

Primer ejercicio

Por motivo de la grave situación económica en el país luego de la pandemia, el Gobierno lanzó un programa social destinado a mejorar los ingresos de hogares pequeños cultivadores de café. En particular, la ayuda consistió en una capacitación acerca de la elección y uso de fertilizantes, ofrecida al jefe de hogar, junto con la entrega de una cantidad fija de dinero. Sin embargo, los recursos para desarrollar esta política eran limitados, por lo que sólo se podía brindar el programa a algunos de los hogares registrados en las bases de datos del Gobierno (los hogares no se tenían que inscribir para ser candidatos a recibir el programa). Por lo tanto, buscando evitar favorecimientos indebidos, se eligió aleatoriamente un subconjunto de hogares que recibirían el programa. Todos las familias seleccionadas para recibir la ayuda participaron activamente.

En busca de evaluar la efectividad de la medida, el Gobierno recolectó información acerca de la cosecha posterior al programa para todos los n hogares que fueron elegibles para el programa (esto es, tanto para los que recibieron el programa como para aquellos que no). Entre la información recolectada, se cuenta con la cantidad de café producido por hectárea cultivada en kilogramos por cada hogar C_i , así como una dummy D_i que toma el valor de uno si el hogar participó en el programa y cero de lo contrario.

Confiando en las habilidades de los estudiantes de econometría avanzada, a ustedes los contrata el Departamento Nacional de Planeación para llevar a cabo la evaluación de impacto respectiva.

Su jefe les dice que una manera de modelar el problema es a través de un modelo de regresión lineal dado por

$$C_i = \alpha + \tau D_i + \epsilon_i \quad (1)$$

donde ϵ_i es un componente aleatorio idiosincrásico de media cero ($\mathbb{E}[\epsilon_i] = 0$) y con varianza σ^2 ($\mathbb{E}[\epsilon_i^2] = \sigma^2$).

- a) Sean $C_i(1)$ y $C_i(0)$ los resultados potenciales de haber participado o no en el programa respectivamente. Similarmente, sean $\epsilon_i(1)$ y $\epsilon_i(0)$ los resultados potenciales análogos del error idiosincrático. Suponga además que $\mathbb{E}[\epsilon_i(1)] = \mathbb{E}[\epsilon_i(0)] = 0$.
- ¿Cuáles son las formas funcionales de $C_i(1)$ y $C_i(0)$ inducidas por el modelo (1)?
 - ¿Cuál parámetro del modelo captura el ATT del programa? Justifique matemáticamente.
 - Imagine un escenario donde el programa, en lugar de ser asignado aleatoriamente, se entregaba a las familias que viven en lugares con climas menos favorables para el cultivo. ¿En este escenario el parámetro del inciso ii) sigue capturando el ATT?

Pista: Recuerde que

$$C_i = D_i C_i(1) + (1 - D_i) C_i(0); \quad \epsilon_i = D_i \epsilon_i(1) + (1 - D_i) \epsilon_i(0)$$

Solución:

- i) Siguiendo la pista, el modelo escogido implica que:

$$D_i C_i(1) + (1 - D_i) C_i(0) = C_i = \alpha + \tau D_i + (D_i \epsilon_i(1) + (1 - D_i) \epsilon_i(0))$$

Como la anterior ecuación es válida para cualquier valor de D_i , entonces en particular cuando $D_i = 1$ se tiene que

$$C_i(1) = \alpha + \tau + \epsilon_i(1)$$

y cuando $D_i = 0$

$$C_i(0) = \alpha + \epsilon_i(0)$$

ii) Por definición:

$$ATT = \mathbb{E}[C_i(1) - C_i(0)|D_i = 1] = \mathbb{E}[(\alpha + \tau + \epsilon_i(1)) - (\alpha + \epsilon_i(0))|D_i = 1]$$

Por lo que

$$ATT = \tau + \mathbb{E}[\epsilon_i(1) - \epsilon_i(0)|D_i = 1]$$

Ahora, como la asignación del tratamiento fue aleatoria, entonces $D_i \perp \epsilon_i(1), \epsilon_i(0)$, luego

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

de ahí que el parámetro de interés sea τ .

iii) No, pues es plausible pensar que, entre los factores que conforman los errores idiosincráticos, está el clima. De esta manera, tendríamos que $D_i \not\perp \epsilon_i(1), \epsilon_i(0)$, luego

$$\mathbb{E}[\epsilon_i(1) - \epsilon_i(0)|D_i = 1] \neq 0$$

por lo que

$$\tau = ATT - \mathbb{E}[\epsilon_i(1) - \epsilon_i(0)|D_i = 1] \neq ATT$$

- b) Bajo el cumplimiento de los supuestos del modelo clásico lineal, demuestre que el estimador de MCO $\hat{\tau}$ es un estimador consistente del ATT y derive explícitamente su distribución asintótica (la varianza asintótica debe depender únicamente de n , σ^2 y $p = \mathbb{P}(D_i = 1)$).

Pista: Recuerde el siguiente teorema visto en clase:

Una secuencia de vectores aleatorios $\{x_N : N = 1, 2, \dots\}$ de $K \times 1$ converge en distribución al vector aleatorio x , si para cualquier vector no aleatorio c de $K \times 1$,

$$c^T x_N \xrightarrow{d} c^T x$$

Solución:

En este caso se cumple el supuesto de no correlación: puesto que $D_i \perp \epsilon_i$

$$\mathbb{E}[D_i \epsilon_i] = \mathbb{E}[D_i] \mathbb{E}[\epsilon_i] = 0$$

Luego tenemos que

$$\hat{\tau} \xrightarrow{P} \tau$$

Pero sabemos que a) que $\tau = ATT$, entonces $\hat{\tau}$ es un estimador consistente.

Ahora, bajo los supuestos del modelo clásico lineal, sabemos también que

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} \mathcal{N}(0, \sigma^2 Q_{XX}^{-1})$$

donde

$$x_i = \begin{pmatrix} 1 \\ D_i \end{pmatrix}; \quad \beta = \begin{pmatrix} \alpha \\ \tau \end{pmatrix}; \quad \hat{\beta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\tau} \end{pmatrix}$$

y donde $Q_{XX} = \mathbb{E}[x_i x_i^T]$.

Note entonces que

$$Q_{XX} = \mathbb{E}[x_i x_i^T] = \mathbb{E}\left[\begin{pmatrix} 1 \\ D_i \end{pmatrix} (1 \ D_i)\right] = \mathbb{E}\left[\begin{pmatrix} 1 & D_i \\ D_i & D_i^2 \end{pmatrix}\right] = \begin{pmatrix} 1 & p \\ p & p \end{pmatrix}$$

Entonces

$$Q_{XX}^{-1} = \frac{1}{p-p^2} \begin{pmatrix} p & -p \\ -p & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{(1-p)} & -\frac{1}{(1-p)} \\ -\frac{1}{(1-p)} & \frac{1}{p(1-p)} \end{pmatrix}$$

Sea $c = (0 \ 1)^T$. Notemos que

$$(\hat{\tau} - \tau) = c^T(\hat{\beta} - \beta) = (0 \ 1) \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\tau} - \tau \end{pmatrix}$$

Entonces, por la pista:

$$\sqrt{n}(\hat{\tau} - \tau) = c^T[\sqrt{n}(\hat{\beta} - \beta)] \xrightarrow{D} c^T Z; \quad Z \sim \mathcal{N}(0, \sigma^2 Q_{XX}^{-1})$$

Así las cosas, si $Z = (Z_1 \ Z_2)^T$, entonces $c^T Z = Z_2$. Sabemos además por las propiedades de la normal multivariada que Z_2 debe ser normal. Por lo tanto, para completar el ejercicio calculamos la media y varianza de $c^T Z$:

$$\mathbb{E}[c^T Z] = c^T \mathbb{E}[Z] = 0$$

y,

$$\text{Var}(c^T Z) = c^T \text{Var}(Z)c = \sigma^2(c^T Q_{XX}^{-1} c)$$

pero

$$c^T Q_{XX}^{-1} c = (0 \ 1) \begin{pmatrix} \frac{1}{(1-p)} & -\frac{1}{(1-p)} \\ -\frac{1}{(1-p)} & \frac{1}{p(1-p)} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{p(1-p)}$$

Concluimos entonces que

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{D} Z_2 \sim \mathcal{N}\left(0, \frac{\sigma^2}{p(1-p)}\right)$$

Contentos por los reveladores hallazgos de los incisos a) y b), ustedes van por un tinto a la cafetería. En ella, encuentran dos colegas debatiendo acerca de cómo estimar correctamente el ATT. Uno de ellos argumenta que, debido a que hay aleatorización, el estimador debe ser una diferencia “ingenua” de medias de la variable dependiente entre los hogares participantes y aquellos que, aunque elegibles, no accedieron al programa. Su otro compañero lo contradice, pues afirma que la manera correcta de hacerlo es estimando τ del modelo (1) por MCO dadas las conclusiones del inciso a) y b). ¿Quién tiene la razón?

Queriendo resolver esta encucijada, a ustedes se les ocurre una idea: si logran probar que ambos estimadores son numéricamente equivalentes, entonces no debería importar cuál procedimiento utilicen.

- c) Prueben que la diferencia ingenua de medias es numéricamente equivalente al estimador de MCO del parámetro τ del modelo (1).

