Tarea 3: Econometría

Pablo Braga [1], Cristóbal Donoso[2], Sofía Muñoz [3], Esteban Puentes [4], Camila Carrasco [5], y Santiago Garcıa [6]

1-3 Estudiantes de Magíster en Economía
4 Profesor
5-6 Ayudantes

May 16, 2025

1 Varianza con asignación individual (10 puntos)

Suponga que un programa se asigna aleatoriamente a individuos, con probabilidad P de recibir tratamiento. Cada individuo tiene un resultado Y_i con varianza σ^2, y las observaciones son independientes entre sí.

Demuestre que el estimador de diferencias de medias entre el grupo tratado y el grupo de control, definido como:

$$\hat{\beta} = \bar{Y}_1 - \bar{Y}_0 \tag{1}$$

tiene varianza:

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{NP(1-P)} \tag{2}$$

donde N es el tamaño total de la muestra.

En el contexto dado, es posible utilizar un modelo de regresión simple dado por:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \tag{3}$$

donde X_i es una variable dummy que indica si el individuo (i) recibió el tratamiento $(X_i = 1)$ o no $(X_i = 0)$. En este modelo, α representa la media del grupo de control (cuando $X_i = 0$), y β representa la diferencia de medias entre el grupo tratado y el de control (el efecto del tratamiento).

Note que,

$$\begin{aligned} \mathbb{E}[Y_i|D_i=0] &= \alpha+\beta(0)+\mathbb{E}[\varepsilon_i|X_i=0] \\ &= \alpha \qquad (4) \\ \mathbb{E}Y_i|D_i=1 &= \alpha+\beta(1)+\mathbb{E}[\varepsilon_i|X_i=1]=\alpha+\beta \\ &= \alpha+\beta \qquad (5) \end{aligned}$$

Reemplazando (4) en (5),

$$\begin{aligned} \mathbb{E}Y_i|D_i=1 &= \alpha+\beta \\ &= \mathbb{E}[Y_i|D_i=0]+\beta \\ \beta &= \mathbb{E}[Y_i|D_i=1]-\mathbb{E}[Y_i|D_i=0] \qquad (6) \end{aligned}$$

lo cual coincide con la expresión 1 afirmando que el estimador $\hat{\beta}$ corresponde a la diferencia entre las medias condicionales.

Ahora calculamos la varianza del estimador $\hat{\beta}$ como,

$$\text{Var}(\hat{\beta})=\text{Var}(\bar{Y_1})+\text{Var}(\bar{Y_0}). \qquad (7)$$

Bajo el supuesto de homocedasticidad[a],

$$\text{Var}(\hat{\alpha})=\frac{\sigma^2}{n_0} \quad \text{Var}(\hat{\beta})=\frac{\sigma^2}{n_1}$$

Entonces,

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \frac{\sigma^2}{n_0}+\frac{\sigma^2}{n_1} \\ &= \frac{\sigma^2}{N(1-P)}+\frac{\sigma^2}{N(P)}=\frac{\sigma^2}{NP(1-P)} \qquad (8) \end{aligned}$$

donde N es el numero total de individuos. ■

[a] En esta derivación se utiliza que $n_1 = NP$ y $n_0 = N(1-P)$, lo cual no es exacto sino que representa el valor esperado bajo asignación aleatoria. Si $D_i \sim \text{Bernoulli}(P)$, entonces $n_1 = \sum_{i=1}^{N} D_i$ es aleatorio, pero por la Ley de los Grandes Números se cumple que $\frac{n_1}{N} \to P$ en probabilidad. Por ello, cuando N es grande, es habitual suponer $n_1 \approx NP$ y $n_0 \approx N(1-P)$ para simplificar la notación y los cálculos.

2 Varianza con asignación por grupos (15 puntos)

Ahora suponga que el tratamiento se asigna aleatoriamente a grupos (clústeres) en lugar de individuos, y que hay J grupos de tamaño n cada uno. Los individuos dentro de un grupo tienen resultados correlacionados, con:

- varianza individual σ^2, y
- correlación intra-cluster $\rho = \text{Corr}(Y_{ij}, Y_{kj})$ para j (mismo grupo), $i \neq k$.

