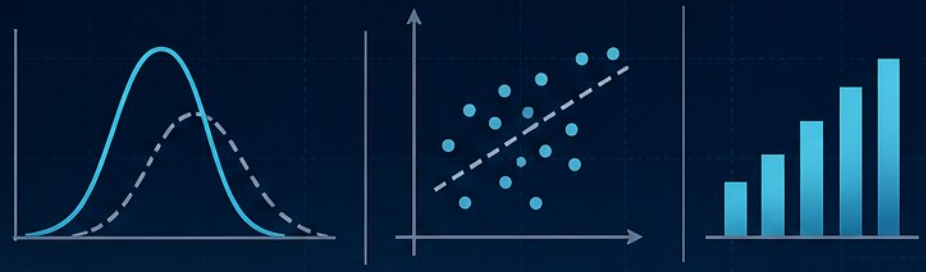
Power in Statistics:
Concepts, Purpose, and Examples



- Statistical power: the probability of detecting a true effect
- A pre-study design tool for experiments and analyses
- Connects significance testing, sample size, and decision quality
- Essential for analysts, researchers, and experiment designers

What Is Statistical Power?

- Probability of correctly rejecting a false null hypothesis
- Power = 1 – β, where β is Type II error probability
- A pre-study design property, not a post-result justification

Hypothesis Testing Foundation

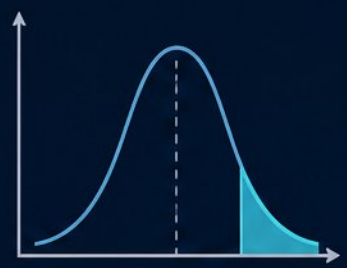
- Null and alternative hypotheses frame all possible outcomes
- Type I error: false positive; Type II error: missed true effect
- Lower alpha reduces power but cuts false positives
- Significance tests frame decisions via p-values and rejection regions

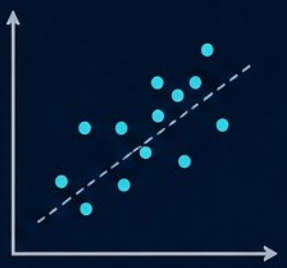
DECISION MATRIX

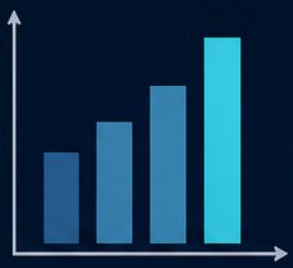
	RETAIN NULL HYPOTHESIS (Do Not Reject H_0)	REJECT NULL HYPOTHESIS (Reject H_0)
NULL HYPOTHESIS IS TRUE (H_0 is true)	Correct Decision	Type I Error: False Positive
ALTERNATIVE HYPOTHESIS IS TRUE (H_a is true)	Type II Error: Missed True Effect	Correct Decision (Discover True Effect)

REJECT NULL HYPOTHESIS
Seek evidence for an effect

RETAIN NULL HYPOTHESIS
Insufficient evidence for an effect







The Four Interlocking Quantities

- Power, effect size, alpha, and sample size are mathematically linked
- Fixing any three quantities determines the fourth
- Power analysis and sample size calculation are the same procedure in reverse
- Treating sample size as an isolated choice leads to poor study design

Effect Size: The Magnitude Worth Detecting

- Quantifies the practical difference or relationship of interest
- Common metrics: Cohen's d, r, correlation coefficients
- Sources: meta-analyses, prior studies, pilot data, literature
- Alternatives: smallest effect size of interest (SESOI)
- Realistic effect size assumptions often matter more than sample size