Ayuda: Considere el modelo vectorial

$$\mathbf{C} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$

donde

$$\mathbf{X} = \begin{pmatrix} 1 & D_1 \\ 1 & D_2 \\ \vdots & \vdots \\ 1 & D_n \end{pmatrix}; \quad \mathbf{C} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix}; \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}; \quad \beta = \begin{pmatrix} \alpha \\ \tau \end{pmatrix}$$

Puede usar sin probar que

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{\left(\sum_{i=1}^n D_i\right) \left(\sum_{i=1}^n (1-D_i)\right)} \begin{pmatrix} \sum_{i=1}^n D_i & -\sum_{i=1}^n D_i \\ -\sum_{i=1}^n D_i & n \end{pmatrix}$$

y que

$$\mathbf{X}'\mathbf{C} = \begin{pmatrix} \sum_{i=1}^n C_i \\ \sum_{i=1}^n D_i C_i \end{pmatrix}$$

Solución:

De hecho, ambos tienen la razón, pues los estimadores son numéricamente equivalentes:

Note que el modelo (1) se puede escribir vectorialmente como:

$$\mathbf{C} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$

donde

$$\mathbf{X} = \begin{pmatrix} 1 & D_1 \\ 1 & D_2 \\ \vdots & \vdots \\ 1 & D_n \end{pmatrix}; \quad \mathbf{C} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix}; \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}; \quad \beta = \begin{pmatrix} \alpha \\ \tau \end{pmatrix}$$

Sabemos que el estimador MCO del modelo es:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}$$

Por un lado note que

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & \cdots & 1 \\ D_1 & \cdots & D_n \end{pmatrix} \begin{pmatrix} 1 & D_1 \\ 1 & D_2 \\ \vdots & \vdots \\ 1 & D_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n D_i \\ \sum_{i=1}^n D_i & \sum_{i=1}^n D_i \end{pmatrix}$$

pues $\sum_{i=1}^n D_i^2 = \sum_{i=1}^n D_i$, dado que $D_i \in \{0, 1\}$.

Entonces, se sigue que

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{\left(\sum_{i=1}^n D_i\right) \left(n - \sum_{i=1}^n D_i\right)} \begin{pmatrix} \sum_{i=1}^n D_i & -\sum_{i=1}^n D_i \\ -\sum_{i=1}^n D_i & n \end{pmatrix}$$

Note además que

$$n - \sum_{i=1}^n D_i = \sum_{i=1}^n 1 - \sum_{i=1}^n D_i = \sum_{i=1}^n (1 - D_i)$$

Ahora, por otro lado

$$\mathbf{X}'\mathbf{C} = \begin{pmatrix} 1 & \cdots & 1 \\ D_1 & \cdots & D_n \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n C_i \\ \sum_{i=1}^n D_i C_i \end{pmatrix}$$

Así las cosas

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C} = \frac{1}{\left(\sum_{i=1}^n D_i\right)\left(\sum_{i=1}^n (1-D_i)\right)} \begin{pmatrix} \sum_{i=1}^n D_i & -\sum_{i=1}^n D_i \\ -\sum_{i=1}^n D_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n C_i \\ \sum_{i=1}^n D_i C_i \end{pmatrix}$$

En particular, nos interesa saber $\hat{\tau}$, por lo que necesitamos encontrar el segundo componente del vector. Realizando las operaciones matriciales correspondientes se tiene que

$$\hat{\tau} = \frac{-\left(\sum_{i=1}^n D_i\right)\left(\sum_{i=1}^n C_i\right) + n\left(\sum_{i=1}^n D_i C_i\right)}{\left(\sum_{i=1}^n D_i\right)\left(\sum_{i=1}^n (1-D_i)\right)}$$

Finalmente, observe que

$$\sum_{i=1}^n C_i = \sum_{i=1}^n D_i C_i + \sum_{i=1}^n (1-D_i) C_i$$

Reagrupoando términos se tiene entonces que

$$\hat{\tau} = \frac{\left(\sum_{i=1}^n D_i C_i\right)\left(n - \sum_{i=1}^n D_i\right) - \left(\sum_{i=1}^n D_i\right)\left(\sum_{i=1}^n (1-D_i) C_i\right)}{\left(\sum_{i=1}^n D_i\right)\left(\sum_{i=1}^n (1-D_i)\right)}$$

Y simplificando

$$\hat{\tau} = \frac{\sum_{i=1}^n D_i C_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n (1-D_i) C_i}{\sum_{i=1}^n (1-D_i)} = C_i | D_i = 1 - C_i | D_i = 0$$

Ahora una compañera le comenta que, según lo que ha leído, ella cree que el éxito del programa depende en gran medida de la calidad de la tierra donde se siembra. De manera que, pueden existir posibles *efectos heterogéneos* dependiendo de las dotaciones de este factor en cada hogar. Asimismo, le comenta que afortunadamente se cuenta con una variable Z_i en la base de datos que captura esta información. Por lo tanto, ella propone que un modelo más completo estaría dado por

$$C_i = \beta_0 + \beta_1 D_i + \beta_2 \bar{Z}_i + \beta_3 D_i \bar{Z}_i + \epsilon_i \quad (2)$$

donde $\bar{Z}_i = (Z_i - \bar{Z})$ y \bar{Z} es el promedio muestral. Finalmente, suponga que $\mathbb{E}[\epsilon_i(1)|Z_i] = 0$ y $\mathbb{E}[\epsilon_i(0)|Z_i] = 0$, esto es, que la calidad de la tierra es exógena.

d) Partiendo del modelo (2):

i) Calcule el efecto esperado en los tratados como función de Z :

$$ATT(Z) = \mathbb{E}[C_i(1) - C_i(0)|D_i = 1, Z]$$

¿Cuál es la interpretación de β_1 y β_3 ?

- ii) Halle el parámetro que ahora captura el ATT del programa.
- iii) ¿Qué pasaría si Z_i no fuese exógena?

Solución:

i) Partiendo del modelo (2), y siguiendo una lógica similar al inciso a), entonces

$$C_i(1) = \beta_0 + \beta_1 + \beta_2 \bar{Z}_i + \beta_3 \bar{Z}_i + \epsilon_i(1)$$

y que

$$C_i(0) = \beta_0 + \beta_2 \tilde{Z}_i + \epsilon_i(0)$$

de manera que

$$C_i(1) - C_i(0) = \beta_1 + \beta_3 \tilde{Z}_i + (\epsilon_i(1) - \epsilon_i(0))$$

Entonces

$$\mathbb{E}[C_i(1) - C_i(0)|D_i = 1, Z_i] = \beta_1 + \beta_3 \mathbb{E}[\tilde{Z}_i|D_i = 1, Z_i] + \mathbb{E}[\epsilon_i(1) - \epsilon_i(0)|D_i = 1, Z_i]$$

Pero como la asignación al programa es aleatoria:

$$\mathbb{E}[\epsilon_i(1) - \epsilon_i(0)|D_i = 1, Z_i] = \mathbb{E}[\epsilon_i(1) - \epsilon_i(0)|Z_i] = 0$$

Finalmente

$$\mathbb{E}[\tilde{Z}_i|D_i = 1, Z_i] = \mathbb{E}\left[Z_i - \frac{1}{n} \sum_{j=1}^n Z_j \mid Z_i\right] = Z_i - \mu$$

donde $\mu = \mathbb{E}[Z_i]$ dado que son variables aleatorias i.i.d. Así las cosas:

$$ATT(Z_i) = \beta_1 + \beta_3(Z_i - \mu)$$

Bajo esta perspectiva, $\beta_1 = ATT(\mu)$, esto es, el efecto que percibe la persona con calidad de tierra promedio, mientras que β_3 es el efecto adicional para un individuo que se desvíe en $(Z_i - \mu)$ unidades de la media.

ii) Por la ley de esperanzas iteradas:

$$ATT = \mathbb{E}_{Z_i}[ATT(Z)] = \mathbb{E}_{Z_i}[\beta_1 + \beta_3(Z_i - \mu)] = \beta_1$$

iii) Si Z_i no fuese exógena entonces $\beta_1 + \beta_3(Z_i - \mu)$ no capturaría el ATT condicional en Z puesto que

$$\mathbb{E}[\epsilon_i(1) - \epsilon_i(0)|Z_i] \neq 0$$

por lo que

$$ATT(Z_i) - \mathbb{E}[\epsilon_i(1) - \epsilon_i(0)|Z_i] = \beta_1 + \beta_3(Z_i - \mu)$$

La moraleja de esto es que se pueden examinar efectos heterogéneos en un problema siempre que la fuente de heterogeneidad sea exógena (características geográficas, sexo, variables predeterminadas, etc.)

Finalmente, su jefe les tiene un último reto: probar si para al menos el 95 % de los hogares participantes el efecto de participar fue positivo.

e) Suponga que $Z_i \sim N(\mu, 1)$. Usando la forma funcional del modelo (2):

- i) Encuentre la distribución de $ATT(Z)$
- ii) Encuentre el percentil 5 ($p_{0.05}$) de la distribución de $ATT(Z)$ en términos de los parámetros del modelo.
- iii) Suponga que usted estimó por MCO que

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} 0.6 \\ 2 \\ 0.3 \\ -0.03 \end{pmatrix}; \quad S = \begin{pmatrix} 0.6 & -0.26 & 0.33 & 0 \\ -0.26 & 0.7 & 0.5 & 0 \\ 0.33 & 0.5 & 0.35 & -0.25 \\ 0 & 0 & -0.25 & 0.08 \end{pmatrix}$$

donde S es la matriz de varianza-covarianza estimada de los parámetros.

Construya un estimador consistente de $p_{0.05}$ y diseñe un test que le permita probar:

$$\begin{cases} H_0 : p_{0.05} \leq 0 \\ H_a : p_{0.05} > 0 \end{cases}$$

Use los datos disponibles para ejecutar su test. Concluya para un nivel de signifancia $\alpha = 0.05$.

- iv) Verdadero o falso: Si el p -valor del test es mayor a 0.05, entonces existe evidencia estadística de que el efecto de que el hogar en el percentil 5 de la distribución de efectos fue negativo o nulo. Justifique.

