



ABDUL LATIF JAMEEL

Poverty Action Lab



J-PAL LAC TRADUCIENDO LA INVESTIGACIÓN EN ACCIÓN

# Curso Ejecutivo de Evaluación de Impacto de Programas Sociales

10 al 14 de Junio del 2013  
Ciudad de Guatemala

## Contenido

Agenda.....	1
Objetivos.....	7
Perfiles Equipo .....	9
Casos de Estudio .....	11
Caso 1: Exponiendo Políticos Corruptos.....	11
Caso 2: Capacitaciones Profesionales .....	19
Caso 3: Programa de Profesores Adicionales .....	32
Ejercicios .....	39
Ejercicio 1: La mecánica de la aleatorización y el balance estadístico.....	39
Ejercicio 2 - Cálculos del Tamaño Muestral .....	47
Preguntas para el diagnóstico de evaluabilidad .....	55
Términos Importantes.....	57
Apuntes.....	63
Mapas.....	69
Otras informaciones.....	70

Curso Ejecutivo de J-PAL LAC |  
Ciudad de Guatemala, 10 al 14 de abril de 2013

## Agenda

### LUNES 10 DE JUNIO

Inicio	Fin	Tiempo	Sesión	Descripción	Expositor
8:30 AM	9:00 AM	0:30		Inscripción de Participantes	
9:00 AM	10:00 AM	1:00	Introducción al Curso	Bienvenida e introducción de participantes.	Anne Thibault
10:00 AM	10:30 AM	0:30		<i>Evaluación de conocimientos inicial</i>	
10:30 AM	11:45 AM	1:15	Clase 1: ¿Qué es una evaluación?	En esta unidad se discutirá la importancia de la evaluación de impacto de programas sociales. Se introducirá el concepto de evaluaciones aleatorias y las condiciones óptimas para realizar este tipo de evaluación de impacto.	Magdalena Valdés
11:45 AM	12:00 PM	0:15		Coffee Break	
12:00 PM	1:30 PM	1:30	Caso de Estudio 1: Teoría de Cambio	Caso: Exponiendo Políticos Corruptos en Brasil.	Profesores Asistentes
1:30 PM	2:30 PM	1:00		Almuerzo	
2:30 PM	4:00 PM	1:30	Clase 2: Resultados, indicadores y midiendo impacto	Esta sección trata la teoría de cambio como una guía al momento de implementar y evaluar programas sociales. Una evaluación bien diseñada tiene como meta resolver no sólo si un programa tiene impacto, sino que también, de qué manera ocurre ese impacto. Se presentarán los pasos necesarios para desarrollar una evaluación de impacto.	Magdalena Valdés
4:00 PM	4:15 PM	0:15		Coffee Break	
4:15 PM	5:30 PM	1:15	Proyecto en Grupo: Escoger un proyecto	En esta primera sesión de trabajo, los integrantes del grupo escogerán el proyecto sobre el cual trabajarán a lo largo de la semana.	Profesores Asistentes
<b>Horas</b>		<b>5:30</b>			

## MARTES 11 DE JUNIO

Inicio	Fin	Tiempo	Sesión	Descripción	Expositor
9:00 AM	10:30 AM	1:30	Proyecto en Grupo: Teoría de Cambio	Durante esta sesión, los grupos revisarán la teoría de cambio de sus proyectos y armarán una cadena causal para presentar al final del taller.	Profesores Asistentes
10:30 AM	10:45 AM	0:15	Coffee Break		
10:45 AM	12:00 PM	1:15	Caso de Estudio 2: Por qué aleatorizar	Caso Capacitaciones Profesionales en Colombia.	Profesores Asistentes
12:00 PM	1:15 PM	1:15	Clase 3: ¿Por qué aleatorizar?	Se introducirá el concepto del contrafactual y se abordarán las razones por las cuales el diseño experimental es la metodología preferida para evaluar impacto.	Diego Aycinena
1:15 PM	2:15 PM	1:00	Almuerzo		
2:15 PM	3:30 PM	1:15	Clase 4: ¿Cómo aleatorizar?	Se discutirán algunos elementos prácticos de aleatorización, incluyendo las restricciones al aleatorizar, la unidad de aleatorización y tratamientos múltiples.	Ximena Peña
3:30 PM	3:45 PM	0:15	Coffee Break		
3:45 PM	5:00 PM	1:15	Trabajo en Grupo: Diseño de Evaluación	Los grupos definirán su estrategia de identificación y el diseño que usarán para realizar una evaluación experimental.	Profesores Asistentes
<b>Horas</b>		<b>5:00</b>			

## MIÉRCOLES 12 DE JUNIO

Inicio	Fin	Tiempo	Sesión	Descripción	Expositor
9:00 AM	10:15 AM	1:15	Ejercicio 1: Ejercicio de Aleatorización	En esta sesión, los profesores asistentes guiarán a los participantes a través de ejercicio de aleatorización. Se observará la diferencias al aleatorizar grupos de diferentes tamaños y se verificará como aleatorizar asegura balance en grupos.	Thomas Vargas
10:15 AM	10:30 AM	0:15	Coffee Break		
10:30 AM	12:00 PM	1:30	Clase 5: Muestreo y tamaño de muestra	Esta sesión discutirá la importancia de tener un tamaño de muestra adecuado, con especial énfasis en conceptos como poder y efecto mínimo detectable. Se tratarán estos temas a través de algunos ejemplos experimentales realizados por académicos de J-PAL.	Diego Aycinena
12:00 PM	1:15 PM	1:15	Ejercicio 2: Cálculos de poder	En esta sesión, los profesores asistentes guiarán a los participantes a través de un ejercicio de cálculos de poder utilizando diferentes softwares estadísticos.	Thomas Vargas
1:15 PM	2:15 PM	1:00	Almuerzo		
2:15 PM	3:45 PM	1:30	Caso de Estudio 3: Amenazas y análisis	Amenazas y análisis en Kenia.	Profesores Asistentes
3:45 PM	4:00 PM	0:15	Coffee Break		
4:00 PM	5:30 PM	1:30	Clase 6: Amenazas y análisis	Durante esta sesión se discutirá el análisis de datos. Se presentarán las amenazas al caso ideal y algunos remedios. Se hablará sobre los siguientes conceptos: intención de tratar y tratamiento en los tratados, efectos promedio, efectos promedio de tratamiento local, estimación estructural y resultados múltiples.	Sebastián Galiani
<b>Horas</b>		<b>5:15</b>			

## JUEVES 13 DE JUNIO

Inicio	Fin	Tiempo	Sesión	Descripción	Expositor
9:00 AM	10:30 AM	1:30	Clase 7: Aspectos operativos de la contratación de una evaluación de impacto	En esta sección se revisarán los elementos más relevantes para aquellos que quieran contratar a una empresa encuestadora para trabajo de campo. Además se revisarán, mediante ejemplos, el costo de levantar datos para dar una idea de los órdenes de magnitud de costos. Todo esto con el fin de proporcionar rangos para aquellos dispuestos a realizar y presupuestar una evaluación aleatoria.	Alejandra Aponte
10:30 AM	10:45 AM	0:15	Coffee Break		
10:45 AM	12:00 PM	1:15	Trabajo en Grupo: Diseño de evaluación y cálculos de poder	En esta sesión los grupos discutirán los requisitos muestrales para poder detectar impacto.	Profesores Asistentes
12:00 PM	1:15 PM	1:15	Clase 8: Evaluación aleatoria de principio a fin	Durante esta unidad se presentará una evaluación de impacto de principio a fin. Se revisarán todos los pasos, desde la inepción de la idea y el diseño de la evaluación, hasta la administración del proyecto, concluyendo con el análisis de resultados.	Sebastián Galiani
1:15 PM	2:15 PM	1:00	Almuerzo		
2:15 PM	3:45 PM	1:30	Trabajo en Grupo: Trabajo en presentación	Los grupos de trabajo finalizarán los detalles de la presentación.	Profesores Asistentes
3:45 PM	4:00 PM	0:15	Coffee Break		
4:00 PM	5:30 PM	1:30	Trabajo en Grupo: Trabajo en presentación	Los grupos de trabajo finalizarán los detalles de la presentación.	Profesores Asistentes
<b>Horas</b>		<b>5:15</b>			

## VIERNES 14 DE JUNIO

Inicio	Fin	Tiempo	Sesión	Descripción	Expositor
9:00 AM	10:30 AM	1:30	Presentaciones finales	Presentaciones finales del trabajo en grupo.	Participantes
10:30 AM	10:45 AM	0:15		Coffee Break	
10:45 AM	12:15 PM	1:30	Presentaciones finales	Presentaciones finales del trabajo en grupo.	Participantes
12:15 PM	12:45 PM	0:30		<i>Evaluación de conocimientos final</i>	
12:45 PM	1:30 PM	0:45	Ceremonia de clausura	Entrega de certificados y clausura de la ceremonia.	Anne Thibault
				Almuerzo de cierre	
<b>Horas</b>		<b>3:30</b>			



## Objetivos

### Lunes 10

#### Clase 1: ¿Qué es una evaluación?

- Definir lo qué es una evaluación de impacto y situarla en el mundo de evaluación y monitoreo.
- Explicar a grandes rasgos las razones por las cuales es importante evaluar impacto de programas sociales, o de programas con un enfoque social.

#### Caso de estudio 1: Brasil

- Evidenciar la complejidad de medir el impacto de un programa social.
- Poder pasar de una teoría implícita a una cadena causal.
- Entender la importancia de una cadena causal para definir la pregunta de evaluación y poder interpretar los resultados de una evaluación de impacto.

#### Clase 2: Resultados, Indicadores y Midiendo Impacto

- Definir conceptos estándares para producir una cadena causal.
- Observar la creación de una cadena causal en la práctica.
- Poder distinguir entre una cadena causal bien construida de una que no se sustenta lógicamente.

### Martes 11

#### Caso de estudio 2: Colombia

- Aprender a distinguir entre diferentes métodos de evaluación.
- Explorar el problema de identificación causal e introducir el concepto de contrafactual.
- Introducir la intuición detrás del concepto de sesgo de selección.
- Demostrar como la asignación aleatoria elimina el sesgo de selección.

#### Clase 3: ¿Por qué aleatorizar?

- Explorar las opciones metodológicas a la hora de medir impacto.
- Diferenciar entre la evaluación experimental y las cuasi-experimentales.
- Entender las fortalezas y debilidades del diseño experimental con respecto a las técnicas cuasi-experimentales.

#### Clase 4: ¿Cómo aleatorizar?

- Detallar los aspectos prácticos de la aleatorización:
- Saber cómo decidir el nivel de aleatorización apropiado.
- Presentar opciones con respecto al método de aleatorización.
- Presentar las diferentes restricciones que pueden afectar el proyecto de aleatorizar.

## Miércoles 12

### Ejercicio 1: Aleatorización

- Mostrar el proceso de aleatorización en Excel.
- Observar cómo el tamaño de la muestra afecta el balance entre las muestras.
- Practicar la estratificación de muestra.

### Clase 5: Muestreo y tamaño de muestra

- Entender la importancia de tener un tamaño de muestra adecuado.
- Entender cómo realizar cálculos de poder ayudan a determinar el tamaño de muestra.
- Introducir el concepto de Efecto Mínimo Detectable (EMD).

### Ejercicio 2: Cálculos de Poder

- Saber utilizar OptimalDesgin para realizar cálculos de poder.
- Observar cómo ciertos parámetros afectan el poder.

### Caso de estudio 3: Kenia

- Explorar cómo un diseño experimental puede ser usado para contestar múltiples preguntas de investigación.
- Explorar diferentes diseños de aleatorización.

### Clase 6: Amenazas y análisis

- Explorar las diferentes amenazas al diseño experimental.
- Identificar cómo éstas afectan los resultados que observamos.
- Explorar diferentes tipos de análisis que se pueden emplear para identificar correctamente el impacto causal de una intervención.

