

Metodos de Evaluacion de Impacto en R

Clase 4: Regression Discontinuity Design

Dr. Sergio Bejar Lopez

CIDE - Escuela de Metodos

29 de Enero, 2026

Agenda de Hoy

- ① Motivacion: Becas Benito Juarez
- ② Intuicion de RDD
- ③ El Grafico Clave
- ④ Supuesto de Continuidad
- ⑤ Estimacion en R
- ⑥ Validacion del Diseno
- ⑦ Limitaciones
- ⑧ Ejercicio Practico

Objetivo: Entender cuando y como usar RDD

Becas Benito Juarez en Mexico

Contexto del programa:

- Apoyo economico para estudiantes de preparatoria
- 1,600 pesos mensuales
- Objetivo: reducir desercion escolar

Criterio de asignacion:

- Basado en **puntaje de examen de admision**
- Cutoff: 700 puntos (de 1000)
- Si puntaje $\geq 700 \rightarrow$ BECA
- Si puntaje $< 700 \rightarrow$ NO beca

Pregunta de investigacion

¿La beca aumenta la tasa de graduacion?

El Problema de Evaluacion

NO podemos comparar todos los que reciben vs no reciben:

Graduacion	
Con beca (puntaje ≥ 700)	75%
Sin beca (puntaje < 700)	58%
Diferencia	17 pp

Problema

Los estudiantes con puntaje alto son **diferentes**:

- Mas preparados academicamente
- Mas motivados
- Mejores condiciones familiares

La diferencia NO es solo por la beca!

La Solucion: RDD

Idea central de RDD:

Regression Discontinuity Design

Comparar estudiantes **justo arriba** y **justo debajo** del cutoff

¿Por que funciona?

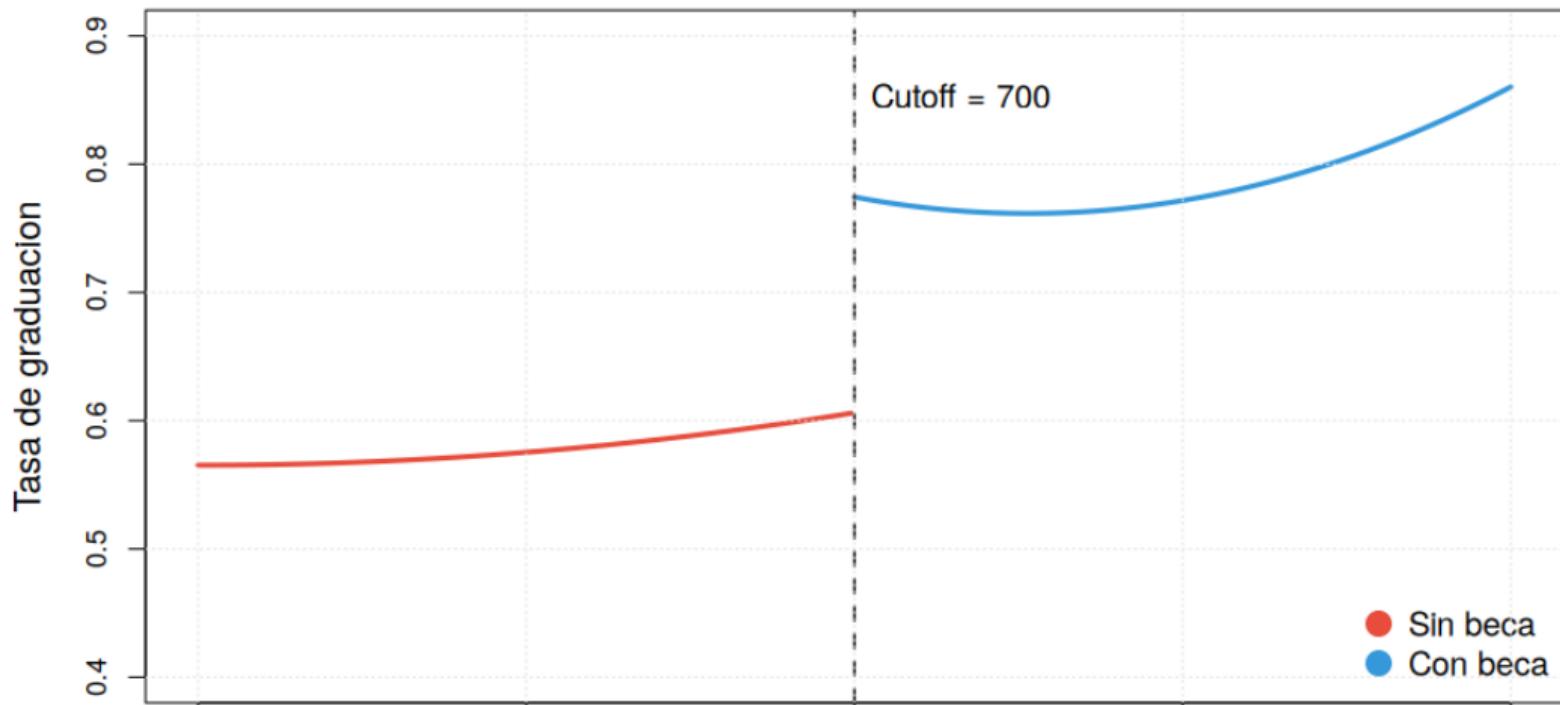
- Quien saca 699 vs 700 es prácticamente **identico**
- La única diferencia: uno recibe beca, otro no
- Es como un experimento aleatorio **en el cutoff**

