

---

**Declaración de uso de inteligencia artificial.** En la elaboración de este trabajo se utilizó Claude (Anthropic) como herramienta de asistencia. El uso incluyó corrección de errores de codificación en  $\text{\LaTeX}$  y python, y sugerencias de redacción.

## Primer Ejercicio

Desde la perspectiva del capital humano, estudios como el de [Cunha & Heckman \(2007\)](#) 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. Describa el problema de estimación utilizando el lenguaje de resultados potenciales. Para esto:

- 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.

Los resultados potenciales del estudiante  $i$  se definen como:

- $Y_i(1)$ : puntaje en el test de habilidades cognitivas si el estudiante  $i$  fue asignado al aula tecnológica.
- $Y_i(0)$ : puntaje en el test de habilidades cognitivas si el estudiante  $i$  permaneció en el salón regular.

El resultado observado se escribe como  $Y_i = D_i \times Y_i(1) + (1 - D_i) \times Y_i(0)$ , donde  $D_i \in \{0, 1\}$  indica la asignación al tratamiento. En este caso, el problema de inferencia causal está en que, para cada estudiante, solo uno de los dos resultados potenciales es observable. Si  $D_i = 1$ , se observa  $Y_i = Y_i(1)$  pero  $Y_i(0)$  no. Si  $D_i = 0$ , ocurre lo contrario. Esta restricción hace imposible calcular el efecto individual  $Y_i(1) - Y_i(0)$  para cualquier estudiante.

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

Los tres parámetros causales se definen como:

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

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

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

- 1) ATE: El efecto promedio del tratamiento. Es el promedio del efecto del programa en el puntaje cognitivo para un estudiante tomado al azar de la muestra, sin distinción de a qué salón fue asignado.
- 2) ATT: El efecto promedio en los tratados. Es el efecto promedio del programa para los estudiantes que efectivamente fueron asignados al Aula Tecnológica. Responde a la pregunta de si el programa fue beneficioso para quienes lo recibieron.
- 3) ATU: El efecto promedio en los no tratados. Es el efecto promedio del programa para los estudiantes que permanecieron en el salón regular. Responde a la pregunta de qué habría ocurrido con el puntaje de esos estudiantes si hubieran participado en el programa.

2. Usted se encuentra interesada en estimar el siguiente efecto causal:

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

Para esto plantea una diferencia de medias ingenua:

$$\tau_{\text{ing}} = E[Y_i | D_i = 1] - E[Y_i | D_i = 0] \quad (2)$$

Muestre formalmente que:

$$a) \tau_{\text{ing}} = \text{ATT} + (\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0])$$

Partiendo de la definición del estimador de diferencia de medias:

$$\tau_{\text{ing}} = \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0].$$

Se reemplaza la ecuación  $Y_i = D_i \times Y_i(1) + (1 - D_i) \times Y_i(0)$  dentro del estimador:

$$\begin{aligned}\tau_{\text{ing}} &= \mathbb{E}[D_i \times Y_i(1) + (1 - D_i) \times Y_i(0) | D_i = 1] \\ &\quad - \mathbb{E}[D_i \times Y_i(1) + (1 - D_i) \times Y_i(0) | D_i = 0] \\ \tau_{\text{ing}} &= \mathbb{E}[(1) \times Y_i(1) + (1 - (1)) \times Y_i(0) | D_i = 1] \\ &\quad - \mathbb{E}[(0) \times Y_i(1) + (1 - (0)) \times Y_i(0) | D_i = 0] \\ \tau_{\text{ing}} &= \mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0]\end{aligned}$$

Se suma y resta un contrafactual, concretamente  $\mathbb{E}[Y_i(0) | D_i = 1]$ :

$$\tau_{\text{ing}} = \mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 1] + \mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0].$$

Reordenando, el primer par de términos es el ATT por definición:

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

El segundo par de términos es el sesgo de selección:

$$\text{Sesgo} = \mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0].$$

Por tanto:

$$\tau_{\text{ing}} = \underbrace{\mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 1]}_{\text{ATT}} + \underbrace{\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0]}_{\text{Sesgo de selección}}.$$

El sesgo de selección captura la diferencia en el resultado potencial sin tratamiento entre los que efectivamente se tratan y los que no. Si los estudiantes asignados al aula tecnológica hubieran tenido, en ausencia del programa, puntajes distintos a los del salón regular, entonces  $\tau_{\text{ing}} \neq \text{ATT}$ .

b)  $\tau_{\text{ing}} = \text{ATU} + (\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(1)|D_i = 0])$

Partiendo del mismo punto de la demostración anterior:

$$\tau_{\text{ing}} = \mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0].$$

Se suma y resta un contrafactual, concretamente  $\mathbb{E}[Y_i(1) | D_i = 0]$ :

$$\tau_{\text{ing}} = \mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(1) | D_i = 0] + \mathbb{E}[Y_i(1) | D_i = 0] - \mathbb{E}[Y_i(0) | D_i = 0].$$

Reorganizando, se ve que el segundo par de términos es el ATU por definición:

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

Y el primer par de términos es el sesgo de selección en  $Y_i(1)$ :

$$\text{Sesgo en } Y_i(1) = \mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(1) | D_i = 0].$$

Reordenando:

$$\tau_{\text{ing}} = \underbrace{\mathbb{E}[Y_i(1) | D_i = 0] - \mathbb{E}[Y_i(0) | D_i = 0]}_{\text{ATU}} + \underbrace{\mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(1) | D_i = 0]}_{\text{Sesgo de selección en } Y_i(1)}.$$

Este sesgo captura si los tratados obtendrían puntajes con el programa distintos a los del grupo de control. Si los estudiantes del aula tecnológica, inherentemente, tuvieran puntajes con tratamiento más altos por capacidad propia,  $\tau_{\text{ing}}$  sobreestima el ATU.

c)  $\text{ATE} = \pi \text{ATT} + (1 - \pi) \text{ATU}$  donde  $\pi \equiv \mathbb{P}(D_i = 1)$

Bajo la definición del ATE:

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

Y bajo la ley de esperanzas iteradas, condicionando en  $D_i$ :

$$\text{ATE} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1] \frac{|\{j \in \mathbb{N} | D_j = 1\}|}{N} + \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 0] \frac{|\{j \in \mathbb{N} | D_j = 0\}|}{N}$$

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

Reconocemos los términos por definición. El primer factor de cada sumando es, respectivamente, el ATT y el ATU:

$$\text{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1], \quad \text{ATU} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 0].$$

Definimos  $\pi \equiv \mathbb{P}(D_i = 1)$ , de modo que  $\mathbb{P}(D_i = 0) = 1 - \pi$ . Sustituyendo:

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

Esta expresión muestra que el ATE es un promedio ponderado del ATT y el ATU, con pesos iguales a la proporción de la población en cada grupo. En el caso de estudio,  $\pi = N_1/N$  es la fracción de estudiantes en el aula Tecnológica.

Interprete cada expresión, haciendo énfasis en el significado de cada término. Luego responda, ¿es cierto que si  $\tau_{\text{ing}} = \text{ATT}$ , entonces  $\tau_{\text{ing}} = \text{ATU}$ ? Si su respuesta es afirmativa, demuéstrelo. En caso contrario, dé un contra ejemplo de acuerdo al caso de estudio.

No es necesariamente cierto que  $\tau_{\text{ing}} = \text{ATT}$  implique  $\tau_{\text{ing}} = \text{ATU}$ .

Si  $\tau_{\text{ing}} = \text{ATT}$ , del resultado del punto (a) se ve que el sesgo de selección es cero:

$$\mathbb{E}[Y_i(0) | D_i = 1] = \mathbb{E}[Y_i(0) | D_i = 0].$$

Pero del resultado del punto (b):

$$\tau_{\text{ing}} = \text{ATU} + (\mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(1) | D_i = 0]).$$

Para que  $\tau_{\text{ing}} = \text{ATU}$  también se cumpla, se requeriría adicionalmente que:

$$\mathbb{E}[Y_i(1) | D_i = 1] = \mathbb{E}[Y_i(1) | D_i = 0].$$

Estas condiciones son independientes entre sí. Como ejemplo, se puede suponer que los estudiantes del aula tecnológica son intrínsecamente más hábiles y, por tanto, obtendrían puntajes más altos incluso con el programa ( $\mathbb{E}[Y_i(1) | D_i = 1] > \mathbb{E}[Y_i(1) | D_i = 0]$ ), mientras que en ausencia del programa sus puntajes coinciden con los del salón regular ( $\mathbb{E}[Y_i(0) | D_i = 1] = \mathbb{E}[Y_i(0) | D_i = 0]$ ). En ese caso  $\tau_{\text{ing}} = \text{ATT}$  pero  $\tau_{\text{ing}} > \text{ATU}$ , por lo que la implicación no se cumple.

3. 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:

$$Y_i = \beta_0 + \tau D_i + u_i. \tag{3}$$

- a) 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

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

$$\text{y } N_1 = \sum_i D_i \text{ y } N_0 = \sum_i (1 - D_i).$$

Basandose en el modelo  $Y_i = \beta_0 + \tau D_i + u_i$  con  $D_i \in \{0, 1\}$ , las condiciones de primer orden del estimador MCO son:

$$\frac{\partial}{\partial \hat{\beta}_0} \sum_i (Y_i - \hat{\beta}_0 - \hat{\tau} D_i)^2 = 0 \implies \sum_i (Y_i - \hat{\beta}_0 - \hat{\tau} D_i) = 0,$$

$$\frac{\partial}{\partial \hat{\tau}} \sum_i (Y_i - \hat{\beta}_0 - \hat{\tau} D_i)^2 = 0 \implies \sum_i D_i(Y_i - \hat{\beta}_0 - \hat{\tau} D_i) = 0.$$

Se recuelve para  $\hat{\beta}_0$ . De la primera condición:

$$\sum_i Y_i = N\hat{\beta}_0 + \hat{\tau} \sum_i D_i \implies \hat{\beta}_0 = \bar{Y} - \hat{\tau} \bar{D},$$

donde  $\bar{Y} = \frac{1}{N} \sum_i Y_i$  y  $\bar{D} = \frac{N_1}{N}$ .

