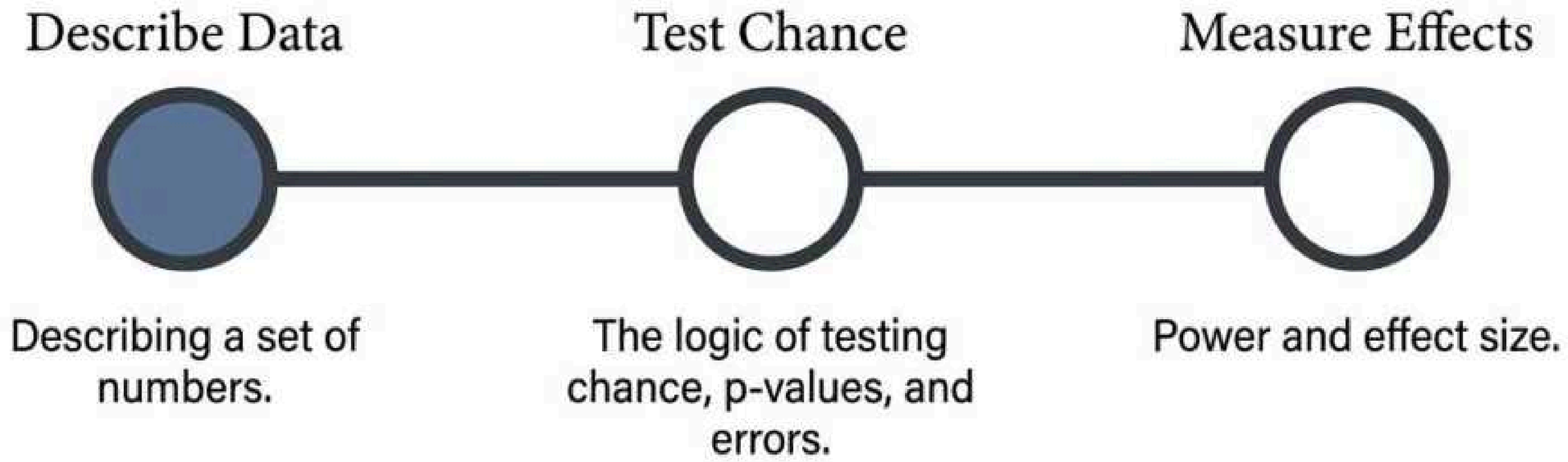
5.1

Statistics in plain language. Understanding the logic of testing results without advanced mathematics—only the core ideas.



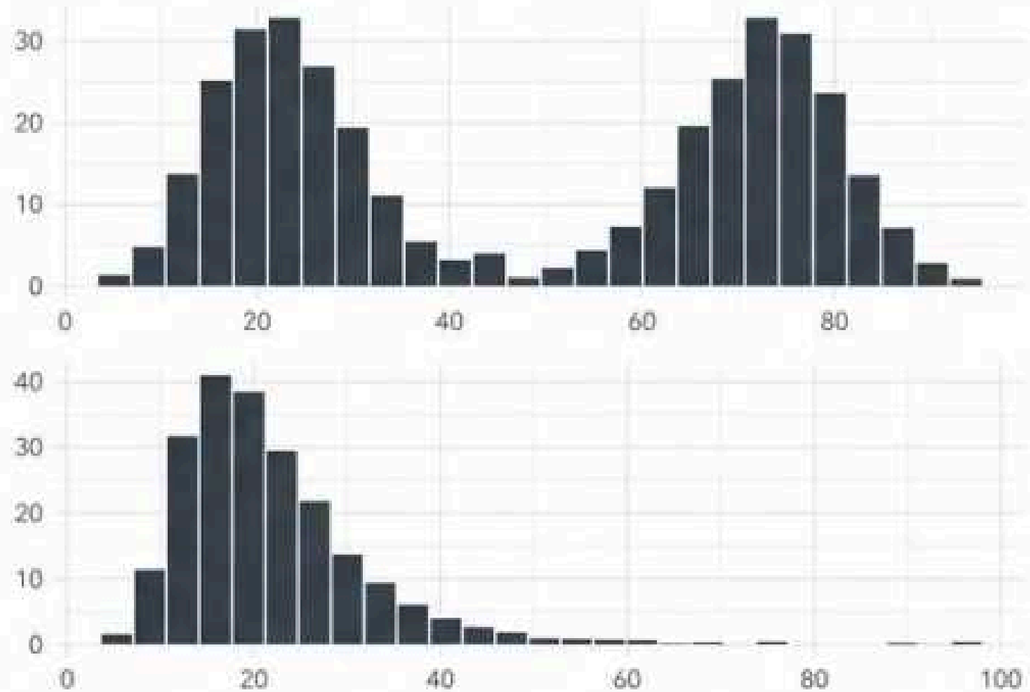
Gemini Notebook

5.2

Descriptive statistics summarize data, but shape must come first. Before calculating anything, always plot the data.

Mean: 50

A single average can hide important features.

