Clase 1: Regression Discontinuity Design

Tomás Rau Binder
QLAB

Julio 2023

Contenidos

Presentación

- Tomás Rau, MA in Statistics y PhD in Economics, University of California Berkeley
- Profesor Titular y Director del Instituto de Economía, PUC-Chile
- Research Fellow, IZA Institute of Labor Economics
- Co-Editor *Economia*, journal de LACEA
- Investigación: Econometría, Economía Laboral, Desarrollo Económico
- Docencia: Econometría (paramétrica y no-paramétrica), Evaluación de Impacto, Economía Laboral
- Email: trau@uc.cl
- Página: https://sites.google.com/site/tomasraubinder

Presentación

- Algunas publicaciones:
  - **Peer Effects in the Adoption of a Youth Employment Subsidy** (with Claudio Mora-García). *The Review of Economics and Statistics.*, Vol. 105(3), pp. 614-625, May 2023.
  - **The Effects of Equal Pay Laws on Firm Wage Premiums: Evidence from Chile** (with Gabriel Cruz). Labour Economics. Vol. 75, April 2022.
  - **The Children of the Missed Pill** (with Miguel Sarzosa and Sergio Urzúa). *Journal of Health Economics, Vol. 79, September, 2021.*
  - **The Effects of a Maternity Leave Reform on Children's Abilities and Maternal Outcomes in Chile**, (con Pinjas Albagli). *The Economic Journal*, Vol. 129(619), pp. 1015-1047, April, 2019.
  - **A Matching Estimator Based on a Bilevel Optimization Problem**, (con Juan Díaz y Jorge Rivera). *The Review of Economics and Statistics*, Vol. 97(4), pp. 803-812, 2015.

Introducción

En econometría existen 5 grandes tipos de "estrategias de identificación" para la inferencia causal:

- MCO/Matching/Efectos fijos
- Diferencias en diferencias
- Variables Instrumentales
- Regression Discontinuity
- Evaluación experimental (RCT)

En estas dos clases nos centraremos en *Regression Discontinuity*.

Introducción

- Este es un método muy usado en evaluación de programas cuando la asignación al tratamiento es una función discontinua de una variable observable.
- Ejemplos de este tipo de asignaciones son becas con puntaje de corte (como universidades que becan un 50 % si el puntaje es sobre 700, pero 0 % si es 699 o menos).
- Asignaciones de programas sociales como acceso a seguro médico en Colombia: puntaje del Sisbén menor a 47 obtiene beneficio. Chile Solidario: puntaje de Ficha de Protección Social menor que 11.734 obtiene el beneficio
- El paramétro de interés es la discontinuidad <u>en sí</u> en la variable de resultado (outcome) generado por la asignación del programa.
- Este método tiene buena validez interna. De acuerdo a Lee, puede ser tan bueno como un experimento, pero local.

Regression Discontinuity Design

- Formalmente, observamos para cada individuo i: D_i si fue tratado o no, el outcome Y_i y la variable (continua) de asignación X_i que determina si se recibe tratamiento o no en función del umbral x_0.
- Así, las personas **justo debajo** del umbral x_0 debieran ser muy parecidas a las personas **justo encima** de x_0 excepto porque unas reciben tratamiento mientras que las otras no.
- Por tanto, los últimos podrían ser un **buen contrafactual** de los primeros de modo que la comparación de ambos grupos nos entrega un estimador válido del impacto del programa para quienes están cerca de x_0.

Regression Discontinuity

Luego, siguiendo a Hahn, Todd y Van der Klaauw (2001), sea Y_{1i} el outcome potencial si es tratado y Y_{0i} el outcome potencial si no es tratado. Además, sea $D_i = 1$ si el individuo es tratado y $D_i = 0$ si no lo es.

Luego,

$$Y_i = Y_{1i}D_i + Y_{0i}(1 - D_i)$$

Sea $\alpha_i = Y_{1i} - Y_{0i}$ el efecto del tratamiento, entonces podemos escribir (esta notación es usada de manera usual):

$$Y_i = Y_{0i} + \alpha_i D_i$$

Regression Discontinuity

Ahora, existen dos tipos de diseños de *Regression Discontinuity*.