Se resuelve para  $\hat{\tau}$ . Sustituimos  $\hat{\beta}_0$  en la segunda condición y usamos que  $D_i^2 = D_i$  para  $D_i \in \{0, 1\}$ :

$$\begin{aligned} 0 &= \sum_i D_i Y_i - \hat{\beta}_0 \sum_i D_i - \hat{\tau} \sum_i D_i^2 \\ &= \sum_i D_i Y_i - (\bar{Y} - \hat{\tau} \bar{D}) N_1 - \hat{\tau} N_1 \\ &= \sum_i D_i Y_i - N_1 \bar{Y} + \hat{\tau} N_1 \bar{D} - \hat{\tau} N_1 \\ &= \sum_i D_i Y_i - N_1 \bar{Y} - \hat{\tau} N_1 (1 - \bar{D}). \end{aligned}$$

Despejando  $\hat{\tau}$ :

$$\hat{\tau} = \frac{\sum_i D_i Y_i - N_1 \bar{Y}}{N_1 (1 - \bar{D})}.$$

Separamos  $\bar{Y} = \frac{N_1 \bar{Y}_1 + N_0 \bar{Y}_0}{N}$ :

$$\begin{aligned} \sum_i D_i Y_i - N_1 \bar{Y} &= N_1 \bar{Y}_1 - N_1 \cdot \frac{N_1 \bar{Y}_1 + N_0 \bar{Y}_0}{N} \\ &= N_1 \bar{Y}_1 \left(1 - \frac{N_1}{N}\right) - N_1 \frac{N_0}{N} \bar{Y}_0 \\ &= N_1 \frac{N_0}{N} \bar{Y}_1 - N_1 \frac{N_0}{N} \bar{Y}_0 \\ &= \frac{N_1 N_0}{N} (\bar{Y}_1 - \bar{Y}_0) \end{aligned}$$

Como  $\bar{D} = N_1/N$ :

$$N_1 (1 - \bar{D}) = N_1 \left(1 - \frac{N_1}{N}\right) = N_1 \cdot \frac{N_0}{N} = \frac{N_1 N_0}{N}$$

Dividiendo:

$$\hat{\tau}_{MCO} = \frac{N_1 N_0 / N (\bar{Y}_1 - \bar{Y}_0)}{N_1 N_0 / N} = \bar{Y}_1 - \bar{Y}_0 = \hat{\tau}_{DM}$$

- b) ¿Qué supuestos deben cumplirse para que  $\hat{\tau}_{DM}$  sea un estimador insesgado del efecto causal promedio?

Para que  $\hat{\tau}_{DM}$  sea un estimador insesgado del efecto causal promedio se requiere que:

$$\mathbb{E}[\hat{\tau}_{DM}] = \text{ATE} = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)].$$

Dado que  $\mathbb{E}[\hat{\tau}_{DM}] = \mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0]$ , esta condición exige que el tratamiento sea independiente de los resultados potenciales:

$$\{Y_i(0), Y_i(1)\} \perp D_i.$$

Esto se satisface cuando la asignación al tratamiento es aleatoria. En el contexto del caso, se necesita que el mecanismo por el cual los estudiantes fueron asignados a cada salón sea independiente de su capacidad cognitiva potencial. Si el salón fue elegido de forma no aleatoria, como seleccionar el salón con mejores estudiantes, el sesgo de selección sería distinto de cero y  $\hat{\tau}_{DM}$  no identificaría el ATE.

4. 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.

Partiendo de que un estimador  $\hat{\theta}_N$  es consistente para  $\theta$  si converge en probabilidad al parámetro verdadero:

$$\hat{\theta}_N \xrightarrow{P} \theta \quad (N \rightarrow \infty),$$

es decir, para todo  $\varepsilon > 0$ ,  $\lim_{N \rightarrow \infty} \mathbb{P}(|\hat{\theta}_N - \theta| > \varepsilon) = 0$ .

Suponiendo aleatorización,  $\{Y_i(0), Y_i(1)\} \perp D_i$ . En particular:

$$\mathbb{E}[Y_i(1) | D_i = 1] = \mathbb{E}[Y_i(1)] \quad \text{y} \quad \mathbb{E}[Y_i(0) | D_i = 0] = \mathbb{E}[Y_i(0)].$$

Los  $\{Y_i(1) : D_i = 1\}$  son i.i.d. con media  $\mathbb{E}[Y_i(1) | D_i = 1] = \mathbb{E}[Y_i(1)]$ . Por la ley de grandes números, cuando  $N_1 \rightarrow \infty$ :

$$\bar{Y}_1 = \frac{1}{N_1} \sum_{i: D_i=1} Y_i(1) \xrightarrow{P} \mathbb{E}[Y_i(1) | D_i = 1] = \mathbb{E}[Y_i(1)].$$

De la misma forma, cuando  $N_0 \rightarrow \infty$ :

$$\bar{Y}_0 = \frac{1}{N_0} \sum_{i: D_i=0} Y_i(0) \xrightarrow{P} \mathbb{E}[Y_i(0) | D_i = 0] = \mathbb{E}[Y_i(0)].$$

Por el teorema de mapeo continuo, que intuitivamente quiere decir que la resta es una función continua:

$$\hat{\tau}_{DM} = \bar{Y}_1 - \bar{Y}_0 \xrightarrow{P} \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$$

La última igualdad usa que bajo independencia  $\mathbb{E}[Y_i(1) - Y_i(0) | D_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$ . La consistencia garantiza que, con muestras grandes en ambos grupos, la diferencia de medias converge al efecto causal real del programa, eliminando el sesgo de selección asintóticamente.

## 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.

Nota: Omitir variable `id_salon` ya que no fue incluida en la base de datos

1. 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.

**Homocedasticidad:** El supuesto dice que  $\mathbb{V}[\varepsilon_i \mid X_i] = \sigma^2$  para todo  $i$ ; es decir, la varianza del error es constante e independiente del valor de las características observables o variables independientes. En este contexto,  $X_i = D_i$  indica si el estudiante  $i$  tuvo un profesor negro en kínder y los errores recogen todos las características no observables que afectan el resultado ( $Y_i$ ). El supuesto implica que estas características no observables que tienen relación en el puntaje SAT o en la asistencia universitaria tienen la misma dispersión para los estudiantes con y sin profesor negro. Esto es debatible. Los estudiantes cuyo profesor en kínder fue negro pueden tener características no observables que estén afectando los resultados de las variables de estudio como entornos distintos, como colegios con mayor movilidad social, lo que hace que la variabilidad de los resultados difiera frente a la del grupo de control.

**No correlación:** El supuesto establece que  $\mathbb{E}[\varepsilon_i \varepsilon_j \mid X_i, X_j] = 0$  para todo  $i \neq j$ ; es decir, los errores de distintos individuos no están relacionados entre sí. Así, los estudiantes del mismo colegio comparten recursos, docentes y entorno socioeconómico. Un colegio con factores no observados favorables, como financiación adicional, se esperaría que aumentara el resultado de todos sus estudiantes de manera simultánea. Esto produce correlación positiva entre los errores dentro del mismo colegio, lo que viola el supuesto de no correlación.

2. Por el momento, suponga que usted desea estimar el modelo sin intercepto  $Y_i = \beta X_i + \varepsilon_i^1$ , donde  $X_i$  es una variable determinística. Determine la varianza del estimador de  $\beta$  bajo los siguientes escenarios:

- **Caso 1:** Se cumplen los supuestos de homocedasticidad y no-correlación del término de error.
- **Caso 2:** Se cumple el supuesto de no-correlación, pero existe heterocedasticidad.
- **Caso 3:** Existe heterocedasticidad y correlación entre las características no-observadas de los individuos que asisten al mismo colegio.

Para cada caso, identifique el estimando poblacional correspondiente y proponga un estimador adecuado de la varianza. Justifique brevemente su respuesta.

Para el modelo  $Y_i = \beta X_i + \varepsilon_i$  (sin intercepto), el estimador MCO es

$$\hat{\beta} = \frac{\sum_i x_i Y_i}{\sum_i x_i^2}.$$

Sustituyendo  $Y_i = \beta x_i + \varepsilon_i$ :

$$\begin{aligned}\hat{\beta} &= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2} \\ &= \frac{\beta \sum_i x_i^2 + \sum_i x_i \varepsilon_i}{\sum_i x_i^2} \\ &= \beta + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2},\end{aligned}$$

de modo que el error de estimación es

$$\hat{\beta} - \beta = \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}.$$

Aplicando el operador de varianza y como  $x_i$  es no estocástico,  $(\sum_i x_i^2)^2$  es una constante y puede sacarse de la esperanza:

$$\mathbb{V}[\hat{\beta}] = \mathbb{E}[(\hat{\beta} - \beta)^2] = \mathbb{E}\left[\left(\frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}\right)^2\right] = \frac{\mathbb{E}[(\sum_i x_i \varepsilon_i)^2]}{(\sum_i x_i^2)^2}.$$

Como  $\mathbb{E}[\varepsilon_i] = 0$  para todo  $i$ , se tiene  $\mathbb{E}[\sum_i x_i \varepsilon_i] = \sum_i x_i \mathbb{E}[\varepsilon_i] = 0$ , por lo que

$$\mathbb{E}\left[\left(\sum_i x_i \varepsilon_i\right)^2\right] = \mathbb{V}\left[\sum_i x_i \varepsilon_i\right].$$

Expandiendo la varianza por la propiedad  $\mathbb{V}[A + B] = \mathbb{V}[A] + \mathbb{V}[B] + 2 \text{Cov}(A, B)$ , aplicada iterativamente:

$$\mathbb{V}\left[\sum_i x_i \varepsilon_i\right] = \sum_i \mathbb{V}[x_i \varepsilon_i] + 2 \sum_{i < j} \text{Cov}(x_i \varepsilon_i, x_j \varepsilon_j).$$

Como  $x_i$  es una constante,  $\mathbb{V}[x_i \varepsilon_i] = x_i^2 \mathbb{V}[\varepsilon_i]$  y  $\text{Cov}(x_i \varepsilon_i, x_j \varepsilon_j) = x_i x_j \text{Cov}(\varepsilon_i, \varepsilon_j)$ . Esto puede escribirse como:

$$\mathbb{V}\left[\sum_i x_i \varepsilon_i\right] = \sum_i \sum_j x_i x_j \text{Cov}(\varepsilon_i, \varepsilon_j).$$

Sustituyendo, la varianza es:

$$\mathbb{V}[\hat{\beta}] = \frac{\sum_i \sum_j x_i x_j \text{Cov}(\varepsilon_i, \varepsilon_j)}{(\sum_i x_i^2)^2}. \quad (*)$$

Los tres casos se diferencian por el supuesto sobre  $\text{Cov}(\varepsilon_i, \varepsilon_j)$ .