Demuestre que la varianza del estimador de diferencia de medias entre grupos tratados y de control es:

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{JP(1-P)} \cdot \left(\rho + \frac{1-\rho}{n}\right)$$

Suponemos que cada grupo se asigna al tratamiento (o al control) de forma independiente de los demás, y que los resultados entre distintos grupos no están correlacionados una vez que sabemos qué grupos fueron tratados. Es decir,

$$\text{Cov}(Y_{ij}, Y_{kl} \mid D) = 0 \quad \text{para } j \neq l.$$

Sea J_1 el número de grupos tratados y $J_0 = J - J_1$ el número de grupos de control. En promedio, bajo asignación aleatoria a nivel de grupo:

$$\mathbb{E}[J_1] = JP, \quad \mathbb{E}[J_0] = J(1-P).$$

Definimos el estimador como la diferencia de medias entre los grupos tratados y de control:

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

donde:

$$\bar{Y}_1 = \frac{1}{J_1 n} \sum_{j \in \mathcal{T}} \sum_{i=1}^{n} Y_{ij}, \quad \bar{Y}_0 = \frac{1}{J_0 n} \sum_{j \in \mathcal{C}} \sum_{i=1}^{n} Y_{ij},$$

y $\mathcal{T}$, $\mathcal{C}$ son los conjuntos de grupos tratados y de control, respectivamente.

Como la asignación es a nivel de grupo y los grupos son independientes entre sí, se tiene:

$$\text{Var}(\hat{\beta}) = \text{Var}(\bar{Y}_1) + \text{Var}(\bar{Y}_0).$$

Ahora, sea la media del grupo j:

$$\bar{Y}_j = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}.$$

Entonces, podemos escribir:

$$\bar{Y}_1 = \frac{1}{J_1}\sum_{j\in\mathcal{T}} \bar{Y}_j, \quad \bar{Y}_0 = \frac{1}{J_0}\sum_{j\in\mathcal{C}} \bar{Y}_j.$$

Dado que los grupos son independientes, las varianzas de estas medias son:

$$\text{Var}(\bar{Y}_1) = \frac{1}{J_1}\text{Var}(\bar{Y}_j), \quad \text{Var}(\bar{Y}_0) = \frac{1}{J_0}\text{Var}(\bar{Y}_j), \quad \forall j \in \mathcal{J}$$

donde $\mathcal{J}$ el conjunto de todos los grupos. Por lo tanto:

$$\text{Var}(\hat{\beta}) = \left(\frac{1}{J_1} + \frac{1}{J_0}\right)\text{Var}(\bar{Y}_{j\in\mathcal{J}}). \tag{9}$$

Luego para calcular la varianza de la media del grupo,

$$\text{Var}(\bar{Y}_{j\in\mathcal{J}}) \;=\; \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_{ij}\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} Y_{ij}\right). \tag{10}$$

Ahora usamos la estructura de correlación intra-cluster,

$$\begin{aligned}
\text{Var}\left(\sum_{i=1}^{n} Y_{ij}\right) &= \sum_{i=1}^{n}\text{Var}(Y_{ij}) + \sum_{i\neq k}\text{Cov}(Y_{ij}, Y_{ik}) \\
&= n\sigma^2 + n(n-1)\rho\sigma^2 \\
&= \sigma^2 n[1 + (n-1)\rho], \qquad (11)
\end{aligned}$$

Ahora, reemplazando 11 en la Ecuación 10

$$\text{Var}(\bar{Y}_j) = \frac{1}{n^2} \cdot \sigma^2 n[1 + (n-1)\rho] = \frac{\sigma^2[1 + (n-1)\rho]}{n}.$$

Finalmente, reemplazamos en la expresión 9 y obtenemos:

$$\text{Var}(\hat{\beta}) = \left(\frac{1}{J_1} + \frac{1}{J_0}\right) \cdot \frac{\sigma^2[1 + (n-1)\rho]}{n}.$$

