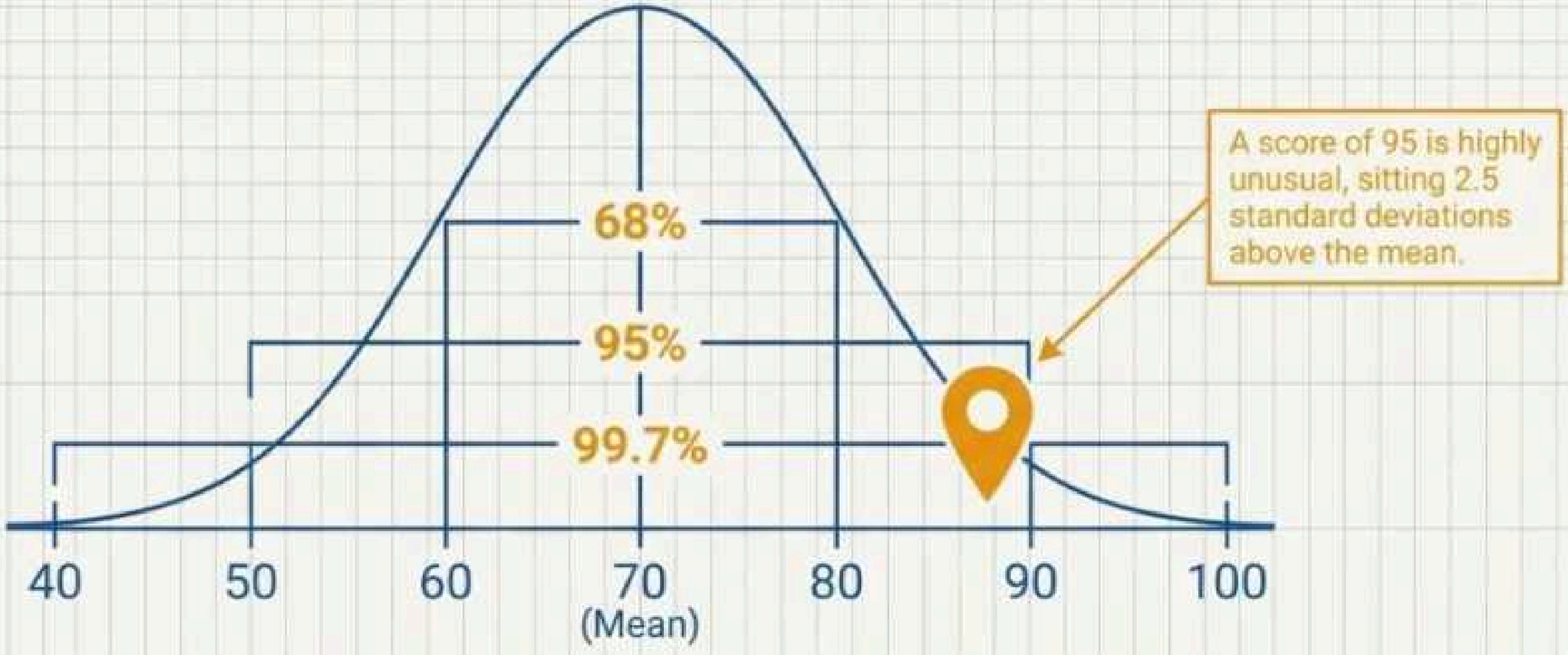
3.6

The 68, 95, and 99.7 Rule



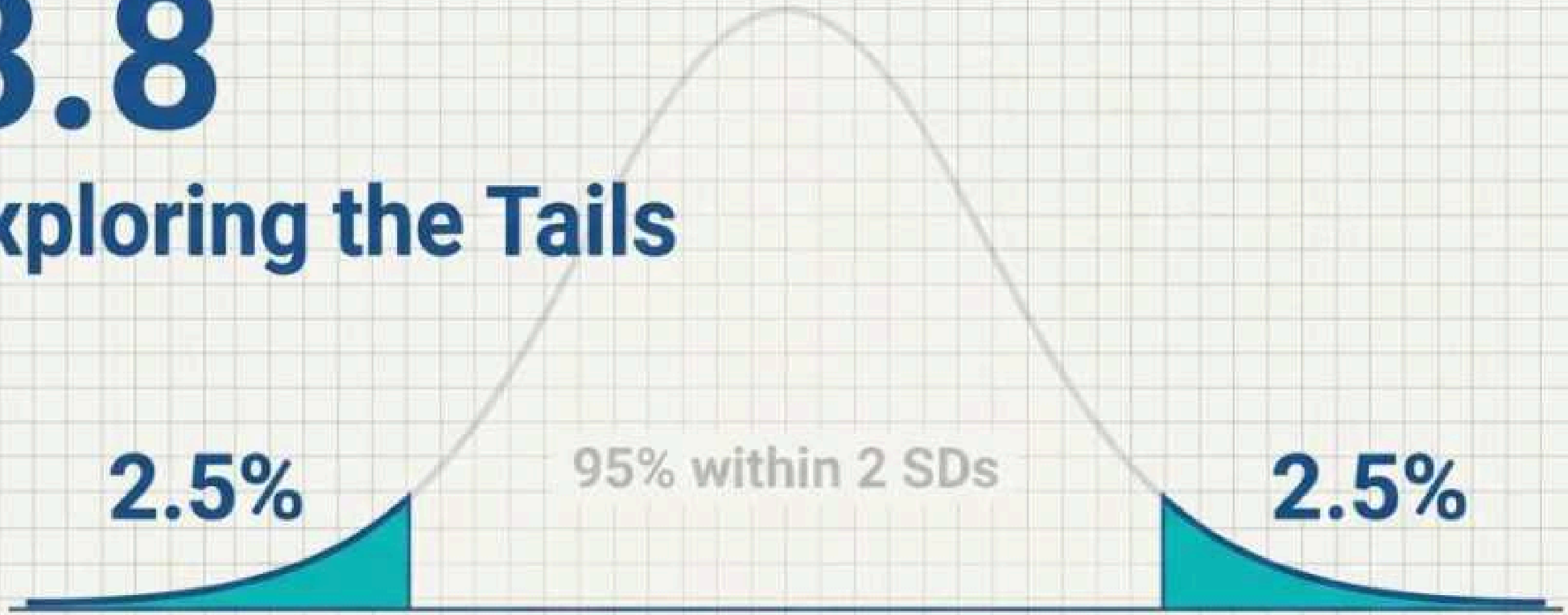
3.7

Applied Example: Exam Scores



3.8

Exploring the Tails



If 95% lie within, 5% lie outside.
By symmetry, the 5% splits evenly into **2.5% above** and **2.5% below**.

About **1 student in 40** scores **above 90**.

