

Sobre el abordaje de la metodología de muestreo en evaluaciones de políticas públicas

Andrés Gutiérrez¹, M.Sc. Ph.D(c)

La realidad de las encuestas

Una evaluación es un proceso objetivo de entendimiento acerca de cómo fue implementada una política, u otra intervención, qué efectos tuvo, para quiénes, cómo y por qué. Existen muchas razones por las cuales el gobierno debe poner en marcha la evaluación de sus políticas, pero una razón de peso por la cual no evaluar una política es inadecuado, es porque esto significaría que no se podría proveer ningún tipo de evidencia significativa que apoye cualquier afirmación que se quiera realizar acerca de la efectividad de una política.

Las encuestas han tenido una gran trascendencia en las evaluaciones que el gobierno realiza para medir los resultados o el impacto de las diferentes intervenciones del gobierno en la sociedad². De esta manera, el sector gubernamental está interesado en investigar, no sólo, la efectividad de la intervención, sino su implementación y desarrollo; también querrá determinar las mejoras de las condiciones sociales y/o económicas de los distintos actores sociales.

Aunque no constituyen el único método existente, sí es posible afirmar que las encuestas, hoy por hoy, son una de las más poderosas herramientas para la evaluación de las políticas públicas. En este sentido, mediante la implementación de una encuesta, es posible realizar evaluaciones que indaguen acerca del impacto de una política, en subgrupos específicos de la población, mediante la estimación de sus efectos en términos de lo que se esperaba al principio de la intervención, comparada con otra intervención, o comparada con un escenario contra factual. De la misma forma, también es posible realizar evaluaciones que indaguen cómo, por qué y bajo qué condiciones una política es exitosa o no³.

Dada la importancia de las encuestas, es muy pertinente que la entidad encargada de evaluar las intervenciones estatales establezca procesos óptimos de valoración, en términos de metodologías de muestreo, que permitan proponer, ejecutar y analizar las estrategias de muestreo que induzcan cifras confiables y precisas. Este corto ensayo tiene como objetivo establecer algunos parámetros fundamentales que el evaluador debe tener en cuenta al momento de diseñar una estrategia de muestreo.

¹ Asesor en muestreo. Dirección de Evaluación de Políticas Públicas. Departamento Nacional de Planeación. contacto@gutierrezandres.com www.gutierrezandres.com

² The Magenta Book: Guidance notes for Policy Evaluation and Analysis. (2011). HM Treasury.

³ Silva, P. (2009). Policy evaluation and sample surveys: a brief introduction. Southampton Statistical Sciences Research Institute.

La complejidad de la estrategia

Dependiendo del tipo de evaluación que se requiera, el plan metodológico de recolección de datos y análisis de información se hace más complejo. Silva (2009) aclara que, para realizar una óptima evaluación, es menester definir claramente tanto la intervención y sus alcances, como la población objetivo, de tal forma que sean identificables *a priori*; los resultados de la política deben ser claramente medibles y el diseño de la evaluación debe ser consensuado y aceptado. De esta forma, existe una pregunta importante que debe ser respondida: ¿cuál hubiese sido la situación si la intervención no se hubiese realizado? , la cual es imposible de responder en la práctica, existen varios abordajes para tratar de acercarse, por lo menos empíricamente, a simular una realidad similar. Uno de ellos es crear grupos *contrafactuales*, de no participantes en el programa (elegibles o no), que permita diseñar una evaluación de tipo experimental y contrastarlo con un grupo de participantes.

Ya sea que se decida crear grupos de participantes y no participantes, por ejemplo en el caso de una evaluación de impacto, o sólo crear un grupo de participantes, en el caso de una evaluación de resultados, la asignación de las unidades de muestreo a los grupos debe ser realizada de forma aleatoria. Es valioso señalar la diferenciación existente entre azar y aleatoriedad, porque si bien muchos estudios son realizados al azar, estos no están fundamentados sobre la base científica que sustenta los procesos de muestreo estadístico, la aleatoriedad. Esta base permite asegurar que la selección de las unidades de muestreo esté libre de cualquier tipo de sesgo y que la confiabilidad y precisión de los resultados sean aceptables.

