

Mejorar el rendimiento escolar de estudiantes entre 11 y 16 años

Disclaimer: Comenzamos el ejercicio asumiendo que el grupo de tratamiento estaba por encima del umbral. Por lo tanto, los gráficos, las cifras y los resultados están invertidos. Pero las conclusiones se han adaptado.

1. Que é bunching?

“Uno de sus compañeros que no asistió a la ayudante’ia le menciona que para implementar una regresión discontinua, primero deberían revisar si hay evidencia de “bunching” en la running variable, pero no entiende qué significa esto.”

1a) Explicación

“Explique con sus propias palabras a que se puede estar refiriendo su compañero y por qué sería relevante descartar este fenómeno antes de implementar la metodología en cuestión”

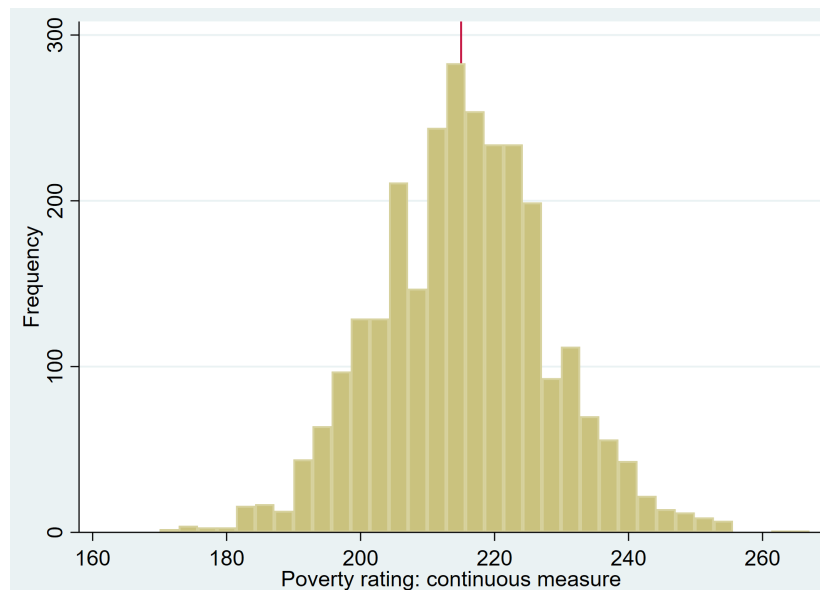
Bunching se refiere a una alta frecuencia de observaciones alrededor del punto de corte. Si esto ocurre, indica que podría haber habido una manipulación de los datos, es decir, que, en este caso, el índice de pobreza de los estudiantes fue manipulado para incluir a más estudiantes por debajo del umbral, o por arriba. Esto significa que una evaluación RD arrojaría resultados poco fiables, y no sería posible utilizar esta metodología.

Para comprobar si hay manipulación o "bunching" deberíamos ver una distribución homogénea de la frecuencia alrededor del corte, con un histograma de frecuencias o con lo test de McCary (2008), por ejemplo.

