Power calculations in RD designs
Northwestern Causal Inference Workshop

Gonzalo Vazquez-Bare

Department of Economics, UC Santa Barbara

August 1, 2024

Overview: power and sample size calculations

- Statistical power: prob of rejecting $\mathsf{H}_0 : \tau = 0$
  - Prob of detecting a non-zero effect
- Power calculation:
  - Statistical power given a sample size and $\tau = \tau_{\mathsf{A}} \neq 0$
- Sample size calculation:
  - Number of obs required to achieve a certain power for a given effect
- Minimum detectable effect (MDE) calculation:
  - Smallest effect detectable at a given power and sample size

Introduction: Experimental Design

- ATE estimator:

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0 = \frac{\sum_{i=1}^n Y_i D_i}{N_1} - \frac{\sum_{i=1}^n Y_i (1 - D_i)}{N_0}$$

where

$$N_1 = \sum_{i=1}^n D_i, \qquad N_0 = \sum_{i=1}^n (1 - D_i)$$

- Central limit theorem:

$$\sqrt{n}(\hat{\tau} - \tau_{\mathsf{ATE}}) \to_{\mathcal{D}} \mathcal{N}(0, \mathsf{V}), \quad \mathsf{V} = \frac{\sigma_1^2}{p} + \frac{\sigma_0^2}{1-p}$$

where $p = \mathbb{P}[D_i = 1]$

Power function

$$\mathsf{H}_0: \quad \tau_{\mathsf{ATE}} = 0 \qquad vs \qquad \mathsf{H}_\mathsf{A}: \quad \tau_{\mathsf{ATE}} \neq 0$$

- Statistic:

$$T = \frac{\hat{\tau}}{\sqrt{\mathsf{V}/n}}$$

- Power function:

$$\beta(\tau) = \mathbb{P}_\tau[\text{reject } \mathsf{H}_0] = \mathbb{P}_\tau\left[\frac{|\hat{\tau}|}{\sqrt{\mathsf{V}/n}} \geq z_{1-\alpha/2}\right]$$

Approximating the power function

- We can use the CLT to approximate the power function $\beta(\tau)$:

$$\beta(\tau) \approx 1 - \Phi\left(\frac{\sqrt{n}\tau}{\sqrt{\mathsf{V}}} + z_{1-\alpha/2}\right) + \Phi\left(\frac{\sqrt{n}\tau}{\sqrt{\mathsf{V}}} - z_{1-\alpha/2}\right)$$

- Power increases when:
  - Sample size n is large
  - Variance of the estimator V is small

Power and sample size calculation

$$\beta(\tau) \approx 1 - \Phi\left(\frac{\sqrt{n}\tau}{\sqrt{\mathsf{V}}} + z_{1-\alpha/2}\right) + \Phi\left(\frac{\sqrt{n}\tau}{\sqrt{\mathsf{V}}} - z_{1-\alpha/2}\right)$$

- Suppose we know or can estimate V
- Power calculation:
  - Set τ, n and V, find β
- Sample size calculation:
  - Set τ, β and V, find n

Relative sample size

- How do we divide treated and controls given a total sample size n?
- Asymptotic variance of $\hat{\tau}$:

$$\frac{\mathsf{V}}{n} = \frac{1}{n}\left(\frac{\sigma_1^2}{p} + \frac{\sigma_0^2}{1-p}\right)$$

- Minimizing the variance with respect to $p = \mathbb{P}[D_i = 1]$,

$$p^* = \frac{\sigma_1}{\sigma_1 + \sigma_0}$$

  - Assign more people to group with higher s.d.
  - If $\sigma_1 = \sigma_0$, $p^* = 1/2$

Power calculations in RD designs

- Same general idea: use CLT to approximate power function
- But need to incorporate robust bias-corrected inference
- Robust bias-corrected inference:

$$\frac{\hat{\tau}^{\mathsf{bc}} - \tau_{\mathsf{RD}}}{\sqrt{\widehat{\mathsf{V}}^{\mathsf{bc}}}} \to_{\mathcal{D}} \mathcal{N}(0,1), \quad \hat{\tau}^{\mathsf{bc}} = \hat{\tau} - \hat{\mathsf{B}}$$

  - $\hat{\tau}^{\mathsf{bc}}$ is the bias-corrected estimator
  - $\widehat{\mathsf{V}}^{\mathsf{bc}}$ is the robust variance estimator

Power calculations in RD designs

- Power function:

$$\beta^{\mathsf{bc}}(\tau) = 1 - \Phi\left(\frac{\tau}{\sqrt{\widehat{\mathsf{V}}^{\mathsf{bc}}}} + z_{1-\alpha/2}\right) + \Phi\left(\frac{\tau}{\sqrt{\widehat{\mathsf{V}}^{\mathsf{bc}}}} - z_{1-\alpha/2}\right)$$

  where

$$\widehat{\mathsf{V}}^{\mathsf{bc}} = \frac{\widehat{\mathsf{V}}^{\mathsf{bc}}_{+}}{nh_{+}} + \frac{\widehat{\mathsf{V}}^{\mathsf{bc}}_{-}}{nh_{-}}$$

- Power depends on:
  - Effective sample size: n and h_{+}, h_{-}
  - Size of the effect τ
  - Outcomes variances, kernel, order of polynomial, etc: $\widehat{\mathsf{V}}^{\mathsf{bc}}_{+}$, $\widehat{\mathsf{V}}^{\mathsf{bc}}_{1}$

Sampling in RD designs

- $\beta^{\mathsf{bc}}(\tau)$ can be used to calculate required sample size
- Relative sample sizes:

$$p^* = \frac{\sqrt{\widehat{\mathsf{V}}_+^{\mathsf{bc}}}}{\sqrt{\widehat{\mathsf{V}}_+^{\mathsf{bc}}} + \sqrt{\widehat{\mathsf{V}}_-^{\mathsf{bc}}}}$$

- How to sample:
  - Order obs according to distance to cutoff
  - Begin with closest obs, continue in order of distance tu cutoff