Al realizar la asignación de las unidades a los diferentes grupos, es importante tener en cuenta que existen varias restricciones éticas y prácticas que impedirán que el diseño experimental sea completamente aleatorizado. Por ejemplo:

1. En un estudio de impacto en el sector salud, surgen problemas éticos cuando la intervención (que se supone positiva) cubre a los participantes del grupo tratamiento, pero no a los participantes del grupo control.
2. En un estudio de intervención a madres gestantes, surgen complicaciones prácticas debido a que la intervención no puede durar más que el periodo de gestación. Esto supone que los afectados por las políticas públicas pueden cambiar su estatus fácilmente en periodos cortos de tiempo.

Por estas y otras razones, el investigador preferirá en algunos casos definir metodologías cuasi-experimentales o no experimentales, sin que esto afecte la aleatorización de las unidades a los nuevos grupos, que resultan más eficientes en términos de costos y tiempos. Por lo tanto, los participantes de los distintos grupos se observan antes y después de la intervención, lo cual puede, en algunas situaciones, redundar en dificultades de análisis puesto que la caracterización de los grupos puede haber cambiado, de manera estructural, no necesariamente por la intervención per se, sino por alguna otra razón ajena a la política pública. Por

ejemplo, si el gobierno está interesado en evaluar el impacto, en términos de calidad de vida, de algún tipo de subsidio económico en una subpoblación específica, después de la intervención, los resultados del estudio pueden arrojar índices más altos en términos de calidad de vida, pero esto puede no deberse al subsidio económico, sino, por ejemplo, a la recuperación general de la economía en ese periodo de tiempo, entre otros.

Por consiguiente, el diseño de muestreo, los métodos de estimación y el abordaje estadístico a las posibles complicaciones que se puedan presentar durante la evaluación, deben estar supeditadas directamente al tipo de metodología que se utilice para afrontar el estudio. Por ejemplo, si la evaluación es de resultados (o se está levantando una línea base), bastará con proponer un diseño de muestreo y un estimador de medias; desde luego las expresiones teóricas para el cálculo de la precisión y confiabilidad son simples. Por otro lado, si la evaluación es de impacto, y la metodología de la encuesta es de tipo panel o semipanel, es muy posible que el diseño de muestreo sea más complejo y que el estimador sea una función, muchas veces no lineal, de medias. Es así como surgen los estimadores de diferencias en diferencias y todas sus variantes al momento de calcular tasas. Por ejemplo, si en una evaluación de impacto, de algún programa en una población en condición de desplazamiento forzado, es de interés estimar la proporción de personas desplazadas que no están siendo intervenidas por el programa, entonces el estimador que se debe utilizar debe ser un estimador de razón, puesto que el número de personas desplazadas no es conocido y debe ser estimado. Más aún si la evaluación está en un periodo avanzado, es posible estimar la diferencia entre esas proporciones en diferentes tiempos; luego la expresión teórica del estimador es mucho más compleja, junto con las expresiones de varianza.

Con base en lo expuesto anteriormente, es muy pertinente que el equipo técnico auditor de las evaluaciones, tenga un entendimiento diáfano de los criterios que se deben tener en cuenta para aceptar una propuesta metodológica en términos del diseño de muestreo, la escogencia del estimador y el cálculo de los errores que permiten tener acceso a intervalos de confianza.

Diseño de muestreo

Cuando el gobierno decide evaluar una política por medio de una encuesta, el investigador debe, después de decidir el tipo de metodología apropiada para realizar la evaluación, escoger un diseño de muestreo que permita obtener la información pertinente de las unidades finales de muestreo – que deberían coincidir con los beneficiarios de la política. El diseño de muestreo es la herramienta más poderosa que permite obtener una muestra de beneficiarios. Es importante tener en cuenta que no siempre se tiene un marco de muestreo que permita seleccionar directamente las unidades y por lo tanto, en muchas ocasiones, se acude a diseños de muestreo en varias etapas o en varias fases.

Dependiendo del tipo de diseño de muestreo, se debe asegurar que se cumpla con el principio de representatividad⁴, que constituye el andamiaje de las encuestas probabilísticas. Por lo tanto, si la política ha tenido mayor intervención en algunas áreas, municipios, hogares, o personas, es conveniente utilizar diseños de muestreo con probabilidades desiguales. En esta etapa, el investigador debe definir un tamaño de muestra óptimo que asegure un error mínimo, por lo general menor al cinco por ciento.