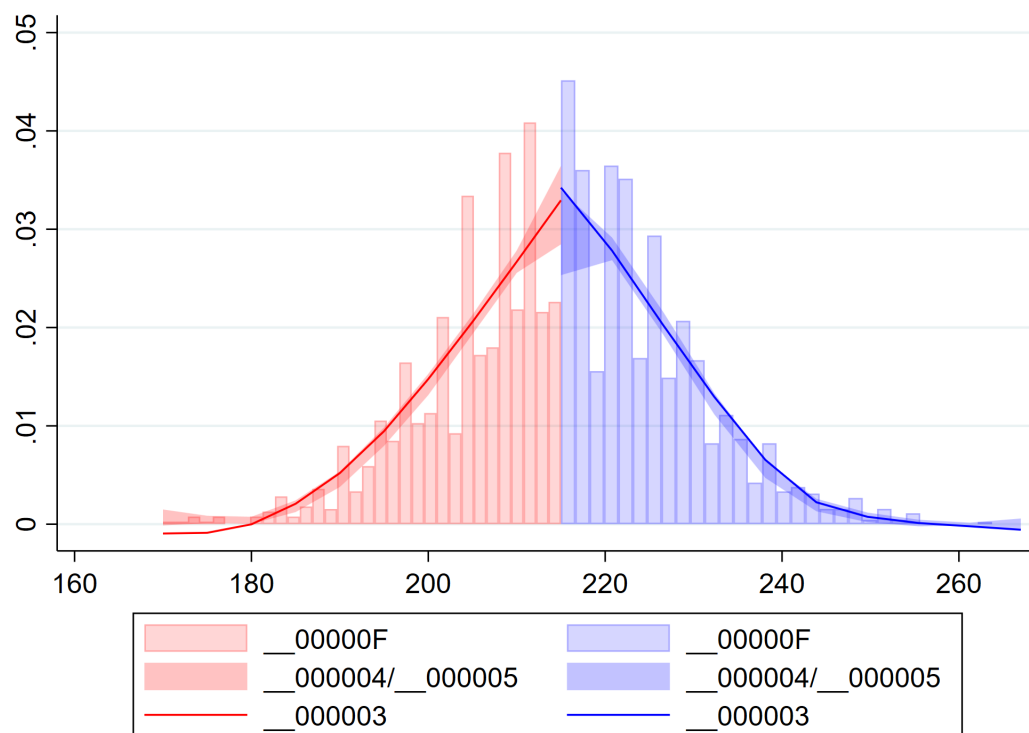
1b) Confirmación de bunching

“Demuestre si existe “bunching” en la running variable presentando evidencia gráfica y estadística para justificar su respuesta.”

Primero se hace un histograma de frecuencias:



De este gráfico, parece que hay un cierto "amontonamiento" de la "variable en curso" poverty_rating antes del umbral. Sin embargo, para una mejor confirmación haremos un test de McCarty, para un nivel de significancia de 5%:



Aquí, si ve que los IC se tocan pero que no la diferencia no será estadísticamente significativa, y se puede seguir con la evaluación.

