

Evaluación de Impacto

María Camila Caraballo¹

Laura Sarif Rivera²

Taller Regresión Discontinua

En su artículo “Islamic Rule and the Empowerment of the Poor and Pious”, Meyersson (2014) investiga si la llegada al poder por parte del Partido Islámico tiene algún efecto sobre el empoderamiento de las mujeres en Turquía. Para esto, implementa la metodología de Regresión Discontinua, explotando información de: (1) elecciones locales de alcalde en Turquía del año 1994 y (2) mujeres con educación secundaria completa en el año 2000. Concretamente, estima por MCO la siguiente ecuación

$$y_i = \alpha + m_i + f(x_i) + \epsilon_i$$

donde Y_i es la proporción de mujeres entre 15 y 20 años con educación secundaria completa en el año 2000, x_i es el margen de votos con el que ganó o perdió el candidato del partido islámico, $f(\cdot)$ es un polinomio de grado n de la variable x_i , m_i es una dicótoma que toma el valor de uno si $x_i > 0$, es decir, si el alcalde que llegó al poder en 1994 era del partido Islámico y ϵ_i es el término del error. La ecuación es estimada en un vecindario alrededor del corte, el cual, en este caso, es cero.

1 Análisis introductorio

¿Por qué el autor usa la metodología de Regresión Discontinua para identificar los efectos de interés? ¿Cuál es la intuición detrás? ¿Cuál es el supuesto de identificación?

Erik Meyersson (2014) en su trabajo *Islamic Rule and the Empowerment of the Poor and Pious* busca identificar el efecto causal de que un municipio sea gobernado por un alcalde islámico sobre el empoderamiento de las mujeres, medido a través de la proporción de jóvenes entre 15 y 20 años con educación secundaria completa. El desafío metodológico radica en que la elección de un partido islámico no es aleatoria, ya que los municipios que votan por estos partidos pueden diferir en características culturales, religiosas o socioeconómicas que también influyen en los resultados educativos. Si se compararan directamente municipios con y sin alcalde islámico, las estimaciones resultarían sesgadas por esas diferencias estructurales.

La metodología de regresión discontinua ofrece una solución porque permite aislar la variación que puede considerarse casi aleatoria en torno al umbral de victoria electoral. Cuando un candidato gana o pierde por un margen muy pequeño, el resultado depende

¹Código: 201613424

²Código: 202421904

más del azar o de factores impredecibles que de diferencias estructurales entre los municipios. En ese punto de corte, ganar o perder una elección es casi como producto del azar, lo que genera un entorno cuasi experimental en el que las unidades a ambos lados del umbral son muy parecidas. Conceptualmente, los municipios que se ubican justo por encima o por debajo del margen de victoria del partido islámico tienen el mismo perfil social, económico y cultural, por lo que cualquier salto en los resultados educativos observado exactamente en el punto de corte puede interpretarse como consecuencia directa de haber tenido un alcalde islámico.

El supuesto fundamental de identificación es el de continuidad, consiste en que en ausencia del tratamiento el resultado esperado variaría de manera continua con el margen de victoria electoral. Esto implica que no existen saltos sistemáticos en el resultado alrededor del punto de corte que no estén causados por la victoria del partido islamista. Es decir, el supuesto sostiene que cerca del umbral los municipios son similares en todo aspecto salvo en la variable de tratamiento, lo que permite interpretar la discontinuidad observada en los resultados como un efecto causal del triunfo electoral del partido.

2 Para cada género, presenten en una tabla los resultados de estimar las siguientes especificaciones:

a. Ecuación principal, para toda la muestra, sin incluir controles

Table 1: Efecto de alcaldes islámicos sobre educación secundaria

	<i>Dependent variable:</i>	
	Mujeres	Hombres
	(1)	(2)
Alcalde islámico (D)	-0.00001 (0.0001)	0.008 (0.006)
Constant	0.002*** (0.0001)	0.196*** (0.005) (0.005)
Observations	2,498	2,498
R ²	0.014	0.002
Adjusted R ²	0.013	0.001
Residual Std. Error (df = 2494)	0.001	0.077
F Statistic (df = 3; 2494)	12.095***	1.661

Note:

*p<0.1; **p<0.05; ***p<0.01

Errores estándar agrupados por provincia.

b. Ecuación principal, para toda la muestra, con controles.