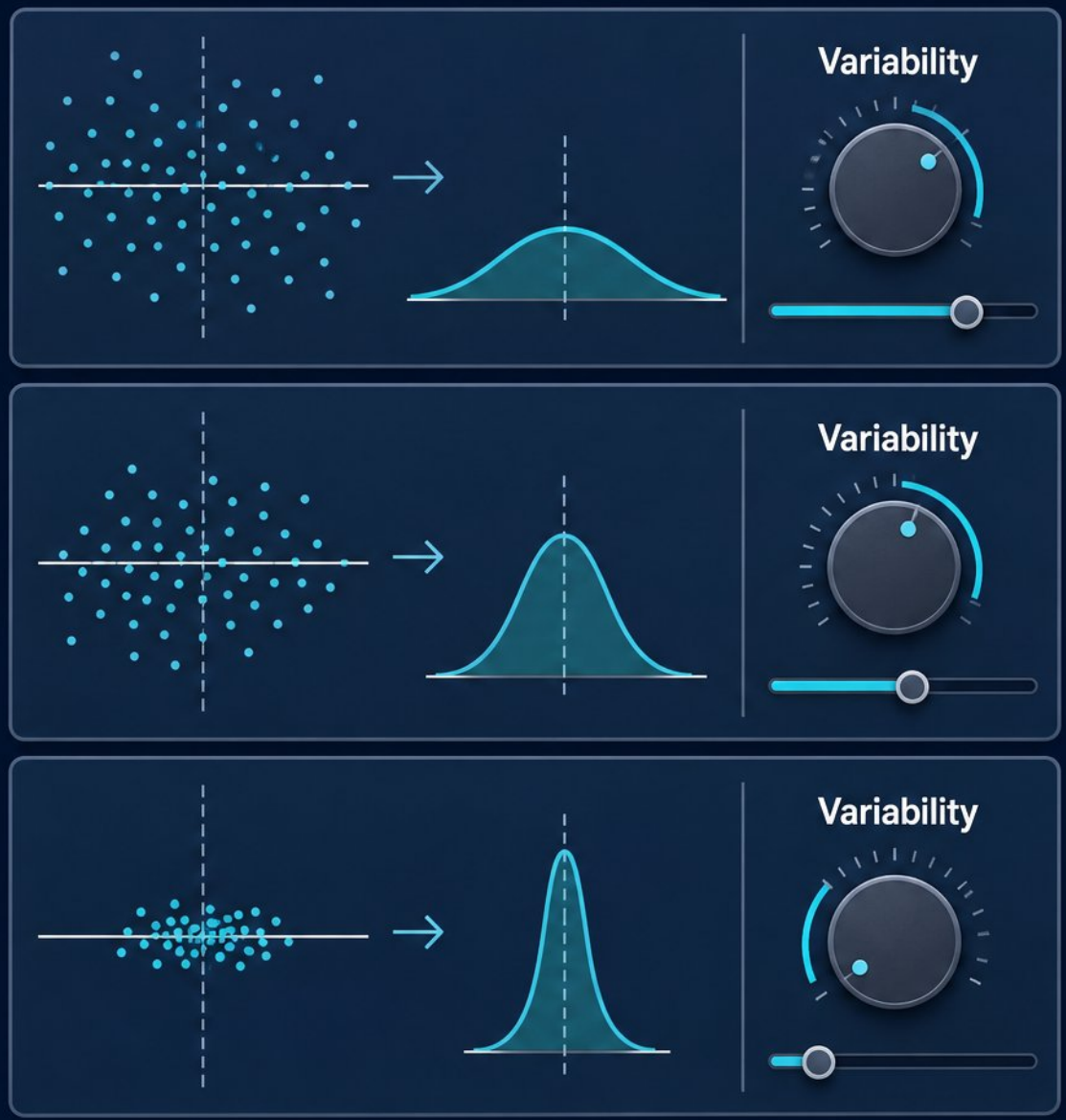
SPECIMEN CARD

Detecting and Measuring a Meaningful Difference

Effect size is the **magnitude** of the difference or relationship we aim to detect—not just whether it exists.