2. Distribución homogénea de variables alrededor del corte

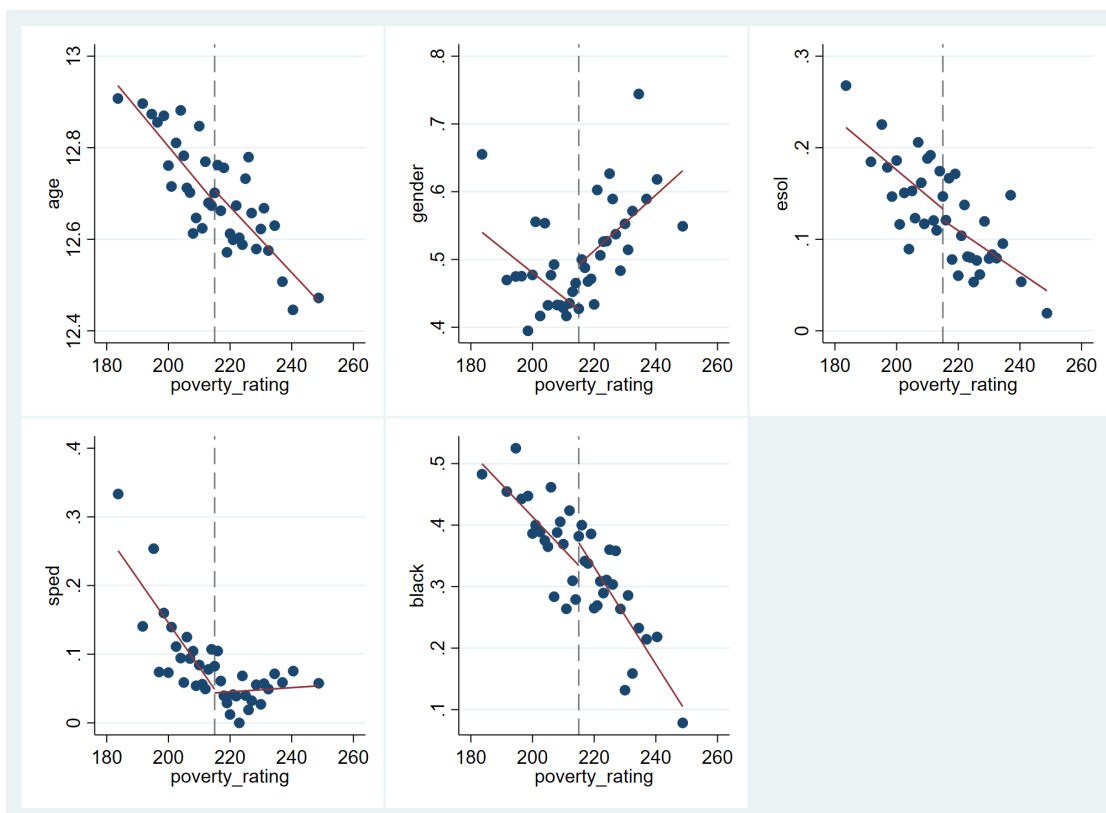
2a) Explicación

Al igual que cualquier otra metodología de evaluación de impacto, es necesario garantizar que existe un contrafactual fiable para el grupo de tratamiento. En otras palabras, debemos comprobar si el contrafactual (alumnos por debajo del umbral) es comparable con el grupo de tratamiento (alumnos por encima del umbral), para un determinado intervalo de tiempo. Esto puede hacerse observando cómo se comportan las variables observables en torno al umbral, es decir, si hay homogeneidad en su distribución. Si encontramos diferencias, entonces los grupos no son comparables y la DR no es fiable, o, entonces, se debe controlar la regresión por las variables que no son comparables entre los grupos.

Podemos comprobar si esto ocurre gráfica y estadísticamente, como se hace en la siguiente pregunta.

2b) Confirmación de homogeneidad

Para evaluar gráficamente si hay homogeneidad alrededor de lo “threshold”, se puede graficar un “binscatter” (se utilizó 50 quantiles) de las variable independientes que se pretenden analizar:



En estos gráficos, se muestra relevante profundar la análisis de las variables “gender”, “esol” y “black” por que las rectas son claramente no coincidentes.

Sin embargo, se realizará una evaluación estadística a la homogeneidad de todas las variables mencionadas (age, gender, esol, sped y black), a través de un "t-test" a los promedios de los grupos completos: tratados ($\text{poverty_rating} > 215$) y no tratados ($\text{poverty_rating} < 215$). Como se van utilizar todas las observaciones, podemos esperar diferencias significativas en las variables, ya que las personas muy pobres son muy diferentes de las muy ricas

Así, para evaluar si existen diferencias estadísticamente significativas entre las mismas variables de los distintos grupos, realizamos un "t-test" para las medias de distribución de estas variables para los distintos grupos, con las siguientes hipótesis:

- H_0 : No hay diferencias entre las medias
- H_1 : Hay diferencias entre las medias

En el caso de las variables que rechazan la hipótesis nula, es decir. valor $p < \alpha$ (0.05), hay pruebas estadísticamente significativas de que la media de estas variables es diferente entre los grupos, por lo que no se pueden comparar los grupos de tratamiento y control.

Los resultados del test t se pueden ver en la siguiente tabla (para un nivel de significancia de 5%), y como esperado, todas las variables rechazan la hipótesis nula (hay diferencias entre los grupos):

Variable	P-Value ($\Pr(T > t)$)	Interpretación
age	0.0000	Rechaza H_0
gender	0.0000	Rechaza H_0
esol	0.0000	Rechaza H_0
sped	0.0000	Rechaza H_0
black	0.0000	Rechaza H_0

Esto significa que los grupos no son comparables y se debe controlar por estas variables, pero esto no significa que no se puede concluir un efecto causal del programa evaluado, solo significa que los grupos no son similares.

3) Evaluación con ancho de banda de 5 puntos

Para una estimación más fiable, deberíamos disminuir el "bandwidth" para tener mayores similitudes entre los grupos de control y de tratamiento. De hecho, el razonamiento

subyacente del diseño de discontinuidad de la regresión es que en el límite del umbral los individuos tienen (en promedio) características estadísticamente similares.

