

Taller 1. Econometría Avanzada

Daniel Ricardo Amaya Alba

Table of contents

1	Primer Ejercicio	1
1.1	Punto1	2
1.2	Punto2	4
1.3	Punto3	6
1.4	Punto4	10

1 Primer Ejercicio

Desde la perspectiva del capital humano, estudios como el de Cunha & Heckma (2007) encuentran que las intervenciones durante los primeros años de escolaridad son fundamentales para el desarrollo temprano de habilidades y, con ello, para la trayectoria de aprendizaje posterior de los individuos. Una forma de evaluar el éxito de estas intervenciones es mediante el nivel de habilidades cognitivas que desarrollan los estudiantes, dado que estas capacidades constituyen la base de cualquier aprendizaje futuro.

En este contexto, los gobiernos suelen mostrar interés por los programas de digitalización, ya que prometen acelerar el desarrollo del razonamiento y facilitar la adaptación al mundo moderno. Con base en lo anterior, la pregunta de investigación que guiará este ejercicio, inspirada en Cristia et al. (2012), es: ¿cuál es el efecto de participar en un programa de tecnología educativa sobre el desarrollo de habilidades cognitivas de razonamiento abstracto?

Para responder esta pregunta, usted analizará una política implementada por la Secretaría de Educación de Macondo en conjunto con el colegio Sonrisas, ubicado en una zona rural del país. La iniciativa buscaba transformar el modelo educativo tradicional mediante la entrega de laptops diseñadas específicamente para el aprendizaje autodidacta en entornos con limitaciones severas de infraestructura. El programa se enfocó exclusivamente en estudiantes de tercer grado de primaria, nivel que en este colegio cuenta con dos salones. El programa seleccionaba un salón para convertirlo en Aula Tecnológica.

Los estudiantes asignados al Aula Tecnológica realizaban actividades guiadas tres veces por semana utilizando las laptops. Estas actividades se desarrollaban únicamente durante el horario de clases, las cuales fueron adaptadas para incluir ejercicios lógicos y juegos de asociación visual orientados a fortalecer el razonamiento abstracto. Cada actividad estaba vinculada con el tema correspondiente a la clase del día. Adicionalmente, los estudiantes no podían llevar las laptops a sus hogares. Los estudiantes del otro salón, por su parte, continuaron con sus clases de manera habitual.

Considere una muestra de N estudiantes observados en un diseño de corte transversal. Defina Y_i como el puntaje del estudiante i en un test de habilidades cognitivas de razonamiento abstracto. Este test se aplica anualmente a nivel nacional a estudiantes de tercer grado, está definido por el Ministerio de Educación y sus puntajes oscilan entre 0 y 100. Por otro lado, D_i es una variable dicotómica que toma el valor de 1 si el estudiante i fue asignado al Aula Tecnológica y 0 en caso contrario.

Para comenzar, suponga que el mecanismo de asignación al tratamiento es desconocido; es decir, no se conocen los criterios que determinaron la elección de un salón sobre el otro ni la forma en que los estudiantes son asignados a cada salón.

1.1 Punto1

Describe el problema de estimación utilizando el lenguaje de resultados potenciales

a) Teniendo en cuenta el contexto presentado, defina formalmente los resultados potenciales $Y_i(1)$ y $Y_i(0)$. Luego, describa el problema de inferencia causal en este contexto.

R/: En este caso el resultado potencial $Y_i(1)$ se define como el puntaje que obtendría el estudiante i dado que fue asignado al aula tecnologica. Por su parte el resultado potencial $Y_i(0)$ es el resultado que obtendría el estudiante i dado que no fue asignado al aula tecnologica.

Para recuperar el efecto causal de la participación en las aulas tecnologicas en el puntaje obtenido por el estudiante i , nos gustaria ver la dferencia en sus resultados potenciales, es decir de sus puntajes dado que participio en el aula y que no participo en el aula:

$$\tau_i = Y_i(1) - Y_i(0)$$

Sin embargo, acá nos encontramos con el problema de la inferencia causal. En nuestra muestra N para el estudiante i solo se puede observar uno de los dos caminos, o el estudiante fue asignado ($D_i = 1$) o no fue asignado ($D_i = 0$). Es decir, lo que realmente tenemos en nuestra base de datos es:

$$Y_i = Y_i(1)D_i + (1 - D_i)Y_i(0)$$

De tal manera que el resultado observado Y_i para el estudiante con $D_i = 1$ sera su resultado potencial dado que participó $Y_i(1)$ y para el estudiante con $D_i = 0$ observaremos su resultado potencial dado que no participó $Y_i(0)$. Pero nunca seremos capaces de ver para el mismo estudiante i su resultado dado que participo $D_i = 1$ y su contrafactual $D_i = 0$.