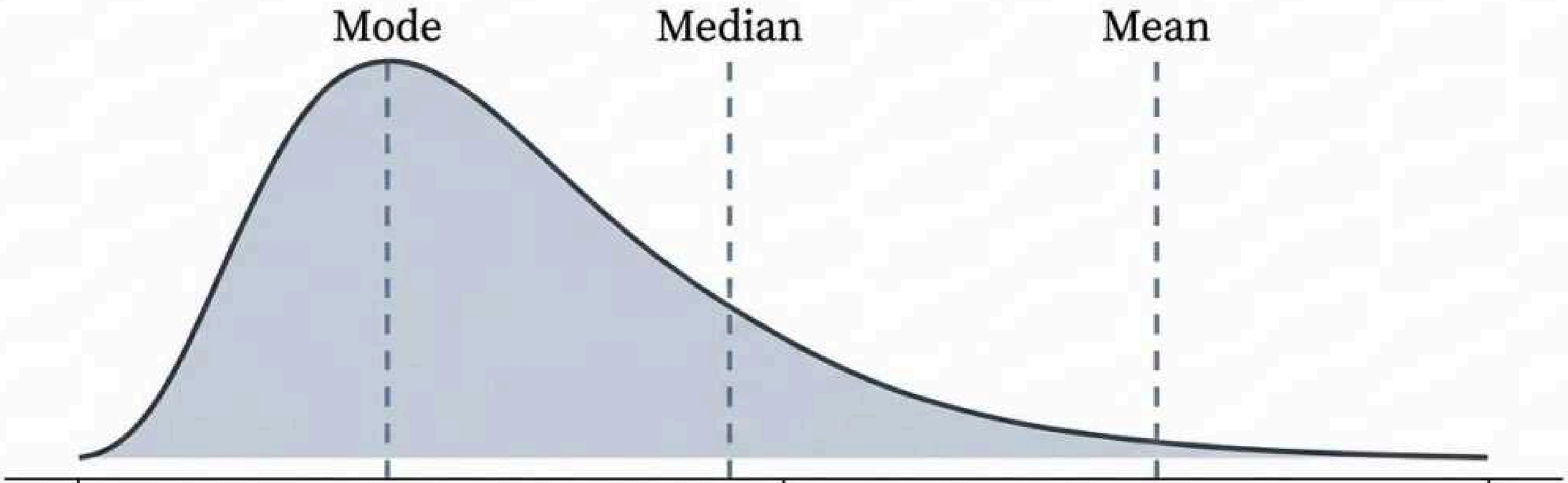


A lopsided or two-peaked distribution tells a story no single number can.

Gemini Notebook

5.3

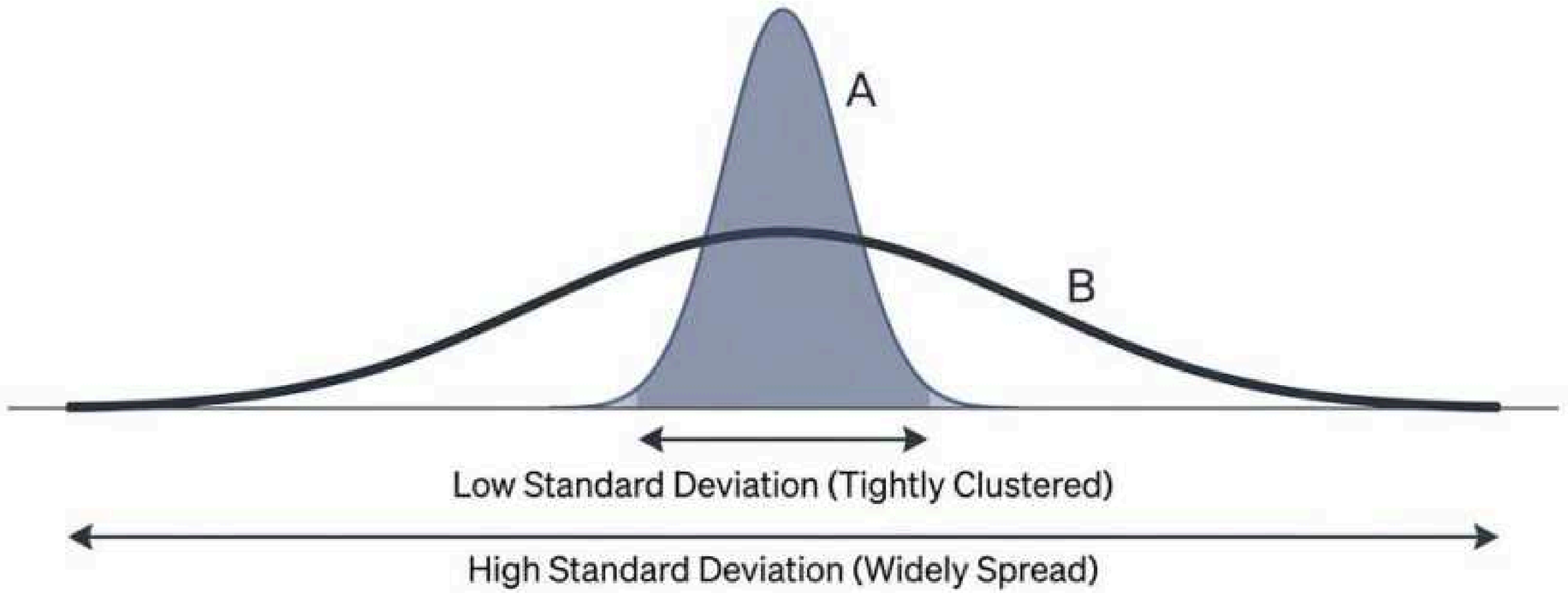
Measures of central tendency give a typical value, but behave differently when data are skewed.



