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:

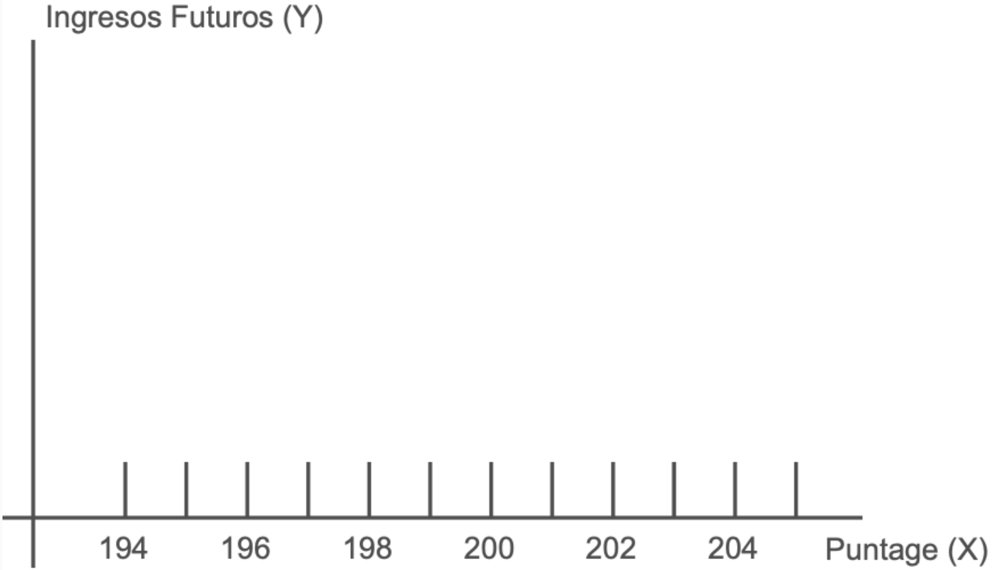
- Seguimos a todas las personas que rindieron el examen.
- Agrupamos a todos según su puntaje (entre 190 y 191, entre 191 y 192, ...).
- Calculamos el ingreso promedio de cada grupo.
- Graficamos esos promedios.

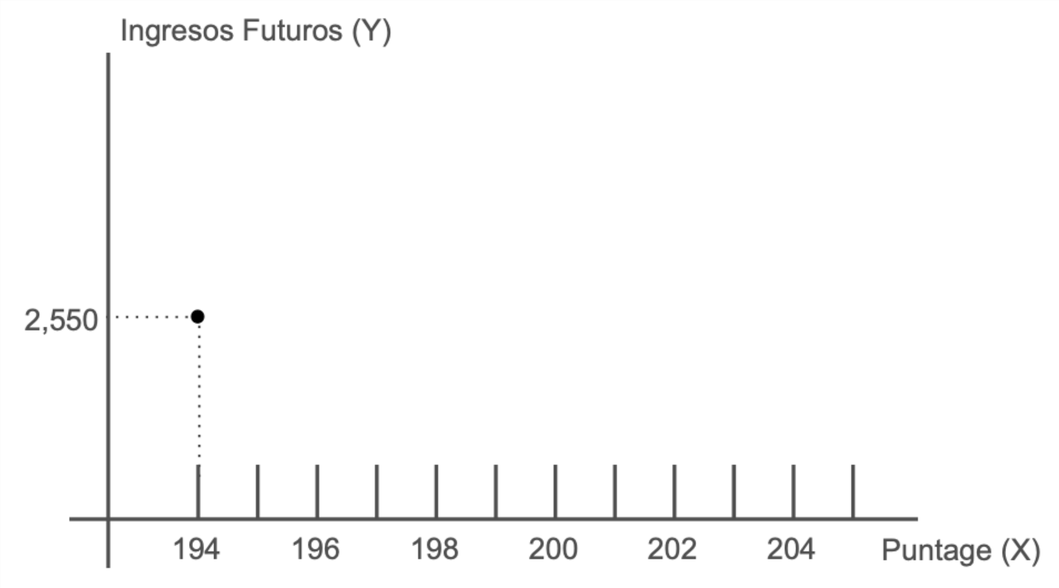
¿Cómo esperarías que se viera el gráfico?

