

Actividad 6: Resultados Potenciales

Fecha de entrega: Domingo 17 de Abril, antes de las 11:59pm, por SICUA.

Valor en la nota final de curso: 10%.

Evaluación tipo X: la calificación de esta actividad pondera el desempeño grupal y el esfuerzo individual. Además de la entrega grupal, cada integrante debe evaluar individualmente el esfuerzo de sus compañeros en una escala de 0 a 100 por medio del enlace [Eval - X Actividad 6](#), en la [página web del curso](#).¹ La nota de la actividad no será entregada hasta que todos los integrantes del grupo evalúen a sus compañeros.

1 Introducción

La educación de calidad es un poderoso instrumento de transformación social. Uno de los grandes retos que tienen las políticas de educación es disminuir las tasas de deserción escolar y aumentar la tasa de asistencia a clase de niños, niñas y adolescentes.

Para muchos investigadores y formuladores de políticas públicas la menstruación es uno de los factores que explica la ausencia de clase de las niñas en los colegios en países en desarrollo. La hipótesis es que, al no contar con los productos sanitarios adecuados, las niñas faltan a clase en los días de su periodo. Esto ha llevado a numerosos esfuerzos de política pública para llevar productos sanitarios adecuados a estas población. Los costos de estos esfuerzos rondan los millones de dólares (?).

En este taller vamos a estudiar la relación empírica entre asistencia escolar, menstruación y productos sanitarios. Los datos que vamos a usar son una versión modificada de los datos del artículo de ?.² En particular, los datos provienen de encuestas realizadas entre el 2006 y el 2008 en el distrito de de Chitwan en Nepal.

2 Análisis Preliminar

1. Para motivar este trabajo, explique tres impactos económicos negativos de la inasistencia o el abandono del sistema educativo escolar por parte de mujeres jóvenes.

¹Recuerden, del programa del curso, que 0 indica que su compañero de grupo no realizó ningún esfuerzo en el desarrollo del trabajo, mientras que 100 indica que realizó el mismo esfuerzo que el resto de los integrantes.

²Las bases de datos de este taller están disponibles en la página web de Nicolás. Pueden consultar el artículo y la base de datos original aquí: <https://www.aeaweb.org/articles?id=10.1257/app.3.1.91>

2. Cargue la base de datos *he1-ndr-taller-datos-apreliminar.dta*. Cada observación corresponde a una niña y las variables son:

- $falto_i$: variable dummy igual a 1 para las niñas que faltaron a clase durante por lo menos uno de sus periodos menstruales. i es un subíndice para cada niña.
- uso_i : variable dummy igual a 1 para las niñas que usan productos sanitarios durante su periodo.
- $ingreso_i$: categoría del nivel de ingresos del hogar de la adolescente, donde 0 es la categoría más baja y 6 la más alta.
- $edad_i$: edad de la niña.

3. Estamos interesados en establecer el efecto de uso_i en $falto_i$. Estime por Mínimos Cuadrados Ordinarios (MCO) la siguiente ecuación:

$$falto_i = \beta_0 + \beta_1 uso_i + e_i \quad (E1)$$

donde e_i es el término del error. Reporte los resultados en una tabla.³

4. El valor estimado de β_1 (llamémoslo $\hat{\beta}_1$) es positivo. ¿Ese valor representa el efecto causal de la variable uso_i en $falto_i$? Si no, ¿qué representa entonces?
5. Piense en un formulador de política que le da una interpretación causal a su $\hat{\beta}_1$. ¿Qué recomendación de política daría para reducir la ausencia escolar en los días del periodo menstrual?
6. Preocupado por el signo de $\hat{\beta}_1$, un econometrista ingenuo le sugiere que use controles en la regresión y en particular le sugiere que estime por MCO:

$$falto_i = \alpha_0 + \alpha_1 uso_i + \alpha_2 edad_i + u_i \quad (E2)$$

donde las variables de la ecuación son las descritas arriba y u_i es el término del error.

