ACADEMIA HUMANITARIA · HUMANITAIRES JOBS

Mini MBA Programas y MEAL: diseñar, dirigir y demostrar el impacto

Material del curso · Semana 10 : Medir el impacto: contrafactual, métodos y contribución

Programa de exactamente 3 meses (12 semanas, 144 horas) para diseñar, dirigir y evaluar programas humanitarios: evaluación de necesidades y severidad, cadena de resultados e indicadores, focalización y transferencias monetarias, planificación y presupuesto, plan MEAL y calidad de los datos, muestreo, análisis, rendición de cuentas, evaluación, medición del impacto, optimización de recursos y gestión adaptativa. Ejercicios numéricos corregidos, tres exámenes mensuales con estudios de caso y un diploma verificable de la Academia Humanitaria.

https://humanitairejobs.org/es/academia/mini-mba-programmes-meal/semaine-10/

Documento gratuito de la Academia Humanitaria. Puede imprimirse y compartirse libremente con fines de formación, sin modificaciones.

Semana 10 : Medir el impacto: contrafactual, métodos y contribución

Objetivos

- Comprender el problema del contrafactual y por qué una comparación antes-después no mide un efecto.
- Elegir y aplicar un método cuasiexperimental o experimental adaptado al contexto humanitario, con sus condiciones de validez.
- Utilizar los enfoques de contribución cuando la atribución es imposible, y decidir qué es razonable demostrar.

El problema fundamental: ¿qué habría ocurrido sin el programa?

El **impacto** de un programa es la diferencia entre lo que ocurrió con el programa y lo que habría ocurrido sin él: el **contrafactual**. Nunca se puede observar directamente el mismo hogar con y sin programa; todos los métodos de evaluación de impacto son maneras de construir una estimación creíble de ese contrafactual.

Dos comparaciones ingenuas son engañosas:

- **Antes-después** en los beneficiarios: el cambio observado mezcla el efecto del programa y todo lo que cambió al mismo tiempo (cosecha, precios, seguridad, otras ayudas). Si la cosecha fue buena, la seguridad alimentaria mejora incluso sin programa; si fue mala, un programa eficaz puede parecer sin efecto.
- **Beneficiarios frente a no beneficiarios** después del programa: los dos grupos eran diferentes al principio, precisamente porque la focalización eligió a los más vulnerables. La comparación subestima entonces el efecto, o lo sobrestima si fueron los más motivados los que se inscribieron (sesgo de selección).

> Pregunta que hay que plantear ante cualquier afirmación de impacto: «¿en comparación con qué?». Si la respuesta es «en comparación con antes», no es una medición de impacto.

El ensayo controlado aleatorizado y su uso en el ámbito humanitario

En un **ensayo controlado aleatorizado** (ECA), se sortea, entre las personas elegibles, quiénes reciben el programa (grupo de tratamiento) y quiénes no lo reciben o lo reciben más tarde (grupo de control). El sorteo hace que los dos grupos sean comparables en promedio en todas las características, observadas o no. El efecto es simplemente la diferencia de los resultados medios al final.

En el ámbito humanitario, negar una ayuda vital a personas elegibles para constituir un grupo de control es **contrario a la ética**. Pero los ECA siguen siendo posibles y han producido conocimientos importantes (sobre el efectivo, la nutrición, la salud mental) en situaciones concretas:

- **Recursos insuficientes** de todos modos para todo el mundo: el sorteo es entonces una forma equitativa de repartir, y el grupo no atendido sirve de control.
- **Implantación progresiva**: todo el mundo recibirá el programa, pero el orden se sortea (diseño escalonado, «stepped wedge»).
- **Comparación de dos variantes** de un programa, ambas consideradas aceptables: efectivo mensual o en un solo pago, con o sin acompañamiento.

Los ECA son caros, llevan tiempo y responden a preguntas concretas; se justifican cuando la respuesta servirá a numerosos programas. Para la mayoría de los proyectos, bastan los métodos cuasiexperimentales.

La diferencia en diferencias

La **diferencia en diferencias** compara la evolución de un grupo asistido con la evolución de un grupo de comparación no asistido: efecto = (después − antes) en el grupo asistido − (después − antes) en el grupo de comparación. Elimina las diferencias iniciales estables entre los grupos y las tendencias comunes (cosecha, precios).