Figures come from the idealized model and are approximate for real data.

3.9

Standardizing with Z-Scores

$$Z = \frac{\text{Observation - Mean}}{\text{Standard Deviation}}$$

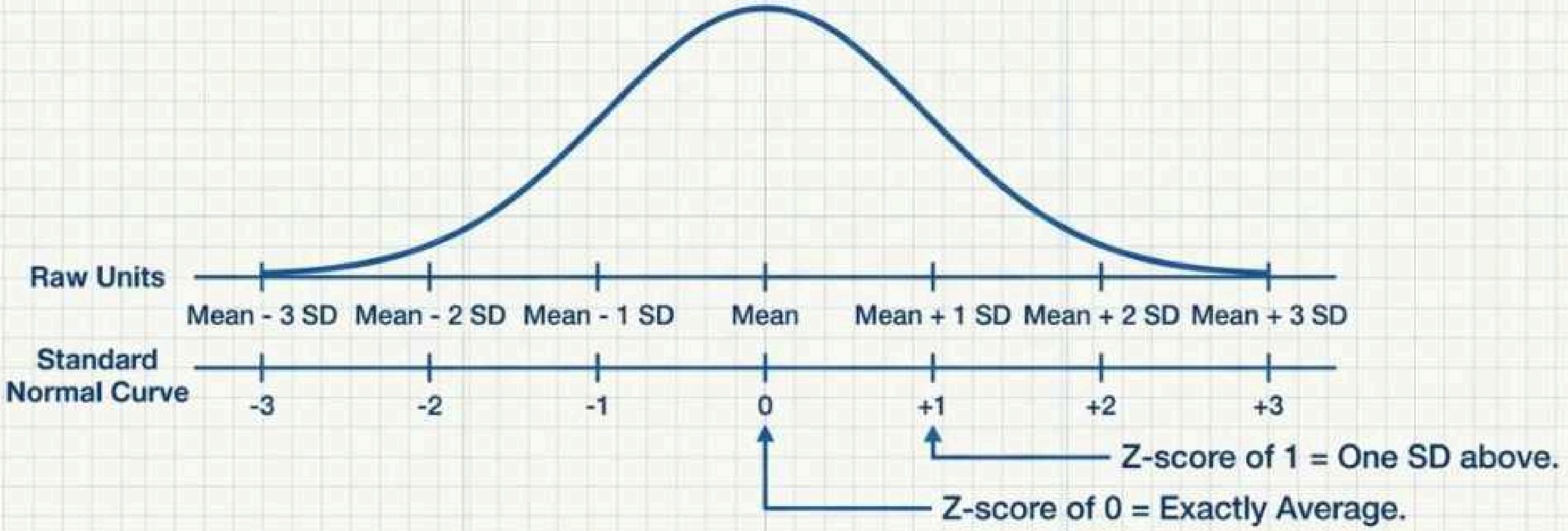
Tells us how many standard deviations a value lies above or below the mean.

