

Taller 5 - Regresión Discontinua

Maria Camila Caraballo, Laura Sarif Rivera Sanabria

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

$$y_i = \alpha + \beta m_i + f(x_i) + \varepsilon_i$$

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 $x_i \geq 0$, 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.

0. Paso inicial

Para esto se deberá eliminar aleatoriamente el 5% de las observaciones y usar la base restante. La semilla que deben usar para que sus resultados sean replicables es su código de estudiante.

1) ¿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 paper Islamic Rule And The Empowerment of the Poor and Pious busca identificar un efecto causal de las elecciones de partidos islámicos de los municipios en Turquía sobre los resultados de empoderamiento de las mujeres medido como la proporción de jóvenes entre 15 y 20 años con educación secundaria completa. Para ello utiliza la metodología de Regresión Discontinua (RD) aprovechando que la asignación del tratamiento, es decir, si es un alcalde islámico no es aleatoria, dado que los municipios que eligen a este tipo de líderes pueden diferir en factores culturales, religiosos o socioeconómicos, una simple comparación entre municipios con y sin alcalde islámico generaría estimaciones sesgadas. La estrategia de RD permite aproximarse a un experimento natural. La intuición es que los municipios donde el candidato islámico ganó o perdió por un margen pequeño puede ser comparable en todas sus características observables y no observables, de modo que el hecho de haber obtenido la alcaldía o no puede considerarse casi aleatorio.

El supuesto de identificación implica que, en ausencia del tratamiento, es decir, de haber tenido un alcalde del partido islámico, los resultados potenciales, como la proporción de mujeres jóvenes con educación secundaria completa serían continuos en torno al punto de corte del margen electoral. Esto implica que los municipios donde el candidato islámico ganó o perdió por muy poco son, en promedio, comparables en todas las características relevantes que podrían afectar el empoderamiento femenino.

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

- Ecuación principal, para toda la muestra, sin incluir controles.
- Ecuación principal, para toda la muestra, con controles.

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

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

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

$$f(x_i) = \gamma x_i + \delta x_i \times m_i$$

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.

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

a. ¿Por qué cambian los coeficientes entre especificaciones?

Los coeficientes cambian entre los modelos porque al incluir controles adicionales se corrige por diferencias observables entre municipios, reduciendo potenciales sesgos de omisión y permitiendo que la estimación del efecto de tener un alcalde islámico sea más precisa y cercana al efecto causal. Además, al reducir el ancho de banda y enfocarse en municipios más cercanos al umbral, las estimaciones se vuelven más comparables, aunque aumentan los errores estándar por el menor número de observaciones.

El coeficiente para mujeres es negativo y significativo, mientras que con controles cambia a positivo y significativo, lo que sugiere que la estimación inicial estaba sesgada por diferencias preexistentes entre municipios. Para los hombres, el coeficiente con controles es cercano a cero y no significativo, pero aumenta sin controles, aunque sigue sin ser significativo.

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?