Terminología:

- **Running variable:** puntaje de examen
- **Cutoff:** 700 puntos
- **Tratamiento:** recibir beca
- **Outcome:** tasa de graduacion

EL GRAFICO CLAVE DE RDD

RDD: Efecto de la Beca Benito Juarez



Leyendo el Grafico RDD

Paso a paso:

1. Relacion general

- Mayor puntaje → mayor graduacion (esperado)
- Relacion continua y suave

2. El cutoff (700 puntos)

- Linea punteada vertical
- Izquierda = sin beca (rojo)
- Derecha = con beca (azul)

3. La discontinuidad

- Hay un **SALTO** en 700
- Graduacion salta de 60% a 75%
- Ese salto = **efecto causal de la beca**

Logica: Justo antes y justo despues del cutoff, los estudiantes son casi identicos. ▶ El salto



Sharp vs Fuzzy RDD

Nuestro caso: Sharp RDD

Sharp RDD

Si puntaje $\geq 700 \rightarrow 100\% \text{ recibe beca}$

Si puntaje $< 700 \rightarrow 0\% \text{ recibe beca}$

Regla se cumple **perfectamente**

Fuzzy RDD (no cubierto hoy):

- Regla no se cumple 100%
- Algunos con puntaje < 700 reciben beca (excepciones)
- Algunos con puntaje ≥ 700 no la toman
- Requiere estimacion con variables instrumentales

Hoy nos enfocamos en Sharp RDD

El Supuesto Clave: Continuidad

Supuesto de Continuidad

En **ausencia del tratamiento**, el outcome seria una funcion continua (suave) de la running variable

En palabras:

- Si NO existiera la beca
- La graduacion cambiaria suavemente con el puntaje
- NO habria saltos en 700 puntos

Consecuencia:

- Cualquier discontinuidad en el cutoff
- DEBE ser causada por el tratamiento (beca)
- No por otros factores

¿Como verificamos? Ver otras variables (covariables, placebo outcomes)



¿Cuando se Viola Continuidad?

Amenazas al supuesto:

1. Otras politicas en el mismo cutoff

- Si hay otro programa que tambien empieza en 700
- No sabemos cual causa el salto

2. Manipulacion de la running variable

- Estudiantes manipulan su puntaje para pasar 700
- Los justo arriba son diferentes a los justo abajo
- Ejemplo: trampa en examen, correccion favorable

3. Auto-seleccion alrededor del cutoff

- Solo cierto tipo de estudiante se presenta si espera sacar cerca de 700

En nuestro caso: Puntaje de examen es dificil de manipular, buen diseño



Metodo 1: Comparacion local (ventana)

```
# Definir bandwidth (ventana alrededor del cutoff)
bandwidth <- 50 # +/- 50 puntos

# Muestra local
muestra_local <- datos[abs(datos$puntaje_centrado) <= bandwidth, ]

# Diferencia simple
rdd_simple <- mean(muestra_local$tasa_graduacion[muestra_local$recibe_beca]) -
               mean(muestra_local$tasa_graduacion[!muestra_local$recibe_beca])

cat("Efecto RDD:", rdd_simple)
# Resultado: 0.172 (17.2 puntos porcentuales)
```