## Jueves 13

### Clase 7. Aspectos operativos.

- Revisar pasos importantes para aquellos que vayan a contratar, subcontratar o trabajar en una evaluación de impacto.
- Presentar consejos para el desarrollo de Términos de Referencia para subcontratar evaluaciones de impacto.
- Analizar diferentes propuestas de presupuesto para una evaluación de impacto.

### Clase 8. Evaluación de principio a fin

- Entender las diferentes etapas que existen en las evaluaciones aleatorias, desde su diseño hasta el trabajo en campo, análisis de datos e impactos en política pública.

## Perfiles Equipo

### Perfil de Profesores



**Diego Aycinena**  
Profesor Asistente de Economía, Universidad  
Francisco Marroquín  
(Ph.D. George Mason University)

Diego Aycinena es Profesor Asistente de Economía en la Facultad de Ciencias Económicas y Director del Centro Vernon Smith de Economía Experimental en la Universidad Francisco Marroquín en Guatemala. También es afiliado del *Economic Science Institute* en la *Chapman University*. Sus investigaciones utilizan experimentos económicos para entender el comportamiento humano y las propiedades de varias instituciones. Ha utilizado experimentos en laboratorios para testear mecanismos alternativos de Mercado. Tiene una Maestría y Ph.D. en Economía de la George Mason University y un pregrado de la Universidad Francisco Marroquín.



**Sebastián Galiani**  
Profesor de Economía,  
University of Maryland  
(Ph.D. Oxford University)

Sebastián Galiani es Profesor de Economía en la Washington University en St. Louis y Profesor Visitante en la Universidad de San Andrés, Argentina. Fue el Presidente de la *Network of Inequality and Poverty* de LACEA durante los años 2004 y 2005 y miembro de su Comité Ejecutivo entre los años 2004 y 2008. Obtuvo su Ph.D. en Economía de la Oxford University y trabaja en áreas de Desarrollo Económico y Microeconomía Aplicada. También ha trabajado como Consultor para Naciones Unidas, el Banco Interamericano de Desarrollo, el Banco Mundial y los gobiernos de Argentina, México, Panamá y Sudáfrica.



**Magdalena Valdés**  
Subdirectora J-PAL LAC  
(MPA, London School of Economics)

Magdalena Valdés posee un MPA en Políticas Públicas y Económicas de la *London School of Economics and Political Science* y un MS en Finanza de la Universidad Adolfo Ibáñez. Ella se unió a J-PAL LAC en marzo de 2011 como Gerente de Políticas Públicas y Capacitación. Durante sus años de estudio en Londres, tuvo la oportunidad de trabajar como Consultora Externa para la *Organization for Economic Cooperation and Development (OECD)*, analizando el impacto de diferentes *Vocational Education and Training Systems (VET)* en Alemania, Inglaterra y Chile. Sus áreas de interés se relacionan con temas de género, ética y empoderamiento de la sociedad civil.



**Ximena Peña**  
Profesora Asistente de Economía, Universidad  
de los Andes  
(Ph.D., Georgetown University)

Ximena Peña es profesora asistente en la Facultad de Economía de la Universidad de Los Andes. Obtuvo su doctorado en economía en la Universidad de Georgetown en 2007-8. Su investigación se centra en el área de economía laboral, en temas como las decisiones de educación de hombres y mujeres, la brecha salarial de género, la brecha salarial enfrentada por las madres, entre otras cosas. Escribió el libro 'Guía Práctica para la evaluación de impacto', que ha sido adoptado como libro texto en Universidades de varios países de América Latina. Está participando en las evaluaciones de impacto de los principales programas de primera infancia en Colombia.

**Coordinadora del Curso**



**Anne Thibault**  
Training Manager  
J-PAL LAC



**Alejandra Aponte**  
J-PAL LAC  
Guatemala



**Camila Uribe**  
Universidad de los Andes  
Colombia

**TEACHING  
ASSISTANTS  
(TAs)**



**Marialejandra Guzmán**  
J-PAL LAC  
República Dominicana



**Thomas Vargas**  
J-PAL LAC  
Chile



**Diego Verdugo**  
J-PAL LAC  
Chile

## Casos de Estudio

### Caso 1: Exponiendo Políticos Corruptos

#### Teoría de Cambio, Medición y Resultados

Este caso está basado en "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes," de Claudio Ferraz y Frederico Finan (The Quarterly Journal of Economics, mayo 2008)

#### Definiciones

**Impacto:** El objetivo final del programa o el objetivo de largo plazo.

**Intervención:** La suma de insumos utilizados y actividades emprendidas para producir los bienes y servicios que constituyen la intervención o programa a entregar al grupo de tratamiento.

**Necesidades:** Los problemas que motivan una intervención y sobre los cuales se desea realizar un impacto.

**Producto:** Los bienes y servicios que resultan de la combinación de insumos y actividades.

**Resultados Intermedios:** Las actividades o uso que la población beneficiaria da a el producto.

**Supuestos:** Las condiciones que se deben cumplir para que la lógica de la cadena causal se mantenga intacta.

**Teoría de cambio:** Describe la forma en que la implementación de un programa se traduce en resultados, por medio de una cadena de relaciones causales.

\*Fuentes:

3ie (2011), 3ie Glosario para la Evaluación de Impacto. Iniciativa Internacional para la Evaluación de Impacto. Nueva Delhi, India.

Gertler, Paul, Sebastián Martínez, Patrick Premand, Laura B. Rawlings, Christel M. J. Vermeersch. La evaluación de impacto en la práctica. Banco Mundial. Washington, DC: 2010.

## Parte 1 – El contexto

### Antecedentes

En respuesta a la corrupción generalizada, el Gobierno de Brasil ha aumentado las sanciones a políticos corruptos, les ha prohibido la postulación a cargos públicos y ha aumentado el número de auditorías financieras en el sector público. Al mismo tiempo, los medios de comunicación brasileños han sacado a la luz varios casos de corrupción de alto perfil y han perseguido celosamente el tema. Esto sin embargo, puede no ser suficiente para resolver el problema al nivel local, donde los casos de corrupción no atraen la atención de la prensa nacional y la gente no tiene acceso fácil a la información sobre cómo sus líderes municipales hacen uso de los fondos y recursos públicos. Este caso de estudio trata sobre una intervención cuyo propósito era disminuir los niveles de corrupción y hacer que los políticos rindan cuenta de sus acciones. La intervención incluía auditorías aleatorias y la difusión de esta información a través de los medios de comunicación.

### Corrupción y Auditorías Aleatorias

A nivel local, la corrupción adopta muchas formas. Los políticos locales usan recibos falsos en sus proyectos, crean compañías fantasmas, aumentan los valores de productos y servicios en facturas, o simplemente utilizan recursos del gobierno para su propio beneficio. Aunque muchas personas en los municipios pueden sospechar de estas actividades fraudulentas, pocas tienen la información necesaria para evaluar el comportamiento de sus líderes locales y recompensarlos o castigarlos en las urnas al momento de votar. Como resultado, no existe control sobre el comportamiento de los políticos locales, haciendo de la corrupción un problema endémico.

En 2003, en un esfuerzo por crear un sistema más transparente y responsable, el Gobierno de Brasil inició auditorías aleatorias de los gastos de los gobiernos municipales. Cada mes, se escogía aleatoriamente cierto número de municipalidades pequeñas (aquellas con una población menor a 450 mil) por medio de un sorteo nacional. Luego, un equipo de 10 a 15 auditores de la *Contraloría General da União* (Oficina del Contralor General; CGU), eran enviados a las municipalidades seleccionadas para examinar cuentas, registros, la existencia y calidad de las obras públicas y la entrega de servicios públicos. Además, el equipo se reunía con miembros de la comunidad local para recoger quejas directas de corrupción.

Basándose en la información recopilada durante este proceso, el equipo de auditoría realizó informes que se publicaron en línea y se enviaron a la CGU, a la oficina de rendición de cuentas de Brasil, a los fiscales y al poder legislativo municipal.

### Discusión

1. Mencione tres herramientas o mecanismos que existen para enfrentar el problema de la corrupción a nivel local. ¿Cuáles de estas herramientas son efectivas?
2. ¿Cuáles son las fortalezas y debilidades del programa de auditorías implementado por el Gobierno de Brasil? ¿Cree usted que dicho programa puede ser exitoso? ¿Por qué?
3. Tomando en cuenta las debilidades del programa, ¿qué intervenciones complementarias implementaría usted?



3. Qué factores pueden intervenir, ya sea positiva o negativamente, en el impacto de la publicación de los reportes de auditoría?

### Parte 3 – Teoría de Cambio

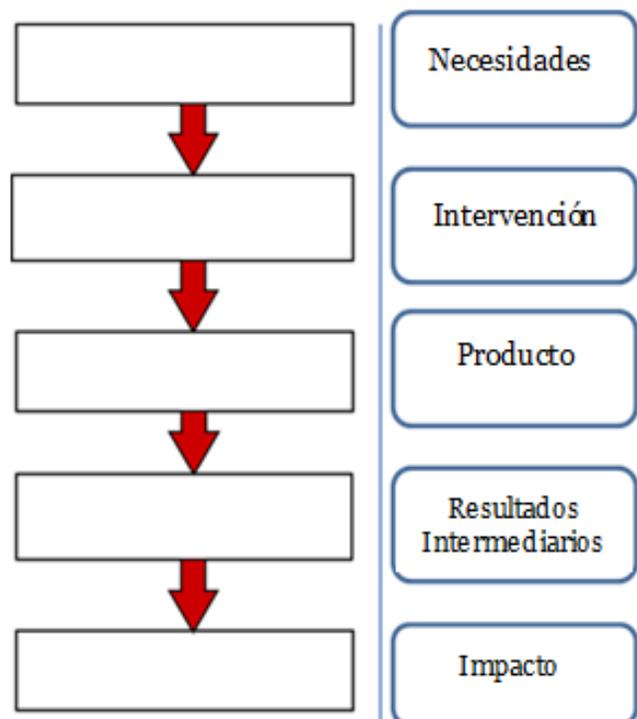
Figura 1

La teoría de cambio identifica la cadena de causalidad entre la intervención y el resultado final. La Figura 1 muestra una manera en que la teoría de cambio puede estructurarse.

#### Discusión

Dibuje una cadena causal con el formato que se presenta en la Figura 1 que conecte la difusión de los informes de auditoría a través de los medios de comunicación con los resultados finales.

¿Cuáles son las condiciones/supuestos necesarios que subyacen esta teoría de cambio? ¿Qué tiene que ocurrir para que esta cadena funcione?



## Parte 4 – ¿Qué datos se deben recopilar? Recopilación de datos y medición

Antes de decidir qué datos recopilar, se deben tener muy claros los resultados que se desean obtener, y de qué manera se cree que la intervención impactará este resultado. En otras palabras, la definición de una hipótesis principal y de la teoría de cambio al inicio de una evaluación ayuda a decidir qué información recopilar.

Para cada paso de la teoría de cambio se necesita identificar indicadores que reflejen los resultados en cada etapa e instrumentos para recopilar esta información. Además, se necesita recopilar los datos sobre sus supuestos para verificar su veracidad.

## Discusión

1. ¿Qué indicadores mediría en cada paso de su teoría de cambio? Piense en los indicadores para sus resultados intermedios, finales y para sus supuestos.
2. ¿Cómo recopilaría los datos para estos indicadores? ¿Qué desafíos cree que deberá enfrentar durante este proceso?
3. Si tuviera recursos para recolectar 3 de los indicadores sugeridos anteriormente, ¿cuáles escogería para la evaluación? ¿Por qué?



## Caso 2: Capacitaciones Profesionales

### Comparando diferentes métodos de evaluación

Este caso de estudio está basado en la publicación “Subsidizing Vocational Training for Disadvantaged Youth: Evidence from a Randomized Control Trial” de Orazio Attanasio, Adriana Kugler y Costas Meghir (2011), American Economic Journal: Applied Economics.

