



# CRS Muestras

GUÍA PARA EL CÁLCULO DEL TAMAÑO DE LAS MUESTRAS CUANTITATIVAS

Esta guía fue escrita por Stacy Prieto, PhD. Se basa en gran medida en *Going beyond simple sample size calculations: a practitioner's guide* de Brendon McConnell y Marcos Vera-Hernández, lo que la hace más accesible para el personal de CRS y aplicable al contexto operativo de CRS.

Foto de portada por Katie Price

©2025 Catholic Relief Services. Todos los derechos reservados. 21MK-328766MT

Este documento está protegido por derechos de autor y no puede ser reproducido total o parcialmente sin autorización. Póngase en contacto con [stacy.prieto@crs.org](mailto:stacy.prieto@crs.org) para obtener autorización. Cualquier "uso justo" bajo la ley de derechos de los EE. UU. debe contener la referencia apropiada a Catholic Relief Services.



# Tabla de contenidos

Tabla de contenidos.....	i
Lista de figuras.....	iii
Lista de tablas.....	iv
Agradecimientos.....	v
Abreviaturas.....	vi
Introducción.....	1
Objetivo.....	1
Cuándo usar esta guía.....	1
Cómo está organizada esta guía.....	2
Información necesaria antes de usar esta guía.....	2
1. Árbol de decisión del tamaño de la muestra.....	4
2. Las ocho ecuaciones principales.....	6
2.1 Múltiples Grupos de Estudio.....	6
2.2 Las ecuaciones.....	6
2.2.1 Comparaciones entre grupos.....	6
2.2.2 Grupos individuales.....	11
2.2.3 Resumen.....	12
3. Revisión requerida: Elementos que aumentarán el tamaño de la muestra.....	14
3.1 Pérdida de datos.....	14
3.2 Disminución.....	14
3.3 Indicadores Específicos para Subpoblaciones.....	14
3.4 Grandes cambios desde la línea de base hasta la línea final.....	15
3.5 Resumen.....	15
4. Otras consideraciones.....	16
4.1 El factor de corrección por población finita (FPC).....	16
4.2 Establecer objetivos alcanzables y ser conscientes de los costes asociados a la recopilación de datos.....	16
4.3 Estratificación.....	17
4.4 Utilizar el tratamiento desigual y el tamaño de los grupos de control con indicadores binarios.....	18
4.5 Uso de conjuntos de datos de panel.....	18
4.6 Resumen.....	19
5. Mitos sobre el tamaño de la muestra.....	20
5.1 El tamaño de la muestra depende del tamaño de la población subyacente.....	20
5.2 Los indicadores binarios requieren tamaños de muestra más grandes.....	20
5.3 Resumen.....	21
6. Diseño de la muestra y análisis.....	22
6.1 Selección de la muestra.....	22

6.1.1 Utilizar muestras aleatorias y documentar cualquier sesgo muestral debido a un muestreo no aleatorio.....	22
6.1.2 La selección de grupos de población proporcional al tamaño (PPT) puede no ser apropiada en el contexto de CRS.....	23
6.1.3 Resumen.....	24
<b>6.2 Uso de tamaños de muestra en el análisis de datos .....</b>	<b>24</b>
6.2.1 Ponderación de la muestra: cálculo .....	24
6.2.2 Pesos de la muestra - uso.....	25
6.2.3 Medias/proporciones ponderadas y totales .....	25
6.2.4 Ponderaciones de la falta de respuesta .....	26
6.2.5 Muestras agrupadas o estratificadas y análisis de regresión .....	27
6.2.6 Intervalos de confianza .....	27
6.2.7 FPC.....	28
6.2.8 Resumen.....	28
<b>Apéndice 1. Coeficiente de correlación intraclase .....</b>	<b>29</b>
<b>A1.1 Efecto de diseño frente a CCI .....</b>	<b>29</b>
<b>A1.2 Cálculo del CCI .....</b>	<b>29</b>
A1.2.1 Varianza general.....	30
A1.2.2 CCI para indicadores continuos.....	30
A1.2.3 CCI para indicadores binarios.....	31
<b>A1.3 Cálculo de la desviación estándar .....</b>	<b>31</b>
<b>A1.4 Valores CCI y desviación típica para determinados indicadores .....</b>	<b>31</b>
<b>Apéndice 2. Calculadora de tamaño de muestra – Excel.....</b>	<b>33</b>
<b>Apéndice 3. Calculadora de tamaño de muestra – R .....</b>	<b>34</b>
<b>Apéndice 4. Descripciones formales de los cálculos del tamaño de la muestra .....</b>	<b>37</b>
<b>Apéndice 5. Guía de referencia rápida .....</b>	<b>39</b>
<b>Confidencial - Apéndice 6. Otras guías de tamaño de muestra .....</b>	<b>41</b>
<b>Bibliografía .....</b>	<b>42</b>

# Lista de figuras

**FIGURA 1. ÁRBOL DE DECISIÓN DEL TAMAÑO DE LA MUESTRA..... 4**

# Lista de tablas

<b>Tabla 1. Ejemplo de rendimiento de zanahoria en Honduras.....</b>	<b>8</b>
<b>Tabla 2. Ejemplo de SILC en Sierra Leona / Burkina Faso .....</b>	<b>9</b>
<b>Tabla 3. Ejemplo de profesor de Burkina Faso.....</b>	<b>11</b>
<b>Tabla 4. Datos de ejemplo de peso de muestra .....</b>	<b>26</b>
<b>Tabla 5. Ejemplo de presentación de tamaño de muestra .....</b>	<b>38</b>
<b>Tabla 6. Ejemplo de lista preparatoria para cálculos.....</b>	<b>40</b>

# Agradecimientos

El autor desea agradecer a Benjamin Allen, James Campbell, Tony Castleman, Heather Dolphin, John Hembling y TD Jose de CRS, Stephanie Martin de TANGO International, Eric Gerber de Northeastern University y Tim Smith de la Oficina del Censo de los Estados Unidos por sus contribuciones al contenido de esta guía.

Esta guía está dedicada al personal de Monitoreo, Evaluación, Rendición de Cuentas y Aprendizaje (MEAL, por sus siglas en inglés) de CRS que trabaja incansablemente para mejorar su conocimiento y capacidad para implementar prácticas MEAL globales de alta calidad.

# Abreviaturas

ALNP	Alimentación del Lactante y del Niño Pequeño
BHA	Oficina de Asistencia Humanitaria de USAID
CCI	Coefficiente de correlación intraclase
CI	Intervalo de confianza
DMA	Dieta Mínima Aceptable
EMD	Efecto Mínimo Detectable
FPC	Factor de corrección por población finita (FPC)
FTF	Iniciativa Alimentar el Futuro de USAID
IMC	Índice de Masa Corporal
IPTT	Tabla de seguimiento del rendimiento de los indicadores
MEAL	Control, Evaluación, Rendición de Cuentas y Aprendizaje
MIRA	Indicadores de medición para el análisis de la resistencia
ONG	Organización no gubernamental
PPT	Probabilidad proporcional al tamaño
SD	Desviación estándar
SE	Error estándar
SILC	Grupo de Autoahorro y Préstamo
USAID	Agencia de los Estados Unidos para el Desarrollo Internacional
USDA	Departamento de Agricultura de los Estados Unidos

# Introducción

CRS, y la comunidad internacional de desarrollo en general, se ha vuelto cada vez más rigurosa en su evaluación cuantitativa de los proyectos de desarrollo. CRS, en sus Políticas y Procedimientos internos de Monitoreo, Evaluación, Rendición de Cuentas y Aprendizaje (MEAL, por sus siglas en inglés), requiere estudios de línea de base y evaluaciones finales para todos los proyectos de valor superior a \$1 millón (Catholic Relief Services 2023). Después de recopilar los datos de la evaluación final, los valores de la tabla de seguimiento del rendimiento de los indicadores (IPTT) para los indicadores a nivel de resultados deben compararse con sus respectivos valores de referencia, para saber si las medias<sup>1</sup> o las proporciones de los indicadores difieren significativamente entre los dos períodos de tiempo.

La detección de una diferencia estadísticamente significativa entre dos medias o proporciones requiere: 1) Que realmente exista una diferencia y, 2) Que la muestra representativa a partir de la cual se calculan las medias o proporciones sea suficientemente grande.

Además, los datos de muchos indicadores anuales de seguimiento del desempeño se recopilan de una muestra de beneficiarios del proyecto, y esa muestra debe ser de tamaño suficiente para obtener una estimación razonablemente precisa del valor del indicador entre todos los beneficiarios del proyecto.

Esta guía resume las consideraciones clave necesarias para determinar el tamaño de las muestras, con ejemplos específicos del trabajo de CRS. Su objetivo es llenar un vacío en el conocimiento de la agencia en torno a los cálculos del tamaño de la muestra, reducir la confusión en torno a qué ecuaciones son apropiadas para su uso en qué contextos y proporcionar un punto de referencia coherente para toda la agencia.

## Objetivo

Esta guía debe utilizarse durante la etapa de diseño del proyecto para asignar recursos suficientes en el presupuesto del proyecto para la recopilación de datos y para revisar las necesidades de recopilación de datos durante la puesta en marcha y ejecución del proyecto. Esta guía incluye ecuaciones de tamaño de muestra que calculan un tamaño de Efecto Mínimo Detectable (MDE), ya que los proyectos de CRS generalmente buscan detectar un cambio significativo entre dos puntos (línea de base y final).