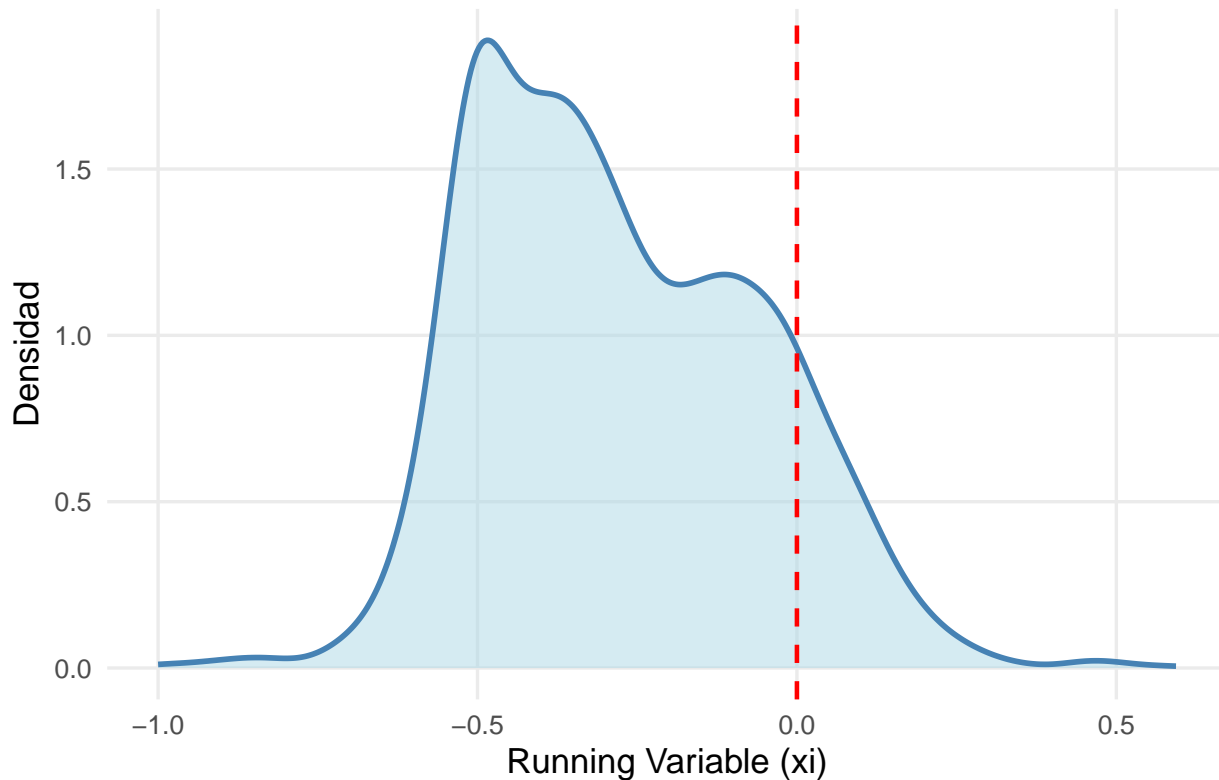
En la estimación sin controles, el coeficiente es negativo y significativo, lo que sugiere inicialmente que la llegada de un alcalde islámico podría reducir la proporción de mujeres con educación secundaria. Sin embargo, al incluir controles, el efecto cambia de signo y se vuelve positivo, lo que indica que la estimación inicial estaba sesgada por características preexistentes de los municipios. En conjunto, la evidencia sugiere que, tras ajustar por esos sesgos, los municipios gobernados por alcaldes islámicos presentan una mayor educación para las mujeres.

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

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

El gráfico 1 muestra la densidad kernel del margen de votos del partido islámico en 1994 (x_i) y evidencia el supuesto de identificación de la metodología RD debido a que en ausencia del tratamiento que es tener un alcalde islámico, los resultados potenciales sobre la educación secundaria en las mujeres, habría sido continua en torno al umbral. Por lo tanto, sugiere que no hubo manipulación electoral en los municipios donde el partido islámico gana o pierde entonces las características son comparables.

Densidad kernel del margen electoral (x_i)



- 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 created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com
 % Date and time: lun, oct. 20, 2025 - 2:32:27 p. m.

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

Los resultados obtenidos respaldan la validez del supuesto de identificación en el diseño de regresión discontinua. La evidencia sugiere que no existe manipulación de la variable running, ya que la distribución de la variable alrededor del umbral se mantiene continua y sin saltos evidentes, de acuerdo con el gráfico de densidad. Así mismo, la elección de un alcalde del partido Islámico en 1984 y el logaritmo de la población en 1994 no son significativos.

Table 1: Regresión Discontinua con Controles: efecto de alcaldes islámicos sobre educación secundaria

