

0.0.1. ¿Qué es una Evaluación de Impacto?

Los programas y políticas de desarrollo suelen estar diseñados para cambiar resultados, como aumentar los ingresos, mejorar el aprendizaje o reducir las enfermedades [2]. Saber si estos cambios se logran o no es una pregunta crucial para las políticas públicas [2], y aquí es donde entra en juego lo que se conoce como **evaluación de impacto**.

Una evaluación de impacto mide los cambios en el bienestar de los individuos que se pueden atribuir a un proyecto, un programa o una política específicos [2]. Su sello distintivo radica en que pueden proporcionar **evidencia** robusta y creíble sobre si un programa concreto ha alcanzado o está alcanzando sus resultados deseados [2].

Este tipo de evaluaciones pone un fuerte énfasis en los resultados y busca responder una pregunta específica de causa y efecto: ¿cuál es el impacto (o efecto causal) de un programa en un resultado de interés? [2]. Es decir, se focalizan en los cambios directamente atribuibles al tratamiento [2] sobre un conjunto de **variables de resultado o de interés** en un conjunto de individuos [1]. Estas variables son aquellas sobre las cuales se espera que el programa tenga un efecto en los beneficiarios [1]. Por ejemplo, podrían ser la estatura y peso de los individuos, o la cantidad de empleados o de ventas en una empresa.

Por lo tanto, el objetivo final de la evaluación de impacto consiste en establecer la diferencia entre la variable de resultado del individuo participante en el programa en presencia del programa, y la variable de resultado de ese individuo en ausencia del programa [1], lo que se conoce como **efecto del tratamiento**. Sin embargo, es evidente que la respuesta a “¿qué habría pasado con los beneficiarios en ausencia del programa?” se refiere a una situación que no es observable. Este resultado hipotético se denomina **contrafactual** y es lo que se debe estimar en cualquier evaluación de impacto.

Algunas de las principales razones por las que se debería promover el uso de estas evaluaciones como herramientas de gestión provienen del hecho que permiten mejorar la rendición de cuentas, la inversión de recursos públicos o la efectividad de una política, obtener financiamiento, como así también probar modalidades de programas alternativos o innovaciones de diseño [2] y revelar la realidad de muchas políticas públicas para de esta forma contribuir a la fiscalización mediática [1].

Un ejemplo claro de por qué son necesarias las evaluaciones de impacto es el que describe Howard White con respecto al Programa Integrado de Nutrición en Bangladesh (PINB) [4]. Este programa identificaba, mediante mediciones en campo, a los niños desnutridos y los asignaba a un tratamiento que incluía alimentación suplementaria a los menores y educación nutricional a las madres [1]. Inicialmente, el programa fue considerado como un éxito ya que los datos de monitoreo mostraban caídas importantes en los niveles de desnutrición. El Banco Mundial decidió, con base en esta evidencia y previo a cualquier tipo de evaluación, aumentar los recursos destinados al programa. Sin embargo, las primeras evaluaciones de impacto, realizadas por el Grupo Independiente de Evaluación del mismo Banco Mundial y por la ONG inglesa *Save the Children*, mostraron que la mejoría de los indicadores de los beneficiarios era similar o inferior a la de otros niños

con características comparables que no hacían parte del programa [1]. Estos resultados reflejaron que las percepciones de los administradores del programa y de las entidades financiadoras eran erradas, y sugirieron algunos correctivos al programa [1].

0.0.2. Estimación del Tratamiento

El marco teórico estándar para formalizar el problema de la evaluación de impacto se basa en el modelo de resultado *potencial* o modelo de Roy-Rubin [3].

Formalmente, definimos dos elementos para cada individuo $i = 1, \dots, N$, donde N denota la población total:

- Por un lado, el indicador de tratamiento D_i , tal que $D_i = 1$ implica que el individuo i participó del tratamiento, y $D_i = 0$ en caso contrario.
- Por otro lado, las variables de resultado las definimos como $Y_i(D_i) = Y_i|D_i$ - se lee como “el valor de Y_i dado D_i ”. De esta forma, $Y_i(1)$ es la variable de resultado si el individuo i es tratado, e $Y_i(0)$ es la variable de resultado si el individuo i no es tratado.

Con esto, el **efecto del tratamiento** para un individuo i se puede escribir como:

$$\tau_i = Y_i(1) - Y_i(0) = (Y_i|D_i = 1) - (Y_i|D_i = 0) \quad (1)$$

Según esta fórmula, el impacto causal (τ) de un programa (D) en una variable resultado (Y) para un individuo i es la diferencia entre la variable con el programa ($Y_i(1)$) y la misma variable sin el programa ($Y_i(0)$).

