Regresión en Discontinuidad

2025-05-28

Introducción

Introducción

Cuando usamos variables instrumentales (IV), tratamos de encontrar una forma "exógena" de afectar nuestra variable de interés.

Pero ¿qué sucede cuando **conocemos** las reglas que determinan la asignación del tratamiento?

Diseño de regresión discontinua

Diseño de regresión discontinua

Imaginemos que queremos estudiar los efectos de asistir a la USACH.

Existe un examen y la admisión se decide por un corte simple, por ejemplo 200.

Un estudiante que obtiene 199 no es admitido; uno que obtiene 201 sí lo es.

Queremos saber cómo ingresar a la USACH afecta los ingresos futuros.

Diseño de regresión discontinua

En algunas situaciones, la asignación del tratamiento se basa en un punto de corte de una "variable de puntuación" X conocida.

$$D_i = 1 \text{ si } X_i \geq c$$

$$D_i = 0 \text{ si } X_i < c$$

En este caso, X es el puntaje del examen y c es el corte de admisión (200).

Diseño de regresión discontinua

Supongamos que sabemos que ir a la USACH aumenta los ingresos futuros en 800 mil pesos mensuales.

Imaginemos que:

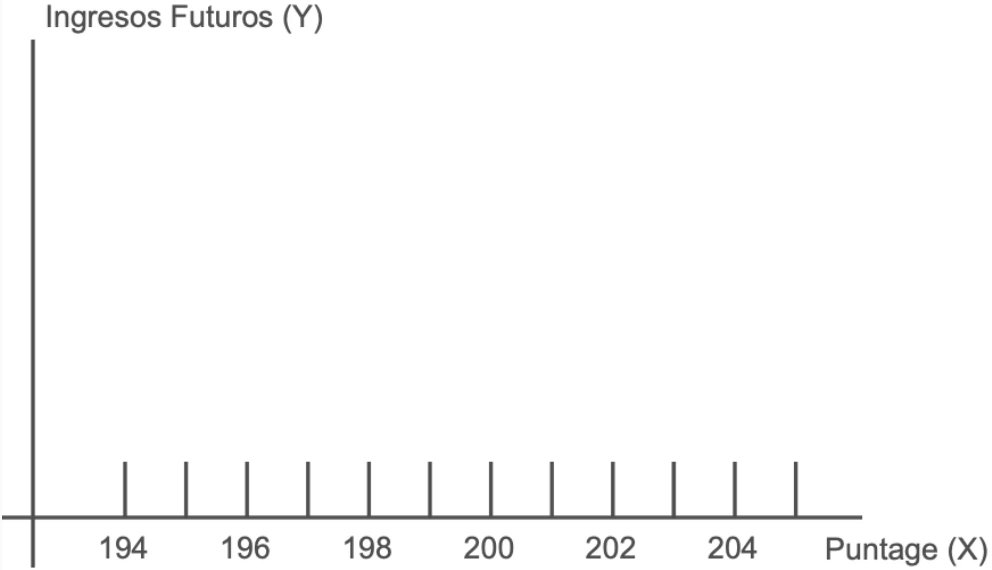
- Seguimos a todas las personas que rindieron el examen.
- Agrupamos a todos según su puntaje (entre 190 y 191, entre 191 y 192, ...).
- Calculamos el ingreso promedio de cada grupo.
- Graficamos esos promedios.

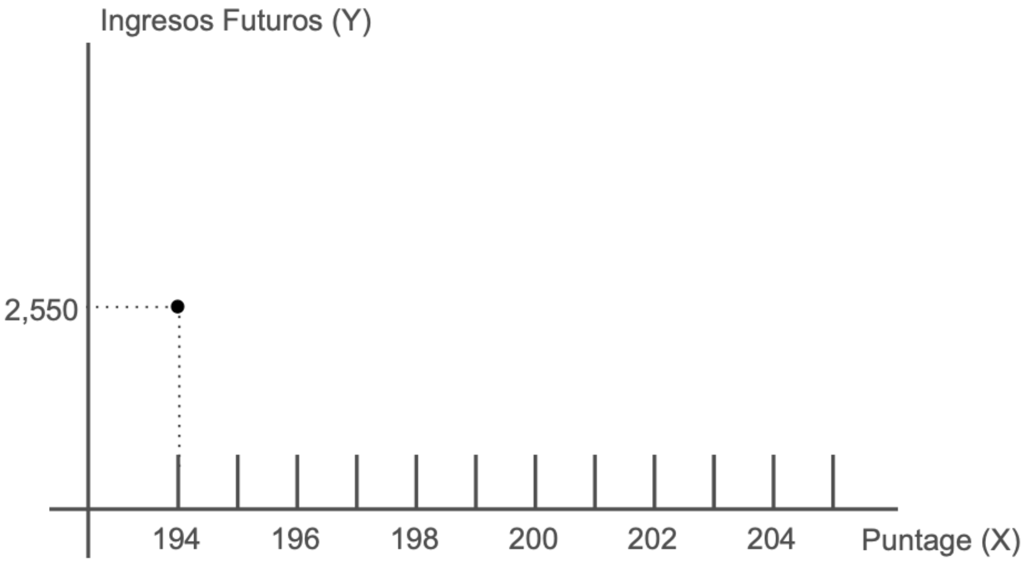
¿Cómo esperarías que se viera el gráfico?

