

Curso: Técnicas estadísticas de evaluación de impacto socioeconómico de políticas públicas y programas sociales

Humberto Soto

Oficial de Asuntos Ambientales

Comisión Económica para América Latina y el Caribe

Oficina subregional en México

Email: humberto.soto@cepal.org

Julio de 2010

1 Introducción. Diseño y evaluación

Los métodos estadísticos han resultado de gran utilidad desde mitad del siglo pasado en el diseño y evaluación de las políticas públicas y privadas orientadas a impulsar el desarrollo económico y social de los países del mundo, los cuales después de la segunda guerra mundial se encontraron con la necesidad de planificar un proceso integral de desarrollo (ya sea por rezagos históricos o por tareas de reconstrucción).

Las políticas antes mencionadas involucraron, y siguen involucrando muy distintos aspectos del desarrollo, en lo económico por ejemplo, al buscar el mayor aprovechamiento de los recursos humanos disponibles, se crean programas de capacitación especializada, o de educación a todos los niveles. En lo social se implementan programas de transferencias monetarias en la búsqueda de redistribuir los recursos disponibles para reducir la desigualdad social, para combatir la pobreza y para igualar las oportunidades de desarrollo de los individuos.

Muchos otros ejemplos pueden mencionarse, programas implementados para mejorar la salud de la población, para incrementar la capacidad productiva de las empresas, para generar incentivos a favor del consumo de cierto tipo de bienes o servicios deseables (alimentación, educación, etc.), para regular, controlar o desincentivar la adquisición de otro tipo de bienes no deseables (contaminantes, drogas, etc.). La lista puede extenderse considerablemente, sin embargo, en todos los casos hay dos elementos en común que son de particular interés de estudio para el área estadística: i) la selección de los beneficiarios, y ii) la medición del impacto.

El primer elemento se refiere a la parte del diseño de los programas socioeconómicos que establece su población objetivo. En este punto vale la pena mencionar la vinculación con la ciencia económica que establece en su definición la necesidad de administrar recursos escasos en presencia de necesidades ilimitadas. Esto conduce a la necesidad de optimizar el uso de los recursos, lo cual en el contexto de la implementación de un programa socioeconómico se traduce en lograr llevar atinadamente los recursos disponibles a la población objetivo.

El segundo elemento establece la necesidad de determinar si, bajo el supuesto de que se ha logrado llegar a la población objetivo, las acciones realizadas por el programa, que en la literatura especializada se han denominado tratamiento, han logrado su cometido en el plazo de tiempo especificado en la etapa de diseño.

Si bien pareciera que ambos elementos, la selección de beneficiarios y la medición del impacto, son temas tratables por separado, se verá más adelante que tratarlos conjuntamente puede facilitar la adecuada consecución del segundo. Por ello es que, si bien el presente curso se enfoca especialmente en el tema de evaluación de impacto, toca tangencialmente el tema de selección de los beneficiarios.

2 ¿Y la estadística cuando?

Las herramientas estadísticas están tan presentes en el análisis de los impactos de política económica a partir de datos, que existe todo un conjunto de métodos enfocados a ese tipo de análisis catalogados como métodos econométricos, (que van más allá de los análisis de series temporales). En muchos casos existe una sutil línea que diferencia los métodos econométricos, por ello no se ahondará más al respecto, sin embargo vale la pena resaltar este tipo de nomenclatura, muy presente en la literatura especializada de evaluación de impactos socioeconómicos de programas.

Respondiendo a la pregunta que titula esta sección, los métodos estadísticos son ampliamente utilizados en la determinación de la población objetivo. Desde las herramientas sencillas de análisis exploratorio utilizadas para caracterizar a los individuos, hasta herramientas sofisticadas de tipo geoestadístico que permiten localizar territorialmente a aquellos individuos con cierto tipo de características, pasando por las técnicas de agrupación y los métodos de clasificación, estas técnicas son ampliamente utilizadas para determinar a los individuos o grupos poblacionales que se espera sean impactados por el programa que se pretende implementar.

El análisis de regresión logística, por ejemplo, es utilizado para modelar la probabilidad de selección de un individuo de acuerdo con sus características.

En algunas ocasiones los tratamientos planificados pueden ser graduales, siendo mayores en cierto grupo poblacional, medios en otro y bajos o nulos en un tercer grupo. En estos casos la regresión logística multinomial suele ser utilizada.

