

Microeconometría aplicada

Taller 2

Mauricio Romero

23 de septiembre de 2021

1. Instrucciones generales

- Pueden usar R, Stata o cualquier otro programa
- Deben entregar un archivo PDF. Algo así como un memo con los que hicieron y lo que encontraron.
- Adicionalmente, deben entregar un archivo ZIP con los datos crudos (raw data), y todos los códigos (scripts) que usen.
- Yo debería poder replicar sus resultados descomprimiendo el archivo zip y cambiando (únicamente) el directorio principal al principio y corriendo el código.
- Envíen sus respuestas (PDF y ZIP) a: mtromero+microeconometria@itam.mx
- Van a ser evaluados no solo en la respuesta final. También tendré en cuenta el proceso que usen para llegar a sus respuestas (i.e., su código y los comentarios de ese código).
- Es muy importante que escriban su código (y los comentarios) de manera que sea fácil para otros (yo en este caso) seguir el código y entender lo que se hizo.
- Usar recursos adicionales (internet, libros, blogs, etc.) está bien (y muchas veces es necesario), pero por favor menciones los recursos que usan y cítelos adecuadamente.
- La fecha de entrega es octubre 8 de 2021 a las 12:00 (medio día)
- La tarea se puede hacer en grupos (del tamaño que quieran)

2. Problema 1 - Experimento en educación

Suponga que el gobierno está considerando duplicar el salario de los docentes. Antes de expandir esta política a todas las escuelas públicas, deciden hacer un piloto y que sea evaluado mediante un experimento aleatorio. Uds. están encargados de diseñar el experimento. Suponga que Ud. tiene cierto margen de elegir cuantas escuelas pueden estar en este piloto (entre 100 y 200).

1. Primero debemos diseñar el experimento

- a) Discuta los méritos de asignar maestros aleatoriamente a que se les duplique el salario o no, sin importar la escuela (i.e., asignar el tratamiento a nivel individual). [Hint: ¿Se cumpliría SUTVA?]
- b) Discuta los méritos de asignar escuelas aleatoriamente a que se les duplique el salario a todos sus maestros o no (i.e., asignar el tratamiento a nivel escuela). [Hint: ¿Se pierde algo en términos de poder?]

- c) ¿Qué diseño preferiría, y por qué?
2. Suponga que se hizo el segundo diseño (i.e., se asignaron escuelas aleatoriamente). Antes de hacer cualquier análisis, ¿Ud. cree que comparar los resultados de los estudiantes en las escuelas de tratamiento con las de control unos años después del tratamiento nos va a indicar la efectividad de esta política pública? [Pista: Piense en las decisiones de largo plazo de quien se vuelve maestro]
 3. Suponga que efectivamente comparando los resultados de los estudiantes en diferentes escuelas puede estudiar el efecto del tratamiento. Suponga que Ud. tiene una base de datos que tiene resultados a nivel estudiante para diferentes escuelas: ¿Que regresión utilizaría para medir el efecto del tratamiento (indique el nivel de cada variable, como manejaría los errores estándar, si va a incluir controles y cuales, etc.)?
 4. Haga cálculos de poder para este experimento usando mínimos cuadrados ordinarios para analizar los datos. Es decir, simule diferentes efectos del tratamiento (incluido cero) y muestre con que probabilidad rechazaría la hipótesis nula. Para esto, utilice los datos de PLANEA para grado Sexto de 2017-2018 en matemáticas a nivel estudiantes, los cuales puede encontrar aquí https://www.inee.edu.mx/wp-content/uploads/2019/01/Planea06_2018_Alumnos.zip (más información aquí <https://www.inee.edu.mx/evaluaciones/planea/sexto-primaria-ciclo-2017-2018/>). Para ser más específico. Elija una muestra de 200 escuelas públicas, asigne las escuelas aleatoriamente a tratamiento y control, y después simule diferentes efectos de tratamiento.
 5. Repita los cálculos de poder, utilizando inferencia de aleatorización
 6. Un experimento muy similar a este se llevó a cabo en Indonesia (ver <https://academic.oup.com/qje/article-abstract/133/2/993/4622956?redirectedFrom=fulltext>). Los datos de ese experimento están en: <https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/MTVM50/HVY2EB&version=1.1>. Note que la variable que tiene los puntajes en la base de estudiantes se llama “*ITT_score*”. Adicionalmente, la muestra tiene pesos (“weight”) que se deben tener en cuenta a la hora de hacer la estimación. El experimento fue estratificado, y la variable “*triplet_id*” muestra los estratos. Puede usar el código de Stata (hacen la Tabla 5 en la línea 334 aproximadamente) para guiarse.
 - a) Estime el efecto del tratamiento (y el p-valor) en matemáticas usando mínimos cuadrados ordinarios.
 - b) Estime el efecto del tratamiento (y el p-valor) en matemáticas usando inferencia de aleatorización.

