

Taller 2 (EVIMP2020II)

Juan Sebastián Cañas

Universidad Nacional de Colombia

jscanass@unal.edu.co

September 23, 2020

1 Descripción de RCT en Safaera

- RCT en Safaera
- Lotería en Safaera

2 Análisis de Resultados del RCT en Safaera

- Impacto causal
- Regresión
- SUTVA
- Impacto en resultados educativos
- Comentarios finales

Objetivo

Determinar la efectividad de un programa de eliminación de bacterias y parásitos, enfocado en niños y niñas de edad de 6-10 en los municipios con alta tasa de pobreza.

Diseño Experimental

RCT. Asignación a los hogares a tratamiento o control, a través de una lotería. El tratamiento implicaba una semana de toma de píldoras, con efectos colaterales leves.

Población de estudio

200 hogares con niños en escuelas de Safaera.

Línea base

- Tamaño del hogar
- Ingreso mensual
- Dónde viven (rural/urbano)
- Mujer cabeza de hogar
- Escuela

Datos recolectados 6 meses después

- Índice de salud: 0 - 10 (10 mejor estado de salud)
- Resultados escolares: promedio sencillo de 0 a 5 (5 mejor nota)

Los resultados, el análisis de estos y sus posibles implicaciones dependen del proceso de **aleatorización** del RCT. Para ver esto exploramos los datos de línea base buscando homogeneidad entre tratados y no tratados:

Variable	Tratados	No tratados
Rural	.53	.51
Mujer cabeza de familia	.51	.56
Tamaño hogar	6.3 (1.6)	6.7 (1.5)
Ingreso	677 (200)	631 (166)

Table: Medias para variables de línea base del RCT en Safaera para la población tratada y no tratada en el estudio. Entre paréntesis está la desviación estándar. Para rural y mujer cabeza de familia se reporta la fracción de la población que tiene este atributo.

Impacto Causal

Para medir el impacto causal del tratamiento, calculamos la diferencia entre los promedios del índice de salud entre tratados y no tratados en la población de estudio. Es decir,

$$SDO = \frac{1}{N_T} \sum_i^n (y_i | d_i = 1) - \frac{1}{N_C} \sum_i^n (y_i | d_i = 0) = 6.826 - 1.951 = 4.875$$

El proceso de aleatorización del RCT nos quita el problema del sesgo de selección y nos asegura que $SDO = ATE$, es decir, las diferencia de medias nos da el efecto promedio del tratamiento [Cunningham, 2018]. Entonces el medicamento tiene un impacto causal positivo en la salud de la población de estudio, en una escala del 0 a 10, de 4.9 para los niños tratados respecto a los no tratados.

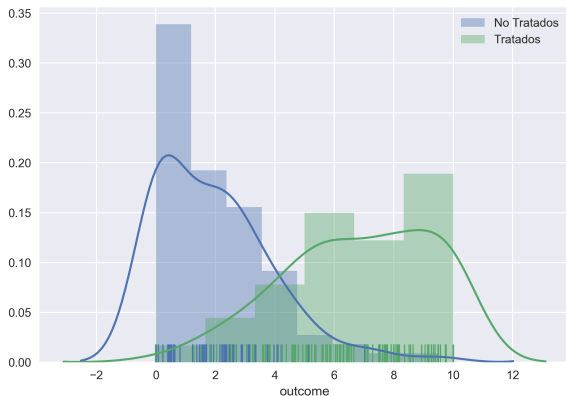


Figure: Distribución de probabilidad de los outcomes del índice de salud en la población de estudio tratada y no tratada en el RCT de Safaera

Resultados de Regresión

Variable explicativa	(1)	(2)	(3)
tratamiento	4.769 (.308)	4.779 (.307)	4.780 (.313)
mujer cabeza de familia	-1.358 (.307)	-1.389 (.307)	-1.389 (.367)
rural	-	-.461 (.306)	-.461 (.315)
ingreso hogar	-	-	-3.322e-07 (0.001) (p = 1)
tamaño hogar	-	-	0.0007 (0.097) (p = 0.994)
R ²	.580	.584	.584

Table: Resultados de la regresión que estima el efecto del medicamento en la salud de 200 niños y niñas en un municipio con alta tasa de pobreza

A pesar que fue un RCT podemos incluir variables de control que pueden tener un efecto explicativo y mejorar las estimaciones de los efectos causales (disminuir error y desviaciones) [Angrist, 2008]. La primera variable que se incluye es la de mujer cabeza de hogar ya que es la que más está desbalanceada entre los grupos y puede afectar el efecto de aleatorización. Después agregamos la variable rural teniendo en cuenta que la asignación aleatoria se hizo para el medicamento pero no para las escuelas. Las condiciones de los estudiantes atendiendo colegios rurales pueden variar respecto a los colegios urbanos y estas variaciones pueden afectar la salud. Por ejemplo, el acceso al servicio de agua potable o el acceso a servicios de salud. Siguiendo con el argumento anterior, las variables de ingreso y tamaño de hogar se incluyeron teniendo en cuenta que pueden tener un poder explicativo. Por ejemplo, el ingreso es una medición directa de la pobreza monetaria y el tamaño del hogar podría medir condiciones de hogar como hacinamiento, variables que influyen la salud [Marmot, 2005].