Solución:

- i) Partiendo del modelo (2) y usando las propiedades de la distribución normal:

$$ATT(Z_i) = \beta_1 + \beta_3(Z_i - \mu) \sim \mathcal{N}(\beta_1, \beta_3^2)$$

- ii) Note que $p_{0.05}$ es tal que

$$0.05 = \mathbb{P}(ATT(Z_i) \leq p_{0.05}) = \mathbb{P}\left(\frac{ATT(Z_i) - \beta_1}{\beta_3} \leq \frac{p_{0.05} - \beta_1}{\beta_3}\right)$$

Pero $Y_i = \frac{ATT(Z_i) - \beta_1}{\beta_3} \sim \mathcal{N}(0, 1)$, luego sabemos que

$$\frac{p_{0.05} - \beta_1}{\beta_3} = -1.645$$

de manera que:

$$p_{0.05} = \beta_1 - 1.645\beta_3$$

- iii) Podemos construir un estimador consistente de la cantidad de interés usando los resultados de MCO:

$$\hat{p}_{0.05} = \hat{\beta}_1 - 1.645\hat{\beta}_3$$

Como cada $\hat{\beta}_j$ es asintóticamente normal, sabemos que $\hat{p}_{0.05} \xrightarrow{D} \mathcal{N}(0, V)$, donde

$$V = \text{Var}(\hat{p}_{0.05}) = \text{Var}(\hat{\beta}_1) + (1.645)^2 \text{Var}(\hat{\beta}_3) - 1.645 \text{Cov}(\hat{\beta}_1, \hat{\beta}_3)$$

Si bien no conocemos las varianzas poblacionales, podemos construir un estimador de la varianza usando sus análogos muestrales. Así las cosas:

$$\hat{V} = \widehat{\text{Var}}(\hat{\beta}_1) + (1.645)^2 \widehat{\text{Var}}(\hat{\beta}_3) - 2 * 1.645 \widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_3)$$

Usando los valores de S :

$$\hat{V} = 0.7 + 1.645^2 * 0.08 - 2 * 1.645 * 0 = 0.916$$

además

$$\hat{p}_{0.05} = 2 + 0.03 * 1.645 = 2.05$$

Para testear la hipótesis entonces aprovechamos que, bajo la hipótesis nula, y por el Teorema de Slutsky:

$$\frac{\hat{p}_{0.05}}{\sqrt{\hat{V}}} = \frac{\hat{p}_{0.05}}{\sqrt{V}} \frac{\sqrt{V}}{\sqrt{\hat{V}}} \xrightarrow{p} 1 * N(0, 1)$$

Luego si

$$\frac{\hat{p}_{0.05}}{\sqrt{\hat{V}}} = 2.05/0.957 = 2.141$$

entonces el p-valor asociado (test de cola derecha) es 0, por lo que rechazamos la hipótesis nula.

Así las cosas, para un nivel de significancia de $\alpha = 0.05$, existe evidencia estadística de que, para al menos el 95 % de los hogares participantes, el efecto del programa fue positivo.

- iv) Falso. En una prueba de hipótesis no podemos “aceptar” H_0 . En la práctica, los tests de hipótesis se centran en “probar” H_a no H_0 : En breve, lo que se hace en los tests es suponer que H_0 es verdad (es decir que H_a no lo es), y miran si los datos son coherentes con este supuesto. Si no lo son, es poco probable (donde poco probable lo determina la significancia) que H_0 sea cierta (o de lo contrario los datos se acomodarían). Por lo tanto, preferimos tomar como cierta H_a sobre H_0 .

¿Ahora bien, qué pasa si los datos sí se “ajustan”? Pues lo que en realidad estamos probando es que no existe evidencia para a descartar H_0 . Pero... de ahí a que H_0 sea verdad hay un largo trecho...

Un ejemplo que ilustra el punto:

Usted cree que una especie de reptil del Amazonas está extinta. Por lo tanto usted plantea el siguiente test de hipótesis:

$$\begin{cases} H_0 : \text{La especie está extinta} \\ H_a : \text{La especie no está extinta} \end{cases}$$

Su plan entonces es internarse en el Amazonas. Bajo la hipótesis nula, ustedes no deberían encontrar rastro del reptil. Supongamos que ustedes hacen el viaje y efectivamente no encuentran rastros del animal. Entonces, ¿concluirían ustedes que la especie está extinta? Si bien esta conclusión es consistente con el hecho de no haber encontrado nada, lo cierto es que pueden haber otras. Por ejemplo, pudo también pasar que, por mala suerte, ustedes no vieran el animal en su viaje pero que, en realidad, todavía exista. Simplemente no sabemos, y no podemos saber con certeza (porque es imposible saber cuáles animales hay en el Amazonas, siempre inferimos a partir de pequeñas muestras, experiencias, etc.). Lo mismo pasa en estadística: cuando los datos se ajustan a una H_0 , decimos que “no se puede rechazar H_0 ”. Lo anterior, es una manera graciosa de decir que H_0 es consistente con los datos, pero no sabemos si es la única que lo es.

Ahora bien, lo que sí nos permite hacer el test es “probar” H_a : Suponga que efectivamente usted en realidad encuentra un ejemplar vivo. ¡Eureka! H_0 no puede ser cierta (en este caso con probabilidad 1), por lo que favorecemos H_a .

Segundo ejercicio

Una curva de aprendizaje se puede pensar como el cambio en la productividad o eficiencia con la que se hacen las cosas en el transcurso del tiempo. Este concepto es utilizado en Economía, por ejemplo, para modelar el cambio de los costos reales de las firmas según el tiempo que llevan operando. En particular, se supone que las

firms van aprendiendo a producir de manera más eficiente, por lo que el costo real de producción debería, *ceteris paribus*, decrecer.

Supongan que la curva generalizada de aprendizaje de una firma hipotética está dada por

$$C_i = C_0 N_i^{\frac{\alpha}{\gamma}} Y_i^{\frac{1-\alpha}{\gamma}} \exp(u_i), \quad (1)$$

donde C_i corresponde a los costos reales unitarios que enfrenta una firma i ; Y_i es su nivel de producción; N_i es la producción acumulada a lo largo del tiempo; C_0 corresponde a una medida de costos iniciales; y u_i es un término estocástico de media cero desconocido para el economista. El parámetro α determina la dirección de la elasticidad del costo unitario con respecto a la producción acumulada. Este es el principal parámetro de interés. Finalmente, el parámetro $\gamma \in \mathbb{R}^+$ caracteriza los retornos a escala de la función de aprendizaje: si $\gamma = 1$ la curva tiene retornos constantes a escala, si $\gamma < 1$ la curva tiene retornos decrecientes a escala, y si $\gamma > 1$ la curva tiene retornos crecientes a escala.¹

Usted, como investigador, está interesado en estimar la curva de aprendizaje que enfrentan las firmas de la industria manufacturera colombiana. Para ello, cuenta con información de C_i , N_i y Y_i para N firmas de la industria.

- a) Propongan un modelo de regresión lineal que capture la forma funcional dada en (1). Muestre cómo los parámetros de su nuevo modelo dependen de los parámetros de la curva de aprendizaje. Mencionen y discutan los supuestos necesarios para que los estimadores por MCO de los parámetros del modelo sean insesgados y consistentes. ¿Son estos supuestos plausibles en el contexto del problema?

Solución:

Para estimar los parámetros de interés, puede ser aplicada una transformación logarítmica a la ecuación (1). Concretamente:

$$\begin{aligned} \ln(C_i) &= \ln(C_0 N_i^{\frac{\alpha}{\gamma}} Y_i^{\frac{1-\alpha}{\gamma}} \exp(u_i)) \\ \ln(C_i) &= \ln(C_0) + \frac{\alpha}{\gamma} \ln(N_i) + \frac{1-\alpha}{\gamma} \ln(Y_i) + u_i \end{aligned}$$

Así, dada la transformación logarítmica, la ecuación a estimar por MCO está dada por

$$\ln(C_i) = \beta_0 + \beta_1 \ln(N_i) + \beta_2 \ln(Y_i) + u_i$$

donde $\beta_0 = \ln(C_0)$, $\beta_1 = \frac{\alpha}{\gamma}$ y $\beta_2 = \frac{1-\alpha}{\gamma}$.

Sean

$$X_i = [1, \ln(N_i), \ln(Y_i)]_{3 \times 1}^T \quad \forall \quad i = 1, 2, \dots, N,$$

$$\mathbf{X} = [X_1, X_2, \dots, X_N]_{N \times 3} \quad \text{y}$$

$$y = [\ln(C_1), \ln(C_2), \dots, \ln(C_N)]_{N \times 1}^T.$$

[Opcional] Primero, para que los parámetros estimados por MCO, $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2]$, sean únicos y puedan ser identificados debe ser que $X^T X$ es invertible (i.e., $\text{Rango}(\mathbf{X}^T \mathbf{X}) = 3$), de manera que $(\mathbf{X}^T \mathbf{X})^{-1}$ existe y, por tanto, $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T y$ está definido de manera única.

Segundo, para que los estimadores por MCO sean insesgados y consistentes se debe cumplir el supuesto de independencia condicional, según el cual

$$\mathbb{E}[\mathbf{u} | \mathbf{X}] = 0$$

donde $\mathbf{u} = [u_1, \dots, u_N]_{N \times 1}^T$.

El primer supuesto se puede garantizar incluyendo variables que no presenten multicolinealidad perfecta. El segundo supuesto en este contexto, sin embargo, puede no cumplirse. En particular, puede que las empresas/firmas que han producido más a lo largo del tiempo (N_i) y tienen una

¹Si la tecnología de producción muestra rendimientos constantes a escala, los costos unitarios reales no deben variar con el nivel de producción. Por el contrario, si los rendimientos son crecientes a escala, los costos unitarios deberían disminuir a medida que aumenta el nivel de producción, y si son decrecientes, se esperaría lo contrario.

mayor producción en determinado momento del tiempo (Y_i) cuenten con un cuerpo de trabajo más grande, más productivo, más motivado, más calificado, etc. Dado que es razonable que los costos operacionales de una empresa sean proporcionales también a las características de su cuerpo de trabajo y no solo de su producción, es plausible que $E[\mathbf{u}|\mathbf{X}] \neq 0$, dado que no estamos controlando por tales variables. Así, se espera que los estimadores de MCO sean sesgados e inconsistentes.

Nota: Esta respuesta no es única. Los estudiantes pueden justificar que el supuesto es razonable. Lo que importa es que cada supuesto esté correctamente discutido/justificado.

- b) Supongan que se cumplen los supuestos que ustedes discutieron en el inciso anterior. Propongan un estimador consistente de α y γ .