De nuevo, el problema fundamental de la evaluación de impacto es que se intenta medir una variable en un mismo momento del tiempo para la misma unidad de observación pero en dos realidades diferentes [2]. Sin embargo, claramente solo se da uno de los dos resultados potenciales $Y_i(1)$ o $Y_i(0)$, pero no ambos. Es decir, en los datos queda solamente registrado $Y_i(1)$ si $D_i = 1$ y $Y_i(0)$ si $D_i = 0$; no se dispone de $Y_i(1)$ si el individuo no fue tratado ($D_i = 0$), ni tampoco de $Y_i(0)$ si el individuo fue tratado ($D_i = 1$). De esta manera, el **resultado observado** de Y_i se puede expresar como:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) = \begin{cases} Y_i(1) & \text{si } D_i = 1 \\ Y_i(0) & \text{si } D_i = 0 \end{cases} \quad (2)$$

Si nos concentramos en las unidades tratadas, en la ecuación 1, el término $Y_i(0) = (Y_i|D_i = 0)$ representa el contrafactual, es decir *cuál habría sido el resultado si la unidad no hubiera participado en el programa*. Al ser imposible observar directamente el contrafactual, es necesario **estimarlo**. La forma más directa de solucionar este problema sería hallando un “clon perfecto” para cada uno de los individuos participantes del programa, lo

cual resulta bastante difícil. Por lo tanto, el primer paso para lograr esta estimación consiste en **desplazarse desde el nivel individual al nivel del grupo** [2], concentrando el análisis en el impacto o efecto *promedio* (y no individual).

En primera instancia, se puede estimar el *impacto promedio del programa sobre la población* (o *ATE*):

$$ATE = \mathbb{E}[Y_i(1) - Y_i(0)] \quad (3)$$

El *ATE* se interpreta como el cambio promedio en la variable de resultado cuando un individuo escogido al azar pasa aleatoriamente de ser participante a ser no participante [1]. Esta medida es relevante en el caso de la evaluación de un programa universal.

Sin embargo, en la mayoría de los casos, el tratamiento solo está disponible para un subconjunto de la población, por lo cual resulta más preciso utilizar un estimador que únicamente promedie el efecto sobre la población elegible [1]. De esta forma, definimos el *impacto promedio del programa sobre los tratados* (o *ATT*), que es en general el parámetro de mayor interés en una evaluación de impacto, y representa el efecto esperado del tratamiento en el subconjunto de individuos que fueron efectivamente tratados:

$$ATT = \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 1] \quad (4)$$

Es decir, el *ATT* es la diferencia entre la media de la variable de resultado en el grupo de los participantes y la media que hubieran obtenido los participantes si el programa no hubiera existido [1].

0.0.3. El Grupo de Control como Estimador del Contrafactual

En la ecuación 4, el término $\mathbb{E}[Y_i(1)|D_i = 1]$ es un resultado observable mientras que $\mathbb{E}[Y_i(0)|D_i = 1]$ es el promedio contrafactual que debemos aproximar. Para ello, se construye lo que se conoce como **grupo de control**, formado por individuos que no participan del programa pero que, idealmente, son estadísticamente similares [2] al **grupo de tratamiento**, compuesto por quienes sí recibieron el programa.

Por lo tanto, en la práctica, el reto de una evaluación de impacto es definir un *buen* grupo de control, es decir uno que sea estadísticamente idéntico al de tratamiento, en promedio, en ausencia del programa [2]. Si se lograra que los dos grupos sean idénticos, con la única excepción que un grupo participa del programa y el otro no, entonces sería posible estar seguros que cualquier diferencia en los resultados tiene que deberse al programa [2].

Para que un grupo de comparación sea válido, debe satisfacer lo siguiente [2]:

- Las características *promedio* del grupo de tratamiento y del grupo de comparación deben ser idénticas en ausencia del programa. Cabe resaltar que no es necesario que las unidades individuales en el grupo de tratamiento tengan clones perfectos en el grupo de control.
- No debe ser afectado por el tratamiento de forma directa ni indirecta.