Es por lo anterior que para trabajar con nuestra muestra utilizaremos no el resultado del individuo i sino que trabajaremos con el promedio de los resultados potenciales, es decir buscaremos estimar los efectos promedio, bien sea el efecto promedio del tratamiento ATE , el efecto promedio del tratamiento en los tratados ATT y el efecto promedio del tratamiento en los no tratados ATU .

b) Define en lenguaje matemático el ATE , el ATT y el ATU . Interprete cada uno de estos conceptos de acuerdo al contexto del caso

R/:

El ATE se define como el efecto promedio del tratamiento:

$$\begin{aligned}\tau_{ATE} &= E[Y_i(1) - Y_i(0)] \\ \tau_{ATE} &= E[Y_i(1)] - E[Y_i(0)]\end{aligned}$$

Que para el caso de nuestra investigación representa la diferencia entre el puntaje promedio en el test de habilidad cognitivas si toda la población de estudiantes de tercer grado hubieran asistido a las aulas tecnológicas, frente al promedio si ninguno hubiera asistido a dichas aulas.

El ATT se define como el efecto promedio del tratamiento en los tratados:

$$\begin{aligned}\tau_{ATT} &= E[Y_i(1) - Y_i(0) | D_i = 1] \\ \tau_{ATT} &= E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 1]\end{aligned}$$

En este caso representa la diferencia en el puntaje promedio del test de habilidades cognitivas del salón que fue asignado a las aulas tecnológicas y el resultado que esos mismos estudiantes hubieran obtenido si no hubieran sido asignados. Es decir, en este caso estamos restringiendo el efecto promedio al grupo que fue tratado.

El ATU se define como el efecto promedio del tratamiento en los no tratados:

$$\begin{aligned}\tau_{ATU} &= E[Y_i(1) - Y_i(0) | D_i = 0] \\ \tau_{ATU} &= E[Y_i(1) | D_i = 0] - E[Y_i(0) | D_i = 0]\end{aligned}$$

En este caso representa la diferencia en el puntaje promedio del test de habilidades cognitivas del salón que no fue asignado a las aulas tecnológicas y el resultado que esos mismos estudiantes hubieran obtenido si hubieran sido asignados. Es decir, en este caso estamos restringiendo el efecto promedio al grupo que no fue tratado.

1.2 Punto2

Usted se encuentra interesada en estimar el siguiente efecto causal:

$$\tau = E[Y_i(1)] - E[Y_i(0)]$$

Para esto plantea una diferencia de medias ingenua

$$\tau_{naive} = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

a). Muestre formalmente que:

$$\tau_{naive} = ATT + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

R/:

$$\begin{aligned}\tau_{naive} &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\ \tau_{naive} &= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 1] \\ \tau_{naive} &= \underbrace{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]}_{ATT} + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0] \\ \tau_{naive} &= ATT + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]\end{aligned}$$

b). Muestre formalmente que:

$$\tau_{naive} = ATU + E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0]$$

R/:

$$\begin{aligned}\tau_{naive} &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\ \tau_{naive} &= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] + E[Y_i(1)|D_i = 0] - \\ &E[Y_i(1)|D_i = 0] \\ \tau_{naive} &= \underbrace{E[Y_i(1)|D_i = 0] - E[Y_i(0)|D_i = 0]}_{ATU} + \\ &E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0] \\ \tau_{naive} &= ATU + E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0]\end{aligned}$$

c). Muestre formalmente que:

$$ATE = \pi ATT + (1 - \pi) ATU$$

R/: Esta demostración es diferente a las anteriores, es menos intuitiva. Sin embargo, podemos aplicar la ley de esperanzas iteradas a la definición de ATE :

$$\begin{aligned}\tau_{ATE} &= E[Y_i(1)] - E[Y_i(0)] \\ \tau_{ATE} &= E[E[Y_i(1) - Y_i(0)|D_i]] \\ \tau_{ATE} &= E[Y_i(1) - Y_i(0)|D_i = 1]P(D_i = 1) + E[Y_i(1) - Y_i(0)|D_i = 0]P(D_i = 0)\end{aligned}$$

Donde:

$$\tau_{ATE} = \underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{ATT} \cdot \underbrace{P(D_i = 1)}_{\pi} + \underbrace{E[Y_i(1) - Y_i(0)|D_i = 0]}_{ATU} \cdot \underbrace{P(D_i = 0)}_{(1-\pi)}$$

