

Taller 5

Evaluación de Impacto

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.

En este taller, ustedes realizarán algunas de las estimaciones hechas por el autor e interpretarán los resultados. Para esto, deben usar una submuestra aleatoria de la base de datos *turquia.dta*. Así, justo después de abrir la base de datos – i.e., la base completa, **deben eliminar aleatoriamente el 5% de las observaciones y usar la base restante**.¹ La semilla que deben usar para que sus resultados sean replicables es su código de estudiante.²

Como su solución a este conjunto de problemas, proporcione un **documento PDF** con sus respuestas y el **do file** (u otro programa) que usó para resolverlo. Se recomienda tener respuestas tan breves como sea posible. Ciertas preguntas, sobre todo aquellas que preguntan por su opinión, pueden no tener una única respuesta correcta. Así, lo importante es que argumenten de manera coherente. No seguir estas instrucciones básicas resultará en penalizaciones en su calificación.

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?
2. Para cada género, presenten en una tabla³ los resultados de estimar las siguientes especificaciones
 - a. Ecuación principal, para toda la muestra, sin incluir controles.

¹ Para esto, en STATA pueden crear una variable binaria (i.e. que se distribuya Bernoulli) con la función `rbinomial()` tal que tome el valor de uno si la observación debe ser eliminada y cero de lo contrario. La probabilidad de que tome el valor de uno es, entonces, 0.05. En R pueden usar la función `sample()`.

² Para esto deben usar el comando `set seed` en STATA o la función `set.seed()` en R.

³ Es decir, presenten una tabla para hombres y otra tabla para mujeres. En cada tabla, la variable dependiente es la proporción de hombres o mujeres, según corresponda, entre 15 y 20 años con educación secundaria completa.

- b. Ecuación principal, para toda la muestra, con controles⁴.
- c. Ecuación principal, para la submuestra a \hat{h} unidades alrededor del corte, con controles.
- d. Ecuación principal, para la submuestra a $\hat{h}/2$ unidades alrededor del corte, con controles.

donde $\hat{h} = 0.24$ 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.⁵ Para las especificaciones sin controles, no incluyan ningún polinomio. Todas las especificaciones deben usar errores estándar clúster a nivel de provincia.⁶

3. A partir de los resultados encontrados en el anterior punto, respondan:
 - a. ¿Por qué cambian los coeficientes entre especificaciones?
 - 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?
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 histograma⁷ de la variable de asignación x_i . ¿Parece haber manipulación?
 - 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?
 - c. Dados sus resultados en los incisos a y b, ¿es plausible el supuesto de identificación? Expliquen por qué.

⁴ La lista de controles la pueden encontrar en la nota al pie de la Tabla II, en la página 248 del artículo. Las dicótomas de tipos de municipios corresponden a las siguientes variables: *buyuk*, *merkezi*, *merkezp* y *subbuyuk*.

⁵ Como ejercicio adicional, pueden intentar usar polinomios de mayores grados. Noten que en este caso el corte de la variable es cero, de manera que podemos incluir la variable en la regresión sin problema. Si el corte fuera $c \neq 0$, la variable a incluir en la regresión sería $\tilde{x}_i = x_i - c$.

⁶ Como ejercicio adicional, pueden usar el comando `rdplot` en STATA o la función `rdplot()` de R para representar el efecto de forma gráfica.

⁷ Para eso pueden usar los comandos `histogram` o `twoway kdensity` en STATA. En R pueden usar la función `hist()` para hacer el histograma o las funciones `density()` y `plot()` para la distribución *kernel*. Para más información, pueden revisar el siguiente [enlace](#). En caso de que decidan presentar un histograma, es importante que tenga bastantes barras. Siguiendo a los autores, recomendamos usar 99 barras o *bins*.