Allows us to compare values from completely different scales.

$$Z = \frac{85 - 70}{10} = 1.5$$

3.10

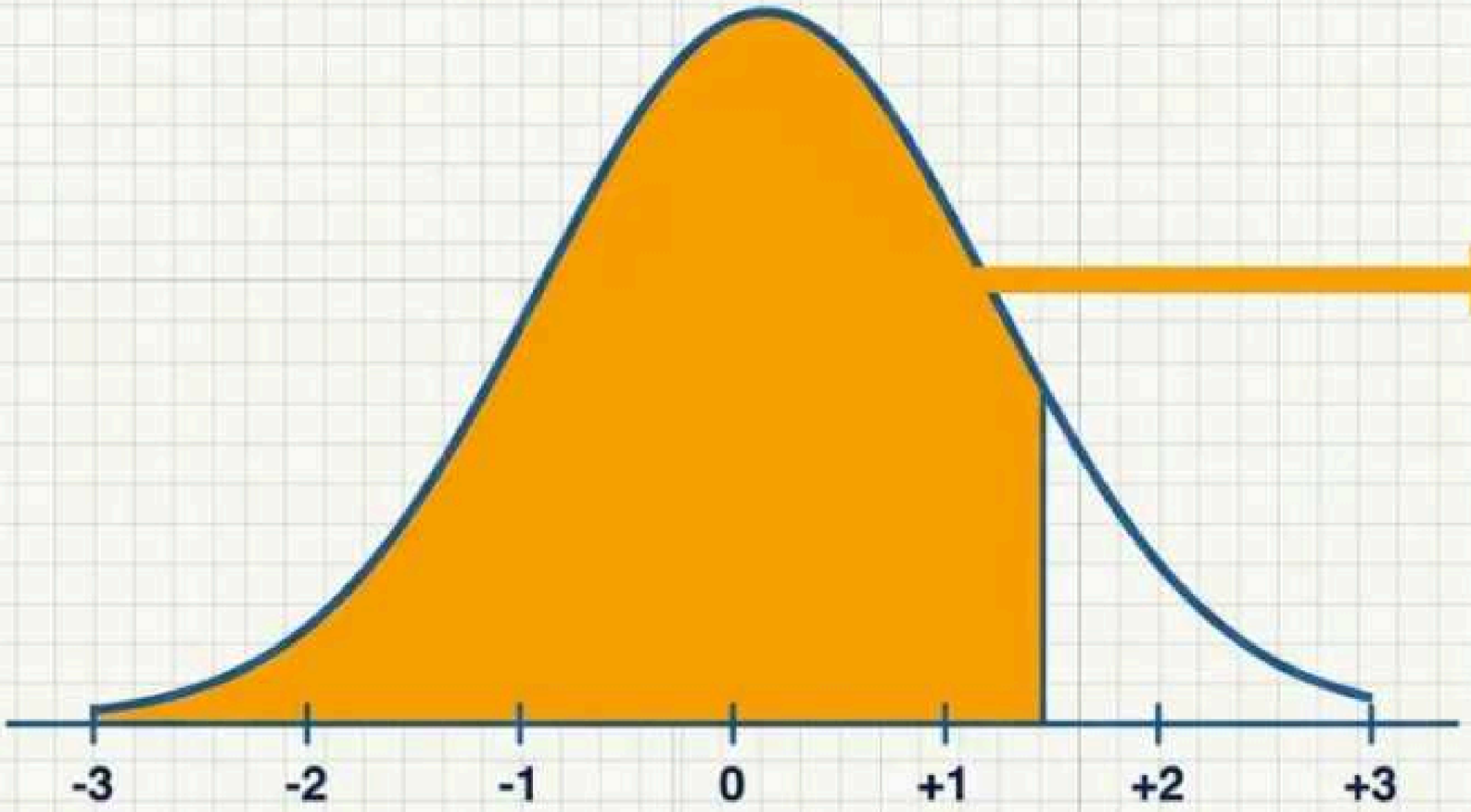
The Standard Normal Curve



Standardizing changes the units but NOT the shape. If original data are normal, z-scores follow the Standard Normal Curve (Mean = 0, SD = 1).

3.11

Calculating Exact Percentages



Using a table or calculator for the standard normal curve gives the exact proportion of observations below any z-score.

Area = 0.9332

Translating to Percentile:
A z-score of 1.5 leaves 93.32% of the area below it. The student with an 85 did better than ~93.3% of the class. They are in the **93rd percentile**.

3.12

Percentile vs. Percent Correct



Percentile

Rank among the group.