En el Anexo 6 se examinan las guías sobre el tamaño de las muestras de los donantes y se destacan algunas diferencias fundamentales entre esas guías y ésta.

El público al que va dirigida esta guía son los líderes de propuestas de MEAL (para la preparación del presupuesto de MEAL); coordinadores MEAL del proyecto (para informar la recopilación de datos durante la puesta en marcha y la implementación); y los gerentes de MEAL del programa de país y los asesores técnicos de MEAL (para ayudarlos en su apoyo a lo anterior).

## Cuándo usar esta guía

El personal del proyecto debe calcular un tamaño de muestra para cada marco muestral, o tipo de encuestado, que encuestarán utilizando métodos cuantitativos. Por ejemplo, si un proyecto

<sup>1</sup> Esta guía utiliza el término "media" en todo momento porque es el término utilizado por matemáticos y estadísticos.

---

*Los destinatarios de esta guía son los líderes de la propuesta MEAL, los coordinadores MEAL del proyecto, los gerentes de MEAL del programa de país y los asesores técnicos de MEAL.*

va a encuestar a los agricultores para hacer un seguimiento de un indicador, a los grupos de productores para hacer un seguimiento de otro y a los niños menores de cinco años para hacer un seguimiento de otro, deben identificar la ecuación adecuada y calcular el tamaño de muestra necesario para cada uno.

---

*Si se van a recopilar datos para varios indicadores de un marco de muestra, calcule el tamaño de muestra necesario para cada indicador y, a continuación, elija el tamaño más grande calculado.*

Si los datos de varios indicadores se recopilarán de un marco de muestra, la mejor práctica es calcular el tamaño de muestra necesario para cada indicador y, a continuación, elegir el tamaño más grande calculado, dentro de lo razonable. Si el tamaño de la muestra más grande fuera exorbitantemente costoso de recolectar, concéntrese en 1) los indicadores más importantes, generalmente a nivel de resultados, o 2) los indicadores estándar de los donantes. Consulte también a un asesor técnico de MEAL, si es necesario.

Esta guía también se puede utilizar para estimar el tamaño de muestra necesario para los indicadores de nivel de salida. Sin embargo, los proyectos suelen recopilar datos sobre los indicadores de producto a través de registros de distribución o capacitación (seguimiento rutinario), y no una muestra representativa de los beneficiarios del proyecto.

## Cómo está organizada esta guía

Esta guía proporciona un árbol de decisión básico para determinar la ecuación adecuada que se debe utilizar al calcular el tamaño de las muestras. A continuación, tiene una sección obligatoria sobre consideraciones que pueden aumentar el tamaño de la muestra requerida, para garantizar que el tamaño final de la muestra sea adecuado. A esto le sigue una sección opcional sobre los factores que pueden disminuir el tamaño de la muestra. Se recomienda utilizar solo esta última sección si el tamaño de muestra previamente determinado es demasiado caro y es necesario reducirlo. La guía incluye ecuaciones que miden el cambio entre dos períodos de tiempo. También incluye ecuaciones para realizar una evaluación simple (y no detectar un cambio a lo largo del tiempo o entre grupos).

Aunque este documento solo pretende guiar los cálculos del tamaño de la muestra, hay una breve sección sobre los mitos sobre el tamaño de la muestra y cómo el diseño de la muestra afecta el análisis de datos.

La guía tiene seis anexos: 1) coeficientes de correlación intraclase; 2) una hoja de cálculo de Excel y 3) ejemplo de código R para cálculos más complejos; 4) lenguaje repetitivo para describir los cálculos del tamaño de la muestra en documentos escritos formales (por ejemplo, informes de donantes, propuestas de proyectos, etc.); 5) una guía de referencia rápida; y 6) una revisión de otras guías clave sobre el tamaño de la muestra.

## Información necesaria antes de usar esta guía

En esta guía se asume que se han realizado algunos pasos antes de utilizarla.

- 1) Los equipos del proyecto han decidido si realizarán comparaciones estadísticas entre grupos con los resultados. Por ejemplo, comparar a los beneficiarios del proyecto (grupo de tratamiento) con un grupo de control; comparación de los valores iniciales con los finales; comparar a los hombres con las mujeres; comparación de grupos de edad, etc.
- 2) Ya se ha elaborado la tabla de seguimiento del rendimiento de los indicadores del proyecto (IPTT) con valores estimados de línea de base y meta. Por lo general, el IPTT se estima en la etapa de diseño del proyecto, con una revisión documental que determina las estimaciones del valor de referencia de otros proyectos o estudios similares. Estos deben ser estudios

más recientes que cubran las mismas o similares áreas geográficas, estaciones y condiciones climáticas (por ejemplo, sequía / no sequía).

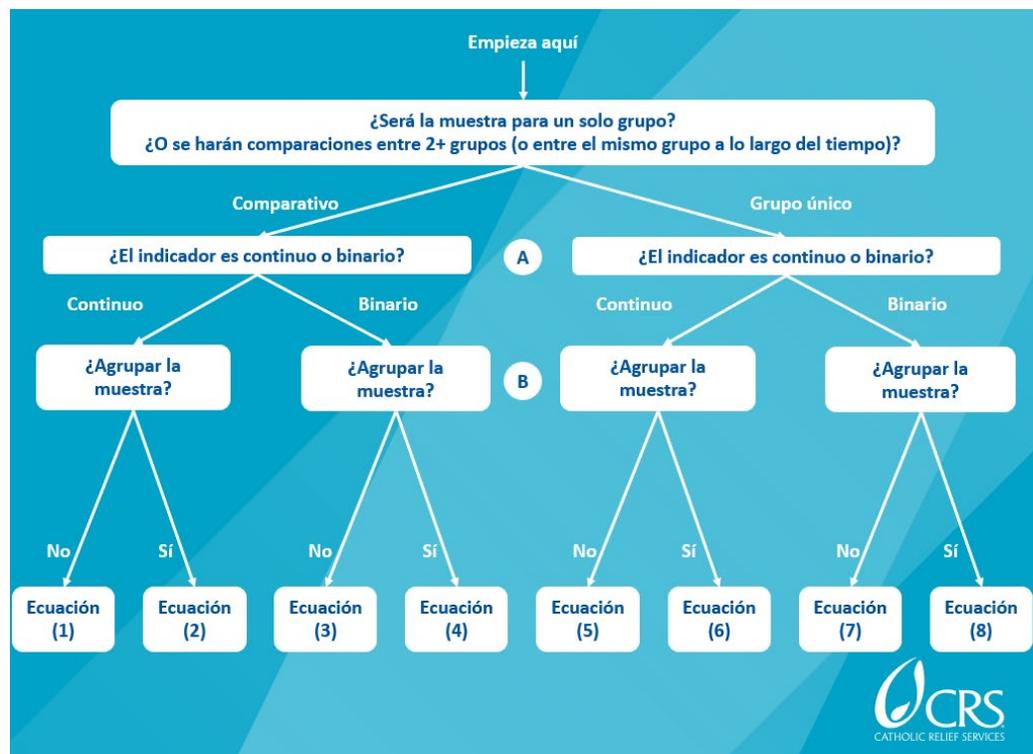
Después de revisar quién debe usar esta guía, cuándo usarla y recopilar la información necesaria, los equipos de proyecto ahora pueden revisar el árbol de decisión del tamaño de la muestra (sección 1), para determinar la ecuación necesaria para cada indicador.



Registro y entrega de asistencia en efectivo multipropósito en Brasil. [Felippe Thomaz]

# 1. Árbol de decisión del tamaño de la muestra

La Figura 1, el Árbol de Decisión del Tamaño de la Muestra, es una representación gráfica de la pregunta clave que debe responderse para cada indicador: "¿Qué ecuación de tamaño de la muestra deben usar los equipos de proyecto?" El árbol de decisión guía a los usuarios a la ecuación apropiada (hay ocho), en función de si compararán estadísticamente los datos recopilados entre los grupos; están recopilando datos para un indicador continuo o binario; y si están agrupando la muestra durante la recolección de datos.<sup>2</sup>



Los puntos (A) y (B) de la Figura 1 conducen a información adicional para ayudar en la toma de decisiones.

**FIGURA 1. ÁRBOL DE DECISIÓN DEL TAMAÑO DE LA MUESTRA**

Los puntos (A) y (B) de la Figura 1 conducen a información adicional para ayudar en la toma de decisiones. Con respecto a (A), los indicadores continuos y binarios siguen diferentes distribuciones de probabilidad, por lo que requieren diferentes ecuaciones de tamaño muestral:

(A) Los indicadores continuos, para fines de CRS, recopilan datos que tienen una amplia gama de valores posibles. Algunos ejemplos:

<sup>2</sup> En esta guía se utiliza el término "grupo" para describir lo que en la literatura estadística suele denominarse "conglomerado". El objetivo es mejorar la accesibilidad del lenguaje para el personal de CRS, aunque se reconoce que no es técnicamente correcto.