<sup>1</sup>Considere que es un modelo donde las variables  $Y$  y  $X$  se han re-centrado con respecto a sus medias. En general, note que eliminar el intercepto tiene consecuencias importantes si no se realiza dicho proceso.

### Caso 1: homocedasticidad y no-correlación.

Partiendo de los supuestos

$$\mathbb{V}[\varepsilon_i] = \sigma^2 \quad \forall i, \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j.$$

En la suma los términos con  $i \neq j$  se anulan porque  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ . Solo se computan los términos diagonales  $i = j$ , donde  $\text{Cov}(\varepsilon_i, \varepsilon_i) = \mathbb{V}[\varepsilon_i] = \sigma^2$ :

$$\mathbb{V}[\hat{\beta}] = \frac{\sum_i x_i^2 \sigma^2}{(\sum_i x_i^2)^2} = \frac{\sigma^2 \sum_i x_i^2}{(\sum_i x_i^2)^2} = \frac{\sigma^2}{\sum_i x_i^2}.$$

**Estimando poblacional:**  $\frac{\sigma^2}{\sum_i x_i^2}$ .

**Estimador:**  $\sigma^2$  no es observable, pero puede estimarse de forma insesgada con los residuos MCO:

$$\hat{\sigma}^2 = \frac{\sum_i \hat{\varepsilon}_i^2}{N - 1}, \quad \hat{\mathbb{V}}[\hat{\beta}] = \frac{\hat{\sigma}^2}{\sum_i x_i^2}.$$

### Caso 2: heterocedasticidad, sin correlación entre individuos.

Se mantiene la no-correlación entre individuos distintos, pero ahora la varianza del error depende de  $i$ :

$$\mathbb{V}[\varepsilon_i] = \sigma_i^2, \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j.$$

Los términos fuera de la diagonal siguen siendo nulos. Los términos diagonales ahora contribuyen con  $\sigma_i^2$  distinto para cada  $i$ :

$$\mathbb{V}[\hat{\beta}] = \frac{\sum_i x_i^2 \mathbb{V}[\varepsilon_i]}{(\sum_i x_i^2)^2} = \frac{\sum_i x_i^2 \sigma_i^2}{(\sum_i x_i^2)^2}.$$

**Estimando poblacional:**  $\frac{\sum_i x_i^2 \sigma_i^2}{(\sum_i x_i^2)^2}$ .

**Estimador (White, 1980):**  $\sigma_i^2 = \mathbb{V}[\varepsilon_i]$  no es observable. Usando que  $\mathbb{E}[\varepsilon_i] = 0$ :

$$\mathbb{V}[\varepsilon_i] = \mathbb{E}[\varepsilon_i^2] - (\mathbb{E}[\varepsilon_i])^2 = \mathbb{E}[\varepsilon_i^2].$$

White (1980) demostró que el residuo al cuadrado  $\hat{\varepsilon}_i^2$  es un estimador consistente de  $\mathbb{E}[\varepsilon_i^2]$ , ya que  $\hat{\varepsilon}_i \xrightarrow{p} \varepsilon_i$  cuando  $N \rightarrow \infty$ . Sustituyendo  $\sigma_i^2$  por  $\hat{\varepsilon}_i^2$ :

$$\hat{\mathbb{V}}_{HC}[\hat{\beta}] = \frac{\sum_i x_i^2 \hat{\varepsilon}_i^2}{(\sum_i x_i^2)^2}.$$

Este estimador es consistente aunque no se cumpla la homocedasticidad.

### Caso 3: heterocedasticidad y correlación intra-colegio.

Se permite ahora que los errores de individuos en el mismo colegio  $g$  estén correlacionados:

$$\text{Cov}(\varepsilon_i, \varepsilon_j) \neq 0 \quad \text{si } i, j \in g, \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{si } i, j \text{ en colegios distintos.}$$

En la suma, los términos con  $i, j$  en colegios distintos se anulan; solo contribuyen los pares dentro del mismo colegio. Reorganizando la suma por clusters  $g \in \{1, \dots, G\}$ :

$$\mathbb{V}[\hat{\beta}] = \frac{\sum_g \sum_{i \in g} \sum_{j \in g} x_i x_j \text{Cov}(\varepsilon_i, \varepsilon_j)}{(\sum_i x_i^2)^2}.$$

Como  $\mathbb{E}[\varepsilon_i] = 0$ , se tiene  $\text{Cov}(\varepsilon_i, \varepsilon_j) = \mathbb{E}[\varepsilon_i \varepsilon_j]$ . Así, el numerador del cluster  $g$  puede reescribirse como, teniendo en cuenta  $S_g := \sum_{i \in g} x_i \varepsilon_i$ :

$$\sum_{i \in g} \sum_{j \in g} x_i x_j \mathbb{E}[\varepsilon_i \varepsilon_j] = \mathbb{E} \left[ \left( \sum_{i \in g} x_i \varepsilon_i \right)^2 \right] = \mathbb{E}[S_g^2].$$

Sustituyendo:

$$\mathbb{V}[\hat{\beta}] = \frac{\sum_g \mathbb{E}[S_g^2]}{\left(\sum_i x_i^2\right)^2}.$$

**Estimando poblacional:**  $\frac{\sum_g \mathbb{E}[S_g^2]}{\left(\sum_i x_i^2\right)^2}.$

**Estimador (Liang & Zeger, 1986):** Siguiendo la misma lógica que White,  $\mathbb{E}[S_g^2]$  puede estimarse por el cuadrado del residuo ponderado agregado por cluster,  $\hat{S}_g^2 = \left(\sum_{i \in g} x_i \hat{\varepsilon}_i\right)^2$ . Sustituyendo:

$$\hat{\mathbb{V}}_{CL}[\hat{\beta}] = \frac{\sum_g \left( \sum_{i \in g} x_i \hat{\varepsilon}_i \right)^2}{\left(\sum_i x_i^2\right)^2}.$$

Este estimador es consistente en  $G \rightarrow \infty$  y, al agregar los productos  $x_i \hat{\varepsilon}_i$  antes de elevar al cuadrado, captura automáticamente tanto la heterocedasticidad como la correlación dentro de cada colegio. Además, si cada individuo forma su propio cluster ( $n_g = 1$  para todo  $g$ ), la suma doble intra-cluster se agrupa a un solo término y  $\hat{\mathbb{V}}_{CL}$  se reduce al estimador HC de White.

3. Uno de sus colegas le indica que no es necesario tener en cuenta el cluster a nivel de colegio y le propone usar la siguiente expresión como estimador de la varianza en el **caso 3**:

$$\frac{\sum_i \sum_j x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j}{\left(\sum_i x_i\right)^2}.$$

- Utilizando sus conocimientos de teoría de regresión lineal, explique brevemente por qué este estimador es inválido.
- Compare el estimador que usted propuso para el **caso 3** con el que le propuso su colega. ¿Cómo soluciona su estimador el problema que identificó?

El estimador propuesto por el colega es:

$$\frac{\sum_i \sum_j x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j}{\left(\sum_i x_i\right)^2}.$$

El numerador puede reescribirse como  $(\sum_i x_i \hat{\varepsilon}_i)^2$ . Sin embargo, por la condición de primer orden del estimador MCO:  $\sum_i x_i \hat{\varepsilon}_i = 0$ . Por esto, el numerador es cero y el estimador entrega varianza igual a cero para cualquier muestra, lo que es incorrecto.

El estimador del inciso 2, caso 3 es:

$$\hat{\mathbb{V}}_{CL}[\hat{\beta}] = \frac{\sum_g \left( \sum_{i \in g} x_i \hat{\varepsilon}_i \right)^2}{\left(\sum_i x_i^2\right)^2}.$$

La diferencia es que la suma del numerador se realiza primero dentro de cada cluster  $g$  y luego se suman los cuadrados de esas sumas por cluster. Aunque  $\sum_g \sum_{i \in g} x_i \hat{\varepsilon}_i = 0$  (la suma total sigue siendo cero), los términos  $\sum_{i \in g} x_i \hat{\varepsilon}_i$  para cada colegio  $g$  son, en promedio, distintos de cero. El estimador clusterizado captura la variación de estos residuos ponderados a nivel de colegio, que refleja la correlación intra-cluster en los errores. Además, el denominador correcto es  $(\sum_i x_i^2)^2$ , no  $(\sum_i x_i)^2$ , ya que el denominador del estimador MCO es  $\sum_i x_i^2$ .

4. ¿Compare los estimadores robustos a heterocedasticidad (**caso 2**) y robustos a clusters (**caso 3**) que usted propuso. Para ello:

- Explique por qué el estimador robusto a clusters también es robusto a heterocedasticidad.

- ¿Es razonable esperar que la varianza estimada mediante el estimador robusto a clusters sea mayor que la obtenida con el estimador robusto a heterocedasticidad? Justifique su respuesta.