Gemini Notebook

5.4

Averages alone are insufficient without understanding variability.



Two classes can have the identical mean score, yet one ranges from very low to very high, while the other clusters tightly.

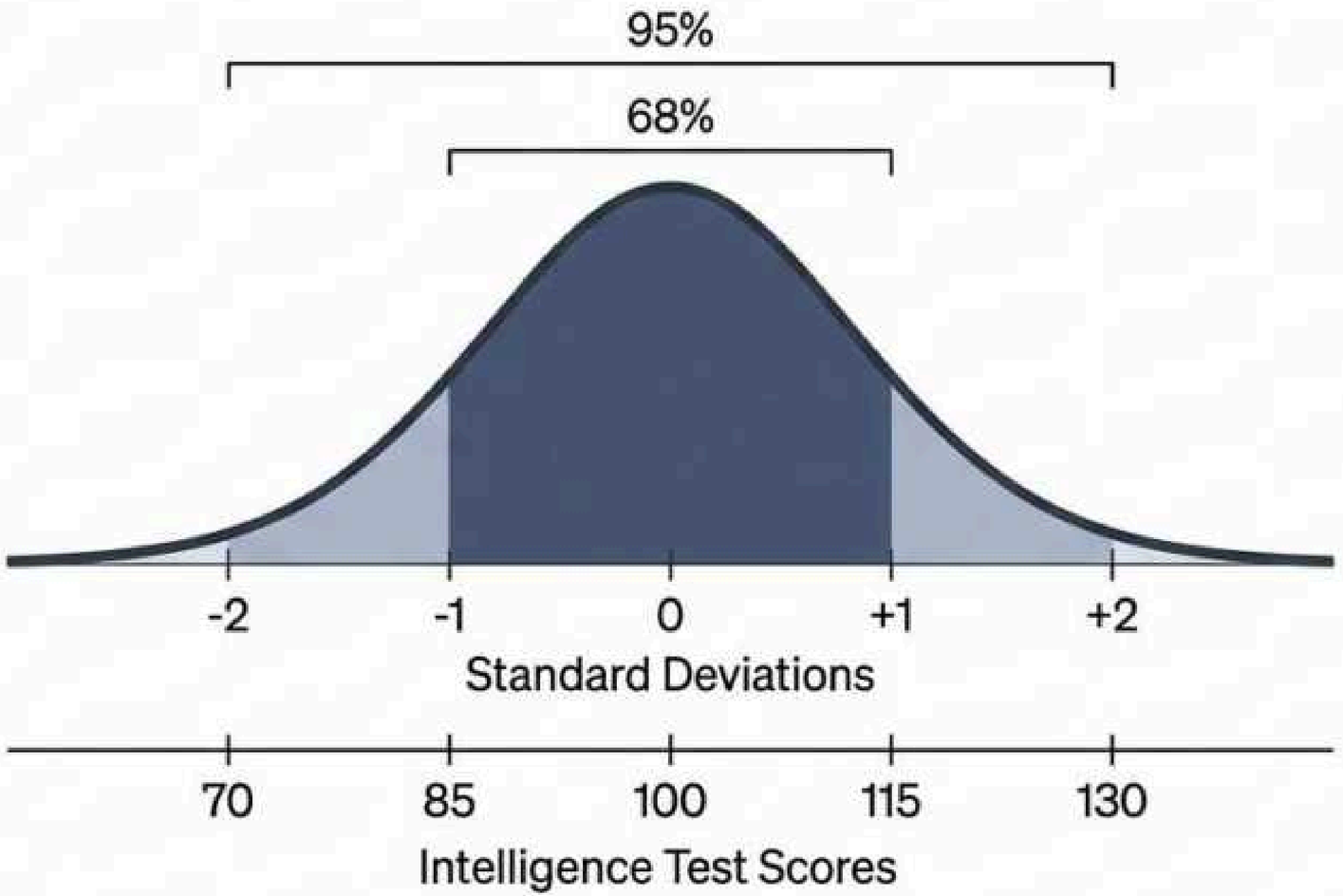
The standard deviation measures how far scores typically fall from the mean. A smaller deviation means tighter clustering.

Gemini Notebook

5.5

Many human traits follow the normal distribution (the bell curve).