- Valor medio de las ventas anuales de explotaciones agrícolas y empresas<sup>3</sup>
- Número medio de hectáreas sometidas a prácticas o tecnologías de gestión mejoradas
- Rendimiento medio de los productos agrícolas seleccionados
- Volumen medio de mercancías vendidas por las explotaciones agrícolas y las empresas
- Tasa media de asistencia a clase
- Número medio de grupos de alimentos consumidos por un hogar
- Índice de capital social a nivel de los hogares

Los indicadores binarios recopilan datos como respuesta de sí/no. Algunos ejemplos:

- Porcentaje de personas u organizaciones que utilizan una práctica mejorada
- Porcentaje de estudiantes que saben leer
- Porcentaje de niños menores de 5 años que sufren retraso en el crecimiento
- Porcentaje de personas que acceden a financiamiento relacionado con la agricultura

(B) Agrupamiento significa seleccionar al azar una aldea, escuela, asociación de productores, etc. como el grupo, y luego seleccionar al azar a las personas dentro de ese grupo para la muestra. Por el contrario, el muestreo aleatorio simple significa tener una lista de todas las personas en todas las aldeas y seleccionar aleatoriamente a las personas de esta lista, sin seleccionar primero un grupo.

A menudo, las muestras se agrupan para ahorrar dinero cuando los encuestados se distribuyen en una amplia área geográfica. La agrupación de la muestra evita la necesidad de visitar cada pueblo, lo que podría ser más costoso.<sup>4</sup>

Tenga en cuenta que, con la agrupación, las ecuaciones de tamaño de las muestras indican tanto el número de grupos a encuestar como el número de individuos dentro de cada grupo. Es importante mantener el número de individuos por grupo lo más similar posible. Por ejemplo, si las escuelas en el área objetivo son pequeñas y no siempre tienen 15 estudiantes que cumplan con los criterios de selección, entonces no use 15 individuos/grupo en las ecuaciones de tamaño de la muestra.<sup>5</sup>

---

<sup>3</sup> Obsérvese que cuando se utiliza una muestra para informar de los valores de indicadores estándar como "Valor de las ventas, Número de hectáreas sometidas a prácticas de gestión mejoradas, número de personas que utilizan prácticas de gestión mejoradas, etc.", estos indicadores requieren extrapolarse a partir del resultado medio de una muestra. Por lo tanto, calcule el tamaño de muestra necesario para el valor medio de estos indicadores.

<sup>4</sup> La agrupación en clústeres, si bien reduce los costos de recopilación de datos, complica el análisis. Consulte la Sección 6.2 para obtener más información.

<sup>5</sup> Los diseños de evaluación utilizados por los profesionales del desarrollo utilizan con frecuencia métodos de muestreo de probabilidad proporcional al tamaño (PPS, por sus siglas en inglés) para aumentar la probabilidad de que se incluyan en la muestra grupos con poblaciones más grandes. Consulte la sección 6.1.2 sobre análisis de datos para conocer las limitaciones de la metodología PPS.

# 2. Las ocho ecuaciones principales

Las ocho ecuaciones del tamaño de la muestra se explican en la sección 2 y asumen que se selecciona aleatoriamente<sup>6</sup> una muestra representativa de la población.

## 2.1 Múltiples Grupos de Estudio

Antes de sumergirse en las ecuaciones, es importante tener en cuenta que las ocho ecuaciones proporcionan el tamaño de muestra necesario para cada grupo de comparación. Ejemplos de grupos de comparación son la línea de base y final, o el tratamiento y el control.<sup>7</sup> Si se recopilan datos al inicio del estudio para compararlos con los datos de la evaluación final, se debe aplicar en cada caso el número final proporcionado. Por lo tanto, si el tamaño de la muestra es de 100, se necesitan 100 observaciones al inicio y 100 al final. O 100 observaciones en el grupo de tratamiento y 100 en el grupo de control. Además, si se analizan los resultados entre estratos<sup>8</sup>, como las pruebas de diferencias estadísticas entre áreas geográficas o género, el número final proporcionado debe aplicarse a cada estrato.

## 2.2 Las ecuaciones

Esta sección guiará a los usuarios a través de las ecuaciones (1-8), explicará cada uno de los términos de cada ecuación y proporcionará un ejemplo práctico para cada ecuación. Tenga en cuenta que el apéndice 2 es una hoja de cálculo de Excel con las ecuaciones (1-8) ya programadas, para ayudar en los cálculos.

### 2.2.1 Comparaciones entre grupos

La referencia para las primeras cuatro ecuaciones es McConnell y Vera-Hernández (2015). Esta referencia es una buena opción para el contexto de CRS porque 1) proporciona las ecuaciones exactas necesarias, en lugar de hacer referencia a comandos empaquetados en software estadístico; 2) Deriva matemáticamente las ecuaciones del tamaño de la muestra y/o proporciona referencias para las derivaciones; 3) Proporciona ecuaciones de efecto mínimo detectable (MDE) tanto para indicadores binarios como continuos.

Las ecuaciones (1-2) son las mismas que las utilizadas en la guía Alimentar el futuro para evaluadores externos (Stukel 2018a), mientras que las ecuaciones (3-4) difieren un poco. Véase el apéndice 6 para una comparación más detallada de esta guía con la guía Alimentar el futuro.

Tenga en cuenta que si realiza comparaciones entre grupos:

- Para un indicador continuo recogido de una muestra no agrupada, utilice la ecuación (1)
- Para un indicador continuo de una muestra agrupada, utilice la ecuación (2)

<sup>6</sup> Las muestras no aleatorias, y su sesgo asociado, se abordan en la sección 6.1.1.

<sup>7</sup> Tenga en cuenta que una versión futura de esta guía puede hacer referencia a ecuaciones diferentes, probablemente más eficientes, a utilizar para determinar el tamaño total de la muestra y la asignación entre 2 o más grupos de tratamiento (Duflo, Glennerster y Kremer 2007).

<sup>8</sup> Véase la sección 4.3 al estratificar la muestra con fines organizativos (no para analizar las diferencias entre estratos). En este caso, no es necesario aumentar el tamaño de la muestra.

*Las ocho ecuaciones proporcionan el tamaño de muestra necesario para cada grupo de comparación.*

- Para un indicador binario de una muestra no agrupada, utilice la ecuación (3)
- Para un indicador binario de una muestra agrupada, utilice la ecuación (4)

Consulte la sección 1 si no está claro sobre los términos "continuo", "binario" o "clúster".

### 2.2.1.1 Ecuación (1)

Al hacer comparaciones entre grupos para un **indicador continuo recogido de una muestra no agrupada**, utilice la ecuación (1):

$$n^* = \frac{2(t_\beta + t_{\alpha/2})^2 SD^2}{\delta^2} \quad (1)$$

Donde

- $t_\beta$  (beta) es el valor crítico de la cola izquierda de la distribución t inversa<sup>9</sup> con ( $n^*-1$ ) grados de libertad.<sup>10</sup> Normalmente, el valor crítico de  $t_\beta$  elegido es del 80 % y representa la potencia de la muestra. Por lo tanto, existe un 20 % de probabilidad de no encontrar una diferencia con respecto a la intervención, a pesar de que exista (también conocido como error de tipo II). El valor crítico necesario puede obtenerse de una tabla t apropiada o utilizando el comando INV.T de Excel (como se utiliza en el apéndice 2).<sup>11</sup>
- $t_{\alpha/2}$  (alfa) es el valor crítico de dos colas de la distribución t inversa. Normalmente, el valor elegido de  $t_\alpha$  es el 5 % y representa el nivel de significación. Con este valor, hay un 5 % de posibilidades de rechazar la hipótesis nula (generalmente sin diferencia entre períodos de tiempo/grupos de comparación), cuando no debería rechazarse (también conocido como error de tipo I). El valor crítico necesario puede obtenerse a partir de una tabla t apropiada o utilizando el comando INV.T.2C de Excel (como se utiliza en el apéndice 2).<sup>12</sup>
- $SD$  es la desviación estándar del indicador. Si esto se desconoce, consulte el apéndice 1.4 para obtener una lista de valores posibles, así como una guía sobre cómo calcular la desviación estándar de los datos existentes. Obtención de datos de otros proyectos dentro del mismo programa de país, región o global; así como la realización de una revisión documental de las desviaciones estándar de indicadores similares de estudios anteriores puede ser útil.
- $\delta$  (delta) es el cambio objetivo en el indicador debido a la programación. Normalmente,  $\delta$  es la diferencia entre los valores objetivo de referencia y los valores objetivo de la vida útil del proyecto para el indicador.

Para trabajar con un ejemplo de la vida real, en la etapa inicial del proyecto, el Proyecto de Adquisición de Ayuda Alimentaria Local y Regional financiado por el Departamento de Agricultura de los Estados Unidos (USDA) en Honduras necesitaba estimar el tamaño de la muestra requerido para el indicador "Aumento medio del rendimiento para los participantes en el proyecto" que cultivan zanahorias. A través de la investigación de fuentes externas, el personal del programa estimó que el rendimiento medio de zanahoria era de 20.324 kg/ha, con una desviación estándar de

<sup>9</sup> Cuando se utiliza una desviación estándar de la muestra (y no la verdadera desviación estándar de toda la población, que se desconoce), se toman los valores críticos de la distribución t (y no de la distribución normal o z).