#### Definiciones

**Asignación Aleatoria:** Un diseño de intervención en el cual los miembros de la población elegible son asignados aleatoriamente bien sea al grupo de tratamiento o al de control (i.e. asignación aleatoria).

**Contrafactual:** Una estimación de lo que hubiera ocurrido a la unidad de análisis, en la ausencia de una intervención.

**Grupo de Tratamiento:** Grupo que recibe el programa.

**Grupo de Control:** Grupo de comparación con el grupo de tratamiento, que no recibe el programa.

**Impacto:** La diferencia entre lo que ocurrió con el programa y lo que habría ocurrido sin el programa en la población de interés.

**Variables de interés** (o dependiente): Una variable que se cree que puede ser predicha por o causada por una o más variables distintas (variables independientes).

\*Fuentes:

3ie (2011), 3ie Glosario para la Evaluación de Impacto. Iniciativa Internacional para la Evaluación de Impacto. Nueva Delhi, India.

Gertler, Paul, Sebastián Martínez, Patrick Premand, Laura B. Rawlings, Christel M. J. Vermeersch. La evaluación de impacto en la práctica. Banco Mundial. Washington, DC: 2010.

## ¿Por qué capacitar?

Muchos jóvenes alrededor del mundo, tanto en países en vías de desarrollo como en países desarrollados, tienen muchas dificultades para encontrar un empleo estable. Según la Organización Internacional del Trabajo (OIT), hacia finales del año 2011 alrededor de 74,5 millones de jóvenes estaban desempleados a nivel mundial. En países en desarrollo, la alta tasa de desempleo juvenil generalmente es atribuida a que los sistemas educativos no logran entregarles a los jóvenes la educación y las habilidades necesarias para desempeñarse en el mercado laboral. Se cree que una mejor educación y la entrega de habilidades claves podrían incrementar la probabilidad de que los jóvenes obtengan un trabajo estable.

Para enfrentar el problema de desempleo juvenil, en el año 2001 el Gobierno de Colombia comenzó un programa de capacitaciones para jóvenes entre 18 y 25 años de escasos recursos en las siete ciudades más grandes del país. El programa consistía en tres meses de capacitación en la sala de clases y tres meses de capacitación en una empresa. Las capacitaciones en las salas de clases eran impartidas por instituciones privadas seleccionadas mediante un proceso de licitación, mientras que las capacitaciones en los lugares de trabajo eran impartidas a través de compañías registradas que operaban en diversos sectores económicos.

Los participantes recibieron US\$2,20 por día para ayudar a costear gastos de transporte y almuerzo; las mujeres con niños menores de siete años recibieron US\$3 diarios.

¿Qué impacto tiene la capacitación?

¿Funcionó el programa de capacitación? ¿Qué se requiere para medir si un programa tuvo impacto?

En general, preguntarse si un programa funciona es preguntarse si el programa logra los cambios que se había propuesto realizar sin que estos cambios sean provocados por otros factores o eventos que ocurren al mismo tiempo. Para poder demostrar que un programa o intervención es el causante de los cambios observados, necesitamos poder comparar estos cambios con los que hubieran ocurrido en ausencia del programa.

¿Pero cómo sabemos qué hubiera ocurrido? Medir lo que hubiera ocurrido, en la ausencia de un programa, requiere ingresar a un mundo imaginario en el que el programa nunca se aplicó a los participantes de una intervención. Los resultados de estos mismos participantes en este mundo imaginario representan el contrafactual. Dado que es imposible observar el verdadero contrafactual, lo mejor que podemos hacer es estimarlo o buscar replicarlo.

El desafío clave de una evaluación de impacto es construir o replicar el contrafactual. Normalmente, esto se hace mediante la selección de un grupo de personas lo más similar posible al grupo de participantes, pero que no se beneficia del programa. Este grupo se denomina el grupo de control. Construir un buen grupo de control es difícil pero muy importante. Para medir impacto de manera rigurosa, necesitamos que la única diferencia entre el grupo de control y los participantes sea que el grupo de control no haya participado del programa. En la medida en que esto se cumpla, se puede argumentar con mayor confianza que una intervención es responsable de las diferencias entre los resultados de ambos grupos; por lo tanto, se puede estimar el impacto de una intervención como la diferencia observada al final del programa entre los resultados del grupo de control y los de los participantes del programa.

La precisión con la cual estimamos impacto depende de la precisión con la cual replicamos al contrafactual. Si el grupo de control representa de forma deficiente al contrafactual, el impacto se calcula de forma deficiente. Por lo tanto, el método utilizado para construir el grupo de control es una decisión clave en el diseño de cualquier evaluación de impacto. Esto nos devuelve a la pregunta: ¿Qué impacto tiene este programa de capacitación en el empleo de jóvenes vulnerables de Colombia?

En este caso, la intención del programa es entregar a los participantes ciertas habilidades valoradas por los empleadores. La variable de interés principal es la condición de empleo entre los jóvenes participando en el programa. Preguntarse si el programa funcionó es preguntarse si éste incrementó el empleo entre los jóvenes que participaron en el programa. Por lo tanto, el impacto es la diferencia entre la tasa de empleo de aquellos que participaron en el programa y la tasa de empleo de los mismos participantes si nunca hubieran participado en el programa.

Como no podemos observar la situación de los participantes en el caso de que no hubiesen participado, ¿qué grupo de control se debe usar para estimar esto? A continuación se muestran diferentes métodos para evaluar impacto usados por distintos expertos (ver tabla en la última página con una lista de diferentes métodos de evaluación y los supuestos en los cuales se basan).

### Método 1:

#### Artículo de periódico: “Grandes ganancias para mujeres en programa de capacitación”

“Datos publicados hoy por una agencia de gobierno indican que el programa de capacitación del gobierno, el cual ha estado funcionando desde el año 2001 en las siete ciudades más grandes de Colombia, incrementó la tasa de empleo de las mujeres participantes en un 49,66%. Antes de participar en el programa, sólo 46,92% de las mujeres que realizaron la capacitación tenía empleo, y meses después de completar el programa de capacitación, 70,22% tenía empleo. Este incremento es sustancial e importante para muchas mujeres jóvenes y estos números son evidencia de que los programas de capacitación del gobierno funcionan. Los gobiernos en todo el mundo deberían aprender de esta evaluación y comenzar sus propios programas de capacitación o expandir programas existentes”.

	Promedio	Error Estándar
Empleo Línea Base (antes del programa)	46,92%	0,017
Empleo Línea Final (después del programa)	70,22%	0,015
Diferencia	23,30**	
Incremento	49,66%	

Nota: Estadísticamente significativo con un 95% confianza (\*\*). Muestra = 910 mujeres.

## Tema de Discusión 1

1. ¿En qué método de evaluación se basa este artículo del periódico?
2. ¿Cómo intenta este método estimar el contrafactual?
3. ¿Cuál es el supuesto que se tiene que cumplir para que el contrafactual esté bien estimado?
4. ¿Cuáles podrían ser los problemas con este método de evaluación?

## Método 2

### Carta al Editor: “No saltemos a conclusiones”

“Los periódicos tienden a exagerar muchas cosas y esto es precisamente lo que el artículo “Grandes ganancias para mujeres en programa de capacitación” hizo la semana pasada al reportar sobre el impacto del programa de capacitación del gobierno. Como economista interesado en mercados laborales, he estado siguiendo de cerca el programa de capacitación desde que el gobierno lo anunció. Obviamente, estoy feliz de ver resultados positivos de este programa, pero las afirmaciones en cuanto a la dimensión del impacto pueden ser un tanto engañosas. Por ejemplo, muchas cosas le podrían haber pasado a las mujeres participantes entre el comienzo y el final del programa. La economía colombiana ha tenido un crecimiento sostenido desde 2002 y la seguridad ha aumentado en varias ciudades del país. Estos factores pueden haber afectado los resultados de la evaluación del programa ya que los resultados que vemos pueden haber sido causados en parte por estos otros factores. Propongo que la mejor manera de medir el impacto real del programa es enfocarnos en cómo las mujeres que participaron en el programa se comparan con mujeres que no participaron en el programa. He hecho este cálculo y verán que con este método el programa incrementó la tasa de empleo de las mujeres que participaron en el programa en un 10%, una gran diferencia al 50% expresado en el artículo, pero sigue siendo una mejora”.



## Tema de Discusión 2

1. ¿Qué método de evaluación se propone en este artículo del periódico?
2. ¿Cómo intenta este método estimar el contrafactual?
3. ¿Cuál es el supuesto que se tiene que cumplir para que el contrafactual esté bien estimado?
4. ¿Cuáles podrían ser los problemas con este método de evaluación?

### Método 3

#### Reporte de Donantes: “Comparando manzanas con manzanas”

“En los últimos días el programa de capacitación del gobierno ha recibido mucha cobertura de prensa. Algunos han afirmado que el programa tiene un impacto enorme, mientras que otros argumentan que los impactos son más bien moderados. Este reporte busca proveer una medida más correcta del impacto del programa usando un método más apropiado. Los análisis anteriores han usado métricas erróneas para calcular el impacto del programa, posiblemente sobrestimando el efecto del programa en el incremento de las tasas de empleo. Por ejemplo, se han comparado las tasas de empleo de las mujeres que participaron en el programa y las que no participaron, lo cual podría introducir un sesgo de selección ya que estos dos grupos pueden ser muy diferentes por muchas razones, y no sólo por haber o no participado en el programa.

Lo que se necesita para obtener una estimación más adecuada de impacto es comparar cambios en las tasas de empleo de ambos grupos. De esta manera, podremos ver por cuánto cambiaron estas tasas para cada grupo. Cuando repetimos este análisis usando esta métrica, vemos que las mujeres que participaron en el programa experimentaron un incremento de 5,85% en sus tasas de empleo, lo que demuestra que el programa sí incrementa las tasas de empleo, pero no por las magnitudes que antes se creía”.

### Tema de Discusión 3

1. ¿Qué método de evaluación propone este artículo del periódico?
2. ¿Cómo intenta este método estimar el contrafactual?
3. ¿Cuál es el supuesto que se tiene que cumplir para que el contrafactual esté bien estimado?
4. ¿Cuáles podrían ser los problemas con este método de evaluación?

## Método 4

### Los números no mienten siempre y cuando el investigador no esté durmiendo

“Durante las últimas semanas, el público colombiano ha recibido información contradictoria sobre el impacto del programa de capacitación del gobierno colombiano. Aquellos que apoyan el programa afirman que el programa de capacitación provee a mujeres jóvenes habilidades que necesitan para encontrar empleo, lo que incrementa su probabilidad de estar empleadas. Otros, sin embargo, creen que este impacto está inflado y que el impacto verdadero del programa es mucho más modesto.

Desafortunadamente, ambos lados están usando métodos equivocados para medir el impacto del programa y la pregunta de si el programa de capacitación incrementó la probabilidad de tener empleo entre las mujeres que participaron sigue sin ser contestada.

Por lo tanto, este reporte utiliza un método estadístico sofisticado para medir el verdadero impacto del programa. Nuestra mayor preocupación eran otras variables que pudieran estar confundiendo resultados previos. Por lo tanto, realizamos una encuesta para recolectar información sobre edad, estado civil, niveles de educación y la ciudad donde viven las participantes. Todas estas variables pueden potencialmente afectar la empleabilidad de una persona, así que nuestro análisis controla por estas variables, lo que nos permite identificar el verdadero impacto del programa.

Como podemos observar en la Tabla 1, los resultados cambian y nuestra estimación de impacto cae cuando controlamos por variables adicionales. Los resultados de la columna (1) sugieren que el programa incrementó la tasa de empleo por 6,5%, significativa al nivel del 10%. Si vemos la columna (2), la cual incluye las variables de control, el impacto es disminuido a 5,7%, significativa al nivel del 10%. Es importante notar que tanto las variables estado civil y ciudad son estadísticamente significantes también.

Al controlar por estas variables que pueden afectar la tasa de empleo, descubrimos que el impacto efectivo del programa de capacitación es modesto. Aunque este incremento sugiere que la capacitación no es una panacea para el desempleo juvenil, es un incremento que puede hacer una gran diferencia en la vida de muchos”.