Interpretacion: La beca aumenta graduacion en 17.2 pp

RDD en R: Regresion

Metodo 2: Regresion lineal local

```
# Modelo RDD con interaccion  
1 modelo_rdd <- lm(tasa_graduacion ~ puntaje_centrado * recibe_beca,  
2                     data = muestra_local)  
3  
4 summary(modelo_rdd)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.6008	0.0103	58.46	<2e-16 ***
puntaje_centrado	0.0003	0.0004	0.73	0.468
recibe_becaTRUE	0.1691	0.0140	12.05	<2e-16 ***
puntaje_centrado:recibe_becaTRUE	-0.0004	0.0005	-0.79	0.428

Coeficiente clave: `recibe_becaTRUE = 0.169 (16.9 pp)`

Interpretando el Modelo RDD

Modelo:

$$Y_i = \beta_0 + \beta_1(X_i - c) + \beta_2 T_i + \beta_3(X_i - c) \times T_i + \epsilon_i$$

Coeficiente	Valor	Significado
β_0	0.601	Graduacion en el cutoff (sin beca)
β_1	0.0003	Pendiente a la izquierda
β_2	0.169	Efecto RDD
β_3	-0.0004	Cambio en pendiente

Resultado

Recibir la beca incrementa la graduacion en **16.9 puntos porcentuales** ($p < 0.001$)

Bandwidth: ¿Que Ventana Usar?

Trade-off del bandwidth:

Bandwidth pequeño (ej: ± 20):

- **Pro:** Mas comparables (mas cerca del cutoff)
- **Contra:** Menos datos, mas varianza

Bandwidth grande (ej: ± 100):

- **Pro:** Mas datos, menos varianza
- **Contra:** Menos comparables, mas sesgo

Solucion: Bandwidth optimo (paquete rdrobust)

- Minimiza MSE (bias + varianza)
- Calculado automaticamente
- Tipicamente entre 30-80 para nuestro caso

Limitaciones de RDD

Ventajas:

- No requiere aleatorización
- Diseño creíble y transparente
- Fácil de visualizar y comunicar
- No necesita supuestos sobre observables

Limitaciones IMPORTANTES:

- ① **Efecto local:** Solo en el cutoff, no generalizable
- ② **Requiere cutoff arbitrario:** Si cutoff es endógeno, no funciona
- ③ **Poder estadístico:** Necesita muchos datos cerca del cutoff
- ④ **Forma funcional:** Sensible a especificación (lineal, cuadrática)
- ⑤ **Manipulación:** Si hay, invalida el diseño

¿Cuando Usar RDD?

RDD es apropiado cuando:

- Hay un cutoff claro y **arbitrario**
- Asignacion se basa estrictamente en running variable
- Running variable es dificil de manipular
- Hay suficientes observaciones cerca del cutoff
- No hay otras politicas en el mismo cutoff

Ejemplos buenos:

- Becas por puntaje de examen
- Programas por umbral de pobreza
- Regulaciones por tamaño de empresa
- Edad legal para votar/beber/conducir

NO usar RDD si:

- Cutoff es endogeno (decidido por actores)

Ejercicio Practico

Tu turno: 20 minutos

Tarea: Estimar efecto en **promedio final** (no graduacion)

Pasos:

- ① Crear grafico RDD para promedio final
- ② Calcular efecto con comparacion local
- ③ Estimar modelo de regresion
- ④ Interpretar resultados
- ⑤ **BONUS:** Probar diferentes bandwidths

Dataset: datos_rdd_becas.csv

Variable outcome: promedio_final (escala 0-10)

Hipotesis: La beca deberia aumentar el promedio final

Resumen de lo Aprendido

Conceptos clave:

- ① RDD compara justo arriba vs justo debajo del cutoff
- ② Discontinuidad = efecto causal
- ③ Supuesto: continuidad sin tratamiento
- ④ Validacion: placebo tests, density test

Habilidades tecnicas:

- Crear graficos RDD
- Estimar efecto con regresion
- Elegir bandwidth
- Validar supuestos

Mensaje final:

- RDD es muy creible cuando diseño es bueno
- Pero solo identifica efecto LOCAL (en el cutoff)

Comparacion de Metodos

	Matching	DiD	RDD
Datos	Cross-section	Panel	Cross-section
Supuesto	CIA	Tendencias paralelas	Continuidad
Identificacion	Observables	Fijos + tendencias	Cutoff arbitrario
Efecto	ATT	ATT	LATE (local)
Validez externa	Media	Media	Baja

¿Cual usar? Depende del diseño y datos disponibles

Gracias

Clase 4: Regression Discontinuity Design

Dr. Sergio Bejar Lopez
Escuela de Metodos - CIDE

sergio.bejar@cide.edu