Otra alternativa comunmente utilizada para determinar el conjunto de individuos a impactar por un programa es el análisis discriminante que permite clasificar a un conjunto amplio de la población a partir de la información (en ocasiones difícil o costosa de obtener, como pueden ser los ingresos) de un conjunto reducido pero representativo de individuos. Las técnicas de reducción de dimensionalidad también son frecuentemente utilizadas, cuando lo que se busca es resumir un conjunto de atributos en un valor continuo a partir del cual, por medio de un punto de corte se puede segmentar a la población, identificando así al grupo a beneficiar.

Sin ser un listado exhaustivo, la mención de las metodologías previamente enunciadas otorga una visión aproximada del alcance de los métodos estadísticos en el tema. Ahora bien, para su aplicación es indispensable partir de información, la cual en el ámbito de estudio socioeconómico suele provenir de encuestas realizadas a los individuos. Esto tiene varias repercusiones que pueden incidir tanto en la precisión de la selección de beneficiarios como en la medición posterior de los impactos (lo que se verá posteriormente).

Más aún, en frecuentes ocasiones las encuestas que proporcionan los datos han sido realizadas bajo un esquema de diseño muestral complejo, lo que implica que los métodos antes mencionados no pueden ser directamente utilizados, sino que requieren las adecuaciones para considerar que no se tenga un conjunto de datos independientes y en ocasiones tampoco idénticamente distribuidos. Muchos esfuerzos se han realizado para que los paquetes computacionales incorporen ya las adecuaciones teóricas al respecto, sin embargo hay aún muchos desarrollos teóricos por realizar y por implementar, lo cual da un espacio amplio de desarrollo a futuro.

3 ¿Tratamiento vs. control? ¿Papas o individuos?

Profundizando en el tema de la evaluación de impacto, destaca de la literatura especializada el hecho de denominar a las acciones realizadas por un programa socioeconómico con miras a generar un impacto benéfico en cierto aspecto como el tratamiento. Esto no es casual, y la explicación se da en el contexto del área de diseño de experimentos, en la cual se plantea generalmente tener un grupo de tratamiento y un grupo de control, medidos antes y después de aplicado el tratamiento, obteniéndose el impacto a partir de la diferencia entre lo observado antes y lo observado después en cada grupo.

Este tipo de diseños de experimentos es ampliamente utilizado en la búsqueda de mejoras de procesos agrícolas e industriales. Un ejemplo es la aplicación de distintos tipos de fertilizantes, esquemas de riego y exposición solar al cultivo de las papas intentando mejorar su productividad, tamaño y sabor. Pero cuando los experimentos se realizan con personas (incluso con animales), aparece un componente ético que puede generar dilemas en su implementación. Es claro que los individuos no son papas, y que un impacto negativo por la aplicación de un tratamiento equívoco es mucho más trascendente en este caso.

Estableciendo las adecuadas diferencias, seguiremos hablando de grupo de tratamiento para referirnos al grupo que recibe los beneficios o las regulaciones de un programa socioeconómico y grupo de control a los que no lo reciben. Es natural pensar que una posibilidad para realizar la medición de impactos será utilizar las herramientas disponibles del área de diseño de experimentos. Si bien en la literatura se plantea esto como una posibilidad, en la práctica existen muchas limitantes para ello, siendo la componente ética la que más pesa en este sentido.

Tratar desde un enfoque de diseño experimental a los individuos a beneficiar de un programa socioeconómico básicamente significa que se puede asignar aleatoriamente la pertenencia al grupo de tratamiento y de control. En algunos casos, esto va simplemente en contra de los objetivos de un programa socioeconómico, cuya finalidad generalmente es identificar una población objetivo a atender. Y aún cuando el objetivo sea atender a toda la población es muy complicado explicar políticamente porque se escoge a algunos individuos para excluirlos.

Por estas razones en las siguientes secciones nos enfocaremos a analizar métodos de medición de impacto partiendo del supuesto que la asignación de los individuos a los grupos de tratamiento y control no es aleatoria.

4 Los supuestos de modelización. Notación

Para realizar la medición de impacto de un programa socioeconómico, el primer supuesto a realizar es que se dispone de una variable de interés medible que refleje, en su variación, los resultados que se esperan del programa, asumiendo además que dicha variable ha sido medida exitosamente para cada individuo una vez finalizado el periodo de tratamiento planificado por el programa. Este primer supuesto se refleja en que para cada individuo se dispondrá de (Y_i) .

Para poder definir la medida del impacto del programa de manera exacta se requeriría tener una situación hipotética en la cual se pudieran observar dos

escenarios, el primero en el que el programa es implementado, y el segundo en el que el programa no es implementado. Se adopta entonces la notación $(Y_{\beta,i})$ donde el indicador binario $\beta \in \{0, 1\}$ identifica si la medición del atributo Y se ha realizado en presencia o ausencia del programa.