<sup>10</sup> Los grados de libertad son el número de observaciones utilizadas en el análisis. Dado que  $n^*$  aún es desconocido, realice algunas iteraciones, cambiando la estimación de  $n^*$ , hasta que el  $n^*$  utilizado para encontrar el valor crítico de la distribución t sea igual al  $n^*$  recomendado por la ecuación. Un ejemplo de este proceso iterativo se muestra en la hoja de cálculo del Apéndice 2.

<sup>11</sup> Si su interfaz Excel está en inglés, utilice la función T.INV.

<sup>12</sup> Si su interfaz Excel está en inglés, utilice la función T.INV.2T.

Utilice la ecuación (1) cuando realice comparaciones entre grupos para un indicador continuo recogido de una muestra no agrupada.

7.848 kg/ha (Lana 2012). Luego, el personal del programa examinó un rango de valores objetivo del proyecto, para ver cómo variaban los tamaños de muestra resultantes. El proyecto esperaba ver grandes aumentos en el rendimiento debido a la mejora de las técnicas, por lo que estimó que solo necesitarían encuestar a 21 agricultores.

$$n^* = \frac{2(0,88 + 2,26)^2 * 7.848^2}{(20.324 * 0,35)^2}$$

donde  $t_\beta = 0,88$  (a partir de 10 grados de libertad;  $t_{\alpha/2} = 2,26$ ;  $SD = 7.848$ ; y  $\delta = (20.324 * 0,35)$ . Sin embargo, para un cambio menor, como un aumento medio del 10 % en el rendimiento, el proyecto necesitaría encuestar a 236 agricultores. En resumen, para detectar cambios más pequeños con un indicador continuo, se necesitan tamaños de muestra más grandes. A menudo es útil elaborar un cuadro, como se hace en las hojas de cálculo del anexo 2, para examinar las distintas opciones. La Tabla 1 es un ejemplo de una tabla de este tipo

**TABLA 1. EJEMPLO DE RENDIMIENTO DE ZANAHORIA EN HONDURAS**

CAMBIO ( $\delta$ )	35%	30%	25%	20%	15%	10%
TAMAÑO TOTAL DE LA MUESTRA ( $n^*$ )	22	29	40	61	106	237

Con la Tabla 1 en la mano, el personal de MEAL puede ayudar a identificar metas alcanzables, para las cuales se pueden detectar cambios estadísticos con respecto a la línea de base, dentro de las limitaciones presupuestarias.

### 2.2.1.2 Ecuación (2)

Al hacer comparaciones entre grupos para un **indicador continuo recopilado de una muestra agrupada**, utilice la ecuación (2), que añade el término  $(1 + (m - 1)\rho)$  a la ecuación (1). Este término aumenta el tamaño de la muestra para compensar los efectos de la agrupación.

$$n^* = m^*k^* = \frac{2(t_\beta + t_{\alpha/2})^2 SD^2}{\delta^2} (1 + (m - 1)\rho) \quad (2)$$

Dónde

- $m$  es el número de personas muestreadas en cada grupo.
- $k$  es el número de grupos muestreados.
- $\rho$  es el coeficiente de correlación intraclase (CCI) prevista en la línea de base del proyecto. El CCI es una medida de cuánta variabilidad en el indicador se debe a las diferencias entre grupos frente a individuos dentro de grupos. Se explica con más detalle en el apéndice 1, incluyendo cómo se correlaciona con el efecto de diseño y una lista de posibles valores CCI.
- $t_\beta$ ,  $t_{\alpha/2}$ ,  $SD$  y  $\delta$  se definen como se indica arriba.

Para simplificar la escritura, la ecuación (2) se presenta así. Sin embargo, para facilitar el cálculo en los ejemplos del apéndice 2,  $m$  se resuelve en función de  $k$ . Por lo tanto, los usuarios ingresarán  $k$  (grupos) y calcularán  $m$  (individuos). Obsérvese que, en el caso agrupado, la distribución t tiene  $2*(k-1)$  grados de libertad. Es útil colocar las distintas opciones en una tabla y jugar con diferentes números de grupos para determinar el mejor tamaño de muestra dadas las limitaciones de recursos.

Utilice la ecuación (2) cuando haga comparaciones entre grupos para un indicador continuo recopilado de una muestra agrupada.

En otro ejemplo, los proyectos Alimentación para la Educación de McGovern-Dole International, financiados por el USDA en Sierra Leona y Burkina Faso, querían ver cómo el gasto educativo difería entre los hogares que eran o no miembros de las Comunidades de Ahorro y Préstamos Internos (SILC, por sus siglas en inglés). Utilizando una revisión documental, el proyecto asumió  $\rho = 0,28$ , como se encontró en un estudio similar realizado en Uganda sobre las características del hogar versus las de la comunidad que contribuyen al ahorro (Chowa, Ansong, and R. Despard 2014). Utilizando los valores medios de un estudio de SILC en Zambia, el equipo observó que los miembros de SILC frente a



Miembro de SILC en Burkina Faso junto a la caja de seguridad del grupo.

los que no lo eran gastaban un 152 % más en gastos de educación que la media de 10,47 USD (Noggle 2017; Author's calculations from study data). La desviación estándar de los datos de Zambia fue alta, de 24,87 USD. El proyecto consideró que el presupuesto podría respaldar la detección de un cambio menor al 152 %, en caso de que no se observara una diferencia tan grande en el estudio de Sierra Leona/Burkina Faso. Decidieron encuestar a 7 miembros de SILC en cada uno de los 27 SILC, lo que permitiría un aumento del 115 % (11,95 USD) sobre la media del grupo de control del estudio de Zambia.

$$n^* = \frac{2(0,88 + 2,26)^2 * 24,87^2}{11,95^2} (1 + (7 - 1)0,28)$$

Donde  $t_\beta = 0,88$ ;  $t_{\alpha/2} = 2,26$ ;  $SD = 24,87$ ;  $\delta = 11,95$ ;  $k = 7$ ; y  $\rho = 0,28$ . La Tabla 2 es un ejemplo.

**TABLA 2. EJEMPLO DE SILC EN SIERRA LEONA / BURKINA FASO**

CAMBIO ( $\delta$ )	15,91	14,13	13,09	12,04	10,47
GRUPOS ( $k$ )	27	27	27	27	27
INDIVIDUOS ( $m$ )	2	3	4	7	41
TAMAÑO TOTAL DE LA MUESTRA ( $n^*$ )	54	81	107	162	1.080

Como se muestra en la Tabla 2, si se mantiene constante el número de grupos en 27 y la SD en 24,87, el equipo del proyecto elige el equilibrio entre el cambio ( $\delta$ ) y el tamaño de la muestra individual ( $m$ ) y, por extensión, el presupuesto. Veintisiete grupos, con 2 individuos cada uno, es el tamaño total de muestra más pequeño, pero limita el tamaño del cambio que el equipo puede detectar.

### 2.2.1.3 Ecuación (3)

Al realizar comparaciones entre grupos para un **indicador binario recopilado de una muestra no agrupada**, utilice la ecuación (3):

Utilice la ecuación (3) cuando realice comparaciones entre grupos para un indicador binario recogido de una muestra no agrupada.

$$n^* = (p_1(1 - p_1) + p_0(1 - p_0)) \frac{(z_\beta + z_{\alpha/2})^2}{\delta^2} \quad (3)$$

Dónde

- $p_1$  es la meta del proyecto para el indicador
- $p_0$  es el valor de referencia
- $z_\beta$  (beta) es el valor crítico de una cola de la distribución normal inversa. Normalmente, el valor crítico  $z_\beta$  elegido es del 80 % y representa la potencia de la muestra. El valor crítico necesario se puede obtener de una tabla z adecuada o mediante el uso de Excel INV.NORM (como se utiliza en el apéndice 2).<sup>13</sup>
- $z_{\alpha/2}$  es el valor crítico de dos colas de la distribución normal inversa. Normalmente, el valor  $z_\alpha$  elegido es el 5 % y representa el nivel de significación. El valor crítico necesario se puede obtener de una tabla z adecuada o mediante el uso de Excel INV.NORM (como se utiliza en el apéndice 2).
- $\delta$  (delta) es el cambio objetivo en el indicador y es  $p_1 - p_0$ .<sup>14</sup>

Para demostrarlo, el proyecto Alimentación para la Educación de McGovern-Dole Internacional en Burkina Faso, financiado por el Departamento de Agricultura de los Estados Unidos (USDA, por sus siglas en inglés), quiso encuestar a las madres para medir el indicador "Porcentaje de participantes en intervenciones nutricionales a nivel comunitario que practican comportamientos de Alimentación del Lactante y del Niño Pequeño (ALNP)". Anticiparon que el valor de referencia sería del 40 % y establecieron la meta final en el 55 %. El equipo llegó a la conclusión de que encuestaría a 171 madres de niños menores de 2 años.

$$n^* = (0,55(1 - 0,55) + 0,40(1 - 0,40)) \frac{(0,84 + 1,96)^2}{0,15^2}$$

donde  $p_1 = 0,55$ ;  $p_0 = 0,40$ ;  $z_\beta = 0,84$ ;  $z_{\alpha/2} = 1,96$ ; y  $\delta = 0,15$ .

#### 2.2.1.4 Ecuación (4)

Utilice la ecuación (4) cuando realice comparaciones entre grupos para un indicador binario recogido de una muestra agrupada.