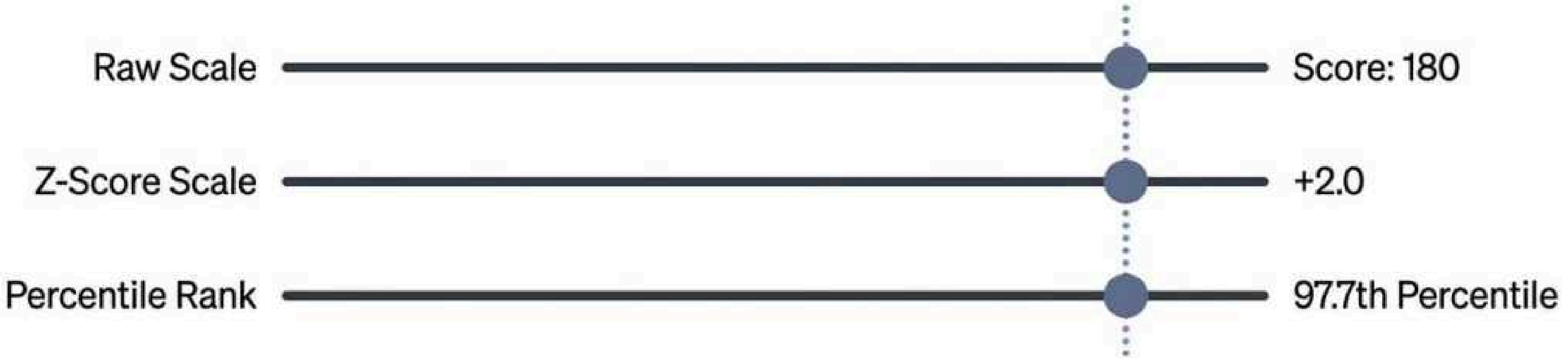
- ~68% of scores fall within 1 standard deviation; ~95% within 2 standard deviations.
- Example: Intelligence tests are scaled to a mean of 100 and an SD of 15. Thus, roughly two-thirds of people score between 85 and 115.



Gemini Notebook

5.6

Standard scores place any result into universal context, allowing cross-scale comparison.



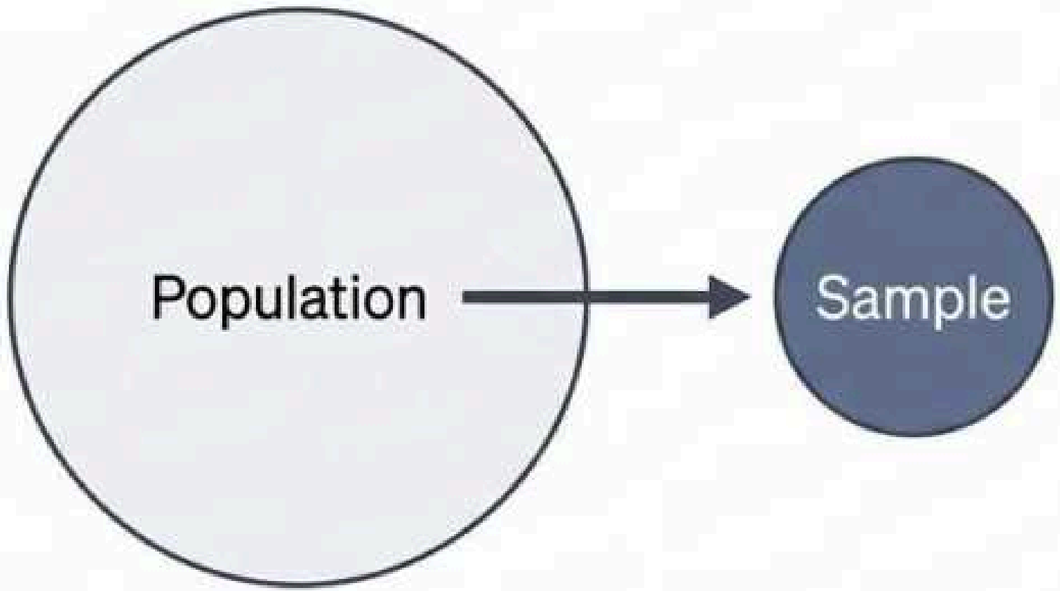
A z-score measures how many standard deviations a score lies from the mean. A z-score of 2 is high on any scale.

A percentile rank identifies the share of the population who scored lower than that point.

Gemini Notebook

5.7

Inferential statistics draw conclusions about a population from a smaller sample.



