

Tarea 1

Evaluación de Impacto

Fecha de entrega: Lunes 21 de abril de 2025

1. Instrucciones

- Esta area debe ser entregada el día Lunes 21 de abril, en cualquier horario.
- La entrega se realiza via Canvas en la sección de Tareas.
- Debe entregar dos archivos
 1. El pdf con las respuestas, tablas y figuras y los análisis solicitados.
 2. El código que genera los análisis, independiente del lenguaje que ha utilizado.
- Si tienen dudas o preguntas, pueden escribir al profesor o al ayudante.

2. Ejercicios

Ejercicio 1. En este ejercicio analizaremos el impacto que tiene un programa que subsidia el seguro de salud a población vulnerable en Tanzania. Uno de los objetivos principales del programa es reducir la carga de los gastos asociados con la salud en los hogares de bajos ingresos. El objetivo de este ejercicio será evaluar el impacto que tiene el programa sobre los gastos en salud.

Cuando se realizó el proceso de selección de comunidades que podrían recibir el subsidio para el seguro de salud, hubo un mayor número de hogares elegibles de los que se podía cubrir con el presupuesto disponible. Debido a esta situación, las autoridades decidieron realizar una lotería para seleccionar a los hogares que participarían del programa, por lo que todas los hogares tenían una oportunidad justa de ser seleccionadas. Para llevar a cabo el análisis se utilizará la base de datos titulada *data_tarea1_aleatorizacion.csv*. Abajo incluyo una descripción de las variables que se incluyen en la misma. Note que a menos que se diga lo contrario las variables están definidas al inicio o en la línea base. Presten atención a tres variables: *elegible*, *treatment* y *enrolled*. La variable *eligible* es igual a 1 si el hogar tiene las características necesarias para participar del programa. La variable *treatment* es igual a uno para aquellos que fueron favorecidos en la lotería (se les ofreció el programa), cero en cualquier otro caso. Por definición, todos los hogares no elegibles, no reciben ofrecimiento.

La variable *enrolled* captura participación efectiva: es igual a uno si el hogar efectivamente participó del programa, cero en cualquier otro caso.

1. Usando los hogares elegibles, compare el gasto en salud y los valores de las demás variables presentes¹ en la base de datos en línea base entre los hogares *seleccionados* para recibir el tratamiento y aquellos seleccionados como controles. ¿Es la muestra balanceada? ¿Es esto lo que esperaría o no?
2. Como ya le advirtiera en el enunciado, la lotería no se respetó perfectamente: algunos de los controles recibieron el programa (variable *enroled*). Describa qué tan grande es el *incumplimiento* en este experimento. Compare formalmente qué tan distintos son los hogares que se matuvieron en el grupo de control, versus los que terminaron recibiendo tratamiento.
3. Para el período post tratamiento, compare el gasto promedio en salud para los hogares a los que se les ofreció tratamiento y los que no. ¿Es este un buen estimador del efecto del programa? ¿por qué si o por qué no?
4. Ahora utilice una regresión por Mínimos Cuadrados Ordinarios para estimar el efecto del ofrecimiento del programa en los gastos en salud explotando la asignación experimental y concluya: ¿cuál fue el impacto del programa?. Recuerde de estimar el modelo usando sólo los hogares elegibles y evalúe distintas especificaciones que abajo se mencionan. Discuta sus resultados.
 - a) Sin controles.
 - b) Incluyendo la variable gasto rezagada.
 - c) Incluyendo la variable gasto rezagada y las edades del jefe de hogar y su pareja.
 - d) Incluyendo otros controles que a su juicio pudieran mejorar su estimación. Justifique su inclusión.

¹Por ejemplo, puede usar las siguientes variables: *health_expenditures_{base}*, *age_hh*, *age_sp*, *educ_sp*, *hospital_base*, *poverty_index*, *female_hh*, *indigenous*, *hhsiz*, *dirtfloor*, *bathroom*, *land*, y *hospital_distance*.

