EqualStrength power analysis

MODIFIED
July 10, 2023

Table of contents

1 Introduction

This analysis estimates the statistical power for EqualStrength tests across different sample sizes. Power is the probability of rejecting the null hypothesis when it is, in fact, false. In other words, the probability of correctly rejecting the hypothesis of *non-discrimination* when there is discrimination.

1.1 Limitations

As with any power analysis, the scope of this analysis is limited to the statistical significance of the test only. Apart from the general limitations of the NHST approach[1], this analysis ignores the probability of the estimate being in the wrong direction or exaggerated [2].

2 Main parameters

We estimate power as a function of a given i) sample size, ii) significance level, and iii) effect size.

▶ Code

For the sample size, we use a sequence of different sizes varying from 100 to 6000. The *significance* level is determined by convention [3]:

- **Alpha**: 0.05.

The effect size can be specified either from previous studies or by *assuming* a "minimum effect that would be substantively important" [2] or interesting [4]. An overview of recent studies indicate a substantial variation of effect sizes depending on the setting and country[1]. Several studies also highlight that any postulation of effect sizes based on previous studies is severely limited by publication bias, population effect size heterogeneity, and model error [7], which can lead to "overly optimistic" expectations [4]. Here we posit the following callback rates for the main treatment:

▶ Code

- **Majority group call-back rate**: 0.4
- **Minority group call-back rate**: 0.3

We then use these rates to estimate the effect size (expressed as Cohen's h[3]):

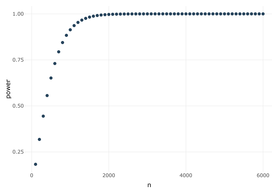
▶ Code

- **Effect size of the main treatment**: -0.21.

2.1 Simple estimate (PWR)

Based on these parameters, it is possible to estimate power assuming a basic design with two groups of the same size. The package pwr [8] contains functions based on the formulas developed by Jacob Cohen[3] to calculate power. The following estimates are based on a power calculation for two proportions (with same sample sizes).

▶ Code



3 Specific parameters

Here we use three different approaches to incorporate specific features from the EqualStrength research design. First we incorporate a second treatment and an interaction term between the two treatments.

▶ Code

The effect size of the second treatment is postulated to be the same as the one of the main treatment (-0.21). The interaction effect size is assigned as 25% or 50% the effect size of the main treatment, which results in the following values:

- **Interaction effect size (lower)**: 0.05.
- **Interaction effect size (higher)**: 0.11.

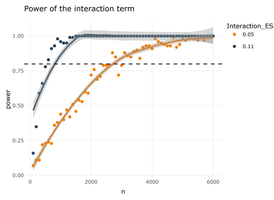
3.1 Gelman, Hill, and Vehtari (GHV)

In GHV[9], the authors highlight that it is necessary "4 times the sample size to estimate an interaction that is the same size as the main effect" and estimate a 16 times larger sample for an interaction half the size of the main effect. Using the previous estimates from Section 2.1 and a power >80%, we would get a sample size of 3,200 for an interaction effect as large as the main effect and about 12,800 for an effect half the size.

3.2 Package InteractionPoweR (IPR)

Using the package InteractionPoweR (IPR)[10] we can use the interaction effect size as a parameter for simulations.[2] It is assumed that the treatment variables are not correlated but this can be changed in the r.x1.x2 argument of the function.

▶ Code

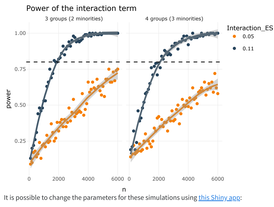


3.3 Edvard N. Larsen (ENL)

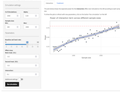
Using the R code shared by Edvard Nergård Larsen, it is possible to calculate the power with additional parameters. The estimates result from linear models [11] of simulated datasets incorporating different combinations of parameters present in the EqualStrength research design.

Here we also get estimates for main treatments using two or three minority groups in addition to the majority group.