Por otra parte, en evaluaciones de impacto, es muy común que no se pueda identificar a los beneficiarios de una política. Por ejemplo, considere una evaluación de impacto de una política que provee atención en materia de salud, sólo a las familias pertenecientes a un subgrupo de la población con características socioeconómicas muy específicas (Silva, 2009). Luego, en el levantamiento de la línea base, dado que no existe un marco de muestreo de estas familias, es conveniente utilizar un diseño de muestreo en dos fases. En la primera fase se selecciona una muestra probabilística de hogares (elegibles y no elegibles), mediante un diseño de muestreo en varias etapas. Luego, después de preguntar a los hogares seleccionados por sus características socioeconómicas y por su membresía en el programa de la política, se procede a seleccionar, en la segunda fase, una submuestra estratificada de hogares; el primer estrato definido por los hogares de la submuestra que sí son beneficiarios (grupo tratamiento), el segundo estrato conformado por los hogares elegibles que no son beneficiarios (primer grupo control) y el último estrato determinado por los hogares que no son elegible y por lo tanto no reciben ningún beneficio (segundo grupo control). Estos son ejemplos claros en donde el diseño de muestreo se ciñe rigurosamente a la metodología de evaluación seleccionada por el equipo técnico.

Escogencia del estimador

La estrategia de muestreo se complementa con la escogencia del estimador. Existen variadas opciones que permiten estimar los parámetros trazadores que permiten verificar el éxito o fracaso de la intervención. El estimador es una expresión matemática que presenta ventajas o desventajas al utilizarlo en conjunto con un diseño de muestreo.

En esta etapa el investigador debe responder a las siguientes preguntas: ¿es más conveniente ponderar los resultados de la encuesta? Si es así, ¿de qué forma se deben ponderar? La respuesta a estas preguntas determina el tipo de estimador que se debe utilizar. Por lo general, se escogen estimadores naturales inducidos por el diseño de muestreo. Sin embargo, cuando existe información secundaria a la mano y esta información está correlacionada estructuralmente con el comportamiento de la población objetivo, en términos de la respuesta a la intervención, entonces es conveniente considerar estimadores que induzcan nuevas ponderaciones.

⁴ Gutiérrez, A. (2011). El principio de representatividad en la escogencia de la mejor estrategia de muestreo. IB Revista de la Información Básica. DANE.

La escogencia del estimador define la longitud de los intervalos de confianza generados en la evaluación de la política pública, sujetos a un mismo nivel de confiabilidad. Por ejemplo, en una evaluación de resultados, en donde se quiere tener conocimiento de un indicador definido como la proporción de personas satisfechas con la intervención sobre las personas intervenidas, es posible que con una mala escogencia del estimador se obtenga un intervalo de confianza, al 95%, que estime que esta proporción está entre el 20% y el 80%. Asumiendo el mismo escenario, una buena escogencia del estimador, puede inducir una estimación por intervalo, al 95% de confianza, entre el 34% y el 36%. Por supuesto, es deseable que se presente la segunda situación y no la primera. Es importante recalcar que el investigador debe asegurar que la información secundaria esté relacionada con las características de interés. De otra forma, aunque el estimador sea estadísticamente insesgado, la variabilidad de la estrategia de muestreo puede verse aumentada dramáticamente⁵.

Medidas de confiabilidad y precisión

Aunque la escogencia del diseño de muestreo y el estimador sean de libre elección para los investigadores, no lo es el cálculo de las medidas de confiabilidad y precisión. Dado que la base científica sobre la cual descansa el muestreo es la inferencia estadística⁶, se deben respetar las normas básicas para la asignación y posterior cálculo del margen de error, que constituye una medida unificada del error de muestreo que cuantifica la incertidumbre acerca de las estimaciones en una encuesta⁷.

