Regression Discontinuity: A Causal Modeling Approach to Non-randomized Interventions

Ryan Johnson, Ph.D.
Principal Research Analyst
University of California, Riverside
Agenda

- Useful references for the Regression Discontinuity (RD) method
- Conceptual fundamentals of RD
- Comparison of RD vs. other common quasi-experimental methods
- Description of steps involved in RD
- 10-15-minute break
- Guided data analysis activity

After the conference, I will e-mail you the slides, Stata scripts, log files, and graphs
• RD approach first proposed by Thistlethwaite & Campbell (1960)
• RD only became notably popular the last two decades
• For a detailed academic review of RD and list of published economic studies using the method, see Lee and Lemieux (2010)
• For a more practical review of RD, see William Trochim’s demonstration at: http://www.socialresearchmethods.net/kb/quasird.php
Useful References


Random Assignment Method

- The gold standard for internal validity
- When using a method that is inherently random (e.g., coin flip) to assign members of a sample into groups, those groups will be equal on all possible characteristics
- If a treatment is applied to one group, any post-treatment differences must have been caused by the treatment or something related to that treatment
Non-randomized Methods

• Necessary because randomization is often not ethical or even desirable in higher education research
• Goal
  • Without random assignment, determine if a manipulation caused a change in some outcome
• Example non-randomized methods
  • RD
  • Multiple Regression/Matching on Observables
  • Instrumental Variables
• Of the methods above, only RD is comparable in internal validity to true random assignment
Underlying Premise of RD

• Random assignment still works regardless of the degree of sample heterogeneity, assuming sufficient sample size
• Those close to each side of an arbitrary cutoff are functionally randomly assigned to their side
• Body weight example
  • 150.0000000 lbs vs. 150.0000001 lbs
  • Imprecise measurement
    • Scales are not perfect (e.g., +/- 1 lb)
  • Individual variation
    • Countless random variables affect true scores (e.g., how recently last meal was eaten)
Random Assignment Is A Form Of RD

![Graph showing overall college GPA vs. randomly assigned number](image-url)
Example Data Appropriate For RD

Overall College GPA

Writing Placement Score (Assignment Variable)
Example Data Appropriate For RD

Overall College GPA vs. Writing Placement Score (Assignment Variable)
Random Assignment Is A Form Of RD
Random Assignment Is A Form Of RD

Overall College GPA

Randomly Assigned Number
Example Data Appropriate For RD

Overall College GPA vs. Writing Placement Score (Assignment Variable)

- Scatter plot showing the relationship between Overall College GPA and Writing Placement Score.
- A red line represents the trend, while a green line shows a different relationship.
- The plot includes a vertical line at the median placement score.
Interpretation of RD Effect at Cutoff

• A model that compares two regression lines at the cutoff does not compare two groups of actual people (or units) close to the line on either side.…

• It compares the trends of the entire treatment and control groups for predicted people who are literally equal on the assignment variable and only ended up as treated or control due to randomness

• This is an important distinction, because it highlights why large samples directly on top of the cutoff are not necessary to draw valid conclusions
Interpretation of RD Effect at Cutoff

- The effect at the cutoff is a weighted average effect for the entire sample
  - The average is weighted on the probability that each individual is close to the cutoff
  - The more random the assignment variable, the more generalizable the effect at the cutoff
- Although RD has internal validity equivalent to random assignment, the weighted average effect issue gives RD potentially lower external validity
Random Assignment Is A Form Of RD
Example Data Appropriate For RD

Overall College GPA vs. Writing Placement Score (Assignment Variable)
Necessary Conditions for Basic RD

- Treatment group must be assigned based on a specific cutoff in a previously measured continuous variable (the “assignment variable”)
  - No exceptions for those above or below the cutoff to switch groups after the assignment variable is measured
  - It must be impossible for the people in the study to precisely manipulate the assignment variable in an attempt to gain (or avoid) treatment
- The cutoff for the assignment variable must be “arbitrary” (i.e., no natural gaps at the cutoff)
Necessary Conditions for Basic RD

- Data distribution must be able to be described as a polynomial function.
- Intervention must be uniformly delivered.
- Model must be correctly specified.
Example Data Appropriate For RD

- Overall College GPA
- Writing Placement Score (Assignment Variable)

Scatter plot showing a relationship between Overall College GPA and Writing Placement Score. The data points are distributed across the plot, with a trend line indicating a positive correlation.
RD vs. Multiple Regression/Matching (MR/M)

- Goal with MR/M is to control for differences in characteristics between those in treated and control groups.
- MR/M assumes the unlikely premise that all causal covariates are accounted for in the model.
- RD requires no assumptions regarding covariates. In fact, inclusion of covariates does not even improve RD estimates, although they can be used to test certain assumptions about RD.
• The IV method assumes that the chosen IV is uncorrelated with unobserved variables
• Unobserved variables are irrelevant to well-designed RD studies
Steps in RD Design and Analysis

• Design
  • Ensure all assumptions are met. Proper study design is more important than which analytic method you use

• Analysis
  • Center the assignment variable at the cutoff
  • Graph the relationship
    • Graph all points to get a sense of total distribution/outliers
    • Use a binned approach if there is too much noise to see patterns clearly
  • Specify higher-order terms and interactions
  • Estimate initial full model
  • Refine the model to include only essential parameters
Optional Additional Steps

- If there is concern about manipulation of the assignment variable:
  - Look for irregular frequency distributions above and below the cutoff
  - Conduct an RD analysis on baseline covariates
  - Include baseline covariates in the original RD analysis
• Scenario 1 based on real data collected at LMU
• Scenario 2 based on data manipulated by me for a particular purpose
• Loyola Law School (LLS) wanted to help struggling students perform better in their second year
• Those with first-semester GPAs below 2.75 were required to take an additional course in Spring designed to build skills
• All students with low GPAs were required to take the course, and no students with higher GPAs were allowed
• The outcome was GPA in Fall of their second year
If you still need the data…

• Stata data: [http://goo.gl/qTM0IR](http://goo.gl/qTM0IR)
• CSV data: [http://goo.gl/0InuGJ](http://goo.gl/0InuGJ)
For questions, e-mail ryan.johnson@ucr.edu