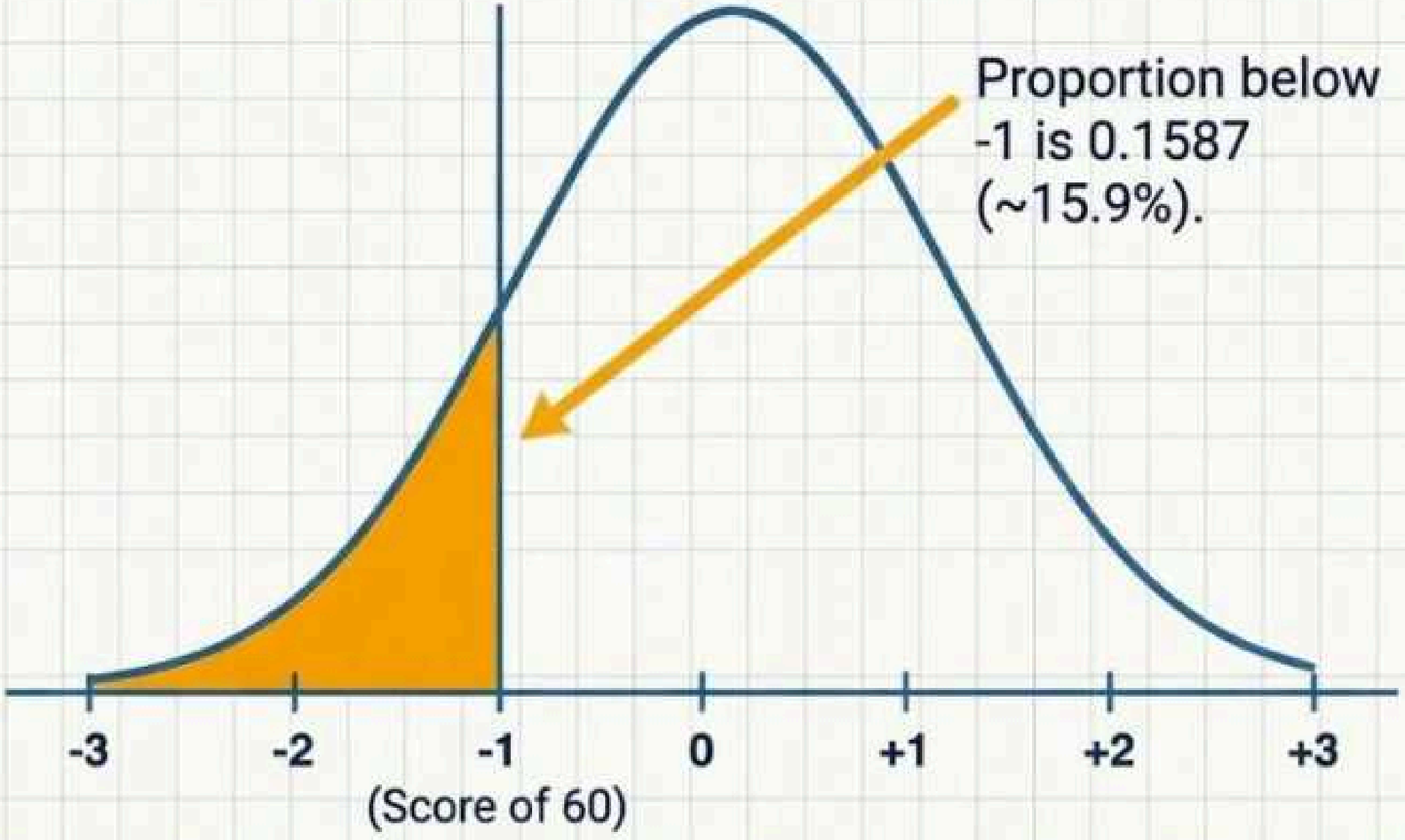
Percent Correct

Performance against the test.

Do not confuse the two. A student can score only 60% correct on a very hard test, but still rank at the 90th percentile compared to their peers.

3.13

The Negative Z-Score



Checking the math:

1. The 68% rule leaves 32% outside.

2. 32% split in half = 16% in the bottom tail.

Conclusion: The math agrees perfectly.

3.14
Finding Cutoffs (Working Backwards)

Top 10% (Needs 90% below it) → Find Z-score for 90th Percentile = 1.28 → Multiply by SD (1.28 × 10 = 12.8) → Add to Mean (70 + 12.8 = 82.8)

