Workshop

November 3, 2023

Power Analysis using G*Power

Tom Ledermann, Ph.D.

FLORIDA STATE UNIVERSITY
COLLEGE OF EDUCATION,
HEALTH, AND HUMAN SCIENCES

In designing a study, we commonly want to know what sample size is required to detect a specific effect or in the case the data collection is completed what is the power for a specific effect.

Overview

1. Power
2. Effect Size
3. Effect Size Measures
4. Classification of Effect Sizes
5. Power Analysis
6. G*Power
7. Examples
8. Possible Scenarios
9. Non-Centrality Parameter

Power

Power $(1 - \beta)$ is

- the probability of correctly assuming there is an effect (i.e., correctly rejecting the null hypothesis).
- inversely related to the probability of making a Type II Error (concluding there is no effect when there is one).

	In the population	
	H_0 is true	H_0 is false (H_1 is true)
Decision (Empiry)		
Do not reject H_0	Correct (1 - α)	**Type II Error (β)**
Reject H_0	Type I Error (α)	**Correct (1 - β)**

If power is high, the probability of making a Type II Error (i.e., false negative) is low.

Power

How much power is enough?

A power of .80 is probably the most reasonable compromise between the chance of detecting a substantial effect without involving more recourses in raising the power any higher and the risk of committing a Type I error (i.e., incorrectly concluding there is an effect).

- Studies with low power are statistically underpowered and have a high risk for false negative conclusions (Type II error).
- Studies with high power have a high risk for false positive conclusions (Type I error).

$\Rightarrow$ Each study should be adequately powered.