Cuadro 1: Descripción de Variables

Variable	Tipo	Descripción
locality_identifier	Numérica	Identificador de la localidad
household_identifier	Numérica	Identificador único del hogar
expenditures_out_base	Numérica	Gasto anual per capita en salud en la línea de base (periodo 0)
age_hh	Numérica	Edad del jefe de hogar
age_sp	Numérica	Edad de la pareja del jefe de hogar
educ_hh	Numérica	Años de educación del jefe de hogar
educ_sp	Numérica	Años de educación de la pareja del jefe de hogar
hospital_base	Numérica	Indicador relacionado a hospital en el periodo 0
expenditures_out_end	Numérica	Gasto anual per capita en salud en el periodo posterior al tratamiento (periodo 1)
hospital_end	Numérica	Indicador relacionado a hospital en el periodo 1
eligible	Binaria (0/1)	Hogar elegible para participar en el programa (0=no, 1=sí)
enrolled	Binaria (0/1)	Hogar inscrito en el programa (0=no, 1=sí)
treatment	Binaria (0/1)	Hogar que recibió la invitación a participar (0=no, 1=sí)
poverty_index	Numérica (1-100)	Índice de pobreza del hogar (1=menos pobre, 100=más pobre)
female_hh	Binaria (0/1)	El jefe de hogar es mujer (0=no, 1=sí)
indigenous	Binaria (0/1)	El jefe de hogar habla lengua indígena (0=no, 1=sí)
hhsiz	Numérica	Tamaño del hogar (número de miembros)
dirtfloor	Binaria (0/1)	Vivienda con piso de tierra al inicio (0=no, 1=sí)
bathroom	Binaria (0/1)	Vivienda con baño privado al inicio (0=no, 1=sí)
land	Numérica	Hectáreas de tierra que posee el hogar al inicio
hospital_distance	Numérica	Distancia al hospital más cercano

Ejercicio 2. Un organismo público desea evaluar el impacto de un programa de capacitación laboral orientado a jóvenes desempleados. El resultado principal de interés es el salario mensual de los participantes, medido 12 meses después del inicio del programa. La variable salario es continua. Para ello, se propone realizar un experimento aleatorizado en el que un grupo de jóvenes desempleados será asignado al tratamiento (acceso al programa de capacitación) y otro al grupo de control (sin acceso). Se desea calcular el **tamaño mínimo de muestra necesario** para detectar un efecto esperado de aumento salarial mensual de \$60.000 CLP, con una desviación estándar del salario de \$200.000 CLP, poder estadístico de 80 % y nivel de significancia de 5 %. A partir de este contexto, se pide calcular el tamaño de muestra bajo distintos escenarios ²:

- a) Aleatorización individual sin incumplimiento ni atrición. ¿Cuál sería el número total de personas en la muestra si asumimos tamaños iguales para el grupo tratamiento y grupo de control? Escriba la fórmula que usó para realizar este cálculo.
- b) Suponga ahora que sólo el 80 % de los asignados al tratamiento efectivamente participa del programa ($c = 0,8$). Calcule el tamaño de la muestra necesario para mantener igual mínimo efecto detectable. Si mantuvieramos el tamaño de la muestra en lo que encontró en (a), qué pasa con el mínimo efecto detectable? Escriba la fórmula que usó para realizar este cálculo.
- c) Suponga ahora que sigue con cumplimiento perfecto, pero ahora un 15 % de los participantes (tanto tratamiento como control) no puede ser encuestado al final del estudio. ¿Cuál es el tamaño de la muestra inicial antes de la atrición? Escriba la fórmula que usó para realizar estos cálculos.
- d) Volviendo a la situación en (a), suponga ahora que la intervención se implementará en centros de empleo, y que la unidad de aleatorización serán los centros y no las personas. Cada centro tiene en promedio 20 participantes, y se espera un coeficiente de correlación intracluster (ICC) de 0,15. ¿Cómo afecta esto al tamaño de muestra total requerido? ¿Cuántos clústeres se necesitarían por grupo para alcanzar el mismo poder estadístico y el mismo mínimo efecto detectable? (tenga en cuenta que sólo puede reclutar 20 personas por centro). Escriba la fórmula que usó para realizar estos cálculos.

²Puede usar como guía el sitio Power Calculator de JPAL que no sólo hace cálculos interactivos, sino también le muestra el código de R que necesita ejecutar para realizar dicho cálculo. Puede también programar la fórmula usted mismo en R o incluso en Excel. Las fórmulas las vimos en clase; tal vez necesite despejar N o MDE, dependiendo del ejercicio. Se le pide mostrar su trabajo, pero la calculadora de Jpal le permite verificar su resultado.