Con base en lo anterior, es realmente imperativo, y merece ser afirmado con vehemencia, que el cálculo del error de muestreo esté acorde con el diseño de muestreo que haya sido utilizado en la evaluación. Por ejemplo, si para evaluar la penetración de una política pública, por medio de una evaluación de resultados, se propone un diseño de muestreo en varias etapas, con selección proporcional y estratificado; entonces, la varianza del estimador está supeditada a las etapas del muestreo, a la desigualdad en las probabilidades de selección/inclusión y al proceso de estratificación. En los países latinoamericanos, es muy común encontrar prácticas, si bien no malintencionadas, sí equivocadas, en donde el error de muestreo se calcula sistemáticamente usando expresiones de muestreo aleatorio simple sin reemplazo, sin tener en cuenta que el proceso de recolección de información no es así de simple.

Ahora, este proceso puede llegar a ser más complejo cuando se presentan estimaciones en un periodo avanzado de una evaluación de impacto. Máxime si se consideran muestras panel o semipanel, en donde se hace necesario utilizar

⁵ Gutiérrez, A. (2011) Una nota sobre algunas estrategias representativas en muestreo y su validez práctica. Revista Comunicaciones en Estadística.

⁶ Zhang, H. & Gutiérrez, A. (2011). Teoría estadística, aplicaciones y métodos. Editorial de la Universidad Santo Tomás.

⁷ Scheuren, F. (2004). What is a Survey. Survey Research Methods Section - American Statistical Association.

expresiones anidadas, muchas veces de diseños de muestreo en fases, que den cuenta del efecto longitudinal de las mediciones en los periodos de comparación.

Las complicaciones en el cálculo de los errores de muestreo pueden ser mayores dependiendo de la escogencia de la estrategia de muestreo (diseño de muestreo y estimador). Por lo tanto, en muchas ocasiones es posible acceder a un cálculo computacional aproximado que permite, pasar por alto el proceso teórico de definición de las cantidades que estiman la varianza de la estrategia. Cuando esto ocurre, es muy importante que el investigador presente la sintaxis computacional que utilizó para obtener las estimaciones. Asimismo, es indispensable que el equipo evaluador coteje junto con el investigador que las expresiones teóricas de varianzas, o su correspondiente sintaxis computacional, efectivamente coincidan con el cálculo del error de muestreo inducido por el diseño de muestreo.

El cálculo del error de muestreo es la garantía que el investigador le impone a este tipo de estudios, que son considerados productos intangibles, y por consiguiente se hace inaceptable que existan disparidades en las expresiones teóricas o computacionales.

Enfrentando la ausencia de respuesta

Una de las dificultades de la implementación de una encuesta es la ausencia de respuesta. Este es un fenómeno que constituye una regla marcada en la mayoría de las encuestas⁸. Existen variados tipos de ausencia de respuesta. Por ejemplo, hay una clara diferenciación entre un cuestionario vacío, puesto que el respondiente se negó a colaborar, o porque la ubicación o identificación de la unidad de observación son errados, y un cuestionario en donde algunas preguntas sí son contestadas y otras no, posiblemente debido a la longitud del cuestionario, indisponibilidad parcial del entrevistado u otras circunstancias.

Ante este fenómeno, el investigador puede tomar diversos tipos de posiciones. En el contexto de la evaluación de políticas públicas, la actitud más perjudicial es hacer caso omiso a esta situación; en otras palabras, ignorar que se presenta la ausencia de respuesta. De una manera u otra, se hace necesario que el investigador proponga algún tipo de abordaje metodológico para paliar las graves consecuencias de esta anomalía.

Es así como el equipo técnico debe vigilar la correspondiente validación de las estimaciones, que deben pasar por un proceso técnico en donde es posible mantener acotado el error no muestral causado por esta situación. A manera de sugerencia, es posible realizar un análisis estadístico exploratorio para corroborar las posibles diferencias entre la población de respondientes y no respondientes. Si existe alguna evidencia de esta diferenciación, como usualmente ocurre, entonces se debe proceder a realizar una intervención en el análisis de la información.

⁸ Gutiérrez, A. (2009). Estrategias de muestreo, diseño de encuestas y estimación de parámetros. Editorial de la Universidad Santo Tomás.