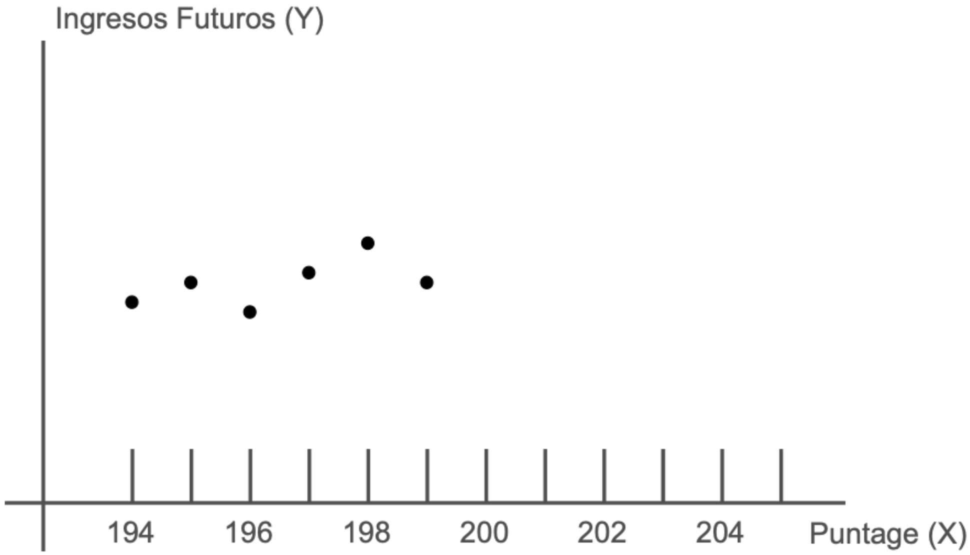
Diseño de regresión discontinua

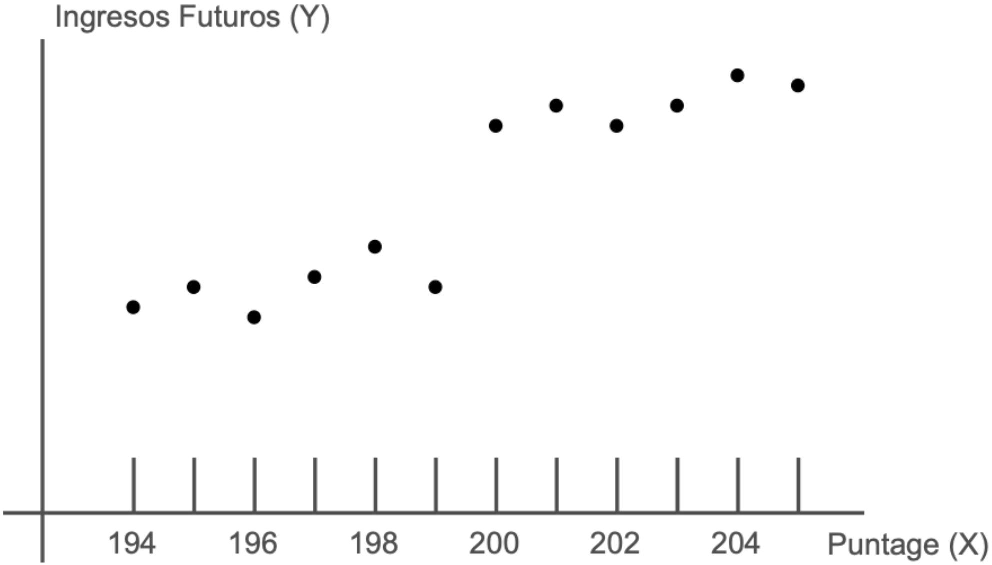
Visualicemos el gráfico:

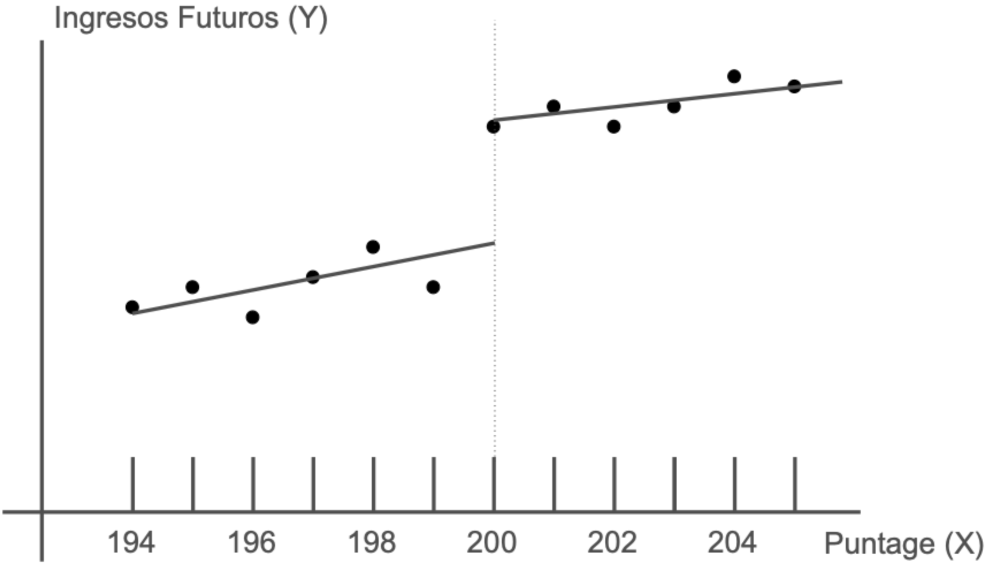
Comenzamos con el ingreso promedio de quienes tuvieron puntaje más bajo y avanzamos hacia arriba.


Ingresos Futuros (Y)
194
196
198
200
202
204
Puntage (X)


Ingresos Futuros (Y)
2,550
194
196
198
200
202
204
Puntage (X)


Ingresos Futuros (Y)
194
196
198
200
202
204
Puntage (X)


Ingresos Futuros (Y)
194
196
198
200
202
204
Puntage (X)


Ingresos Futuros (Y)
194
196
198
200
202
204
Puntage (X)

Características clave del DRD:

- La asignación del tratamiento es una función conocida de la variable de puntuación.
- El tratamiento es discontinuo en X: no se recibe por debajo de c y sí por encima de c.

Diseño de regresión discontinua

Normalmente X está correlacionada con el resultado Y, por lo que no podemos simplemente comparar tratados y controles.