## Análisis de una tabla de regresión

**Variable de interés (independiente):**  
Queremos averiguar el impacto de la capacitación sobre la probabilidad de empleo.

**Variable a explicar (o dependiente):**  
Nos interesa ver el impacto sobre la probabilidad de empleo.

**Modelo o especificación:** En este modelo (2), la regresión incluye la variable de interés y las variables independientes.

Tabla 1: Probabilidad de Empleo

	(1)	(2)
Capacitación	0,065 * (0,022)	0,057* (0,022)
Edad		0,004 (0,005)
Estado Civil		-0,066* (0,026)
Nivel Educativo		0,007 (0,006)
Ciudad		-0,036*** (0,005)
Constante	0,64 ** (0,02)	0,63 ** (0,14)

**Significancia estadística:** Por medio de la regresión, estamos tratando de inferir un coeficiente en la población total (o de interés) a partir de una muestra. Las estrellitas señalan coeficientes que son creíbles. Por convención dos estrellitas señala que el coeficiente es estadísticamente significativo al 5%. Esto indica, con un 95% de confianza, que el coeficiente estimado es distinto de cero.

#### Tema de Discusión 4

1. ¿Qué método de evaluación se propone en este artículo del periódico?
2. ¿Cómo intenta este método estimar el contrafactual?
3. ¿Cuál es el supuesto que se tiene que cumplir para que el contrafactual esté bien estimado?
4. ¿Cuáles podrían ser los problemas con este método de evaluación?

Nota: Los artículos en este caso no son reales – fueron creados para este caso. Los datos sí son de una evaluación real.

	Metodología	Descripción	¿Quién es el grupo de comparación?	Supuestos de la metodología	Datos requeridos
Métodos Cuasi	Pre-Post	Mide variación en resultados de participantes en el tiempo. La diferencia entre la situación anterior a un tratamiento y la situación posterior a este.	Los mismos participantes del programa justo antes de aplicarse el programa- representan el grupo de comparación.	Se asume que el programa es el único factor que influyó cualquier cambio en el resultado obtenido.	Datos de los participantes antes y después del programa.
	Diferencia Simple	Mide las diferencias entre aquellos que participaron en el programa y otros que no lo hicieron, después de completado el programa.	Los individuos que no participaron en el programa (por alguna razón), pero para los cuales tenemos datos después de completado el programa.	Los no-participantes en promedio son idénticos a los participantes excepto por la participación en el programa, y tenían la misma probabilidad de participar en el programa antes de que comenzara.	Datos de los participantes y no participantes para después del programa.
	Diferencias en Diferencias	Compara la variación en resultados a lo largo del tiempo de los participantes con la de los que no participaron del programa.	Los individuos que no participaron en el programa (por alguna razón), pero de quienes se recolectó información tanto antes como después del programa. Hay que buscar idealmente no participantes que sean "similares" a los participantes.	Asume que si el programa no existiera, los dos grupos tendrían idénticas trayectorias a lo largo de este periodo.	Datos de ambos participantes, antes y después del programa.
	Regresión Multivariada	Los individuos que recibieron el programa son comparados con los que no recibieron, y otros factores que pueden explicar las diferencias de resultados son tomados en consideración (controlados).	Los individuos que no participan en el programa (por cualquier motivo), pero de quienes se recolectó la información antes y/o después del programa. En este caso la información no esta comprometida sólo con indicadores de resultados, sino también por otras variables explicativas.	Los factores que fueron excluidos (porque eran no observables y/o que no fueron medidos) no sesgan los resultados porque no se correlacionan con los resultados o no difieren entre los que participan y los que no (al menos una de estas condiciones debe cumplirse para que no sesguen la estimación).	Los resultados así como las "variables control" para los que participan y los que no.
	"Matching"	Los individuos del grupo tratado son comparados con individuos similares en el grupo de comparación.	<u>Matching exacto</u> : Para cada participante, se escoge al menos un no-participante que es idéntico en las características seleccionadas. <u>Propensity score matching</u> : los que no participan tienen una serie de características muy parecidas a los participantes, lo que predice que hubiesen podido ser parte del programa. Se busca que el promedio ponderado de las características de los individuos de cada par sean similares.	Los no-participantes en promedio son idénticos a los participantes excepto por la participación en el programa, y tenían la misma probabilidad de participar en el programa antes de que comenzara.	Resultados, así como también "variables de coincidencia o matching" para ambos grupos, los que participan y los que no.
	Diseño de Regresión Discontinua	Los individuos son clasificados en base a criterios medibles específicos. Hay una línea de corte que determina si un individuo es elegible o no. Los participantes que se encuentran en la vecindad del límite después son comparados entre si y el criterio de elegibilidad es controlado.	Los individuos que están cerca de la línea de corte, pero que caen en el lado "equivocado", y por tanto no pueden participar del programa.	Después de controlar el criterio (y otras medidas de decisión), las diferencias restantes entre los individuos directamente debajo de la línea de corte o por encima de ella no son estadísticamente significativas y no producirán sesgo en los resultados. Un requerimiento necesario y suficiente para mantener este supuesto es que la línea de corte es un criterio estricto de adherencia.	Resultados, así como también todas las medidas para el criterio de selección (y cualquier otra medida de control).
	Variables instrumentales	La participación puede ser predicha por un factor accidental (aleatorio o casi aleatorio), o variable "instrumental", que no está correlacionada con el resultado, de manera distinta al hecho de que predice la participación (y la participación afecta el resultado).	Individuos que, por su cercanía al factor aleatorio, se predice que no participaran y (posiblemente como resultado) no participan.	Si la variable instrumental no tuviese la habilidad de predecir la participación, este "instrumento" no tendría efecto sobre o no estaría correlacionado con el resultado.	Requiere información del resultado, el "instrumento", y otras variables de control.
Método experimental	Evaluación Aleatoria	Método experimental para medir relaciones causales entre dos variables comparando participantes y no participantes cuando la participación fue determinada por sorteo.	La participación en el programa es asignada de forma aleatoria a un grupo de potenciales participantes elegibles mayor al número de cupos para el programa	La Aleatorización "funciona", es decir, los dos grupos son estadísticamente idénticos (en factores observable y no observables)	Requiere información posterior para los grupos tratamiento y control. El tamaño de muestra debe ser lo suficientemente grande para detectar un resultado significativo.

## Caso 3: Programa de Profesores Adicionales

### Diseñando una evaluación para responder tres preguntas claves sobre políticas educativas

Este estudio se basa en la publicación: "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya", de Esther Duflo (MIT), Pascaline Dupas (Stanford) y Michael Kremer (Harvard).

En 2005, la ONG International Child Support Africa (ICS) se ofreció para apoyar el sistema escolar de Kenia occidental mediante la contratación de profesores en 190 escuelas primarias para poder confrontar el problema de escuelas con demasiados estudiantes y con escasez de profesores. En su programa de dos años, ICS proporcionó fondos a estas escuelas para que cada una contratara un profesor adicional. A diferencia de los funcionarios civiles contratados por el Ministerio de Educación, los profesores del programa son contratados localmente por comités escolares. ICS esperaba que este programa mejorara el aprendizaje de los estudiantes mediante la disminución del tamaño de las clases y la utilización de profesores que rindieran cuentas directamente a las comunidades a las que les prestan servicios. Sin embargo, los profesores adicionales tienden a tener menos capacitación y reciben un salario mensual inferior al de los funcionarios civiles. Por esta razón, existía preocupación acerca de si estos profesores estaban lo suficientemente motivados, dada su remuneración, o calificados, dadas sus credenciales.

¿Qué diseños experimentales podrían comprobar el impacto de esta intervención en el rendimiento educativo? ¿Cuál de estos cambios en el sistema escolar puede ser el principal responsable de un mejor desempeño de los estudiantes?

#### Escuelas con exceso de alumnos

Al igual que muchos otros países en vías de desarrollo, Kenia ha realizado rápidos progresos hacia el cumplimiento del objetivo de educación primaria universal de los Objetivos de Desarrollo del Milenio. Después de la eliminación de las tarifas escolares en 2003, la matrícula en las escuelas primarias aumentó casi un 30 por ciento, de 5,9 millones a 7,6 millones entre 2002 y 2005.

Sin embargo, sin una inversión correspondiente por parte del gobierno, este progreso ha dado origen a un conjunto de nuevos desafíos en Kenia:

### **1) Tamaño de clases demasiado grandes:**

Debido a restricciones presupuestarias, el aumento en la matrícula escolar no ha coincidido con un aumento proporcional en el número de profesores. Esto ha resultado en clases muy grandes, particularmente en los grados menores. Por ejemplo, en una muestra de escuelas en Kenia occidental, las clases de primer grado tenían un promedio de 83 estudiantes en 2005. Esto es preocupante porque típicamente se cree que las clases pequeñas son más importantes para los estudiantes más jóvenes, quienes aún se están acostumbrando al entorno escolar. El Sindicato Nacional de Profesores de Kenia calcula que el país necesita 60.000 profesores de escuela primaria adicionales, además de los 175.000 ya existentes para servir a todos los estudiantes de escuela primaria y reducir los tamaños de las clases.

### **2) Ausentismo de los profesores:**

Exacerbando aún más el problema de la proporción de profesores-alumnos, las tasas de ausentismo de los profesores es muy alta: cerca del 20% en algunas áreas de Kenia.

Normalmente no existen reemplazos para los profesores ausentes, de modo que los estudiantes simplemente dan vueltas, se van a casa o asisten a otra clase, por lo general de otro grado. Las escuelas pequeñas, que prevalecen en áreas rurales de los países en vías de desarrollo, pueden tener que cerrarse completamente por la ausencia de un profesor. Las familias deben considerar si la escuela estará abierta al decidir si envían o no a sus hijos a la escuela. Un resultado obvio es la baja asistencia de los estudiantes, incluso en los días en que la escuela está abierta.

### **3) Clases heterogéneas:**

Las clases en Kenia también son muy heterogéneas, con estudiantes que varían ampliamente en términos de preparación escolar y el apoyo que reciben en el hogar.

Agrupar a los estudiantes en clases clasificadas por aptitud (subdivisión o tracking) es una idea controversial entre los académicos y los encargados de políticas públicas. Por un lado, si los profesores encuentran más fácil enseñar a un grupo homogéneo de estudiantes, la subdivisión podría mejorar la efectividad de las escuelas y los puntajes en los exámenes. Por otra parte, si los estudiantes aprenden en parte de sus compañeros, la subdivisión podría convertirse en una desventaja para los estudiantes con malos resultados y un beneficio para los que tienen buenos resultados, lo que exacerbaría la desigualdad.

#### **4) Escasez de materiales para la escuela:**

Debido a los altos costos de materiales educativos y al creciente número de estudiantes, los recursos educacionales se han quedado cortos, y en algunos casos hasta cuatro estudiantes deben compartir un solo libro. Además, la ya sobrecargada infraestructura se deteriora más rápido cuando se ve obligada a servir a más niños.

#### **5) Bajos índices de graduación:**

Como resultado de estos factores, las tasas de graduación de la escuela primaria en Kenia son muy bajas, con sólo 45,1% de los niños y 43,3% de las niñas graduándose de primaria.

Considerando todos estos problemas, se presentan nuevos desafíos para las comunidades: ¿Cómo garantizar la calidad mínima de la educación dadas las restricciones de presupuesto de Kenia?

#### **¿Qué son los Profesores Adicionales?**

Los gobiernos de varios países en vías de desarrollo han respondido a desafíos similares llenando las vacantes de profesores con profesores por contrato locales que no son funcionarios civiles. Las cuatro características principales de estos profesores adicionales son:

- (1) Se contratan con renovación anual, sin garantía de renovación del empleo (a diferencia de lo que ocurre con los profesores tradicionales);
- (2) Con frecuencia son menos calificados que los profesores regulares y es menos probable que tengan un certificado o grado de capacitación de profesor;
- (3) Reciben salarios menores que los de los profesores regulares (normalmente reciben menos de una quinta parte de los salarios que se pagan a los profesores);
- (4) Es más probable que sean del área local donde se encuentra la escuela.

