

IN4143 – Análisis de Datos e Inferencia Causal

Primavera 2024

Profesores: Omar Pérez, Daniel Schwartz

Auxiliares: Antonia Aceituno, Camila Galarce

Ayudantes: Guillermo Escobar, Bastián Medina

LABORATORIO 4

Entrega: Subir respuestas en U-Cursos hasta las 17:45 hrs en un archivo pdf.

Esta evaluación es individual. Para responder puede revisar el material del curso disponible en u-cursos y las cápsulas de clase.

El Ministerio de Economía está evaluando un programa empresarial de fomento a la Investigación y Desarrollo (I+D), de forma que puedan aumentar su capital de inversión. Este subsidio sólo se entregaría a las empresas que formulen y presenten los mejores proyectos de I+D del concurso; específicamente que obtengan un puntaje de postulación igual o superior a 75 puntos (solamente).

La base de datos del programa contiene la información de múltiples firmas que concursaron al subsidio. Se cuenta con información acerca de:

- **puntaje** → puntaje obtenido en el concurso
- **tratamiento** → si es que se adjudicaron el subsidio
- **inversión** → inversión realizada por todas las empresas que concursaron, posterior al periodo del concurso.
- **size** → cantidad de trabajadores de la empresa

Con esta información, se le pide a usted estimar el efecto que tuvo adjudicarse el subsidio para I+D sobre el capital de inversión de las firmas, mediante un modelo de regresión discontinua.

Utilizando la base de datos adjunta “Lab4_RD” responda las siguientes preguntas.

1) ¿Qué supuestos se deben cumplir para que el modelo de RDD esté identificado? Sólo menciónelos (4 supuestos) ¿son todos comprobables? (indique cuál(es) no). [4pts]

2) Determine a partir de gráficos si se cumplen los supuestos (e incluya los gráficos)

- I. ¿Es un diseño sharp o fuzzy (nítido o difuso)? Construya un gráfico entre la variable de tratamiento y el puntaje (running variable), además ilustre el puntaje de corte con una línea vertical. (no centre aún la variable running en torno a cero). [4pts]
- II. ¿Existe evidencia para sospechar que las empresas pueden manipular su variable running? Realice un gráfico de densidad de la variable running, indicando el punto de corte con una línea vertical. (no centre aún la variable running entorno a cero) [4pts]
- III. **Bonus:** ¿Existe una discontinuidad en la variable dependiente en torno al puntaje de corte (para la running variable)? Genere un gráfico de dispersión (scatterplot), incluyendo una

línea vertical en el punto de corte, y dos ajustes lineales (uno a la izquierda del punto de corte y otro a la derecha [ganadores del concurso vs. no ganadores]). Visualmente ¿pareciera haber un resultado del tratamiento (ganarse el concurso sobre la inversión)? (suponga que se cumplen todos los supuestos) Para este punto (III.) y los siguientes debe centrar la variable running en cero. [4pts]

IV. Las empresas que se ganaron el concurso ¿son diferentes en “cantidad de trabajadores” a las que no se lo ganaron? Para comprobar esto construya dos gráficos entre la variable size y la variable puntaje ajustando dos modelos; uno ajustando un polinomio de grado 1 (reg. lineal entre size y puntaje); y otro gráfico ajustando un polinomio de grado 4. Cabe destacar que en cada gráfico es necesario ajustar el modelo por separado al grupo control y al grupo tratamiento (una curva distinta para cada grupo). (centre la variable running entorno a cero).[6pts]

V. Con la evidencia encontrada hasta este punto ¿se cumplen todos los supuestos para aplicar RDD? [2pts]

3) Regresión Discontinua: Elabore una tabla de síntesis con las siguientes cinco regresiones lineales [6pts]:

I. Efecto del tratamiento sobre la variable de inversión, controlando por el puntaje (dada la evidencia de 3 debe determinar si incluir o no la variable “size” [cantidad de trabajadores] como control).

- a. Polinomio de orden 1
- b. Polinomio de orden 3

II. Repita I.b. (polinomio de orden 3) pero para distintos tamaños de “ventana.”

- a. Vecindad de +/- 20 puntos en torno al puntaje de corte
- b. Vecindad de +/- 10 puntos en torno al puntaje de corte
- c. Vecindad de +/- 5 puntos en torno al puntaje de corte

***** Recuerde mostrar todos los resultados (de a y b) en una sola tabla (se sugiere utilizar stargazer)**

4) Conclusiones: A partir de la tabla estimada en el ítem 3): ¿Son robustos los resultados encontrados ante distintas especificaciones? ¿Se puede afirmar algo sobre el efecto del tratamiento (positivo, negativo, no concluyente)? [2pts]

5) **Bonus:** En el contexto del subsidio para fomentar la inversión en I+D, reflexione sobre cómo podría implementarse un enfoque de matching para estimar el efecto del tratamiento, explique qué supuestos se deben cumplir y qué tipo de matching usaría. Justifique [5pts].