Al hacer comparaciones entre grupos para un **indicador binario recopilado de una muestra agrupada**, utilice la ecuación (4) que añade el término  $(1 + (m - 1)\rho)$  a la ecuación (3). Este término aumenta el tamaño de la muestra para compensar los efectos de la agrupación.

$$n^* = m^*k^* = (p_1(1 - p_1) + p_0(1 - p_0)) \frac{(z_\beta + z_{\alpha/2})^2}{\delta^2} (1 + (m - 1)\rho) \quad (4)$$

donde todos los parámetros se definen como se indica arriba.

Para usar otro ejemplo del equipo del USDA de Burkina Faso, estaban utilizando una evaluación de desempeño para determinar el "Número de maestros/educadores/asistentes de enseñanza que demuestran el uso de técnicas o herramientas de enseñanza nuevas y de calidad". Para calcular el

<sup>13</sup> Si su interfaz Excel está en inglés, utilice la función NORM.INV.

<sup>14</sup> Obsérvese que, en el caso binario de los grupos de tratamiento y control, el único momento en que una división uniforme de 50/50 entre los grupos de tratamiento y control es óptima es cuando  $p_0 = 1 - p_1$ , por ejemplo, cuando el valor basal se anticipa en el 40 % y el valor de evaluación final anticipado en el grupo de tratamiento (valor objetivo) es del 60 %. Consulte la Sección 4.4 sobre cómo determinar la división óptima cuando  $p_0 \neq 1 - p_1$ .

tamaño de la muestra necesario al inicio del estudio, el equipo utilizó los datos de la evaluación final de una fase anterior para calcular el CCI pertinente de 0,44. En su IPTT, estimaron el valor de referencia en un 49 %. Su meta para el indicador era del 65 %. El equipo llegó a la conclusión de que necesitaban tomar una muestra de 2 maestros en cada una de las 105 escuelas para detectar un cambio estadístico desde la línea de base hasta el final.

$$n^* = (0,65(1 - 0,65) + 0,49(1 - 0,49)) * \frac{(0,84 + 1,96)^2}{0,16^2} (1 + (2 - 1)0,44)$$

donde  $p_1 = 0.65$ ;  $p_0 = 0.49$ ;  $z_\beta = 0.84$ ;  $z_{\alpha/2} = 1.96$ ;  $\delta = 0.16$ ,  $m = 2$ ; y  $\rho = 0.44$ . Una vez más, especialmente con los diseños agrupados, es útil elaborar una tabla para encontrar la mejor combinación de grupos y el número de personas por grupo para ajustarse a las restricciones de recursos, como se hace en la Tabla 3:

**TABLA 3. EJEMPLO DE PROFESOR DE BURKINA FASO**

CAMBIO ( $\delta$ )	16 %	16 %	16 %	16 %	16 %
GRUPOS ( $k$ )	146	105	92	85	81
INDIVIDUOS ( $m$ )	1	2	3	4	5
TAMAÑO TOTAL DE LA MUESTRA ( $n^*$ )	147	211	276	340	405

Como se muestra en la Tabla 3, si se mantiene constante el cambio deseado del 16 %, el equipo del proyecto elige el equilibrio entre el tamaño de la muestra del grupo ( $k$ ) frente al tamaño de la muestra individual ( $m$ ). Ciento cuarenta y seis grupos, con 1 individuo en cada uno, es el tamaño total de muestra más pequeño. Sin embargo, viajar a 105 escuelas se alineó más estrechamente con otros indicadores que requerían visitas a esas mismas escuelas, lo que redujo los costos generales de recopilación de datos.

### 2.2.2 Grupos individuales

Si un equipo de proyecto está seguro de que solo necesita evaluar un contexto para un grupo, y no anticipa el uso de los datos para detectar cambios a lo largo del tiempo, entonces se deben usar las ecuaciones (5-8). Esto podría ser apropiado, por ejemplo, para una evaluación de necesidades en la que no se realizarán comparaciones estadísticas entre puntos temporales, áreas geográficas o género.

Las ecuaciones (5-8) son muy similares a las ecuaciones (1-4), pero cuando no se detecta un cambio entre los grupos de comparación, el término  $\delta$  ya no es relevante. Las ecuaciones no incluyen un término  $\delta$ , ya que ya no hay preocupación por los errores de Tipo II (no encontrar una diferencia, a pesar de que la hay).

Las ecuaciones (5-8) siguen muy de cerca a Sullivan (2019), y son las mismas que se utilizan en la guía Alimentar el futuro para los socios implementadores (Stukel 2018b). Véase el apéndice 6 para una comparación más detallada de esta guía con la guía Alimentar el futuro.

Si recopila datos que se utilizarán para analizar datos de un solo grupo, utilice las ecuaciones (5-8)

- Para un indicador continuo recopilado de una muestra no agrupada, utilice la ecuación (5)
- Para un indicador continuo de una muestra agrupada, utilice la ecuación (6)
- Para un indicador binario de una muestra no agrupada, utilice la ecuación (7)
- Para un indicador binario de una muestra agrupada, utilice la ecuación (8)

### 2.2.2.1 Ecuaciones (5 y 6)

Ya no hay dos  $SD$  de muestra, por lo que el 2 en el numerador de ecuaciones (1-2) ya no es necesario. Sin embargo, para estimar la media poblacional dentro de un rango significativo, las ecuaciones incluyen un margen de error ( $E$ ). Por ejemplo, si se supone que el peso medio es de 60 kg y  $E$  se establece en 30 kg, entonces la muestra estimaría el rango de peso de la población adulta entre 30 y 90 kg, lo que no es significativo. En su lugar, establezca  $E$  en 5 kg para producir una estimación del peso real de la población adulta como 55-65 kg.

Por lo tanto, la ecuación (1) para los indicadores continuos se convierte en:

$$n^* = \frac{t_{\alpha/2}^2 SD^2}{E^2} \quad (5)$$

Multiplique la ecuación (5) por  $(1 + (m - 1)\rho)$  para usarla con muestras agrupadas.<sup>15</sup> Este término aumenta el tamaño de la muestra para compensar los efectos de la agrupación.

$$n^* = m^* k^* = \frac{t_{\alpha/2}^2 SD^2}{E^2} (1 + (m - 1)\rho) \quad (6)$$

### 2.2.2.2 Ecuación (7 y 8)

En las ecuaciones (3-4) para indicadores binarios, solo se necesita la probabilidad estimada para el indicador en la población evaluada y, por lo tanto, no hay un término  $p_1$  y  $p_0$ . Obsérvese que, como se describe en el apéndice 3, la varianza con un indicador binario es mayor cuando  $p = 50\%$ ; también cuando el tamaño de la muestra para las ecuaciones (7-8) es mayor. En ausencia de otros datos, calcule el tamaño de la muestra para las ecuaciones (7-8) con  $p = 50\%$ .

Por lo tanto, la ecuación (3) se convierte en:

$$n^* = (p(1 - p)) \frac{z_{\alpha/2}^2}{E^2} \quad (7)$$

Multiplique la ecuación (7) por  $(1 + (m - 1)\rho)$  para su uso con muestras agrupadas, ya que este término aumenta el tamaño de la muestra para compensar los efectos de la agrupación.

$$n^* = (p(1 - p)) \frac{z_{\alpha/2}^2}{E^2} (1 + (m - 1)\rho) \quad (8)$$

### 2.2.3 Resumen.

En la sección 2.2 se presentan las ecuaciones (1-4) para calcular el tamaño mínimo de la muestra necesario para detectar una diferencia estadística entre dos grupos de comparación (p. ej., basal y final; tratamiento y control, etc.). La ecuación apropiada dependía de si el indicador era a) continuo o binario y b) recopilado de una muestra agrupada.

Si los equipos del proyecto están llevando a cabo evaluaciones sencillas y no tienen la intención de detectar diferencias estadísticas entre los grupos de comparación (por ejemplo, género, área geográfica, control/tratamiento, línea de base/final, etc.), entonces es apropiado utilizar las ecuaciones (5-8) en lugar de las ecuaciones (1-4). Dependiendo del margen de error o de la

<sup>15</sup> Al igual que con la ecuación (2) anterior, para simplificar la escritura, la ecuación (6) se presenta como tal. Sin embargo, para la simplicidad computacional en el apéndice 2,  $m$  se resuelve en función de  $k$ . En este caso, la distribución  $t$  tiene  $k-1$  grados de libertad.

Utilice la ecuación (5) para las evaluaciones de un solo grupo de un indicador continuo recopilado de una muestra no agrupada. La ecuación (6) es la versión agrupada correspondiente.

Utilice la ecuación (7) para las evaluaciones de un solo grupo de un indicador binario recopilado de una muestra no agrupada. La ecuación (8) es la versión agrupada correspondiente.

Antes de finalizar los tamaños de muestra necesarios, los equipos de proyecto deben revisar la sección 3 para determinar si se aplica algún criterio adicional a su situación.

probabilidad estimada elegida, esto puede dar lugar a tamaños de muestra calculados más pequeños.

Antes de finalizar los tamaños de muestra necesarios, los equipos de proyecto deben revisar la sección 3 para determinar si se aplica algún criterio adicional a su situación. Si es así, tendrán que aumentar aún más el tamaño de su muestra.

Después de finalizar los cálculos de las secciones 2 y 3, si los equipos de proyecto están preocupados por el tamaño de la muestra calculado es demasiado grande dadas las limitaciones presupuestarias, revise la sección 4. Es posible reducir el tamaño de la muestra sin afectar al tamaño del cambio que se va a detectar.

