Statistical Power Calculations

Urban Economics

Ignacio Sarmiento-Barbieri

Universidad de los Andes

November 12, 2025

Statistical Power and Study Design

- Statistical power tells us how likely we are to detect economically relevant effects before investing in field work.
- Power analysis is something we do **before** we run a study.
  - Helps you figure out the sample you need to detect a given effect size.
  - Or helps you figure out a minimal detectable difference given a set sample size.
  - May help you decide whether to run a study.

Four results from hypothesys testing

What we really want to
know but cannot observe

Reality/underlying truth

What we actually
measure/learn

Evaluation results

	No impact	**Impact**
No impact detected		
Impact detected		

Four results from hypothesys testing

		Reality/underlying truth	
		No impact	Impact
Evaluation results	No impact detected	GREAT!	False negative: you conclude there is NO impact when there is Mismatch!
	Impact detected	False positive: you conclude there is impact when there is not Mismatch!	GREAT!

Type I and Type II Errors

Type I Error (α)

Rejecting H_0 even though it is true. Intuition: we interpret pure noise as a real effect. Common convention: $\alpha = 0.05$.

Type II Error (β)

Failing to reject H_0 when it is false. Intuition: the study is not sensitive enough to capture the real effect.

- **Example**: testing whether a transit subsidy reduces travel time by 5 minutes. If we report "no effect" even though travel time drops, we commit a Type II error.

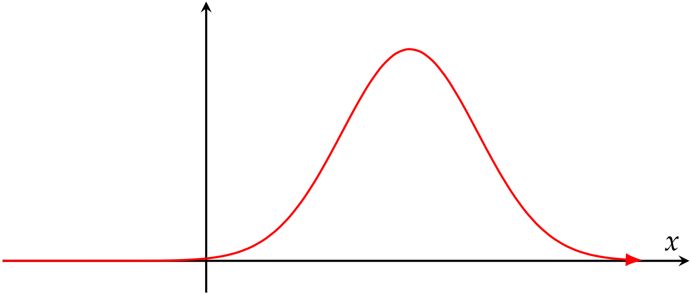
Linking Errors and Power

- Statistical power: $P(\text{reject } H_0 \mid H_1 \text{ true}) = 1 - \beta$.
- Lowering α (being stricter) usually raises β if the design stays the same—there is a trade-off.
- Goal: balance risks while accounting for the cost of wrong policy decisions.

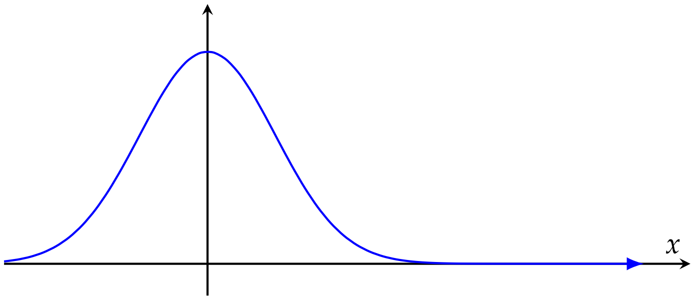
Table 1: Decision Matrix

	Do not reject H_0	**Reject** H_0
H_0 true	Correct decision	Type I error (α)
H_0 false	Type II error (β)	Power ($1 - \beta$)

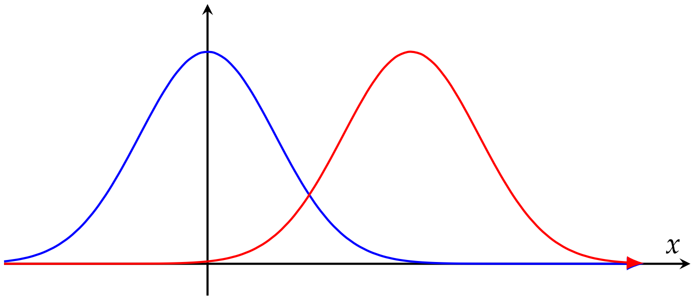
Many Experiments: a distribution of Estimates



Distribution of Estimates if true effect is 0



Two distributions under two hypothesis



What Drives Power?

- **Effect size**: larger effects are easier to detect; define the minimum detectable effect that matters.
- **Sample size**: more observations reduce estimator variance and raise power.
- **Significance level (α)**: stricter thresholds reduce power unless other elements change.
- **Outcome variability**: less dispersion (e.g., stratifying or adding covariates) improves precision and power.

Example: doubling the number of treated neighborhoods in an infrastructure pilot can raise power from 0.55 to 0.80 while keeping $\alpha = 0.05$.

Why Simulate?

- Complex models (interactions, nonlinear effects, spillovers) rarely yield closed-form power formulas.
- Simulations translate realistic urban behavior assumptions into sampling distributions of statistics.
- Advantages: flexibility, ability to incorporate unbalanced panels, heteroskedasticity, or adaptive assignment rules.

Basic Simulation Workflow

1. Specify the data-generating model (sample sizes, expected effect, intra-cluster correlations).
2. Draw thousands of synthetic replications and estimate the statistic of interest each time.
3. Compute the share of replications that reject H_0; that share is the simulated power.
4. Adjust the design (number of units, assignment, instruments) until reaching the target power, e.g., 0.8.

Key Takeaways

- Designing with enough power keeps us from mistaking noise for evidence or missing real effects.
- Understanding α, β, and power helps plan the appropriate study size.
- Simulations expand our toolkit when classical assumptions break down.
- Documenting assumptions eases replication, auditing, and communication of experimental design quality.