

11.2: Prospective and retrospective power

Introduction

Statistical power is defined as $p = 1 - \beta$. The power of a test is the probability of rejecting a null hypothesis when it is false. There is a relationship between Type I error and Type II error. We want β to be large, BUT β is generally not known when we are performing the statistical test. We do know that α is inversely related to β . The smaller the α value we use to reject the null hypothesis, the MORE likely we will accept a FALSE Null Hypothesis. If we make α very small (one in a billion) then there would be a very HIGH chance of accepting α Null Hypothesis when it is false (β is high). Note that this is the same discussion that we had about sensitivity of an assay test and the specificity of that assay. If we increase sensitivity of the assay such that we approach 100% detection of the true positives will be detected, we necessarily will increase the number of false positives. The BEST way to reduce both Type I and Type II statistical errors is to INCREASE the sample size!

Power of any statistical test (z-test, t-test, one-way ANOVA...) can be determined BEFORE the experiment is done and data are gathered or AFTER an experiment is completed (Cohen 1992). For now, here's our first real taste of experimental design — we can evaluate how we can make an experiment to test a particular hypothesis.

Prospective power analysis

Good experimental design should include considerations of power. The design will determine the size of the effect your experiment will be able to detect and the probability of correctly rejecting the null hypothesis. In large part, this will involve decisions of sample size and ways to reduce error variance. Moreover, one needs to decide ahead of time, just how large of an effect does the experiment need to detect? A one-gram change in body mass before and after a diet treatment is of no concern whatsoever if your study subjects are African elephants, but may be a very large effect for a study of shrews!

Retrospective power analysis

Power analysis can also be done after a test has been conducted and the null hypothesis failed to be rejected. One interpretation that might follow from a retrospective power analysis is that, if the study had low power, the lack of statistical significance could be viewed merely as the result of low sample size. However, as forcefully argued by Hoenig and Heisey (2001) (see also Colegrave and Ruxton 2003), retrospective or post-hoc power tests provide no more information than does the p-value and therefore are redundant. At worse, retrospective power analysis can be misleading as to how well it predicts true power, i.e., biologically meaningful differences between different treatment groups (Zhang et al., 2019).

Prospective power is more than effect size between groups

Large effect size, i.e., large differences between groups, is not necessarily evidence of important biological differences. The concept of power has limits (Hoenig and Heisey 2001, Yuan and Maxwell 2005). On the other hand, small effect sizes can be important differences, especially if the treatment is difficult or expensive. We work through this conclusion with an example, but emphasize here that providing confidence intervals for the effect size (Colegrave and Ruxton 2003). Suppose the null hypothesis was not rejected from a [two-sample independent t-test](#) conducted on the test of differences in plant height between two samples of 'ohi'a found at different elevations. Was there really no difference or was the sample size in the study simply too small to detect the real and important difference between the samples? By conducting a power analysis, one can determine if a slight increase in sample size would have yielded a statistically significant difference, or it may suggest that the effect size is small enough not to warrant further attention.

R code

Three options for conducting power analysis in R are provided in [Section 11.5](#).

Questions

1. Assuming that a study was done by randomly sampling from a population, and the the primary outcome is found not statistically different with p-value 0.13 between placebo and treatment groups, what can be gained made from a retrospective power analysis?

This page titled [11.2: Prospective and retrospective power](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.