El estimador clusterizado es:

$$\hat{V}_{\text{CL}}[\hat{\beta}] = \frac{\sum_g \left( \sum_{i \in g} x_i \hat{\varepsilon}_i \right)^2}{(\sum_i x_i^2)^2}.$$

Para cada cluster  $g$ , expandiendo el cuadrado:

$$\left( \sum_{i \in g} x_i \hat{\varepsilon}_i \right)^2 = \sum_{i \in g} x_i^2 \hat{\varepsilon}_i^2 + \sum_{\substack{i,j \in g \\ i \neq j}} x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j.$$

Sumando sobre todos los colegios  $g$ :

$$\sum_g \left( \sum_{i \in g} x_i \hat{\varepsilon}_i \right)^2 = \underbrace{\sum_i x_i^2 \hat{\varepsilon}_i^2}_{\text{términos diagonales}} + \underbrace{\sum_g \sum_{\substack{i,j \in g \\ i \neq j}} x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j}_{\text{productos cruzados intra-colegio}}.$$

Dividiendo por  $(\sum_i x_i^2)^2$  se obtiene:

$$\begin{aligned} \hat{V}_{\text{CL}}[\hat{\beta}] &= \frac{\sum_i x_i^2 \hat{\varepsilon}_i^2}{(\sum_i x_i^2)^2} + \frac{\sum_g \sum_{\substack{i,j \in g \\ i \neq j}} x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j}{(\sum_i x_i^2)^2}. \\ &= \hat{V}_{\text{HC}}[\hat{\beta}] \end{aligned} \quad (*)$$

El primer término es exactamente el estimador HC de White. El segundo captura los productos cruzados intra-colegio.

Por otro lado, ser robusto a heterocedasticidad significa que  $\hat{V}_{\text{CL}}$  converge en probabilidad a la varianza verdadera de  $\hat{\beta}$  incluso cuando  $\mathbb{V}[\varepsilon_i] = \sigma_i^2$  varía entre individuos.

La varianza verdadera del estimador MCO bajo heterocedasticidad sin correlación entre individuos es (Caso 2):

$$\mathbb{V}[\hat{\beta}] = \frac{\sum_i x_i^2 \sigma_i^2}{(\sum_i x_i^2)^2}.$$

Tomamos el comportamiento asintótico de cada sumando:

Según White (1980),  $\hat{\varepsilon}_i^2 \xrightarrow{p} \varepsilon_i^2$  ya que  $\hat{\varepsilon}_i \xrightarrow{p} \varepsilon_i$  conforme  $N \rightarrow \infty$ . Como  $\mathbb{E}[\varepsilon_i] = 0$ , se tiene  $\mathbb{E}[\varepsilon_i^2] = \mathbb{V}[\varepsilon_i] = \sigma_i^2$ . Por tanto:

$$\frac{\sum_i x_i^2 \hat{\varepsilon}_i^2}{(\sum_i x_i^2)^2} \xrightarrow{p} \frac{\sum_i x_i^2 \sigma_i^2}{(\sum_i x_i^2)^2} = \mathbb{V}[\hat{\beta}].$$

Bajo solo heterocedasticidad (sin correlación intra-colegio),  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$  para todo  $i \neq j$ , incluso si pertenecen al mismo colegio. Como  $\mathbb{E}[\varepsilon_i] = 0$ , se tiene  $\text{Cov}(\varepsilon_i, \varepsilon_j) = \mathbb{E}[\varepsilon_i \varepsilon_j]$ . Por tanto:

$$\mathbb{E} \left[ \sum_g \sum_{\substack{i,j \in g \\ i \neq j}} x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j \right] \approx \sum_g \sum_{\substack{i,j \in g \\ i \neq j}} x_i x_j \underbrace{\mathbb{E}[\varepsilon_i \varepsilon_j]}_{=\text{Cov}(\varepsilon_i, \varepsilon_j)=0} = 0.$$

Combinando ambos resultados:

$$\hat{V}_{\text{CL}}[\hat{\beta}] \xrightarrow{p} \mathbb{V}[\hat{\beta}] + 0 = \mathbb{V}[\hat{\beta}],$$

lo que demuestra que  $\hat{V}_{\text{CL}}$  es consistente bajo heterocedasticidad.

Si cada observación forma su propio cluster ( $n_g = 1$  para todo  $g$ ), no existen pares  $i \neq j$  dentro de ningún cluster. Los términos cruzados desaparecen identicamente:

$$\sum_g \sum_{\substack{i,j \in g \\ i \neq j}} x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j = 0 \quad (\text{no hay pares intra-cluster}).$$

Por tanto:

$$\hat{\mathbb{V}}_{\text{CL}}^{n_g=1}[\hat{\beta}] = \frac{\sum_i x_i^2 \hat{\varepsilon}_i^2}{(\sum_i x_i^2)^2} = \hat{\mathbb{V}}_{\text{HC}}[\hat{\beta}].$$

El estimador HC es el caso particular del clusterizado cuando el nivel de agrupación es el individuo.

La diferencia entre ambos estimadores es:

$$\hat{\mathbb{V}}_{\text{CL}} - \hat{\mathbb{V}}_{\text{HC}} = \frac{\sum_g \sum_{\substack{i,j \in g \\ i \neq j}} x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j}{(\sum_i x_i^2)^2}.$$

Si la correlación intra-colegio es positiva ( $\text{Cov}(\varepsilon_i, \varepsilon_j) > 0$  para  $i \neq j$  en el mismo colegio), los residuos de compañeros de colegio tienden a tener el mismo signo. El producto  $\hat{\varepsilon}_i \hat{\varepsilon}_j > 0$  en promedio, y dado que  $x_i, x_j > 0$ , porque la variable de tratamiento es un indicador, los términos cruzados son positivos en promedio, haciendo  $\hat{\mathbb{V}}_{\text{CL}} > \hat{\mathbb{V}}_{\text{HC}}$ .

En este caso, los estudiantes del mismo colegio comparten entorno institucional, recursos pedagógicos, calidad del director y composición de pares. Estas características comunes generan factores no observados compartidos que se trasladan al término de error, induciendo correlación intra-colegio positiva. Ignorar lo anterior y usar solo  $\hat{\mathbb{V}}_{\text{HC}}$  produce errores estándar artificialmente pequeños, estadísticos  $t$  inflados e intervalos de confianza demasiado estrechos, lo que lleva a rechazar  $H_0$  con mayor frecuencia que el nivel nominal.

Habiendo adquirido una mayor claridad sobre la teoría detrás de los errores estándar clusterizados, usted procede a intentar replicar algunos de los resultados del trabajo de [Gershenson et al. \(2022\)](#), integrando los conocimientos sobre errores estándar clusterizados desarrollados en los incisos anteriores.

En particular, recuerde que el experimento se llevó a cabo sobre estudiantes de kínder de raza negra y que la exposición a profesores de distinta raza fue aleatorizada *dentro de cada colegio*, asignando estudiantes y docentes a diferentes salones.<sup>2</sup> El experimento se implementó en múltiples condados y estados de E.E.U.U.

Bajo este diseño, los estudiantes de kínder de raza negra fueron expuestos de manera aleatoria a profesores de la misma raza o de razas distintas *condicional en el colegio*. En consecuencia, haciendo uso de sus conocimientos sobre diseños experimentales, usted estima la siguiente ecuación:

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i. \tag{4}$$

5. Usando el caso estudio, identifique qué factores pueden generar una mayor varianza estimada por parte de los errores estándar clusterizados a nivel de colegio. Para ello:

- Usando el estimador que usted propuso en el **caso 3 inciso 3**, enuncie qué mecanismo matemáticos incrementan la varianza estimada.
- Por cada uno de los mecanismos hallados, de un ejemplo concreto usando el contexto del caso estudio que lo respalde.
- Estime el modelo (4) utilizando como variable dependiente: i) el puntaje en matemáticas del SAT y ii) la probabilidad de ingresar a la universidad. Reporte los resultados en una tabla, incluyendo errores estándar robustos a heterocedasticidad y errores estándar clusterizados a nivel de colegio.

<sup>2</sup>Una ilustración intuitiva de este contexto puede verse en la serie *Abbott Elementary*, que retrata un colegio con una población estudiantil predominantemente negra cuyos docentes pertenecen a distintas razas.

Del estimador  $\hat{V}_{CL}[\hat{\beta}] = \sum_g \left( \sum_{i \in g} x_i \hat{\varepsilon}_i \right)^2 / (\sum_i x_i^2)^2$ , tres factores amplían la varianza estimada respecto a la robusta a heterocedasticidad:

- Colegios con muestra grande ( $n_g$  elevado): el residuo ponderado agregado del cluster  $\sum_{i \in g} x_i \hat{\varepsilon}_i$  acumula más términos, lo que eleva su magnitud. En el estudio, colegios con muchos estudiantes negros en la muestra contribuirán más a la varianza estimada.
- Alta correlación intra-colegio en los errores: si los factores no observados que afectan los puntajes SAT son similares dentro de un colegio (mismos recursos, misma dirección, mismas redes), los productos cruzados  $x_i x_j \hat{\varepsilon}_i \hat{\varepsilon}_j$  dentro del colegio serán positivos y amplificarán el residuo ponderado agregado.
- Varianza del regresor dentro del colegio: si los valores de  $x_i = D_i - \bar{D}$  son grandes en magnitud dentro del colegio (alta dispersión del tratamiento dentro del colegio), los residuos ponderados individuales  $x_i \hat{\varepsilon}_i$  son mayores.

Por otra parte, la siguiente tabla reporta los resultados de la estimación del modelo (4) para los dos resultados de interés.

	SAT Matemáticas	Asistencia universitaria
Tiene prof. negro	4.6072***	0.1980***
(SE Robusto HC1)	(0.0640)	(0.0030)
[SE Cluster colegio]	[0.0649]	[0.0044]
N	99,204	99,204

\*\*\* p<0.01. SE robustos entre paréntesis; SE clusterizados por colegio entre corchetes.

