Taller 1. Econometría Avanzada

Daniel Ricardo Amaya Alba

Table of contents

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, Di 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

Describa 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 asigando 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 inddividuo i sino que trabajaremos con el promedio de los resultados potenciales, es decir buscaremos estumar los efectos promedio, bien sea el efecto promedio del tramaiento 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 efectro promedio del tratamiento:

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

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 asisitido a dichas aulas.

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

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

En este caso representa la diferencia en el puntaje promedio del test de habilidades cognitivas del salon 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 efectro promedio del tratamiento en los no tratados:

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

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

En este caso representa la diferencia en el puntaje promedio del test de habilidades cognitivas del salon 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/:

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

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:

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

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 tro lado esta expresión nos indica que el τ_{naive} recupera el efecto promedio causal sobre los no tratado (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 dle salon que hubiera sido seleccionado.

c).

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

Por úlimo esta expresión nos dice que el efecto promedio del tratamiento es un promedio ponderado del efecto promedio en los tratado 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 esta 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 descomposicón que se hace en libro Causal Inference: The Mixtape de Scott Cunningham podemos demostrar la relación entre estos parametros:

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

Despejando ATT:

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

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

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

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 hterogeneidad 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 β_0:

$$\hat{\beta}_0 = \bar{Y} - \hat{\tau}\frac{N1}{N}$$

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:

$$\hat{\tau} = \bar{Y_1} - \left(\bar{Y} - \hat{\tau}\,\frac{N_1}{N}\right)$$

$$\hat{\tau} - \hat{\tau}\frac{N1}{N} = \bar{Y_1} - \bar{Y}$$

$$\hat{\tau}\left(1 - \frac{N_1}{N}\right) = \bar{Y_1} - \bar{Y}$$

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 esitimador insesgadod el efecto causal promedio?

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

$$E[\hat{\tau}_{DM}] = E[\bar{Y_1} - \bar{Y_0}]$$
$$E[\hat{\tau}_{DM}] = E[\bar{Y_1}] - E[\bar{Y_0}]$$

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:

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

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:

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

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

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

Simplficando:

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

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

2 Segundo Ejercicio

Tal como señalan Abadie et al. (2022), los errores estándar clusterizados han adquirido un rol central en el trabajo empírico y esto no es casualidad. Si bien una parte sustancial de los esfuerzos de los investigadores se concentra en la obtención de estimadores consistentes e insesgados, resulta también fundamental poder realizar inferencia estadística válida sobre dichos estimadores. En este contexto, una de las prácticas más extendidas en la literatura empírica es el reporte de errores estándar clusterizados. No obstante, su uso plantea varias preguntas recurrentes en la práctica: ¿en qué situaciones es apropiado utilizar este tipo de errores estándar?, ¿cuándo tienen un impacto importante sobre la inferencia estadística? y ¿a qué nivel debería realizarse la clusterización?

Teniendo en cuenta lo anterior, suponga que usted desea estudiar **el impacto de que un estudiante de raza negra tenga un profesor de la misma raza**. Es bien sabido que existen brechas raciales en los resultados educacionales, lo cual resulta problemático por diversas razones: la movilidad social, los ingresos futuros y el bienestar social están estrechamente vinculados al nivel educativo. En este contexto, usted decide seguir la línea de investigación propuesta por Gershenson et al. (2022) y busca replicar algunos de sus principales resultados, en particular aquellos relacionados con el desempeño en exámenes y la entrada a la universidad.

Para realizar estos ejercicios, usted dispone de una base de datos de corte transversal denominada `race_teaching.dta` que proviene de un experimento realizado en EE.UU. que aleatorizaba la raza de los profesores en diferentes salones de kínder. Sus datos están a nivel individual, donde cada observación es una persona de raza negra que participó en el experimento.

Esta base contiene información sobre los resultados en el SAT (examen estandarizado para el ingreso a la universidad) y una variable indicadora que señala si el estudiante asistió a la universidad. A estas variables de interés se las denota por Y_i. Adicionalmente, la base incluye una variable indicadora D_i, que toma el valor de uno si el estudiante i tuvo un profesor de raza negra en kínder y cero en caso contrario. Asimismo, se incorporan características geográficas, tales como el salón, el colegio, el condado y el estado al que pertenece cada individuo.