Utilizando los valores de esperados,

$$J_1 \approx JP, \quad J_0 \approx J(1-P),$$

entonces:

$$\text{Var}(\hat{\beta}) \approx \left(\frac{1}{JP} + \frac{1}{J(1-P)}\right) \cdot \frac{\sigma^2[1 + (n-1)\rho]}{n} = \frac{1}{JP(1-P)} \cdot \frac{\sigma^2[1 + (n-1)\rho]}{n}. \quad (12)$$

Desarrollando el termino en (12), la varianza del estimador es:

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{JP(1-P)} \cdot \left(\rho + \frac{1-\rho}{n}\right).$$

Casos extremos:

- Si $\rho = 0$, no hay correlación intra-grupo y obtenemos la fórmula clásica para aleatorización individual:

  $$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{JnP(1-P)} = \frac{\sigma^2}{NP(1-P)},$$

  donde $N = Jn$ es el número total de individuos.

- Si $\rho = 1$, todos los individuos del grupo son idénticos, y la varianza depende sólo del número de grupos:

  $$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{JP(1-P)}.$$

3 Cálculo del MDE usando datos agrupados (5 puntos)

Suponga que tiene un experimento con aleatorización a nivel de grupo. Hay $J = 40$ grupos, cada uno con $n = 10$ personas. La proporción de tratamiento es $P = 0.5$, la varianza individual del resultado es $\sigma^2 = 1$, y la correlación intra-grupo es $\rho = 0.05$. Calcule el tamaño mínimo detectable (MDE) para un test bilateral con nivel de significancia $\alpha = 0.05$ y poder estadístico del 80%.

La fórmula general para el MDE en este contexto es:

$$\text{MDE} = (z_{1-\alpha/2} + z_{1-\beta}) \cdot \text{SE}$$

donde el error estándar ajustado por el diseño de clúster es:

$$\text{SE} = \sqrt{\frac{\sigma^2}{JnP(1-P)} \cdot (1 + (n-1)\rho)}$$

Con los valores entregados:

$$\text{SE} = \sqrt{\frac{1}{40 \cdot 10 \cdot 0.5 \cdot 0.5} \cdot (1 + 9 \cdot 0.05)} = \sqrt{\frac{1.45}{100}} = \sqrt{0.0145} \approx 0.1204$$

Los valores críticos para los cuantiles normales estándar son:

$$z_{0.975} \approx 1.96, \quad z_{0.80} \approx 0.84$$

Por lo tanto:

$$\text{MDE} = (1.96 + 0.84) \cdot 0.1204 = 2.80 \cdot 0.1204 \approx 0.337$$

Conclusión: el efecto mínimo que este diseño puede detectar con 80% de poder y 5% de significancia es aproximadamente **0.337**.

4 Simulación Monte Carlo para detectar efectos pequeños (20 puntos)

Simule un experimento con las siguientes características:

- 40 grupos, 10 individuos por grupo
- Varianza individual del resultado: 1
- ICC (ρ) = 0.05
- Proporción de tratamiento: 0.5

a) Simule 1000 experimentos donde el efecto verdadero es de 0.3 desviaciones estándar. ¿En qué proporción se rechaza H_0?

> El poder de un test se define como la probabilidad de rechazar H_0 cuando H_1 es verdadera. En términos generales se expresa como:
>
> $$\text{Poder} = P[T > t_\alpha | H_1 \text{ verdadero}]$$
>
> Bajo este contexto, el experimento de Monte Carlo para datos agrupados, con 1000 simulaciones y un valor verdadero de 0.3 desviaciones estándar, arroja un poder efectivo del 66.5%. Esto significa que, de las 1000 simulaciones, el test logra detectar el efecto en 665 de ellas. Sin embargo, este resultado aunque cercano, se encuentra por debajo del 80% típicamente deseado.

b) Repita el experimento con un efecto verdadero de 0.4 desviaciones estándar. Compare los resultados. Explique sus resultados.