Técnicamente, existen diversas maneras de abordar este fenómeno. A continuación se enuncian algunas:

- Una alternativa puede ser realizar un ajuste de los factores de expansión inducidos por el diseño muestral, que originalmente fue planeado con un tamaño de muestra más grande que el efectivo.
- Si se tiene acceso a información auxiliar, podría ajustarse un modelo de regresión, ponderado por las probabilidades de inclusión, para imputar los valores faltantes.
- Es posible suponer que la muestra de respondientes corresponde a una submuestra de la población, definida como una mezcla de respondientes y no respondientes, y utilizar los principios de los diseños en dos fases.
- Estimar un modelo para la ausencia de respuesta, por lo general dado por una distribución Bernoulli o Binomial, y combinarlo con un estimador de calibración.

Acerca de la representatividad

La representatividad es el sello de calidad en el mundo de las encuestas. Todos quieren esta caracterización en sus estudios, pero obtenerla no es tarea fácil. Este término es muy difundido entre todas las partes involucradas en el desarrollo de una encuesta, y por lo tanto, en la evaluación de políticas públicas, es imprescindible que este término no se convierta en un cliché técnico.

En términos *non sanctos*, la representatividad se consigue cuando, para la estimación de los parámetros de interés, dentro de los diferentes subgrupos poblacionales de interés (grupos etarios, regiones geográficas, niveles de escolaridad, grupos socioeconómicos, entre otros) el error de muestreo está por debajo del establecido en los términos de referencia. Generalmente, si la evaluación requiere altos niveles de desagregación, y si se exige que para estos niveles el estudio sea representativo, entonces el costo de la encuesta subirá substancialmente puesto que se requerirá que el tamaño de muestra sea alto.

Por otro lado, es pertinente recalcar una fácil regla de asociación, derivada del muestreo estratificado: si existe representatividad en los subgrupos poblacionales, entonces existe representatividad en la población; pero si existe representatividad en la población, esto no necesariamente indica que haya representatividad dentro de los subgrupos. Esta regla es esencial para abordar adecuadamente el diseño metodológico de una evaluación.

Por ejemplo, asumiendo que se quiere levantar la línea base para la evaluación de una intervención en la población de menores de cinco años, y que debido a la ausencia de un marco de muestreo en esta población objetivo, se propone una muestra de hogares - de tal manera que se entreviste a los niños menores de cinco años dentro de los hogares - entonces, bajo estas circunstancias no basta

con obtener representatividad de los hogares, sino que se debe garantizar representatividad en los niños menores de cinco años, los cuales constituyen la población objetivo. Por supuesto, hay dos maneras de satisfacer estos requerimientos: la primera, que consiste en aumentar sustancialmente el tamaño de muestra de hogares, lo cual redundaría en obtener un tamaño de muestra más elevado de niños menores de cinco años y la segunda, flexibilizar el condicionamiento de representatividad en los subgrupos, lo cual puede constituir una grave falta a los términos de referencia de la evaluación.

Discusión

Este ensayo constituye una caracterización primaria y no exhaustiva de la realidad de las evaluaciones de políticas públicas. Por supuesto, es posible profundizar mucho más, en el contexto de la evaluación de una política, en el desarrollo metodológico, en la propuesta de la estrategia de muestreo y en las muchas dificultades que el investigador enfrenta cuando diseña, ejecuta y analiza una encuesta.

En términos generales, el equipo técnico debe velar por una correcta caracterización de los elementos de la encuesta. Desde el diseño de muestreo, para que induzca probabilidades de selección/inclusión acordes a las respuestas esperadas; en la escogencia del estimador, para que efectivamente tenga las características deseadas de insesgamiento y precisión; cotejando que el cálculo del error de muestreo coincida con la definición de la estrategia de muestreo y teniendo en cuenta el proceso de modelamiento estadístico al enfrentarse a disyuntivas comunes como la ausencia de respuesta.

Por supuesto, la aceptación de la propuesta metodológica en términos de muestreo, dependerá de la capacidad del investigador al conciliar los anteriores esquemas de análisis. Si la propuesta del investigador está mal definida o incompleta, el equipo técnico debe indicar las falencias de la misma para que sean corregidas oportunamente. De lo contrario, la sociedad fácilmente podrá indagar, no solamente acerca de su capacidad profesional, sino también de su idoneidad para laborar en este contexto, en donde la mayoría de recursos para la ejecución de la evaluación vienen del erario público.