

Efficient power analysis methods for complex statistical models

Filippo Gambarota¹ Enrico Toffalini²

¹Department of Developmental and Social Psychology

²Department of General Psychology

University of Padova

@AIP Sperimentale Venezia 2026

Motivation and computational challenge

Power analysis turns assumptions into a design

Power analysis turns assumptions about the effect and the data-generating process into a study design with a desired level of power.

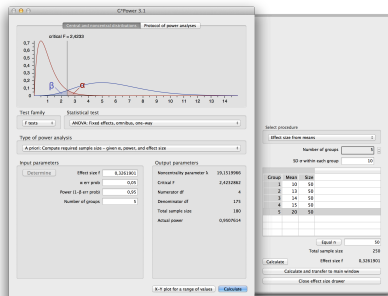
hypothesis $\longrightarrow \theta \longrightarrow$ model assumptions $\longrightarrow N^*$

For example, given an effect θ and a significance level α , what total sample size N yields 80% power?

$$\pi(N^*) = .80$$

Simple models can be solved directly

G*Power



jamovi

The screenshot shows the jamovi software interface. The 'Analyses' tab is selected, and the 'T-Test Power Analysis' window is open. The 'Calculate' dropdown is set to 'N'. The 'T-Test type' is 'Independent samples'. The 'Effect Information' section shows 'Expected Cohen's d (d)' as 0.5, 'N (Group 1 size)' as 20, and 'Relative size of groups' as 1. The 'Parameters' section shows 'Minimum desired power' as 0.9, ' α (type I error rate)' as 0.05, and 'Tails' as 'Two-tailed'. The 'Equivalence Testing' section has 'Perform equivalence testing' and 'Explanatory text' checkboxes, both unchecked. The 'Options' section has 'Sensitivity Analysis' and 'Power Parameters plot' buttons, both highlighted.

Complex models require Monte Carlo simulations

When no closed-form solution is available, power is often estimated by Monte Carlo simulation.

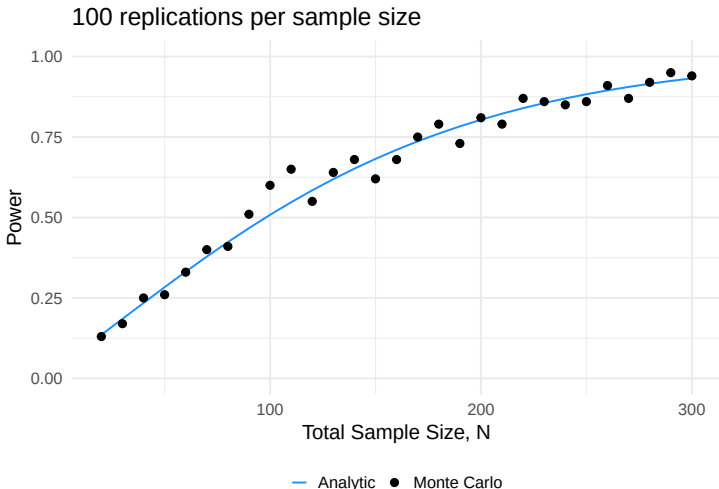
- ▶ Define the data-generating process and statistical model
- ▶ Evaluate multiple candidate values of N
- ▶ At each N , repeatedly simulate, fit, and test

$$\text{cost} = \text{candidate designs} \times \text{replications}$$

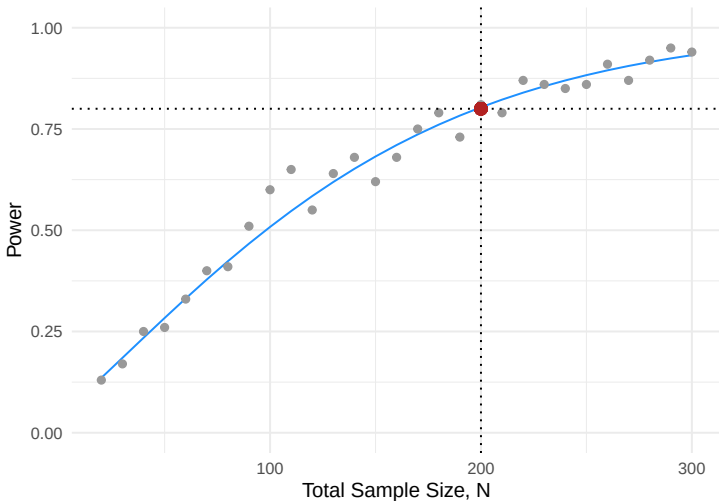
For complex models, even a single fit can be expensive, making brute-force grid search impractical.