- (a) ¿Por qué tiene sentido controlar por $edad_i$?
- (b) Reporte los resultados de estimar (E2) por MCO.
- (c) Note que el valor estimado de α_1 (es decir, $\hat{\alpha}_1$) es diferente al de $\hat{\beta}_1$. Comente sobre el $\hat{\alpha}_1$ teniendo en cuenta su significancia estadística. ¿Por qué cree que cambiaron los resultados al controlar por $edad_i$?
- (d) ¿Tiene $\hat{\alpha}_1$ una interpretación causal?
- (e) ¿Qué condición se tiene que cumplir para $\hat{\alpha}_1$ tenga una interpretación causal? Use los términos de la ecuación (E2) para escribir esta condición.
- (f) Explique la intuición de la condición anterior. Puede usar un dibujo. Use dicha intuición para argumentar si se cumple o no se cumple la condición.
7. Viendo sus resultados anteriores el econometrista ingenuo se preocupa y le sugiere que vuelva a estimar la ecuación pero que esta vez controle por $ingreso_i$. Y le promete que esta vez si controla por $ingreso_i$, el coeficiente estimado de uso_i sí tiene una interpretación causal. Esta vez usted se niega a hacerlo. Explique por qué.

³Noten que aquí estamos estimando una regresión que involucra variables dicótomas. Aunque en Econometría 2 van a ver los métodos de estimación *Probit* y *Logit* para modelos con estas variables, en este taller hacemos caso omiso de eso y tratamos estas variables como si fueran continuas.

3 Análisis del Experimento

Emily Oster y Rebecca Thornton se pusieron en la tarea de establecer, con la ayuda de un experimento aleatorio, el efecto del acceso a implementos sanitarios en la asistencia escolar. El experimento se corrió desde noviembre de 2006 hasta enero de 2008 en cuatro colegios de Chitwan en Nepal. Las investigadoras invitaron a las estudiantes de grados séptimo y octavo junto con sus madres a una reunión. En ella hicieron una encuesta con preguntas sociodemográficas, de escolaridad y relacionadas con la menstruación.⁴ Al final de la reunión, se realizó una lotería donde 25 estudiantes de cada colegio fueron asignadas al grupo de tratamiento. Al grupo de tratamiento (y sus madres) se les regaló una copa menstrual con las instrucciones de uso. En el 2008 se regresó a los colegios a recoger información adicional.

3.1 Diseño del Experimento

En esta sección trataremos de establecer si el experimento de Emily Oster y Rebecca Thornton quedó bien hecho.

1. Cargue la base *he1-ndr-taller-datos-baselinesurvey.dta*. Esta base contiene información de la primera encuesta que hicieron las investigadoras. En ella se entrevistaron 198 niñas y se le dio la copa a 101. Con el comando *tab* de STATA, verifique que este sea el caso y reporte los resultados.
2. En clase vimos que en un experimento aleatorio bien diseñado, el promedio de cualquier variable debe ser el mismo entre el grupo de tratamiento y el grupo de control. ¿Qué resultado que vimos en clase garantiza que esto se cumpla?⁵ Explique en detalle con una variable (por ejemplo, la edad de la niña).
3. La base de datos contiene las siguientes variables:
 - *edumadre_i*: Años de educación de la madre.
 - *edupadre_i*: Años de educación del padre.
 - *ingreso_i*: Categoría del nivel de ingresos del hogar de la adolescente, donde 0 es la categoría más baja y 6 la más alta.
 - *edad_i*: Edad de la niña.
 - *hindupadre_i*: Variable binaria que toma el valor de 1 si el padre es hindú, 0 de lo contrario.

⁴Las respuestas a algunas de esas preguntas fueron las que usaron en el numeral 2. de este taller.

⁵Pista: es una “ley” muy importante.