▶ Code



It is possible to change the parameters for these simulations using this Shiny app:



4 Key points

- There are large differences depending on the method adopted. GHV suggests 4,000 and IPR indicates less than 1,000 using similar parameters. But the method suggested by ENL is the most appropriate as it incorporates attributes of the EqualStrength research design.
- Following this method, we estimate that at least 1,700 observations equally distributed across 3 ethnic groups are necessary to get >= 80% power with a main effect size of -0.21 and an interaction effect size of 0.11.
- After trying different parameters in the shiny app, the most influential component for the sample size seems to be the effect size of the interaction term:

Maj. call-back	Main treat.	Interaction	Min. sample 80%
0.3	-0.3	0.075	600
0.3	-0.3	0.15	600
0.3	-0.1	0.025	17,600
0.3	-0.1	0.05	5,600
0.4	-0.3	0.075	2,600
0.4	-0.3	0.15	1,100
0.4	-0.1	0.05	10,100
0.5	-0.3	0.075	3,100
0.5	-0.3	0.15	11,00
0.5	-0.1	0.05	12,600

5 References

1. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon Statistical Significance. The American Statistician [Internet]. 2019 Mar [cited 2023 Jul 4];73(sup1):235–45. Available from: https://doi.org/10.1080/00031305.2018.1527253
2. Gelman A, Carlin J. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspectives on Psychological Science [Internet]. 2014 Nov [cited 2023 Jul 4];9(6):641–51. Available from: https://doi.org/10.1177/1745691614551642
3. Cohen J. The Concepts of Power Analysis. In: Statistical Power Analysis for the Behavioral Sciences [Internet]. Elsevier; 1977 [cited 2023 Jul 4]. p. 1–17. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780121790608500062
4. Lakens D. Sample Size Justification. Collabra: Psychology [Internet]. 2022 Mar [cited 2023 Jul 10];8(1):33267. Available from: https://doi.org/10.1525/collabra.33267
5. Ramos M, Thijsen L, Coenders M. Labour market discrimination against Moroccan minorities in the Netherlands and Spain: A cross-national and cross-regional comparison. Journal of Ethnic and Migration Studies [Internet]. 2021 Apr [cited 2023 Jul 10];47(6):1261–84. Available from: https://doi.org/10.1080/1369183X.2019.1622824
6. Larsen EN, Di Stasio V. Pakistani in the UK and Norway: Different contexts, similar disadvantage. Results from a comparative field experiment on hiring discrimination. Journal of Ethnic and Migration Studies [Internet]. 2021 Apr [cited 2023 Jul 10];47(6):1201–21. Available from: https://doi.org/10.1080/1369183X.2019.1622777
7. Anderson SF, Kelley K, Maxwell SE. Sample-Size Planning for More Accurate Statistical Power: A Method Adjusting Sample Effect Sizes for Publication Bias and Uncertainty. Psychological Science [Internet]. 2017 Nov [cited 2023 Jul 5];28(11):1547–62. Available from: http://journals.sagepub.com/doi/10.1177/0956797617723724
8. Champely S, Ekstrom C, Dalgaard P, Gill J, Weibelzahl S, Anandkumar A, et al. Pwr: Basic functions for power analysis. 2017;
9. Gelman A, Hill J, Vehtari A. Regression and Other Stories [Internet]. 1st ed. Cambridge University Press; 2020 [cited 2023 Jul 5]. Available from: https://www.cambridge.org/highereducation/product/9781139161879/book
10. Baranger DA, Finsaas M, Goldstein B, Vize C, Lynam D, Olino TM. Tutorial: Power analyses for interaction effects in cross-sectional regressions [Internet]. PsyArXiv; 2022 Aug [cited 2023 Jul 4]. Available from: https://osf.io/5atd7
11. Gomila R. Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. Journal of Experimental Psychology: General [Internet]. 2021 Apr [cited 2023 Jul 4];150(4):700–9. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/xge0000920

Footnotes

1. Looking at some previous studies, Ely found baseline callback rates ranging from 60-20%, CB rate differences from -31% (-0.31) to +7.3 (0.073). Ethnicity differences in some GEMM papers [5,6] runs from -17 to -6, and gender from +7 to +4.[6]
2. The developers indicate that they would like to include 3-way interactions in a future version, but emphasise that "There are several steps in the analysis where we do not know exactly what to do for 3-way interactions" ↩