Como notamos en la tabla con los resultados de regresión, las últimas dos variables no tienen ningún poder explicativo en nuestro caso. Lo anterior se puede explicar por el balance entre grupos producido por el efecto de aleatorización del RCT. Se incluyen porque una hipótesis que se tuvo fue la de su poder explicativo en los resultados. Finalmente nuestra regresión quedaría con las variables de mujer cabeza de hogar y rural.

$$\text{IndiceSalud} = \beta \text{Tratamiento} + \text{MujerCabezaHogar} + \text{Rural}$$

Con esto tenemos que $\beta = 4.779(.307)$, el cual coincide en su rango de desviación estándar con el valor calculado de $ATE = 4.875$.

SUTVA(Stable Unit Treatment Value Assumption)

Son los **supuestos** que se hacen para poder calcular los impactos causales bajo el modelo de Rubin. La idea general es que el **efecto del tratamiento para los individuos se mantiene fijo sobre toda la población** durante el proceso de generación de los datos o el experimento. Se puede dividir en los siguientes casos: tratamientos homogéneos; efecto de red; efecto de equilibrio general [Cunningham, 2018].

SUTVA en RCT de Safaera

En la siguiente tabla se observa una posible violación de SUTVA: para las escuelas 1, 2, 3 y 4 hay mezcla en la misma escuela entre niños tratados y no tratados. Como consecuencia, se pueden presentar efectos de red tales como niños no tratados tienen menor probabilidad de enfermarse si están con niños tratados.

Escuela	1*	2*	3*	4*	5	6	7	8	9	10
Sin tratamiento	7	8	9	10	0	22	0	24	0	12
Tratamiento	4	15	15	5	27	0	22	0	20	0

Table: Conteo de tratados y no tratados según la Escuela donde pertenece en la población de estudio del RCT de Safaera.

Para las escuelas 5,6,7,8,9 y 10 se tienen solamente poblaciones tratadas o no tratadas en cada escuela. En este caso no existe el efecto de vacuna que podría generar la mezcla en una misma escuela de tratados y no tratados. Tomando esta submuestra calculamos el efecto causal y obtenemos que $SDO_{SUTVA} = 4.9$, que coincide con el SDO para toda la población.

Impactos en resultados educativos

Ahora revisamos un efecto adicional del tratamiento, este es, el efecto en los resultados educativos para los niños y niñas tratados. Tenemos que efectos positivos $ATE_{SUTVA} = .78$ Y $ATE = 0.82$. Esto se puede explicar teniendo en cuenta que una salud deteriorada puede tener consecuencias en el rezago estudiantil al aumentar la inasistencia escolar e incluso la deserción. También se podría explicar teniendo en cuenta la indisposición mental para aprender que produce los efectos negativos de la salud.

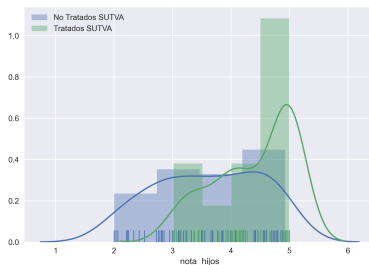


Figure: Distribución de probabilidad de los outcomes de las notas promedio los niños tratados y no tratados que cumplen SUTVA en el RCT de Safaera

Fortalezas

- Estudio aleatorizado
- Se tienen medidas de pobreza monetaria
- Se observan efectos positivos en la salud y efectos positivos adicionales en los resultados educativos

Debilidades

- En la línea base no se calcula el índice de salud. Se puede tener el caso que la población tratada es, en promedio, más saludable que la no tratada.
- Para poder generalizar este estudio en el país de Safaera tendríamos que saber la ubicación en la línea de pobreza municipal de Safaera y si el contexto de nuestra población de estudio representa bien a la población de los municipios pobres del país de Safaera. También necesitamos saber el efecto por edades para medir el efecto del tratamiento según la edad.

En la última diapositiva, escriba la información adicional que necesitaría para poder calcular si ese impacto justifica el costo del programa

- Determinar que tan *leves* son los efectos colaterales
- Condiciones de salud de los niños y niñas antes del tratamiento y edad
- Si el contexto de nuestra población de estudio representa el conjunto de hogares con pobreza en Safaera y el país de Safaera

Todos los resultados de este análisis están disponibles para replicar en el siguiente enlace https://github.com/jscanass/EVIMP2020/blob/master/talleres/EVIMP2020_taller2_RCT_Safaera.ipynb



Cunningham, Scott (2018)

Causal Inference: The Mixtape

https://scunning.com/cunningham_mixtape.pdf



Angrist, Joshua D and Pischke, Jörn-Steffen (2008)

Mostly harmless econometrics: An empiricist's companion

Princeton university press



Marmot, Michael (2005)

Social determinants of health inequalities

The Lancet, volumne 365, number 9464, year 2005