Monte Carlo estimates power one design at a time

This is an example of power for an independent-samples t-test with Cohen's $d = 0.4$ and varying N .

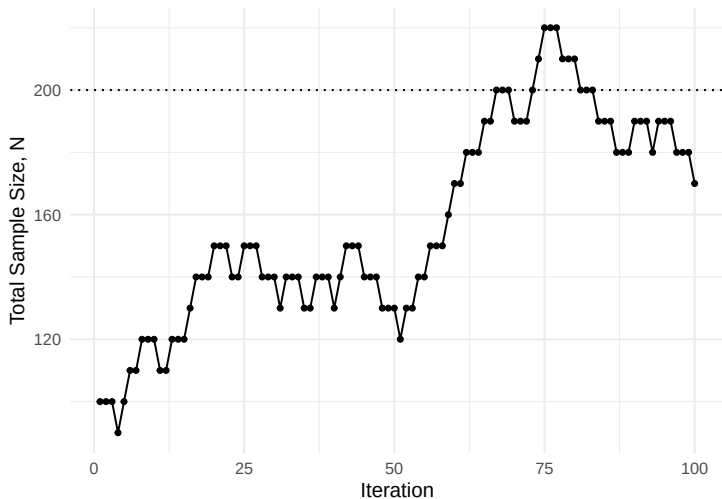


We usually need one point, not the full curve

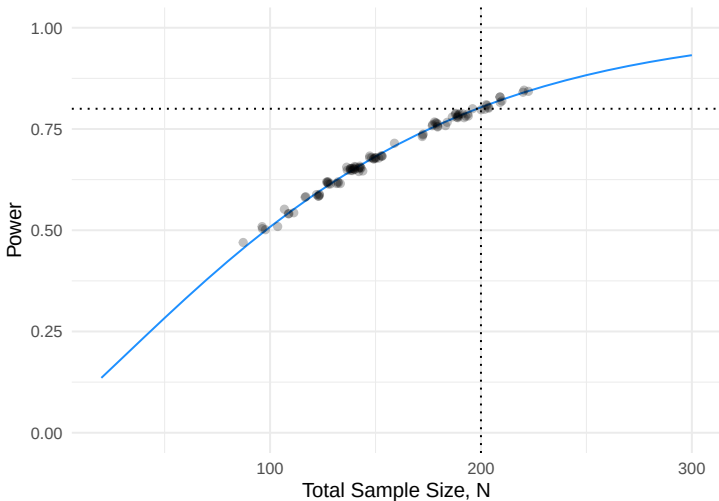


Power analysis as adaptive threshold estimation

Psychophysical staircases move toward a threshold



Adaptive sampling concentrates iterations near N^*



The standard-error-based search procedure

Power depends on the signal-to-noise ratio

For a parameter of interest θ :

$$\hat{\theta} \sim \mathcal{N}(\theta, SE^2)$$

The standardized signal is:

$$\delta = \frac{|\theta|}{SE}$$

For fixed α :

$$\text{Power} = f(\delta) = f\left(\frac{|\theta|}{SE}\right)$$

Larger $\frac{|\theta|}{SE} \implies$ larger power.

Power defines a target standard error

For a two-sided test, approximately:

$$\frac{|\theta|}{SE^*} = z_{1-\alpha/2} + z_{\text{Power}}$$

where

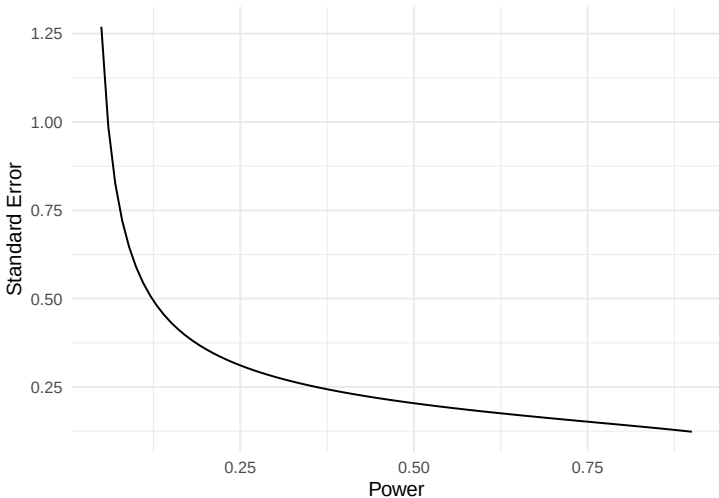
$$z_{\text{Power}} = \Phi^{-1}(\text{Power})$$

Therefore:

$$SE^* = \frac{|\theta|}{z_{1-\alpha/2} + \Phi^{-1}(\text{Power})}$$

$$\text{Target Power} \iff \text{Target } SE^*$$

Power defines a target standard error, graphically



Standard errors shrink predictably with sample size

In many regular models,

$$SE(\hat{\theta}) \approx \frac{C}{\sqrt{N}}$$

where C depends on the model, outcome variability, and design.