#### **¿Son efectivos los profesores adicionales?**

El incremento del uso de los profesores adicionales ha sido una de las innovaciones más importantes en las políticas públicas que buscan entregar educación primaria en los países en vías de desarrollo. Sin embargo, también ha sido una idea altamente controversial.

Los partidarios señalan que contratar profesores adicionales es una forma eficiente de ampliar el acceso a la educación y mejorar su calidad para una gran cantidad de alumnos que conforman una primera generación de estudiantes. Al saber que la recontractación al año siguiente depende de una decisión del comité de la escuela, el cual se basa en el desempeño del profesor, es posible que los profesores adicionales estén más motivados

que sus colegas con cargos gubernamentales. Además, los profesores adicionales provienen de un contexto similar a sus estudiantes en cuanto a su cultura y situación socioeconómica, permitiéndoles entender mejor la situación de sus estudiantes.

Los oponentes argumentan que al utilizar profesores no capacitados y sin las certificaciones adecuadas se puede suplir la falta de personal en las salas de clases, pero no producir resultados positivos de aprendizaje. Además, argumentan que el uso de profesores adicionales resta profesionalismo a la enseñanza, reduce el prestigio de toda la profesión y reduce la motivación de todos los profesores. Incluso si ayuda en el corto plazo, puede perjudicar los esfuerzos para reclutar a profesores calificados en el futuro.

Aunque el uso de los profesores adicionales ha sido controversial, existe muy poca evidencia respecto a la efectividad de estos profesores para mejorar los resultados de aprendizaje de los estudiantes.

#### La evaluación experimental del Programa de Profesores Adicionales

En enero de 2005, *International Child Support Africa* comenzó un programa de dos años para examinar el efecto de los profesores adicionales en la educación en Kenia. Bajo el programa, ICS entregó fondos a 190 comités de escuela locales para que contrataran a un profesor adicional que enseñara una clase de primer grado adicional. El propósito de esta intervención era abordar tres desafíos importantes: el tamaño de la clase, la rendición de cuentas del profesor y la heterogeneidad de las aptitudes. La evaluación se diseñó para medir el impacto de la reducción del tamaño de las clases, la efectividad relativa de los profesores adicionales y el impacto de una subdivisión de las clases por aptitudes o tracking (es decir, unas clases con los estudiantes de bajos rendimientos y otras con aquéllos que tienen buenos rendimientos).

Abordar varias preguntas de investigación a través de un diseño experimental

Se pueden utilizar diferentes estrategias de aleatorización para responder diferentes preguntas.

**Tema de discusión 1:**

1. ¿Qué estrategia de aleatorización se podría utilizar para evaluar las siguientes preguntas? ¿Quiénes estarían en el grupo de tratamiento y en el grupo de comparación? ¿Cómo serían asignados aleatoriamente a estos grupos?
2. ¿Cuál es la efectividad relativa de los profesores adicionales en comparación con la de los profesores tradicionales del gobierno?
3. ¿Cuál es el efecto de las clases más pequeñas en el desempeño de los estudiantes?
4. ¿Cuál es el impacto de agrupar a los estudiantes según sus aptitudes en el desempeño de los estudiantes?





## Ejercicios

### Ejercicio 1: La mecánica de la aleatorización y el balance estadístico.

#### Definiciones

**Balance estadístico:** Se refiere al balance entre grupos de tratamiento y control en variables de interés. En una evaluación aleatoria ambos grupos deben ser comparables.

Aleatorizar garantiza que en promedio tengamos muestras estadísticamente idénticas o balanceadas.

**Aleatorización simple:** Aleatorizar a una población elegible en dos grupos idénticos sin restricciones.

**Aleatorización estratificada:** Una aleatorización simple dentro de diferentes estratos.

\*Fuentes:

3ie (2011), 3ie Glosario para la Evaluación de Impacto. Iniciativa Internacional para la Evaluación de Impacto. Nueva Delhi, India.

Gertler, Paul, Sebastián Martínez, Patrick Premand, Laura B. Rawlings, Christel M. J. Vermeersch. La evaluación de impacto en la práctica. Banco Mundial. Washington, DC: 2010.

En este ejercicio revisaremos el proceso de asignación aleatoria de unidades de análisis a grupo de control y grupo de tratamiento. Utilizaremos el programa Microsoft Excel. Luego de realizar la asignación, revisaremos los resultados de esas asignaciones, con el objetivo de entender aspectos claves de la asignación aleatoria de un programa:

- ¿Logra la asignación aleatoria generar grupos de tratamiento y de control balanceados? ¿En qué sentido están balanceados?
- ¿Cómo influye el tamaño de muestra en el balance entre grupos?
- ¿Cómo influye la estratificación de la asignación aleatoria en el balance?

Para estos fines, contaremos con una muestra de 2.000 escuelas de 10 municipios de Guatemala, las que asignaremos aleatoriamente a los grupos de tratamiento (T) y de control (C). Con el objetivo de responder las preguntas planteadas, haremos la asignación aleatoria para 4 casos:

- Caso 1: Aleatorización Simple con muestra de 20 escuelas.
- Caso 2: Aleatorización Simple con muestra de 200 escuelas.
- Caso 3: Aleatorización Estratificada con muestra de 200 escuelas.
- Caso 4: Aleatorización Simple con muestra de 2.000 escuelas.

Y para cada caso veremos:

- Si hay balance en términos de promedios de cada grupo. La variable que usaremos para ello será el puntaje promedio de las escuelas en una prueba.
- Si las distribuciones de cada grupo se asimilan a las distribuciones de la población de escuelas.

### **Parte 1: Aleatorización Simple.**

Excel cuenta con la función generación de número aleatorio, que es la que usaremos para este ejercicio. Repita el siguiente proceso para las primeras 3 hojas del archivo \*.xls correspondiente a este ejercicio ("Simple N=20", "Simple N=200" y "Simple N=2000").

#### **(1) ¿Cómo asignar el programa?**

Se debe fijar una regla para asignar el programa. En este caso, generaremos un número aleatorio y en base a este número asignaremos escuelas a tratamiento y control:

- Control: La mitad de las escuelas con el número aleatorio más pequeño.
- Tratamiento: La mitad de las escuela con el número aleatorio más grande.

## (2) Tenemos toda la lista de escuelas.

Para cada escuela, la planilla muestra su código, su nombre, su municipio y el puntaje promedio de sus alumnos en una prueba.

The screenshot shows an Excel spreadsheet with the following data:

	B	C	D	E	F	G
	Nombre de escuela	Municipalidad	Puntaje Prueb	Número Aleatori	Orden Aleatorio	Asignación
2	Escuela 801	Tarrazú	64		#N/A	#N/A
3	Escuela 402	Desamparados	48		#N/A	#N/A
4	Escuela 1802	Curridabat	59		#N/A	#N/A
5	Escuela 1001	Mora	54		#N/A	#N/A
6	Escuela 1202	Goicoechea	60		#N/A	#N/A
7	Escuela 802	Tarrazú	46		#N/A	#N/A
8	Escuela 1201	Goicoechea	40		#N/A	#N/A
9	Escuela 1002	Mora	78		#N/A	#N/A
10	Escuela 1401	Santa Ana	39		#N/A	#N/A
11	Escuela 202	Escazú	76		#N/A	#N/A
12	Escuela 602	Puriscal	46		#N/A	#N/A
13	Escuela 601	Puriscal	70		#N/A	#N/A
14	Escuela 2	San José	30		#N/A	#N/A
15	Escuela 201	Escazú	30		#N/A	#N/A
16	Escuela 1	San José	66		#N/A	#N/A
17	Escuela 1801	Curridabat	48		#N/A	#N/A
18	Escuela 401	Desamparados	80		#N/A	#N/A
19	Escuela 1402	Santa Ana	64		#N/A	#N/A
20	Escuela 1602	Alajuelita	56		#N/A	#N/A
21	Escuela 1601	Alajuelita	30		#N/A	#N/A

## (3) Asigne un número aleatorio a cada escuela:

La función **ALEATORIO()**, o **RAND()** si tiene Excel en inglés, es el generador de número aleatorio de Excel. Para usarlo, en la Columna E, escriba lo siguiente: **=ALEATORIO()**. Luego, copie y pegue esa función en toda la columna.

The screenshot shows the same Excel spreadsheet as above, but with the formula **=aleatorio()** entered in cell E2. The formula bar at the top shows the function being applied.

	B	C	D	E	F	G
	Nombre de escuela	Municipalidad	Puntaje Prueb	Número Aleatori	Orden Aleatorio	Asignación
2	Escuela 801	Tarrazú	64	=aleatorio()	#N/A	#N/A
3	Escuela 402	Desamparados	48		#N/A	#N/A
4	Escuela 1802	Curridabat	59		#N/A	#N/A
5	Escuela 1001	Mora	54		#N/A	#N/A
6	Escuela 1202	Goicoechea	60		#N/A	#N/A
7	Escuela 802	Tarrazú	46		#N/A	#N/A
8	Escuela 1201	Goicoechea	40		#N/A	#N/A
9	Escuela 1002	Mora	78		#N/A	#N/A
10	Escuela 1401	Santa Ana	39		#N/A	#N/A
11	Escuela 202	Escazú	76		#N/A	#N/A
12	Escuela 602	Puriscal	46		#N/A	#N/A
13	Escuela 601	Puriscal	70		#N/A	#N/A
14	Escuela 2	San José	30		#N/A	#N/A
15	Escuela 201	Escazú	30		#N/A	#N/A
16	Escuela 1	San José	66		#N/A	#N/A
17	Escuela 1801	Curridabat	48		#N/A	#N/A
18	Escuela 401	Desamparados	80		#N/A	#N/A
19	Escuela 1402	Santa Ana	64		#N/A	#N/A
20	Escuela 1602	Alajuelita	56		#N/A	#N/A
21	Escuela 1601	Alajuelita	30		#N/A	#N/A

Escribir = ALEATORIO() genera un número aleatorio de 15 dígitos entre 0 y 1 en la celda. La función =ALEATORIO() volverá a aleatorizar cada vez que usted realice cambios en cualquier otra parte de la planilla de cálculo. (Usted también puede inducir el re-cálculo y por lo tanto la re-aleatorización presionando la tecla F9.)

Esto no es útil para nuestros propósitos ya que una vez que hayamos generado nuestra columna de números aleatorios no necesitamos volver a aleatorizar. Para hacer que Excel deje de recalcular, usted puede reemplazar las "funciones" en esta columna con "valores".

Para hacer esto, seleccione todos los valores en la Columna E. Luego, haga click con el botón derecho en cualquier parte de la columna destacada y escoja la opción "copiar". Luego, haga click con el botón derecho en cualquier parte de esa columna y escoja la opción "Pegado Especial". Aparecerá la ventana de "Pegado Especial". Haga click en "Valores".

#### (4) Asigne a las escuelas a los grupos de Control y de Tratamiento:

De acuerdo a la regla de asignación ya definida en el paso (1), se ordena a las escuelas según su número aleatorio y se asigna a la mitad con números menores al grupo control y a la mitad restante al grupo tratamiento. Esto lo hacemos utilizando las funciones JERARQUIA y SI respectivamente. En inglés estas funciones son RANK e IF.