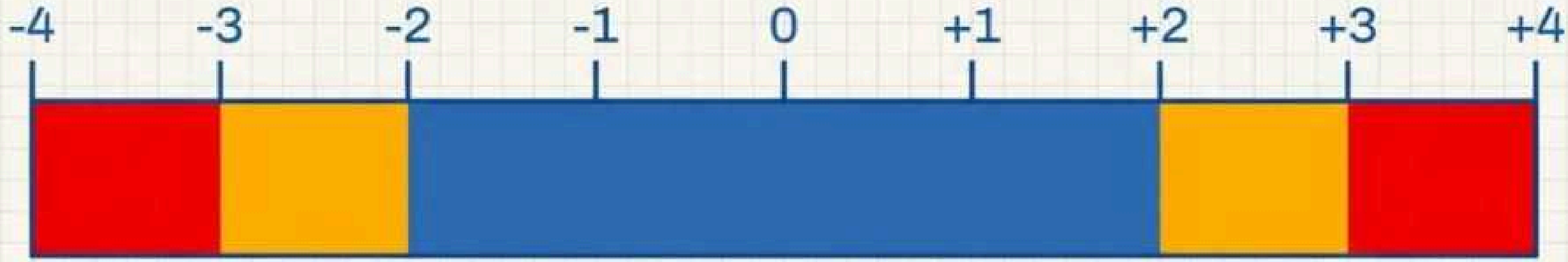
To find the cutoff for an "A" grade given to the top 10%, we reverse-engineer the z-score formula to find a raw score of 82.8.

3.15 Comparing Apples and Oranges

Student	Raw Score	Test Profile (Mean / SD)	Calculated Z-Score
Maria	650	Mean: 500, SD: 100	1.5
Ben	28	Mean: 21, SD: 5	1.4

Even though the scales are completely different, z-scores prove that **Maria did slightly better** relative to her respective group.

