

**IN4143**  
**Análisis de datos e inferencia causal**  
**(Parte 2)**

Profesor: Omar D. Perez  
Semestre: Otoño 2024

**LABORATORIO 3**  
**(Variables Instrumentales)**

**Entrega:** Subir respuestas a U-Cursos hasta las 23:59 hrs en un documento PDF tipo informe. No se aceptarán entregas posteriores y se aplicarán descuentos en caso de enviar el documento en otro formato. No es necesario adjuntar códigos.

El objetivo de este laboratorio es que usted sea capaz de definir los conceptos vistos en clase asociados al método de variables instrumentales, y poder implementarlo computacionalmente. Además, se espera que usted sea capaz de analizar críticamente la validez de los supuestos del diseño presentado.

Para responder puede revisar el material del curso en U-Cursos, googlear, preguntarle a chat GPT o buscar información donde estime conveniente; es la mejor manera de aprender.

**Día del asfalto**

Para el día del asfalto, el gremio internacional de reparadores de calles decide secretamente realizar una conmemoración, modernizando las calles de 31 megaciudades del mundo; un día lunes a las 7:30 de la mañana de cada ciudad. Esto generó un incremento en el tiempo de recorrido de millones de ciudadanos, provocando que varios llegasen atrasados a sus lugares de trabajo. Para elegir qué espacios modernizar, el gremio usó un algoritmo que eligió cuadras aleatoriamente.

Después de unos días, un conjunto de analistas de datos que trabajan en una compañía con múltiples sedes en cada megaciudad, accedió a los datos de movilización de cada uno de sus trabajadores durante el día del asfalto. Además, accedieron a un conjunto de datos de productividad de ese lunes, y algunas variables invariantes en el tiempo por persona, asociadas a productividad.

La persona que lidera este equipo de investigación cursó la asignatura IN4143, por lo que en vez de lamentarse por la terrible situación ve una oportunidad de usar lo que aprendió en el curso, usando el método de variables instrumentales para identificar si los atrasos en la hora de llegada tienen un efecto en la productividad de sus funcionarios

**Utilizando la base de datos adjunta y el glosario de variables expuesto en ANEXOS, responda las siguientes preguntas.**

1. Indique el efecto pedido (y errores estándar en los casos que se pueda) para los siguientes casos, asumiendo que  $X$  es la variable endógena (atraso), y preséntelos en

una tabla como la que se muestra en Anexos (Tabla 1). En todas las regresiones la variable dependiente es Y (productividad por persona).

- i. OLS sin instrumentalizar (sin controles) [regresión con endogeneidad].
- ii. ITT (sin controles).
- iii. LATE a partir de ITT e ITTD (sin controles).
- iv. LATE mediante dos regresiones OLS (sin controles).
- v. LATE utilizando package estadístico para 2SLS<sup>1</sup> (sin controles).
- vi. LATE utilizando package estadístico para 2SLS<sup>2</sup> (controlando por Q1, Q2 y Q3).

2. A partir de los resultados obtenidos en 1., responda brevemente:

- a) ¿Se observa evidencia de endogeneidad en la estimación de i.? En caso de que exista, explique..
- b) En relación a los errores estándar asociados a la variable endógena de las regresiones iv., v. y vi., ¿son diferentes? ¿por qué?
- c) Compare los coeficientes asociados a la variable endógena de i., iii., iv., v. y vi. ¿Son diferentes? ¿por qué?

3. En relación a los supuestos requeridos para que la estimación de LATE sea válida (mediante IV), compruebe si se cumplen.

4. Olvidándose de los coeficientes estimados en las etapas anteriores. Asuma que puede contactarse con sus versiones de 999 universos paralelos, en los que se realizó la conmemoración a la misma población ante condiciones externas casi idénticas. En cada uno de estos multiversos una versión de usted calculó el coeficiente de interés mediante OLS (regresión endógena) y mediante 2SLS. Luego de obtener los resultados ilustró en histogramas los coeficientes estimados, uno para OLS y otro para 2SLS, y obtuvo un gráfico, adjunto en la sección de Anexos (Ilustración 1).