En este punto se puede plantear una primera definición de la medida del impacto del programa para cada individuo, que intuitivamente sería $I_i = Y_{1,i} - Y_{0,i}$, es decir la diferencia entre el valor del atributo si el programa es implementado $\beta = 1$ y si no se implementa $\beta = 0$. El impacto promedio es:

$$I = E(Y_1 - Y_0) \quad (1)$$

Un análisis de esta primera definición establece su clara inviabilidad de aplicación en el contexto de la implementación de un programa socioeconómico, pues la ecuación (1) supone que se puede medir a un mismo individuo en presencia y en ausencia del programa al mismo tiempo. Este problema se manifiesta incluso en los más básicos diseños experimentales y refleja que el mayor problema que se enfrenta cuando se quiere evaluar un programa es en realidad un problema de falta de información.

En la práctica es común además que solo un conjunto de la población participe en el programa. Esto establece la necesidad de ampliar la notación, partiendo de el escenario en que el total de la población de tamaño N es dividido en dos grupos, el de tratamiento de tamaño N_T y el de control de tamaño N_C de donde $N = N_T + N_C$. Posteriormente se realiza un comentario sobre las implicaciones de tener más de un tratamiento. En este escenario el vector de información disponible para cada individuo es conformado por $(Y_{i,\beta}, D_i)$ donde D_i es un indicador binario que tomará valor uno si el individuo pertenece al grupo de tratamiento $\rightarrow D_i = 1$ si $i \in T \leftarrow$ y cero en otro caso $\rightarrow D_i = 0$ si $i \in C \leftarrow$. En otras palabras, D_i es la variable que determina si el individuo es seleccionado o no para recibir el programa.

Dado que el interés al plantear estas medidas de impacto está concentrado en la población que efectivamente recibirá los beneficios del programa, la medida del impacto definida en la ecuación (1) se redefine a partir de los elementos previamente mencionados como $I_T = E(Y_1 - Y_0 | D = 1)$ que con un poco de simple matemática resulta ser:

$$I_T = E(Y_1 | D = 1) - E(Y_0 | D = 1) \quad (2)$$

es decir, la diferencia entre los valores esperados calculados solo para el grupo que recibió el programa, del dato observado de la variable de interés y el dato que hubiera sido observado si el programa no hubiera sido implementado. Pero el hubiera no existe. Es importante señalar que esta medida de impacto en la literatura especializada suele denominarse como impacto promedio en los tratados.

Una propuesta natural que surge para realizar la medición de los impactos es estimar el valor de I_T a partir del cálculo de:

$$G = E(Y_1|D = 1) - E(Y_1|D = 0)$$

En este caso ambas esperanzas son calculables, y lo que se compara es el promedio estadístico de la variable de interés para los que recibieron el programa con el mismo promedio obtenido en el grupo que no recibió el tratamiento. Esta estimación solo es correcta cuando el grupo de control es un adecuado contra-factual del grupo de tratamiento. Esta suposición suele ser errónea, como demostraremos a continuación, pudiéndose incluso presentar un caso extremo en que el valor G sea cero pero el impacto del programa sí exista.

Establecer la relación entre I_T y G en este punto es clarificante, puesto que si planteamos que:

$$I_T = G + S \tag{3}$$

y determinamos S , estaremos en condiciones de saber cuál es el error (denominado sesgo de selección) cometido al estimar I_T con G . El álgebra conduce a que:

$$S = E(Y_1|D = 0) - E(Y_0|D = 1)$$

en este escenario se requeriría estimar este sesgo para así calcular el impacto, o bien validar el supuesto de que $S = 0$.

Por lo anterior, dos opciones se abren en este momento para poder estimar el valor de I_T de manera adecuada. La primera es buscar que $S = 0$ de alguna forma, la segunda, es buscar otro método para estimar I_T (o alternativamente $S \neq 0$). Algunos de estos métodos contemplan el hecho de que pueden existir otras variables que influyan en los cambios en la variable de interés.

En las siguientes secciones se plantearán las herramientas estadísticas disponibles para realizar ambas alternativas.