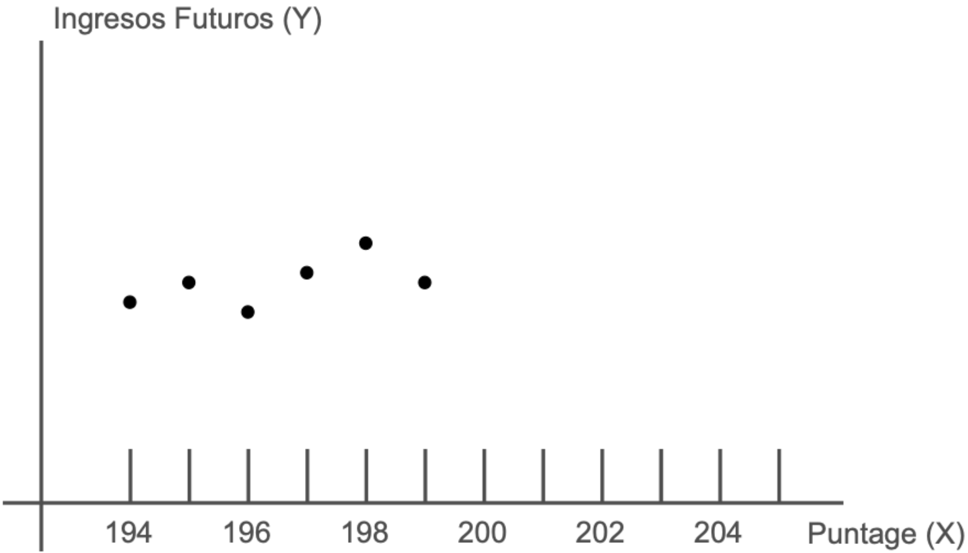
Diseño de regresión discontinua

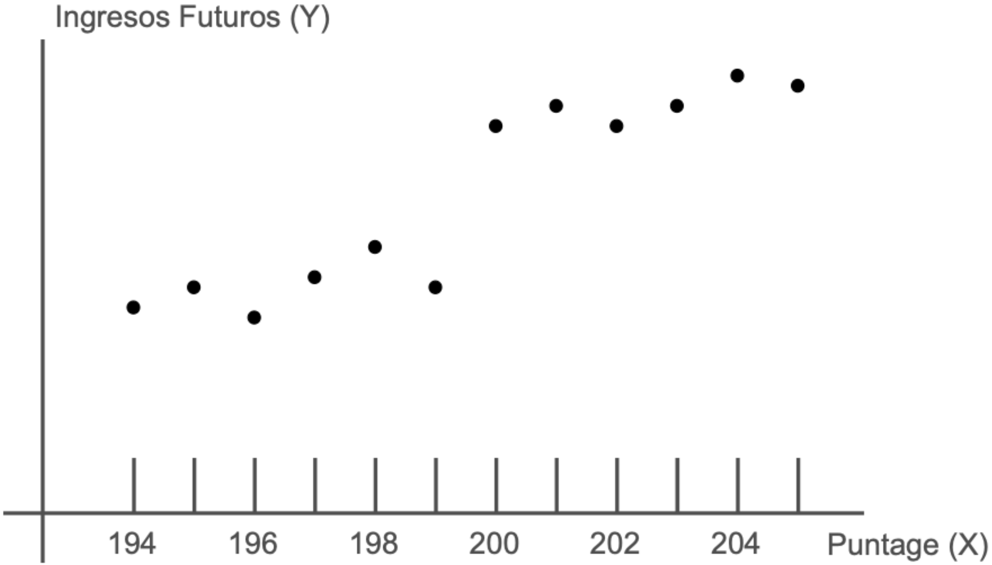
Visualicemos el gráfico:

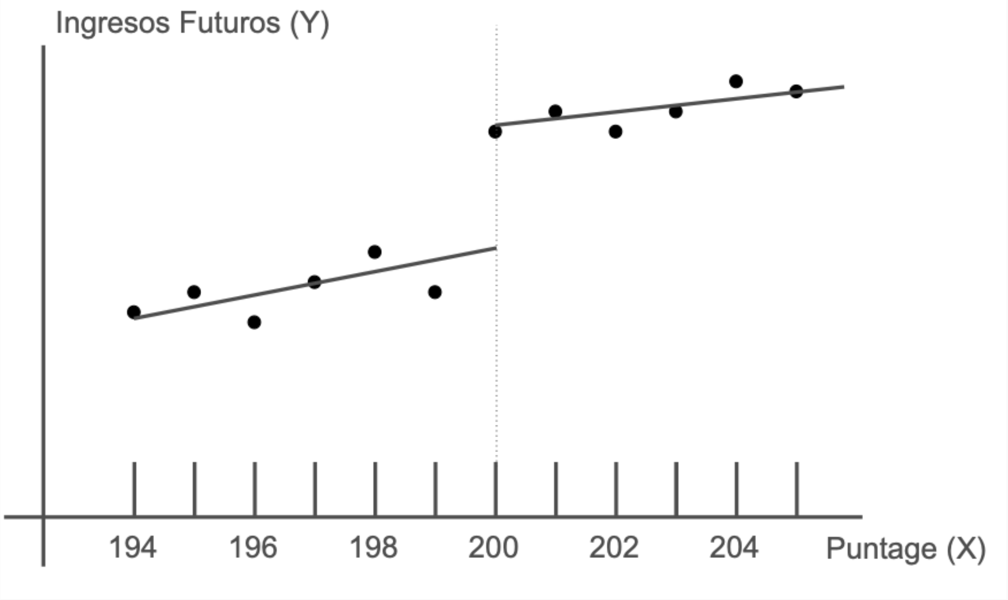
Comenzamos con el ingreso promedio de quienes tuvieron puntaje más bajo y avanzamos hacia arriba.


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)

Diseño de regresión discontinua

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.

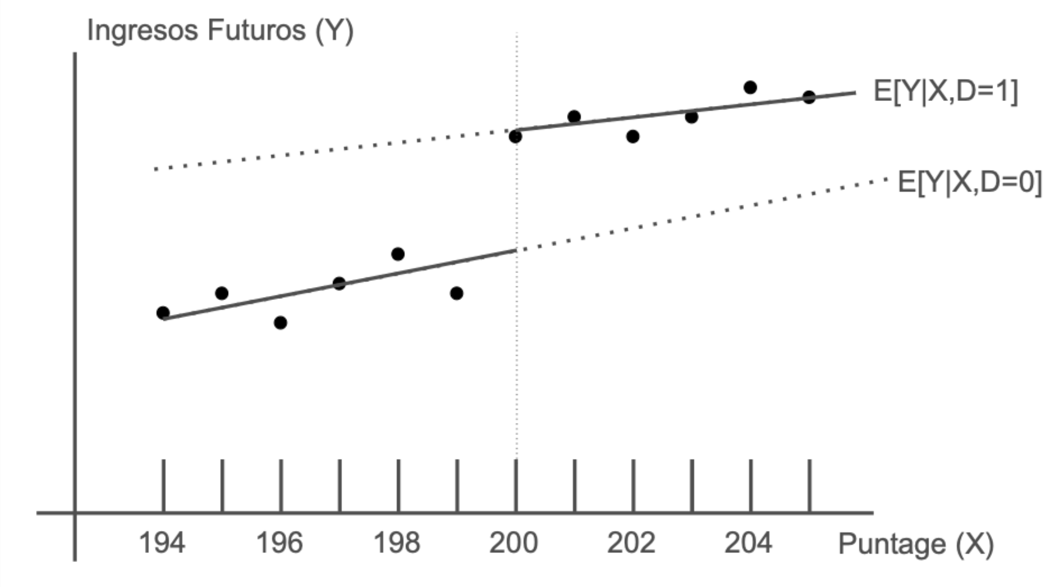
- Los estudiantes que entran a la USACH tienen puntajes más altos que quienes no lo hacen; probablemente les iría mejor aunque la USACH no existiera.
- Pueden existir otros factores que afecten Y y estén correlacionados con X.
- Pero cerca del corte estos factores deberían ser muy similares.