El primero es el *Sharp Design* en el cual $D_i = f(x_i)$ es una función determinística de x_i (variable de asignación) y continua excepto en x_0. Por ejemplo,

$$D_i = \begin{cases} 1 & \text{si } X \leq x_0 \\ 0 & \text{si } X > x_0 \end{cases}$$

Note que también podría escribirse de otra forma, $D_i = 1$ si $X > x_0$. Dependerá de la política a evaluar.

El otro tipo de diseño es el *Fuzzy Design* en el cual la probabilidad de tratamiento es discontinua en x_0, i.e. $Pr(D_i = 1|x)$ es discontinua en x_0. La discontinuidad es usada como IV.

Regression Discontinuity

Supuesto (RD):

(*i*) Los siguientes límites existen:

$$\lim_{x \to x_0^+} E(D_i | X_i = x)$$

$$\lim_{x \to x_0^-} E(D_i | X_i = x)$$

(*ii*) $x_0^+ \neq x_0^-$.

Regression Discontinuity

Supuesto (A1): $E(Y_{0i}|X_i = x)$ es continua en $X = x_0$.

Teorema 1*: Si $\alpha_i = \alpha$ es constante y se cumplen los supuestos RD y AI, entonces:*

$$\alpha = \frac{\lim_{X \to x_0^-} E(Y_i|X_i = x) - \lim_{X \to x_0^+} E(Y_i|X_i = x)}{\lim_{X \to x_0^-} E(D_i|X_i = x) - \lim_{X \to x_0^+} E(D_i|X_i = x)}$$

Recordando que $\alpha = Y_1 - Y_0$.

Regression Discontinuity

La demostración es trivial: tome las diferencias

$$\lim_{X \to x_0^+} E(Y_i|X_i = x) - \lim_{X \to x_0^-} E(Y_i|X_i = x) =$$

$$\underbrace{\lim_{X \to x_0^-} E(y_{0i}|X_i = x) - \lim_{X \to x_0^+} E(y_{0i}|X_i = x)}_{0}$$

$$+\alpha(\lim_{x \to x_0^-} E(D_i|X_i = x) - \lim_{X \to x_0^+} E(D_i|X_i = x))$$

$$= \alpha(\lim_{x \to x_0^-} E(D_i|X_i = x) - \lim_{X \to x_0^+} E(D_i|X_i = x))$$

QED.

Regression Discontinuity

El resultado anterior es independiente del tipo de diseño (Sharp o Fuzzy).

Note que cuando tenemos *Sharp Design*:

$$\lim_{x \to x_0^-} E(D_i|x_i = x) = 1$$

$$\lim_{x \to x_0^+} E(D_i|x_i = x) = 0$$

luego

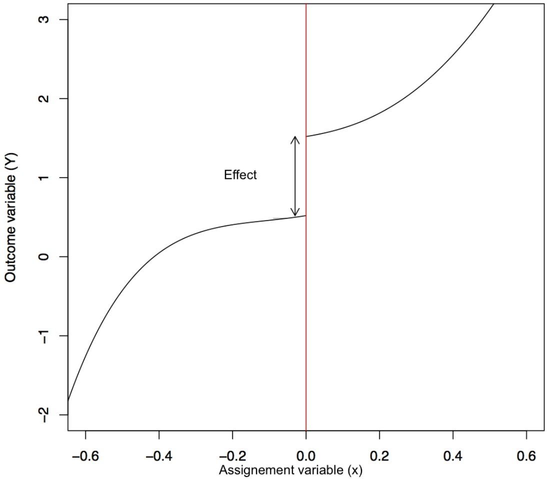
$$\alpha = \lim_{x \to x_0^-} E(y_i|x_i = x) - \lim_{x \to x_0^+} E(y_i|x_i = x)$$

Solo implica restar dos esperanzas para obtener el ATE en el corte x_0.

Regression Discontinuity

- Note que el resultado anterior nos permite encontrar un *efecto causal* explotando la discontinuidad en la asignación del tratamiento
- Esto es muy potente porque no necesitamos más controles que el x de la regla de asignación
- Un problema: estamos asumiendo $\alpha_i = \alpha$ (constant treatment effect) y eso puede ser muy restrictivo

Regression Discontinuity



Regression Discontinuity

Supuesto (A2): $E(\alpha_i | x_i = x)$ es continua en $x = x_0$.