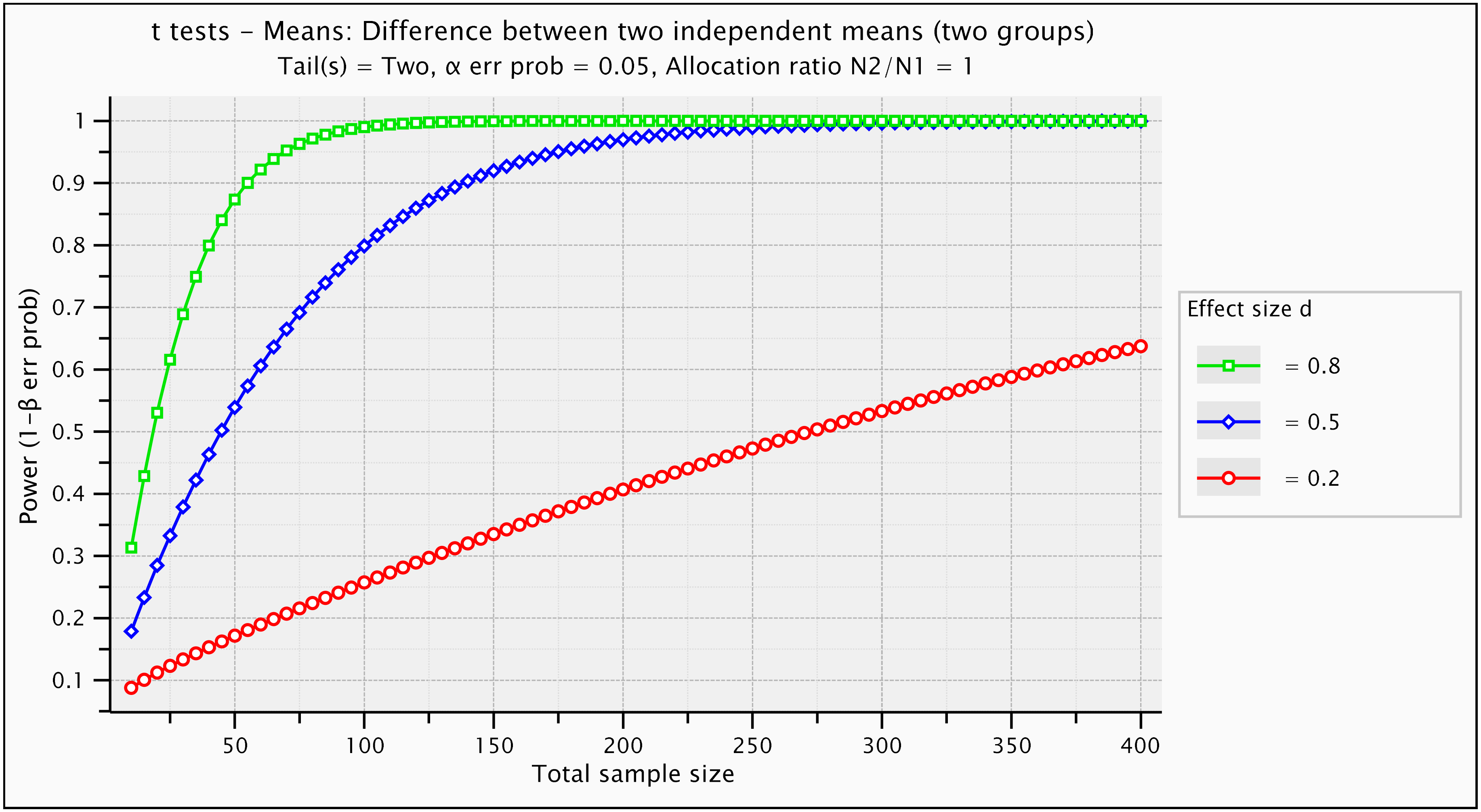
Power

Statistical Power depends on:

- the Type of Statistical Test (statistical model)
- the Size of the Effect (↑)
- the Sample Size N (↑)
- the Significance Level α (↑)