	B	C	D	E	F	G	H	K	L	M	N
1	Nombre de escuela	Municipalidad	Puntaje Prueb	Número Aleatori	Orden Aleatorio	Asignación					
2	Escuela 1802	Curridabat	59	0.887674067	2	C					
3	Escuela 1001	Mora	54	0.777721335	3	C					
4	Escuela 802	Tarrazú	46	0.658246937	5	C					
5	Escuela 1002	Mora	78	0.357498354	10	C					
6	Escuela 1401	Santa Ana	39	0.614865383	6	C					
7	Escuela 602	Puriscal	46	0.75917848	4	C					
8	Escuela 2	San José	30	0.897074784	1	C					
9	Escuela 201	Escazú	30	0.553310789	9	C					
10	Escuela 1801	Curridabat	48	0.561441733	7	C					
11	Escuela 1602	Alajuelita	56	0.554222449	8	C					
12	Escuela 801	Tarrazú	64	0.112516812	14	T					
13	Escuela 402	Desamparados	48	0.053717059	18	T					
14	Escuela 1202	Goicoechea	60	0.007161951	20	T					
15	Escuela 1201	Goicoechea	40	0.118155364	13	T					
16	Escuela 202	Escazú	76	0.15092466	12	T					
17	Escuela 601	Puriscal	70	0.067704812	16	T					
18	Escuela 1	San José	66	0.037807573	19	T					
19	Escuela 401	Desamparados	80	0.289952216	11	T					
20	Escuela 1402	Santa Ana	64	0.109403909	15	T					
21	Escuela 1601	Alajuelita	30	0.058447026	17	T					

Utilizando la opción DATOS>ORDENAR, ordene los datos de diversas formas para observar cómo las distintas escuelas fueron asignadas aleatoriamente a los grupos de tratamiento y control.

## Parte 2: Aleatorización Estratificada.

Una aleatorización estratificada de un programa no es muy diferente a una aleatorización simple. De hecho, una aleatorización estratificada no es más que una serie de aleatorizaciones simples, una al interior de cada estrato. En este caso, diremos que los estratos son cada uno de los municipios que tenemos en la muestra. ¿Cómo se podría apoyar esta decisión?

En cuanto a los pasos a seguir, trabaje sobre la hoja 4 del archivo Excel correspondiente, llamada "Estratificado N=200", y lleve a cabo los pasos (1), (2) y (3) que ya realizó para los casos anteriores. Una vez hecho eso, continúe con el paso (4):

### (4) Asigne a las escuelas a los grupos de Control y de Tratamiento:

De acuerdo a la regla de asignación ya definida en el paso (1), al interior de cada estrato, es decir, al interior de cada comuna, se ordena a las escuelas según su número aleatorio y se asigna a la mitad con números menores al grupo control y a la mitad restante al grupo tratamiento. En esta muestra para Alajuelita, tendremos a 10 escuelas en el grupo tratamiento y a 10 escuelas en el grupo control y para Curridabat, 2 y 2 respectivamente. Esto lo hacemos, al igual que para el caso de la aleatorización simple, utilizando las funciones =JERARQUIA() y =SI() respectivamente.

	A	B	C	D	E	F	G
1	Código de Escuela	Nombre de escuela	Municipalidad	Puntaje Prueba	Número Aleatorio	Orden Aleatorio	Asignación
2	1601	Escuela 1601	Alajuelita	40		#N/A	#N/A
3	1602	Escuela 1602	Alajuelita	30		#N/A	#N/A
4	1603	Escuela 1603	Alajuelita	28		#N/A	#N/A
5	1604	Escuela 1604	Alajuelita	75		#N/A	#N/A
6	1605	Escuela 1605	Alajuelita	75		#N/A	#N/A
7	1606	Escuela 1606	Alajuelita	80		#N/A	#N/A
8	1607	Escuela 1607	Alajuelita	25		#N/A	#N/A
9	1608	Escuela 1608	Alajuelita	35		#N/A	#N/A
10	1609	Escuela 1609	Alajuelita	70		#N/A	#N/A
11	1610	Escuela 1610	Alajuelita	41		#N/A	#N/A
12	1611	Escuela 1611	Alajuelita	45		#N/A	#N/A
13	1612	Escuela 1612	Alajuelita	81		#N/A	#N/A
14	1613	Escuela 1613	Alajuelita	68		#N/A	#N/A
15	1614	Escuela 1614	Alajuelita	47		#N/A	#N/A
16	1615	Escuela 1615	Alajuelita	62		#N/A	#N/A
17	1616	Escuela 1616	Alajuelita	29		#N/A	#N/A
18	1617	Escuela 1617	Alajuelita	64		#N/A	#N/A
19	1618	Escuela 1618	Alajuelita	61		#N/A	#N/A
20	1619	Escuela 1619	Alajuelita	48		#N/A	#N/A
21	1620	Escuela 1620	Alajuelita	27		#N/A	#N/A
22	1801	Escuela 1801	Curridabat	88		#N/A	#N/A
23	1802	Escuela 1802	Curridabat	25		#N/A	#N/A
24	1803	Escuela 1803	Curridabat	64		#N/A	#N/A
25	1804	Escuela 1804	Curridabat	15		#N/A	#N/A
26	1805	Escuela 1805	Curridabat	60		#N/A	#N/A
27							
28							
29							

Utilizando la opción DATOS>ORDENAR, ordene los datos por las variables "Municipalidad" y "Orden Aleatorio" para así poder observar la forma en que el programa ha sido asignado al interior de cada municipio.

### Parte 3: Revisando el Balance entre Grupos.

¿A qué nos referimos con balance? Fundamentalmente a que ambos grupos sean estadísticamente similares. En esta parte del ejercicio vamos a revisar si los grupos construidos en las partes previas están balanceados o no. Y ello lo vamos a hacer de dos formas: revisando sus promedios y las diferencias en dicha dimensión entre grupos, y revisando sus distribuciones.

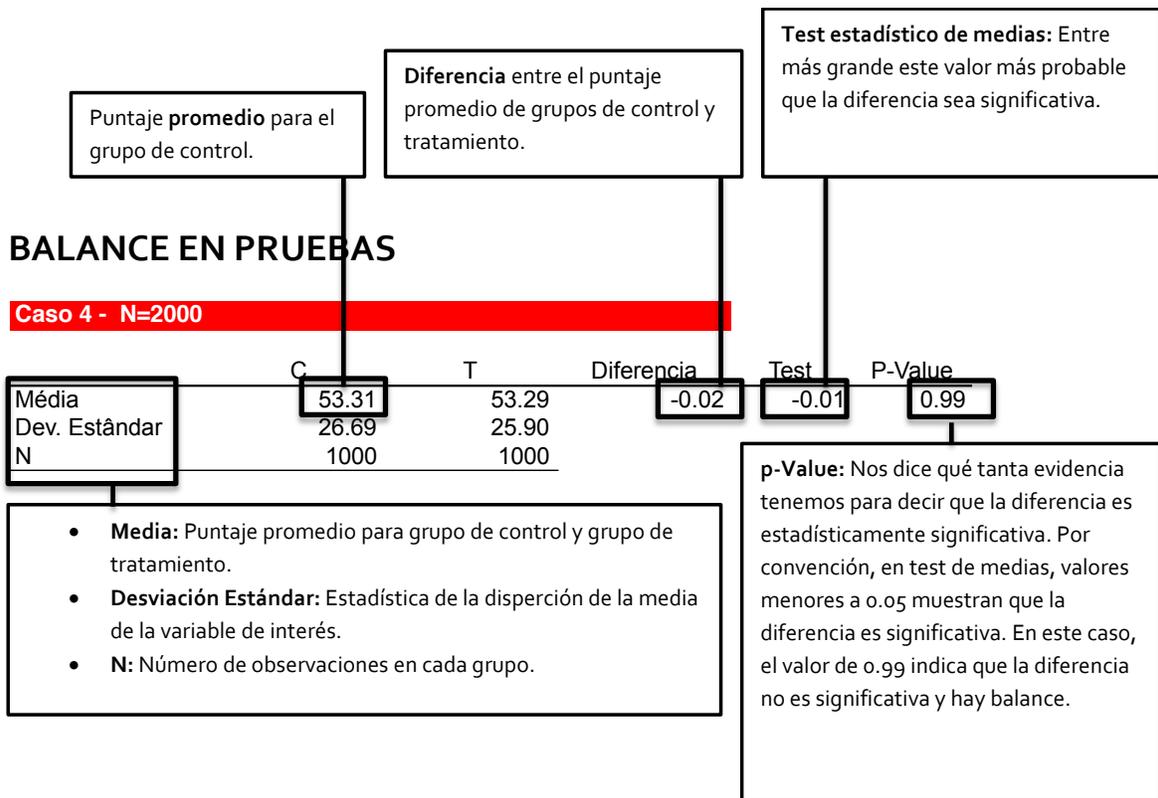
En la hoja 5 del archivo Excel, llamada "Balance", se calculan diversos atributos de las muestras que se han construido aleatoriamente en cada uno de los casos. En primer lugar, se calculan el promedio, la desviación estándar de los puntajes y el tamaño de los grupos control y tratamiento en cada caso. A partir de esos cálculos, hacen tests de "diferencia de medias" que constituyen una medida estadística respecto de la significancia de tales diferencias entre dos muestras. Al hacer una evaluación de impacto, se espera que ambos grupos sean estadísticamente igual antes de ser tratados. ¿Se cumple eso en este caso? ¿Para qué casos se cumple?

En la hoja 6 del archivo Excel, llamada "Gráficos", se ordenan los puntajes promedio de los grupos tratamiento y control de cada caso en intervalos, con el objetivo de graficar la distribución de los mismos, de compararlas entre sí, y de compararlas con la de la población de escuelas. ¿Qué patrón se observa al comparar las distintas aleatorizaciones llevadas a cabo?

### Parte 4: ¿Qué aprendemos?

En términos generales, este ejercicio nos permite aprender:

- Cómo asignar aleatoriamente un programa.
- Cómo cambios en el tamaño de muestra afectan los resultados de tal asignación.
- Cómo hacer una aleatorización estratificada puede influir en los resultados de la asignación aleatoria de un programa.





## Ejercicio 2 - Cálculos del Tamaño Muestral

### Definiciones

**Cálculos de poder:** El cálculo de la muestra requerida para la evaluación de impacto, el cual depende de: el tamaño de efecto mínimo y el nivel de confianza requerido.

**Poder:** La probabilidad que, si el programa tiene un efecto, lo podamos distinguir de cero dado un tamaño de muestra.

**Significancia:** La probabilidad de que el efecto que hemos medido no haya ocurrido por chance. Utilizamos tests estadísticos para determinar si un grupo (tratamiento) es diferente de otro (control) con respecto a una (o varias) variables de interés.

**Desviación estándar:** Una medida estandarizada de la variación en una población muestra con respecto a la media de una variable.

**Efecto estandarizado:** Una medida estandarizada de la magnitud esperada del efecto de un programa.

**Clúster:** El nivel de observación que nos permite medir el tamaño de muestra. Un clúster contiene varias observaciones. Por lo general, observaciones altamente correlacionadas deberían ser medidas al nivel de clúster.

**Correlación Intra-clúster:** Una medida de la correlación entre observaciones en un clúster.

**Tamaño de efecto mínimo:** El tamaño mínimo de efecto que el investigador considera necesario detectar en la evaluación de impacto. Empleado para desarrollar el cálculo de potencia necesario para determinar el tamaño de muestra requerido.

\*Fuentes:

3ie (2011), 3ie Glosario para la Evaluación de Impacto. Iniciativa Internacional para la Evaluación de Impacto. Nueva Delhi, India.

Gertler, Paul, Sebastián Martínez, Patrick Premand, Laura B. Rawlings, Christel M. J. Vermeersch. La evaluación de impacto en la práctica. Banco Mundial. Washington, DC: 2010.

El estudio de caso del Programa “Extra Teacher” (ETP) discutía el concepto de ensayos aleatorios por conglomerado. El ejemplo Balsakhi usado en la clase anterior introdujo el concepto de cálculos de poder. En este último estábamos interesados en medir el efecto de un tratamiento (balsakhis en las salas de clase) sobre los resultados medidos a nivel individual (puntajes de prueba por niño). Sin embargo, la aleatorización de Balsakhis se realizó a nivel de sala de clase. Podría ser que nuestro resultado de interés está correlacionado con los estudiantes en la misma sala de clase por razones que no tienen conexión con el Balsakhi. Por ejemplo, todos los estudiantes en la sala de clase serán afectados por su profesor, por si su sala de clase está inusualmente oscura o si tienen pizarrón para tiza; estos factores significan que cuando a un estudiante en la sala le va particularmente bien por esta razón, a todos los estudiantes en esa sala probablemente les iría mejor también, lo cual podría no tener conexión con un Balsakhi.