Por lo tanto, si volvemos a realizar las pruebas t, con un ancho de banda de 5 puntos, no deberíamos esperar diferencias significativas entre los grupos (que son, ahora, de menor tamaño):

- Grupo Control: $210 < \text{poverty_rating} < 215$
- Grupo Tratamiento: $215 < \text{poverty_rating} < 220$

Los resultados siguen en la tabla abajo, para un nivel de significancia de 5%:

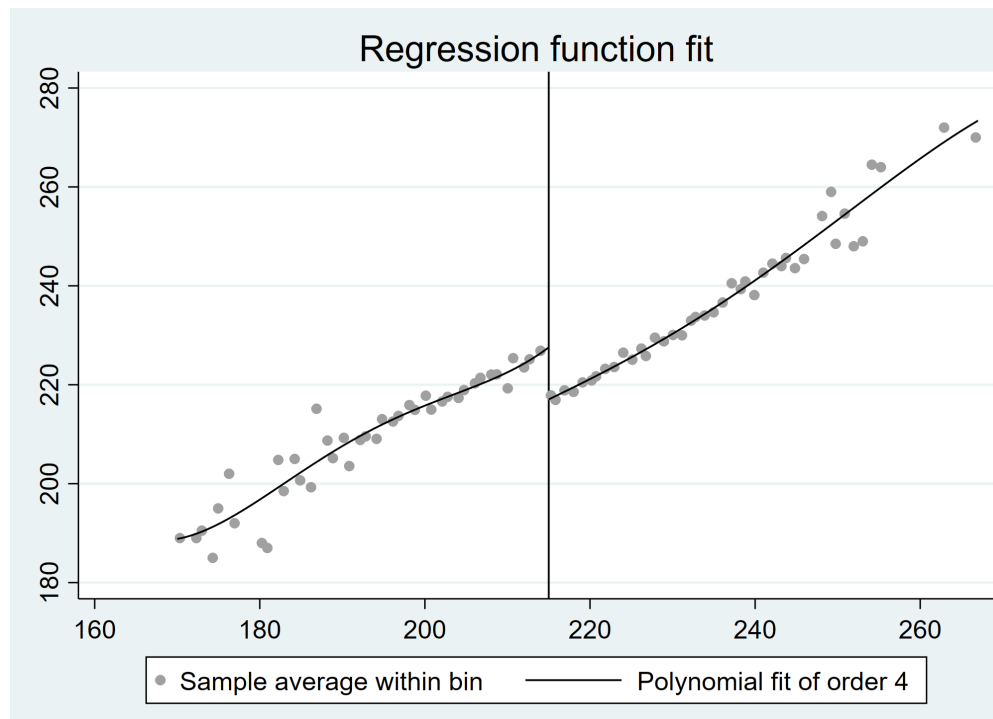
Variable	P-Value ($\Pr(T > t)$)	Interpretación
age	0.9867	No rechaza H0
gender	0.2374	No rechaza H0
esol	0.5789	No rechaza H0
sped	0.4256	No rechaza H0
black	0.3874	No rechaza H0

Ahora, con este ancho de banda se puede, con confianza, evaluar el impacto del programa, con el ejercicio de regresión discontinua.

4. Estimación del efecto

Con los comandos “`rdrobust`” y “`rdplot`” se pueden evaluar estadísticamente y gráficamente lo efecto de lo programa (respectivamente).

En primer lugar, gráficamente se puede ver una diferencia entre los grupos de tratamiento y de control, por ende, el hecho de que una persona haya participado en el programa aumenta la rentabilidad del estudiante. Así que el programa tiene el efecto esperado.



Estadísticamente, sin embargo, la participación en el programa aumenta la rentabilidad de los estudiantes en un 10.28 puntos. (Valor p é significativo a 1%).

Sharp RD estimates using local polynomial regression.

Cutoff $c = 215$	Left of c	Right of c	Number of obs =	2767
			BW type =	mserd
			Kernel =	Triangular
			VCE method =	NN
Number of obs	1295	1472		
Eff. Number of obs	633	757		
Order est. (p)	1	1		
Order bias (q)	2	2		
BW est. (h)	8.672	8.672		
BW bias (b)	14.170	14.170		
rho (h/b)	0.612	0.612		
Unique obs	42	42		

Outcome: math_score. Running variable: poverty_rating.

