

Tarea 5

Joao Garcia

Fecha de entrega: 2025-07-02

Ejercicio 1: El Efecto Causal de la Titularidad en las Elecciones de EE. UU.

Esta guía de ejercicios lo guiará a través de un análisis completo de Diseño de Regresión Discontinua (RDD). Replicaremos los hallazgos centrales del influyente artículo de David Lee (2008), “Randomized experiments from non-random selection in U.S. House elections”.

a)

Una pregunta central en la ciencia política es si ocupar un cargo le da a un partido político una ventaja en elecciones futuras. Esto se conoce como la “ventaja de la titularidad” (incumbency advantage). Explique por qué la simple comparación del margen de votación (o la probabilidad de victoria) entre un candidato titular (es decir, uno que ganó la última elección y está actualmente en el poder) y un retador es poco probable que resulte en efectos causales. ¿Qué podría causar un sesgo de variable omitida? ¿Espera que esta comparación subestime o sobreestime la verdadera ventaja de la titularidad?

b)

Lee (2008) ideó una forma ingeniosa de aislar el efecto causal de ganar una elección. Se centró en elecciones que fueron muy reñidas. La idea es que en una elección decidida por un puñado de votos, si ganó el candidato demócrata o el republicano es esencialmente aleatorio. Al comparar los resultados de la siguiente elección para los distritos donde un demócrata ganó por un margen muy estrecho con aquellos donde un demócrata perdió por un margen muy estrecho, podemos estimar el efecto causal de la titularidad.

Datos: Usaremos el conjunto de datos `lee2008`, que está disponible públicamente en el paquete `RATest` de R. Puede cargarlo ejecutando el siguiente código:

```
# install.packages("RATest") # instalar solo una vez

data <- RATest::lee2008
```

Las variables clave son:

difdemshare: El margen de victoria en la proporción de votos demócratas en la elección t . Un valor positivo significa que el demócrata ganó, y un valor negativo significa que el demócrata perdió. Un valor de 0 significa una elección empatada (lo que en realidad no ocurre). **demsharenext:** La proporción de votos del partido Demócrata en la siguiente elección, $t+1$. **demshareprev:** La proporción de votos del partido Demócrata en la elección *anterior*, $t-1$. **demofficeexp:** La experiencia política del candidato demócrata, medida en años en el cargo. **othofficeexp:** La experiencia política del candidato oponente, medida en años en el cargo.

Antes de comenzar, identifique y explique los elementos clave de la estrategia RDD:

- i) La variable de puntaje (running variable)
- ii) El punto de corte (cutoff)
- iii) El tratamiento

- iv) La variable de resultado de interés
- v) Los supuestos clave
- vi) ¿Es este un RD nítido (sharp) o difuso (fuzzy)?

c)

Grafique los datos usando un diagrama de dispersión del resultado promedio por intervalos (bins) de la variable de corrida. (Pista: puede usar el código que vimos en clase para ayudarse). Basado en la figura, ¿qué concluye?

d)

Estimación paramétrica. Use regresión lineal para estimar el efecto de la titularidad. Estime y muestre los resultados de 4 modelos: i) un ajuste lineal simple con una pendiente diferente a cada lado del punto de corte, ii) un ajuste cuadrático a cada lado del punto de corte, iii) como en i, pero con ponderaciones mayores cerca del punto de corte, iv) como en ii, pero con un ancho de banda (bandwidth) más estrecho de su elección. Comente la variación que observa en el efecto estimado.

e)

Elija su modelo favorito de los estimados en d. Interprete y explique cada coeficiente y grafique la figura correspondiente.

f)

Use el paquete `rdrobust` para estimar el efecto causal de la titularidad e interprete los resultados. ¿Son consistentes con su elección en e?

g)

Use el comando `rddensity` para realizar un test de McCrary. Interprete los resultados y cree un gráfico. ¿Por qué es importante este test?

h)

Basado en su modelo de elección en e, estime el efecto de la titularidad en `demshareprev`. ¿Qué espera encontrar y qué muestran los resultados? ¿Por qué es una prueba importante?

Ejercicio 2: El Efecto Causal de la Disponibilidad de Cuidado Infantil

En Chile, por ley, las empresas con 20 o más empleadas mujeres están obligadas a proporcionar cuidado infantil gratuito para sus hijos hasta los 2 años. Un investigador, preocupado por la baja tasa de fertilidad de Chile, está interesado en cómo esto afecta los salarios de las mujeres y la probabilidad de que abandonen la fuerza laboral después de dar a luz.

El investigador propone usar el número de empleadas como la variable de corrida en un diseño RD. Al comparar empresas con 19 mujeres empleadas y aquellas con 20, argumenta que puede encontrar el impacto causal de la provisión de cuidado infantil.

Piense cuidadosamente en los supuestos de este diseño.

a)

¿Qué supuesto es más probable que no se cumpla en este estudio?

b)

Si este supuesto se viola, ¿cómo causaría un sesgo? Explique.

c)

¿Cómo se puede probar si el supuesto se cumple o no?