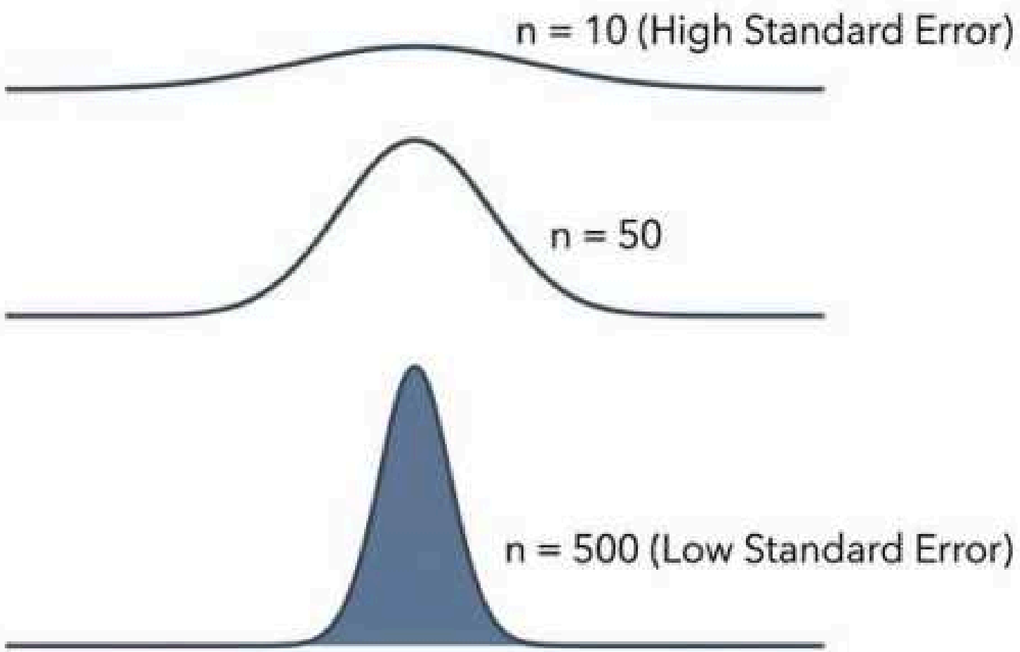


- Core Problem: Sampling error. A sample never matches the population exactly.
- Takeaway: A sample mean will be near the population mean, but rarely equal to it.

Gemini Notebook

5.8

If we drew many samples, their means would form a sampling distribution. Its spread is the standard error.



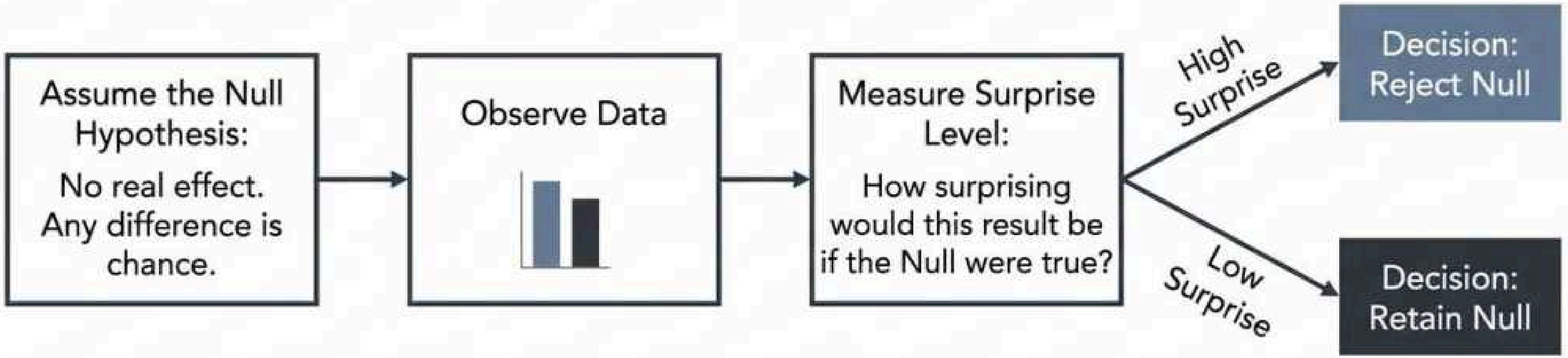
Quadruple the Sample (4x)
→
Halve the Error (0.5x)

Key Fact: Standard error shrinks as samples get bigger. Big samples bounce around less. A study of 500 people is vastly more precise than a study of 10. Precision is expensive.

Gemini Notebook

5.9

The core logic of Null Hypothesis Significance Testing (NHST).



We do not measure the probability of the hypothesis; we measure the surprise level of our data under the assumption that the null hypothesis is true. Conclude an effect is likely only if the data would be highly surprising otherwise.

Gemini Notebook

5.10

The measure of surprise is the p-value. It is the most commonly misunderstood idea in statistics.

p-value

The probability of getting results at least as extreme as ours IF the null hypothesis were true.

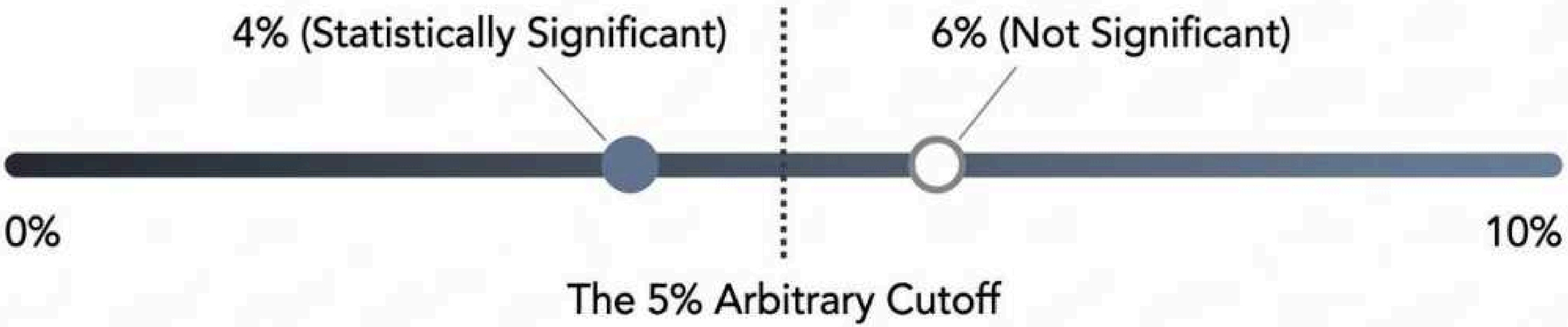
~~NOT the probability that the null hypothesis is true.~~