Solución:

Tenemos que $\beta_1 = \frac{\alpha}{\gamma}$ y $\beta_2 = \frac{1-\alpha}{\gamma}$. Luego,

$$\beta_2 = \frac{1-\alpha}{R\gamma} \longleftrightarrow \beta_2 = \frac{1}{R\gamma} - \frac{\alpha}{\gamma} \longleftrightarrow \beta_2 = \frac{1}{\gamma} - \beta_1 \longleftrightarrow \frac{1}{R\gamma} = \beta_1 + \beta_2 \longleftrightarrow \gamma = \frac{1}{\beta_1 + \beta_2}$$

y

$$\beta_1 = \frac{\alpha}{\gamma} \longleftrightarrow \alpha = \beta_1\gamma \longleftrightarrow \alpha = \frac{\beta_1}{\beta_1 + \beta_2}$$

Entonces, se proponen como estimadores de α y γ ,

$$\hat{\alpha} = \frac{\hat{\beta}_1}{\hat{\beta}_1 + \hat{\beta}_2} \quad \text{y} \quad \hat{\gamma} = \frac{1}{\hat{\beta}_1 + \hat{\beta}_2}.$$

Dado que los supuestos del inciso anterior se cumplen, se tiene que

$$\text{plim } \hat{\beta} = \beta$$

y, por tanto,

$$\text{plim } \hat{\beta}_1 = \beta_1 \quad \text{y} \quad \text{plim } \hat{\beta}_2 = \beta_2.$$

De esta manera, dado que $\hat{\alpha}$ y $\hat{\gamma}$ son funciones continuas en $\hat{\beta}_1$ y $\hat{\beta}_2$ para $\hat{\beta}_1 + \hat{\beta}_2 \neq 0$:

$$\begin{aligned} \text{plim } \hat{\alpha} &= \text{plim } \frac{\hat{\beta}_1}{\hat{\beta}_1 + \hat{\beta}_2} && (\text{Slutsky}) \\ &= \frac{\text{plim } \hat{\beta}_1}{\text{plim } \hat{\beta}_1 + \text{plim } \hat{\beta}_2} && (\hat{\beta}_1 + \hat{\beta}_2 \neq 0) \\ &= \frac{\beta_1}{\beta_1 + \beta_2} = \alpha \end{aligned}$$

$$\begin{aligned} \text{plim } \hat{\gamma} &= \text{plim } \frac{1}{\hat{\beta}_1 + \hat{\beta}_2} && (\text{Slutsky}) \\ &= \frac{\text{plim } 1}{\text{plim } \hat{\beta}_1 + \text{plim } \hat{\beta}_2} && (\beta_2 \neq -1) \\ &= \frac{1}{\beta_1 + \beta_2} = \gamma \end{aligned}$$

Es decir, $\text{plim } \hat{\alpha} = \alpha$ y $\text{plim } \hat{\gamma} = \gamma$, de manera que ambos estimadores son consistentes.

- c) Propongan un estimador de las desviaciones estándar de $\hat{\alpha}$ y $\hat{\gamma}$. Para esto, supongan que la matriz de

varianzas y covarianzas asintótica de $\sqrt{N}(\hat{\beta} - \beta)$ y su respectivo estimador están dados por

$$V = Avar(\sqrt{N}(\hat{\beta} - \beta)) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \text{ y } \hat{V} = \widehat{Avar}(\sqrt{N}(\hat{\beta} - \beta)) = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \hat{\sigma}_{13} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 & \hat{\sigma}_{23} \\ \hat{\sigma}_{13} & \hat{\sigma}_{23} & \hat{\sigma}_3^2 \end{pmatrix} \quad (3)$$

respectivamente. Asegúrense de que sus expresiones estén en términos de $\hat{\beta}_1$, $\hat{\beta}_2$ y los elementos que componen \hat{V} .

Pista: Recuerden que pueden implementar el método delta.

Solución:

Sea $\theta^T = [\alpha \ \gamma]$. El Jacobiano del vector θ^T respecto a β es

$$J = \frac{\partial \theta^T}{\partial \beta} = \begin{pmatrix} 0 & -\frac{\beta_2}{(\beta_1 + \beta_2)^2} & -\frac{\beta_1}{(\beta_1 + \beta_2)^2} \\ 0 & -\frac{1}{(\beta_1 + \beta_2)^2} & -\frac{1}{(\beta_1 + \beta_2)^2} \end{pmatrix}$$

Adicionalmente, sea

$$V = Avar(\sqrt{N}(\hat{\beta} - \beta)) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \quad (4)$$

la varianza asintótica poblacional.

Sabemos entonces que

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, JV J^T)$$

donde

$$JV J^T = \begin{pmatrix} \frac{\sigma_2^2 \beta_2^2 - 2\sigma_{23}\beta_1\beta_2 + \sigma_3^2 \beta_1^2}{(\beta_1 + \beta_2)^4} & \frac{-\sigma_2^2 \beta_2 + \sigma_{23}(\beta_1 - \beta_2) + \sigma_3^2 \beta_1}{(\beta_1 + \beta_2)^4} \\ \frac{-\sigma_2^2 \beta_2 + \sigma_{23}(\beta_1 - \beta_2) + \sigma_3^2 \beta_1}{(\beta_1 + \beta_2)^4} & \frac{\sigma_2^2 + 2\sigma_{23} + \sigma_3^2}{(\beta_1 + \beta_2)^4} \end{pmatrix}$$

de manera que las desviaciones estándar asintóticas de $\hat{\alpha}$ y \hat{R} están dadas por

$$\sqrt{Avar(\hat{\alpha})} = \sqrt{\frac{\sigma_2^2 \beta_2^2 - 2\sigma_{23}\beta_1\beta_2 + \sigma_3^2 \beta_1^2}{N(\beta_1 + \beta_2)^4}}$$

$$\sqrt{Avar(\hat{\gamma})} = \sqrt{\frac{\sigma_2^2 + 2\sigma_{23} + \sigma_3^2}{N(\beta_1 + \beta_2)^4}}$$

y, por tanto, unos posibles estimadores son

$$\sqrt{\widehat{Avar}(\hat{\alpha})} = \sqrt{\frac{\hat{\sigma}_2^2 \hat{\beta}_2^2 - 2\hat{\sigma}_{23}\hat{\beta}_1\hat{\beta}_2 + \hat{\sigma}_3^2 \hat{\beta}_1^2}{N(\hat{\beta}_1 + \hat{\beta}_2)^4}}$$

$$\sqrt{\widehat{Avar}(\hat{\gamma})} = \sqrt{\frac{\hat{\sigma}_2^2 + 2\hat{\sigma}_{23} + \hat{\sigma}_3^2}{N(\hat{\beta}_1 + \hat{\beta}_2)^4}}$$

Usted cuenta con la base de datos “manufacturaCol.dta” para implementar la estimación de los parámetros propuesta en los incisos anteriores. La base de datos cuenta con información de las siguientes variables para cada una de las 11,500 firmas de la muestra:

- **costos:** Costos unitarios reales de producción que enfrenta la firma en el momento que se levantaron los datos medido en miles de millones pesos.

- *producto*: Producción de la firma en miles de millones pesos en el momento que se levantaron los datos.
- *producto_acum*: Producción acumulada histórica de la firma en el momento en que se levantaron los datos medida en miles de millones de pesos .

Resuelvan los siguientes incisos a partir de los datos disponibles en la base de datos.

d) A manera de estadísticas descriptivas, representen mediante un gráfico de dispersión y su respectiva línea de ajuste las siguientes relaciones: 1. La relación de $\log(C_i)$ con $\log(N_i)$ una vez se ha “removido” el efecto de $\log(Y_i)$. 2. La relación de $\log(C_i)$ con $\log(Y_i)$ una vez se ha “removido” el efecto de $\log(N_i)$.

Ayuda: para lograr esto, implementen el concepto de partialling out o residualización. Interprete los gráficos de dispersión.

Solución:

i)

La relación incondicional entre dos variables, la cual puede ser representada por un gráfico de dispersión entre dos variables, generalmente captura muchas más cosas que la relación de interés. En este caso, la relación entre $\log(C_i)$ y $\log(N_i)$ puede ser parcialmente afectada por las relaciones que existen entre $\log(C_i)$ y $\log(Y_i)$ y entre $\log(N_i)$ y $\log(Y_i)$. De esta manera, puede ser que al representar $\log(C_i)$ vs. $\log(N_i)$ observemos una relación incorrecta y contaminada, tal y como indica el investigador.

Así, una metodología a partir de la cual podemos graficar relaciones condicionales es la de *partialling out* o residualización. Esta metodología permite representar la relación entre $\log(C_i)$ y $\log(N_i)$, por ejemplo, mientras que mantenemos fijo $\log(Y_i)$.

Si, entonces, queremos representar $\log(C_i)$ y $\log(N_i)$, manteniendo fija $\log(Y_i)$, la metodología consiste en:

1. Estimar por MCO

$$\log(N_i) = \alpha_0 + \alpha_1 \log(Y_i) + \varepsilon_i$$

2. Obtener los residuales

$$\hat{\varepsilon}_i = \log(N_i) - \hat{\alpha}_0 - \hat{\alpha}_1 \log(Y_i)$$

los cuales corresponden a la parte de $\log(N_i)$ que no depende de $\log(Y_i)$.

3. Graficar $\log(C_i)$ contra $\hat{\varepsilon}_i$.

Como fue mencionado en el paso 2, $\hat{\varepsilon}$ corresponde a la parte de $\log(N_i)$ que no depende de $\log(Y_i)$. Así, movimientos en $\hat{\varepsilon}$ no están relacionados con $\log(Y_i)$, de manera que graficar $\log(C_i)$ contra $\hat{\varepsilon}$ representa la relación entre $\log(C_i)$ y la parte de $\log(N_i)$ que no se mueve ante cambios de $\log(Y_i)$. Esto es, la relación condicional entre $\log(C_i)$ y $\log(N_i)$ una vez controlamos por $\log(Y_i)$. La metodología es análoga para la relación entre $\log(C_i)$ y $\log(Y_i)$.

ii)

A continuación son presentadas los gráficos de dispersión y sus respectivas líneas de ajuste una vez implementamos la metodología de *partialling out* y controlamos por $\log(Y_i)$ y $\log(N_i)$ en cada caso.