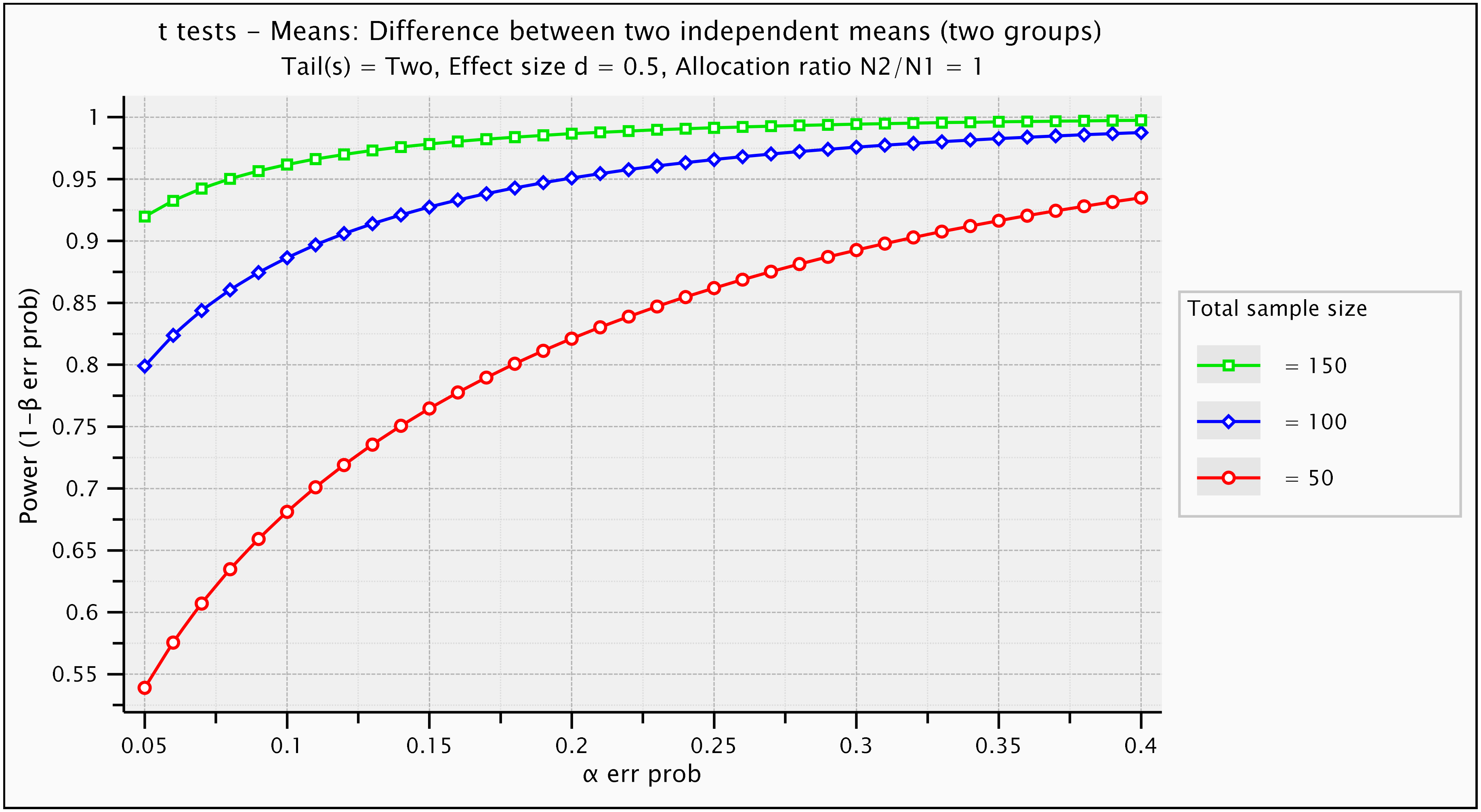
Power

Power as a function of *N* and Effect Size (Cohen's *d*)