Table 2: Efecto de alcaldes islámicos sobre educación secundaria

	Variable Dependiente:	
	Mujeres (1)	Hombres (2)
Alcalde islámico (D_i)	0.0001** (0.0001)	0.005 (0.006)
Número de partidos (1994)	-0.0004*** (0.0001)	0.015 (0.014)
Población (log, 1994)	-0.001*** (0.0004)	-0.069** (0.033)
Proporción 0-19 años	-0.00003** (0.00002)	-0.010*** (0.002)
Proporción 60+ años	0.0001** (0.00003)	0.005 (0.003)
Razón de sexo	-0.0001*** (0.00001)	-0.005*** (0.001)
Tamaño promedio del hogar	-0.0001*** (0.00001)	-0.005*** (0.001)
Categoría: Merkezi	-0.00000*** (0.00000)	-0.0004*** (0.0001)
Categoría: Merkezp	0.00005** (0.00002)	0.006*** (0.002)
Categoría: Sub-Büyük	0.001*** (0.0001)	0.047*** (0.005)
Categoría: Büyük	0.0005*** (0.0001)	0.052*** (0.010)
subbuyuk	0.0003 (0.0002)	0.037*** (0.014)
buyuk	0.0004* (0.0002)	0.079*** (0.020)
Constant	0.005*** (0.0004)	0.472*** (0.039)
Estadísticas de Ajuste		
Observaciones	2,498	2,498
R^2	0.438	0.192
Ajustado R^2	0.436	0.187
Error Estándar Residual	0.001	0.070
Estadístico F	149.220***	45.323***

*Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

Errores estándar agrupados por provincia. Todos los modelos incluyen efectos fijos por provincia.

Donde $h=0.24$ es el ancho de banda óptimo estimado por los autores. Para las especificaciones con controles, supongan que:

$$f(x_i) = x_i + x_i \cdot m_i$$

Es decir, un polinomio de grado uno con pendiente distinta a cada lado del corte. Para las especificaciones sin controles, no incluyan ningún polinomio. Todas las especificaciones deben usar errores estándar clúster a nivel de provincia.

c. Ecuación principal, para la submuestra a h unidades alrededor del corte, con controles

Table 3: Regresión Discontinua con controles (submuestra \hat{h} alrededor del corte)

	Variable Dependiente:	
	Mujeres (1)	Hombres (2)
Alcalde islámico (D_i)	0.0003*** (0.0001)	0.010 (0.008)
Número de partidos (1994)	-0.001** (0.0004)	0.004 (0.040)
Población (log, 1994)	-0.002** (0.001)	-0.084 (0.067)
Proporción 0-19 años	-0.00002 (0.00002)	-0.009*** (0.002)
Proporción 60+ años	0.0001** (0.00004)	0.004 (0.004)
Razón de sexo	-0.0001*** (0.00001)	-0.005*** (0.0005)
Tamaño promedio del hogar	-0.00004*** (0.00002)	-0.005*** (0.001)
Categoría: Merkezi	0.00000 (0.00000)	-0.0001 (0.0003)
Categoría: Merkezp	0.00004*** (0.00001)	0.005*** (0.001)
Categoría: Sub-Büyük	0.001*** (0.0001)	0.052*** (0.006)
Categoría: Büyük	0.001*** (0.0001)	0.047*** (0.011)
subbuyuk	0.0002 (0.0002)	0.027 (0.018)
buyuk	0.0004* (0.0002)	0.071*** (0.022)
Constant	0.004*** (0.001)	0.445*** (0.065)
Estadísticas de Ajuste		
Observaciones	968	968
R^2	0.548	0.268
Ajustado R^2	0.542	0.258
Error Estándar Residual	0.001	0.065
Estadístico F	88.888***	26.828***

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Errores estándar agrupados por provincia. Submuestra: $\pm \hat{h}$ unidades alrededor del punto de corte. Incluye controles y efectos fijos por provincia.

d. Ecuación principal, para la submuestra a $h/2$ unidades alrededor del corte, con controles

Table 4: Regresión Discontinua con controles ($\pm \hat{h}/2$ alrededor del corte)