Interpretando cada expresión:

a).

$$\tau_{naive} = ATT + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{Selection Bias}}$$

Esta expresión indica que el τ_{naive} recupera el efecto causal sobre los tratados (ATT) si y solo si el sesgo de selección es igual a cero. Esto ocurre cuando el resultado promedio contrafactual del grupo tratado $E[Y_i(0)|D_i = 1]$ es idéntico al resultado promedio observado del grupo de control $E[Y_i(0)|D_i = 0]$. Para que esto se cumpla la composición de los salones debería ser homogénea o el proceso de asignación debería haber sido aleatorio, garantizando que no existan diferencias pre-existentes entre ambos grupos.

b).

$$\tau_{naive} = ATU + \underbrace{E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0]}_{\text{Heterogeneity Bias}}$$

Por otro lado esta expresión nos indica que el τ_{naive} recupera el efecto promedio causal sobre los no tratados (ATU) si y solo si el sesgo de heterogeneidad en resultados del tratamiento es igual a cero. Esto ocurre cuando el resultado promedio de los estudiantes tratados $E[Y_i(1)|D_i = 1]$ es igual al resultado promedio que hubieran obtenido los estudiantes del grupo de control si se les hubiera asignado el tratamiento $E[Y_i(1)|D_i = 0]$. Esto quiere decir que el efecto de las computadoras sería el mismo, independientemente del salón que hubiera sido seleccionado.

c).

$$ATE = \pi \cdot ATT + (1 - \pi) \cdot ATU$$

Por último esta expresión nos dice que el efecto promedio del tratamiento es un promedio ponderado del efecto promedio en los tratados ATT y en los no tratados ATU .

¿Es cierto que si $\tau_{naive} = ATT$, entonces $\tau_{naive} = ATU$?

Esta afirmación no es cierta. Como vimos, el $\tau_{naive} = ATT$ sí y solo sí el resultado promedio contrafactual del grupo tratado $E[Y_i(0)|D_i = 1]$ es idéntico al resultado promedio observado del grupo de control $E[Y_i(0)|D_i = 0]$, lo cual no nos está brindando ninguna información al respecto de algún posible sesgo de heterogeneidad del tratamiento, es decir no nos garantiza que el efecto del tratamiento sea el mismo para ambos grupos. Por su parte, si el $\tau_{naive} = ATU$ solo sabemos que hay ausencia de sesgo de heterogeneidad, pero no tenemos certeza al respecto de la comparabilidad de los grupos (sesgo de selección).

Teniendo como base la descomposición que se hace en libro Causal Inference: The Mixtape de Scott Cunningham podemos demostrar la relación entre estos parámetros:

$$\begin{aligned}\tau_{ATE} &= \pi \cdot ATT + (1 - \pi) \cdot ATU \\ \tau_{ATE} &= \pi \cdot ATT + (1 - \pi) \cdot ATU + (1 - \pi) \cdot ATT - (1 - \pi) \cdot ATT \\ \tau_{ATE} &= ATT \cdot \underbrace{(\pi + 1 - \pi)}_1 - (1 - \pi) \cdot ATT + (1 - \pi) \cdot ATU \\ \tau_{ATE} &= ATT - (1 - \pi) \cdot ATT + (1 - \pi) \cdot ATU\end{aligned}$$

Despejando ATT :

$$\tau_{ATT} = ATE + (1 - \pi) \cdot (ATT - ATU)$$

Remplazando la expresión que acabamos obtener de τ_{ATT} en nuestro estimador τ_{naive} , obtenemos la descomposición completa del sesgo:

$$\begin{aligned}\tau_{naive} &= ATT + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{Selection Bias}} \\ \tau_{naive} &= ATE + \underbrace{(1 - \pi) \cdot (ATT - ATU)}_{\text{Heterogeneity Bias}} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{Selection Bias}}\end{aligned}$$

Por lo tanto en nuestro caso, es posible que el salón elegido fuera comparable inicialmente al de control (sin sesgo de selección), pero que las laptops tengan un retorno mayor en el salón tratado debido a factores específicos del grupo o del profesor. Por lo tanto para que $ATE = ATT = ATU$ necesitamos tener ausencia del sesgo de heterogeneidad y de selección.

