Power and Multiple Hypothesis Testing

Garret Christensen[1]

[1]UC Berkeley:
Berkeley Initiative for Transparency in the Social Sciences
Berkeley Institute for Data Science

IDB, March 2018
Slides available online at
http://www.github.com/BITSS/IDBMarch2018

Outline

BITSS

BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

What is Statistical Power?

The power of a statistical hypothesis test is the probability that the test correctly rejects the null hypothesis when it is false.

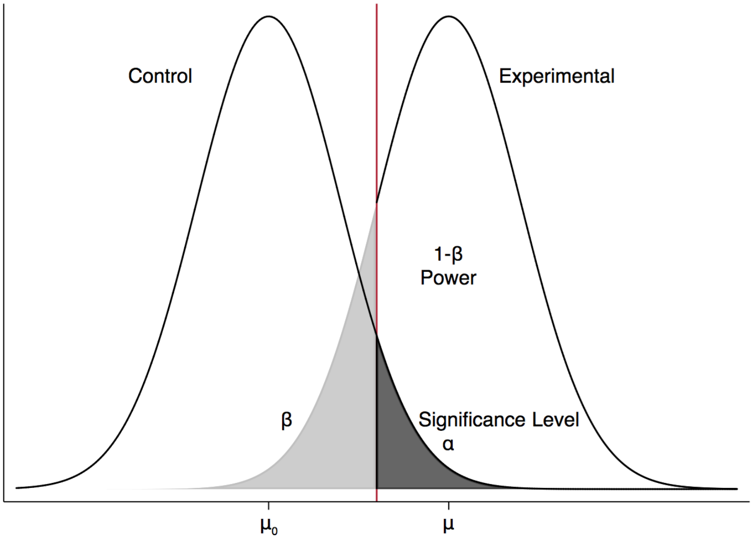
That is, if there's a real effect, what's the likelihood you'll detect it? 80% is the standard.

What is Statistical Power?

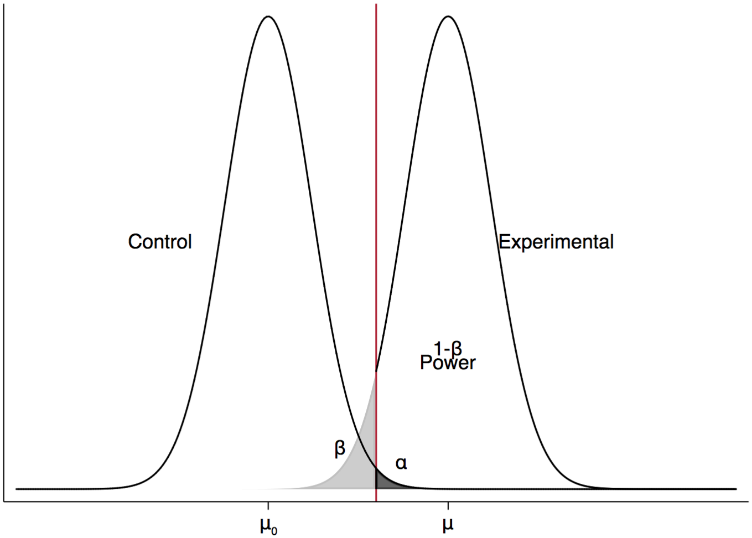
In terms of Type I (false positive) and Type II (false negative) errors:

- Type I error rate is α
- Type II error rate is β
- Power is $1 - \beta$.

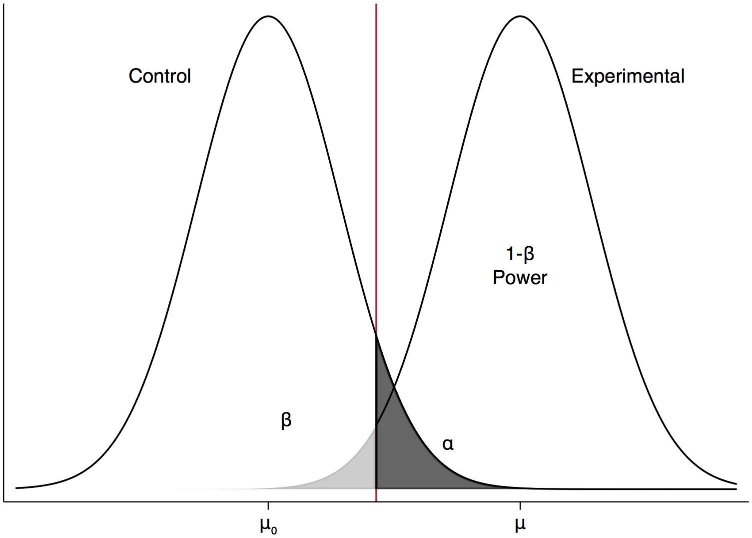
$Power = 1 - \beta$



Less noise, more power



Larger true effect, more power



Derivation

$$Power = 1 - \beta = Pr(Y \geq \ \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}|H_1 : \ \mu > \mu_0)$$

$$= 1 - Pr(Y < \ \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}|H_1)$$

$$= 1 - Pr(\frac{Y - \mu}{\frac{\sigma}{\sqrt{n}}} < \ \frac{\mu_0 + \frac{z_{1-\alpha}\sigma}{\sqrt{n}} - \mu}{\frac{\sigma}{\sqrt{n}}}|H_1)$$

$$= 1 - Pr(\frac{Y - \mu}{\frac{\sigma}{\sqrt{n}}} < \ \frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} + z_{1-\alpha}|H_1)$$

$$= 1 - \Phi(\ \frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} + z_{1-\alpha}|H_1)$$

$$= \Phi(\ \frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} - z_{1-\alpha}|H_1)$$

Increasing Power

$$= \Phi\left(\frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} - z_{1-\alpha} \middle| H_1 \right)$$

Hopefully the equation makes clear that:

- larger n
- lower σ
- larger true effect size $(\mu_0 - \mu)$
- and a larger α, though that's kind of cheating

all increase power.