	Variable Dependiente: Proporción con secundaria completa (15–20 años)	
	Mujeres (1)	Hombres (2)
Alcalde islámico (D_i)	0.0003** (0.0001)	0.019* (0.011)
Número de partidos (1994)	−0.002* (0.001)	−0.005 (0.080)
Población (log, 1994)	0.0001 (0.001)	−0.235 (0.155)
Proporción 0–19 años	−0.00002 (0.00002)	−0.008*** (0.002)
Proporción 60+ años	0.0001 (0.00005)	0.006 (0.005)
Razón de sexo	−0.0001*** (0.00001)	−0.005*** (0.001)
Tamaño promedio del hogar	−0.00004* (0.00002)	−0.004** (0.001)
Categoría: Merkezi	0.00000 (0.00000)	−0.0001 (0.0003)
Categoría: Merkezp	0.00004*** (0.00002)	0.004*** (0.002)
Categoría: Sub-Büyük	0.001*** (0.0001)	0.057*** (0.008)
Categoría: Büyük	0.001*** (0.0002)	0.038** (0.016)
subbuyuk	0.0002 (0.0002)	0.007 (0.013)
buyuk	0.0004 (0.0003)	0.034 (0.023)
Constant	0.004*** (0.001)	0.378*** (0.077)
Estadísticas de Ajuste		
Observaciones	560	560
R^2	0.533	0.283
Ajustado R^2	0.522	0.265
Error Estándar Residual	0.001	0.064
Estadístico F	47.980***	16.539***

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Errores estándar agrupados por provincia. Submuestra: $\pm \hat{h}/2$ unidades alrededor del punto de corte. Incluye controles y efectos fijos por provincia.

3 A partir de los resultados encontrados en el anterior punto, respondan:

a. ¿Por qué cambian los coeficientes entre especificaciones?

Los coeficientes cambian entre especificaciones porque cada modelo incorpora distintos supuestos y grados de control sobre posibles problemas de sesgo. En la primera estimación, que no incluye variables de control, el modelo capta únicamente la discontinuidad simple en torno al umbral de victoria electoral. Al introducir controles socioeconómicos y demográficos en las siguientes especificaciones, el modelo mejora la estimación del efecto del alcalde islámico de otras características municipales que también pueden influir en la educación. Asimismo, al restringir la muestra a bandas más estrechas alrededor del punto de corte, se reduce la heterogeneidad entre observaciones, lo que mejora la validez causal pero puede aumentar la varianza de la estimación. Por estas razones, los coeficientes cambian ligeramente entre especificaciones, reflejando cómo la inclusión de controles y la reducción del ancho de banda afectan la precisión y el sesgo potencial del estimador.

b. ¿Cuál parece ser el impacto de la llegada al poder del Partido Islámico para las mujeres? ¿Parece ser este impacto robusto a las especificaciones?

En cuanto al efecto de la llegada al poder del partido islámico sobre la educación de las mujeres, los resultados muestran que este impacto es positivo y estadísticamente significativo en las especificaciones que integran controles y restringen la muestra alrededor del umbral. En particular, en los modelos de regresión discontinua con controles, tanto en la muestra con ancho de banda óptimo como en la reducida a la mitad, el coeficiente asociado a tener un alcalde islámico es positivo y significativo al 1% o 5%. Esto sugiere que la presencia de un gobierno municipal islámico se asocia con una mayor proporción de mujeres jóvenes que completan la educación secundaria.

Finalmente, el hecho de que los coeficientes se mantengan positivos y significativos para las mujeres en las distintas especificaciones indica que el resultado es robusto. A pesar de los cambios en la inclusión de controles o en el tamaño de la muestra, la dirección y magnitud del efecto son consistentes, mientras que para los hombres los coeficientes no son estadísticamente significativos en la mayoría de los casos. Esto refuerza la interpretación de que el impacto de los alcaldes islámicos es específico al grupo femenino y que los resultados no dependen de una especificación particular del modelo, sino que se sostienen bajo diferentes estrategias de estimación.

4. Finalmente, presenten evidencia a favor (o en contra) del supuesto de identificación, para esto:

a. Presenten en una gráfica de la distribución kernel o el histograma de la variable de asignación x_i ¿Parece haber manipulación?