Alpha, Variability, and Sample Size

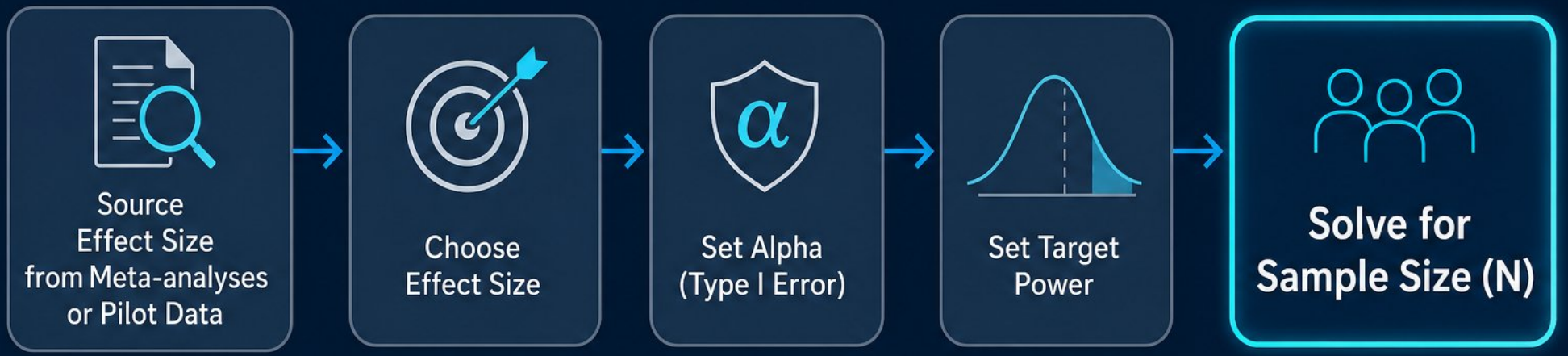
- Lower alpha (e.g., 0.01) means stricter evidence standards
- Stricter alpha reduces power or demands larger samples
- Higher variability hides true effects behind noise
- Use precise measurement or designs to reduce variance
- Larger samples boost power but with diminishing returns



Why Power Matters in Practice

- Low power misses true effects, wasting time and resources
- Significant results from underpowered studies are often inflated
- Inflated estimates contribute to poor replication and reproducibility
- Adequate power planning is a key credibility and ethics decision