Taking logs:

$$\log SE = \log C - \frac{1}{2} \log N$$

More generally:

$$\boxed{\log SE = a + \gamma \log N}$$

γ is estimated empirically; in regular models, $\gamma \approx -\frac{1}{2}$.

A prototype of the algorithm

1. Define the data-generating process
2. Simulate a dataset at N_i and store the estimated standard error $\widehat{SE}_i$
3. Fit the relationship $\log \widehat{SE}_i = a + \gamma \log N_i + \varepsilon_i$
4. Invert the model to estimate the sample size associated with the target standard error: $N^* : \widehat{SE}(N^*) = SE^*$
5. Simulate again at the predicted N^* and update the model
6. Repeat for a fixed number of iterations

This is analogous to an adaptive staircase on the standard error, with sample size as the controlled variable.

Examples and validation

One method, three different models

1. Unequal residual standard deviations between groups
2. Multigroup structural equation model
3. Mixed-effects logistic regression with crossed random-effects

1 — Comparing residual variability

Two groups have different residual standard deviations:

$$\sigma_A \neq \sigma_B$$

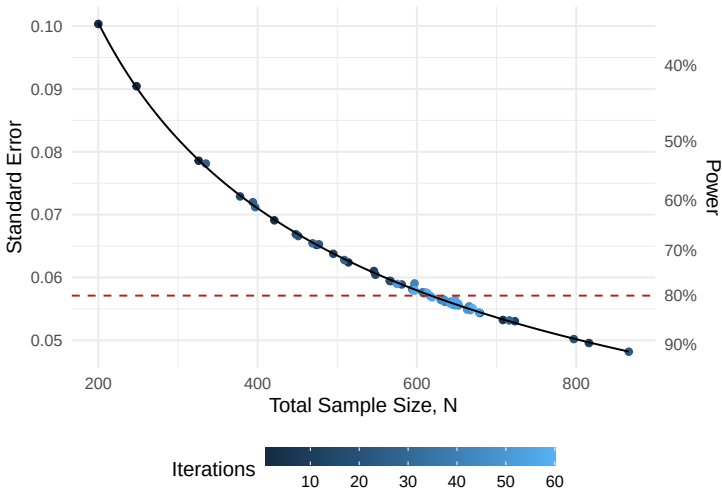
The parameter of interest is the log ratio:

$$\theta = \log \left(\frac{\sigma_B}{\sigma_A} \right) = 0.16$$

The model estimates this parameter directly through the dispersion component.

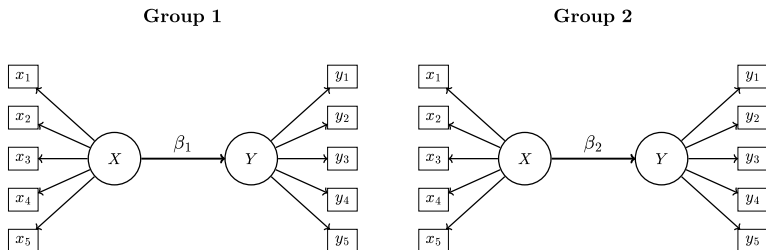
Target: estimate N^* such that power ≈ 0.80

60 adaptive fits suggest $N^* = 618$



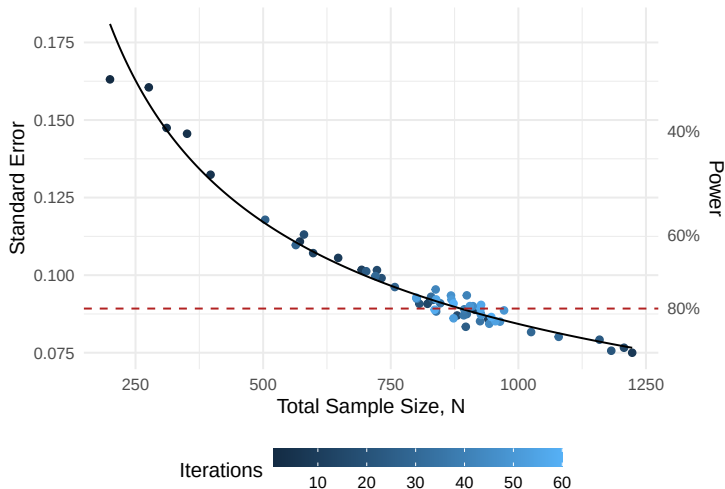
2 — Comparing a structural path across groups