Si el experimento aleatorio quedó bien hecho, ¿qué esperaría usted respecto a la media de esas variables entre los grupo de tratamiento y de control? Haga una prueba estadística para estudiar esto y reporte los resultados. Explique la prueba estadística que utiliza y úsela con cada una de las variables.⁶

4. Hay variables no observables que pueden ser diferentes entre el grupo de tratamiento y el grupo de control. ¿Qué sugieren los resultados de los dos numerales anteriores respecto a esto? Explique en detalle.
5. Note que todas las niñas de los colegios fueron invitadas a las reuniones, independientemente de si recibieron o no la copa. Puede ser que algunas de las niñas que no recibieron la copa igual terminaran usándola. Discuta las implicaciones de esto respecto a la validez de los resultados del experimento.

3.2 Efectos del Tratamiento

1. Cargue la base de datos *he1-ndr-taller-datos-experimento.dta*. Esta base de datos contiene información diaria para cada niña y sus variables son las siguientes:
 - id_i : Identificador de la niña.
 - $asistencia_{it}$: Asistencia al colegio de la niña i en el día t . 0 indica que la estudiante no asistió a clase y 1 que sí asistió.
 - $tratamiento_i$: Variable binaria que toma el valor de 1 si la niña fue tratada (recibió la copa menstrual), 0 si no lo fue.
 - $despues_{it}$: Variable binaria que toma el valor de 1 si la información es de la ventana de tiempo de después del experimento, 0 de lo contrario.
 - $periodo_{it}$: Variable que toma el valor de 1 si la niña i tuvo el periodo en el día t .
 - $agno_{it}$, mes_{it} y dia_{it} : Variables que indican el día de la observación.

Note que algunas de las variables fueron medidas antes del experimento (el día de la primera reunión) y otras fueron medidas después. Esto es lo que indica la variable $despues_{it}$.

2. Usted está interesado en establecer el efecto causal de la copa menstrual en la asistencia escolar, para las niñas que tienen el periodo. Para esto hay que comparar las medias en asistencia escolar entre dos grupos. ¿Cuáles son esos dos grupos para los cuales hay que comparar las medias?⁷ Escriba la diferencia de medias que le permite establecer dicho efecto causal.

⁶Pista: use el comando *ttest* de Stata.

⁷Note que hay tres cosas que hay que tener en cuenta: 1) En qué momento del tiempo se está (valor de la variable $despues_{it}$), 2) si las niñas están o no con el periodo y 3) si las niñas están en el grupo de tratamiento o en el grupo de control.

3. Estime las medias del numeral anterior y presente los resultados de un test de diferencia de medias. ¿Qué puede concluir del resultado obtenido?
4. Explique por qué la comparación de medias del numeral anterior es “Ceteris-Paribus”. Puede referirse a sus respuestas de la Sección 3.1.

5. Ahora considere la ecuación:

$$asistencia_{it} = \alpha_0 + \alpha_1 tratamiento_i + u_i$$

donde $asistencia_{it}$ y $tratamiento_i$ son las mismas variables que arriba y u_i es un término de error. Estime esa ecuación *solo* en la muestra que cumple con $despues_{it} = 1$ y $periodo_{it} = 1$. Presente sus resultados en una tabla.

6. ¿Cómo se comparan los resultados del numeral 3.2.3 y 3.2.5?
7. Tiene el valor estimado de α_1 en la ecuación anterior (llamemos esa estimación $\hat{\alpha}_1$) una interpretación causal? ¿Por qué sí o por qué no?
8. Dados los resultados anteriores, ¿recomendaría usted la repartición de copas menstruales gratis como una política para aumentar la asistencia escolar? Comente en relación a la introducción de este taller.
9. Discuta una alternativa de política pública para mejorar la asistencia escolar de niñas y adolescentes en países en desarrollo.
10. Una de las críticas a los experimentos aleatorios tiene que ver con la validez externa. La preocupación es que los resultados del experimento cambian cuando el experimento se hace en otro contexto. ¿Cree usted que los resultados de este experimento se mantendrían en cualquier contexto? De un ejemplo de una población donde usted cree que no y discuta en detalle por qué.