Ejemplo. Proporción de hogares con un consumo aceptable. Grupo asistido: 28 % antes, 52 % después (+24 puntos). Grupo de comparación: 35 % antes, 44 % después (+9 puntos). Efecto estimado: 24 − 9 = 15 puntos. Una comparación antes-después habría atribuido 24 puntos al programa; una comparación de los resultados finales por sí solos le habría atribuido 8.

Condición de validez: el supuesto de **tendencias paralelas**. Sin el programa, los dos grupos habrían evolucionado de la misma manera. Se verifica, si es posible, con varias mediciones antes del programa, y se comprueba que el grupo de comparación no recibe otra ayuda importante durante el periodo (contaminación) ni sufre un choque específico.

En el ámbito humanitario, el grupo de comparación suele ser una zona vecina no cubierta por falta de financiación, o hogares elegibles inscritos en lista de espera. Nunca se crea un grupo de comparación negando la ayuda para las necesidades del estudio.

Emparejamiento y regresión discontinua

El **emparejamiento** construye un grupo de comparación asociando a cada hogar asistido uno o varios hogares no asistidos que se le parecen en las características medidas antes del programa (tamaño, bienes, ingresos, estatus, zona). El método del **puntaje de propensión** resume estas características en una probabilidad de ser asistido, y se comparan hogares con puntajes cercanos. Su límite: solo corrige las diferencias observadas. Si la motivación o las redes sociales influyeron en la inscripción y no se miden, el sesgo persiste.

La **regresión discontinua** se aplica cuando la elegibilidad depende de un umbral en una puntuación: por ejemplo, se asiste a los hogares con una puntuación de vulnerabilidad de 60 o más, y no a los que están por debajo. Los hogares justo por encima y justo por debajo del

umbral (57-59 y 60-62) son casi idénticos; la diferencia de resultados entre ellos estima el efecto del programa en torno al umbral. Condiciones: el umbral se aplica estrictamente y los hogares no pueden manipular su puntuación para superarlo.

Ejemplo. Hogares con puntuación 57-59: 38 % con un consumo aceptable. Hogares con puntuación 60-62 (asistidos): 51 %. Efecto estimado en torno al umbral: unos 13 puntos. Este resultado vale para los hogares cercanos al umbral, no necesariamente para los más vulnerables.

La **serie temporal interrumpida** compara la tendencia de un indicador antes y después del inicio de un programa, sobre numerosos puntos de medición (por ejemplo, las admisiones semanales en nutrición). Es útil cuando no existe ningún grupo de comparación, siempre que se disponga de suficientes puntos antes de la intervención.

Enfoques de contribución: cuando la atribución es imposible

Muchos efectos importantes no se prestan a los métodos estadísticos: una labor de incidencia que cambia una ley, un fortalecimiento de capacidades de organizaciones locales, una coordinación mejorada. No se pueden comparar con un contrafactual cuantificado. Se busca entonces establecer de forma creíble que el programa **contribuyó** al cambio.

- **Análisis de contribución** (Mayne): se expone la teoría del cambio, se reúnen las evidencias para cada vínculo, se examinan las explicaciones alternativas y se construye un «relato de contribución» que se somete a la crítica de las partes interesadas.
- **Rastreo de procesos** (process tracing): se ponen a prueba hipótesis causales buscando evidencias cuya presencia o ausencia permite confirmar o descartar una explicación (por ejemplo, la reproducción palabra por palabra de una nota de incidencia en un decreto).
- **Cosecha de alcances** (outcome harvesting): primero se identifican los cambios observados en los actores y luego se rastrea hacia atrás para determinar si el programa contribuyó a ellos y cómo.
- **Cambio más significativo**: se recopilan relatos de cambio entre las personas, y unos paneles seleccionan los más significativos explicando por qué. Útil para descubrir efectos no previstos.

> Rigor no quiere decir cuantitativo. Un análisis de contribución bien realizado, que examina seriamente las explicaciones alternativas, es más creíble que una comparación antes-después cuantificada.

Elegir el método y decidir qué hay que demostrar

La elección de un método depende de cuatro preguntas:

- **¿Qué decisión?** Extender un modelo piloto a escala nacional justifica una evidencia sólida; ajustar un programa en curso se conforma con una evidencia suficiente.
- **¿Existen ya las evidencias?** El efecto del efectivo sobre el consumo de alimentos está muy documentado: es inútil volver a demostrarlo con un ECA en cada proyecto. A menudo

basta con verificar que el programa es fiel al modelo y que los efectos intermedios están presentes.