3. Problema 2 - Ivermectina y COVID

La Agencia Digital de Innovación Pública de CDMX (ADIP) sacó un documento de trabajo (<https://osf.io/preprints/socarxiv/r93g4/>) donde encuentra “a significant reduction in hospitalizations among patients who received the ivermectin-based medical kit; the range of the effect is 52%-76% depending on model specification”. Los códigos (<https://t.co/70QIKfSDeg?amp=1>) y los datos (<https://t.co/e0WPYZg2gd?amp=1>) son “públicos”. Pueden ver el hilo de Twitter original que escribió el director de la ADIP sobre el documento de trabajo (<https://twitter.com/PPmerino/status/1393263524139343873?s=20>). En este ejercicio vamos a explorar este estudio.

1. ¿Cuáles son los posibles problemas que puede haber cuando el evaluador de un programa, es el mismo que lo creó y lo implementó? Haga una pequeña reflexión sobre esto.
2. Antes de entrar en la econometría, revisemos los datos <https://t.co/e0WPYZg2gd?amp=1>. Haga un breve resumen sobre si estos datos cumplen con los mejores estándares de “replicación” de análisis que hemos visto en clase. Abajo unas preguntas guía.
 - a) ¿Hay algún diccionario de datos?

- b) ¿Es claro cuál es la fecha de la base de datos?
 - c) ¿Cuál es el universo de la base de datos? Es decir, ¿quién está en la base? ¿Todos las personas que se hicieron una prueba o todos los que dieron positivo?
 - d) ¿Se sabe algo de las comorbilidades o síntomas en sí?
 - e) ¿Se sabe cómo se juntaron los datos de hospitalización con los de la entrega de kits?
 - f) ¿Hasta cuantos meses después se considera que alguien fue hospitalizado?
 - g) ¿Se pueden considerar estos como los datos “crudos” o datos ya procesados?
3. El análisis sugiere que la entrega de un kit reduce la probabilidad de hospitalización. Haga un análisis sobre si la entrega de kits puede tener un sesgo de selección. Para esto revise si las características de quienes recibieron el kit (e.g., género, edad, comorbilidad, fecha, etc.) son iguales (teniendo en cuenta que el kit solo se les entregaba a las personas con síntomas graves). Que les dice este análisis sobre cualquier comparación entre quienes recibieron el kit y quiénes no.
 4. Haga análisis de regresión que estudien que tan robusto es que la entrega del kit reduzca las hospitalizaciones (i.e., pruebe con diferentes variables de control, diferentes submuestras, etc.). Haga un pequeño resumen de lo que encontró.
 5. El kit (por definición) no tenía solo ivermectina. ¿Se puede entonces afirmar que cualquier efecto del kit se debe a la ivermectina?
 6. ¿Que otros méritos o fallas cree usted que tiene la evaluación de la ADIP sobre el efecto de la ivermectina?
 7. Lean este documento con detenimiento: <https://gidmk.medium.com/is-ivermectin-for-covid-19-based-on-fraudulent-research-5cc079278602> y el meta-análisis más completo sobre Ivermectina y COVID (<https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD015017.pub2/full>). Haga un pequeño resumen sobre ambos artículos, y como sopesa usted la evidencia del artículo de la ADIP con la evidencia de otros países.