Teorema 2*: Suponga que D_i es independiente a α_i condicional en x_i cerca de x_0 y se cumplen los supuestos RD, A1 y A2, entonces:*

$$E[\alpha_i | X_i = x_0] = \frac{\text{lím}_{X \to x_0^-} E(Y_i | X_i = x) - \text{lím}_{X \to x_0^+} E(Y_i | X_i = x)}{\text{lím}_{X \to x_0^-} E(D_i | X_i = x) - \text{lím}_{X \to x_0^+} E(D_i | X_i = x)}$$

Recordando que $\alpha_i = Y_{1i} - Y_{0i}$.

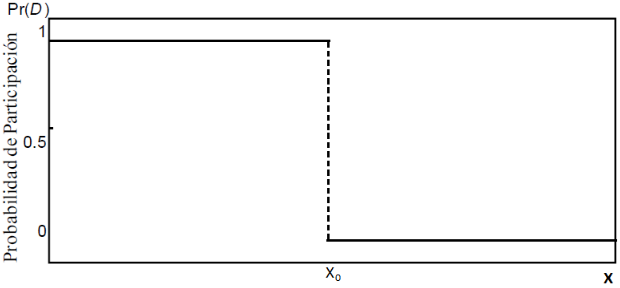
Regression Discontinuity

- Note que sólo podemos identificar la esperanza condicional de α_i en $x = x_0$. Esto limita bastante el análisis, pero sigue siendo un efecto causal de interés: ATE at the cutoff (Sharp) o LATE (en el Fuzzy)
- Si queremos alejarnos de la discontinuidad, debemos asumir "strong ignorability", Angrist y Rokkanen (2015).
- El supuesto de independencia condicional limita a que los individuos no se seleccionen en el programa (o eviten el programa).
- El supuesto anterior es clave, si no se cumple

Diseño “Sharp”, Agudo o Nítido

Sharp RDD: Identificación

Como se discutió, en este caso la probabilidad de recibir tratamiento es discontinua en x_0 y salta de 1 a 0:



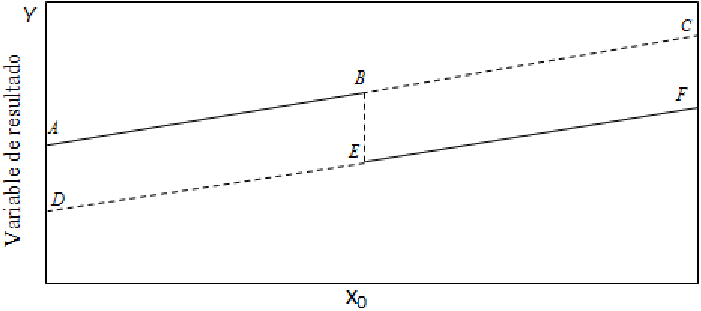
El primer supuesto (RD) para que exista el diseño *Sharp* implica que:

$$\lim_{x \uparrow x_0} Pr(D_i = 1 | X_i = x) \neq \lim_{x \downarrow x_0} Pr(D_i = 1 | X_i = x)$$

donde $X \uparrow x_0$ implica que x que se acerca a x_0 por la izquierda ($X \rightarrow x_0^-$). Se ve gráficamente que se cumple el supuesto RD.

Sharp RDD: Identificación

Veamos ahora el valor esperado del outcome o variable de resultado como función de X_i:



donde $AC = E(Y_1|X)$ y $DF = E(Y_0|X)$ pero solo se observan AB y EF. Por diseño, no existen individuos con $X_i \leq x_0$ para los que se observe Y_0 ya que todos son tratados. Sin embargo, podríamos usar a los no tratados con X_i muy cercano a x_0 como contrafactual si son "similares" a los participantes.