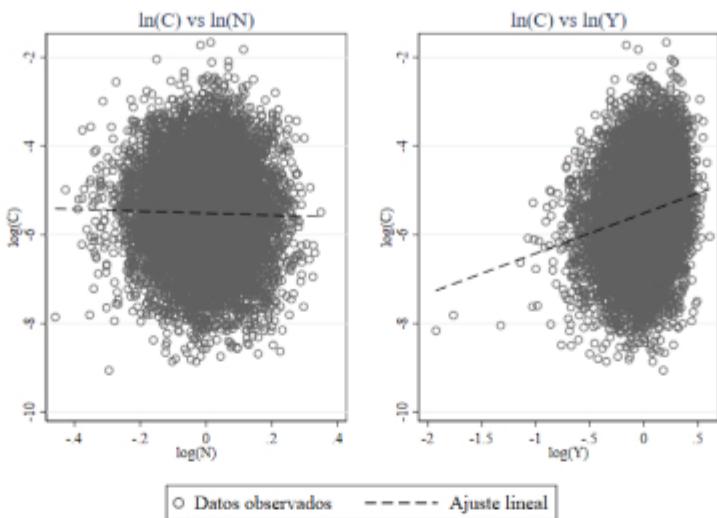


Figura 1. Relaciones parciales

Como se puede observar a partir de los gráficos, parece ser que los costos tienen una relación negativa con la producción acumulada, pero una relación positiva con la producción en el momento de la medición. En primer lugar, la gráfica de la derecha sugiere que la elasticidad de los costos respecto a la producción acumulada es negativa. Sin embargo, parece ser que tal elasticidad es muy baja y probablemente no es significativa. Por otra parte, la gráfica de la derecha indica que la elasticidad de los costos respecto a la productividad del momento es positiva y probablemente significativa. Esta relación sugiere adicionalmente que $\alpha < 1$ dado que $1 - \alpha \geq 0$.

- c) Finalmente, estimen por MCO la ecuación propuesta en el inciso a). A partir de estos resultados, estimen α , γ , y sus respectivos errores estándar siguiendo el planteamiento de los incisos b) y c). Presenten en una tabla los parámetros estimados por MCO. ¿Son estos parámetros consistentes con las gráficas que presentaron en el inciso anterior? Discutan brevemente. Adicionalmente, en una segunda tabla presenten los parámetros α y γ que ustedes estimaron y sus respectivos errores estándar. Interpreten estos últimos. ¿Hay rendimientos crecientes, decrecientes o constantes a escala? ¿Qué tipo de curva de aprendizaje hay en la industria?

Solución:

Tabla 2.1. Regresión por MCO

	(1)
	lnC
lnN	-0.230** (0.0905)
lnY	0.907*** (0.0437)
Constante	-7.825*** (0.865)
Observaciones	11,500
R cuadrado	0.006
Errores estándar entre paréntesis	
*** p<0.01, ** p<0.05, * p<0.1	

A partir de la tabla 2.1. se puede afirmar que, en promedio y a un nivel de significancia del 10%, la elasticidad de los costos respecto a la producción acumulada es de -0.23. Esto indica que efectivamente hay una curva de aprendizaje en la industria. No obstante, parece ser una curva de aprendizaje lenta, dado que -0.23 representa una curva inelástica de costos respecto a la producción acumulada. De otro lado, se puede afirmar que, en promedio y a un nivel de significancia del 1 %, la elasticidad de los costos respecto a la producción del momento es de 0.907.

Esto indica que aumentos en la producción del momento resultan en incrementos proporcionales en los costos marginalmente menores, lo cual sugiere que $\gamma \geq 1$. No obstante, como fue discutido en el inciso b), para conocer el verdadero valor de γ hace falta hacer unos cálculos adicionales. Estos coeficientes son consistentes con lo presentado en la Figura 2.1 del inciso anterior dado que corresponden a la pendiente de la línea de ajuste representada en cada gráfica. Esto gracias al ejercicio de *partialling out* implementado en el inciso anterior.

Tabla 2.2. Parámetros estimados

Parámetro estimado

α	-0.34*
	(0.18)
γ	1.48***
	(0.20)
Errores estándar entre paréntesis	
*** p<0.01, ** p<0.05, * p<0.1	

Ahora bien, la tabla 2.2. presenta los parámetros α y γ y sus respectivos errores estándar. En primer lugar, se puede afirmar que el parámetro α es, en promedio y a un nivel de significancia del 10 %, igual a -0.34. Este coeficiente negativo es el que da forma a la elasticidad negativa que existe entre los costos y la producción acumulada, confirmando así que efectivamente hay una curva de aprendizaje en la industria. De otro lado, se puede afirmar que el parámetro γ es, en promedio y a un nivel de significancia del 1 %, igual a 1.48. Este resultado, tal como fue predicho en el anterior párrafo, indica que en la industria manufacturera hay retornos crecientes a escala, de manera que los costos que enfrentan las firmas de la industria disminuyen a medida que aumenta el nivel de producción.

Tercer ejercicio

Aunque las propiedades asintóticas de los estimadores son de gran utilidad para el ejercicio práctico de la estadística, lo cierto es que muchas veces encontrar resultados teóricos en este ámbito es extremadamente complicado. En estas ocasiones, es usual recurrir a otras herramientas, como lo son los procedimientos de Monte Carlo. En breve, los métodos de Monte Carlo aplicados a la estadística buscan explorar las propiedades de los estimadores (insesgamiento, consistencia, eficiencia, suficiencia, etc.) al observar el comportamiento de los mismos en varias muestras aleatorias simuladas. Este ejercicio los guiará a través de una serie de actividades que les permitirán entender algunos métodos clásicos de simulación de muestras aleatorias para, posteriormente, usarlos para evaluar el comportamiento de uno de los estimadores más importantes que se encontrarán en sus carreras: el estimador de MCO.

Suponga que el Ministerio de Educación ha decidido lanzar un programa de preparación y acompañamiento para la presentación de la prueba ICFES a estudiantes que se encuentren en su último año de bachillerato. En particular, suponga que usted conoce que el puntaje ICFES estandarizado potencial $Y_i(\cdot)$ obedece el siguiente modelo:

$$Y_i(0) \sim N(0, 1)$$

$$Y_i(1) = Y_i(0) + 3$$

donde $Y_i(1)$ y $Y_i(0)$ son los resultados potenciales del puntaje ICFES en caso de participar y de no participar respectivamente, y donde $Y_i(0)$ tiene una distribución normal estándar.

- a) Escriba el puntaje ICFES observado, Y_i , en términos de los resultados potenciales y la exposición al tratamiento, D_i . Suponga que el Ministerio escoge aleatoriamente a los participantes del programa y que todo individuo seleccionado obligatoriamente participa. Proponga un modelo de regresión lineal que le permita estimar el efecto del tratamiento sobre el puntaje, llámese δ .

Sabemos que, reemplazando las definiciones,

$$Y_i = Y_i(1) * D_i + Y_i(0) * (1 - D_i) = Y_i(0) + D_i * \delta \quad (\delta = 3)$$

Consecuentemente, el modelo de regresión apropiado para la estimación de δ viene dado por

$$Y_i = \beta_0 + \delta * D_i + \varepsilon_i$$

donde ε_i satisface los supuestos del modelo clásico lineal.

b). Inicialmente, buscaremos validar a través de simulaciones algunas de las propiedades asintóticas del estimador por MCO. En particular, a usted le interesa saber si a medida que aumenta el tamaño de muestra sus estimativos del efecto del programa se tornan más precisos.

- I. Simule 100 muestras $\{(Y_i(0), Y_i(1), D_i\}_{i=1}^n$ de tamaño $n = 10, 20, \dots, 1000$ donde $D_i \sim Bernoulli(0.3)$. Para cada muestra, estime el efecto de la política del Ministerio sobre el puntaje ICFES y almacene su estimado.

```

1      clear all
2      cap log close
3      set more off
4      cls
5
6
7      global dir "C:\Users\David\OneDrive - Universidad de los Andes\Econ. Av. - COMP
       \2022-1\6. Talleres\Taller 1\Punto 3"
8
9      cd "$dir"
10
11     set seed 202 // Sembramos una semilla
12
13
14
15     * Estudiemos la consistencia
16
17     clear
18
19     // Creamos una matriz para guardar nuestros resultados
20
21     mat B = J(100, 100, .)
22
23     mat S = J(100, 1, .)
24
25     // Hagamos un loop sobre los tamaños de muestra
26
27     forvalues n=10(10)1000{
28
29         forvalues j = 1(1)100{ // Corremos un loop sobre el número de la simulación
30
31             clear
32
33             qui mat N = J(`n', 3, .) // Matriz para ir almacenando las observaciones que
               conforman una misma simulación
34
35
36             forvalues i = 1(`n')`j' // Loop para implementar el algoritmo A-R
37
38             qui mat N[`i', 1] = rnormal() // Almacenamos Y_(i)(0)
39
40             qui mat N[`i', 2] = N[`i', 1] + 3 // Almacenamos Y_(i)(1)
41
42             qui mat N[`i', 3] = (runiform() > 0.7) // Generamos D_(i)
43
44         }
45
46
47     clear
48
49
50     qui svmat N // Convertimos la matriz a datos
51
52     qui gen Y = N3*N2+(1-N3)*N1 // Obtenemos Y
53
54     qui reg Y N3 // Efectuamos la estimación
55
56     qui mat B[`j', `n'/10] = _b[N3] // Almacenamos el resultado
57
58     qui mat S[`n'/10, 1] = `n' // Almacenamos el indicador de tamaño de muestra
59
60     }

```

```
62 } 
```

- II. Obtenga la media y la varianza muestral de los estimadores almacenados para los distintos tamaños de muestra.

```
1  
2  
3 clear  
4  
5 svmat B // Convertimos las estimaciones en datos  
6  
7 mat Bvar = J(100, 1, .) // Creamos la matriz de varianzas de las estimaciones  
8  
9 mat Bmean = J(100, 1, .) // Creamos la matriz de medias de las estimaciones  
10  
11 forvalues i=1(1)100{  
12  
    qui sum B`i'  
    qui mat Bvar[`i', 1] = r(Var)  
    qui mat Bmean[`i', 1] = r(mean)  
13  
}  
14  
15  
16 clear  
17  
18  
19  
20 svmat Bvar // Nos quedamos con las varianzas  
21  
22 svmat Bmean // Nos quedamos con las medias  
23  
24 svmat S // Recuperamos los tamaños de muestra
```