~~NOT the probability that the result is a fluke.~~

Gemini Notebook

5.11

By tradition, a result is "statistically significant" when the p-value falls below a 5% cutoff (a convention from Ronald Fisher).



The Problem: The cutoff is arbitrary. A 4% and a 6% p-value offer practically the same evidence but fall on opposite sides of the line. The number must be a guide, not an absolute verdict.

Gemini Notebook

5.12

Testing can go wrong in two distinct ways, and reducing one tends to increase the other.

	Reality: Effect Exists	Reality: No Effect Exists
Conclusion: Effect Found	Correct Hit	Type I Error (False Alarm)
Conclusion: No Effect Found	Type II Error (Miss)	Correct Rejection

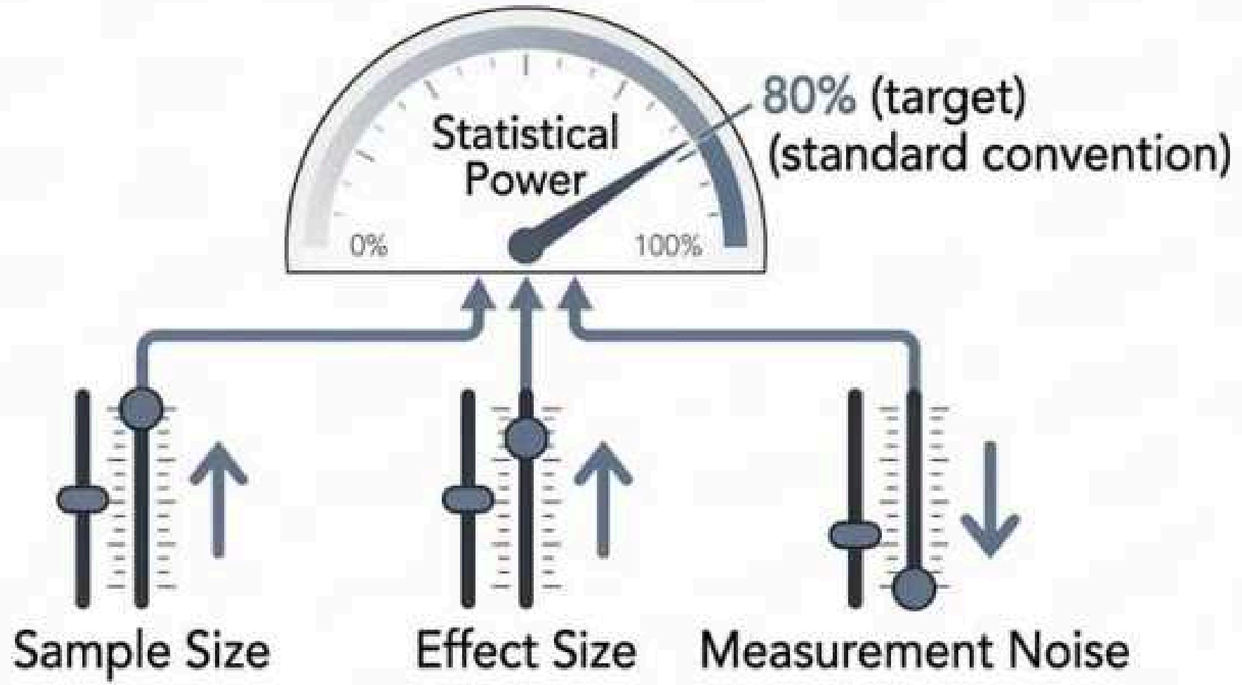
Type I Error: A false alarm. Concluding there is an effect when there is none (the 5% cutoff sets this long-run rate).

Type II Error: A miss. Failing to detect an effect that is real.

Gemini Notebook

5.13

Power is the chance of detecting a real effect (1 minus the chance of a Type II error).