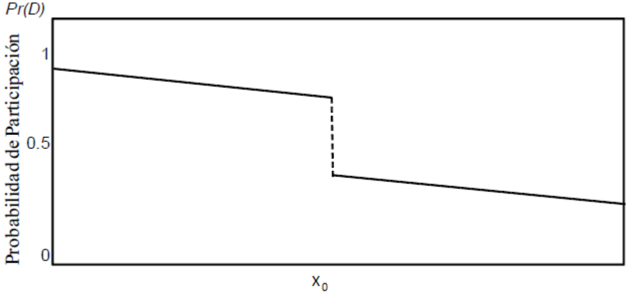
Sharp RDD: Identificación

- Esta es exactamente la idea de un Sharp RDD: comparar a los tratados próximos al umbral, con los controles próximos al umbral.
- Cerca del umbral, las personas serán observacionalmente equivalentes si es que no manipulan la variable de asignación (ejemplo 11724 vs 11744 en la FPS)
- La asignación discontinua, genera un pequeño "experimento local", donde la asignación es cómo si fuese aleatoria
- Esta es una de las razones por que se ha vuelto un método muy popular

Diseño "Fuzzy", Difuso o Borroso

Fuzzy RDD

En este caso es la probabilidad de participación cambia de forma discontínua en el umbral, pero **no de 1 a 0** como el caso sharp:



Esto debido a que la participación efectiva depende de X_i pero también de otras dimensiones observadas y no observadas

Fuzzy RDD

Luego las condiciones de identificación son similares:

- Condición 1: Discontinuidad en la Probabilidad de Participación:

$$\lim_{x \uparrow x_0} Pr(D_i = 1 | X_i = x) \neq \lim_{x \downarrow x_0} Pr(D_i = 1 | X_i = x)$$

- Condición 2: Continuidad de los valores esperados de los outcomes potenciales:

$$\lim_{x \uparrow x_0} E(Y_{1i} | X_i = x) = \lim_{x \downarrow x_0} E(Y_{1i} | X_i = x)$$

$$\lim_{x \uparrow x_0} E(Y_{0i} | X_i = x) = \lim_{x \downarrow x_0} E(Y_{0i} | X_i = x)$$

Fuzzy RDD

- Note que ahora no todos los que están a la izquierda del umbral reciben tratamiento ni tampoco todos los que están a la derecha del umbral son controles. Esto implica que la diferencia de promedios del resultado observado a cada lado del umbral no identifica el ATE (en x_0) como en el diseño Sharp
- Para identificar un efecto de tratamiento en el diseño Fuzzy debemos estimar:

$$LATE = \frac{\text{lím}_{x \uparrow x_0} E(Y_i | X_i = x) - \text{lím}_{x \downarrow x_0} E(Y_i | X_i = x)}{\text{lím}_{x \uparrow x_0} Pr(D_i = 1 | X_i = x) - \text{lím}_{x \downarrow x_0} Pr(D_i = 1 | X_i = x)}$$

  o sea que estamos ajustando la expresión del diseño Sharp por la diferencia en la fracción de individuos que efectivamente son tratados a cada lado del umbral.

Fuzzy RDD: Identificación

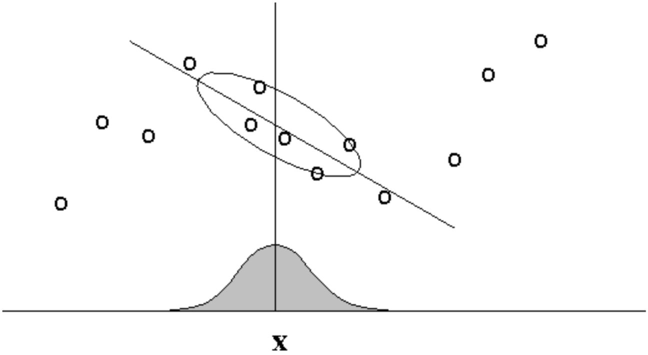
- Note que esta expresión coincide con el estimador de Wald (1940) de Variables Instrumentales
- Hahn, Todd y Van der Klaaw (2001) muestran que bajo el supuesto de monotonicidad local, parecido al supuesto utilizado en IV (Imbens y Angrist 1994), la expresión identifica un efecto doblemente local (LATE): mide el impacto promedio del programa para aquellos individuos cercanos al umbral y cuyo indicador de tratamiento cambiaría de tratado a control si su valor de X cruzara el umbral x_0 (compliers).