> Para comparar, se simula nuevamente el experimento, esta vez considerando un efecto verdadero de 0.4 desviaciones estándar. Como resultado, se obtiene un poder efectivo de un 89.6%, valor que supera el umbral del 80% esperado y contrasta con el resultado obtenido en la parte (a).
> La diferencia entre estos resultados puede explicarse a partir de lo observado en la pregunta 3, en la cual se calculó que, bajo los mismos parámetros y diseño de este ejercicio, el efecto mínimo detectable del test es de 0.337. Esto implica que, si el valor

verdadero está sobre dicho umbral, el poder del test será alto, detectando en una mayor probabilidad su efecto. Ese es el caso de esta simulación, donde $0.4 > 0.337$, y el poder alcanzado es de 89.6%.

En cambio, para un valor verdadero menor al MDE, como ocurre en (a) con $0.3 < 0.337$, el test tiene menores probabilidades de detectar el efecto de forma idónea, lo que se refleja en un poder más bajo de 66.5%.

Incluso sin el valor del MDE disponible, se pueden explicar las diferencias de poder. Esto ya que, por construcción del estadístico T (13), mientras más pequeño sea el valor verdadero, más difícil será para T rechazar H_0.

$$T = \frac{\hat{\beta}}{SE(\hat{\beta})} \tag{13}$$

5 Error al usar fórmula individual cuando hay agrupamiento (20 puntos)

Suponga que se desea detectar un efecto de 0.1 desviaciones estándar (0.1σ) usando la fórmula para el MDE bajo aleatorización individual, con $\alpha = 0.05$ y poder estadístico de 80%. La fórmula es:

$$\text{MDE}_{\text{indiv}} = \frac{z_{1-\alpha} + z_{1-\kappa}}{\sqrt{NP(1-P)}} \cdot \sigma$$

a) Calcule el tamaño muestral N necesario para detectar un efecto de 0.1 usando esta fórmula. Use $P = 0.5$, $\sigma = 1$.

Sustituyendo los valores dados,

$$\begin{aligned}
0.1 \cdot \sigma &= \frac{z_{0.95} + z_{0.80}}{\sqrt{N \cdot 0.5 \cdot (1-0.5)}} \cdot \sigma \\
0.1 &= \frac{1.65 + 0.84}{\sqrt{N \cdot 0.25}} \quad (\text{con } \sigma = 1) \\
0.1 &= \frac{2.49}{\sqrt{0.25N}} = \frac{2.80}{0.5\sqrt{N}} \\
0.5\sqrt{N} &= \frac{2.49}{0.1} \\
\sqrt{N} &= \frac{24.9}{0.5} = 49.8 \Rightarrow N = 2480.04 \qquad (14)
\end{aligned}$$

Por lo tanto, bajo el supuesto de observaciones i.i.d., se necesitarían ~ 2480[a] individuos para detectar un efecto de 0.1 desviaciones estándar con un 80% de poder y un nivel de significancia del 5%.

[a] Asumiendo que no podemos tener una fracción de individuo. Entonces vale redondear.

b) Ahora, suponga que en realidad la aleatorización se hizo a través de 4 grupos iguales (clústeres), cada uno con $N/4$ observaciones, y que existe una correlación intra-grupo de $\rho = 0.05$.

Genere un experimento de Monte Carlo en el que simule esta estructura (aleatorización por grupos con correlación intra-cluster) y repita la estimación 1000 veces. Calcule en qué proporción se rechaza la hipótesis nula de que $\beta = 0$.

Para la prueba de significancia del coeficiente, se utilizó un estadístico t con $G-1$ grados de libertad, donde $G = 4$ es el número de grupos. Aunque el uso de la distribución t con $G-1$ grados de libertad en presencia de un número reducido de grupos (e.g., $G < 50$ o incluso $G < 20$ en algunos contextos) puede llevar a una tasa de rechazo de la hipótesis nula superior a la nominal, esta aproximación es una práctica común y recomendada como un ajuste mínimo en comparación con el uso de la distribución normal estándar [1].

c) Compare el poder efectivo de este diseño con el supuesto de independencia. Comente los resultados.