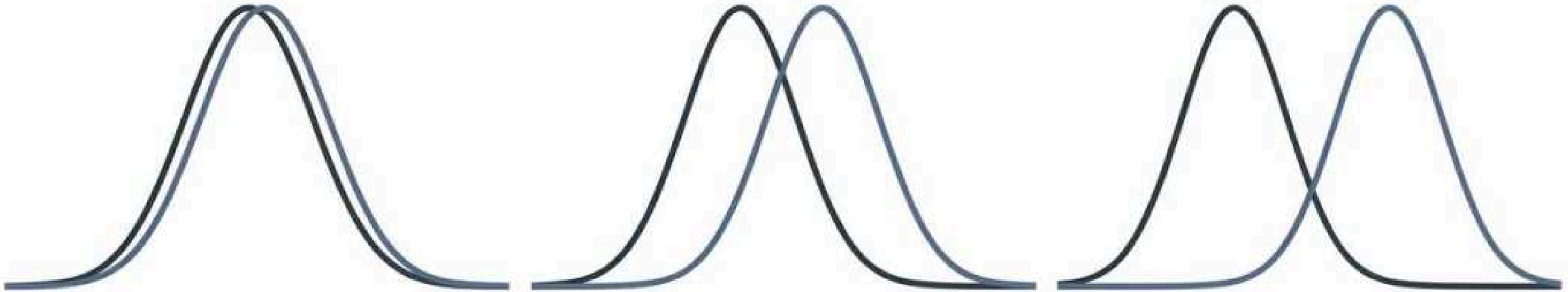


Power rises when you increase sample size, study larger effects, and reduce measurement noise. Classic studies often had low power, meaning real effects were missed and found effects were exaggerated. Modern studies use power calculations to decide participant numbers in advance, rather than guessing.

Gemini Notebook

5.14

A statistically significant result does not necessarily mean an important one. With a huge sample, even a trivial difference becomes significant.



d = 0.2 (Small Effect)

d = 0.5 (Medium Effect)

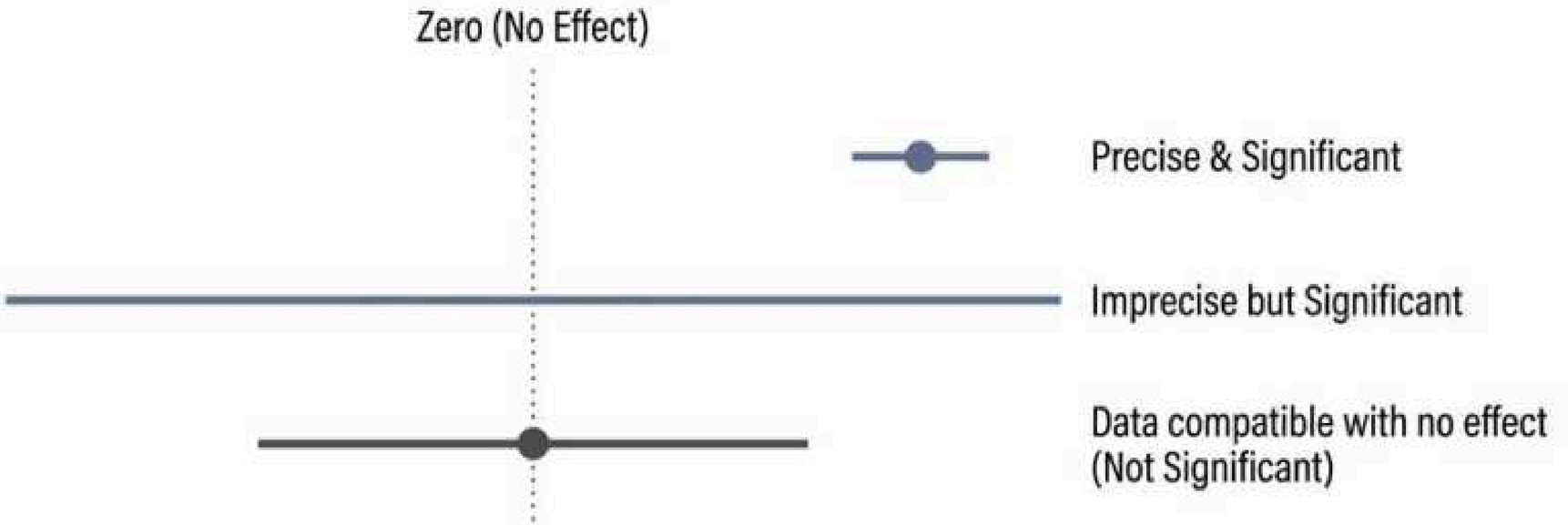
d = 0.8 (Large Effect)

Effect size tells us how big the difference actually is. Cohen's d expresses this difference in standard deviation units.

Gemini Notebook

5.15

A confidence interval gives a range of plausible values for the true effect.



- The 95% Rule: If we repeated the study many times, about 95% of such intervals would contain the true value.
- Interpretation: A wide interval signals imprecision. An interval that includes zero means the data are compatible with no effect.

Gemini Notebook

5.16

Different research questions call for different tests.