Para el puntaje SAT en matemáticas, tener un profesor negro en kínder se asocia con un incremento de 4.61 puntos ( $p < 0.01$ ). El error estándar robusto es 0.0640 y el clusterizado a nivel de colegio es 0.0649, diferencia que sugiere una correlación intra-colegio en los errores baja pero positiva. Para la asistencia universitaria, el efecto es 19.8 puntos porcentuales ( $p < 0.01$ ). El error estándar clusterizado (0.0044) es mayor al robusto (0.0030), lo que indica mayor correlación intra-colegio en este resultado. Ambos efectos son estadísticamente significativos bajo los dos tipos de error estándar, lo que indica que los resultados son robustos a la elección del estimador de varianza.

Ustedes envían sus resultados para consideración en una revista de alto prestigio académico. Tras un tiempo, reciben dos comentarios por parte de los *referees*, es decir, los evaluadores encargados de determinar si la contribución del artículo es publicable. En particular, los comentarios reflejan posiciones contradictorias respecto de la decisión de reportar errores estándar clusterizados a nivel de colegio:

#### Referee #1

“La decisión de clusterizar los errores estándar a nivel de colegio está bien motivada y es consistente con la literatura metodológica existente. Dado el diseño experimental y la evidencia empírica que presentan sobre la correlación intra-colegio, este nivel de clusterización parece apropiado, y los resultados son robustos bajo esta especificación.”

#### Referee #2

“No queda claro por qué los autores se limitan a clusterizar los errores estándar a nivel de colegio cuando disponen de información que permitiría hacerlo a niveles más agregados, como el condado o el estado. En muestras grandes como la suya, aumentar el nivel de agregación rara vez es problemático y puede ofrecer una inferencia más conservadora. Como regla práctica, recomendaría siempre reportar errores estándar clusterizados al nivel más agregado, en lugar de decidir ex-ante si se debe clusterizar y evitar pensar en la existencia de un nivel apropiado.”

6. ¿Qué posible impacto puede tener un nivel de agregación más grande al clusterizar sobre la varianza estimada? Use como guía los siguientes puntos:

- Use el estimador de la varianza que usted propuso en el **inciso 3, caso 3** y compare la varianza teórica que se obtiene al clusterizar por colegio vs por condado.
- Repita las estimaciones del **inciso 5**, concentrándose únicamente en la variable  $Y_i$  correspondiente al puntaje de matemáticas en la prueba SAT. Reporte los coeficientes estimados junto con errores estándar robustos a heterocedasticidad y errores estándar clusterizados a nivel de: i) salón, ii) colegio, iii) condado y iv) estado.
- Interprete los resultados obtenidos. ¿Las estimaciones son consistentes con lo que usted encuentra teóricamente?

Al cambiar de clusterizar por colegio a clusterizar por condado, los clusters se vuelven más grandes (agrupan varios colegios). El estimador de varianza agrega los residuos ponderados de todos los colegios dentro de cada condado antes de cuadrarlos. Si hay correlación positiva entre los residuos ponderados de colegios del mismo condado (lo que ocurre cuando colegios del mismo condado comparten características no observadas), la varianza clusterizada por condado puede ser mayor que la de colegio. Sin embargo, el número de clusters cae (de 1,000 colegios a 200 condados o 50 estados), lo que reduce la precisión del estimador de varianza clusterizado y puede producir resultados no monótonos.

Así, la siguiente tabla muestra los errores estándar del efecto de tener un profesor negro sobre el puntaje SAT bajo distintos niveles de agrupación.

	SAT Matemáticas
Tiene prof. negro	4.6072***
(SE Robusto HC1)	(0.0640)
[SE Cluster colegio]	[0.0649]
[SE Cluster condado]	[0.0667]
[SE Cluster estado]	[0.0598]
N	99,204

\*\*\* p<0.01. SE robustos entre paréntesis; SE clusterizados entre corchetes.

El coeficiente estimado es  $\hat{\beta}_1 = 4.61$  en todos los casos (el estimador MCO no cambia con el tipo de error estándar). Los errores estándar son: HC1 = 0.0640, colegio = 0.0649, condado = 0.0667 y estado = 0.0598.

Los errores estándar clusterizados por colegio y condado son ligeramente mayores al HC1, lo que es consistente con la presencia de correlación intra-cluster positiva en esos niveles. Sin embargo, el error estándar a nivel de estado es menor al de colegio o condado, lo que puede explicarse por el bajo número de clusters (50 estados): con pocos clusters, el estimador de varianza clusterizado tiene alta varianza propia y la corrección de muestra finita puede producir resultados irregulares. Esto muestra que la varianza clusterizada no aumenta de forma monótona con el nivel de agregación: el número de clusters y la correlación intra-cluster a cada nivel determinan conjuntamente el resultado.

7. Usted decide escribirle a su amiga cercana **Susan Athey** para pedirle su opinión sobre el comentario hecho por el referee #2. Ella le envía el siguiente código de *R*:

```

1 # Housekeeping
2 #-----
3
4 # Semilla para replicación.
5 set.seed(31416)
6
7 # Cargar las librerías necesarias.
8 library(sandwich)
9 library(lmtest)
10
11
12 # Parámetros de la simulación
13 #-----
14
15 # Número de individuos en la población.
16 n_population <- 1e7
17
18 # Número de colegios.
19 schools <- 100
20
21 # Número de estudiantes por colegio

```

```

22 | stud_by_school <- n_population / schools
23 |
24 | # Probabilidad de aparecer en la muestra.
25 | p_sample <- 0.01
26 |
27 | # Número de simulaciones.
28 | n_sim <- 500
29 |
30 |
31 | # Construir base de datos poblacional
32 | #-----
33 |
34 | # Base para la construcción.
35 | population <- data.frame(id = 1:n_population)
36 |
37 | # Determinar el tratamiento.
38 | population$D <- rbinom(n_population, size = 1, prob = 0.5)
39 |
40 | # Determinar un colegio para cada individuo.
41 | population$schools <- rep(1:schools, each = stud_by_school)
42 |
43 | # Determinar el efecto por colegio.
44 | population$school_effect <- c(rep(1, n_population / 2), rep(-1, n_population / 2))
45 |
46 | # Construir la variable de resultado.
47 | population$Y <- population$school_effect * population$D + runif(n_population)

```

Listing 1: Código de Susan Athey

Usando ese código, realice el siguiente procedimiento:

- Tome 1'000 muestras de la población de interés y estime el modelo (4) usando errores estándar robustos a heterocedasticidad y clusterizados a nivel de colegio.<sup>3</sup>
- Para cada una de las muestras, determine el intervalo de confianza al 95 % usando ambos errores estándar.
- Para cada una de las muestras, determine si el intervalo de confianza “cubren” el cero (es decir, el intervalo de confianza incluye el cero).<sup>4</sup>
- Determine los errores estándar robustos a heterocedasticidad y clusterizados promedio.
- Determine el porcentaje de veces que los intervalos de confianza “cubren” el cero.
- Reporte sus resultados completando la siguiente tabla:

Heterocedasticidad		Cluster	
Ŷ Promedio	% Cobertura	Ŷ Promedio	% Cobertura
0.000023	94.7 %	0.010124	100.0 %

Tabla 1: Cobertura teórica de heterocedasticidad vs clusters

- Interprete sus resultados. ¿Es cierto que siempre es mejor clusterizar al nivel más agregado?

<sup>3</sup>Pista: Se recomienda usar los comandos vcovHC y vcovCL del paquete *sandwich* para optimizar los resultados.

<sup>4</sup>Note que, en la simulación, el efecto esperado es cero por construcción.

La simulación toma 1,000 muestras del 1 % de una población de 10 millones de estudiantes distribuidos en 100 colegios. Los primeros 50 colegios tienen un efecto de tratamiento de +1 y los últimos 50 de -1; el efecto promedio del tratamiento es cero por construcción. La Tabla ?? reporta los resultados.

Heterocedasticidad		Cluster	
$\hat{V}$ Promedio	% Cobertura	$\hat{V}$ Promedio	% Cobertura
0.000023	94.7 %	0.010124	100.0 %

El error estándar robusto a heterocedasticidad presenta una varianza promedio de  $\hat{V}_{HC} = 0.000023$  ( $SE \approx 0.005$ ) y una tasa de cobertura del 94.7 %, próxima al nivel nominal del 95 %. El estimador clusterizado a nivel de colegio presenta una varianza promedio mucho mayor,  $\hat{V}_{CL} = 0.010124$  ( $SE \approx 0.101$ ), y una cobertura del 100 %, lo que indica sobrecobertura.

El proceso generador de los datos de la simulación genera heterogeneidad en el efecto del tratamiento entre colegios ( $school\_effect_g = \pm 1$ ), pero no correlación en los errores dentro de un colegio. Dentro de cada colegio, los tratamientos son asignados de forma independiente entre estudiantes, lo que hace que la covarianza intra-cluster de los errores sea nula. El estimador robusto a heterocedasticidad captura correctamente esta estructura y produce cobertura nominal. El estimador clusterizado, en cambio, captura los patrones sistemáticos entre colegios (el hecho de que todos los estudiantes del mismo colegio comparten el mismo efecto) e infla la varianza estimada de forma sustancial, produciendo intervalos de confianza innecesariamente amplios. Esto no es cierto que siempre sea mejor clusterizar al nivel más agregado: cuando la correlación intra-cluster de los errores es baja o inexistente, clusterizar penaliza la inferencia sin ganar validez.

8. ¿Considera usted que el referee #2 tiene razón? Escriba una breve respuesta a este comentario teniendo en cuenta los resultados del inciso anterior. Asegúrese de usar la tabla que construyó (máximo 200 palabras).