1.3 Punto3

Para encontrar el efecto causal promedio de participar en el programa de tecnología educativa sobre el resultado en el test, considere el siguiente modelo lineal:

$$Y_i = \beta_0 + \tau D_i + U_i$$

Demuestre que el estimador $\hat{\tau}_{MCO}$ coincide con la diferencia de medias entre el grupo de tratados y el grupo de no tratados. Es decir, pruebe que $\hat{\tau}_{MCO} = \hat{\tau}_{DM}$, donde:

$$\hat{\tau}_{DM} := \frac{1}{N_1} \sum_{\{i:D_i=1\}} Y_i - \frac{1}{N_0} \sum_{\{i:D_i=0\}} Y_i,$$

y donde $N_1 = \sum_i D_i$; $N_0 = \sum_i (1 - D_i)$.

R/:

Para demostrar que $\hat{\tau}_{MCO} = \hat{\tau}_{DM}$, podemos estimar $\hat{\tau}_{MCO}$ haciendo la minimización de cuadrados a los residuos de nuestro modelo de regresión:

$$\min_{\hat{\beta}_0, \hat{\tau}} \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\tau} D_i)^2$$

FOC respecto de $\hat{\beta}_0$:

$$-2 \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\tau} D_i) = 0$$

$$\sum_{i=1}^N Y_i - N \hat{\beta}_0 - \hat{\tau} \sum_{i=1}^N D_i = 0$$

$$N\bar{Y} - N\hat{\beta}_0 - \hat{\tau}N_1 = 0$$

Despejando $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{Y} - \hat{\tau} \frac{N_1}{N}$$

FOC respecto de $\hat{\tau}$:

$$-2 \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\tau} D_i) D_i = 0$$

Para que la sumatoria exista $D_i = 1$:

$$\sum_{i:D_i=1} (Y_i - \hat{\beta}_0 - \hat{\tau}) = 0$$

$$\sum_{i:D_i=1} Y_i - N_1 \hat{\beta}_0 - N_1 \hat{\tau} = 0$$

Despejando $\hat{\tau}$:

$$\hat{\tau} = \frac{1}{N_1} \sum_{i:D_i=1} Y_i - \hat{\beta}_0$$

$$\hat{\tau} = \bar{Y}_1 - \hat{\beta}_0$$

Reemplazamos el β_0 que obtuvimos:

$$\hat{\tau} = \bar{Y}_1 - \left(\bar{Y} - \hat{\tau} \frac{N_1}{N} \right)$$

$$\hat{\tau} - \hat{\tau} \frac{N_1}{N} = \bar{Y}_1 - \bar{Y}$$

$$\hat{\tau} \left(1 - \frac{N_1}{N} \right) = \bar{Y}_1 - \bar{Y}$$

Si $\frac{N_1}{N}$ es la proporción de tratados, $1 - \frac{N_1}{N}$ es la proporción de controles:

$$\hat{\tau} \left(\frac{N_0}{N} \right) = \bar{Y}_1 - \bar{Y}$$

El promedio es la suma del promedio de los grupos ponderado:

$$\hat{\tau} \left(\frac{N_0}{N} \right) = \bar{Y}_1 - \left(\bar{Y}_1 \frac{N_1}{N} + \bar{Y}_0 \frac{N_0}{N} \right)$$

$$\hat{\tau} \left(\frac{N_0}{N} \right) = \bar{Y}_1 \left(1 - \frac{N_1}{N} \right) - \bar{Y}_0 \frac{N_0}{N}$$

$$\hat{\tau} \left(\frac{N_0}{N} \right) = \bar{Y}_1 \left(\frac{N_0}{N} \right) - \bar{Y}_0 \left(\frac{N_0}{N} \right)$$

$$\hat{\tau}_{MCO} = \bar{Y}_1 - \bar{Y}_0$$

$$\hat{\tau}_{MCO} = \frac{1}{N_1} \sum_{\{i:D_i=1\}} Y_i - \frac{1}{N_0} \sum_{\{i:D_i=0\}} Y_i$$

Cómo vemos, $\hat{\tau}_{MCO} = \hat{\tau}_{DM}$.

b). ¿que supuesto debe cumplirse para que $\hat{\tau}_{DM}$ sea un estimador insesgado el efecto causal promedio?