Method	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Conventional	-10.228	1.0777	-9.4906	0.000	-12.3398	-8.11544
Robust	-	-	-7.9685	0.000	-13.0831	-7.91767

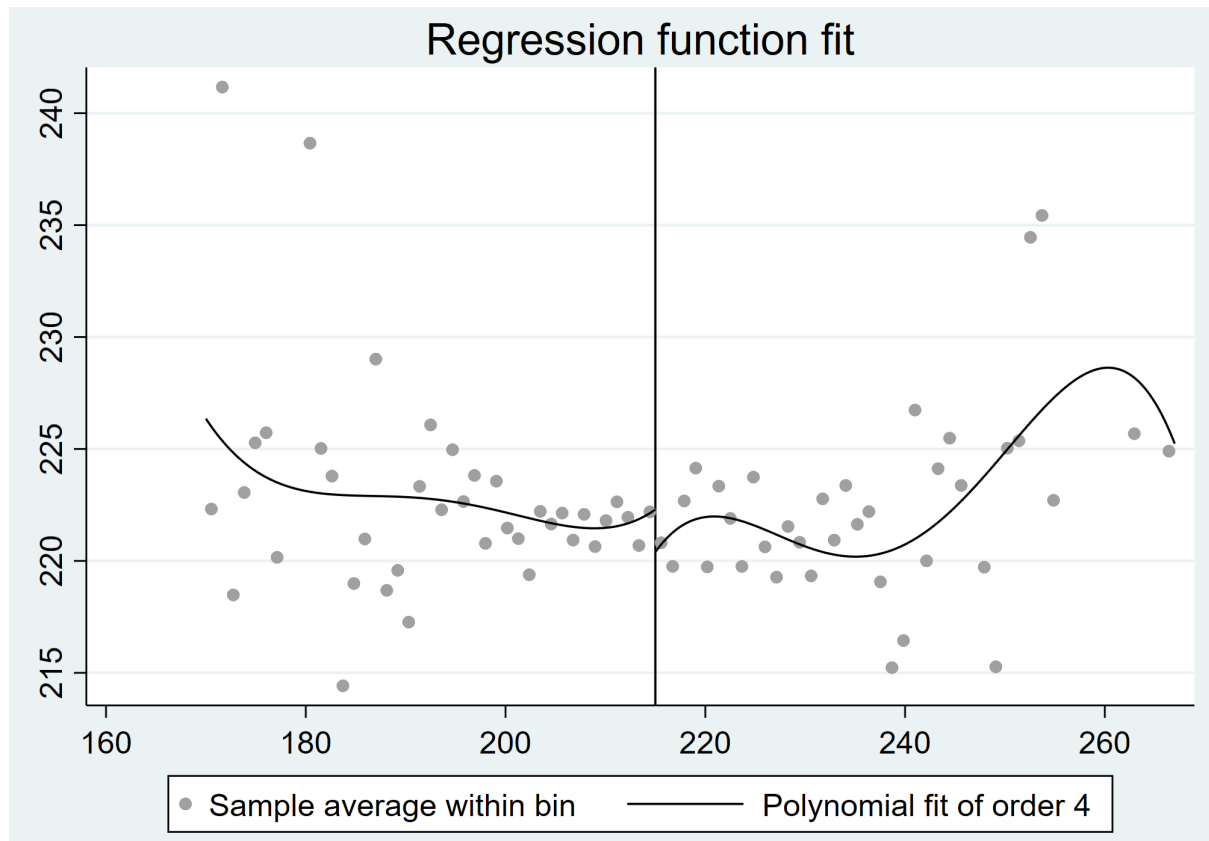
Estimates adjusted for mass points in the running variable.

5. Bônus:

Luego de insertar una semilla, y generar la variable “science score” la cual distribuye normal con media = 222 y sd = 12, ocupamos el mismo código que en la pregunta anterior para estimar si el programa tiene efecto en el rendimiento escolar.

Como se puede observar en el siguiente gráfico, el programa tuvo efectos negativos en el rendimiento escolar, esto se puede observar en el corte (215).

Y también en el coeficiente negativo de la tabla que presentamos.



Sharp RD estimates using local polynomial regression.

Cutoff c = 215	Left of c	Right of c	Number of obs =	2767
			BW type =	mserd
Number of obs	1295	1472	Kernel =	Triangular
Eff. Number of obs	774	908	VCE method =	NN
Order est. (p)	1	1		
Order bias (q)	2	2		
BW est. (h)	10.423	10.423		
BW bias (b)	17.037	17.037		
rho (h/b)	0.612	0.612		
Unique obs	42	42		

Outcome: science_score. Running variable: poverty_rating.

Method	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Conventional	-1.0924	1.2384	-0.8821	0.378	-3.5196	1.33485
Robust	-	-	-0.8843	0.377	-4.26989	1.61488

Estimates adjusted for mass points in the running variable.