Rearranging

Rather than solving for power, you may want to solve for the minimum detectable effect (MDE).

$$MDE = (t_\beta + t_\alpha) * \sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{n}}$$

Or, if you've got unlimited funds, pick the minimum biologically or practically meaningful effect, (or your estimate from previous literature of how big the effect will be) and solve for n.

Design Effect

We've so far assumed independent observations, which isn't the case if we cluster treatment. Multiply MDE by the Design Effect:

$$\sqrt{1 + (n - 1)\rho}$$

Where n is households per sampling unit, and ρ is the intracluster correlation–variance between clusters divided by sum of within and between.

Complications

Clusters not equal sized? Use the coefficient of variation, but it may not matter much. (Eldridge, Ashby, Kerry 2006)

You get the most power with equal proportions of treated/control. If treatment is very expensive, maximize power subject to your budget constraint. (Randomization Toolkit: Duflo, Glennerster, and Kremer 2007)

Panel with serial correlation? (Burlig, Preonas, Woerman 2017)

Complicated? Just simulate it. (Arnold et al. 2011)

Problem of Low Power

So what happens if we have low power?

- More false negatives (Type II error, just β).
- More false positives! More precisely, the likelihood that a reported effect represents a true finding decreases.

Ioannidis 2005

"Why most published research findings are false" (Ioannidis 2005), cited 5600 times.

$$PPV = Pr(True|T > t_\alpha)$$

$$= \frac{(1-\beta)\cdot R}{(1-\beta)R+\alpha}$$

- R is ratio of true relationships to non-relationships tested in a literature.

Derivation

How Common in Economics?

Quite

THE POWER OF BIAS IN ECONOMICS RESEARCH*

John P. A. Ioannidis, T. D. Stanley and Hristos Doucouliagos

We investigate two critical dimensions of the credibility of empirical economics research: statistical power and bias. We survey 159 empirical economics literatures that draw upon 64,076 estimates of economic parameters reported in more than 6,700 empirical studies. Half of the research areas have nearly 90% of their results under-powered. The median statistical power is 18%, or less. A simple weighted average of those reported results that are adequately powered (power ≥ 80%) reveals that nearly 80% of the reported effects in these empirical economics literatures are exaggerated; typically, by a factor of two and with one-third inflated by a factor of four or more.

> Statisticians routinely advise examining the power function, but economists do not follow the advice.
>
> McCloskey (1985, p. 204)

Ioannidis, Stanley, Doucouliagos 2017

If we adopt the conventional 5% level of statistical significance and 80% power level, as well, then the 'true effect' will need to be 2.8 standard errors from zero to discriminate it from zero. The value of 2.8 is the sum of the usual 1.96 for a significance level of 5% and 0.84 that is the standard normal value that makes a 20/80% split in its cumulative distribution. Hence, for a study to have adequate power, its standard error needs to be smaller than the absolute value of the underlying effect divided by 2.8. We make use of this relationship to survey adequate power in economics.

Questions?

Thank you!

Derivation of PPV I

$$PPV = Pr(True|T > t_\alpha)$$

Prior to the study, the quantities involved are as follows:

- Probability of a relationship being true: $\frac{R}{R+1}$
- Probability of a relationship being false: $1 - \frac{R}{R+1} = \frac{1}{R+1}$
- Probability of finding a positive statistical association given that the relationship is false: α
- Probability of finding a positive statistical association given that the relationship is true (i.e., power): $1 - \beta$

Derivation of PPV II

Bayes' law says that $Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)}$, though it is almost always the case that the denominator is more useful when written out with the law of total probability, as follows: $Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B|A)Pr(A)+Pr(B|\neg A)Pr(\neg A)}$.
By using Bayes' law, we know that:

$$Pr(True|T > t_\alpha) = \frac{Pr(T > t_\alpha|True) \cdot Pr(True)}{Pr(T > t_\alpha|True) \cdot Pr(True) + Pr(T > t_\alpha|False \cdot Pr\,(\text{False}))}$$

Derivation of PPV III

Substituting, we find:

$$Pr(True|T > t_\alpha) = \frac{(1-\beta)\frac{R}{R+1}}{(1-\beta)\frac{R}{R+1} + \alpha \cdot \frac{1}{R+1}}$$

$$Pr(True|T > t_\alpha) = \frac{\frac{(1-\beta)\cdot R}{R+1}}{\frac{(1-\beta)R+\alpha}{R+1}}$$

Simplifying:

$$Pr(True|T > t_\alpha) = \frac{(1-\beta)\cdot R}{(1-\beta)R+\alpha} = \frac{(1-\beta)R}{R-\beta R+\alpha}$$

This is the same as the formula in Ioannidis (2005) and equation 1 above. Back