Antes de abordar las decisiones empíricas que deberá tomar en su investigación, usted decide analizar primero, desde una perspectiva teórica, los errores estándar clusterizados. De este modo, podrá comprender su formulación, sus principales propiedades y sentar las bases para determinar en qué contextos resulta apropiado utilizarlos y cómo deben implementarse.

2.1 Punto1

Realice un análisis crítico de los supuestos de homocedasticidad y no-correlación del término de error. Para ello:

- Defina ambos supuestos, tanto desde un punto de vista matemático como intuitivo, en el contexto del caso de estudio.
- Discuta si resulta razonable asumir el cumplimiento de estos supuestos dada la pregunta de investigación. Justifique su respuesta utilizando un ejemplo intuitivo para cada supuesto.

R/:

El supuesto de homocedasticidad implica que $E[e_i^2|x] = \sigma^2$ que tambien podría ser visto como $E[e_i^2|D_i = 1] = \sigma^2$ y $E[e_i^2|D_i = 0] = \sigma^2$, lo que quiere decir que la varianza de los errores es constante para cada estudiante independientemente de sus caracteristicas (X), en este caso, quiere decir que la dispersión en los resultados de la prueba SAT para cada estudiante es la misma para el grupo de estudiantes tratados y no tratados.

Por otro lado, el supuesto de no-correlación implica que $E[e_i e_j] = 0$. Esto quiere decier que conocer el error del estudiante i no me dice nada al respecto del resultado del estudiante j. Es decir, si a un estudiante le va muy bien en la prueba, por algun factor diferente al tratamiento, eso no quiere decir que a otro estudiante le vaya a ir igual de bien o que le vaya a ir mucho peor. Es decir si un estudiante el dia del examen se desperto con dolor de cabeza, eso no nos dice nada respecto de otro estudiante.

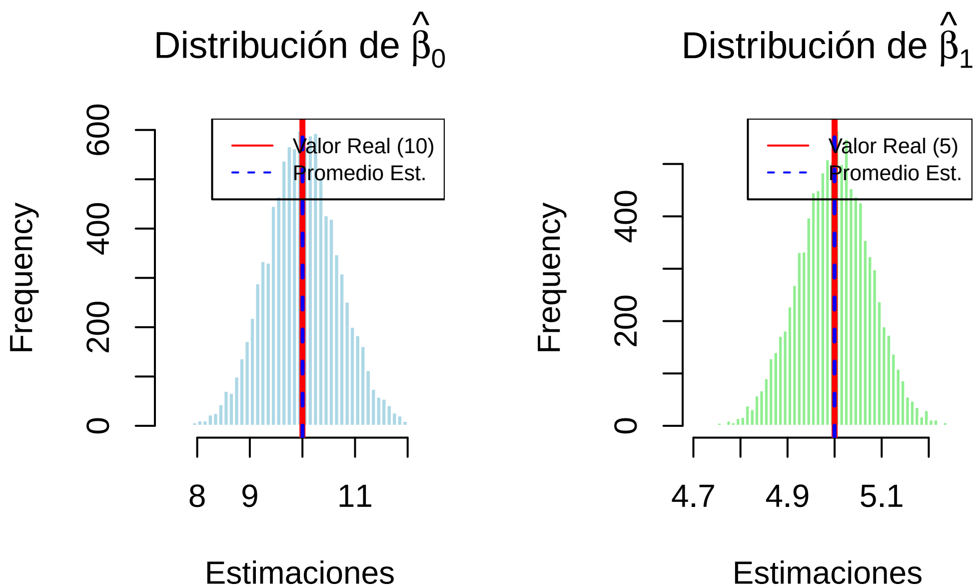
El supuesto de homocedasticidad en nuestro RCT no tiene muchos sentido. Tal como lo señalan Angrist y Pischke (2009) en la vida real, aunque no conocemos la CEF es viable pensar que esta no es lineal necesarimente, por lo que el hecho que no sea lineal ya va a generar por si mismo que tengamos variabilidad en la forma en que la línea de regresión se aproxima. Así mismo hay que tener en cuenta que una de nuestras variables dependientes es binaria (asístencia a la universidad), de acuerdo con los autores mencionados esto anula el supusto de homocedasticidad porque por simple construcción va a haber variación en los errores.