La recomendación del árbitro #2 no es correcta. La simulación muestra que clusterizar a nivel de colegio, cuando la fuente de variación es heterogeneidad en el efecto del tratamiento entre colegios y no correlación intra-cluster en los errores, produce sobrecobertura del 100 % frente al 94.7 % del estimador robusto a heterocedasticidad. La varianza estimada con cluster es 440 veces mayor que la robusta. En términos prácticos, aplicar la regla del árbitro llevaría a reportar errores estándar inflados, reducir el poder estadístico para detectar efectos verdaderos y, en experimentos como el de Gershenson et al. (2022) donde el tratamiento se aleatorizó dentro de colegios, reportar inferencia excesivamente conservadora.

La decisión de clusterizar debe estar motivada por el diseño del estudio y la presencia efectiva de correlación intra-cluster en los errores, no por una regla de maximizar el nivel de agregación. Abadie et al. (2022) establecen que la clusterización es apropiada cuando el proceso que genera los datos produce correlación entre las observaciones dentro de un grupo. En este experimento, dado que la aleatorización fue dentro de cada colegio, clusterizar a nivel de colegio es la elección justificada por el diseño; clusterizar a niveles más agregados, sin evidencia de correlación a esos niveles, solo infla los errores sin ganar validez de la inferencia.

## Tercer Ejercicio

Los experimentos aleatorizados controlados (RCTs, por sus siglas en inglés) se han consolidado como el estándar para la evaluación de políticas públicas y programas de desarrollo. A diferencia de los estudios observacionales, los RCTs permiten establecer relaciones causales mediante la aleatorización del tratamiento, lo cual elimina el sesgo de selección y garantiza que las diferencias observadas entre grupos sean atribuibles únicamente a la intervención evaluada. No obstante, la credibilidad de estos estudios depende de manera crítica de la **replicabilidad**: la capacidad de otros investigadores de reproducir los resultados utilizando los mismos datos y métodos.

El estudio de **Dupas et al. (2025)**, publicado en la *American Economic Review*, implementa un RCT a gran escala en Burkina Faso con 14,545 mujeres distribuidas en 499 aldeas, 100 centros de salud y 20 provincias, empleando múltiples niveles de aleatorización y estratificación. Los resultados desafían décadas de políticas públicas: **el acceso completamente gratuito a anticonceptivos no redujo la fecundidad ni incrementó su uso**. Este hallazgo nulo tiene profundas implicaciones para programas de desarrollo que representan miles de millones de dólares en inversión. Dada la naturaleza contraintuitiva de estos resultados, resulta particularmente valioso comprender cómo los investigadores garantizaron la robustez de sus conclusiones, desde el diseño experimental hasta el análisis de datos. Este ejercicio le permitirá verificar la replicabilidad de los resultados principales del estudio de **Dupas et al. (2025)**. Para ello, deberá familiarizarse tanto con el artículo como con su paquete de replicación. Se recomienda enfáticamente la lectura del texto original para facilitar la solución de este ejercicio, a la hora de evaluar sus respuestas vamos a buscar evidencia de lo anterior.

**Datos disponibles:** Para este ejercicio utilizará muestras aleatorias del 60 % de los datos originales, porque el resto se perdieron (vamos a asumir esto para el ejercicio):

- `baseline_sample.dta`: Muestra del 60 % de la encuesta de línea base (primavera 2018).
- `endline_sample.dta`: Muestra emparejada del 60 % de la encuesta de seguimiento (primavera 2021). Menor N que `baseline` debido a atrición.
- `listing_sample.dta`: Censo de hogares en las aldeas incluidas en la muestra.
- Puede usar el `codebook.xlsx` del paquete de replicación para identificar las variables.

1. **Diseño experimental:** Antes de poder verificar los resultados de un estudio, es fundamental entender cómo fue diseñado e implementado el experimento. El diseño experimental no solo determina qué preguntas pueden responderse, sino también qué supuestos son necesarios para la identificación causal y qué amenazas a la validez interna deben considerarse.

Describa el diseño experimental del estudio de Dupas et al. (2025). Explique cómo fue implementada la aleatorización (incluyendo los diferentes niveles y la estratificación), qué tratamientos fueron aleatorizados y los principales desafíos metodológicos que enfrentaron los autores durante la implementación del estudio. (máx. 300 palabras)

El estudio de Dupas et al. (2025) evaluó si el acceso gratuito a anticonceptivos reduce la fecundidad en Burkina Faso. El experimento incluyó 14,545 mujeres en 499 aldeas, distribuidas en 100 centros de salud y 20 provincias. La aleatorización operó en dos niveles. Primero, los centros de salud se asignaron al azar a una de dos condiciones: subsidio completo, es decir 100 % del costo cubierto mediante vouchers o subsidio parcial, 10 % del costo cubierto, grupo de control. Esta asignación se realizó de forma estratificada por provincia, garantizando representación en las 20 provincias. Segundo, dentro de cada centro de salud, las mujeres elegibles fueron escogidas durante la encuesta de línea base (primavera 2018) e inscritas en el estudio. El seguimiento se realizó aproximadamente tres años después (primavera 2021). El tratamiento principal fue el nivel de subsidio al costo de los anticonceptivos modernos es decir subsidio del 100 % (tratamiento) versus subsidio del 10 % (control). La elección del 10 % como umbral del grupo de control, en lugar de cero, respondió a la decisión ética de garantizar que todas las participantes tuvieran acceso mínimo a anticonceptivos. Los autores identificaron tres desafíos principales. Primero, en julio de 2020, el gobierno de Burkina Faso anunció la gratuidad nacional de los anticonceptivos, lo que representa una posible violación del supuesto SUTVA. Segundo, la pérdida de seguimiento entre línea base y seguimiento fue del 13.8 %, con diferencia entre grupos (15.0 % en tratamiento vs. 12.6 % en control). Tercero, el diseño implicó múltiples variables de resultado, lo que exige cautela para evitar interpretaciones derivadas de comparaciones múltiples.

## 2. Marco de Resultados Potenciales:

- a) Defina formalmente la variable de asignación al tratamiento de subsidio ( $D_i$ ) donde  $i$  indexa a cada mujer. Presente los posibles valores que puede tomar esta variable.

La variable de asignación al tratamiento se define como:

$$D_i = \begin{cases} 1 & \text{si la mujer } i \text{ fue asignada al centro de salud con subsidio del 100 \%} \\ 0 & \text{si la mujer } i \text{ fue asignada al centro de salud con subsidio del 10 \%} \end{cases}$$

donde  $i$  refiere a cada mujer participante en el estudio. La variable toma dos valores posibles:  $D_i \in \{0, 1\}$ . La asignación fue determinada a nivel de centro de salud (no a nivel individual): todas las mujeres reclutadas en un mismo centro de salud recibieron el mismo nivel de subsidio. En la muestra de línea base,  $N_1 = 4,313$  mujeres pertenecen al grupo tratado ( $D_i = 1$ ) y  $N_0 = 4,376$  al grupo de control ( $D_i = 0$ ), para un total de  $N = 8,689$  mujeres.

- b) Defina los resultados potenciales para la variable de interés principal:  $Y_i$  = tuvo al menos un nacimiento durante los 3 años del estudio. ¿Cuántos resultados potenciales existen para cada mujer?

Los resultados potenciales para la mujer  $i$  se definen como:

- $Y_i(1)$ : indicador de si la mujer  $i$  tuvo al menos un nacimiento entre la línea base (abril 2018) y el seguimiento (junio 2021), dado que fue asignada al subsidio del 100 %.
- $Y_i(0)$ : indicador de si la mujer  $i$  tuvo al menos un nacimiento en el mismo período, dado que fue asignada al subsidio del 10 %.

Para cada mujer existen dos resultados potenciales:  $Y_i(1)$  y  $Y_i(0)$ . El principal problema de inferencia causal implica que solo uno de ellos es observable: si  $D_i = 1$  se observa  $Y_i = Y_i(1)$  y  $Y_i(0)$  es contrafactual; si  $D_i = 0$  se observa  $Y_i = Y_i(0)$  y  $Y_i(1)$  queda sin observar.

El resultado observado se escribe  $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ .

- c) Los autores decidieron que el grupo “control” recibiera un subsidio del 10 % en lugar de no recibir nada. ¿Por qué esta decisión tiene sentido desde una perspectiva ética? ¿Qué parámetro causal están estimando entonces? (máx. 100 palabras)

La decisión de usar un subsidio del 10 % como grupo de control, en lugar de no ofrecer ningún subsidio, responde a una razón ética: dado que los anticonceptivos modernos tienen beneficios de salud documentados, privar completamente de acceso a las participantes del grupo de control habría sido éticamente cuestionable. El subsidio del 10 % garantiza un nivel mínimo de acceso para todas las mujeres.

Esta decisión tiene consecuencias directas sobre el parámetro causal que se estima. El parámetro identificado es el efecto ITT del subsidio completo relativo al subsidio parcial  $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$ , donde  $Y_i(1)$  y  $Y_i(0)$  corresponden al resultado bajo subsidio del 100 % y del 10 %, respectivamente. El experimento no identifica el efecto del acceso a anticonceptivos comparado con ningún acceso (el ATE absoluto frente a un grupo sin ningún subsidio).

## 3. Supuestos de Identificación:

- a) Liste los tres supuestos principales necesarios para identificar el efecto causal del subsidio completo sobre fecundidad. Incluya las expresiones matemáticas para aquellos supuestos que sea posible.

- 1) Independencia (aleatorización). La asignación al tratamiento es independiente de los resultados potenciales:

$$\{Y_i(0), Y_i(1)\} \perp D_i.$$

En este estudio, la aleatorización a nivel de centro de salud garantiza este supuesto dentro del diseño experimental.

- 2) SUTVA (Stable Unit Treatment Value Assumption). Existen dos componentes: (i) no hay interferencia entre unidades: el resultado de la mujer  $i$  no depende del tratamiento asignado a otras mujeres; (ii) hay un único valor del tratamiento:  $D_i = 1$  implica siempre el subsidio completo (sin versiones múltiples del tratamiento).