3.16 Spotting Unusual Values



Very Surprising (happens ~0.3% of the time)

Noteworthy (happens ~5% of the time)

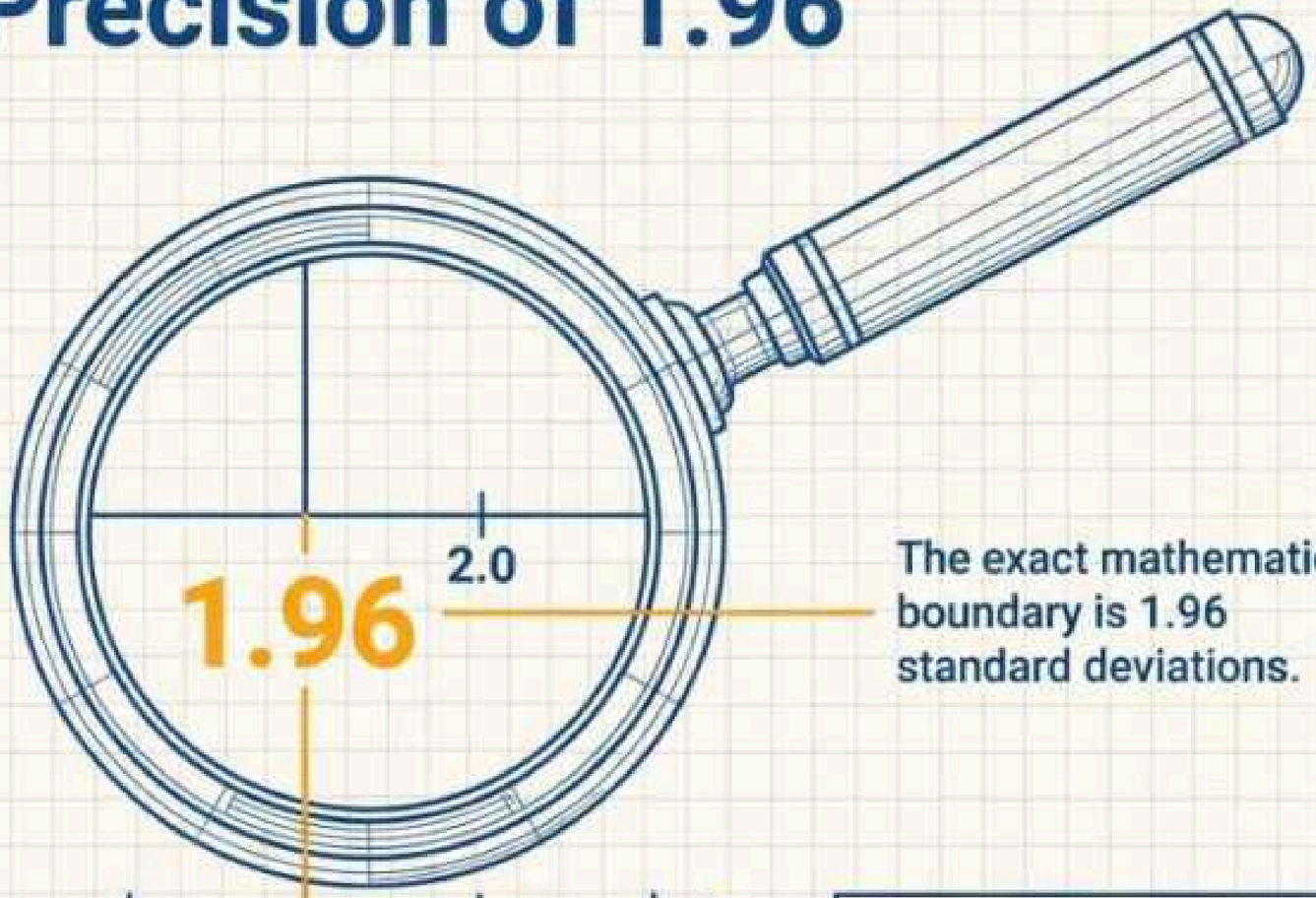
Expected (95% of the time)

Noteworthy (happens ~5% of the time)

Very Surprising (happens ~0.3% of the time)

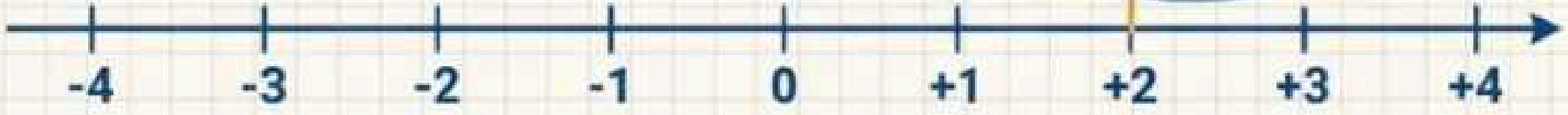
Scientists use this framework to identify outliers.
Part 2 will turn this into formal hypothesis tests.