Estimación

- Para estimar el LATE debemos estimar los 4 elementos de la expresión anterior (solo 2 en el caso Sharp design).
- Se pueden utilizar métodos paramétricos como regresión lineal o polinomios globales y regresiones no lineales para el denominador.
- Sin embargo, el error de especificación de los métodos paramétricos puede traer consigo un sesgo muy importante en la estimación del impacto de un programa; por lo que son preferidos los métodos no paramétricos.
- El método no paramétrico más utilizado en RDD por sus propiedades asintóticas es la **Regresión Lineal Local** o **LLR**.
- Intuitivamente, una **LLR** estima regresiones lineales en vecindades pequeñas de un punto para aproximarse mejor a la verdadera función
- En Stata, el comando más utilizado es *rdrobust* que hace la estimación por **LLR**

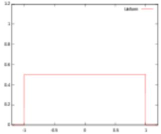
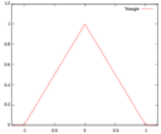
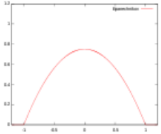
Estimación

Ejemplo de LLR en torno a X



La "campana" sombreada (kernel) indica que le pone más peso a las observaciones cercanas a X y menos peso a las lejanas.

Regression Discontinuity

Kernel Functions, *K*(*u*)		
Uniform	$K(u) = \frac{1}{2}\,\mathbf{1}_{\{\lvert u\rvert \le 1\}}$	
Triangular	$K(u) = (1 - \lvert u\rvert)\,\mathbf{1}_{\{\lvert u\rvert \le 1\}}$	
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2)\,\mathbf{1}_{\{\lvert u\rvert \le 1\}}$	

Regression Discontinuity

Estimación: Local Linear Regression, SHARP Design

Podemos estimar $\lim_{x \to x_0^+} E(Y_i | X_i = x)$

$$(\hat{a}, \hat{b}) = argmin_{a,b} \sum_{i=1}^{n} (Y_i - a - b(X_i - x_0))^2 \mathrm{K}\left(\frac{X_i - x_0}{h}\right) \mathrm{I}_{(X_i > x_0)}$$

Donde $\mathrm{K}(\cdot)$ es una función *kernel* que satisface $\int \mathrm{K}(u) du = 1$, $\mathrm{K}(-u) = \mathrm{K}(u)$. Y $\mathrm{I}_{(X_i > x_0)}$ es una función indicatriz.

Al estimador de a le llamaremos $\hat{a}_R$ en esta regresión.

Regression Discontinuity

y para estimar $\lim_{X \to x_0^-} E(Y_i|X_i = x)$

$$(\hat{a}, \hat{b}) = argmin_{a,b} \sum_{i=1}^{n} (Y_i - a - b(X_i - x_0))^2 \mathrm{K}\left(\frac{X_i - x_0}{h}\right) \mathrm{I}_{(X_i \leq x_0)}$$

Donde el estimador de a le llamaremos $\hat{a}_L$ en este caso.

Es fácil ver que el estimador RD es $E[\alpha_i|X_i = x_0] = \hat{a}_L - \hat{a}_R$ (dado que en el ejemplo el tratamiento se asigna a quienes $< x_0$)

Regression Discontinuity

El resultado anterior es nítido puesto que estamos estimando dos regresiones lineales locales (ponderadas) en torno a x_0:

$$
\begin{aligned}
Y_i &= a_R + b_R(X_i - x_0) + \epsilon_R \text{ si } X_i > x_0 \\
Y_i &= a_L + b_L(X_i - x_0) + \epsilon_L \text{ si } X_i \leq x_0
\end{aligned}
$$

Luego,

$$
\begin{aligned}
\lim_{X \to x_0^+} E(Y_i|X_i = x) &= a_R \\
\lim_{X \to x_0^-} E(Y_i|X_i = x) &= a_L
\end{aligned}
$$

Estimación

Para el Fuzzy design hacemos lo mismo, pero nos falta el denominador.