## 3. Revisión requerida: Elementos que aumentarán el tamaño de la muestra

Las ecuaciones (1-8) presentan el tamaño mínimo necesario para detectar un cambio estadístico a lo largo del tiempo. En la sección 3 se ofrece una breve descripción general de las consideraciones adicionales que puede ser necesario tener en cuenta a la hora de determinar el tamaño final de la muestra que deben utilizar los equipos de proyecto. Se trata de pérdidas de datos, desgaste (conjuntos de datos de panel) y subpoblaciones específicas.

---

*CRS recomienda  
asumir una pérdida de  
datos del 5 %.*

### 3.1 Pérdida de datos

Si los equipos de proyecto anticipan que algunos datos recopilados se perderán debido a un error de la herramienta, el enumerador o la entrada de datos, deben recopilar datos de una muestra más grande que el mínimo calculado. CRS recomienda asumir una pérdida de datos del 5 %, pero los equipos de proyecto deben consultar con colegas de su programa de país sobre la experiencia previa con la pérdida de datos.

### 3.2 Disminución

Si los equipos de proyecto anticipan que algunos datos recopilados se perderán debido a que las personas o los grupos no se pueden rastrear desde la línea de base hasta el final (deserción/disminución);<sup>16</sup> entonces, deben recopilar datos de una muestra más grande que el mínimo calculado. Es probable que este problema sea más grave cuando se recopila un conjunto de datos de panel (en el que se realiza un seguimiento del mismo individuo a lo largo del tiempo) en comparación con una sección transversal repetida (generalmente cuando se realiza un seguimiento del mismo grupo a lo largo del tiempo, pero los individuos dentro de él varían).

En el caso de los conjuntos de datos de panel, debido a que se encuesta a los mismos individuos al menos dos veces, CRS recomienda aumentar el tamaño de la muestra en un 10 % adicional (por encima de las concesiones de pérdida de datos hechas anteriormente), aunque hay algunas excepciones.<sup>17</sup> Los equipos de proyecto deben consultar con colegas de su programa de país sobre la experiencia previa con las tasas de deserción.

### 3.3 Indicadores Específicos para Subpoblaciones

Tenga en cuenta que, para ciertos indicadores, es posible que los equipos de proyecto no puedan identificar a las personas con los criterios necesarios antes de la recopilación de datos. Por ejemplo, el indicador de resultados globales de CRS "Prevalencia de niños de 6 a 23 meses que reciben una dieta mínima aceptable (DMA)" solo se aplica a los cuidadores de niños de 6 a 23 meses. Es posible que los equipos del proyecto no tengan una lista detallada de estos

---

<sup>16</sup> El seguimiento de los beneficiarios a lo largo del tiempo (datos de panel) puede ayudar a reducir el tamaño total de la muestra necesaria. Consulte la sección 4.5 para obtener más detalles.

<sup>17</sup> El estudio MIRA recopiló datos de panel sobre los mismos HH durante más de 24 meses. La deserción fue del 3 % al 5%; esto se debe en gran medida al seguimiento frecuente (mensual) y al uso de CRS de encuestadores integrados que residen en las comunidades encuestadas.

cuidadores antes de la recopilación de datos (especialmente al inicio del estudio) y, por lo tanto, es posible que deban sobre muestrear a la población objetivo con la esperanza de muestrear suficientes cuidadores para alcanzar el tamaño de muestra identificado. Los equipos de proyecto deben consultar con colegas de su programa de país sobre la experiencia previa con subpoblaciones específicas de indicadores.

Por ejemplo, si los equipos del proyecto saben que aproximadamente 1 de cada 3 hogares en sus comunidades participantes tienen niños de 6 a 23 meses, podrían aumentar el tamaño de la muestra en un 300 % con respecto al tamaño calculado. Sin embargo, los encuestadores solo pueden hacer las preguntas relacionadas con el indicador DMA a los hogares una vez que hayan confirmado que tienen un niño de 6 a 23 meses en su hogar.

### 3.4 Grandes cambios desde la línea de base hasta la línea final

Al realizar comparaciones entre grupos, si se estima un gran cambio entre los grupos, el tamaño de la muestra calculado puede ser demasiado pequeño para que cada punto de datos individual sea significativo.

En este ejemplo, se calcula que el 80 % de los agricultores adoptarán una nueva técnica (una vez que se les haya demostrado) porque es fácil de adoptar y supone una gran mejora con respecto a la práctica actual. Según las evaluaciones del diseño del proyecto, se calcula que sólo el 10 % de los agricultores utilizan la técnica. Debido a este gran cambio desde la línea de base hasta la línea final y utilizando la ecuación (3), sólo se deben muestrear cinco agricultores en la línea de base y en la línea final para detectar el cambio del 70 % a lo largo del tiempo. Sin embargo, si se utiliza la ecuación (7) y se ajusta el margen de error hasta que sólo se muestreen 5 agricultores, vemos que el valor inicial tendrá un margen de error del 27 %. Por lo tanto, si la media de la muestra de referencia es del 10%, sólo podemos decir que el verdadero valor de referencia está entre el 0% y el 37 %. Por esta razón, es preferible utilizar la ecuación (7) para calcular el tamaño de la muestra, estimando que el valor basal es del 10 % y con un margen de error del 10 %. Esto recomendaría muestrear a 35 agricultores en la línea de base, y podríamos decir que el verdadero valor de referencia está entre el 0 % y el 20 %.

El tamaño de la muestra final se puede actualizar posteriormente en consecuencia, utilizando los parámetros de los datos del estudio de línea de base.

### 3.5 Resumen

Si, sobre la base de una revisión de la sección 3, los equipos del proyecto determinan que necesitan aumentar el tamaño de la muestra para tener en cuenta la pérdida de datos, la deserción o subpoblaciones específicas, esto se puede hacer aumentando el número de individuos o grupos muestreados. La elección depende de la probabilidad anticipada de que un grupo deje de funcionar durante la vida del proyecto (es poco probable que las escuelas públicas cesen, pero SILC o las asociaciones de productores pueden hacerlo) frente a los individuos que abandonan ese grupo. Los grupos o individuos adicionales de "respaldo" deben seleccionarse al azar de la misma manera que se eligen otros encuestados.

---

*Si los equipos de proyecto utilizan las ecuaciones (1-4), se recomienda cotejar las estimaciones puntuales individuales con las ecuaciones (5-8).*

## 4. Otras consideraciones

Esta sección presenta conceptos más avanzados y hace referencia a las herramientas disponibles fuera de esta guía, sobre las cuales los lectores pueden necesitar más información. Esta sección se centra en 1) el factor de corrección de la población finita; 2) la relación entre los valores de referencia de los indicadores, las metas y el tamaño de las muestras; 3) estratificación de muestras; 4) indicadores binarios en las evaluaciones de impacto; y 5) conjuntos de datos de panel y cálculos del tamaño de la muestra.

### 4.1 El factor de corrección por población finita (FPC)

Si le preocupa que el tamaño de la muestra calculado sea demasiado grande para las restricciones presupuestarias, vea si califica para la reducción por el factor de corrección de población finita (FPC). El FPC se utiliza cuando el tamaño de la muestra es grande en comparación con el tamaño de la población. Utilice el FPC si la muestra calculada es superior al 5% de la población para la que se está recogiendo el indicador (independientemente de que sea continua o binaria). El FPC teórico es  $1 - \left(\frac{n}{N}\right)$ , aunque a veces se escribe como  $\frac{(N-n)}{N}$ , donde  $n$  es el tamaño de la muestra calculado y  $N$  es el tamaño de la población. A continuación, el FPC se multiplica por el intervalo de confianza respectivo a partir del cual se derivan las ecuaciones del tamaño de la muestra (los intervalos de confianza se analizan con más detalle en la sección 6.2.6). No mostramos el álgebra aquí, pero la ecuación (9) se usa para ajustar el tamaño de la muestra inicial por el FPC; nótese que sigue a Thompson (2012):

$$\frac{1}{\frac{1}{n^*} + \frac{1}{N}} \quad (9)$$

siendo  $n^*$  el tamaño de la muestra inicial calculado y  $N$  definido como se ha indicado anteriormente.

El FPC debe aplicarse al tamaño de la muestra calculado antes de realizar los ajustes por pérdida de datos, desgaste o subpoblaciones específicas descritas en la sección 3.

Tenga en cuenta que, en la literatura económica, el FPC generalmente se ignora porque los investigadores asumen que el tamaño de la muestra es pequeño en relación con toda la población (Cameron and Trivedi 2005). Esto sería especialmente cierto cuando se asume la validez externa (generalizabilidad más allá de un proyecto específico) con la evaluación de impacto o las actividades de investigación. Esto suele ocurrir con un diseño experimental en el que se comparan los grupos de tratamiento con un grupo de control. En tales casos, no se recomienda aplicar el FPC.<sup>18</sup>

### 4.2 Establecer objetivos alcanzables y ser conscientes de los costes asociados a la recopilación de datos

Es importante comprender las implicaciones para los presupuestos MEAL durante la etapa de diseño del proyecto. Deben recopilarse y notificarse datos sobre los indicadores estándar de los donantes, y deben asignarse recursos suficientes para detectar un cambio en esos indicadores a lo largo del tiempo.