Por lo tanto, si hacemos una muestra de 100 niños de 10 escuelas escogidas aleatoriamente, esa muestra es menos representativa de la población de escuelas en la ciudad que si seleccionásemos 100 niños al azar de toda la población de escuelas y, por lo tanto, absorbe menos varianza (Esto es algo que ya aprendimos en el ejercicio 1). De hecho, tenemos un tamaño muestral más pequeño de lo que pensamos. Esto llevará a tener más ruido en nuestra muestra y, por lo tanto, un error estándar mayor que en la caso común de muestreo independiente. Necesitamos considerar esto al momento de planificar tanto el tamaño de la muestra como la mejor manera de hacer una muestra de las salas de clases.

Este ejercicio le ayudará a comprender cómo hacer eso. ¿Debiera uno tener una muestra con muchos estudiantes en pocas escuelas? ¿O será preferible tener una muestra de pocos estudiantes distribuidos en muchas escuelas? ¿Cómo decidir?

Trabajaremos estas preguntas determinando el tamaño de la muestra que nos permita detectar un efecto específico con al menos un poder de 80%. Recuerde que el poder es la probabilidad de que cuando el tratamiento tiene un efecto, usted podrá distinguirlo de cero en su muestra.

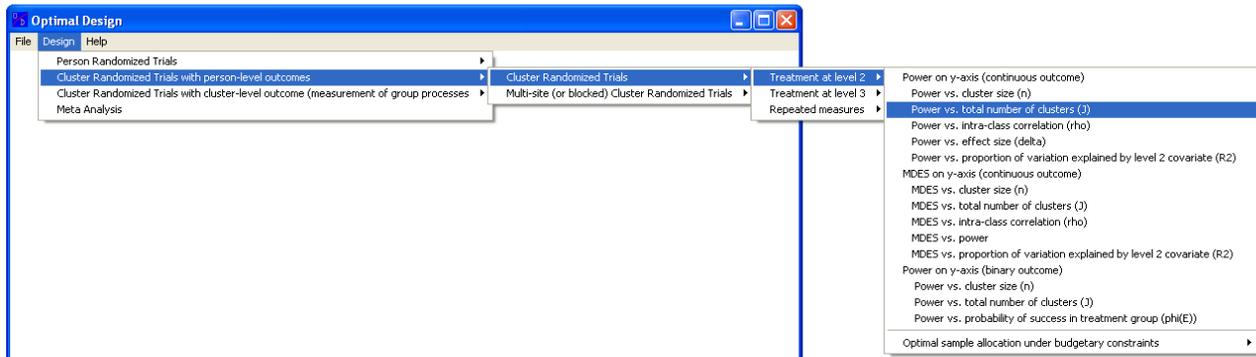
Usaremos un software desarrollado por Steve Raudenbush con fondos de la Fundación William T. Grant. Puede encontrar recursos adicionales sobre diseños agregados en su página Web.

## Parte 1: Usar el Software OD

Primero baje el Software OD del sitio web (se encuentra disponible también el manual del software):

[http://sitemaker.umich.edu/group-based/optimal\\_design\\_software](http://sitemaker.umich.edu/group-based/optimal_design_software)

Cuando lo abra, verá una pantalla que se parece a la de más abajo. Seleccione la opción de menú "Design" para ver el menú principal. Seleccione la opción "Cluster Randomized Trials with person-level outcomes," "Cluster Randomized Trials," y luego "Treatment at level 2". Usted verá varias opciones para generar gráficos, escoja "Power vs. Total number of clusters (J)."



Aparecerá una nueva ventana:



Seleccione  $\alpha$  (alfa). Verá que ya está programada en 0.050 para un nivel de significancia de 5%.

Primero supongamos que deseamos probar 40 estudiantes por escuela. ¿Cuántas escuelas necesita visitar para contar con una respuesta estadísticamente significativa? Presione  $n$  que representa el número de estudiantes por escuela. Ya que estamos testeando solo 40 estudiantes por escuela, sólo complete  $n(1)$  con 40 y presione aceptar.

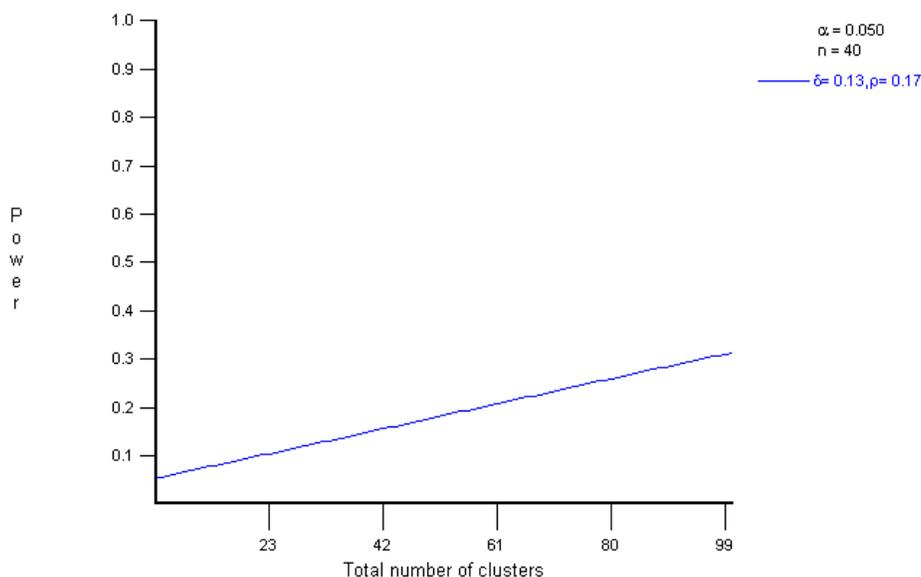
Ahora tenemos que determinar  $\delta$  (delta), el tamaño del efecto que se quiere medir, medido en desviaciones estándar de la variable de interés (este es el MDE). Asumamos que estamos interesados en detectar si existe un aumento de 10% en los puntajes de pruebas (O con mayor precisión, no estamos interesados en detectar menos de 10%) Nuestra encuesta de línea base indicó que el puntaje promedio de prueba es 26 con una

desviación estándar de 20. Deseamos detectar un tamaño de efecto de 10% de 26, el cual es 2,6. Dividimos 2,6 por la desviación estándar para obtener  $\delta$  igual a 2,6/20 o 0,13.

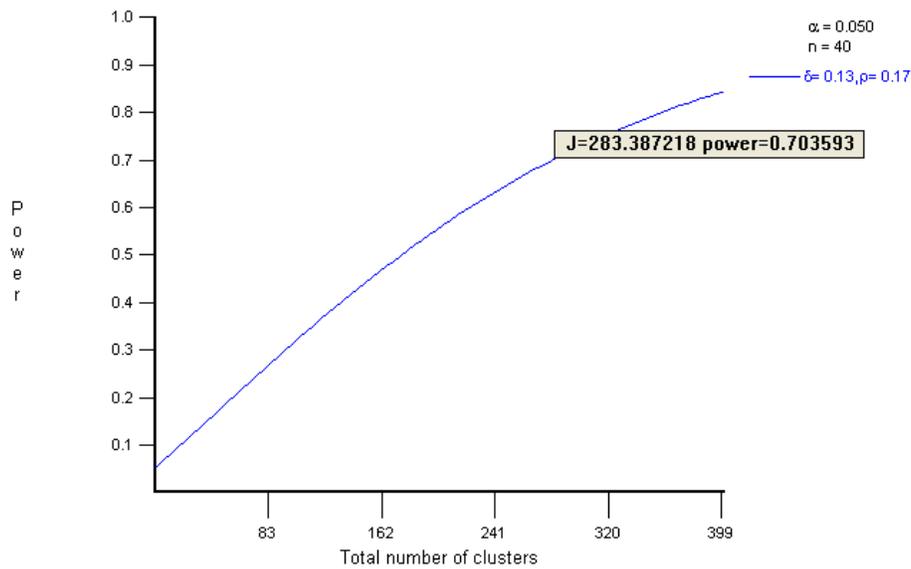
Seleccione  $\delta$  del menú. En el cuadro de diálogo que aparece hay un valor preestablecido de 0,200 para delta(1). Cambie el valor a 0,13 y cambie el valor de delta (2) a vacío. Presione aceptar.

Finalmente necesitamos escoger  $\rho$  (rho), que es la correlación intra-conglomerado.  $\rho$  nos dice que tan fuertemente están correlacionados los resultados para las unidades dentro del mismo conglomerado. Si los estudiantes de la misma escuela fueran clones (sin variación) y todos obtuvieran puntajes iguales en la prueba, entonces  $\rho$  sería igual a 1. Si, por otro lado, los estudiantes de las mismas escuelas son de hecho independientes – y no existiese diferencias entre las escuelas, entonces  $\rho$  será igual a 0.

Usted ha determinado en su estudio piloto que  $\rho$  es 0,17. Complete rho(1) y deje rho (2) vacío. Usted debiera ver un gráfico parecido al que aparece más abajo.



Usted notará que su eje x no es lo suficiente largo para permitirle ver qué números de conglomerado le daría un poder de 80%, que lo que generalmente se usa. Haga click en  para fijar su eje x en un máximo de 400. Luego, puede hacer click en el gráfico para ver el poder exacto y números de conglomerado para un punto en particular.



Ejercicio 2.1:

¿Cuántas escuelas se necesitan para alcanzar un poder de 80%? ¿de 90%?

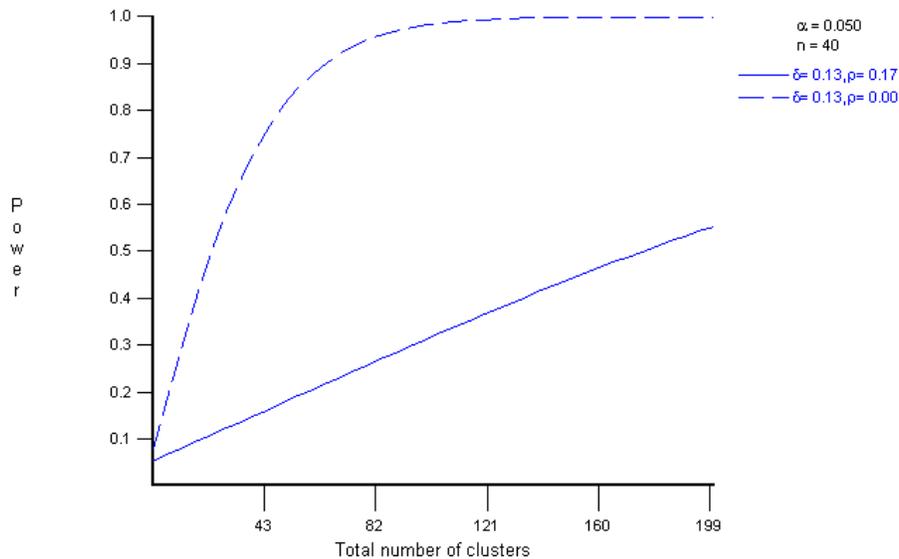
Ahora usted ha visto cuántos conglomerados se necesita para un poder de 80%, haciendo una muestra de 40 estudiantes por escuela. Suponga, en cambio, que usted sólo tiene la capacidad de llegar a 124 escuelas (este es el número real que se utilizó en el programa Balsakhi).

Ejercicio 2.2:

¿Cuántos niños por escuela se necesitan para alcanzar un poder de 80%? ¿de 90%?  
Escoja diferentes valores para n para ver cómo cambia su gráfico.

Finalmente, veamos cómo cambia el poder de una muestra dado del Coeficiente de Correlación Intra-Conglomerado ( $\rho$ ). Deje  $\rho(1)$  en 0,17, pero para la comparación cambie  $\rho(2)$  a 0,0.

