Power calculation for causal inference in social science: Sample size and minimum detectable effect determination

Description

Eric W Djimeu, Deo-Gracias Houndolo, 3ie Working Paper 26, March 2016. Available as pdf

Contents
1. Introduction
2. Basic statistics concepts: statistical logic
3. Power calculation: concept and applications
3.1. Parameters required to run power calculations
3.2. Statistical power and sample size determination
3.3. How to run power calculation: single treatment or multiple treatments?
4. Rules of thumb for power calculation
5. Common pitfalls in power calculation
6. Power calculations in the presence of multiple outcome variables
7. Experimental design
7.1. Individual-level randomisation
7.2. Cluster-level randomisation

1. Introduction

Since the 1990s, researchers have increasingly used experimental and quasi-experimental primary studies – collectively known as impact evaluations – to measure the effects of interventions, programmes and policies in low- and middle-income countries. However, we are not always able to learn as much from these studies as we would like. One common problem is when evaluation studies use sample sizes that are inappropriate for detecting whether meaningful effects have occurred or not. To overcome this problem, it is necessary to conduct power analysis during the study design phase to determine the sample size required to detect the effects of interest. Two main concerns support the need to perform power calculations in social science and international development impact evaluations: sample sizes can be too small and sample sizes can be too large.

In the first case, power calculation helps to avoid the consequences of having a sample that is too small to detect the smallest magnitude of interest in the outcome variable. Having a sample size smaller than statistically required increases the likelihood of researchers concluding that the evaluated intervention has no impact when the intervention does, indeed, cause a significant change relative to a counterfactual scenario. Such a finding might wrongly lead policymakers to cancel a development programme, or make counterproductive or even harmful changes in public policies. Given this risk, it is not acceptable to conclude that an intervention has no impact when the sample size used for the research is not sufficient to detect a meaningful difference between the treatment group and the control group.

In the second case, evaluation researchers must be good stewards of resources. Data collection is expensive and any extra unit of observation comes at a cost. Therefore, for costefficiency and value-for-money it is important to ensure that an evaluation research design does not use a larger sample size than is required to detect the minimum detectable effect (MDE) of interest. Researchers and funders should therefore use power calculations to determine the appropriate budget for an impact evaluation study.

Sample size determination and power calculation can be challenging, even for researchers aware of the problems of small sample sizes and insufficient power. 3ie developed this resource to help researchers with their search for the optimal sample size required to detect an MDE in the interventions they evaluate.

The manual provides straightforward guidance and explains the process of performing power calculations in different situations. To do so, it draws extensively on existing materials to calculate statistical power for individual and cluster randomised controlled trials. More specifically, this manual relies on Hayes and Bennett (1999) for cluster randomised controlled trials and documentation from Optimal Design software version 3.0 for individual randomised controlled trials.

Category

1. Uncategorized

Date
23/11/2024
Date Created
09/04/2016
Author
admin