5 Métodos que buscan eliminar el sesgo de selección. Uso de información adicional

Los métodos que buscan estimar I_T a partir de G validando el supuesto de que $S = 0$ parten de asumir que $E(Y_1|D = 0) = E(Y_0|D = 0)$, bajo la hipótesis de que el valor esperado en la variable de interés en el grupo de control sería el mismo con o sin la implementación del programa. Este supuesto se conoce como de independencia condicional en media. Si bien esta hipótesis es razonable, existe la posibilidad de que el programa tenga un efecto también en la población que no es beneficiada directamente por el programa debido a la interacción social

con el grupo en tratamiento en lo que se denomina como efectos secundarios (spillover effects).

Bajo el supuesto de independencia condicional en media (asumiendo que no existen efectos secundarios) el sesgo es:

$$S = E(Y_0|D = 0) - E(Y_0|D = 1)$$

El valor de S así definido, y que en la literatura suele denominarse sesgo de selección, es la diferencia en los valores esperados entre los grupos de tratamiento y de control, de los valores que hubieran sido observados si no se hubiera implementado el programa.

La manera más simple en la que $S = 0$ se da cuando la asignación de los individuos a los grupos de tratamiento y control es aleatoria, en cuyo caso la variable de interés es independiente de la asignación a los grupos y por ende el sesgo es cero. Ya se ha mencionado que este caso difícilmente se presenta en programas socioeconómicos dadas las implicaciones éticas que ello conlleva.

Una alternativa cuando no se tiene selección aleatoria es recurrir a la disponibilidad de un conjunto de variables \underline{x} que contengan toda la información observable disponible sobre los individuos y que tiene relación o influencia en la variable de interés Y .

Esta información se refleja en que ahora el vector para cada individuo estaría conformado por: (Y, \underline{x}, D) .

Recordando que el impacto del programa a nivel individual se define como $I_i = Y_{1,i} - Y_{0,i}$, si se utiliza la información adicional del vector \underline{x} el impacto individual puede redefinirse como $I_i = (Y_{1,i} - Y_{0,i}|\underline{x})$, o bien $I_i = (Y_{1,i}|\underline{x}) - (Y_{0,i}|\underline{x})$. Si este valor se calcula solo para el grupo en tratamiento se tiene entonces que el impacto es $I_i = (Y_{1,i}|\underline{x}, D_i = 1) - (Y_{0,i}|\underline{x}, D_i = 1)$. Es claro que el segundo componente no es observable, puesto que no se puede medir el valor del atributo de interés de un individuo seleccionado para recibir el programa si el programa no hubiera sido implementado. Una alternativa para estimar dicho valor es encontrar un individuo en el grupo de control con las mismas características \underline{x} que el individuo i -ésimo, de forma tal que sea su contra-factual, de donde se tiene $I_i = (Y_{1,i}|\underline{x}, D_i = 1) - (Y_{1,i^*}|\underline{x}, D_i = 0)$. De este modo el impacto del programa sería:

$$I_T = E(Y_1|\underline{x}, D = 1) - E(Y_{1^*}|\underline{x}, D = 0)$$

Nótese que se hace uso del supuesto de independencia condicional en media a nivel individual. Este procedimiento se conoce con el nombre de empate (matching), y asume que para dos individuos con las mismas características observables influenciando la variable de interés, en ausencia del programa el

valor observado en dicha variable sería el mismo, por lo cual el sesgo S sería cero.

En principio suena razonable buscar individuos de iguales características observables a aquéllos que están siendo beneficiados, sin embargo un empate perfecto será prácticamente imposible de efectuar, sobre todo cuando el vector \underline{x} es de gran dimensión.

Una alternativa para simplificar el procedimiento consiste en calcular la probabilidad de que los individuos sean elegidos para integrar el grupo de tratamiento a partir de sus características (propensity score):

$$p(\underline{x}) = Pr(D = 1|\underline{x})$$

Una vez calculado dicho valor, se busca para cada individuo del grupo de tratamiento, a aquel individuo del grupo de control con la probabilidad $p(\underline{x})$ más similar, convirtiéndose en su contra-factual.

Este método, conocido como propensity score matching, parte del supuesto denominado de traslape que establece que $0 < Pr(D = 1|\underline{x}) < 1$. Este supuesto implica que para cada individuo con las características \underline{x} existe probabilidad distinta de cero de haber sido elegido para el grupo de tratamiento. En la práctica esto significa que es viable encontrar para cada individuo en tratamiento a un contra-factual en el grupo de control. Si esto no es viable se debe considerar otro tipo de métodos como el de discontinuidad en regresión (regression discontinuity), por ejemplo.

Una precaución que debe tenerse al implementar este método es que, una vez calculada la probabilidad $p(\underline{x})$ para todos los individuos, se deben considerar inelegibles para el empate a los pertenecientes al grupo de control que estén fuera del rango de las probabilidades $p(\underline{x})$ observadas para el grupo de tratamiento.