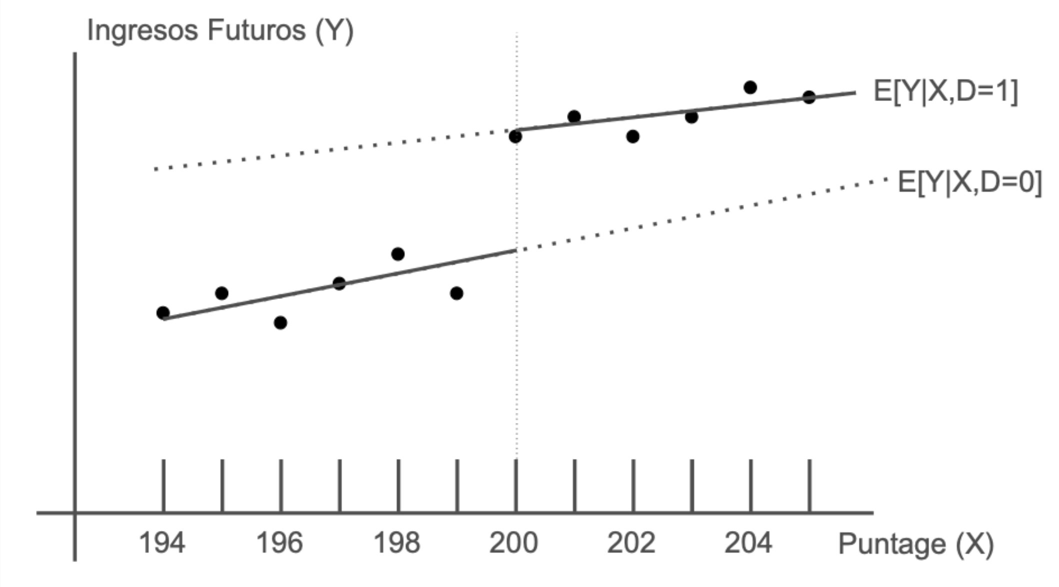
- Los estudiantes que entran a la USACH tienen puntajes más altos que quienes no lo hacen; probablemente les iría mejor aunque la USACH no existiera.
- Pueden existir otros factores que afecten Y y estén correlacionados con X.
- Pero cerca del corte estos factores deberían ser muy similares.

Diseño de regresión discontinua

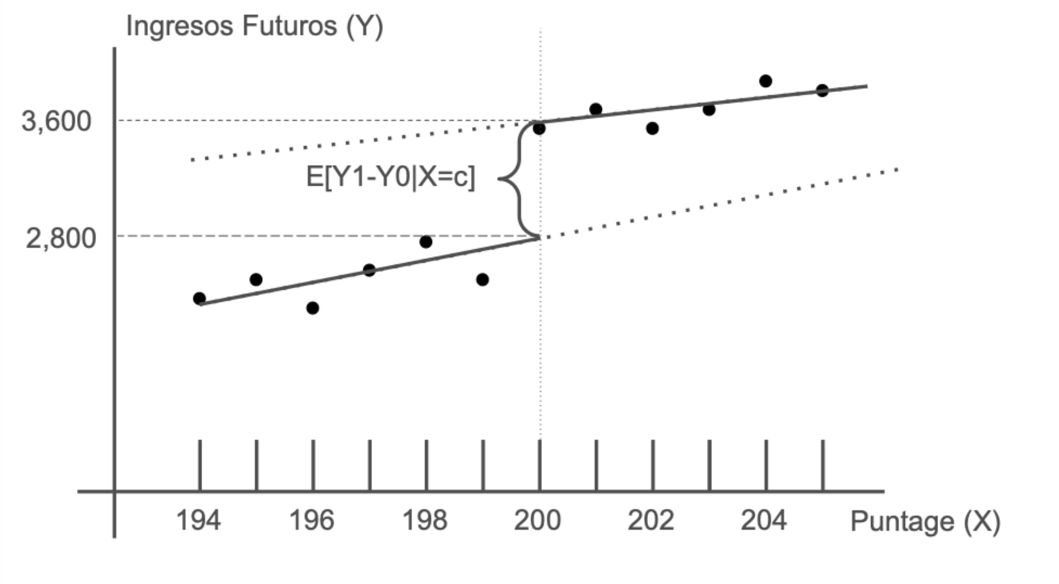
La intuición básica es que, *cerca del corte*, la asignación se parece a un experimento aleatorio.

- Un estudiante que obtuvo 199,5 es muy similar a uno que obtuvo 200: la diferencia es prácticamente aleatoria.

Diseño de regresión discontinua



Diseño de regresión discontinua



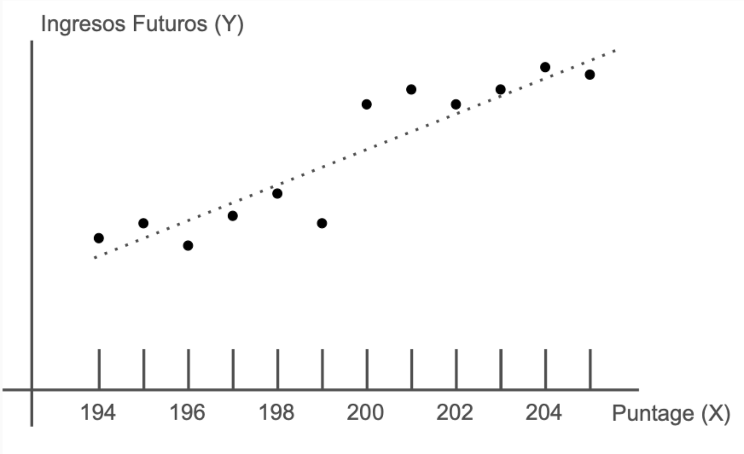
Diseño de regresión discontinua

Entonces, ¿cómo lo estimamos?

¿Qué pasaría si simplemente regresionamos Y sobre X?