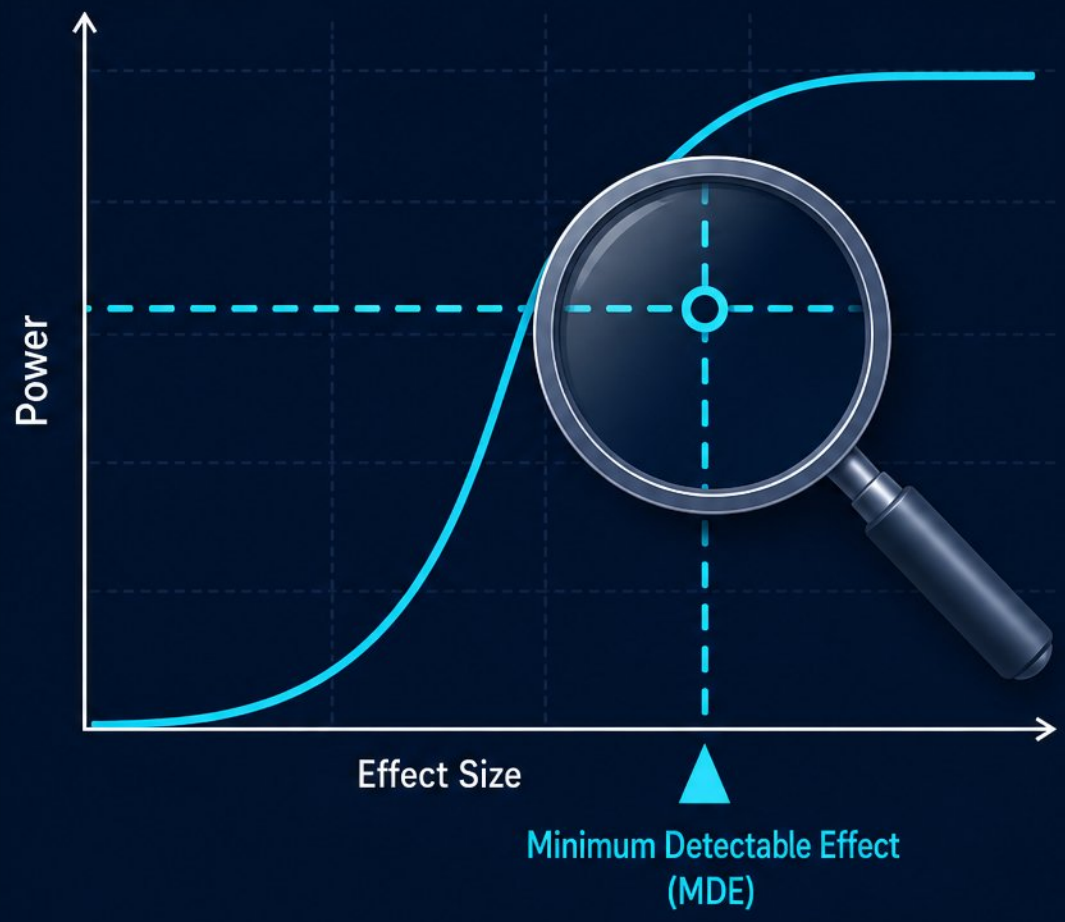
A Priori Power Analysis Workflow



- - Plan sample size before collecting data
- - Fix effect size, alpha, and target power; solve for N
- - Source effect size from meta-analyses or pilot data
- - Document assumptions and inflate for expected attrition

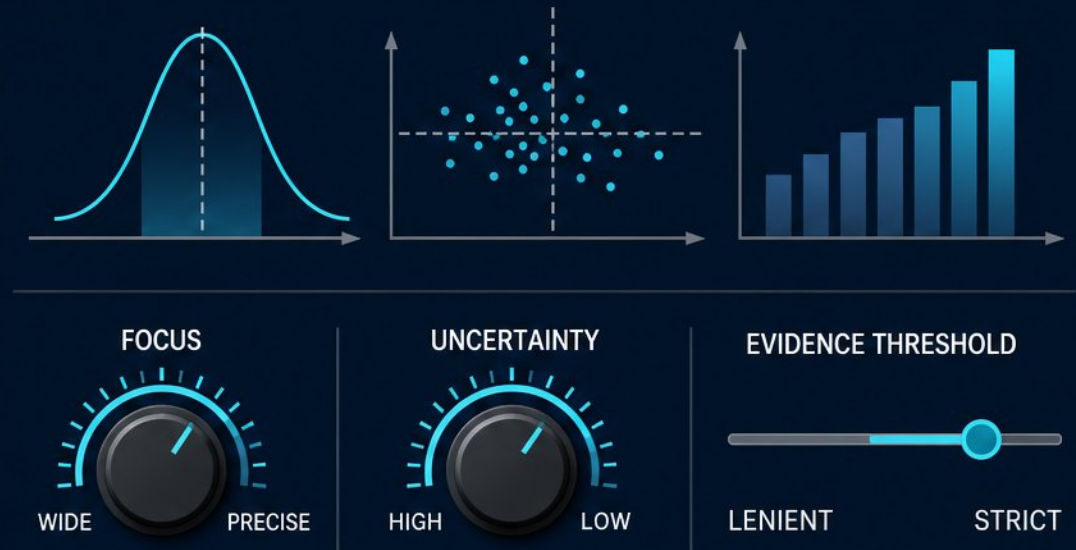
Sensitivity Analysis and Minimum Detectable Effect

- Sensitivity analysis reveals the minimum effect a fixed sample can detect at a given power
- This smallest detectable effect is the Minimum Detectable Effect (MDE)
- MDE assesses whether the design can answer the research question at hand
- Power curves map how power changes with effect size and sample size