	<i>Dependent variable:</i>	
	Proporción con secundaria completa (15–20 años)	hischshr1520m
	Mujeres (1)	Hombres (2)
Alcalde islámico (D)	0.0001** (0.0001)	0.008 (0.005)
Running Variable (X)	−0.0005*** (0.0001)	−0.016 (0.011)
Interacción D · X	−0.001*** (0.0004)	−0.063* (0.032)
Número de partidos (1994)	0.00002 (0.00001)	−0.001 (0.001)
Población (log, 1994)	0.0001** (0.00003)	0.005** (0.002)
Proporción 0–19 años	−0.0001*** (0.00001)	−0.006*** (0.0005)
Proporción 60+ años	−0.00004*** (0.00001)	−0.004*** (0.001)
Razón de sexo	−0.00000*** (0.00000)	−0.0004*** (0.0001)
Tamaño promedio del hogar	0.00005*** (0.00002)	0.004*** (0.001)
Categoría: Merkezi	0.001*** (0.0001)	0.033*** (0.004)
Categoría: Merkezp	0.0004*** (0.0001)	0.024*** (0.009)
Categoría: Sub-Büyük	0.0001 (0.0001)	0.015 (0.009)
Categoría: Büyük	−0.00001 (0.0002)	0.015 (0.017)
Efectos fijos por provincia	−0.0004*** (0.0001)	0.026*** (0.004)
factor(prov_num)3	−0.001*** (0.00002)	−0.030*** (0.002)
factor(prov_num)4	−0.001*** (0.0001)	0.020*** (0.006)
factor(prov_num)5	−0.001*** (0.00003)	−0.025*** (0.003)
factor(prov_num)6	−0.001*** (0.00004)	−0.027*** (0.003)
factor(prov_num)7	4 −0.0001***	−0.009***

Table 2: Regresión Discontinua con Controles (submuestra $\pm\hat{h}$ alrededor del corte)

	<i>Dependent variable:</i>
	Proporción con secundaria completa (15–20 años)
	Mujeres
	(1)
Alcalde islámico (D)	0.0003*** (0.0001)
Running Variable (X)	−0.001*** (0.0004)
Interacción D · X	−0.001 (0.001)
Número de partidos (1994)	0.00002 (0.00002)
Población (log, 1994)	0.0001*** (0.00004)
Proporción 0–19 años	−0.0001*** (0.00001)
Proporción 60+ años	−0.00003 (0.00002)
Razón de sexo	0.00000 (0.00000)
Tamaño promedio del hogar	0.0001*** (0.00001)
Categoría: Merkezi	0.001*** (0.0001)
Categoría: Merkezp	0.0003** (0.0001)
Categoría: Sub-Büyük	0.0001 (0.0001)
Categoría: Büyük	−0.0001 (0.0002)
Efectos fijos por provincia	−0.0001 (0.0001)
factor(prov_num)3	−0.001*** (0.0001)
factor(prov_num)4	−0.0003*** (0.0001)
factor(prov_num)5	−0.0001** (0.0001)
factor(prov_num)6	−0.0004*** (0.0001)
factor(prov_num)7	0.0002*** (0.0001)

Table 3: Regresión Discontinua con Controles (submuestra $\pm\hat{h}/2$ alrededor del corte)

	<i>Dependent variable:</i>
	Proporción con secundaria completa (15–20 años)
	Mujeres
	(1)
Alcalde islámico (D)	0.0003*** (0.0001)
Running Variable (X)	−0.002* (0.001)
Interacción D · X	0.0001 (0.001)
Número de partidos (1994)	−0.00001 (0.00003)
Población (log, 1994)	0.0001** (0.00005)
Proporción 0–19 años	−0.0001*** (0.00001)
Proporción 60+ años	−0.00001 (0.00002)
Razón de sexo	0.00000 (0.00000)
Tamaño promedio del hogar	0.00005*** (0.00002)
Categoría: Merkezi	0.001*** (0.0001)
Categoría: Merkezp	0.0003 (0.0002)
Categoría: Sub-Büyük	0.0001 (0.0002)
Categoría: Büyük	0.00005 (0.0003)
Efectos fijos por provincia	−0.001*** (0.0002)
factor(prov_num)3	−0.001*** (0.0001)
factor(prov_num)4	−0.001*** (0.0002)
factor(prov_num)5	−0.001*** (0.0001)
factor(prov_num)6	−0.001*** (0.0001)
factor(prov_num)7	6 −0.0004***

Table 4: Prueba de continuidad en variables predeterminadas

	<i>Dependent variable:</i>	
	Alcalde islámico en 1984	Log población 1994
	(1)	(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)
Observations	718	968
R ²	0.030	0.004
Adjusted R ²	0.026	0.001
Residual Std. Error	0.244 (df = 714)	1.508 (df = 964)
F Statistic	7.254*** (df = 3; 714)	1.254 (df = 3; 964)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		