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

- 3) Cumplimiento completo / intención de tratar. Todos los centros de salud asignados al tratamiento implementaron efectivamente el subsidio del 100 %, y los del control el 10 %. Bajo este supuesto, el ITT coincide con el efecto del tratamiento asignado. Si hubiera incumplimiento (por ejemplo, centros de control que accedieran al subsidio completo), el ITT subestimaría el efecto causal del tratamiento en los cumplidores.

- b) La aleatorización fue estratificada por provincia. Escriba matemáticamente el supuesto de independencia condicional apropiado para este diseño.

Dado que la aleatorización se realizó dentro de cada provincia, el supuesto de independencia pertinente es el de independencia condicional. Sea  $P_i \in \{1, 2, \dots, 20\}$  la provincia a la que pertenece la mujer  $i$ . El supuesto se escribe:

$$\{Y_i(0), Y_i(1)\} \perp D_i \mid P_i.$$

Esto significa que, dentro de cada provincia, la asignación de centros de salud al tratamiento es independiente de los resultados potenciales. La estratificación garantiza que en cada estrato (provincia) haya tanto centros de tratamiento como de control, lo que elimina la posibilidad de que diferencias sistemáticas entre provincias confundan la estimación. En la práctica, este supuesto se satisface por diseño: la aleatorización se realizó de forma estratificada por las 20 provincias, asegurando representación equilibrada dentro de cada una. La ecuación de regresión incorpora esta estratificación mediante efectos fijos de provincia  $\gamma_p$ .

#### 4. Replicación de Balance Muestral - Tabla A.1:

Los autores reportan pruebas de balance en la Tabla A.1 del Apéndice. Su tarea es replicar el Panel A (características de las mujeres) utilizando `baseline_sample.dta`, una muestra aleatoria del 60 % de los datos originales.<sup>5</sup>

- a) Construya las variables analíticas necesarias para replicar las siguientes variables de la Tabla A.1:
- Edad de la mujer
  - La mujer reporta que su esposo es polígamico
  - Número total de hijos deseados por la mujer
  - Uso actual de anticoncepción moderna
  - La necesidad de la mujer por anticonceptivos está insatisfecha
  - No podría pagar por anticonceptivos si quisiera usarlos
  - El hogar tiene radio

Las variables analíticas se construyeron a partir de `baseline_sample.dta`. Todas las variables se usan sin transformación adicional. Las observaciones con valores perdidos se excluyen a nivel de cada regresión individual, lo que explica la variación en el tamaño muestral entre variables.

- b) Para cada una de las siete variables, estime una regresión de balance de la forma:

$$X_i = \alpha + \beta \cdot \text{Tratamiento}_i + \gamma_{\text{provincia}} + \varepsilon_i,$$

<sup>5</sup>Sus resultados serán similares pero no idénticos a los del paper debido al tamaño muestral.

con errores estándar agrupados (clusterizados) a nivel de centro de salud.

Para cada variable reporte: (i) el coeficiente del tratamiento, (ii) el error estándar entre paréntesis, (iii) el valor-p, (iv) la media del grupo control y (v) el tamaño de muestra (N).

Calcule además el estadístico F conjunto (y su valor-p) para la hipótesis de que todos los coeficientes son iguales a cero.

Los resultados de las regresiones de balance se presentan a continuación:

Variable	Coef.	(SE)	p-valor	Media control	N
Edad de la mujer	0.156	(0.150)	0.300	28.26	8,687
Esposo polígamo	0.030	(0.020)	0.139	0.437	8,689
Hijos deseados	0.056	(0.067)	0.398	5.969	7,855
Uso anticonceptivo moderno	-0.013	(0.014)	0.343	0.324	8,680
Necesidad insatisfecha	0.004	(0.015)	0.782	0.385	8,682
No podría pagar	0.032*	(0.017)	0.065	0.381	7,873
El hogar tiene radio	0.033	(0.020)	0.102	0.478	8,686

F conjunto = 1.705 (p-valor = 0.117). \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

SE clusterizados a nivel de CSPS. Efectos fijos de provincia incluidos.

Para cada variable se estimó la regresión  $X_i = \alpha + \beta \cdot D_i + \gamma_{\text{provincia}} + \varepsilon_i$  con errores estándar clusterizados a nivel de centro de salud (CSPS). La tabla reporta el coeficiente del tratamiento, el error estándar entre paréntesis, el valor-p, la media del grupo de control y el tamaño muestral. El estadístico F conjunto para la hipótesis nula de que todos los coeficientes son cero es  $F = 1.705$  (p-valor = 0.117).

- c) Compare sus resultados con la Tabla A.1 del paper. ¿Hay evidencia de desbalance entre tratamiento y control? ¿Qué concluye sobre la correcta implementación de la aleatorización? (máx. 150 palabras)

Los resultados de la prueba de balance no muestran evidencia de desbalance sistemático entre los grupos de tratamiento y control. El estadístico F conjunto es  $F = 1.705$  con un p-valor de 0.117, por lo que no se rechaza la hipótesis nula de que todos los coeficientes son simultáneamente iguales a cero al nivel convencional del 5 %.

A nivel individual, solo la variable “No podría pagar por anticonceptivos” presenta un coeficiente marginalmente significativo ( $\hat{\beta} = 0.032$ , p-valor = 0.065) al 10 %, lo que indica que el grupo tratado reporta una probabilidad ligeramente mayor de no poder pagar en la línea de base. Dado que se trata de una variable de línea de base y que la diferencia se observa en apenas una de las siete variables con significancia marginal, puede atribuirse a variación muestral aleatoria.

Estos resultados son coherentes con la estratificación por provincia y con el proceso de aleatorización a nivel de centro de salud. La comparación con la Tabla A.1 del artículo original muestra patrones similares (ninguna diferencia estadísticamente significativa al 5 %), lo que confirma que la submuestra del 60 % preserva el balance del diseño experimental. Se concluye que la aleatorización fue implementada correctamente.

- d) Analice las estadísticas descriptivas de la muestra. ¿Qué nos dicen sobre la población estudiada y sobre la plausibilidad y dirección del efecto esperado del tratamiento? Interprete. (máx. 200 palabras).

Las estadísticas descriptivas de la muestra se presentan a continuación:

Variable	N	Media	DE	Media ctrl
Edad de la mujer	8,687	28.281	5.452	28.256
Esposo poligamo	8,689	0.454	0.498	0.437
Numero total de hijos deseados	7,855	6.016	1.885	5.969
Uso actual de anticoncepcion moderna	8,680	0.314	0.464	0.324
Necesidad insatisfecha de anticonceptivos	8,682	0.387	0.487	0.385
No podria pagar por anticonceptivos	7,873	0.403	0.491	0.381
El hogar tiene radio	8,686	0.482	0.500	0.478

Muestra de línea base (submuestra del 60 %). DE = desviación estándar. Media ctrl = media del grupo de subsidio 10 %. Media trat =

La muestra corresponde a mujeres rurales en Burkina Faso con una edad promedio de 28.3 años, es decir, en plena edad reproductiva. El 45.4 % reporta que su esposo es polígamo,

estructura familiar que en contextos de alta fecundidad está asociada a presiones sociales hacia el embarazo. El número promedio de hijos deseados es de 6.0 lo que refleja normas pro-natalistas inherentes. El 31.4 % usa actualmente anticonceptivos modernos, el 38.7 % tiene una necesidad insatisfecha (no desea embarazarse pero no usa método alguno), y el 40.3 % declara que no podría pagar por anticonceptivos. El 48.2 % de los hogares tiene radio, indicador de acceso a medios de comunicación y desarrollo.

##### 5. Replicación de Resultados Principales - Tabla 2:

Use `baseline_sample.dta` y `endline_sample.dta`. Tenga en cuenta que el `endline` tiene menos observaciones debido a la atrición de las participantes.

- a) Construya las variables de resultado, realice el merge entre bases y reporte la tasa de atrición (total y por grupo de tratamiento). Reporte la atrición como tasas (%) y como número de observaciones (N).

La base de `endline` se emparejó con la línea base usando el identificador de mujer. Las observaciones presentes en la línea base pero ausentes en el seguimiento se clasifican como atrición. Los resultados son:

Grupo	Atrición (N)	Tasa de atrición (%)
Tratamiento (100 %)	649	15.04 %
Control (10 %)	551	12.59 %
<b>Total</b>	<b>1,200</b>	<b>13.81 %</b>

La muestra de línea base contiene  $N = 8,689$  mujeres (4,313 en tratamiento, 4,376 en control). Tras el merge, la muestra de seguimiento disponible es de aproximadamente 7,489 observaciones para el resultado principal (nacimientos). La tasa de atrición diferencial entre grupos (15.0 % en tratamiento vs. 12.6 % en control) representa una amenaza potencial a la validez interna del estudio: si las mujeres que abandonaron el seguimiento son sistemáticamente distintas entre grupos, las estimaciones ITT podrían estar sesgadas. Los autores analizan este problema y reportan que los resultados son robustos a distintos supuestos sobre las mujeres que dejaron de reportarse.

- b) Escriba la ecuación econométrica que estima el efecto ITT del subsidio sobre fecundidad. Identifique claramente todos los componentes del modelo.

La ecuación econométrica que estima el efecto ITT del subsidio sobre cada resultado  $Y_i$  es:

$$Y_i = \alpha + \beta \cdot D_i + \sum_{p=1}^{20} \gamma_p \cdot \mathbf{1}[P_i = p] + \varepsilon_i,$$

donde:

- $Y_i$ : resultado de la mujer  $i$  observado en la encuesta de seguimiento.
- $\alpha$ : intercepto (media del grupo de control en la provincia de referencia).
- $\beta$ : coeficiente de interés, el efecto ITT del subsidio del 100 % relativo al 10 %. Bajo el supuesto de independencia condicional e SUTVA,  $\beta = \mathbb{E}[Y_i(1) - Y_i(0)]$ .
- $D_i \in \{0, 1\}$ : indicador de tratamiento (1 = subsidio del 100 %).
- $\gamma_p$ : efecto fijo de la provincia  $p$ ; controla por diferencias sistemáticas entre las 20 provincias que motivaron la estratificación.
- $\mathbf{1}[P_i = p]$ : indicador de que la mujer  $i$  pertenece a la provincia  $p$ .
- $\varepsilon_i$ : término de error. Los errores estándar se clusterizan a nivel de centro de salud (CSPS) para capturar la correlación intra-cluster generada por la aleatorización a ese nivel.