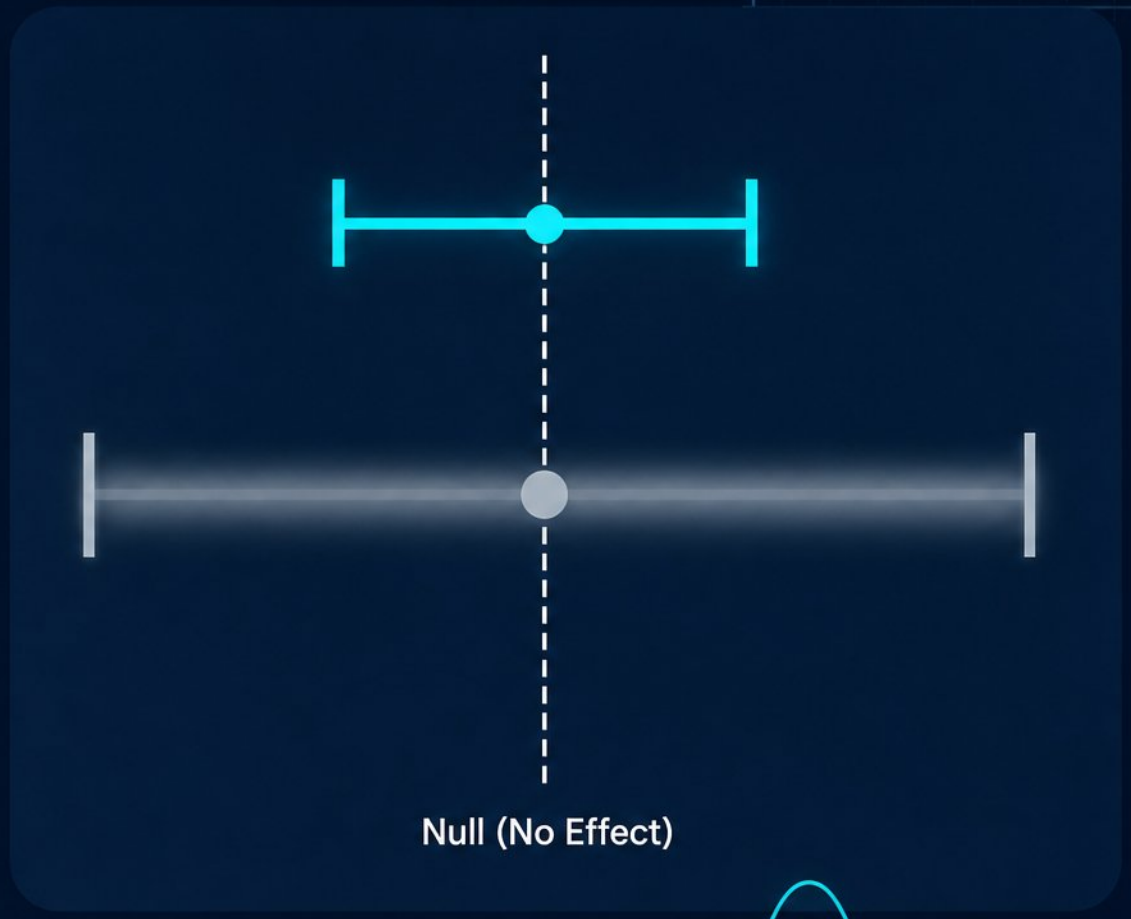
Post Hoc Power and Better Alternatives

- Post hoc power restates the p-value — no extra information
- Observed power is always low for nonsignificant results
- Use 95% confidence intervals to interpret findings instead
- Report the range of effect sizes the study could detect