Para verificar que supuesto se debe cumplir vamos a evaluar la insesgadez tomando la esperanza matematica de nuestro estimador:

$$\begin{aligned} E[\hat{\tau}_{DM}] &= E[\bar{Y}_1 - \bar{Y}_0] \\ E[\hat{\tau}_{DM}] &= E[\bar{Y}_1] - E[\bar{Y}_0] \end{aligned}$$

Si usamos notación de resultados potenciales:

$$E[\hat{\tau}_{DM}] = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

Como en la primera demostración del taller, si sumamos y restamos el contrafactual tenemos que:

$$E[\hat{\tau}_{DM}] = \underbrace{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]}_{ATT} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{Selección Bias}}$$

Por lo tanto, para que nuestro estimador $\hat{\tau}_{DM}$ se insesgado necesitamos evitar el sesgo de selección lo cual se logra por medio de la aleatorización garantizando que $(Y_i(1), Y_i(0))$ D_i

Ahora veamos esto como lo podemos obtener tambien de la recta de regresión. Recordemos que nuestra linea de regresión es:

$$Y_i = \beta_0 + \tau D_i + U_i$$

Partamos del mismo punto que antes:

$$\begin{aligned} E[\hat{\tau}_{DM}] &= E[\bar{Y}_1 - \bar{Y}_0] \\ E[\hat{\tau}_{DM}] &= E[\bar{Y}_1] - E[\bar{Y}_0] \\ E[\hat{\tau}_{DM}] &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \end{aligned}$$

Ahora si tomamos estos valores esperados de nuestra recta obtenemos:

$$\begin{aligned} E[Y_i|D_i = 1] &= \beta_0 + \tau + E[U_i|D_i = 1] \\ E[Y_i|D_i = 0] &= \beta_0 + E[U_i|D_i = 0] \end{aligned}$$

Por lo tanto tenemos:

$$E[\hat{\tau}_{DM}] = \beta_0 + \tau + E[U - i|D_i = 1] - (\beta_0 + E[U_i|D_i = 0])$$

tal que:

$$E[\hat{\tau}_{DM}] = \tau + \underbrace{E[U_i|D_i = 1] - E[U - i|D_i = 0]}_{\text{Selection Bias}}$$

Por lo tanto en terminos de la recta de regresión el supuesto de independencia $(Y_i(1), Y_i(0)) \perp D_i$ se traduce en el supuesto de exogenidad $E[U_i|D_i] = 0$. Esto nos señala que, para tener un estimador insesgado, no debemos tener factores no observables correlacionados con el tratamiento que afecten el resultado, haciendo a los grupos diferentes.”

1.4 Punto4

Suponga ahora que la asignación de estudiantes a cada salón se realiza de manera aleatoria. De igual forma, la decisión sobre qué salón recibe el tratamiento también se tomó aleatoriamente. Además, todos los estudiantes pertenecientes al salón asignado al tratamiento efectivamente participaron en el programa. Demuestre que el estimador

$$\hat{\tau}_{DM} \xrightarrow{p} \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]$$

Puede suponer que el supuesto de SUTVA se cumple. Posteriormente, defina intuitivamente y explique la importancia de la propiedad de consistencia de un estimador.

R/:

Para demostrar que $\hat{\tau}_{DM} \xrightarrow{p} \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]$ empecemos por:

$$\text{plim}[\hat{\tau}_{DM}] = \text{plim} \left(\left(\frac{1}{N_1} \sum_{i:D_i=1} Y_i \right) - \left(\frac{1}{N_0} \sum_{i:D_i=0} Y_i \right) \right)$$

Como vimos en clase, por mapeo continuo (Slutsky):

$$\text{plim}[\hat{\tau}_{DM}] = \text{plim} \left(\frac{1}{N_1} \sum_{i:D_i=1} Y_i \right) - \text{plim} \left(\frac{1}{N_0} \sum_{i:D_i=0} Y_i \right)$$

Así mismo por WLLN:

$$\text{plim}[\hat{\tau}_{DM}] = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

Si escribimos en resultado potencial:

$$\text{plim}[\hat{\tau}_{DM}] = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

Aplicamos la suma y resta del contrafactual:

$$\text{plim}[\hat{\tau}_{DM}] = \underbrace{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]}_{ATT} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{Selección Bias}}$$

Cómo sabemos por el enunciado que la asignación a los salones fue aleatoria entonces podemos estar tranquilo respecto del sesgo de selección, es decir sabemos que $((Y_i(1), Y_i(0)) \perp D_i)$:

$$\text{plim}[\hat{\tau}_{DM}] = \underbrace{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]}_{ATT}$$

Simplificando:

$$\text{plim}[\hat{\tau}_{DM}] = \underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{ATT}$$

que es lo mismo que $\hat{\tau}_{DM} \xrightarrow{p} E[Y_i(1) - Y_i(0)|D_i = 1]$