- c) Estime el efecto ITT del subsidio (100 % vs 10 %) para los cuatro resultados principales. Replique la Tabla 2 del artículo (sin contar la columna 4).

Los efectos ITT estimados para los tres resultados principales (columnas 1–3 de la Tabla 2 del artículo, excluyendo la columna 4 de uso de voucher) se presentan a continuación:

	(1) Nacimiento durante el estudio	(2) Usó anticoncepción moderna en 3 años	(3) Meses de uso moderno (últ. episodio)
<i>Panel A: solo controles de línea final</i>			
Subsidio 100 %	−0.013 (0.013)	−0.006 (0.017)	−0.542 (0.465)
Controles línea base	No	No	No
Efectos fijos provincia	Sí	Sí	Sí
Observaciones	7,489	7,241	7,489
Media control (10 %)	0.630	0.531	9.414

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

SE clusterizados a nivel de CSPS entre paréntesis.

Los tres resultados corresponden a las columnas 1 a 3 de la Tabla 2 del artículo: (1) si la mujer tuvo al menos un nacimiento durante el período de estudio, (2) si usó algún anticonceptivo moderno en los últimos tres años, y (3) los meses de uso moderno del último episodio por método (cero para quienes no usaron). Para cada resultado se estimó la ecuación con efectos fijos de provincia y errores estándar clusterizados a nivel de CSPS, sin controles adicionales de línea base (equivalente al Panel A del artículo).

Ninguno de los tres efectos es estadísticamente significativo: nacimientos  $\hat{\beta} = -0.013$  (SE = 0.013, p = 0.293), uso moderno en 3 años  $\hat{\beta} = -0.006$  (SE = 0.017, p = 0.729), meses de uso  $\hat{\beta} = -0.542$  (SE = 0.465, p = 0.244). Los resultados son cualitativamente consistentes con la Tabla 2 del artículo; las diferencias en magnitud obedecen al uso de la submuestra del 60 %.

## 6. Validez de Supuestos:

- a) En julio de 2020, el gobierno anunció la gratuidad nacional de la anticoncepción. Explique: (i) de qué manera este anuncio puede violar el supuesto SUTVA, (ii) cuál sería la dirección esperada del sesgo que introduce, y (iii) si considera convincente el argumento de los autores basado en la implementación gradual de la política y en el rezago biológico de nueve meses para observar nacimientos. (máx. 200 palabras)
- Violación del supuesto SUTVA. En julio de 2020, el gobierno de Burkina Faso anunció la gratuidad nacional de los anticonceptivos. Esto puede violar el supuesto SUTVA de dos formas. Primero, el “valor único del tratamiento” se ve comprometido ya que a partir del anuncio, las mujeres del grupo de control pueden acceder gratuitamente a anticonceptivos a través de canales gubernamentales, haciendo que  $D_i = 0$  ya no represente exclusivamente el subsidio del 10 % del programa. Segundo, puede haber interferencia indirecta si la política pública afecta la oferta de anticonceptivos en toda la región, alterando la disponibilidad para todos los grupos.
  - Si el grupo de control accede a anticonceptivos gratuitos tras el anuncio, la diferencia efectiva en acceso entre tratamiento y control se reduce. Esto atenúa el contraste entre grupos y sesga el estimador ITT hacia cero.
  - El argumento basado en la implementación gradual de la política es parcialmente robusto. Dado que la política comenzó a implementarse de forma escalonada después del anuncio y el seguimiento terminó en primavera de 2021, el tiempo efectivo de exposición fue limitado. El rezago biológico de nueve meses para nacimientos refuerza que los nacimientos observados al cierre del seguimiento reflejan comportamientos anticonceptivos previos al anuncio. Sin embargo, si el uso de anticonceptivos en el grupo de control aumentó tras julio de 2020, el ITT para uso de anticonceptivos podría estar subestimado. El argumento es más sólido para el resultado de fecundidad que para el uso de anticonceptivos.
- b) Defina y proporcione ejemplos en el contexto de este estudio de los siguientes efectos: Hawthorne, John Henry, demanda del experimentador (experimenter demand) y placebo. ¿En qué dirección sesgaría cada uno de estos efectos los resultados estimados?

- 1) Efecto Hawthorne: Las mujeres cambian su comportamiento por el hecho de saber que están siendo observadas, independientemente del tratamiento recibido. En este contexto, mujeres de ambos grupos podrían declarar un mayor uso de anticonceptivos o adoptar prácticas de planificación familiar al saber que participan en un estudio de salud reproductiva. Esto sesgaría el coeficiente hacia cero, dado que eleva el comportamiento en el grupo de control y reduce la diferencia observada entre grupos.
- 2) Efecto John Henry: El grupo de control, consciente de que recibe menos subsidio, compensa adoptando con mayor esfuerzo métodos alternativos (anticonceptivos tradicionales, abstinencia) para no quedar en desventaja. Lo anterior, reduciría la diferencia en uso de anticonceptivos y fecundidad entre grupos, sesgando el ITT hacia cero.
- 3) Efecto de demanda del experimentador: Las mujeres intuyen que los investigadores esperan que el subsidio reduzca la fecundidad y responden de acuerdo con esa expectativa. En el grupo tratado, podrían sobrereportar uso de anticonceptivos; en el control, no. Por esto, se estaría sesgando el coeficiente estimado hacia abajo (más negativo) para uso de anticonceptivos, sobreestimando el efecto real.
- 4) Efecto placebo: El hecho de ser incluidas en el programa, aun en el grupo de control con subsidio del 10%, genera cambios en comportamiento (mayor concientización sobre planificación familiar). Esto eleva el uso de anticonceptivos en el grupo de control, reduciendo la diferencia con el tratamiento y sesgando el ITT hacia cero.

## 7. Robustez del Hallazgo Nulo:

El resultado principal es nulo: no hay efecto sobre fecundidad. Los nulos requieren cuidado especial.

- a) Revise su Tabla 2. ¿Aumentó el subsidio el uso de anticoncepción? ¿Y el uso de vouchers? Interprete estos resultados conjuntamente.

La pregunta hace referencia al uso de vouchers como cuarta columna. Dado que en el inciso 3.5 se estimaron únicamente las tres primeras columnas de la Tabla 2 del artículo, el análisis se restringe a esos resultados: nacimientos durante el estudio, uso de anticoncepción moderna en los últimos tres años, y meses de uso moderno.

El uso de anticonceptivos modernos en los últimos tres años no aumentó ( $\hat{\beta} = -0.006$ , SE = 0.017, p = 0.729, media control = 0.531), ni los meses de uso moderno ( $\hat{\beta} = -0.542$ , SE = 0.465, p = 0.244, media control = 9.414). Ambos efectos son estadísticamente nulos.

El subsidio completo tampoco redujo los nacimientos ( $\hat{\beta} = -0.013$ , SE = 0.013, p = 0.293, media control = 0.630). Los tres nulos son coherentes entre sí: si el tratamiento no incrementó el uso de anticoncepción moderna, no hay canal mediante el cual pudiera reducir la fecundidad. Una explicación es la posible sustitución: las mujeres asignadas al subsidio completo canjearon vouchers pero eran, en su mayoría, mujeres que ya usaban anticonceptivos modernos y que simplemente sustituyeron su canal de obtención habitual. El subsidio del 10% en el grupo de control fue suficiente para que las mujeres motivadas ya accedieran, de modo que el subsidio adicional no incorporó nuevas usuarias ni redujo nacimientos.

- b) ¿Son los resultados nulos consistentes? ¿Qué tan robusto es el hallazgo? (Pista: leer la lectura)

Los resultados nulos estimados son consistentes y robustos. Los tres efectos ITT apuntan en la misma dirección: nacimientos ( $\hat{\beta} = -0.013$ , p = 0.293), uso moderno en tres años ( $\hat{\beta} = -0.006$ , p = 0.729) y meses de uso moderno ( $\hat{\beta} = -0.542$ , p = 0.244), y el intervalo de confianza al 95% sobre nacimientos [-0.038, 0.011] descarta reducciones superiores a 3.8 pp (6.0% de la media del grupo de control), coherente con el cálculo ex post del artículo que establece una potencia mínima detectable de 3.4 pp. Dupas et al. (2025) extienden estos nulos con verificaciones adicionales: el efecto sobre embarazos es -1.9 pp (no significativo), lo que descarta el canal de abortos espontáneos. La evidencia sobre el mecanismo refuerza los nulos: solo el 17% del grupo de tratamiento canjeó el voucher frente al 14% del control (diferencia de 3 pp, significativa al 1%), y la mayoría ya usaba anticoncepción moderna antes del estudio, por lo que el subsidio sustituyó su canal de compra sin incorporar nuevas usuarias. Los nulos persisten en todos los subgrupos con mayor demanda esperada: mujeres

con necesidad insatisfecha, mujeres que declaran no poder pagar (mayor efecto: -2.5 pp, aun no significativo), y mujeres cercanas al centro de salud (-3.2 pp, no significativo tras corrección por hipótesis múltiples). Finalmente, el artículo combina el subsidio con debates comunitarios sobre normas sociales y corrección de percepciones sobre mortalidad infantil, y ninguna interacción es significativa al 5 %, lo que confirma que la barrera no es de precio ni de información, sino de preferencias: los hombres desean en promedio 9 hijos y las mujeres 6.