<sup>18</sup> Si se informan intervalos de confianza en torno a una media muestral, y si se aplican los criterios de FPC, también se debe utilizar para ajustar la confianza interna. Véase la sección 6.2.7.

---

*No se recomienda aplicar el FPC si se asume una validez externa, como en el caso de la evaluación de impacto o las actividades de investigación.*

Sin embargo, los indicadores personalizados quedan a discreción de los equipos de proyecto. Por ejemplo, durante el reciente diseño de un proyecto educativo en Togo, el equipo quería recopilar índices de masa corporal (IMC) para niños en edad escolar. Es difícil encontrar estos datos para cualquier país, y el equipo de diseño del proyecto supuso que podría ver un pequeño cambio (3-5 %) en este indicador durante el proyecto de 5 años, pero no estaba seguro. Dado el pequeño cambio para un indicador continuo, el tamaño de la muestra requerida fue un 50 % mayor que el de cualquier otro indicador del proyecto.

Por lo tanto, el equipo de diseño del proyecto decidió recopilar los datos a través de un estudio especial, mientras que los equipos de MEAL recopilaron otros datos de los estudiantes, utilizando el tamaño de muestra más pequeño recomendado para otros indicadores. Aunque podría no ser posible detectar un cambio estadístico a lo largo de la vida del proyecto, los datos deberían resultar útiles para informar futuros proyectos en Togo y, potencialmente, otros proyectos educativos de CRS a nivel mundial. Si, inesperadamente, se logra un cambio mayor de lo previsto, el equipo del proyecto podrá detectar una diferencia estadística entre la línea de base y la final. Por lo tanto, el equipo del proyecto eliminó este indicador del IPTT del proyecto, lo que tenía la ventaja añadida de no comprometer al equipo del proyecto con un cambio específico que sería muy costoso de detectar si es que podían hacerlo.

### 4.3 Estratificación

La estratificación simplemente significa organizar o clasificar los datos en grupos. Esto se ha mencionado anteriormente, cuando se habla de los grupos de tratamiento y control, el género o la separación geográfica. Al calcular el tamaño de las muestras, es importante pensar en cualquier estratificación que los equipos de proyecto harán con los datos por dos razones:

---

*Si los equipos de proyecto desean realizar comparaciones estadísticas entre estratos, es necesario aumentar el tamaño de la muestra para tener esto en cuenta.*

- 1) **Comparaciones estadísticas entre estratos.** Si los equipos de proyecto desean realizar comparaciones estadísticas entre estratos, es necesario aumentar el tamaño de la muestra para tener esto en cuenta. Como se describe en la sección 1, esto se suele hacer multiplicando el tamaño final de la muestra por el número de estratos que se van a analizar. Por ejemplo, si se evalúan los resultados educativos de los niños frente a los de las niñas (2 estratos) y el tamaño de la muestra es de 5 niños en cada una de las 50 escuelas, hable con 5 niños y 5 niñas en cada escuela.
- 2) **No se necesitan comparaciones estadísticas.** Si los equipos de proyecto necesitan estratificar los datos, pero no desean hacer comparaciones estadísticas entre los estratos (esto se hace a menudo para la desagregación de datos por género), entonces el tamaño de la muestra no necesita aumentar. Por ejemplo, si el equipo desea conocer los rendimientos de los productores masculinos y femeninos, puede seleccionar al azar a los productores masculinos de una lista y seleccionar por separado a las productoras, para garantizar una representación uniforme de ambos sexos en la muestra, sin aumentar el tamaño total de la muestra.

Los pesos de la muestra, descritos en la sección 6.2.1, deben tener en cuenta cualquier estratificación de la muestra. Esto es cierto incluso cuando no se hace una comparación estadística entre estratos.

- Una potencial forma de evitar la necesidad de ponderaciones es utilizar un muestreo sistemático de intervalos fraccionarios. Se pueden encontrar más detalles sobre esta estrategia de muestreo en la sección 9.4.2 de Stukel (2018b).

---

*Evite la necesidad de pesos de muestra al estratificar una muestra mediante el uso de muestreo sistemático de intervalo fraccionario.*

Al realizar comparaciones estadísticas entre estratos y agrupar la muestra, existe la opción de aumentar el número de grupos, o individuos dentro de cada grupo. Para continuar con el ejemplo del punto 1) anterior, si se evalúan las diferencias entre regiones geográficas (2 estratos), hable con

5 niños en cada una de las 50 escuelas de la Región A y haga lo mismo en la Región B. Si se evalúan las diferencias por región geográfica y género (4 estratos), hable con 5 niños y 5 niñas en cada una de las 50 escuelas de la Región A, y haga lo mismo en la Región B.

#### 4.4 Utilizar el tratamiento desigual y el tamaño de los grupos de control con indicadores binarios

Cuando se utiliza una evaluación de impacto para medir un indicador binario<sup>19</sup> de una muestra agrupada, es más eficaz permitir un número desigual de grupos en el grupo de control frente al grupo de tratamiento. Con los indicadores binarios, el único momento en que una división 50/50 entre los grupos de tratamiento y control es óptima es cuando  $p_0 = 1 - p_1$ , por ejemplo, cuando el valor basal se anticipa en el 40 % y el valor final del grupo de tratamiento se anticipa en el 60 %.

Aunque la división puede dejarse en 50/50, si se realiza una evaluación de impacto (o investigación) con comparaciones entre los grupos de tratamiento y control, puede ser más rentable tener una división desigual. Esto puede ser especialmente cierto si hay un costo elevado para el tratamiento en sí; el grupo de tratamiento puede ser el más pequeño de los dos grupos.

Dada la rareza de esta ocurrencia en el contexto de CRS, esta guía simplemente recomienda que los equipos de proyecto se refieran a las ecuaciones 17, 18 y 21 (y la hoja de cálculo de Excel que las acompaña) en McConnell y Vera-Hernández (2015) para calcular la división óptima entre los grupos de tratamiento y control.

#### 4.5 Uso de conjuntos de datos de panel

Si se realiza un seguimiento de la misma persona a lo largo del tiempo, los equipos de proyecto pueden obtener poder estadístico utilizando el valor basal de cada persona al analizar los datos de la evaluación final. El cálculo de tamaños de muestra que tengan en cuenta explícitamente el poder estadístico adicional proporcionado por los conjuntos de datos de panel requiere que los equipos de proyecto tengan información adicional para las ecuaciones de tamaño de muestra, que puede no estar disponible. La información adicional es:

- En el caso de los indicadores continuos, el CCI tanto a nivel individual como de grupo. Si está interesado en este enfoque, consulte la ecuación (16) en McConnell y Vera-Hernández (2015).
- En el caso de los indicadores binarios, el valor basal de la variable de resultado, o cualquier otra covariable, puede reducir el tamaño de la muestra calculada. El usuario tendrá que estimar empíricamente cómo la covariable afecta al indicador antes de calcular el tamaño de la muestra, y las ecuaciones del tamaño de la muestra en sí mismas pueden ser computacionalmente abrumadoras para aquellos que no están familiarizados con el álgebra lineal. Si está interesado en este enfoque, consulte las ecuaciones (24-25), en McConnell y Vera-Hernández (2015). En el apéndice 3 se proporciona un ejemplo de código R para las ecuaciones (24-25).

En lugar de utilizar ecuaciones de tamaño de muestra específicas para conjuntos de datos de panel, simplemente utilice las ecuaciones (1-4) de esta guía, según corresponda. Ellas no tienen en cuenta el poder estadístico adicional proporcionado por los conjuntos de datos de panel, pero garantizarán que el tamaño de la muestra sea lo suficientemente grande.

---

*Para calcular el tamaño de las muestras que tienen en cuenta el poder estadístico adicional proporcionado por los conjuntos de datos de panel, véase McConnell y Vera-Hernández (2015).*

<sup>19</sup> Tenga en cuenta que, con indicadores continuos, es menos eficiente tener una división desigual entre los grupos de tratamiento y control. Véase la ecuación 3 en McConnell y Vera-Hernández (2015).

## 4.6 Resumen

Esta sección incluye temas especiales sobre los que los usuarios avanzados, probablemente asesores técnicos, deben tener en cuenta con respecto a los cálculos del tamaño de la muestra. Incluyó un debate general sobre el factor de corrección de la población finita; la relación entre el tamaño de la muestra y las metas del proyecto; estratificación de la muestra; indicadores binarios en las evaluaciones de impacto; y el cálculo de los tamaños de muestra para los conjuntos de datos de panel.

# 5. Mitos sobre el tamaño de la muestra

En el campo del desarrollo internacional, existen dos malentendidos generalizados con respecto a los cálculos del tamaño de la muestra: 1) que el tamaño de la muestra depende del tamaño de la población subyacente y 2) que los indicadores binarios requieren muestras más grandes que los continuos. Ambos mitos son desacreditados en esta sección.

## 5.1 El tamaño de la muestra depende del tamaño de la población subyacente

Después de revisar esta guía, tenga en cuenta que la única vez que se considera el tamaño de la población subyacente es cuando se considera apropiado el FPC (sección 4.1). Los factores que influyen en los cálculos del tamaño de la muestra son: 1) el tamaño del cambio que se va a detectar (o margen de error aceptable) y 2) el número de grupos de comparación. Solo cuando el tamaño de la muestra es superior al 5% de la población subyacente debemos considerar reducirlo por el factor FPC.