Diseño de regresión discontinua

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$



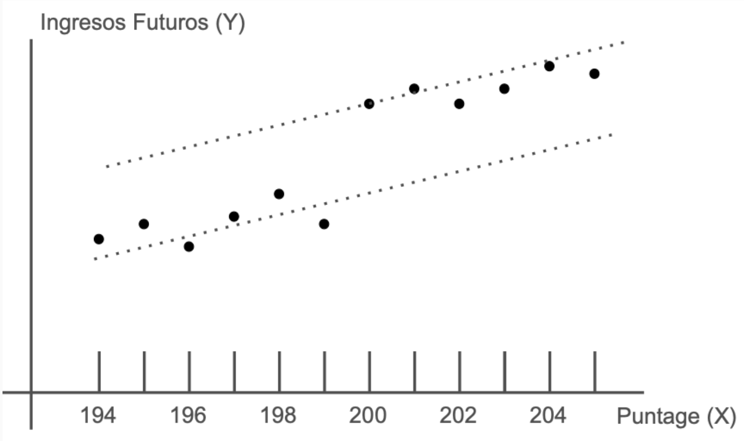
Diseño de regresión discontinua

Si solo regresionamos Y sobre X, "perdemos" la discontinuidad.

¿Qué tal si añadimos una dummy de tratamiento?

Diseño de regresión discontinua

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \mathbf{1}[X_i \geq 0] + u_i$$



Diseño de regresión discontinua

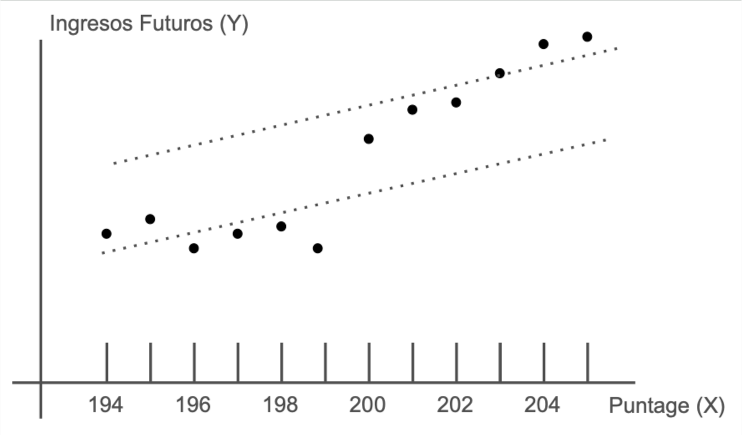
¡Esto es mucho mejor!

Sin embargo, seguimos imponiendo que las pendientes a ambos lados sean iguales. Podrían ser distintas; en ese caso volveríamos a tener un problema.

Añadamos la interacción entre tratamiento y variable de puntuación.

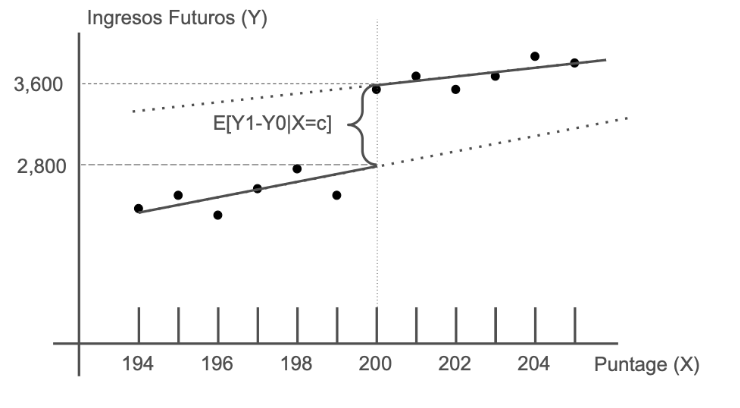
Diseño de regresión discontinua

Ejemplo con pendientes diferentes:



Diseño de regresión discontinua

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \mathbf{1}[X_i \geq 0] + \beta_3 X_i \cdot \mathbf{1}[X_i \geq 0] + u_i$$



¿Cuál es el efecto estimado?

Podemos pensar el parámetro estimado como la diferencia entre:

- el valor predicho para $X = c$ usando solo los datos de los aceptados, y
- el valor predicho para $X = c$ usando solo los datos de los no aceptados.

Queremos obtener el salario promedio para la misma persona, con y sin tratamiento.

Diseño de regresión discontinua

El efecto causal es la diferencia entre el valor predicho de Y para $X = c$ "aproximándonos desde la derecha" y el mismo "aproximándonos desde la izquierda":

$$E[Y|X = c^-] = \beta_0 + \beta_1 c$$

$$E[Y|X = c^+] = \beta_0 + \beta_1 c + \beta_2 + \beta_3 c$$

$$\delta = E[Y|X = c^+] - E[Y|X = c^-] = \beta_2 + \beta_3 c$$

Diseño de regresión discontinua

En la práctica, conviene redefinir la variable de puntuación de modo que $c = 0$.

Por ejemplo, en vez de usar el puntaje del examen (X), usamos la distancia al corte (X'). Así, un valor negativo corresponde a rechazados y uno positivo a admitidos.

Una ventaja es que en la ecuación anterior el segundo término desaparece:

$$\delta = \beta_2 + \beta_3 \cdot 0 = \beta_2$$

Diseño de regresión discontinua

La idea básica es usar una regresión para predecir los valores en c para controles y tratados.

Esto es válido solo porque el tratamiento es lo único que cambia **discontinuamente** en c.

Ambos grupos pueden diferir en promedio, pero son muy similares **alrededor** de c.

Diseño de regresión discontinua

Podemos expresar el supuesto clave así: los resultados potenciales son funciones continuas de X.

- $E[Y^1|X]$ y $E[Y^0|X]$ son continuas en $X = c$.

Si esto se cumple, podemos estimar el efecto **para la población con** $X = c$.

Diseño de regresión discontinua

- $E[Y^1|X]$ y $E[Y^0|X]$ son continuas en $X = c$.

¿Cuándo fallaría este supuesto?

- $E[Y^1|X]$ y $E[Y^0|X]$ son continuas en $X = c$.

¿En qué casos puede fallar?