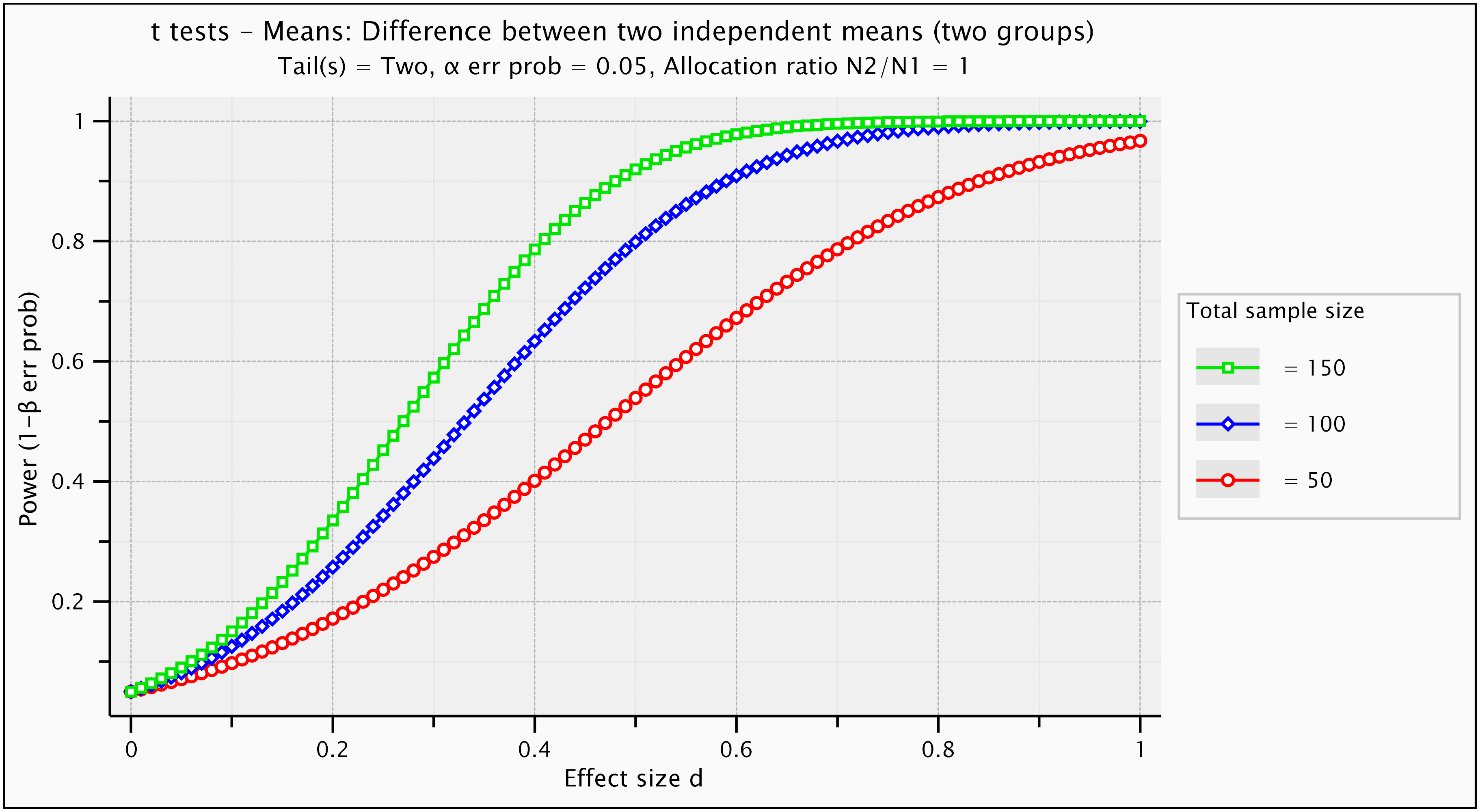
Power

Power as a function of α and *N*



Power

Power as a function of Effect Size (Cohen's *d*) and *N*



Power

Ways to Increase Power

- Increase the sample size.
- Chose a higher alpha level.
- Use a one-tailed test instead of a two-tailed test (not always possible and against consensus).
- Increase the effect size.

In comparing groups, the effect size can be increased by:

- increasing the variability between the groups (bigger difference).
- reducing the variability within groups.

Effect Size

Effect size (ES) is a name given to a family of indices that measure the magnitude of an effect (e.g., the difference between groups or relationships between variables) in a standardized way.

Unlike significance tests, effect sizes are independent of the sample size.

Effect Size

Effect sizes are used

- to judge the practical importance.
- to determine the sample size needed for detecting a specific effect and estimate power in cases where raw data are not available.
- to compare results across studies (and methods).
  $\Rightarrow$ meta-analysis studies

Effect Size Measrues

Effect Size	**Statistical Analysis**
r	Correlation Analysis
R^2	Regression Analysis
η^2	ANOVA family
Cohen's f	Regression Analysis and ANOVA
Odds Ratio	Logistic Regression Analysis
Cohen's d	Difference between 2 means
Δ	Hotelling T^2 test
w	Contingency table (χ^2-test)

Effect size converters:

https://www.escal.site/#

https://www.psychometrica.de/effect_size.html

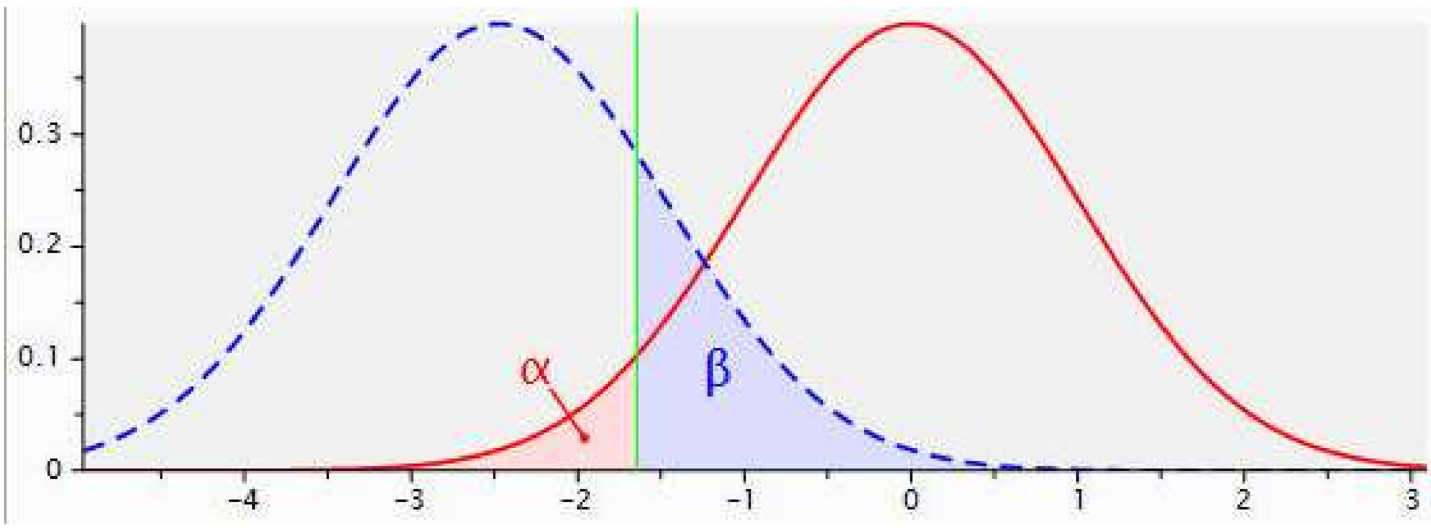
Classification of Effect Sizes

Jacob Cohen (1988) suggested benchmarks for many effect sizes to classify effects as

- small in size
- medium in size
- large in size

Power Analysis

Power analysis refers to statistical methods used to determine the sample size required to detect a specific effect or to estimate power for a specific effect.



Power Analysis

Important types of power analyses

- **A Priori Power Analyses**
  compute *N* as a function of power, α, and the (assumed) population ES.
- **Post Hoc Power Analyses**
  compute power (also know as estimated power) as a function of α, *N*, and ES.
- **Sensitivity Analyses**
  compute the critical population ES as a function of α, power, and *N*.

Power Analysis

Other types of power analyses offered by G*Power

- **Criterion Analyses**
  compute α as a function of power, the population ES, and *N*.
- **Compromise Power Analyses**
  compute both α and power as a function of *N*, the population ES, and the error probability ratio *q*, which is β/α.

G*Power

G*Power was developed by Faul et al. (2009) and is a free-to-use stand-alone power analysis software program.

G*Power can be used for the following

- statistical tests: F test, t test, χ^2 test, and z test.

- statistical methods:
  - correlation and regression analysis
  - analysis of means
  - analysis of proportions
  - analysis of variance

Example 1

Research Design

Three-wave longitudinal study with two groups.

Question

What sample size is required to detect a medium-sized time-by-group interaction effect with power of .80?