Ejercicio 3: Subsidio eléctrico para hogares de bajos ingresos

Un programa de subsidio reduce la tarifa eléctrica a hogares cuyo ingreso per cápita Y está por debajo de un salario mínimo mensual. No obstante, la verificación documental es imperfecta y existen hogares elegibles que no reciben el beneficio, mientras que otros no elegibles lo obtienen. Nos interesa saber como ese subsidio afecta el consumo de energía.

a)

Identifique y explique los elementos clave de la estrategia RDD:

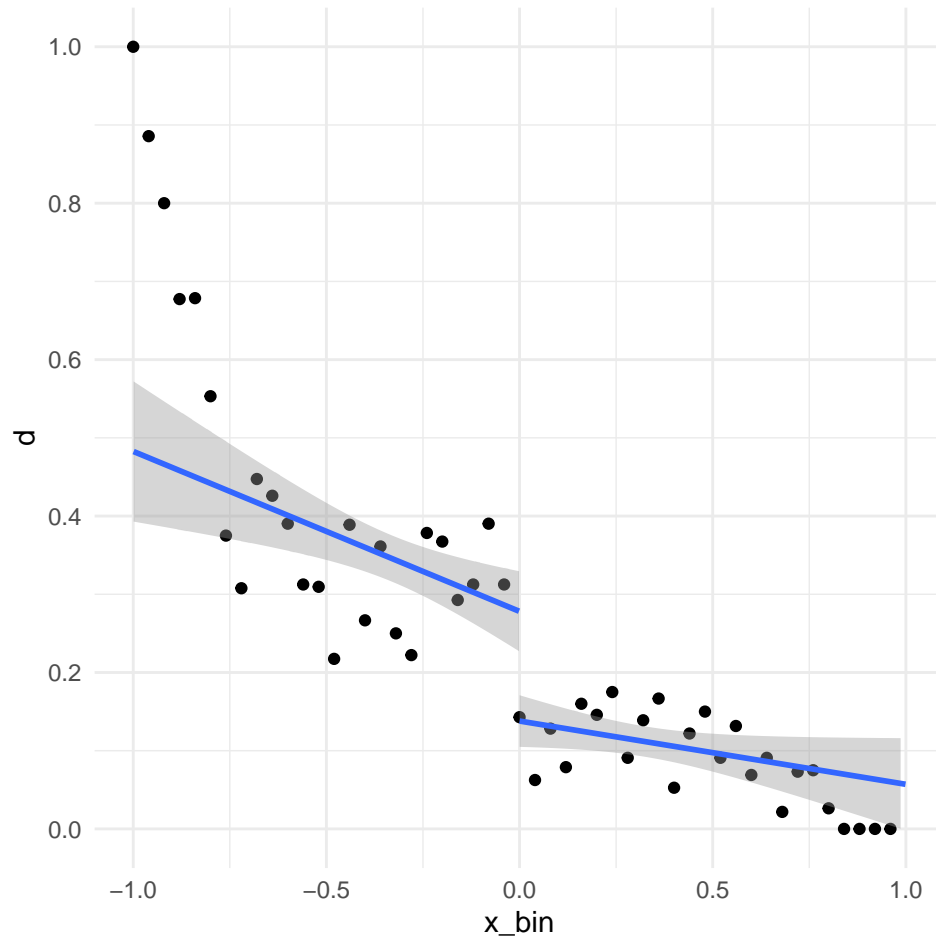
- i) La variable de puntaje (running variable)
- ii) El punto de corte (cutoff)
- iii) El tratamiento
- iv) La variable de resultado de interés
- v) Los supuestos clave
- vi) ¿Es este un RD nítido (sharp) o difuso (fuzzy)?

b)

Un investigador está tratando de estimar el efecto de estar por debajo del umbral en la recepción del subsidio. Propone dos modelos diferentes: uno con un ajuste lineal y otro con un ajuste cúbico. Basado en las estimaciones y las figuras a continuación, ¿cuál elegiría? Explique cómo eligió e interprete los coeficientes en su modelo de elección.