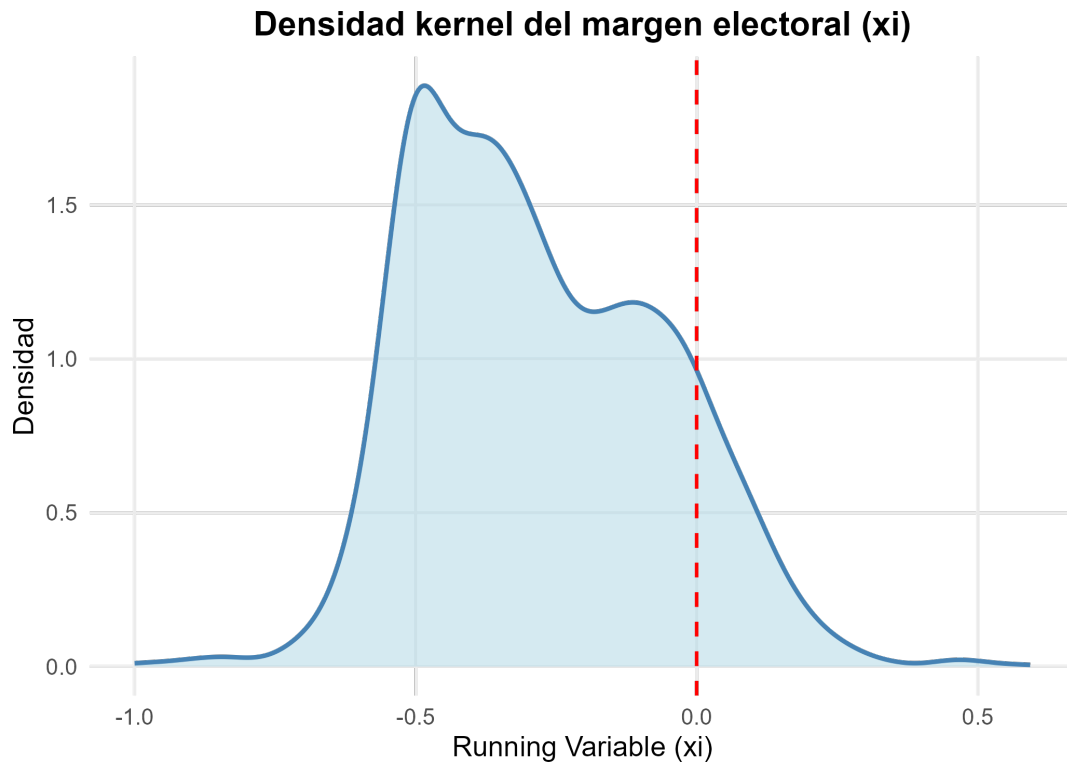


Figure 1: Curva de densidad de la variable de asignación (Kernel).

El gráfico no indica la presencia de manipulación en el umbral de asignación. El supuesto de no manipulación en el diseño de regresión discontinua requiere que la densidad de observaciones en la variable de running sea continua en el punto de corte (que en este caso se marca en el punto 0). La densidad que se observa es suave y no presenta un quiebre, un salto o una caída abrupta en el umbral. Esta continuidad es una evidencia fundamental que sugiere que la probabilidad de ganar o perder por un margen pequeño es prácticamente la misma, lo cual valida la asignación del tratamiento como cuasi-aleatoria.

b. En una tabla presenten los resultados de estimar la ecuación de interés, sin controles, pero incluyendo $f(x_i)$ tomando como variables dependientes: i) la elección de un alcalde del partido Islámico en 1984 (i89) y ii) el logaritmo de la población en 1994 (lpop1994). Para esto, usen únicamente la muestra de elecciones alrededor del ancho de banda óptimo. Dados sus resultados, ¿parece haber continuidad en estas variables?

Table 5: Prueba de continuidad en variables predeterminadas

	Variable Dependiente	
	Alcalde islámico en 1984 (1)	Log población 1994 (2)
Ganó alcalde islámico	-0.005 (0.048)	0.181 (0.183)
Margen electoral	0.375*** (0.117)	-0.947 (0.903)
Interacción margen x ganador	0.064 (0.452)	2.004 (1.904)
Constant	0.090*** (0.020)	8.058*** (0.153)
Estadísticas de Ajuste		
Observaciones	718	968
R^2	0.030	0.004
Ajustado R^2	0.026	0.001
Error Estándar Residual	0.244 ($df = 714$)	1.508 ($df = 964$)
Estadístico F	7.254*** ($df = 3; 714$)	1.254 ($df = 3; 964$)

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

c. Dados sus resultados en los incisos a y b, ¿es plausible el supuesto de identificación? Expliquen por qué.

Los resultados indican una continuidad en las covariables analizadas. Esta prueba permite confirmar que variables que no deberían ser afectadas por el resultado electoral de 1994 no muestran un salto significativo en el punto de corte. La continuidad se verifica al observar que el coeficiente estimado, que captura la discontinuidad en el punto de corte, es estadísticamente indistinguible de cero para las variables analizadas. Específicamente, el valor P amplio para el Alcalde Islámico en 1989 (0.781) y para el Logaritmo de la Población en 1994 (0.866) demuestra que no existe una diferencia significativa en el historial político o el tamaño poblacional entre los municipios que apenas ganaron y los que apenas perdieron la elección de 1994. Esta ausencia de un salto brusco en las covariables antes del tratamiento refuerza la credibilidad de que la asignación del tratamiento es aleatoria en el umbral, lo cual es el supuesto de identificación central del diseño de regresión discontinua.