- III. Haga las siguientes dos gráficas: 1) Grafique el promedio muestral de los estimadores contra el tamaño de muestra. 2) Grafique las varianzas muestrales contra el tamaño de muestra correspondientes. ¿Qué puede decir acerca del comportamiento de las estimaciones a medida que aumenta el tamaño de muestra? ¿Qué propiedades del estimador de MCO se ven reflejadas en el ejercicio?

```
1  
2 line Bvar S1, graphregion(color(white)) plotregion(lcolor(black)) legend(  
    order(1 "Varianza")) ylabel(, nogrid) ytitle("Varianza de los estimadores")  
    xtitle("Tamaño de muestra") // Construimos el gráfico  
3  
4  
5 graph save "$dir/consistencyv.gph", replace  
6 graph export "$dir/consistencyv.pdf", replace  
7  
8 -----  
9 *Latex call  
10  
11 texdoc close  
12 cap erase "$dir/consistencyv.tex"  
13 texdoc init "$dir/consistencyv.tex", force  
14  
15  
16 tex \begin{figure}[H]  
17 tex \caption{Consistencia: varianza de las estimaciones}  
18 tex \centering  
19 tex \includegraphics[scale=0.8]{Talleres/2022-1/Taller 1/Figuras/consistencyv.pdf}  
20 tex \label{fig:my_label}  
21 tex \end{figure}  
22  
23  
24 texdoc close  
25  
26  
27  
28 line Bmean S1, graphregion(color(white)) plotregion(lcolor(black)) legend(order(1  
    "Varianza")) ylabel(, nogrid) ytitle("Media de los estimadores") xtitle("Tamaño de muestra") yline(3, lpattern(dash)) // Construimos el gráfico  
29  
30  
31 graph save "$dir/consistencym.gph", replace  
32 graph export "$dir/consistencym.pdf", replace  
33  
34 -----  
35 *Latex call  
36 
```

```

37 texdoc close
38 cap erase "$dir/consistencym.tex"
39 texdoc init "$dir/consistencym.tex", force
40
41
42 tex \begin{figure}[H]
43 tex \caption{Consistencia: media de las estimaciones}
44 tex \centering
45 tex \includegraphics[scale=0.8]{Talleres/2022-1/Taller 1/Figuras/consistencym.pdf}
46 tex \label{fig:my_label}
47 tex \end{figure}
48
49
50 texdoc close

```

Figura 1: Consistencia: varianza de las estimaciones

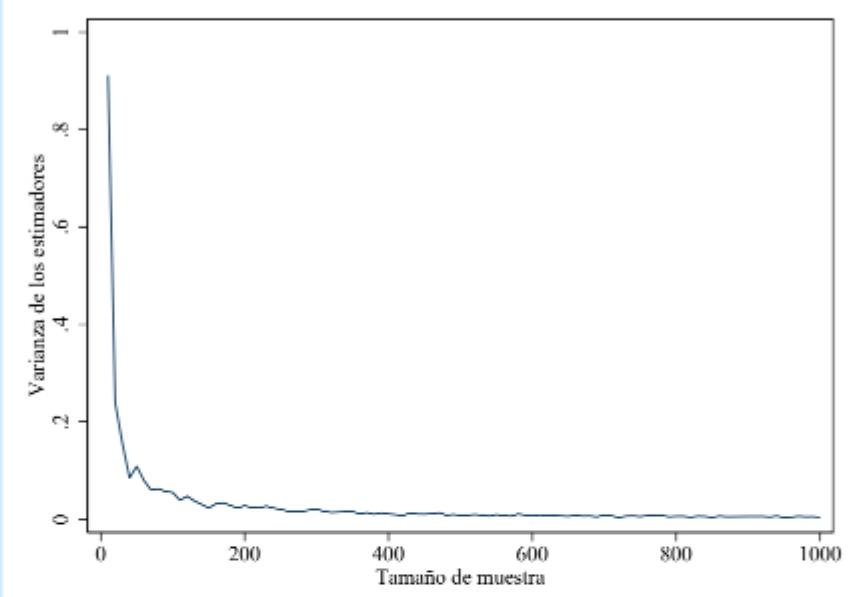
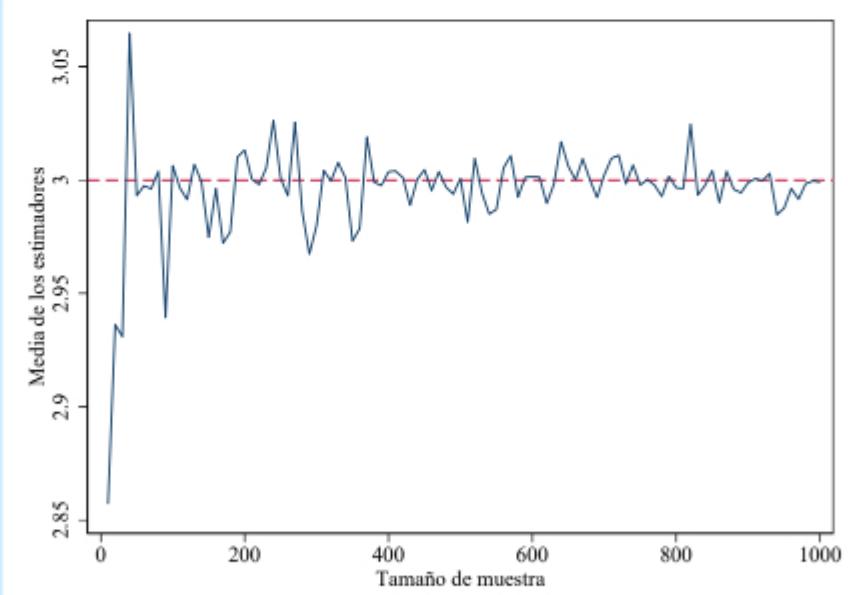


Figura 2: Consistencia: media de las estimaciones



La gráfica del promedio muestral contra el tamaño de muestra nos sugiere que el valor esperado del estimador de MCO es el valor real del parámetro, a saber, $\delta = 3$. Además, esto no depende del tamaño de muestra. Lo anterior, es una manera empírica de comprobar que

el estimador de MCO, bajo los supuestos del modelo clásico lineal, es insesgado.

Similarmente, la gráfica de las varianzas muestra claramente una relación decreciente entre el tamaño de muestra y la varianza muestral de los estimadores de MCO. Lo que nos sugiere esta gráfica es que el estimador está convergiendo a algún valor, pues cada vez las observaciones son más parecidas. Esto, junto con el insesgamiento del estimador, nos garantiza que dicho valor al que converge es en realidad δ . Por lo tanto, hemos comprobado empíricamente que el estimador de MCO es un estimador consistente del parámetro poblacional.

Uno de los propósitos principales de los modelos de regresión lineal es poder estudiar las relaciones que existen entre la variable dependiente y las variables independientes. Más precisamente, nos interesa dilucidar el tamaño y el signo de dicha relación. No obstante, en la práctica, esto es difícil de establecer puesto que desconocemos el proceso generador de los datos y en general, contamos con una sola muestra para nuestra estimación. Así las cosas, si nuestro objetivo es hacer inferencia, debemos preguntarnos qué tan precisas son nuestras estimaciones, esto es, qué tan dependientes son de la muestra particular que tenemos. Para lograrlo, es útil entender a los estimadores como variables aleatorias, que dependen de una muestra, pero que tienen una distribución definida. En particular, por simplicidad, es de nuestro interés dicha distribución cuando el tamaño de muestra es grande, pues conocer la distribución exacta para muestras reducidas puede ser complicado.

c) En este inciso vamos a aproximar la distribución asintótica del estimador de MCO. Para ello, realicen las siguientes instrucciones:

1. Siguiendo los pasos expuestos en el punto b), simule 1000 muestras $\{(Y_i(0), Y_i(1), D_i\}_{i=1}^n$ de tamaño $n = 10, 20, 100, 1000$ donde $D_i \sim Bernoulli(0.3)$. Para cada muestra, estime el efecto del programa del Ministerio y calcule y almacene

$$a_{k,n} = \sqrt{n}(\hat{\delta}_{k,n} - \delta)$$

donde k indexa las muestras de un determinado tamaño.

```
1
2
3
4 * Estudiemos la convergencia en distribuci n
5
6 clear
7
8 // Creamos una matriz para guardar nuestros resultados
9
10 mat C = J(1000, 4, .)
11
12
13 // Hagamos un loop sobre los tama os de muestra
14
15 local k = 1 // Esto nos servir para guardar nuestros resultados
16
17 foreach n in 10 20 100 1000{
18
19
20     forvalues j = 1(1)1000{ // Corremos un loop sobre el n mero de la
21         simulaci n
22
23         clear
24
25         qui mat N = J(`n', 3, .) // Matriz para ir almacenando las observaciones que
26             conforman una misma simulaci n
27
28         forvalues i = 1(1)`n'{ // Loop para implementar el algoritmo A-R
29
30             qui mat N[`i', 1] = rnormal() // Almacenamos Y_fix(0)
31
32             qui mat N[`i', 2] = N[`i', 1] + 3 //Almacenamos Y_(i)(1)
33
34             qui mat N[`i', 3] = (runiform() > 0.7) // Generamos D_(i)
35
36
37         }
38
39         clear
40
41         qui svmat N // Convertimos la matriz a datos
42
```

```

43 qui gen Y = N3*N2+(1-N3)*N1 // Obtenemos Y
44 qui reg Y N3 // Efectuamos la estimaci n
45 qui mat C['j', 'k'] = (_b[N3]-3)*sqrt('n') // Almacenamos el resultado
46
47 }
48
49 local k = `k'+1
50
51 }
52
53 clear
54
55 svmat C // Convertimos las estimaciones en datos