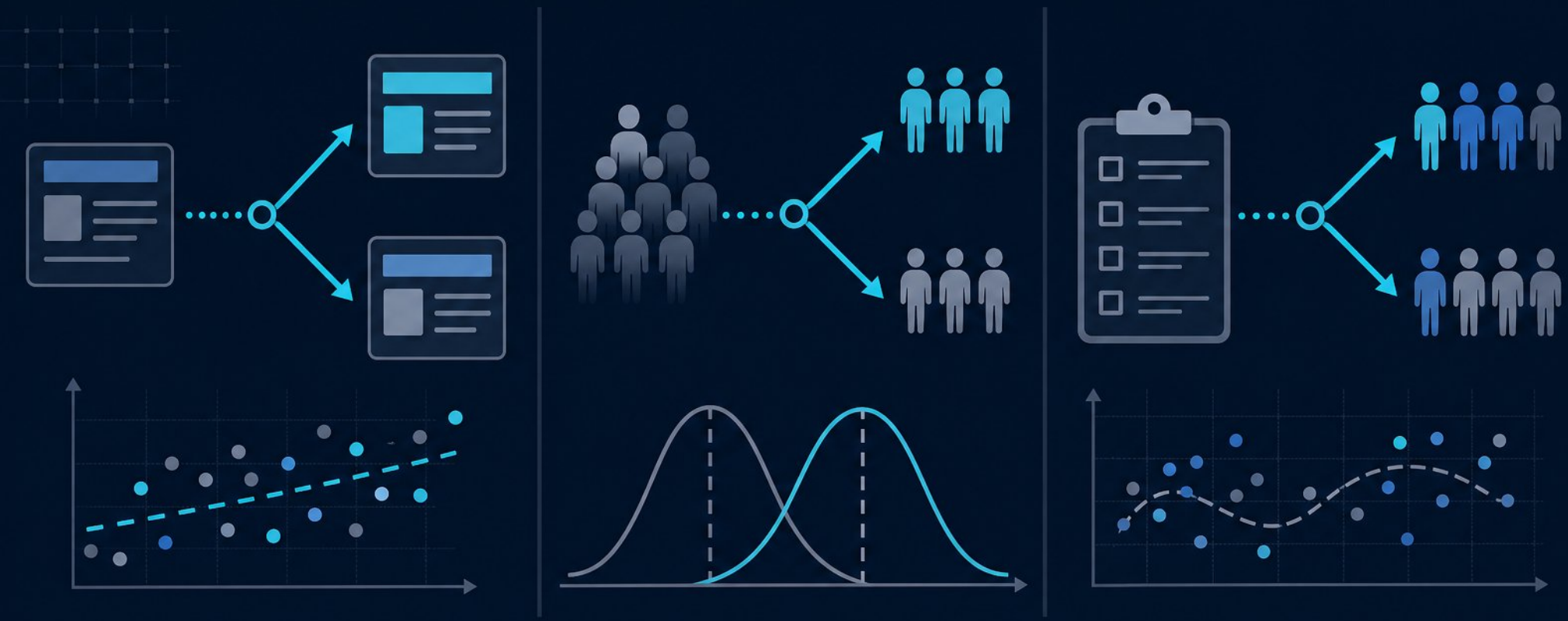


Interpreting Power and Confidence Intervals

- Narrow confidence intervals support more decisive conclusions
- Nonsignificant results with wide intervals are inconclusive
- Check intervals against meaningful effect sizes when judging results



Practical Examples Across Fields



- A/B tests: detect meaningful conversion lifts with adequate samples
- Clinical trials: plan sample size for target effect, alpha, and power
- Surveys and behavioral research: account for design effects
- Simulations: model power for complex real-world designs

Common Misconceptions and Pitfalls

- Confusing statistical significance with practical importance
- Post hoc power restates p-value; is misleading
- Nonsignificant results are not proof of no effect
- Larger samples do not fix poor design or unrealistic effect sizes

Tools and Resources for Power Analysis

- **G*Power:** free software for t, F, χ^2, z, and exact tests
- R packages: pwr, pwrss, ggpower, and Spower
- **Python:** MCPower and ezpwr for flexible power analysis
- Monte Carlo simulation for complex or nonstandard designs
- Match tool to test family, design, and analysis type

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/2dbe012b/public