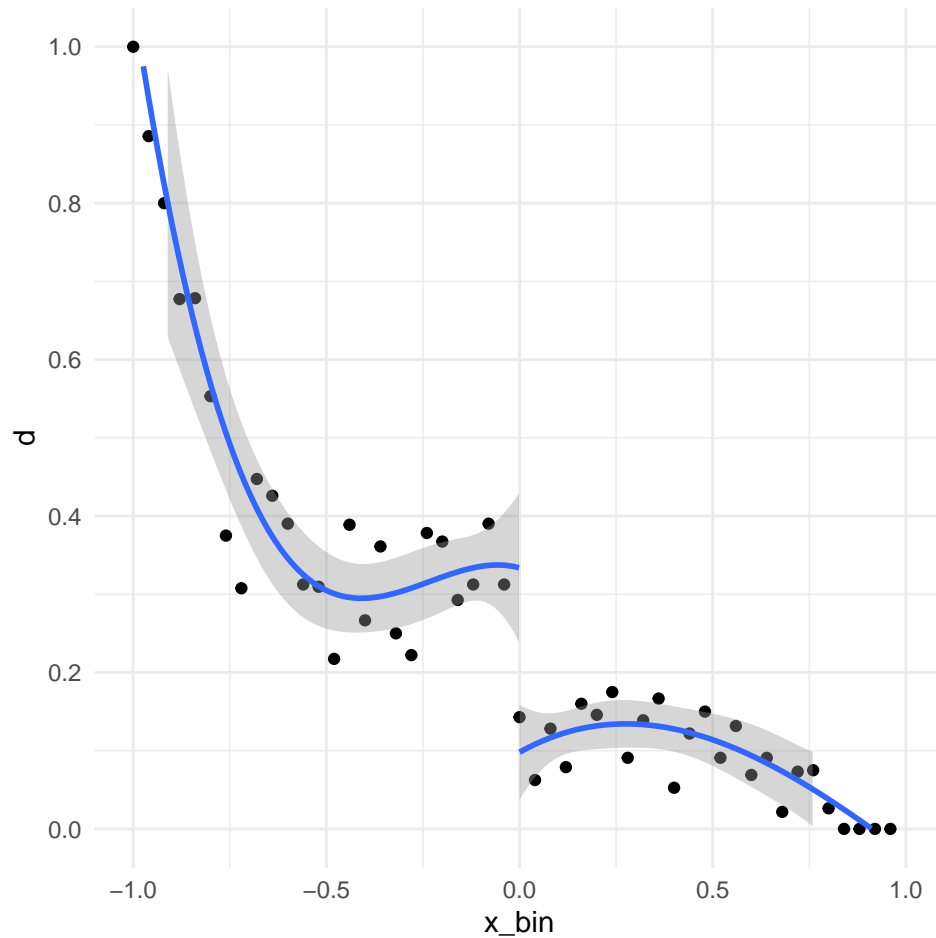
En estos resultados, `ingresos_umbral` corresponde al ingreso del hogar menos el umbral (normalizado para quedar dentre -1 y 1), `eligible` es una dummy igual a 1 si el hogar está *por debajo* del umbral, y `recibe_subsidio` es una dummy igual a 1 si el hogar recibe el beneficio.

Modelo lineal



```
##
## Call:
## lm_robust(formula = recibe_subsidio ~ ingresos_umbral * eligible,
##           data = data_raw, weights = w)
##
## Weighted, Standard error type: HC2
##
## Coefficients:
##               Estimate Std. Error t value      Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)      0.1379   0.0215     6.41 0.000000000187  0.0957  0.180145 1996
## ingresos_umbral  -0.0809   0.0410    -1.97 0.048851562839 -0.1614 -0.000408 1996
## eligible         0.1402   0.0393     3.57 0.000367435703  0.0631  0.217221 1996
## ingresos_umbral:eligible -0.1237  0.0787    -1.57 0.116080530788 -0.2780  0.030609 1996
##
## Multiple R-squared:  0.0854 ,    Adjusted R-squared:  0.084
## F-statistic: 69.1 on 3 and 1996 DF,  p-value: <2e-16
```

Modelo cúbico



```
##
## Call:
## lm_robust(formula = recibe_subsidio ~ ingresos_umbral * eligible +
##           I(ingresos_umbral^2) * eligible + I(ingresos_umbral^3) *
##           eligible, data = data_raw, weights = w)
##
## Weighted, Standard error type: HC2
##
## Coefficients:
##
##               Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)      0.098    0.0423   2.316  0.02068   0.0150   0.181 1992
## ingresos_umbral    0.277    0.4193   0.660  0.50903  -0.5453   1.099 1992
## eligible          0.236    0.0777   3.034  0.00244   0.0834   0.388 1992
## I(ingresos_umbral^2) -0.573    1.0514  -0.545  0.58587  -2.6348   1.489 1992
## I(ingresos_umbral^3)  0.166    0.7311   0.227  0.82038  -1.2678   1.600 1992
## ingresos_umbral:eligible -0.414    0.7398  -0.559  0.57618  -1.8645   1.037 1992
## eligible:I(ingresos_umbral^2) -0.792    1.8403  -0.430  0.66708  -4.4008   2.817 1992
## eligible:I(ingresos_umbral^3) -2.118    1.2807  -1.654  0.09837  -4.6293   0.394 1992
##
## Multiple R-squared:  0.1 , Adjusted R-squared:  0.0971
## F-statistic: 74.5 on 7 and 1992 DF, p-value: <2e-16
```

c)

Usando el mismo método que su elección en **b**, usted encuentra que estar por debajo del umbral aumenta el consumo de electricidad en 4 mil pesos por mes (a precios normales). ¿Cuál es su estimación del efecto de recibir el subsidio sobre el consumo de electricidad?