```

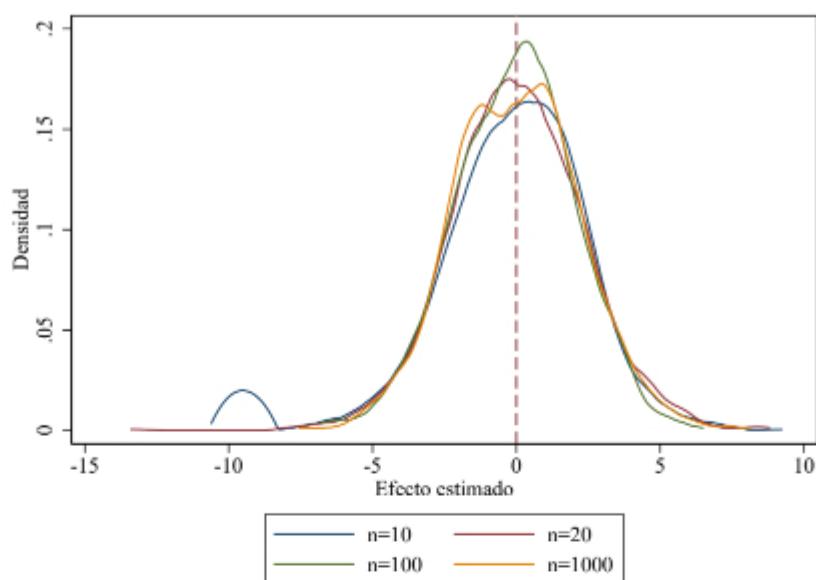
- II. Grafique las densidades estimadas de los $a_{k,n}$ para cada n . ¿Qué puede apreciar a medida que aumenta n ? ¿A qué se debe este resultado?

```

1 clear
2
3 svmat C // Convertimos las estimaciones en datos
4
5
6
7
8 twoway (kdensity C1) (kdensity C2) (kdensity C3) (kdensity C4), legend(order(1 "n
=10" 2 "n=20" 3 "n=100" 4 "n=1000")) graphregion(color(white)) plotregion(
lcolor(black)) ylabel(, nogrid) ytitle(Densidad) xtitle("Efecto estimado")
xline(3, lcolor(maroon) lpattern(dash)) // Construimos el gr fico
9
10
11 graph save "$dir/normal.gph", replace
12 graph export "$dir/normal.pdf", replace
13
14
15 -----
16 *Latex call
17
18 texdoc close
19 cap erase "$dir/normal.tex"
20 texdoc init "$dir/normal.tex", force
21
22
23 tex \begin{figure}[H]
24 tex \caption{Convergencia en distribuci n: distribuci n normal}
25 tex \centering
26 tex \includegraphics[scale=0.8]{Talleres/2022-1/Taller 1/Figuras/normal.pdf}
27 tex \label{fig:my_label}
28 tex \end{figure}
29
30
31 texdoc close

```

Figura 3: Convergencia en distribución: distribución normal



Podemos ver que los $a_{k,n}$ tienen una distribución similar a la normal (de hecho, normal) para todos los tamaños de muestra. Aunque no parece haber diferencias notables en la forma de la densidad estimada conforme incrementa n , sí se presentan menos valores atípicos. Hay dos justificaciones válidas . En primer lugar, la distribución normal de Y implica que los $a_{k,n}$ tienen una distribución exactamente normal. Por otro lado, el Teorema del Límite Central, bajo las condiciones del presente ejercicio, implica que los $\hat{\delta}$ tienen una distribución asintóticamente \sqrt{n} -normal.

Suponga ahora que usted sabe que existen factores ajenos a las políticas del Ministerio que inciden sobre los resultados potenciales de los alumnos. Por ejemplo, sabe que existen alumnos que, independientemente de si son o no seleccionados, igual se inscribirían en cursos de preparación para presentar el ICFES. Similarmente, hay alumnos que por condiciones adversas (enfermedades, salones menos adecuados para la presentación del examen) exhiben un desempeño muy inferior a lo esperado. Estos factores producen una mayor volatilidad en los resultados potenciales de los alumnos.

- d) En este sentido, usted sabe que es más plausible considerar que los resultados potenciales en realidad obedecen la ley $Y_i(0) \sim \text{Cauchy}$ ([Distribución Cauchy estándar](#)) en vez de una normal estándar. Esto es, en ocasiones usted observa datos atípicos que no parecen corresponder con el patrón general de la muestra. Repita los incisos I y II del c) bajo estas condiciones. ¿Por qué en este caso no se satisface el Teorema del Límite Central? ¿Qué consecuencias tiene esto para la inferencia de los parámetros de MCO?

```

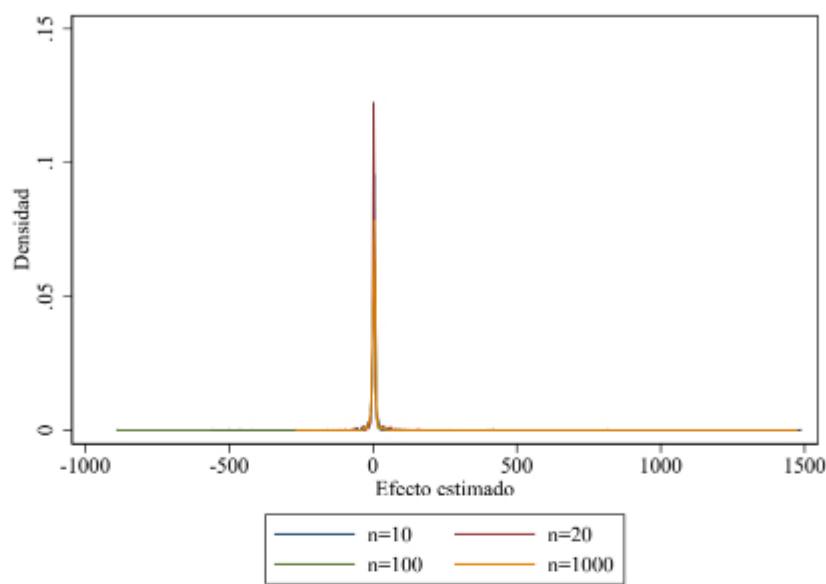
1
2
3
4
5
6 * Generar Cauchy de una uniforme (0,1)
7
8 clear
9
10 // Creamos una matriz para guardar nuestros resultados
11 mat A = J(1000, 3, .)
12
13 // Hagamos un loop sobre los tamaños de muestra
14
15
16 foreach k in 1 2 3{
17     local n = 5*10^`k' // Tamaño de muestra
18     forvalues j = 1(1)1000f // Corremos un loop sobre el número de la simulación
19         A[`j', `k'] = rcauchy(`n')
20 }
```

```

22      clear
23
24      qui set obs `n' // Definimos el n mero de observaciones a simular
25
26      qui gen U = runiform() // Generamos la uniforme
27
28      qui gen Y0 = tan((U-0.5)*_pi) // Generamos Y_{i}(0)^Cauchy usando el teorema
29
30      qui gen Y1 = Y0 + 3 // Generamos Y_{i}(1)
31
32      qui gen D = (runiform() > 0.7) // Generamos D_{i}
33
34      qui gen Y = Y1*D+(1-D)*Y0 // Generamos Y
35
36      qui reg Y D // Efectuamos la estimaci n
37
38      qui mat A[`j', `k'] = _b[D] // Almacenamos el resultado
39
40  }
41
42
43
44
45
46 }
47
48
49
50
51 clear
52 svmat A // Convertimos las estimaciones en datos
53
54
55
56
57
58
59 twoway (kdensity A1) (kdensity A2) (kdensity A3) (kdensity A4), legend(order(1 "n=10"
60   2 "n=20" 3 "n=100" 4 "n=1000")) graphregion(color(white)) plotregion(lcolor(black)
61   ) ylabel(, nogrid) ytitle(Densidad) xtitle("Efecto estimado") title("Par metro
62   estimado: distribuci n de Cauchy", color(black)) xline(0, lcolor(maroon) lpattern
63   (dash)) // Construimos el gr fico
64
65
66 -----
67 *Latex call
68
69 texdoc close
70 cap erase "$dir/cauchy.tex"
71 texdoc init "$dir/cauchy.tex", force
72
73
74 tex \begin{figure}[H]
75 tex \caption{Convergencia en distribuci n: distribuci n de Cauchy}
76 tex \centering
77 tex \includegraphics[scale=0.8]{Talleres/2022-1/Taller 1/Figuras/cauchy.pdf}
78 tex \label{fig:my_label}
79 tex \end{figure}
80
81
82 texdoc close

```

Figura 4: Convergencia en distribución: distribución de Cauchy



En este caso, los resultados clásicos de MCO no se cumplen porque no se tienen las condiciones válidas para aplicar el Teorema del Límite Central. En particular, variables aleatorias cuya distribución es Cauchy tienen varianza infinita, lo cual viola los supuestos necesarios para que el resultado de normalidad asintótica opere.

Finalmente, el problema que esto supone para la inferencia es que la distribución asintótica ya no es normal por lo que los errores estándar calculados bajo las hipótesis clásicas no serán válidos.

Finalmente, suponga que usted sabe que el programa implementado por el ministerio tiene el objetivo de nivelar los conocimientos de los estudiantes que en el participan. Por dicha razón, los resultados de las pruebas de los estudiantes que no participan en el programa son mas volátiles.