Diseño de regresión discontinua

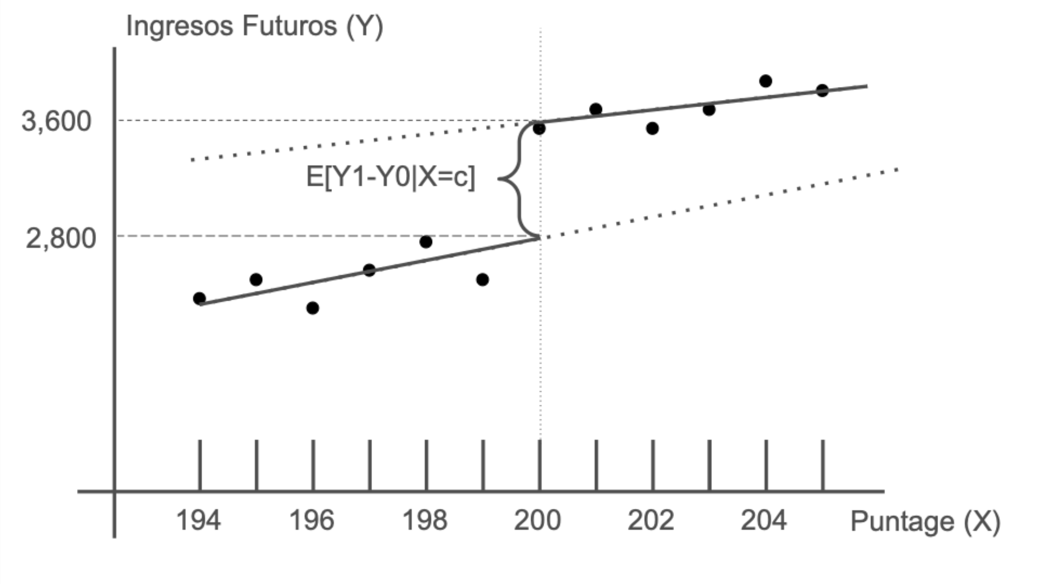
La intuición básica es que, *cerca del corte*, la asignación se parece a un experimento aleatorio.

- 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



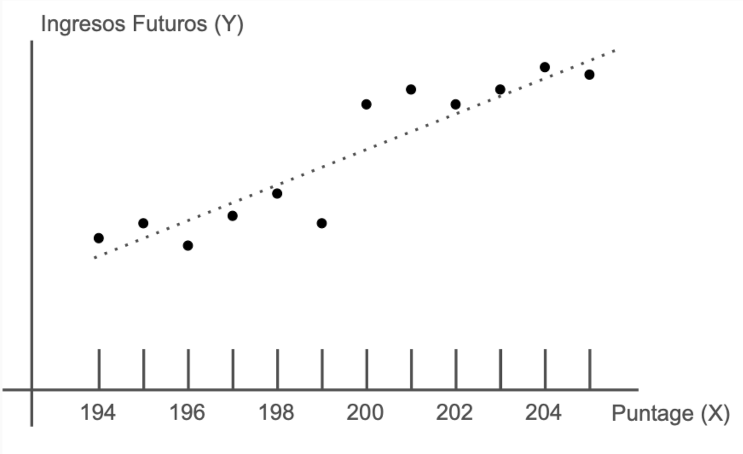
Diseño de regresión discontinua

Entonces, ¿cómo lo estimamos?

¿Qué pasaría si simplemente regresionamos Y sobre X?

Diseño de regresión discontinua

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



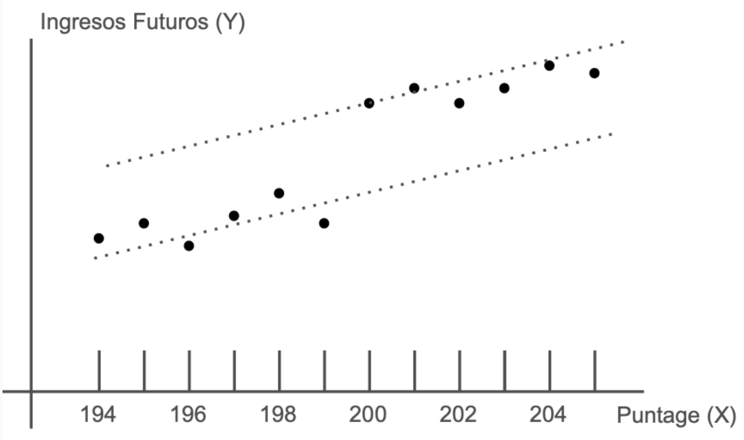
Diseño de regresión discontinua

Si solo regresionamos Y sobre X, “perdemos” la discontinuidad.

¿Qué tal si añadimos una dummy de tratamiento?