Assumptions

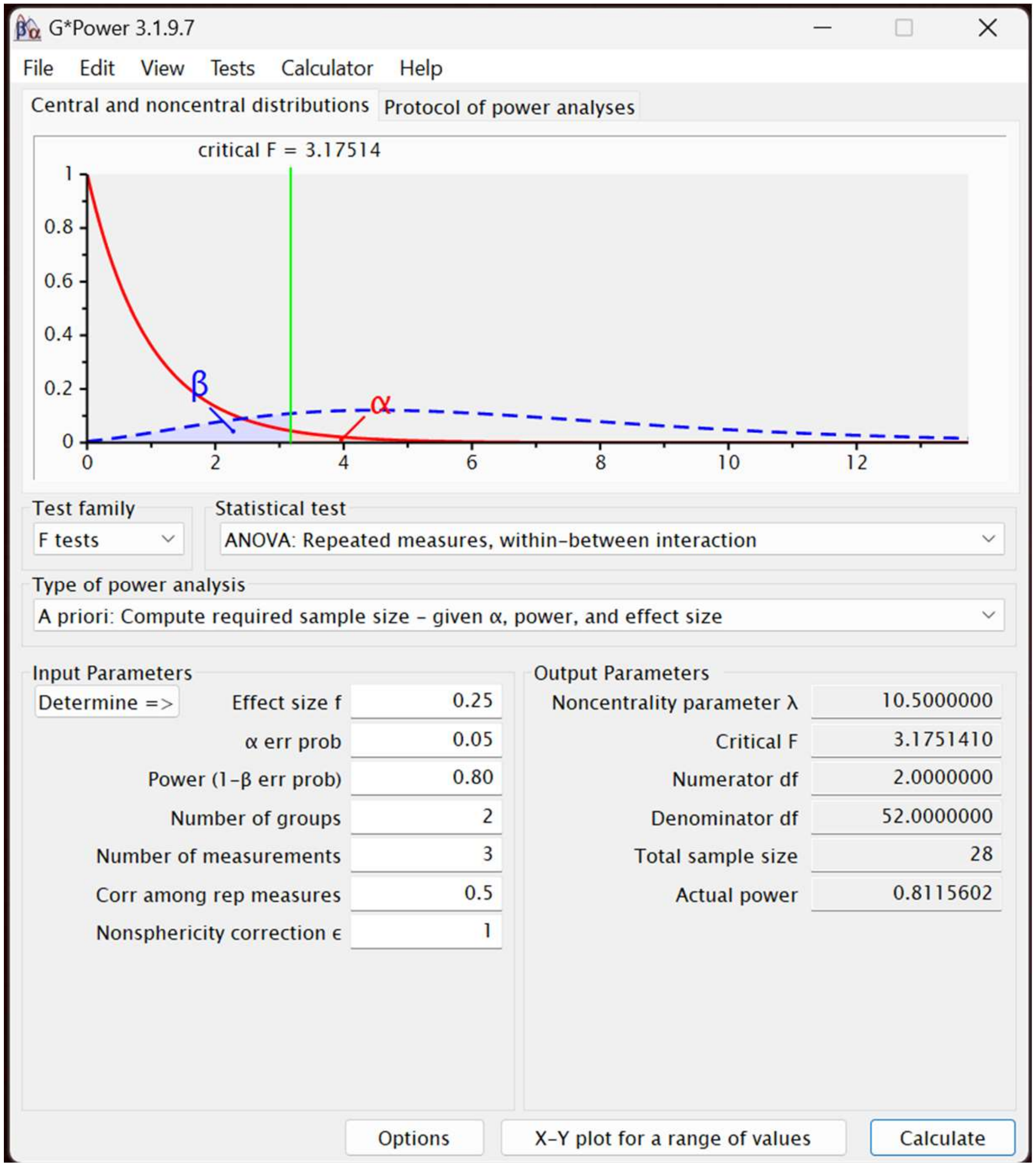
15% missing data

Significance level $\alpha = .05$

Autocorrelation $\rho = .50$

Non-sphericity correction $\varepsilon = 1$

Example 1



Required sample size:

28 complete cases (14 in each group)

Plus 15% missings (28/.85):

32.9 cases total (17 in each group)

The same number of cases would be required for the within-subject (time) effect.

For the between-subject (group) effect, 86 cases would be needed.

Example 2

Research Design

Regression model with 3 covariates and 3 predictor variables.