Una manera de modelar esta observación es a través del modelo:

$$Y_i = 3 * D_i + v_i$$

donde $v_i \sim N(0, 2 - D_i)$.

e) Para este nuevo modelo, realice el siguiente procedimiento:

- I. Simule 1000 muestras $\{(Y_i(0), Y_i(1), D_i\}_{i=1}^{1000}$. Para cada muestra, recupere los intervalos de confianza clásicos al 90 %, 95 % y 99 % obtenidos para δ y codifique en una matriz si dicho intervalo contiene o no a δ .

```

1
2
3 * Errores estandar normales
4
5
6 qui mat D = J(1000, 3, .) // Matriz para almacenar nuestros resultados.
7
8 forvalues i=1(1)1000{ // Loop sobre el numero de la simulacion
9
10    qui clear
11
12    qui set obs 1000 // Definimos el numero de observaciones por simulacion
13
14    qui gen D = (runiform()>0.7) // Generamos el tratamiento
15
16    qui gen v = .
17
18    qui replace v = rnormal(0, 2-D) // Generamos el error
19
20    qui gen Y = 3*D+v // Obtenemos el puntaje ICFES
21

```

```

22 qui reg Y D // Efectuamos la estimaci n
23
24 // Almacenamos en un local los IC para luego ver si el par metro verdadero
25 // se encuentra o no en ellos.
26
27 //90%
28 qui local lbD90 = _b[D] - invtail(e(df_r),0.05)*_se[D] //L mite inferior
29 qui local ubD90 = _b[D] + invtail(e(df_r),0.05)*_se[D] //L mite superior
30
31 //95%
32 qui local lbD95 = _b[D] - invtail(e(df_r),0.025)*_se[D]
33 qui local ubD95 = _b[D] + invtail(e(df_r),0.025)*_se[D]
34
35 //99%
36 qui local lbD99 = _b[D] - invtail(e(df_r),0.005)*_se[D]
37 qui local ubD99 = _b[D] + invtail(e(df_r),0.005)*_se[D]
38
39 // Codificamos en una matriz si el par metro est o no en los IC
40 qui mat D['i', 1] = ('lbD90'<=3 & 'ubD90'>=3)
41 qui mat D['i', 2] = ('lbD95'<=3 & 'ubD95'>=3)
42 qui mat D['i', 3] = ('lbD99'<=3 & 'ubD99'>=3)
43
44 }

```

- II. ¿Qué porcentaje de los intervalos de cada nivel contiene al parámetro verdadero? Presente sus resultados en una tabla ¿A qué se deben estos resultados? ¿Qué le sugiere esto sobre su forma de computar los intervalos?

```

1
2
3 // Convertimos en datos nuestros resultados
4 svmat D
5
6 // Obtenemos los promedios
7 sum D1
8 sum D2
9 sum D3

```

Los resultados se presentan en la siguiente tabla:

Resultados del procedimiento			
Nivel de confianza	90%	95%	99%
% de aciertos	97.3	99.2	100

Como se puede apreciar, el porcentaje de intervalos de confianza que contienen al parámetro verdadero es notablemente superior en todos los casos a lo teóricamente esperado. Esto se debe a que el error es heteroscedástico y, en particular, su varianza está negativamente relacionada con el valor de D_i . Por esta razón, la varianza del error se encuentra sobreestimada, lo que induce que los I.C. obtenidos sean demasiado anchos, lo que posiblemente conduce a ser demasiado conservadores al momento de efectuar pruebas de hipótesis. Dichos resultados sugieren que los intervalos no se están construyendo adecuadamente, por lo cual es necesario encontrar otra manera obtener los errores.

- III. Repita los numerales I y II, esta vez empleando errores estándar de White (robustos) en la construcción de sus intervalos. Concluya.

```

1
2
3
4 qui mat D = J(1000, 3, .) // Matriz para almacenar nuestros resultados.
5
6 forvalues i=1(1)1000{ // Loop sobre el n mero de la simulaci n.
7
8     qui clear
9
10    qui set obs 1000 // Definimos el n mero de observaciones por simulaci n
11
12    qui gen D = (runiform()>0.7) // Generamos el tratamiento
13
14    qui gen v = .
15
16    qui replace v = rnormal(0, 2-D) // Generamos el error
17

```

```

18 qui gen Y = 3*D+v // Obtenemos el puntaje ICFES
19
20 qui reg Y D, r // Efectuamos la estimación con errores robustos
21
22 // Almacenamos en un local los IC para luego ver si el par metro verdadero
   se encuentra o no en ellos.
23
24 //90%
25 qui local lbD90 = _b[D] - invtail(e(df_r),0.05)*_se[D] //Lmite inferior
26 qui local ubD90 = _b[D] + invtail(e(df_r),0.05)*_se[D] //Lmite superior
27
28 //95%
29 qui local lbD95 = _b[D] - invtail(e(df_r),0.025)*_se[D]
30 qui local ubD95 = _b[D] + invtail(e(df_r),0.025)*_se[D]
31
32 //99%
33 qui local lbD99 = _b[D] - invtail(e(df_r),0.005)*_se[D]
34 qui local ubD99 = _b[D] + invtail(e(df_r),0.005)*_se[D]
35
36 // Codificamos en una matriz si el par metro est o no en los IC
37 qui mat D['i', 1] = ('lbD90'<=3 & 'ubD90'>=3)
38 qui mat D['i', 2] = ('lbD95'<=3 & 'ubD95'>=3)
39 qui mat D['i', 3] = ('lbD99'<=3 & 'ubD99'>=3)
40
41 }
42
43
44 // Convertimos en datos nuestros resultados
45 summat D
46
47 // Obtenemos los promedios
48 sum D1
49 sum D2
50 sum D3

```

Los resultados se tabulan a continuación

Resultados del procedimiento			
Nivel de confianza	90 %	95 %	99 %
% de aciertos	89.4	94.8	99.2

Como se puede apreciar, en este caso los porcentajes de intervalos que contienen al parámetro verdadero son notablemente próximos a los teóricos. Esto indica que los errores robustos enfrentan adecuadamente el problema de heteroscedasticidad mencionado en el punto anterior.

Punto doctorado

Usted se encuentra cursando el curso de seminario de investigación doctoral. En las primeras semanas le piden que presente algunas relaciones de causalidad que tenga pensado abordar en su tesis. Su profesora de seminario lo invita a que avance en el planteamiento de sus ideas a través del uso de gráficos acíclicos dirigidos (Directed Acyclic Graph - DAG).² En breve, los DAGs se componen de dos elementos: flechas y variables. Las flechas entre variables indican causalidad en el sentido en el que esta apuntando la flecha. Un **Camino** es cualquier conexión entre dos variables realizada por flechas, sin importar su dirección o si existen otras variables intermedias. Un **Camino por la puerta trasera** de la variable X a la variable Y es un camino que empieza con una flecha dirigida hacia X . Un **Colisionador** es una variable a la que apuntan dos flechas en un camino.

- a) Usando puntos negros para graficar las variables observadas y circunferencias para graficar las variables no observadas (X , Z y Y), realice los DAGs que representan:
 - i) X tiene una relación de causalidad hacia Y . A su vez, Z tiene una relación de causalidad hacia X y otra hacia Y .
 - ii) X tiene una relación de causalidad hacia Z . A su vez, Z tiene una relación de causalidad hacia Y .
 - iii) La variable no observada: U tiene una relación de causalidad hacia Z y otra hacia Y . A su vez, X tiene una relación de causalidad hacia Y y otra hacia Z .
 - iv) La variable X tiene una relación de causalidad hacia Z y otra hacia Y . A su vez, Y tiene una relación de causalidad hacia Z

²Si no se encuentra familiarizado con los DAGs, una excelente referencia y explicación se encuentra en el tercer capítulo del libro de Scott Cunningham: Causal Inference: The mixtape, del año 2021. <https://mixtape.scunning.com/>

- v) La variable no observada: U tiene una relación de causalidad hacia Z y otra hacia Y. A su vez, X tiene una relación de causalidad hacia Y, mientras que Z tiene una relación de causalidad hacia X.

Solución: Los DAGs que representan cada numeral respectivamente, son:

Figura 5: Gráficos acíclicos dirigidos

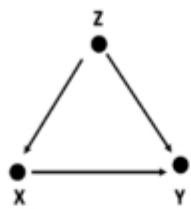


Figura 4: DAG i)

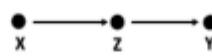


Figura 5: DAG ii)

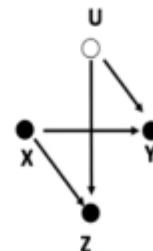


Figura 6: DAG literal iii)

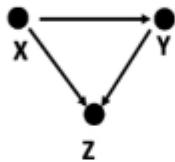


Figura 7: DAG iv)

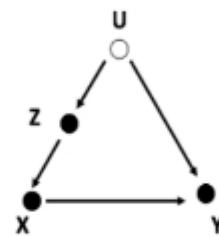


Figura 8: DAG v)

- b) Ahora usted quiere implementar un modelo de regresión lineal simple para evaluar los efectos causales de X sobre Y . Con este objetivo, usted está determinando la conveniencia de incluir la variable Z como variable de control en su modelo. Argumente, para cada caso representado por los DAGs, si la inclusión de Z , como control en el modelo, es apropiada. Para sus respuestas, tenga en cuenta que un variable es un mal control si: bloquea caminos causales entre X y Y o abre otros caminos que no son causales entre X y Y .

(Pista: Bloquear un camino es equivalente a controlar por variables que no son Colisionadoras o no controlar por las Colisionadoras. Abrir un camino es equivalente a controlar por variables que son Colisionadoras o no controlar por no Colisionadoras)

Solución: El caso representado en el literal i) es un buen control. Existe un camino no causal entre X y Y el cual es bloqueado, ya que la variable de control Z no es una colisionadora y se controla por ella. Además, no bloquea caminos causales entre X y Y y no abre otro caminos no causales entre estas dos variables. Mientras tanto, para el el caso representado en ii), Z es un mal control. Este hecho se fundamenta en que se bloquean caminos causales entre X y Y al controlar por la variable no colisionadora Z . En cuanto al DAG del literal iii), Z también es un mal control ya que, en este ejemplo, se abren caminos no causales entre las variables X y Y . En particular, al controlar por Z , que ahora es colisionadora, se abre otro camino no causal entre X y Y . Por otro lado, en el caso del DAG iv), también estamos en presencia de un mal control debido a que Z , nuevamente, es colisionadora y al controlar por esta variable se abre otro camino no causal entre X y Y . Por último, en el quinto DAG, Z es un buen control como consecuencia de que al ser no colisionadora y controlar por ella, se bloquea el camino no causal entre X y Y que pasa por la variable no observada: U . Z no bloquea caminos causales entre X y Y , ni tampoco abre caminos no causales entre estas dos variables.

- c) Finalmente, usted sabe que la variable X es independiente de los resultados potenciales de Y y de Z . Por otro lado, usted sospecha que Z podría ser un mal control y que este hecho puede generarse porque la variable Z es, a su vez, una variable de resultado en su modelo. Plantee una forma de aproximarse empíricamente para evaluar si este hecho puede estar ocurriendo

Solución: Una manera de aproximarse empíricamente a este hecho podría ser estimar una regresión en la que la variable de resultado sea la variable Z . Si se rechaza la hipótesis nula correspondiente a que el parámetro que acompaña a la variable X es igual a cero, se tendrá evidencia de que la variable X tiene incidencia en la variable Z , lo cual se constituiría en un indicio de que Z , en efecto, también es una variable de resultado y puede ser un mal control.