- **¿Qué es factible y ético?** ¿Existe un grupo de comparación sin negar la ayuda? ¿Un umbral de elegibilidad nítido? ¿Una implantación progresiva?
- **¿Qué presupuesto y qué plazo?** Una evaluación de impacto seria exige una línea de base antes del inicio: se diseña con el programa, no al final.

En cuanto al **tamaño de la muestra**, una evaluación de impacto debe tener una potencia suficiente para detectar el efecto esperado. Detectar un efecto de 15 puntos requiere muchos menos hogares que un efecto de 5 puntos. Si el efecto esperado es pequeño y la muestra reducida, el estudio concluirá casi con seguridad «ningún efecto significativo», lo que se interpretará erróneamente como «ningún efecto».

Último punto: presentar con honestidad el **nivel de evidencia**. «Los hogares asistidos mejoraron su consumo en 24 puntos; en comparación con una zona de comparación, el efecto atribuible se estima en 15 puntos, bajo el supuesto de tendencias paralelas que respaldan los datos de dos campañas anteriores» es una frase que inspira confianza.

Ejercicio: estimar un efecto y criticar el método

Situación. Un programa de medios de vida apoyó a 1 200 hogares. Ingreso mensual mediano: hogares asistidos 42 USD antes, 71 USD después. Zona vecina no cubierta (por falta de financiación): 47 USD antes, 58 USD después. Durante el programa, otra ONG distribuyó semillas en la mitad de las aldeas de la zona vecina. El informe final anuncia: «el programa aumentó los ingresos un 69 %».

Trabajo solicitado. 1) ¿De dónde viene la cifra del 69 % y por qué es engañosa? 2) Calcule el efecto por diferencia en diferencias. 3) ¿Cuál es el efecto probable de la distribución de semillas en su estimación, y qué propone?

Corrección. 69 %: (71 − 42) / 42, comparación antes-después que atribuye al programa todo el aumento, incluido el debido al contexto. Diferencia en diferencias: (71 − 42) − (58 − 47) = 29 − 11 = 18 USD al mes, es decir, aproximadamente el 43 % del ingreso inicial de los hogares asistidos. Las semillas probablemente aumentaron los ingresos de la zona de comparación, lo que infla su tendencia y subestima el efecto del programa (la contaminación viola el supuesto de tendencias paralelas). Propuesta: analizar por separado las aldeas de comparación sin semillas, o controlar esta variable; presentar la estimación como un límite inferior e indicar claramente la limitación.

Para recordar

- El impacto se mide en comparación con un contrafactual; antes-después no es una medición de impacto.
- Diferencia en diferencias = evolución del grupo asistido − evolución del grupo de comparación, bajo el supuesto de tendencias paralelas.
- Un ECA en el ámbito humanitario solo es ético si no priva a nadie de una ayuda disponible: recursos limitados, implantación progresiva o comparación de variantes.
- Regresión discontinua: comparar los hogares justo por encima y justo por debajo de un umbral estricto de elegibilidad.
- Cuando la atribución es imposible: análisis de contribución, rastreo de procesos, cosecha de alcances, cambio más significativo.

Cuestionario

1. Grupo asistido: 30 % antes, 50 % después. Comparación: 32 % antes, 40 % después. Efecto por diferencia en diferencias:
   A. 20 puntos
   B. 12 puntos
   C. 10 puntos
   D. 8 puntos

2. ¿Qué supuesto es esencial para la diferencia en diferencias?
   A. Los grupos son idénticos al principio
   B. Sin el programa, los dos grupos habrían evolucionado de la misma manera
   C. La muestra supera los 1 000 hogares
   D. El programa duró más de un año

3. ¿En qué situación es éticamente aceptable un sorteo en el ámbito humanitario?
   A. Para negar una ayuda vital disponible a la mitad de los elegibles
   B. Cuando los recursos no alcanzan de todos modos para todos los elegibles, o para el orden de una implantación progresiva
   C. Nunca
   D. Solo si el donante lo impone

4. El principal límite del emparejamiento por puntaje de propensión es:
   A. Solo funciona con ECA
   B. Solo corrige las diferencias medidas, no las diferencias no observadas como la motivación
   C. Exige un umbral de elegibilidad
   D. Está prohibido por la OCDE

5. Para demostrar que una labor de incidencia contribuyó a la adopción de una ley, el método más adecuado es:
   A. Un ensayo aleatorizado
   B. Una diferencia en diferencias
   C. Un análisis de contribución o un rastreo de procesos
   D. Una encuesta de hogares

Respuestas del cuestionario : 1. B · 2. B · 3. B · 4. B · 5. C