La forma más sencilla es que exista otra variable que también cambie de forma discontinua en el mismo punto (¿admisión a otra universidad?).

Por ejemplo, algunos estudios analizan el efecto de recibir seguro médico gratuito a partir de un programa que lo concede a todos los mayores de 65 años.

¿Cuál es el problema?

Diseño de regresión discontinua

Muchas cosas cambian a los 65 años: edad típica de jubilación, elegibilidad a varios programas, etc.

Debemos ser cautelosos: los efectos estimados serán la combinación de todos los cambios relevantes en el corte.

Diseño de regresión discontinua

- $E[Y^1|X]$ y $E[Y^0|X]$ son continuas en $X = c$.

¿Cuándo más puede fallar el supuesto?

Otro caso es la posible manipulación de la variable de puntuación.

Por ejemplo, en Brasil existe una regla que asigna dinero federal a los municipios; al superar ciertos umbrales poblacionales (10 000, 40 000 ...) el municipio recibe más fondos.

¿Qué crees que ocurre cuando se realiza el censo?

Diseño de regresión discontinua

Resulta que en cada censo hay muchos menos municipios con 9 999 habitantes que con 10 001.
Los alcaldes tienen un fuerte incentivo para "encontrar" habitantes adicionales durante el conteo.

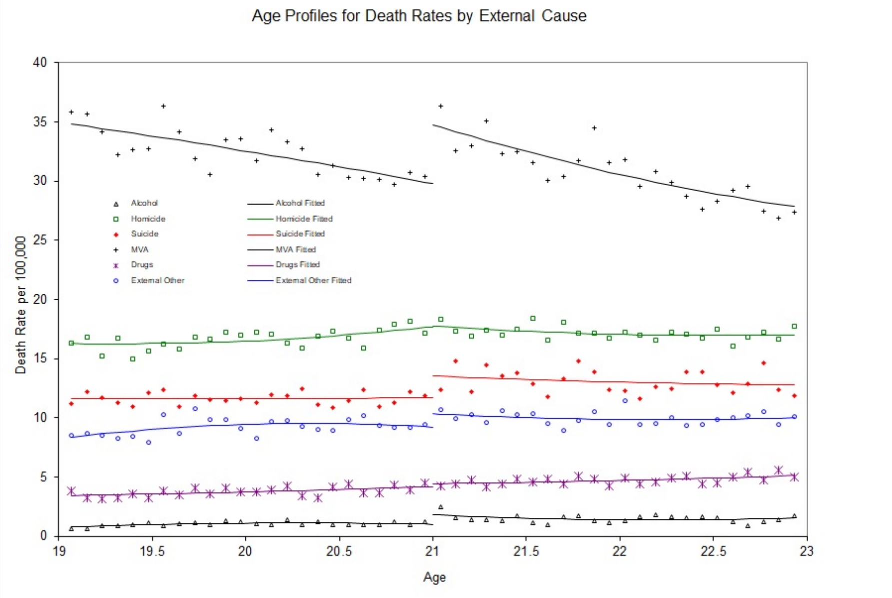
Si esto sucede, tal vez los municipios por encima del corte sean mejores manipulando las reglas.

Diseño de regresión discontinua

Veamos un ejemplo.

En Estados Unidos, la edad legal para beber es 21 años.
Las personas de 20 y 21 son muy parecidas, pero solo las de 21 pueden comprar alcohol legalmente.

Por lo tanto, podemos usar esta discontinuidad para estimar los efectos de poder beber legalmente.


Age Profiles for Death Rates by External Cause
Death Rate per 100,000
Age
Alcohol
Homicide
Suicide
MVA
Drugs
External Other
Alcohol Fitted
Homicide Fitted
Suicide Fitted
MVA Fitted
Drugs Fitted
External Other Fitted
0
5
10
15
20
25
30
35
40
19
19.5
20
20.5
21
21.5
22
22.5
23

Diseño de regresión discontinua

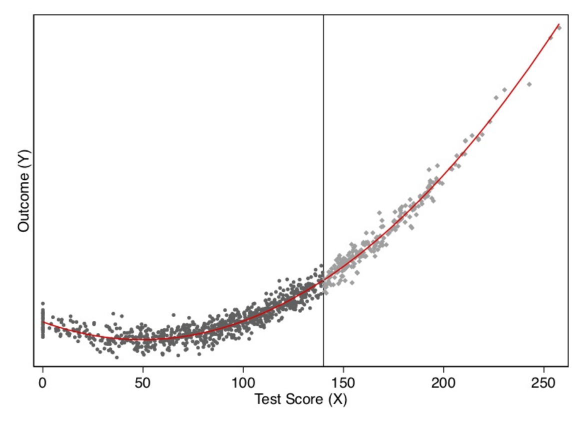
El RDD es especialmente sensible a la no linealidad en X.

Debido a que estamos interesados en pronosticar el resultado en un punto específico, debemos tener cuidado de agregar suficiente flexibilidad para hacerlo bien.

Primero, veamos cómo puede ser un problema y luego qué podemos hacer al respecto.

Diseño de regresión discontinua

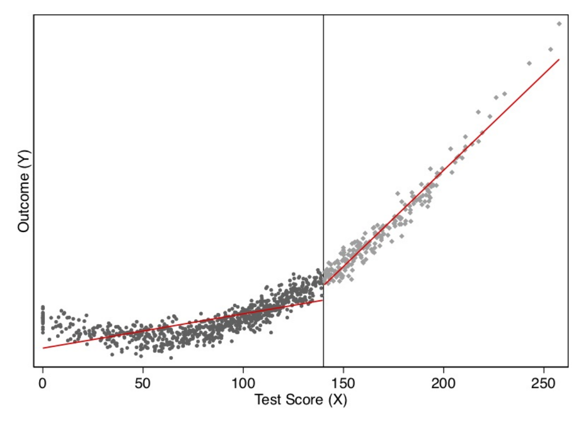
Aquí hay una simulación donde la expectativa condicional de $Y|X$ es cuadrática. No hay discontinuidad en el punto de corte: deberíamos estimar un cero.