Mirando el gráfico (Ilustración 1) ¿Cuál cree usted que es el histograma asociado a las estimaciones de 2SLS? Justifique. (*Hint: Puede aludir a conceptos como varianza del estimador y error estándar*)

---

<sup>1</sup> Se sugiere utilizar el comando "ivreg" del package "ivreg"

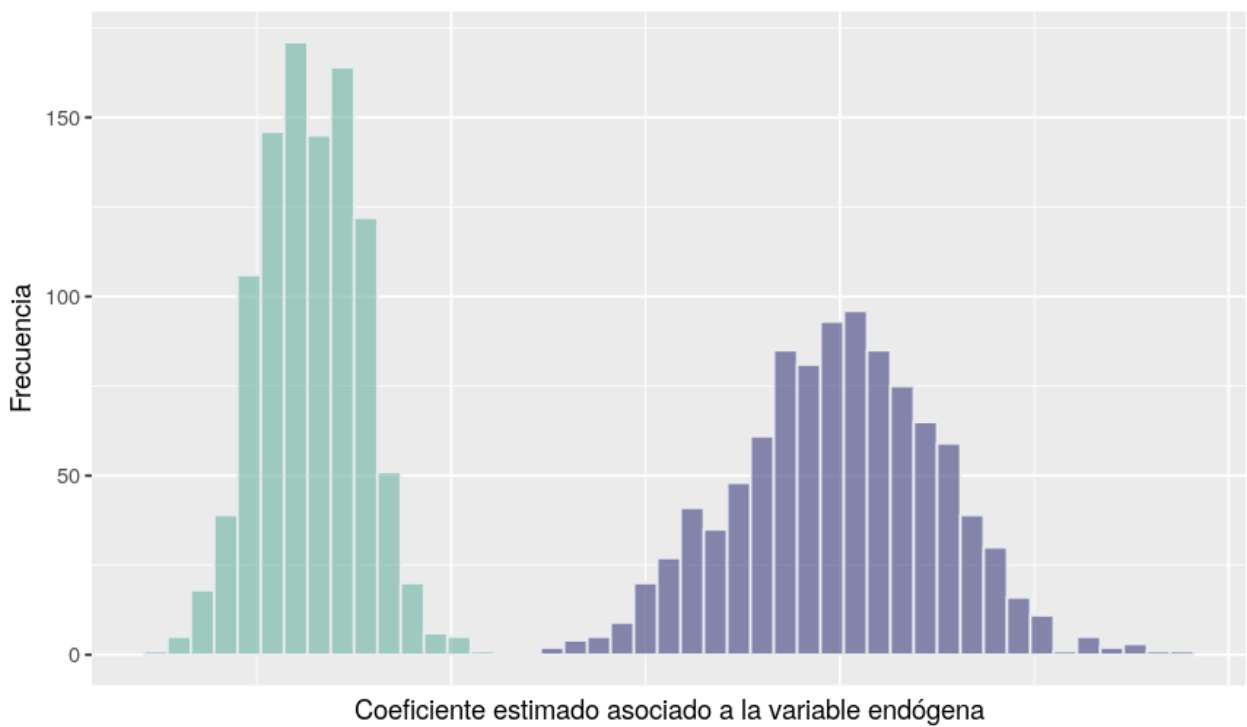
<sup>2</sup> Se sugiere utilizar el comando "ivreg" del package "ivreg"

## ANEXOS

Tabla 1: Formato tabla sugerida resultados pregunta 1

|                | OLS (Sin IV) | ITT         | A partir de<br>ITT e ITTD | Dos regresiones | 2SLS - IV<br>(sin controles) | 2SLS - IV<br>(con controles) |
|----------------|--------------|-------------|---------------------------|-----------------|------------------------------|------------------------------|
| Coeficiente    | <i>BETA</i>  | <i>BETA</i> | <i>BETA</i>               | <i>BETA</i>     | <i>BETA</i>                  | <i>BETA</i>                  |
| Error estándar | (EE)         | (EE)        |                           | (EE)            | (EE)                         | (EE)                         |

Ilustración 1: Estimación de regresiones OLS (endógena) y IV (LATE) en mil universos paralelos



### Glosario

- Z: Variable binaria que indica si se realizó alguna reparación en el camino que el trabajador o trabajadora utiliza para llegar a su trabajo
- X: Variable binaria que indica si el trabajador o trabajadora llegó atrasado el día del asfalto.
- Q1, Q2 y Q3: Variables por persona asociadas a la productividad.
- Y: Medida de productividad por persona.