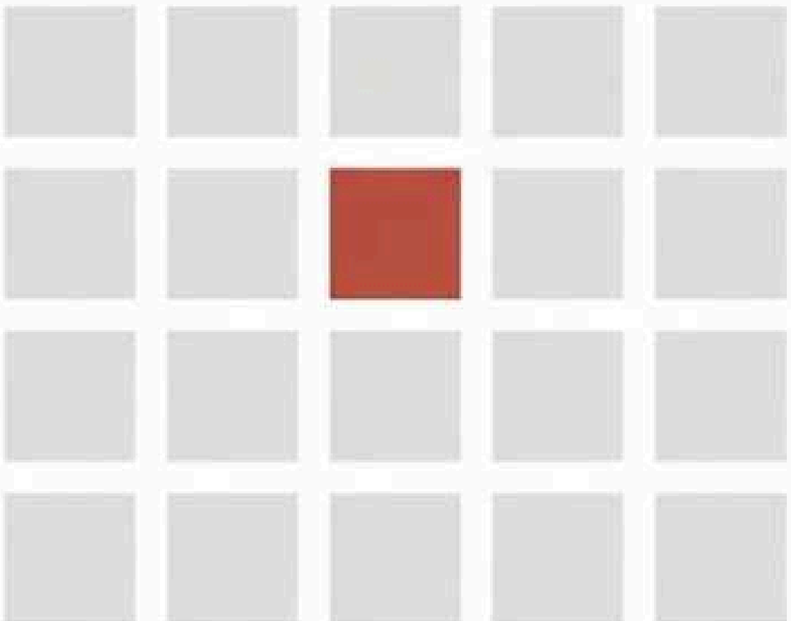
Statistical Test	What It Compares
t-test	Compares the means of two groups.
ANOVA	Compares three or more groups, and handles factorial designs.
Correlation	Asks whether a relationship is larger than chance would produce.
Chi-square	Examines counts in categorical data.

All rely on the same logic of the null hypothesis.

Gemini Notebook

5.17

Beware of common analytical errors and the trap of multiple comparisons.



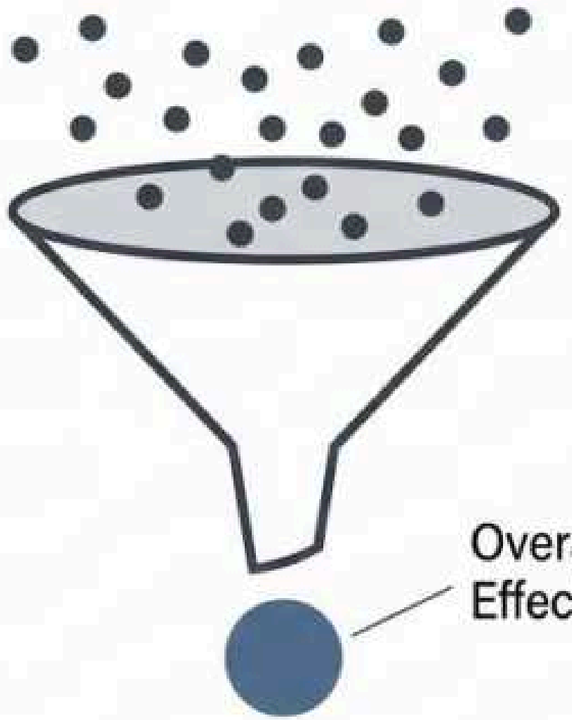
Error: False Alarms. If you test 20 unrelated comparisons at the 5% level, you expect about 1 to appear significant by sheer luck.

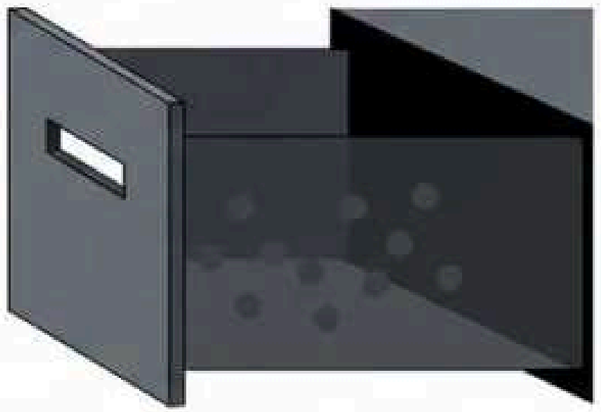
- Error: Confusing non-significance with no effect. A study may simply lack power.
- Takeaway: Honest reports state how many tests were run. Always ask how many chances a result had to appear.

Gemini Notebook

5.18

One study is rarely enough. Researchers combine many in a meta-analysis to pool results and estimate an overall effect size across labs.





The File Drawer
(Unpublished Null Results)

The Threat: A meta-analysis can mislead if the combined studies are biased. If unpublished studies with null results are missing (the file drawer problem), the overall effect will appear falsely inflated.

Gemini Notebook

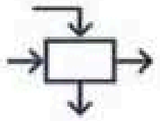
5.19

Lecture Recap: The three pillars of statistical inference.



1. Descriptive

Summarizes data with a center (mean, median) and a spread (standard deviation).



2. Significance Testing

Asks if a result could be chance. The p-value is the probability of such data if there were no actual effect.



3. Evaluation

Effect size (how large), Power (how reliable), and Confidence Intervals (how precise).

Gemini Notebook

From mathematical mechanics to the human side of research.

In the final lecture, we turn to ethics, the replication crisis that shook psychology, and how to read studies and news reports with a critical and fair eye.

Gemini Notebook