Diseño de regresión discontinua

Sin embargo, si usamos una función lineal en cada lado, terminamos estimando una discontinuidad.

La función lineal no pronostica bien, particularmente en el lado izquierdo.



Diseño de regresión discontinua

Una forma de abordar el problema es usar términos de orden superior. Por ejemplo:

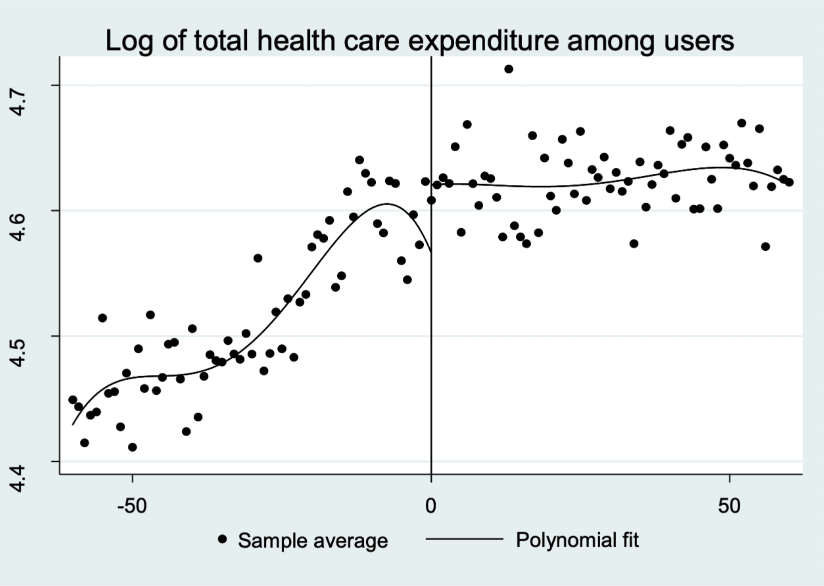
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \gamma_0 1[X_i > 0] + \gamma_1 X_i 1[X_i > 0] + \gamma_2 X_i^2 1[X_i > 0] + \varepsilon_i$$

¿Cuál es el efecto estimado aquí?

Diseño de regresión discontinua

Incluso podemos ir más allá e incluir la tercera, cuarta, quinta potencia.

Sin embargo, esto rápidamente se convierte en un problema. En el margen, las potencias más altas aportan menos al ajuste. Peor aún, pueden empeorar el ajuste en los extremos, *exactamente donde queremos mejorar*.


Log of total health care expenditure among users
4.4
4.5
4.6
4.7
-50
0
50
Sample average
Polynomial fit

Diseño de regresión discontinua

Una alternativa popular es usar la regresión no paramétrica.

Esto significa, esencialmente, calcular promedios locales del valor de Y en cada valor de X.

Por ejemplo, para cada posible valor de X=x, podemos calcular el valor promedio de Y para todos los puntos dentro de 1 unidad de x.

De esta manera, podemos “acercarnos” a la discontinuidad desde cada lado y pronosticar el valor de Y en cada lado.

Diseño de regresión discontinua

Otra elección importante se llama el “ancho de banda” (*bandwidth*).

Los valores de X muy lejanos probablemente tienen poca utilidad para predecir Y en el punto de corte.

Así que, probablemente podamos excluir los valores lejanos.

¿Pero cuánto?

Diseño de regresión discontinua

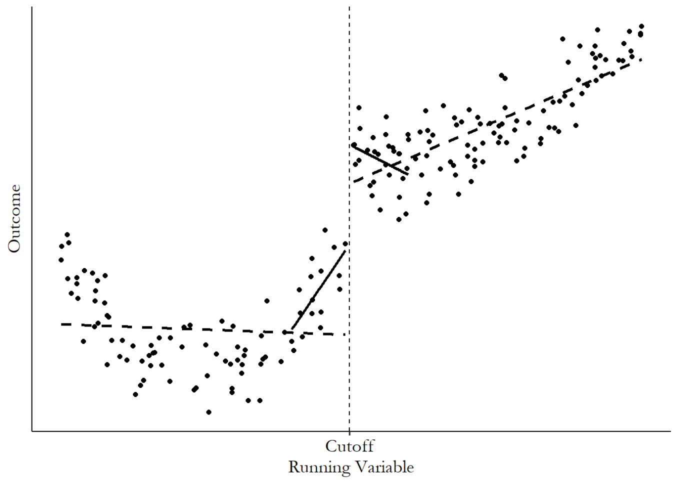
Si excluimos demasiado, nuestro pronóstico será ruidoso.

En la práctica, podemos mirar el gráfico para hacer una elección plausible.

O, más comúnmente, podemos usar algoritmos automatizados que eligen el ancho de banda "óptimo", equilibrando esta compensación. Esto está incluido en el paquete 'rdrobust'.

Diseño de regresión discontinua

Ejemplo de RD con y sin restricción de ancho de banda.



Diseño de regresión discontinua

Finalmente, para intentar predecir bien en la discontinuidad, podemos agregar más peso a las observaciones cercanas al punto de corte.

Esencialmente, le estamos diciendo a la regresión "intenta hacer un buen trabajo alrededor de esta parte".

Diseño de regresión discontinua

¿Qué hacer en la práctica?

El enfoque más común es usar la regresión lineal local.

- Logra un equilibrio entre aproximar la Función de Expectativa Condicional y el *overfitting*.

A menudo se utilizan la elección del ancho de banda y la ponderación.

Diseño de regresión discontinua

Lo más importante es confiar en tus ojos. Haz gráficos y comprueba si parece que hay una discontinuidad.

Las mejores RD son obvias en el gráfico y no necesitan econometría complicada.

Ejercicio Aplicado

Ahora veremos algunos ejemplos de codificación de cómo hacer RDD.

Utilizaremos datos de Lee, Moretti y Butler (2004).