- $\lim_{x \uparrow x_0} \widehat{Pr}(D = 1|X = x) = \widehat{c}_L$, donde

$$(\hat{c}_L, \hat{d}_L) = argmin_{c_L, d_L} \sum_{i=1}^{n} (D_i - c_L - d_L(X_i - x_0))^2 \mathrm{K}\left(\frac{X_i - x_0}{h}\right) \mathrm{I}_{(X_i \leq x_0)}$$

- $\lim_{x \downarrow x_0} \widehat{Pr}(D = 1|X = x) = \widehat{c}_R$, donde

$$(\hat{c}_R, \hat{d}_R) = argmin_{c_R, d_R} \sum_{i=1}^{n} (D_i - c_R - d_R(X_i - x_0))^2 \mathrm{K}\left(\frac{X_i - x_0}{h}\right) \mathrm{I}_{(X_i > x_0)}$$

Estimación

Finalmente,

$$\begin{aligned}\widehat{LATE} &= \frac{\lim_{x\uparrow x_0} \widehat{E}(Y_i|X_i = x) - \lim_{x\downarrow x_0} \widehat{E}(Y_i|X_i = x)}{\lim_{x\uparrow x_0} \widehat{Pr}(D_i = 1|X_i = x) - \lim_{x\downarrow x_0} \widehat{Pr}(D_i = 1|X_i = x)} \\ &= \frac{\hat{a}_L - \hat{a}_R}{\hat{c}_L - \hat{c}_R}\end{aligned}$$

En STATA, el comando más utilizado es *rdrobust*, implementado por Calonico, Cattaneo y Titiunik (2014); no solo porque calcula el ancho de banda óptimo, sino que además propone un error estándar corregido por sesgo e Intervalos de Confianza robustos para el LATE

Veamos 2 ejemplos de aplicaciones para fijar ideas. La próxima clase veremos otros temas relevantes cómo por ejemplo testear los supuestos de identificación.

Ejemplo: Does Medicare Save Lives?

Diseño Sharp

Ejemplo 1 en Stata

- En Card et al. (2009) “Does Medicare Save Lives?” se quiere medir el efecto de Medicare sobre la tasa de mortalidad de los pacientes al ingresar al hospital (por E.R.)
- Las personas al cumplir 65 años pasan a estar cubiertas por Medicare, por lo que la asignación al tratamiento es una función discontinua de la edad.
- Los autores utilizan un método paramétrico:

  $$y_i = f(a_i, \alpha) + Post65_i \beta + \epsilon_i$$

  donde y_i es la mortalidad a los 7 días, a_i es la edad, $f(\cdot, \alpha)$ es una función polinómica con parámetros α y $Post65_i$ es la dummy de tratamiento.
- En esta especificación, β mide el ATE en t=65. Los autores reportan una disminución de 1 % en la tasa de mortalidad a los 7 días.

Ejemplo 1 en Stata

Intentemos replicar los resultados con datos similares (no iguales):

rdrobust d7 age, c(65) kernel(triangular)

Sharp RD estimates using local polynomial regression.

Cutoff c = 65	Left of **c**	Right of **c**
Number of obs	**60**	**60**
Eff. Number of obs	**21**	**21**
Order loc. poly. (p)	**1**	**1**
Order bias (q)	**2**	**2**
BW loc. poly. (h)	**1.716**	**1.716**
BW bias (b)	**2.490**	**2.490**
rho (h/b)	**0.689**	**0.689**

Number of obs	=	**120**
BW type	=	**mserd**
Kernel	=	**Triangular**
VCE method	=	**NN**

Outcome: days_death7_mean. Running variable: age.

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Conventional	**-.00751**	**.00352**	**-2.1344**	**0.033**	**-.014404**	**-.000614**
Robust	**-**	**-**	**-1.6761**	**0.094**	**-.015425**	**.001204**