$$Y = b_0 + b_1 C_1 + b_2 C_2 + b_3 C_3 + b_4 X_1 + b_5 X_2 + b_6 X_3 + e$$

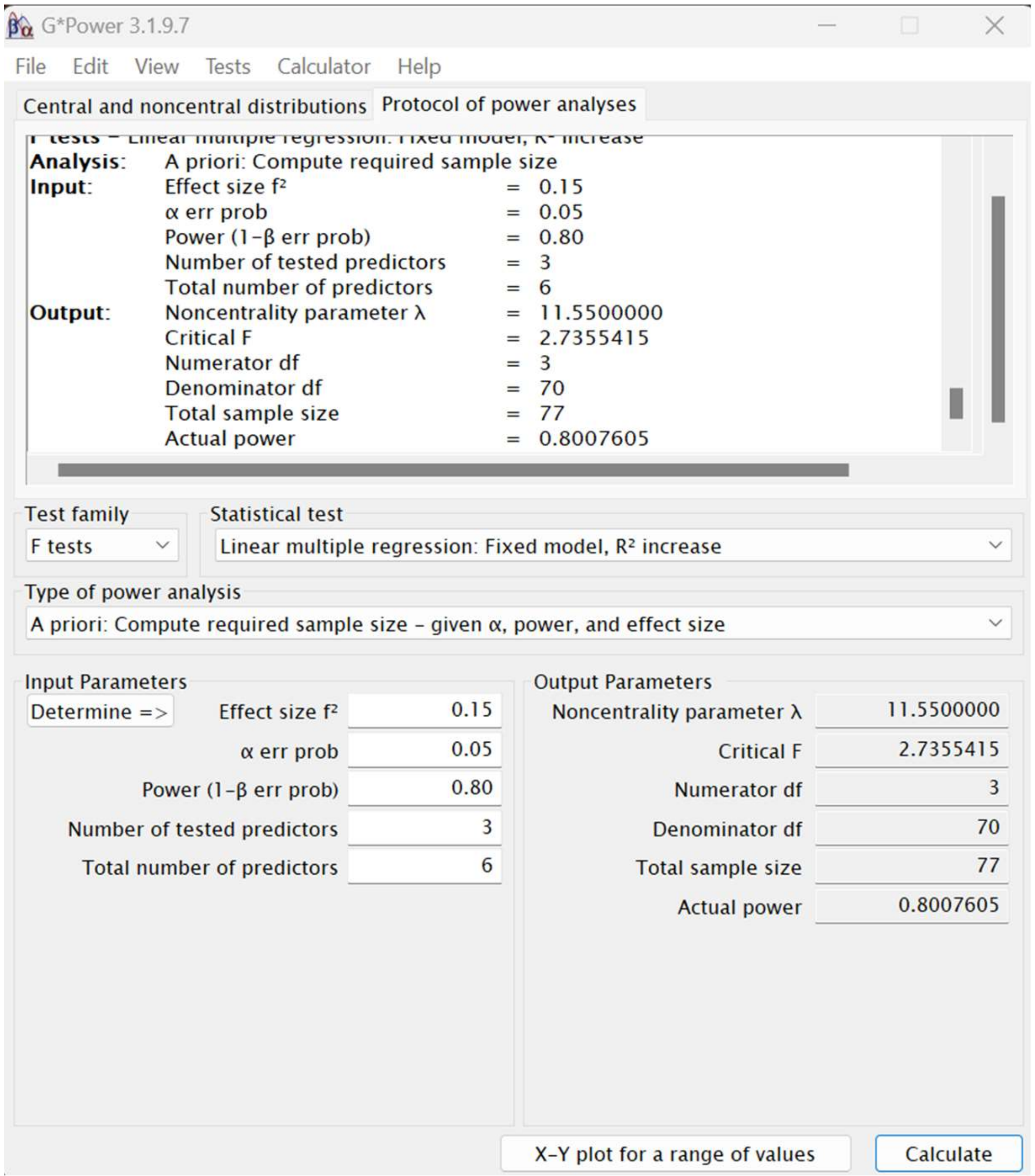
Question

What sample size is needed to detect a medium-sized incremental effect (R^2 increase) for the 3 predictor variables?

Assumption

Significance level α = .05

Example 2



Required sample size:

77 complete cases

Small effect:

550 cases

Large effect:

36 cases

Possible Scenarios

There are two rare but possible scenarios:

- Power is high and the *p* value of an effect is high (not statistically significant).
- The *p* value of an effect is low (statistically significant) although power is low.

In either case, further investigation of the data is warrant.

Noncentrality Parameter

- Many statistical tests involve a test statistic with a normal, *t*, *F*, or chi-square distribution.
- When H_0 is true, then these test statistics have a central distribution (e.g., central *t* distribution).
- When H_0 is false, the test statistics have a non-central distribution, which depends on the size of the effect.

Noncentrality Parameter

The noncentral distributions

- show how the test statistics are distributed when there is an effect (i.e., H_0 is false).
- have an additional parameter, the noncentrality parameter δ (delta).

A central distribution is a special case of a noncentral distribution where the δ is zero.

Noncentrality Parameter

The noncentrality parameter δ represents the degree to which the mean of the sampling distribution of the test statistic departs from its mean when the null is true.