JOURNAL ARTICLE

Do Voters Affect or Elect Policies? Evidence from the U. S. House Get access >

David S. Lee , Enrico Moretti , Matthew J. Butler

The Quarterly Journal of Economics, Volume 119, Issue 3, August 2004, Pages 807–859, https://doi.org/10.1162/0033553041502153

Published: 01 August 2004

Ejercicio Aplicado

En este artículo, intentan distinguir entre dos teorías diferentes de cómo funciona la política.

- A: ¿Los votantes *afectan* a los políticos, haciéndolos votar y legislar de manera diferente según sus preferencias?
- B: ¿O los votantes simplemente *eligen* a los políticos que luego promulgan sus políticas preferidas?

Ejercicio Aplicado

Según ambas teorías, un estado con más votantes de derecha tendrá una política más de derecha.

Pero la teoría A predice que pasar del 49.9% de votantes republicanos al 50.1% de votantes republicanos no debería hacer mucha diferencia. El mensaje es el mismo: este estado está dividido equitativamente y probablemente prefiere un enfoque centrista.

La teoría B predice una gran diferencia: en un caso se elige a un demócrata, que promulga legislación demócrata; en el otro caso se elige a un republicano, que promulga legislación republicana.

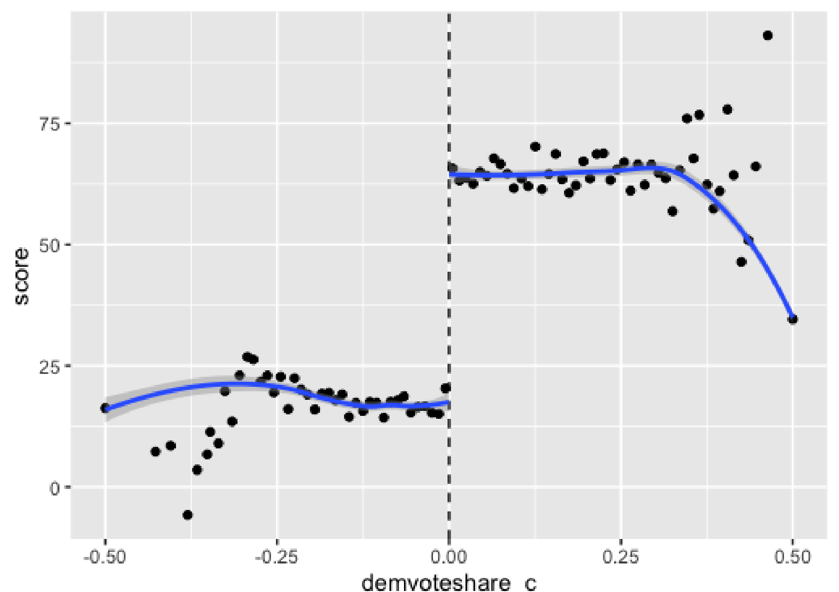
Así, al observar lo que sucede con la legislación en el punto de corte, podemos distinguir las teorías.

Ejercicio Aplicado

Analizaremos datos de elecciones estatales estadounidenses sobre el margen de votos para cada candidato.

Mediremos cuán “de izquierda” fue el registro de votación de cada congresista utilizando una puntuación llamada puntuación ADA. Básicamente: 0 significa que siempre votan a la derecha, 100 significa que siempre votan a la izquierda.

Ejercicio Aplicado



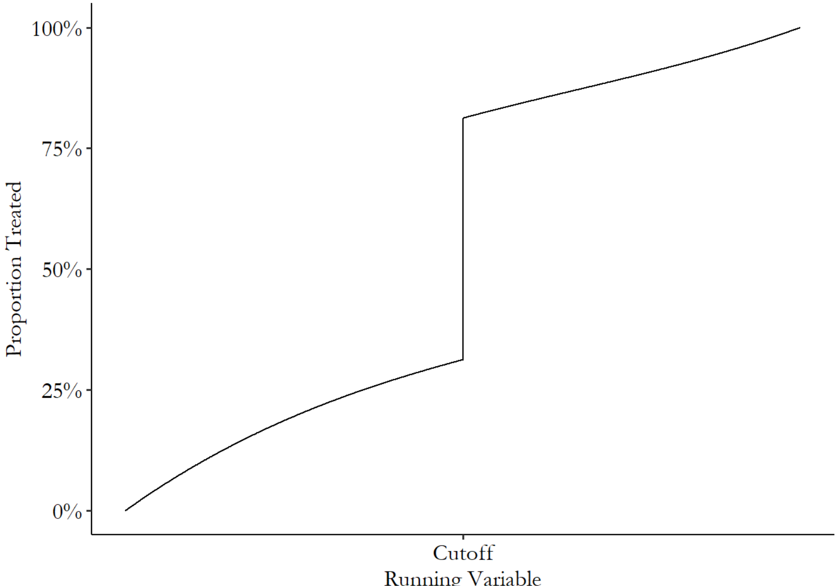
RD Difusa

A veces, la probabilidad de tratamiento no cambia completamente de 0 a 1 en el punto de corte.

- Quizás el punto de corte determina la elegibilidad, pero no todos solicitan un programa.
- Quizás hay cierto grado de arbitrariedad, pero superar el punto de corte aumenta la probabilidad de tratamiento.

En estos casos tenemos lo que se llama una "RD Difusa" (*Fuzzy RD*). El caso en que el tratamiento va de 0 a 1 se llama RD Nítida (*Sharp RD*).

RD Difusa



RD Difusa

Si pensamos en la RD como un “experimento local”, la RD Difusa es el equivalente a un experimento con cumplimiento imperfecto.

Así, podemos tratarla exactamente como trataríamos un experimento.

Esencialmente, aplicamos variables instrumentales para estimar el ITT, la Primera Etapa y el LATE.

