

## Problem Set 10: Regression discontinuity

---

**Fecha de entrega:** 8 de noviembre de 2023

Gastón García Zavaleta y Tomás Pacheco

### Instrucciones

- Agreguen los nombres y apellidos de los integrantes en el encabezado del archivo que entregan.
- El título de los archivos entregados debe tener la siguiente estructura:  
PS10\_Garcia\_Gomez\_Gonzalez\_Perez.pdf
- **UN ÚNICO** miembro del grupo debe subir el PDF y *do-file* al Campus Virtual en la sección "Problem Set 10".
- El código lo tendrán que entregar de dos formas distintas. Primero deberán exportar el *do-file* como PDF y unirlo al informe ([ayuda](#)). Luego, deberán entregar el código en formato .do en el lugar indicado en el Campus Virtual.
- El directorio de trabajo debe tener tres carpetas: programs, input y output. El *do-file* debe ser entregado de tal manera que los tutores podamos correr el código **solamente** modificando el global con el *main directory* (**esto es muy importante**).
- Recuerden que siempre deben presentar la información en tablas y gráficos estéticos. Nunca presenten una tabla o figura sin dedicarle unas líneas a comentar lo que se observa.

Para este problem set vamos a usar *close elections* como fuente de variabilidad. La idea es que cuando alguien gana una elección por muy pocos votos, esto podría considerarse aleatorio. Queremos ver si el hecho de que gobierne un partido demócrata hace que el gasto (*lne*) sea mayor o menor. Los datos corresponden a una elección en el congreso de Estados Unidos.

1. La variable *vote\_share\_democrats* contiene la proporción de votos que ha obtenido el partido demócrata. Generen una variable, *demwon*, en la que el cutoff (50 %) sea 0.
2. Grafiquen el tradicional gráfico de RD utilizando un polinomio de grado 1 y un polinomio de grado 2. Comenten. ¿Cuál ajusta mejor? Elijan el grado con el que trabajarán después.
3. Lleven a cabo los dos primeros falsification tests que vimos en la tutorial. Guarden las siguientes variables de control en un global llamado *covs*: *unemployd* *union* *urban* *veterans*. Comenten los resultados. Nota: para el primer test, no solo muestren los gráficos, sino que también muestren una tabla con los pvalores. Pueden usar el código que se encuentra en la carpeta para generarla<sup>1</sup>.
4. Estimen el modelo sin utilizar las variables de control. Comenten los resultados. ¿Cuál es el bandwidth óptimo? Repitan lo mismo agregando las variables de control. ¿Se mantienen los resultados?
5. Utilizando la especificación con variables de control, estimen tres veces más el RD utilizando como bandwidths 0,025; 0,15 y 0,25. ¿Qué resultados esperarían ver? ¿Qué trade-off hay en esta elección?
6. Si el cutoff fuera en 40 % en vez de 50 %, ¿qué esperarían ver? ¿y en 60 %? ¿Los resultados se condicen con lo que esperarían ver?
7. Si utilizan el enfoque de randomización local utilizando un kernel triangular, ¿los resultados se mantienen? ¿Cuál es la ventana óptima? Nota: usen la seed 444.

---

<sup>1</sup>La idea es que vean cómo se pueden utilizar matrices en Stata para generar tablas y luego exportarlas.