Usted debería ver un gráfico como el siguiente. La curva sólida azul es la que tiene los parámetros que usted ha establecido basado en su estimaciones previas al testeo del efecto de las reservas de mujeres respecto del agua potable. La curva segmentada azul es para propósitos de comparación – para ver cuánto poder obtendría de su muestra si  $\rho$  fuese cero. Mire cuidadosamente el gráfico.



Ejercicio 2.3:

¿Cómo cambia el poder de la muestra con el Coeficiente de Correlación Intra-conglomerado ( $\rho$ )?

## Parte 2: Aplicando lo Aprendido

Ya sabemos utilizar OD para hacer cálculos de tamaño muestral. Procederemos a aplicarlos a la propuesta que el grupo está desarrollando con el objetivo de determinar el tamaño de muestra que sería razonable para la evaluación.

Comenzamos por determinar los parámetros relevantes para nuestros cálculos:

### Ejercicio 2.4:

Discuta y determine valores para los siguientes aspectos:

1. ¿Qué nivel de confianza estadística usamos para nuestros cálculos?
2. ¿Qué nivel de poder estadístico usamos para nuestros cálculos?
3. ¿Qué impacto esperamos que tenga nuestro programa?
4. ¿A qué nivel se asignará nuestro programa?



## Preguntas para el diagnóstico de evaluabilidad

Es importante aclarar que no es cierto, que toda intervención necesite o amerite una evaluación de impacto. No todas las preguntas de políticas públicas, pueden ser esclarecidas utilizando el método experimental. Ante un programa de interés, habrá entonces que verificar, que el mismo cuente con ciertas características que permitan realizar sobre él, un análisis de impacto experimental<sup>1</sup>. A continuación, algunas preguntas que ayudarán a definir si una evaluación puede ser realizada utilizando métodos experimentales:

- ¿Cuál sería el costo total de una evaluación experimental para el programa? ¿Son proporcionados los costos, relativos al tamaño, escala y presupuesto del programa o intervención?
- ¿El programa está maduro o estable en su gestión y métodos?
- ¿Podemos diseñar protocolos de evaluación que no incidan en el funcionamiento normal del programa o de la intervención?
- ¿Qué tipo de preguntas importantes sobre el programa responderá la evaluación? ¿De qué manera? ¿Cómo contribuirá esto al conocimiento de los beneficiarios y de los implementadores y de los hacedores de política pública?
- ¿Podríamos replicar el programa a mayor escala en otros contextos?
- ¿Beneficiará a suficientes personas?

---

<sup>1</sup> Que un programa sea evaluable de acuerdo a nuestro criterio no significa que éste sea mejor o peor que otros.



## Términos Importantes

**Análisis de regresión:** Un método estadístico que determina la asociación entre las variables dependientes y una o más variables independientes.

**Asignación aleatoria:** Un diseño de intervención en el cual los miembros de la población elegible son asignados aleatoriamente bien sea al grupo de tratamiento o al de control (i.e. asignación aleatoria). Es decir, el que alguien esté en el grupo de tratamiento o en el de control es un asunto de azar exclusivamente y no una función de cualquiera de sus características (observadas o no).

**Cálculos de poder:** El cálculo de la muestra requerida para la evaluación de impacto, el cual depende de: el tamaño de efecto mínimo y el nivel de confianza requerido.

**Contaminación:** Cuando los miembros del grupo de control se ven afectados bien sea por la intervención (ver efectos secundarios) u otra intervención que afecte el resultado esperado. La contaminación es un problema común pues existen múltiples intervenciones de desarrollo en la mayoría de las comunidades.

**Contrafactual:** Corresponde al valor que hubiera tenido la variable de resultado para los individuos del grupo de tratamiento en el caso hipotético de no haber recibido la intervención. El contrafactual no puede ser observado por lo que debe ser inferido de un grupo de comparación o grupo de control.

**Correspondencia, Pareamiento (o Matching):** Un método empleado para crear grupos de control, en los cuales los grupos o individuos son hechos corresponder a aquellos en el grupo de tratamiento con base en las características que se consideran relevantes para el resultado o resultados de la intervención.

**Correspondencia de puntaje de propensión (Propensity Score Matching (PSM)):** Un diseño cuasi-experimental para calcular el impacto de una intervención. Los resultados del grupo de tratamiento se comparan con aquellos para un grupo de control, en donde este último se construye a lo largo de correspondencias establecidas sobre puntajes de propensión. Los puntajes de propensión son la probabilidad de participar en la intervención, tal y como son dados por una ecuación probit sobre características observadas. Dichas características no deben verse afectadas por la intervención.

**Desgaste o Atrición:** Bien sea el retiro de los participantes del grupo de tratamiento durante la intervención, o la imposibilidad de recolectar información de una unidad en rondas subsiguientes en un sondeo de información. Cualquiera de las anteriores formas de desgaste puede resultar en estimaciones de impacto sesgadas.

**Diferencia única:** La comparación en el resultado para el grupo de control después de la intervención con el valor de su línea de base (también conocido como "antes versus después" ), o bien una comparación *ex post* en el resultado entre los grupos de tratamiento y control.

**Diseño de evaluación *ex ante*:** Un diseño para una evaluación de impacto preparado antes de que tenga lugar la intervención. Los diseños *ex ante* son más fuertes que los diseños *ex post* por su capacidad para considerar la asignación aleatoria así como la recolección de información de base tanto del grupo de tratamiento como del grupo de control. También conocido como evaluación prospectiva.

**Diseño de evaluación *ex post*:** Un diseño de una evaluación de impacto preparado una vez que la intervención haya comenzado y posiblemente haya sido completada. Salvo que haya habido una asignación aleatoria, entonces debe emplearse un diseño cuasiexperimental.

**Diseño por Regresión Discontinua:** Un diseño de evaluación de impacto en el cual los grupos de tratamiento y control se identifican como aquellos que se encuentran justo a cualquier lado de un valor de umbral de una variable. Esta variable puede ser un puntaje o una característica observada (por ejemplo edad o propiedad raíz) empleada por el personal del programa para determinar la población elegible, o puede ser una variable encontrada para distinguir a los participantes de los no participantes por medio del análisis de datos.

**Diferencia en Diferencia:** La diferencia en el cambio del resultado observado en el grupo de tratamiento comparado con el cambio de observado en el grupo de control; o de manera equivalente, el cambio en la diferencia en el resultado entre tratamiento y control. La doble diferenciación evita el sesgo en la selección que resulta de elementos inobservables invariables en el tiempo. También conocido como diferencia en diferencia. Comparar con diferencia simple y diferencia triple.

**Efecto promedio de tratamiento:** El valor promedio del impacto sobre el grupo de beneficiarios (o sobre el grupo de tratamiento). Ver también intención de tratar y tratamiento sobre los tratados.

**Efectos secundarios o indirectos:** Cuando la intervención tiene un impacto (bien sea positivo o negativo) sobre unidades que se encuentran fuera del grupo de tratamiento. Hacer caso omiso de los efectos secundarios puede resultar en una estimación sesgada del impacto. De existir efectos secundarios, entonces el grupo de beneficiarios es mayor que el de participantes. Cuando los efectos secundarios afecta a miembros del grupo de control, este es un caso especial de contagio.

**Error de muestreo:** El error que se produce cuando se emplean estimaciones para producir información a partir de una muestra en lugar de a partir de la población entera.

**Factores de confusión (*Confounding factors*):** Otras variables o determinantes que afectan el resultado esperado.

**Grupo de comparación o control:** Un grupo de individuos cuyas características son similares a las de los grupos de tratamiento (o participantes) pero que no reciben la intervención.

**Grupo de tratamiento:** El grupo de personas, compañías, instituciones, etc. quienes reciben la intervención. También conocidos como participantes.

**Hipótesis:** una proposición específica acerca de la relación entre dos variables. En una evaluación de impacto, las hipótesis generalmente se plantean como una relación causal entre la intervención y las distintas variables de resultado.

**Impacto:** se define como la diferencia entre lo que ocurrió con el programa y lo que habría ocurrido sin el programa en la población de interés.

**Indicadores:** Métricas usadas para cuantificar o medir las dimensiones que un programa busca afectar, así como la forma en que se implementa un programa.

**Línea base:** Levantamiento de información que se realiza antes de dar inicio a la intervención. Los datos de línea base pueden servir para mejorar la precisión de la estimación de impacto y deberían ser recolectados tanto del grupo de tratamiento como del de control.

**Nivel de aleatorización:** El nivel de agregación (ej. individual, hogar, escuela, aldea) en el cual se realiza la asignación aleatoria a grupos de tratamiento y de control.

**No observables:** Características que no pueden ser observadas ni medidas. La presencia de inobservables puede generar sesgo en la selección en diseños cuasi-experimentales si dichas variables inobservables están correlacionadas con la participación en el programa y la variable resultado.

**Intención de tratar:** El efecto promedio del tratamiento calculado sobre la totalidad del grupo de tratamiento, sin importar si de hecho participaron en la intervención o no. Compárese con tratamiento de los tratados.

**Teoría de cambio:** Describe la forma en que la implementación de un programa se traduce en resultados, por medio de una cadena de relaciones causales. Un enfoque basado en teoría de cambio examina las presunciones que subyacen a las hipótesis de una evaluación de impacto.

**Muestra:** Un subconjunto de la población estudiada. La muestra se obtiene al azar del marco de muestreo. En una muestra aleatoria simple todos los elementos del marco tienen una probabilidad igual de ser seleccionados, pero normalmente se emplean los diseños de muestreo más complejos, los cuales requieren el uso de peso de las muestras en el análisis.

**Prueba aleatoria controlada:** Un diseño para la evaluación de impacto en el cual se emplea la asignación aleatoria para adjudicar la intervención entre miembros de la población elegible. Dado que no debería haber ninguna correlación entre las características de los participantes y el resultado, y que las diferencias entre el resultado entre los grupos de tratamiento y control pueden atribuirse enteramente a la intervención, es decir, que no hay sesgo en la selección. Sin embargo, estas pruebas pueden estar sujetas a diferentes tipos de sesgo y por tanto requieren de protocolos estrictos.

**Sesgo de selección:** Sesgo potencial introducido dentro de un estudio, debido a la selección de diferentes tipos de personas dentro de los grupos de tratamiento y control. Como resultado, las diferencias en los resultados pueden ser explicadas potencialmente como el resultado de las diferencias pre-existentes entre los grupos y no por el tratamiento mismo.

**Tamaño de efecto mínimo:** El tamaño mínimo de efecto que el investigador considera necesario detectar en la evaluación de impacto. Empleado para desarrollar el cálculo de potencia necesario para determinar el tamaño de muestra requerido.

**Tamaño del efecto:** El tamaño de la relación entre dos variables (particularmente entre variables del programa y los resultados).

**Tratamiento de los tratados:** La estimación de tratamiento de los tratados es el impacto (efecto promedio del tratamiento) sólo para aquellos que efectivamente fueron objeto de la intervención. Compárese con intención de tratar.

**Validez externa:** El grado al que los resultados de la evaluación de impacto pueden aplicarse a otros momentos o lugares.

**Validez interna:** La validez del diseño de evaluación, es decir si gestiona adecuadamente temas tales como selección de muestras (para minimizar el sesgo en la selección), efectos secundarios, contagio, y heterogeneidad del impacto.

**Variable dependiente:** Una variable que se cree que puede ser predicha por o causada por una o más variables distintas (variables independientes). El término se utiliza comúnmente en el análisis de regresión.

**Variable independiente:** Una variable que se cree causa cambios sobre la variable dependiente. Usualmente se aplica en el análisis de regresión.

Basado en el glosario para la evaluación de impacto de [Jie](#) (2011), [Jie](#) Glosario para la Evaluación de Impacto. Iniciativa Internacional para la Evaluación de Impacto. Nueva Delhi, India.



## Apuntes











## Mapas