Por otro Abadie et al.(2023) señalan que el ajuste por cluster de errores estandar se debe hacer por nivel de muestreo y de asignación de tratamiento. En este caso el tratamiento fue asignado por salones por lo que en este caso no solo no deberíamos asumir errores homocedasticos sino aplicar cluster. Esto tambien implica que debemos ser concientes que es muy probable que l supuesto de no-correlación implica que $E[e_i e_j] = 0$ no se cumpla, pues en este caso pueden haber factores no observados en cada salon que si correlaciónen los resultados de l estudiante i con el estudiante j que pertenecen al mismo salon, lo que vimos formalmente en clase como $E[e_{gi} e_{gh}] \neq 0$

```
# 1. Configuración de parámetros y "Semilla" para reproducibilidad
set.seed(123)

# Parámetros poblacionales
beta0_true <- 10
beta1_true <- 5
```

```r
N <- 100       # Tamaño de la muestra
R <- 10000     # Número de repeticiones (simulaciones)

# Contenedores para guardar los resultados de cada simulación
beta0_hats <- numeric(R)
beta1_hats <- numeric(R)

# 2. Inicio del Bucle (Algoritmo repetido 10,000 veces)
for (i in 1:R) {

  # --- Paso 1: Generar muestras aleatorias ---
  # X ~ N(8, 16) -> mean=8, sd=4
  xi <- rnorm(n = N, mean = 8, sd = 4)

  # e ~ N(0, 9) -> mean=0, sd=3
  ei <- rnorm(n = N, mean = 0, sd = 3)

  # --- Paso 2: Calcular Yi ---
  # Modelo: Yi =  0 +  1*Xi + ei
  yi <- beta0_true + beta1_true * xi + ei

  # --- Paso 3: Estimar por MCO usando Álgebra Matricial ---
  # Creamos la matriz X (agregando una columna de 1s para el intercepto)
  X_mat <- cbind(1, xi)

  # Fórmula: (X'X)^-1 X'Y
  # t() es transpuesta, %*% es multiplicación matricial, solve() es la inversa
  beta_mco <- solve(t(X_mat) %*% X_mat) %*% t(X_mat) %*% yi

  # --- Paso 4: Guardar los valores estimados ---
  beta0_hats[i] <- beta_mco[1] # El intercepto
  beta1_hats[i] <- beta_mco[2] # La pendiente
}

# 3. Graficar los Histogramas para demostrar Insesgamiento
par(mfrow = c(1, 2)) # Dividir pantalla en 2 paneles

# Histograma para Beta 0
hist(beta0_hats,
     main = expression(paste("Distribución de ", hat(beta)[0])),
     xlab = "Estimaciones",
     col = "lightblue",
```

```
     border = "white",
     breaks = 50)
# Línea roja vertical en el valor verdadero (10)
abline(v = beta0_true, col = "red", lwd = 3)
# Línea azul (promedio de las estimaciones)
abline(v = mean(beta0_hats), col = "blue", lwd = 2, lty = 2)
legend("topright", legend=c("Valor Real (10)", "Promedio Est."), col=c("red", "blue"), lty=c

# Histograma para Beta 1
hist(beta1_hats,
     main = expression(paste("Distribución de ", hat(beta)[1])),
     xlab = "Estimaciones",
     col = "lightgreen",
     border = "white",
     breaks = 50)
# Línea roja vertical en el valor verdadero (5)
abline(v = beta1_true, col = "red", lwd = 3)
# Línea azul (promedio de las estimaciones)
abline(v = mean(beta1_hats), col = "blue", lwd = 2, lty = 2)
legend("topright", legend=c("Valor Real (5)", "Promedio Est."), col=c("red", "blue"), lty=c(
```



```r
# Restaurar configuración de gráficos
par(mfrow = c(1, 1))

# 4. Mostrar resultados numéricos en consola
cat("--- Resultados de la Simulación ---\n")
```

```
--- Resultados de la Simulación ---
```

```r
cat("Promedio de Beta 0 estimado:", mean(beta0_hats), "(Valor real: 10)\n")
```

```
Promedio de Beta 0 estimado: 10.00212 (Valor real: 10)
```

```r
cat("Promedio de Beta 1 estimado:", mean(beta1_hats), "(Valor real: 5)\n")
```

```
Promedio de Beta 1 estimado: 4.999767 (Valor real: 5)
```