Statistical Power for Indirect Effects
in Clustered Regression Discontinuity Designs

Authors and Affiliations:
Ben Kelcey
University of Cincinnati
ben.kelcey@gmail.com

Kyle Cox
University of Cincinnati
coxk5@uc.edu

University of Cincinnati
Purpose

We develop closed-form expressions to estimate the statistical power to detect indirect/mediation effects in clustered regression discontinuity designs. Mapping the sensitivity of such designs is of critical importance because it directly governs the types of evidence researchers can bring to bear on theories of action under different designs and sample sizes. The results provide a set of formulas and software tools (implemented in free software) intended to inform and guide researchers in planning mediation studies with multilevel regression discontinuity designs. Our analyses consider prospective design-based assessments of indirect effects including the asymptotic Sobel test, component-wise joint test, and the resampling-based Monte Carlo interval test. These tests represent classical and modern approaches, capture important differences among tests in terms type one error rates and power levels, and are collectively representative of the range of tests most commonly used in the literature.

Background

In regression discontinuity (RD) designs, units are assigned to treatment conditions on the basis of a continuous scoring rule applied to an observed variable. When correctly implemented and specified, such designs facilitate unbiased inferences concerning the local area effects of a treatment. In education, the RD design often takes on a clustered or hierarchical form to accommodate the multilevel organizational structure of schooling and the school-wide nature of many interventions. For example, researchers may assign to treatment only those schools with an average achievement level below a certain cut point. An important design consideration in RD studies is the power with which mediation effects can be detected if they exist. However, literature is sparse regarding the sensitivity of such designs concerning indirect effects and has not established power formulas governing the detection of indirect effects in clustered RD designs.

Model

For brevity, we outline the development and application of the power analyses using, as an example, the Sobel test for a two-level clustered regression discontinuity design involving students nested within schools, a school-level cut score variable, and a school-level mediator (i.e., 2-2-1 mediation). We note that our complete results accommodate other tests and designs, ground identification and assumptions in the potential outcomes causal framework, and consider more complex model specifications.

Let us consider a clustered RD design where schools are assigned to a dichotomous treatment (T) on the basis of a school-level cut score (S). Our focal estimand is the indirect impact of the treatment on a continuous outcome (Y) through a continuous school-level mediator (M) conditional upon the cut score, individual-level covariates (X), their cluster-level aggregates (X'), and cluster-level covariates (W). The corresponding path model is

\begin{align*}
\text{Mediator model (schools)} & \quad M_j = \pi_j + aT_j + \lambda S_j + \pi_2 W_j + \pi_3 X_j + \epsilon_j^M \\
\text{Outcome model (students)} & \quad Y_{ij} = \beta_{0j} + \beta_1 (X_{ij} - \bar{X}_j) + \epsilon_{ij}^Y \\
\text{(schools)} & \quad \beta_{0j} = \gamma_0 + bM_j + \Delta S_j + c' T_j + \gamma_0 W_j + \gamma_2 X_j + u_{0j} \\
& \quad u_{0j} \sim N(0, \sigma_{\epsilon_j^Y}^2)
\end{align*}

In the mediation equation, we use \( M_j \) as the mediator for school \( j \), \( S_j \) as the regression discontinuity cut score variable with coefficient \( \lambda \), \( W_j \) as school-level covariates, \( \bar{X}_j \) as school-
level aggregates of individual-level covariates with $\pi$ as the respective path coefficients, $T_j$ as the treatment assignment coded as $\pm 1/2$ with associated coefficient $a$, and $\varepsilon^M_j$ as the error term. In the outcome equations, we use a multilevel model such that $Y_{ij}$ is the outcome for individual $i$ in school $j$, $X_{ij}$ are the individual-level covariates with coefficients $\beta$, and $\varepsilon^Y_{ij}$ is the level one error term. At the school-level, we introduce $\gamma$ as the respective path coefficients for covariates, $b$ as the conditional relationship between the mediator and the outcome, $A$ as the cut score coefficient, $c^*$ as the direct effect of the treatment, and $u_{0j}$ as the group-level random effects.

One of the most common tests of statistical significance for indirect effects is the Sobel test that compares the ratio of the product of the $a$ and $b$ coefficients to the standard error ($\sigma_{ab}$) of this product

$$z^{Sobel} = \frac{ab}{\sqrt{\sigma_{ab}^2}}$$

Our omitted derivations demonstrate that we can extend this test to the aforementioned clustered RD design such that the Sobel test for indirect effects under a clustered RD design is

$$z^{Sobel}_{CRD} = \frac{ab}{\sqrt{\sigma_{ab}^2}} = \frac{ab}{\sqrt{a^2(\rho(1-R^2_{M1}) + (1-R^2_{Y1})(1-\rho) / n_1) + b^2(1-R^2_{Y1}) / n_P(1-P)(1-\rho_{TS})}}$$

Here $\rho$ is the unconditional school outcome variance, $R^2_{Y1}$ and $R^2_{Y11}$ are the outcome variance explained by the school- and student-level covariates, $R^2_M$ is the mediator variance explained by the other predictors in the models, $\sigma^2_{M}$ is the unconditional variance of the mediator, $P$ is the proportion of schools receiving treatment, $\rho_{TS}$ is the correlation between the treatment assignment and the cut score variable, $n_1$ and $n_2$ represent the number of students per school and the number of schools, and $\sigma^2_{M}$ is the unconditional variance of the mediator.

Assuming the alternative hypothesis is true, the power of a two-sided test to detect the indirect multilevel mediation effect can be approximated with

$$P(z^{Sobel}_{CRD} > z_{critical}) = 1 - \Phi(z_{critical} - z^{Sobel}_{CRD}) + \Phi(-z_{critical} - z^{Sobel}_{CRD})$$

where $\Phi$ is the normal distribution with $z_{critical}$ as the chosen critical value (e.g., 1.96) corresponding to a nominal type one error rate.

Illustration

Consider an example in which schools with an average prior achievement level less than the 50th percentile are assigned to a whole school reform program using a clustered RD design. Assume that the theory of action is that the program acts on student achievement by improving on school climate (mediator). Let us assume the following parameter values

- $a=0.50$ (treatment-mediator relationship [Cohen’s d scale])
- $b=0.30$ (mediator-outcome relationship [Standardized regression scale])
- $\rho=0.10$ (unconditional intraclass correlation coefficient)
- $R^2_{Y1}=0.50$ (school-level outcome variance explained)
- $R^2_{Y11}=0.50$ (student-level outcome variance explained)
- $R^2_{M}=0.50$ (mediator variance explained)
$P=0.50$ (proportion of schools receiving treatment)

$\rho^2_{ts} = 0.50$ (squared treatment-cut score correlation)

$n_i = 50$ (students/school)

What is a reasonable number of schools to sample and how does this compare to the corresponding randomized design? The resulting power curves as functions of school sample size are described in Figure 1. Based on our derivations, we would expect that sampling approximately 136 schools would supply a 0.80 chance of detecting the indirect effect under the RD design whereas a randomized design would require only about 72 schools to produce a similar level of power. Figure 2 further tracks the relative efficiency of the RD and randomized designs under the assumed parameter values by considering the ratio of their error variances as a function of the squared correlation between the treatment and cut score. Evident from Figure 2, the RD design is equally efficient when there is no correlation between the treatment and cut score and becomes increasingly less efficient as this correlation increases.
Figure 1
Power to detect indirect effects under clustered regression discontinuity design (black) and cluster randomized design (red)
Figure 2
Relative efficiency between regression discontinuity and randomized design