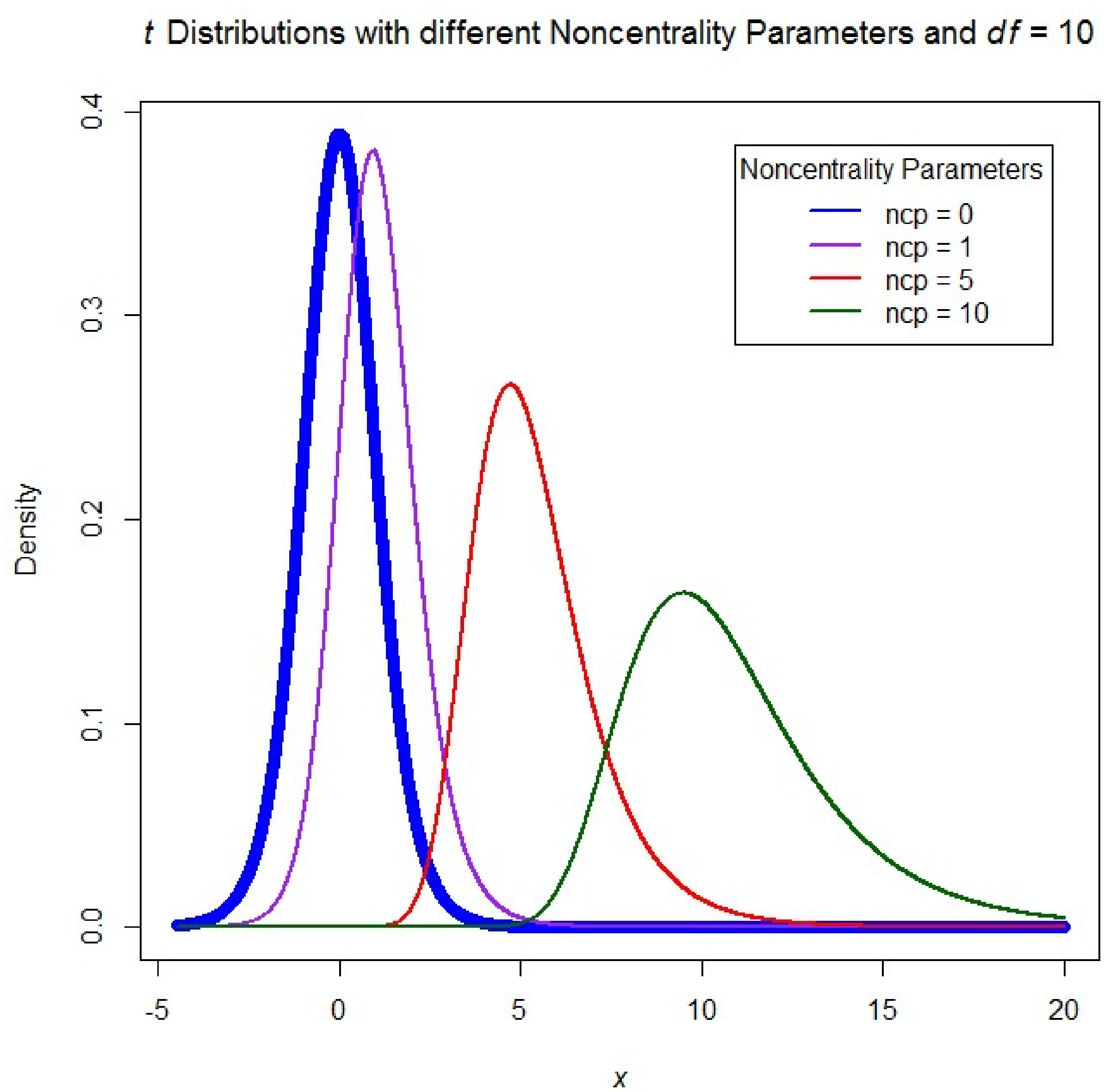
For the two-sample *t* test, the noncentrality parameter is

$$\delta = \text{Cohen's } d \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

When the two samples are equal in size, then

$$\delta = \text{Cohen's } d \sqrt{\frac{n}{2}}$$

Noncentrality Parameter



References and Readings

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112* (1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Fan, X. (November, 1999). *Statistical Significance and Effect Size: Two Sides of a Coin.* Paper presented at the Annual Meeting of the American Evaluation Association, Orlando, FL.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Kang H. (2021). Sample size determination and power analysis using the G*Power software. *Journal of Educational Evaluation for Health Professions, 18*, 17. https://doi.org/10.3352/jeehp.2021.18.17

Lan, L., & Lian, Z. (2010). Application of statistical power analysis–How to determine the right sample size in human health, comfort and productivity research. *Building and Environment, 45*(5), 1202-1213. https://doi.org/10.1016/j.buildenv.2009.11.002

Thank you!

Email: tledermann@fsu.edu

GitHub: https://github.com/thomasledermann/powerAPIM