mserd: One common MSE-optimal bandwidth

Análisis Gráfico

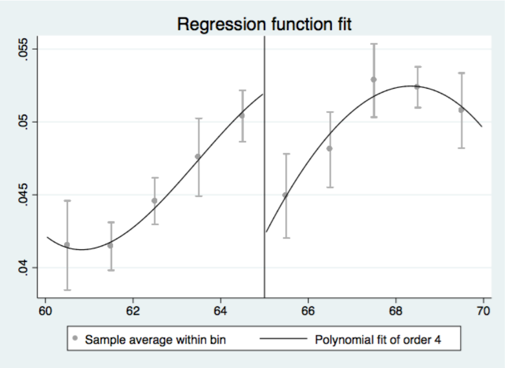
- Uno de los atractivos del diseño RD es su capacidad de mostrar gráficamente el efecto tratamiento
- Si graficamos $E(Y|X)$ respecto a X, tomando en cuenta el umbral x_0 que determina la asignación al tratamiento, entonces tendremos una representación bastante intuitiva del efecto a través del salto que se da en el umbral
- Si la estimación utiliza métodos parámetricos, como regresión lineal o regresión polinomial global, entonces graficamos la predicción del outcome $\widehat{Y}$ vs la variable de asignación X.
- Es recomendable incluir intervalos de confianza para tener una mejor idea de la significancia del efecto estimado.

rdplot en Stata

- Calonico, Cattaneo y Titiunik (2015) proponen un análisis gráfico del diseño RD bajo distintas especificaciones, según el objetivo del usuario
- El gráfico más común es áquel que distribuye los datos en bins del mismo tamaño (*evenly spaced binning of the data*) y los autores proponen formas automáticas de escoger la cantidad de bins
- Sugieren también un gráfico con bins por cuantiles (*quantile spaced binning*)
- En Stata el comando que implementa estos gráficos es *rdplot*

rdplot en Stata

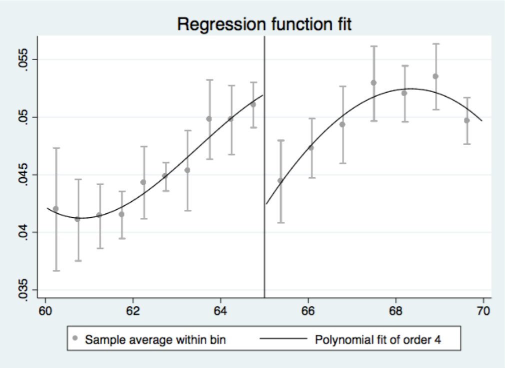
rdplot d7 age, c(65) binselect(es) ci(95)



es: IMSE-optimal evenly-spaced method

rdplot en Stata

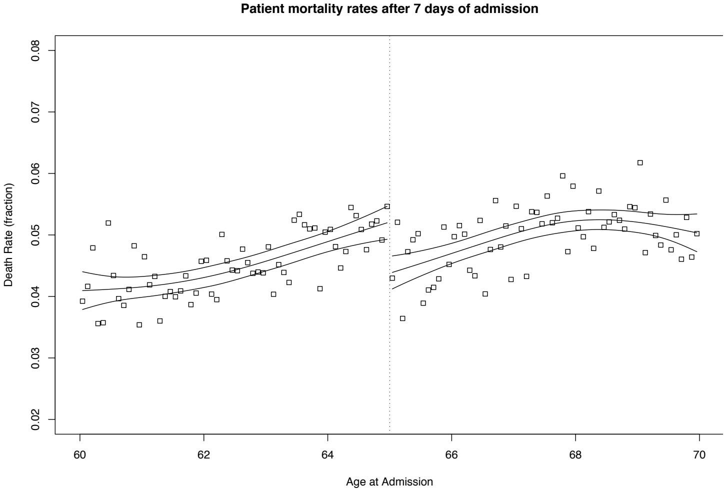
rdplot d7 age, c(65) binselect(esmv) ci(95)



esmv: Mimicking Variance evenly-spaced method

Ejemplo 1: Card (2009)

Otra alternativa es usar métodos no-paramétricos tradicionales (Lowess, Splines, Kernel, etc). Ejemplo para la mortalidad a los 7 días en R:



Does Medicare Save Lives

- El principal resultado es que ser elegible para Medicare reduce la mortalidad.
- Para el grupo que muere a los 7 días de admitidos, la reducción es de un 1 punto porcentual (0.8pp en nuestra muestra). Es un efecto grande dado que la mortalidad era de 0.05 y baja a 0.04...
- La población analizada es una muy enferma, luego hay que ser cautos en la **validez externa** de este hallazgo
- Dado que son admisiones por E.R. es difícil que los enfermos "manipulen" la fecha que van al hospital
- El tamaño de la vecindad se determina de forma de minimizar el error cuadrático medio (MSE) a la Calonico, et. al (2014).