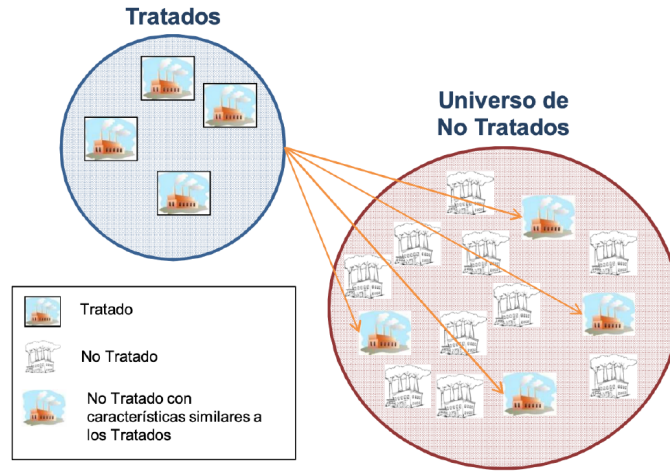


Figura 1:

- El grupo de comparación debería reaccionar de la misma manera que el grupo de tratamiento si fuera objeto del programa.

Cuando el grupo de comparación no produce una estimación precisa del contrafactual, no se puede establecer el verdadero impacto del programa. A continuación, se presentan dos situaciones en las que el grupo de comparación seleccionado conduce a una estimación incorrecta del contrafactual.

Comparaciones antes-después

Este tipo de comparaciones intenta establecer el impacto de un programa a partir de un seguimiento de los cambios en los resultados en los participantes del programa a lo largo del tiempo. Consideran el contrafactual como el resultado para el grupo de tratamiento antes que comience la intervención, momento también conocido como **línea de base**. Esta comparación supone que si el programa no hubiera existido, el resultado para los participantes del programa habría sido igual a su situación antes del programa, lo cual en la mayoría de los tratamientos este supuesto no puede sostenerse [2].

Comparaciones de inscritos y no inscritos: Sesgo de autoselección

Una de las formas más directas de solucionar el problema del contrafactual podría ser simplemente utilizar el promedio de la variable de resultado entre los individuos no participantes del programa pero elegibles para participar. Es decir, podríamos utilizar a $\mathbb{E}[Y_i(0)|D_i = 0]$, que sí se puede observar, como una aproximación de $\mathbb{E}[Y_i(0)|D_i = 1]$ [1]. Sin embargo, esto podría generar estimaciones inexactas del efecto del programa, dado

que los participantes y los no participantes generalmente son diferentes (en características observadas y no observadas), aún en ausencia del programa, y es justamente por tal motivo que unos *eligen* participar y otros no [1]. Este problema se conoce como **sesgo de autoselección**, ya que el grupo se *autoselecciona* para no participar de un programa. Más concretamente, este sesgo se produce cuando los motivos por los que un individuo participa en un programa, usualmente no observables y difíciles de medir, están correlacionados con los resultados [2], y por ende, es muy probable que la variable de resultado del grupo de tratamiento y del grupo de control sean diferentes, *aún si el programa no existiera* [1].

A modo de ejemplo, se puede pensar en un programa de nutrición infantil. Podría ocurrir que las madres de familias participantes del programa sean más proactivas respecto al desarrollo de sus hijos, por lo cual se preocuparon en lograr la participación en el programa. El problema de autoselección en este caso radica en que la motivación de las madres (que no se observa y es difícil de medir) afecta no solo la probabilidad de participar en el programa, sino también el estado nutricional de los niños ya que estas podrían vigilar mejor la dieta de sus hijos. De esta forma, la diferencia observada en el estado nutricional de los niños de los dos grupos podría deberse parcialmente a la diferencia en el nivel de compromiso de las madres, y no exclusivamente a que un grupo participa en el programa y el otro no [1].

Habiendo presentado este problema, volvamos a la fórmula del *ATT* (4), y notemos que la podemos reescribir de la siguiente manera:

$$ATT + \mathbb{E}[Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(1)|D_i = 1] \quad (5)$$

Si ahora restamos $\mathbb{E}[Y_i(0)|D_i = 0]$ a ambos lados, obtenemos:

$$ATT + \mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] = \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] \quad (6)$$

De aquí se deduce que utilizar $\mathbb{E}[Y_i(0)|D_i = 0]$ como aproximación del contrafactual $\mathbb{E}[Y_i(0)|D_i = 1]$ permite recuperar el *ATT* si y solo si:

$$\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] = 0 \quad (7)$$

Y es justamente este el gran desafío de la evaluación de impacto: determinar las condiciones o supuestos bajo los cuales se cumple 7, es decir que $\mathbb{E}[Y_i(0)|D_i = 0]$ sea una aproximación válida de $\mathbb{E}[Y_i(0)|D_i = 1]$ [1].