3.17 The Precision of 1.96



Most people round the middle 95% boundary to 2 standard deviations.

The exact mathematical boundary is 1.96 standard deviations.



This is the exact number behind the margins of error you see in political polls.

3.18 When Data Are Not Normal

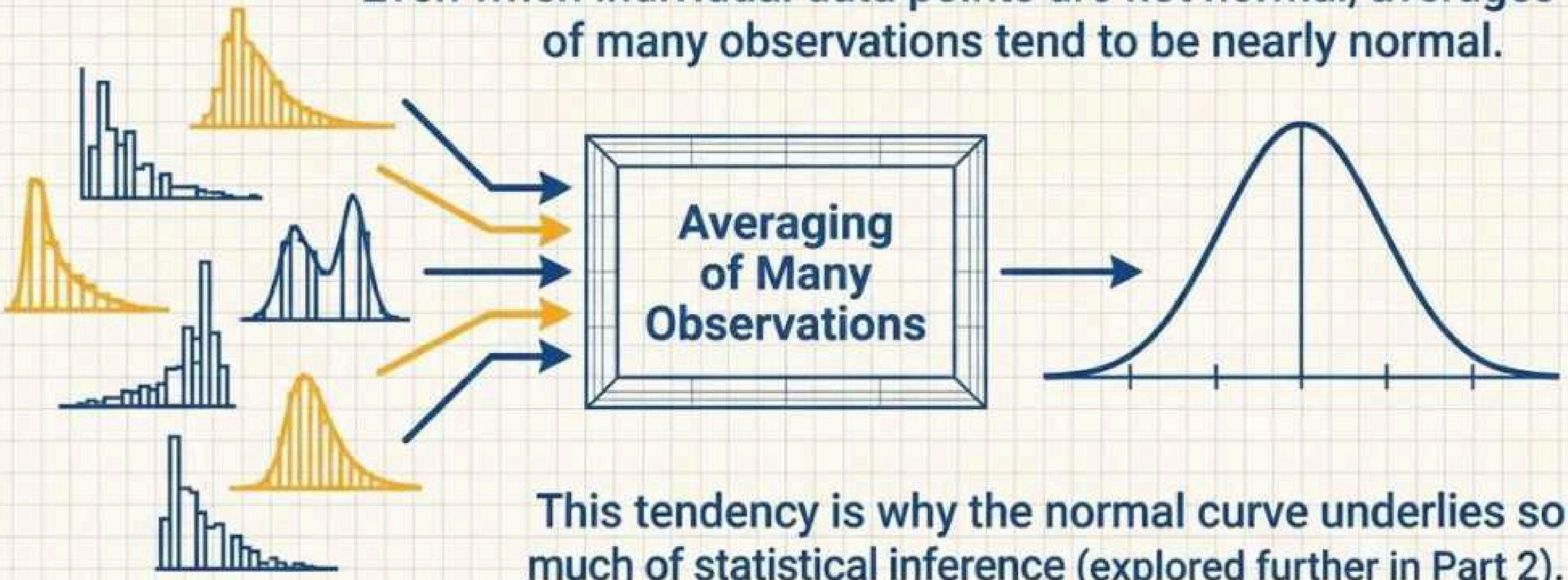


(Also counts of rare events or zero-bounded measurements)

Using normal calculations on skewed data gives wrong answers.
ALWAYS look at a histogram first to see if a bell shape is plausible.

3.19 The Normal Curve Rescue

Even when individual data points are not normal, averages of many observations tend to be nearly normal.



This tendency is why the normal curve underlies so much of statistical inference (explored further in Part 2).

3.20 Lecture Summary: The Normal Distribution

Describing the Curve

- Defined entirely by its Mean (center) and Standard Deviation (width).

The Empirical Rule

- 68% (1 SD), 95% (2 SD), 99.7% (3 SD).

Z-Scores

- Standardized measure of distance from the mean. Used to find exact percentiles and compare completely different scales.

Checking Assumptions

- Data isn't always normal (e.g., income is skewed). Check histograms before calculating.

Next: Lecture 4 turns to two variables at once. →