Una variante a este método consiste en asignar no uno sino varios individuos contra-factuales a cada individuo en el grupo de tratamiento, ponderando sus valores del atributo de interés, en la búsqueda por minimizar los errores.

6 Métodos que utilizan información pre y post intervención

Otra alternativa para calcular el valor del impacto del programa se puede realizar cuando se ha previsto la medición del atributo de interés previo a la implementación del programa. En este caso la notación sobre la información disponible para cada individuo se refleja en el vector $(\alpha \forall \beta, i)$ donde $\alpha = \{a, d\}$ indica si la medición fué hecha antes o después del programa. En este caso la estimación del impacto (conocida como comparación reflexiva) será:

$$I_T = E(dY_1) - E(aY_0)$$

Nótese que en este caso ${}_aY_0 = {}_aY_1$ dado que ambos valores medidos antes del programa aún no han tenido influencia del mismo. Claramente esta medida de impacto se refiere al grupo en tratamiento, y sería invariante cuando el grupo de control fuera un contra-factual válido, en cuyo caso se estaría midiendo el impacto total del programa, pero como se ha visto previamente, esto solo ocurre en caso de selección aleatoria del programa o bien cuando se ha logrado construir de manera cuasi-experimental un grupo de control.

El método reflexivo asume adicionalmente que no existiría una variación en la variable de interés en ausencia del programa, es decir, que ${}_aY_0 = {}_dY_0$. Si este supuesto no es válido, se está nuevamente en presencia de un sesgo.

Para medir efectivamente el impacto del programa, eliminando el sesgo descrito previamente, se puede recurrir a la información disponible sobre el grupo de control, que en el caso de estar disponible también en ambos periodos de tiempo, permite calcular el impacto como:

$$I = E(dY_1 - {}_aY_0 | D = 1) - E(dY_1 - E_aY_0 | D = 0)$$

Este método es conocido como dobles diferencias o de diferencias en diferencias.

7 Métodos que modelan la variable de interés como función del impacto y de otros atributos

Como una alternativa a la construcción de un grupo de control que efectivamente funcione como contra-factual, se puede plantear la modelación de la función que determine el valor de la variable de interés a partir de las características observadas y de la presencia del programa. Un modelo lineal tradicionalmente utilizado para ello es:

$$Y = a + b * D + \underline{c}'x + u$$

En este modelo, el estimador del parámetro b corresponde al estimador del impacto esperado del programa. Como todo modelo estadístico, se tienen supuestos que validar, entre los que sobresale que $E(u|D) = 0$, significando que la variable D es exógena, esto es que no existe sesgo de selección, que como hemos visto hasta ahora es muy difícil de garantizar. Idealmente, la presencia de las variables observadas \underline{x} pueden conducir a que el supuesto sea válido, sin embargo en muchas ocasiones, existen factores no observables no incluidos en el vector \underline{x} que no permiten validar el supuesto.

Una interpretación alternativa de dicho supuesto es que la variable de interés tiene algún impacto en la variable de selección, por lo cual el error no puede ser tratado de la forma tradicional en los modelos de regresión.

Existen alternativas para solucionar este problema de modelación, la más frecuentemente utilizada es el uso de las denominadas variables instrumentales o instrumentos. Un instrumento es una variable z para la cual la variable de asignación al grupo de tratamiento o control depende de dicha variable, es decir:

$$D_i^* = \gamma_0 + \gamma_1 z_i + v_i$$

Con la característica fundamental de que esta variable no es determinante en la variable de interés Y , y es tal que $Cov(z, v) = Cov(u, v) = Cov(x, u) = 0$ pero $Cov(D, z) \neq 0$.

Si se encuentra una variable con tales características, se puede estimar el valor de D^* y con dicho valor estimar el valor del impacto en la ecuación $Y = a + b * D^* + \beta'x + u$ utilizando técnicas usuales de estimación.

Referencias

- [1] Cameron, A. C. y Trivedi, P. K., (2005). Microeconometrics. Methods and Applications. Cambridge University Press.
- [2] Ravallion, M (1999). The Mystery of the Vanishing Benefits. Ms. Speedy Analysts Introduction to Evaluation. Policy Research Working Paper. The World Bank
- [3] Soto, H. (2004). Técnicas estadísticas de solución al problema de agrupación. Aplicación a la clasificación de la población por su condición de pobreza Tesis de maestría - Universidad Nacional Autónoma de México.