A multigroup SEM with the same latent structure in both groups.
The parameter of interest is the difference between the two structural paths: $\theta = \Delta\beta = \beta_1 - \beta_2 = 0.25$



$$\Delta\beta = \beta_1 - \beta_2$$

60 adaptive fits suggest $N^* = 887$

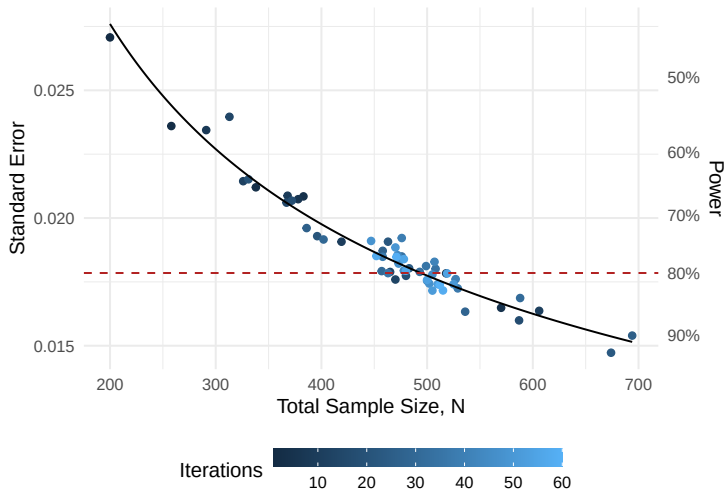


3 — Mixed-effects logistic regression

- ▶ Binary accuracy outcome
- ▶ Within-participant predictor: block
- ▶ Between-participants predictor: group
- ▶ Crossed random effects: participants and items

$$\theta = \beta_{\text{Group} \times \text{Block}} = 0.05$$

60 adaptive fits suggest $N^* = 494$



Did it actually find 80% power?

Model	Estimated N^*	Power at N^*
Residual standard deviation	618	79.1%
Multigroup SEM	887	76.6%
Logistic mixed model	494	81.1%

The selected sample sizes were validated using **5,000 Monte Carlo replications**.

What these examples show — and what they do not

They show

- ▶ The same search procedure can be applied to very different parametric models
- ▶ In these examples, 60 single-dataset fits identified a useful N^*
- ▶ Final power ranged from **76.6% to 81.1%** around the 80% target

They do not yet show

- ▶ General validity across models and effect sizes
- ▶ Robustness to irregular or non-monotonic $SE(N)$ relationships
- ▶ Reliable extension to multiple design parameters

The SE-based method is a search procedure, not a substitute for final Monte Carlo validation.

Selected references on adaptive and efficient algorithms

- Olaniran, O. R., Shamsutdinova, D., Markham, S., Zimmer, F., Stahl, D., Forbes, G., & Carr, E. (2026). Adaptive Gaussian process search for simulation-based sample size estimation in clinical prediction models: validation of the pmsims R package. *BMC Medical Research Methodology*.
<https://doi.org/10.1186/s12874-026-02935-9>
- Chalmers, R. P. (2025). Spower: A general-purpose Monte Carlo simulation power analysis program. *Behavior Research Methods*, 57, 348.
<https://doi.org/10.3758/s13428-025-02787-z>
- Zimmer, F., & Debelak, R. (2025). Simulation-based design optimization for statistical power: Utilizing machine learning. *Psychological Methods*, 30, 513–536. <https://doi.org/10.1037/met0000611>
- Wilson, D. T., Hooper, R., Brown, J., Farrin, A. J., & Walwyn, R. E. (2021). Efficient and flexible simulation-based sample size determination for clinical trials with multiple design parameters. *Statistical Methods in Medical Research*, 30, 799–815. <https://doi.org/10.1177/0962280220975790>
- Ueckert, S., Karlsson, M. O., & Hooker, A. C. (2016). Accelerating Monte Carlo power studies through parametric power estimation. *Journal of Pharmacokinetics and Pharmacodynamics*, 43, 223–234. <https://doi.org/10.1007/s10928-016-9468-y>

✉ filippo.gambarota@unipd.it

🌐 filippogambarota.github.io



Slides

(Last update: 2026-09-03)