RD Difusa

Ejemplo práctico:

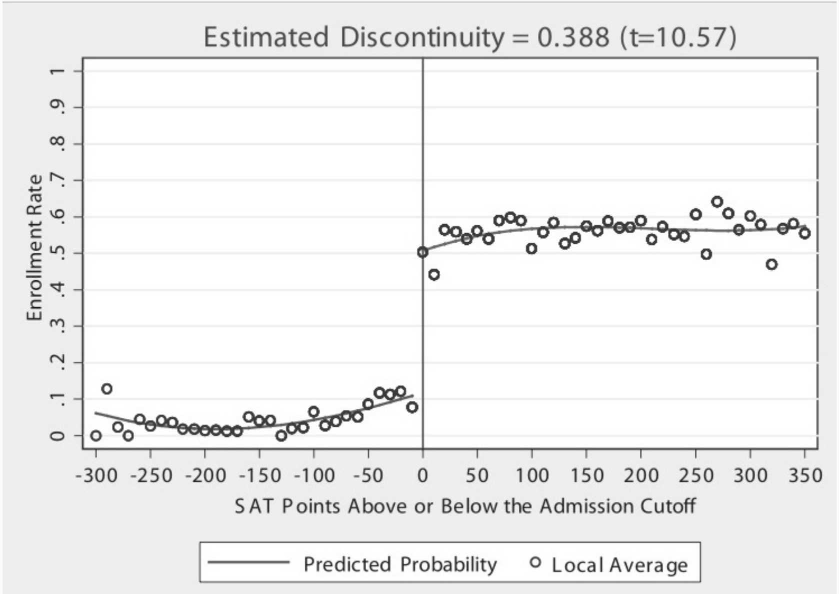
Hoekstra (2009) quiere estimar el efecto causal de asistir a la universidad estatal más grande en cada estado estadounidense.

La idea es que estas universidades son más prestigiosas que otras en el mismo estado.

Pueden usar el punto de corte de admisión del SAT, pero algunas personas por encima del punto de corte eligen ir a una universidad diferente:

- Pueden tener razones personales o estar interesados en un programa específico.
- Además, algunas personas son admitidas con una puntuación más baja por diversas razones.

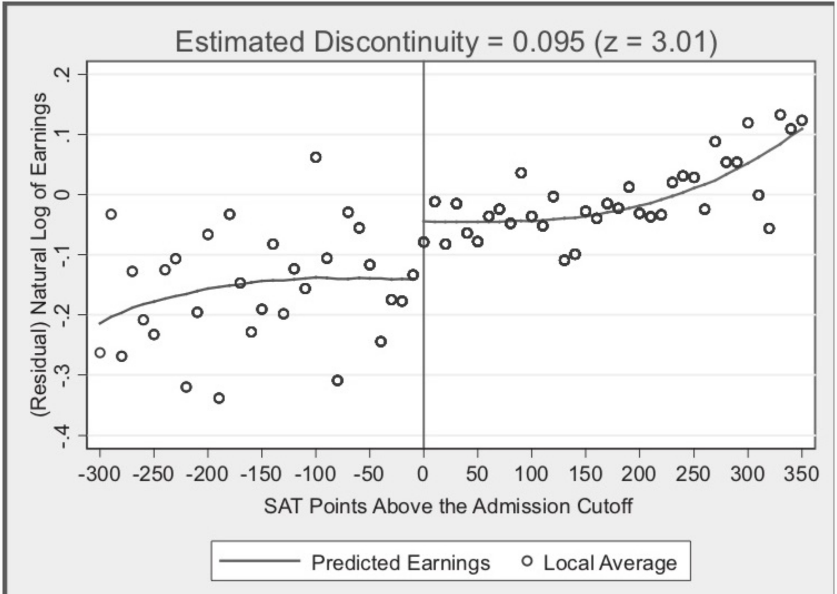
RD Difusa


Estimated Discontinuity = 0.388 (t=10.57)
Enrollment Rate
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
-300
-250
-200
-150
-100
-50
0
50
100
150
200
250
300
350
SAT Points Above or Below the Admission Cutoff
Predicted Probability
Local Average

Entonces, superar el punto de corte aumenta la probabilidad de admisión en aproximadamente un 40%.

Esa es la Primera Etapa.

RD Difusa



Estimated Discontinuity = 0.095 (z = 3.01)
(Residual) Natural Log of Earnings
.2
.1
0
-.1
-.2
-.3
-.4
-300 -250 -200 -150 -100 -50 0 50 100 150 200 250 300 350
SAT Points Above the Admission Cutoff
Predicted Earnings
Local Average

RD Difusa

Y superar el punto de corte aumenta las ganancias logarítmicas en aproximadamente un 0.10.

Así que el ITT es un aumento del 10% en las ganancias.

Combinando estos dos, estimamos que asistir a una universidad pública insignia aumenta las ganancias en...

RD Difusa

Y superar el punto de corte aumenta las ganancias logarítmicas en aproximadamente un 0.10.

Así que el ITT es un aumento del 10% en las ganancias.

Combinando estos dos, estimamos que asistir a una universidad pública insignia aumenta las ganancias en...

25%

Recapitulación

Recapitulación – Idea clave

Asignación discontinua

- Cuando el tratamiento cambia bruscamente en un umbral c de la variable de puntuación X, comparamos unidades de cada lado de c.

Supuesto central: continuidad

- Las funciones de resultado potencial $E[Y0|X]$ y $E[Y1|X]$ son continuas en $X = c$.
- Si otra política o manipulación también cambia en c, el supuesto se rompe.

Interpretación:

- Cerca del punto de corte el diseño se parece a un experimento aleatorio.

Recapitulación – Implementación práctica

Visualiza primero

- Grafica promedios locales de Y contra X.
- La mejor evidencia de un buen RD es “ver” la ruptura.

Regresión lineal local con ancho de banda óptimo.

- Añadir polinomios altos tiene riesgo de sobreajuste.

Recapitulación – Fuzzy RD

Si la probabilidad de tratamiento solo salta parcialmente, usa IV:

$$LATE = \frac{\text{salto en Y}}{\text{salto en D}}$$

Chequeos de robustez:

- Densidad de X (test de manipulación).
- Balance de covariables a cada lado.
- Sensibilidad a distintos anchos de banda y especificaciones.