Diseño de regresión discontinua

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



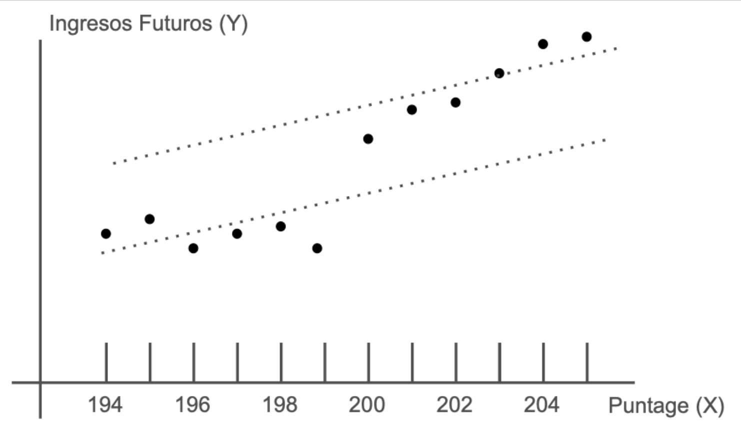
Diseño de regresión discontinua

¡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.

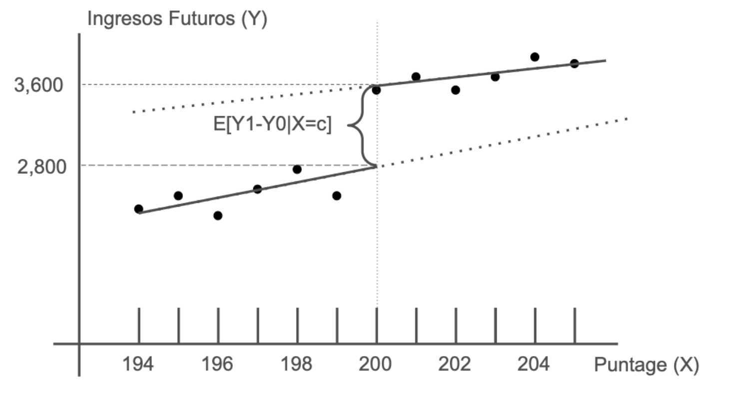
Añadamos la interacción entre tratamiento y variable de puntuación.

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?

Diseño de regresión discontinua

- $E[Y^1|X]$ 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.

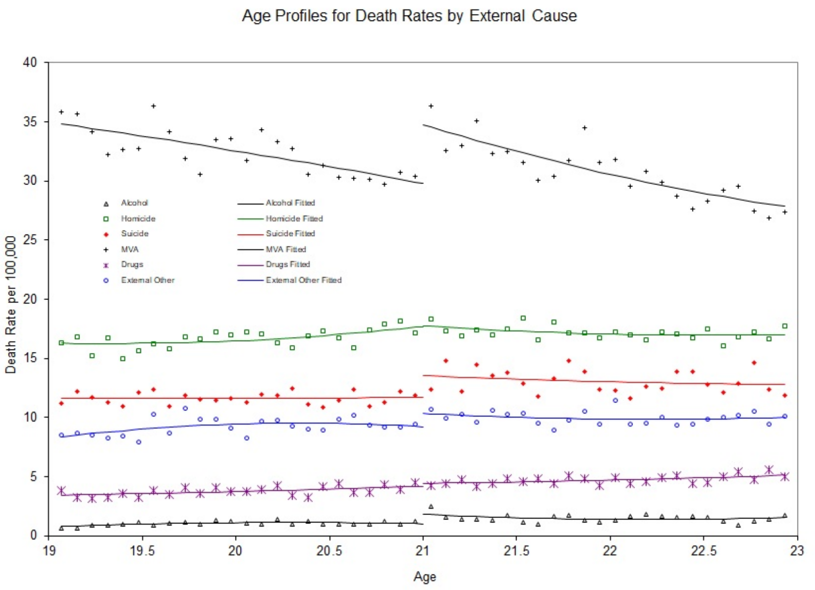
Si esto sucede, tal vez los municipios por encima del corte sean mejores manipulando las reglas.

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
40
35
30
25
20
15
10
5
0
19
19.5
20
20.5
21
21.5
22
22.5
23
Alcohol
Homicide
Suicide
MVA
Drugs
External Other
Alcohol Fitted
Homicide Fitted
Suicide Fitted
MVA Fitted
Drugs Fitted
External Other Fitted