Es importante en este punto identificar claramente los diferentes resultados que hemos mencionado hasta el momento, los cuales se encuentran en la tabla 1. Teniendo en cuenta las fórmulas expuestas en esta tabla, podemos enunciar de otra forma lo que dijimos anteriormente: la calidad de una evaluación de impacto dependerá de los supuestos necesarios para asegurar que no hay sesgo de selección, es decir que $\mathbb{E}[C_i(0) - Y_i(0)] = 0$.

Lo que se desea medir: el <i>ATT</i> .	$\mathbb{E}[Y_i(1) - Y_i(0) D_i = 1]$
Lo que se observa: el promedio de la diferencia entre los resultados de los tratados y el grupo de control.	$\mathbb{E}[Y_i(1) - C_i(0)]$
El sesgo de selección: el promedio de la diferencia potencial entre lo que se observa y lo que se desea medir.	$\mathbb{E}[(Y_i(1) - Y_i(0)) - (Y_i(1) - C_i(0))] = \mathbb{E}[C_i(0) - Y_i(0)]$

Cuadro 1: Distinción entre las diferentes comparaciones presentes en una evaluación de impacto.

Las reglas de un programa para seleccionar a los participantes, también llamadas **reglas de asignación**, constituyen el parámetro clave para determinar el método de evaluación de impacto. A continuación, presentamos las distintas reglas y los métodos de evaluación que se utilizan en la práctica al trabajar con cada una de ellas.

0.0.4. Evaluaciones Experimentales o Aleatorias

En este tipo de evaluaciones, los beneficiarios del programa en cuestión son seleccionados al azar, es decir mediante un sorteo. La asignación aleatoria se considera la regla de oro de la evaluación de impacto ya que no solo proporciona a los administradores del programa una regla imparcial y transparente para asignar recursos escasos entre poblaciones igualmente merecedoras de ellos, sino que también representa el método más sólido para evaluar el impacto de un programa [2].

La asignación aleatoria trae dos consecuencias importantes: las unidades ya no pueden *elegir* si participar o no, y además todas tienen la misma probabilidad de ser seleccionadas para el programa. De esta forma, el resultado potencial de cada individuo es independiente de sus condiciones previas, ya que la asignación al tratamiento no está determinada por sus características, sino por el azar.

Como explicamos anteriormente, el grupo de comparación ideal sería lo más similar posible al grupo de tratamiento en todos los sentidos, excepto con respecto a su participación en el programa que se evalúa. Ahora bien, el proceso de asignación aleatoria producirá dos grupos que tienen una alta probabilidad de ser estadísticamente idénticos en todas sus características (observables y no observables), siempre que el número de unidades sea suficientemente grande [2]. Esto es así ya que en general, si la población de unidades elegibles es lo suficientemente grande, este mecanismo asegura que cualquier rasgo de la población se transfiera a ambos grupos. Dicho esto, lo que ocurre en estos casos es que todos aquellos individuos que no resultaron ser beneficiarios del programa conforman un grupo de control que resulta ser naturalmente un excelente estimador del

contrafactual $\mathbb{E}[Y_i(0)|D_i = 1]$. Es decir, mediante este tipo de asignación, se asegura que:

$$\mathbb{E}[Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(0)|D_i = 0] \quad (8)$$

o, en otros términos:

$$\mathbb{E}[C_i(0) - Y_i(0)] = 0 \quad (9)$$

es decir, que no hay sesgo de selección presente.

La asignación aleatoria puede utilizarse como regla de asignación de un programa en dos escenarios específicos: cuando la población elegible es mayor que el número de plazas disponibles del programa, o cuando sea necesario ampliar un programa de manera progresiva hasta que cubra a toda la población elegible [2]. Además, la asignación aleatoria podría hacerse a nivel individual (por ejemplo, a nivel de hogares), o a nivel de conglomerado (por ejemplo, comunidades) [1].

Una vez que se haya asignado el tratamiento de manera aleatoria, gracias a 8, es bastante sencillo estimar el impacto del programa. Después de que el programa ha funcionado durante un tiempo, tendrán que medirse los resultados de las unidades de tratamiento y de comparación. El impacto del programa es sencillamente la diferencia entre el resultado promedio para el grupo de tratamiento y el resultado promedio para el grupo de comparación, es decir:

$$\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$

0.0.5. Evaluaciones Cuasiexperimentales

Diferencias en Diferencias

Propensity Score Matching