El experimento simulado bajo aleatorización con correlación intra-grupo de $\rho = 0.05$ y 4 clústeres de tamaño 620 arrojó un poder efectivo de 13.5%. Es decir, la hipótesis nula $\beta = 0$ se rechazó en una proproción mucho menor 80%, que es aproximadamente el valor que por construcción se podría haber alcanzado si las observaciones hubiesen sido independientes.

La explicación detrás del resultado reside justamente en el levantamiento de este último supuesto. Matemáticamente, eso se tradujo en un aumento en el varianza del estimador de β (lo que es consistente con la fórmula que demostró en secciones previas del presente documento). A su vez, dicho aumento llevó a que los estadísticos t cayeran y de esa manera se tendieran a alejar del umbral de define la zona de rechazo.

A nivel intuitivo, la diferencia puede pensarse en términos de la información disponible. En el caso simulado, las observaciones de un mismo cluster sorteado compartían una característica que hacía que sus resultados correlacionaran entre sí. En la práctica, eso implica que cada individuo aporta poco para la detección de efectos sistemáticos de un programa o tratamiento. Es por eso que pese a haber usado el tamaño muestral calculado en el ítem (a), es menos probable que podamos identificar la presencia de un β estadísticamente significativo.

En línea con lo anterior, se concluye que si se desea alcanzar un MDE de 0.1 con un mayor poder empírico es indispensable aumentar el número de grupos de tratamiento y control. En efecto, como ilustran la Figura 1, subir el tamaño de cada uno de ellos tiene un impacto marginal en el rechazo de H_0. En cambio, dado un N_g cualquiera, un incremento en G tiene un impacto mucho más grande en la heterogeneidad de los datos

disponibles, lo que facilita la detección de efectos más acotados.

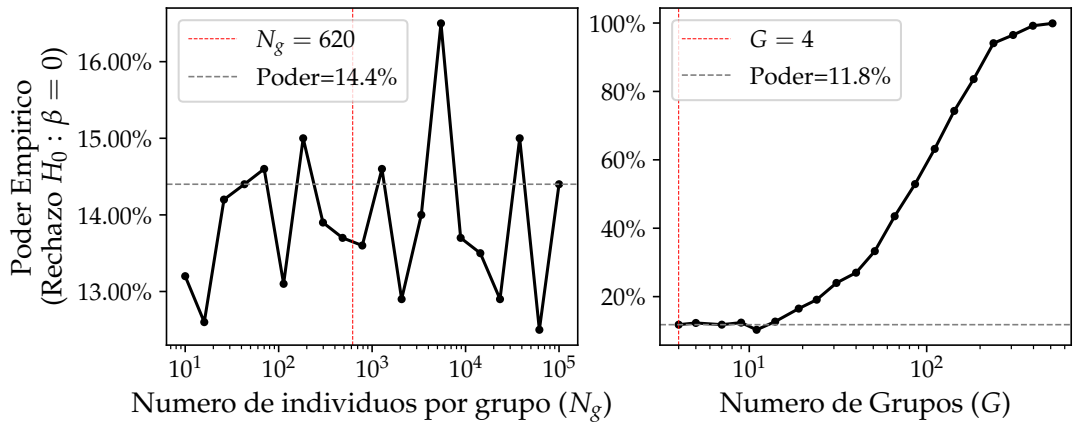

Figure 1: Poder empírico de la prueba t para rechazar la hipótesis nula ($\beta = 0$) en función del número de individuos por grupo (N_g) y el número de grupos (G). El **panel izquierdo** muestra la variación del poder con N_g manteniendo $G = 4$ constante, mientras que el **panel derecho** ilustra el impacto de aumentar G con $N_g = 620$ fijo. Todas las simulaciones se realizaron con una correlación intra-clúster (ρ) de 0.05, un efecto verdadero de 0.1 y un nivel de significancia (α) de 0.05.

References

[1] A Colin Cameron and Douglas L Miller. A practitioner's guide to cluster-robust inference. *Journal of human resources*, 50(2):317–372, 2015.