Ejemplo 2: Camacho y Conover (2013)

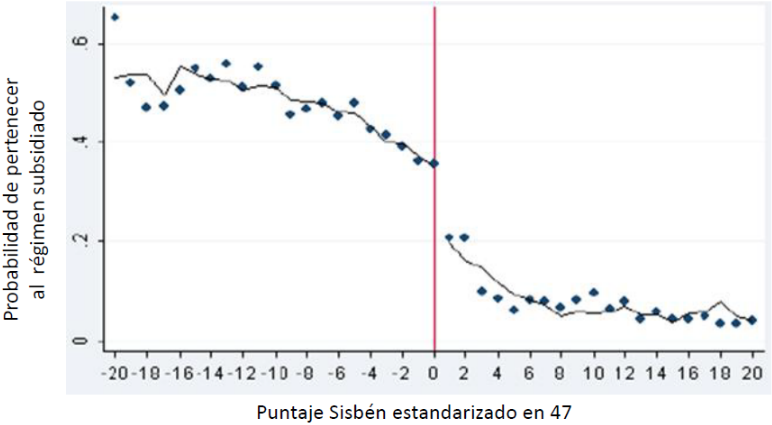
Diseño Fuzzy

Ejemplo 2

- Camacho y Conover (2013) "Effects of Subsidized Health Insurance on Newborn Health in a Developing Country"
- Las autoras evalúan el efecto de tener un seguro de salud (madres) sobre la salud de niños recién nacidos en Colombia.
- Una persona es elegible para recibir el subsidio si tiene un puntaje Sisbén (el equivalente a la FPS en Chile) menor o igual a 47 puntos.
- Sin embargo, hay personas que teniendo puntaje inferior a 47 no acceden al subsidio por tener un empleo formal; mientras que otras personas si acceden aún con puntaje mayor a 47 (indígenas, desplazados e indigentes).
- Lo anterior implica que tenemos un diseño **Fuzzy**

Ejemplo 2

Note que la probabilidad no pasa de 1 a 0 sino que decrece, pero aún así presenta un salto en 47 puntos:

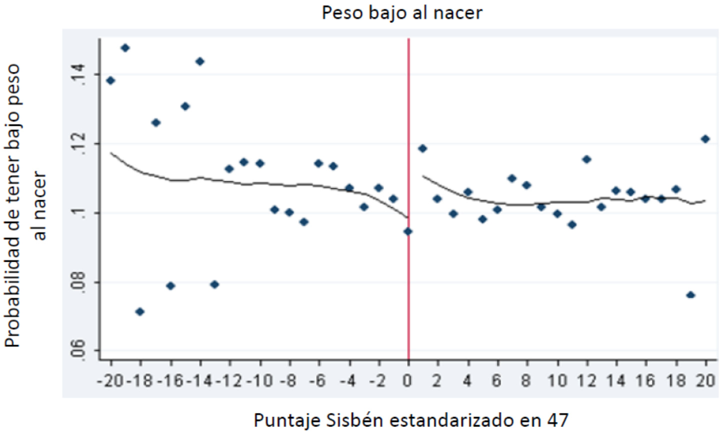


Ejemplo 2

- Clasifican sus variables de resultado en 2 categorías:
    1. Acceso a servicios de salud para madres embarazadas: cantidad de visitas prenatales, si el niño nació en hospital, si hubo médico en el momento del parto
    2. Estado de salud de niños recién nacidos: Logaritmo del peso al nacer, si el niño nació con peso bajo o muy bajo y si el nacimiento fue prematuro
- Encuentran un efecto significativo: una expansión del 10 % en el régimen subsidiado de salud reduce la cantidad de nacimientos con bajo peso entre 1.3 % y 3.7 %
- No debemos olvidar que este efecto es muy local: solo para quienes su puntaje es cercano a 47 y para quienes tener un puntaje de 47 activa el acceso al subsidio

Ejemplo 2

Observe que quienes están justo debajo del umbral tienen una probabilidad de 0.1 de nacer con peso bajo, mientras que luego salta a casi 0.12:



Ejemplo 2

Las autoras presentan un histograma para el puntaje Sisbén (normalizado, S=Sisben-47)