## 5.2 Los indicadores binarios requieren tamaños de muestra más grandes

Algunos datos se pueden convertir "razonablemente" de binarios a continuos, y viceversa. A menudo se dice que es preferible utilizar la versión continua, ya que requerirá un tamaño de muestra más pequeño (Frost 2020). Al comparar las ecuaciones básicas para variables continuas y binarias, ecuaciones (1) y (3) respectivamente, cada una tiene parámetros diferentes y no hay pruebas matemáticas de que una de ellas siempre resulte en un tamaño de muestra más pequeño o grande.

$$n^* = \frac{2(t_{\beta} + t_{\alpha/2})^2 SD^2}{\delta^2} \quad (1) \quad n^* = (p_1(1 - p_1) + p_0(1 - p_0)) \frac{(z_{\beta} + z_{\alpha/2})^2}{\delta^2} \quad (3)$$

Para dar un ejemplo utilizando los resultados de las pruebas de los estudiantes de primaria en Sierra Leona, el indicador estándar del proyecto requerido por los donantes es el porcentaje de niños que aprueban la prueba de lectura, que es un indicador binario. El equipo del proyecto anticipó que el valor de referencia sería del 41 % y estableció un objetivo del 58 %, por lo que quería medir un cambio del 17 %. Se considera que los estudiantes que responden correctamente 3 de las 5 preguntas han aprobado el examen. En teoría, el valor basal y la meta previstos podrían convertirse en una puntuación. Si el 41% de los estudiantes obtuvieron al menos 3 puntos en la prueba (y todos los demás obtuvieron cero), el puntaje promedio de referencia sería  $3 * 0,41 = 1,23$  por estudiante; Por lo tanto, el objetivo sería  $3 * 0,58 = 1,74$  por estudiante, y el cambio sería  $0,51$ . Utilizando los datos reales de la puntuación de la prueba de este ejemplo, la desviación estándar es  $1,83$ . En este ejemplo, la ecuación (1) para la versión continua del indicador calcula el tamaño de muestra más grande.

$$n^* = \frac{2(0,88 + 2,26)^2 1,83^2}{0,51^2} = 205$$

$$n^* = (0.58(1 - 0.58) + 0.41(1 - 0.41)) \frac{(0.84 + 1.96)^2}{0.17^2} = 132$$

Es importante recordar que los indicadores continuos y binarios tienen diferentes distribuciones de probabilidad y, por lo tanto, varianzas. Por esta razón, requieren ecuaciones diferentes y no se pueden hacer afirmaciones generales sobre los cálculos resultantes. Los parámetros necesarios para cada ecuación son diferentes porque miden cosas fundamentalmente diferentes.

### 5.3 Resumen

Utilice siempre la ecuación adecuada para el tipo de indicador, ya que las pruebas estadísticas finales realizadas en la etapa de análisis dependerán del tipo de indicador. En última instancia, el objetivo de los cálculos del tamaño de la muestra es tener una muestra lo suficientemente grande como para detectar cambios estadísticos en la etapa de análisis, y el tamaño de la población subyacente no es un factor en estos cálculos.

# 6. Diseño de la muestra y análisis

Aunque esta guía se centra en los cálculos del tamaño de la muestra, es importante comprender cómo el método elegido de selección de la muestra, realizado antes de determinar qué ecuación de tamaño de muestra utilizar, afecta al análisis. En esta sección se ofrece una breve descripción general del muestreo aleatorio y el sesgo de selección de muestras; selección de la población proporcional al tamaño de la muestra; y el uso de pesos muestrales en la etapa de análisis.

## 6.1 Selección de la muestra

En esta sección se describe la importancia de utilizar muestras aleatorias y los desafíos de utilizar la metodología de probabilidad proporcional al tamaño (PPT).

### 6.1.1 Utilizar muestras aleatorias y documentar cualquier sesgo muestral debido a un muestreo no aleatorio

Las muestras representativas siempre deben seleccionarse al azar, a partir de una lista precargada o de un censo rápido, y debe conocerse la probabilidad de seleccionar a una persona en la muestra.<sup>20</sup> CRS tiene referencias internas para varias formas de seleccionar una muestra cuantitativa aleatoria (Culligan et al. 2019). Tenga en cuenta que la selección aleatoria de muestras es fundamental para lograr la validez externa; será difícil publicar externamente cualquier evaluación o resultados de investigación de una muestra no aleatoria.

Sin embargo, a menudo debido a las limitaciones de recursos, se produce un muestreo no aleatorio y, por lo tanto, un sesgo de selección. Esto puede deberse a limitaciones de seguridad que impiden que los equipos de estudio lleguen a un área prohibida o cuando las listas de las que se seleccionan aleatoriamente las personas o los grupos están desactualizadas, y resultaría demasiado costoso o imposible localizar a los seleccionados al azar. Si falta un 5 % o más de respuestas a una pregunta individual (Cameron and Trivedi 2005), en la sección de limitaciones del informe de evaluación, describa lo mejor posible las fuentes de sesgo debidas a estas omisiones.

Por ejemplo, si los estudiantes no están presentes en la escuela el día en que se supone que deben ser encuestados, ¿en qué se diferencian los estudiantes ausentes de los presentes? ¿Una prueba  $t$  de medias indica que la proporción de grupos clave (género, etnia, área geográfica)<sup>21</sup> en la muestra es la misma que la de los que no fueron incluidos? Si no es así, ¿cómo podría estar sesgada la muestra? ¿De qué otra manera podría ser diferentes los estudiantes que no se presenten ese día? ¿Es posible que no obtengan tan buenos resultados en las pruebas de alfabetización, etc., porque podrían faltar a la escuela con frecuencia?

Otro ejemplo, ¿qué pasa si se puede medir el rendimiento de algunos agricultores, pero no de otros porque a) no plantaron el cultivo que se está encuestando o b) plantaron el cultivo, pero no lo cosecharon?

<sup>20</sup> Por esta razón, no se recomienda el uso de una "caminata aleatoria" para la selección de hogares, a menos que se conozca el número total de hogares en una comunidad.

<sup>21</sup> Es posible que el analista no tenga mucha información sobre los estudiantes que no están presentes. Sin embargo, según los nombres de los estudiantes y la ubicación de las escuelas, es posible que al menos tengan esta información.

---

*Si falta el 5 % o las respuestas a una pregunta individual, en la sección de limitaciones del informe del estudio, describa cualquier fuente de sesgo debida a estas omisiones.*

Imagine otros escenarios en los que esto podría ocurrir y reflexione sobre lo que se puede decir sobre las diferencias entre los que podrían y no podrían ser encuestados.

### 6.1.2 La selección de grupos de población proporcional al tamaño (PPT) puede no ser apropiada en el contexto de CRS

El PPT es un método para seleccionar grupos de estudio. Se utiliza comúnmente para tener en cuenta el tamaño de los grupos al seleccionarlos en la primera etapa de la recopilación de datos, en la que cada individuo de cada grupo tiene la misma probabilidad de ser seleccionado en la muestra. Si, en la segunda etapa, se utiliza una muestra aleatoria simple o sistemática para seleccionar a cada individuo entre todos los individuos del grupo, entonces la muestra es "auto ponderada" y no es necesario aplicar pesos de muestra en la etapa de análisis.

Los analistas de datos recopilados a través de una muestra seleccionada por PPT deben comprender cuatro cosas:

- 1) Si la muestra fue estratificada (como se describe en la sección 4.3), o si no se utilizó una muestra aleatoria simple o sistemática en la segunda etapa, entonces la muestra no se auto pondera y se deben usar pesos de muestra (ver sección 6.2.1 para una discusión de los pesos de la muestra).
- 2) Para utilizar PPT, la medida del tamaño debe ser la misma que la unidad de muestreo utilizada en la etapa de análisis. Tenga en cuenta que a menudo se necesitan diferentes unidades (hogares, productores, cuidadores, etc.) para diferentes indicadores en el mismo proyecto, por lo que será necesario extraer una muestra diferente para cada unidad de muestreo. Por ejemplo, es incorrecto utilizar el número total de hogares de una aldea como el "tamaño" en PPT, y luego utilizar a los cuidadores de niños de 6 a 23 meses como unidad de muestreo. Si se utiliza PPT en este ejemplo, es necesario saber primero cuántos cuidadores de niños de 6 a 23 meses hay en cada aldea y utilizarlo como medida de "tamaño".
- 3) Con PPT, se deben utilizar los estimadores de Hansen-Hurwitz o Horvitz-Thompson para estimar la media de la muestra. Además, esos estimadores deben utilizarse al calcular la varianza en cualquier modelo de regresión (Hansen and Hurwitz 1942; Horvitz and Thompson 1952). Este punto no suele tratarse en otras guías de muestreo.
- 4) Además, cuando se utiliza PPT, la medida del tamaño debe ser precisa, de lo contrario se sobreestimarán o subestimarán la varianza de la muestra, en comparación con la simple selección aleatoria de grupos (Thomsen, Tesfu, and Binder 1986). Incluso si las medidas de línea de base del tamaño son precisas, si se utiliza una sección transversal repetida en los mismos grupos en la evaluación final y el "tamaño" de los grupos cambia notablemente con el tiempo, se producirá el mismo problema de estimación errónea de la varianza de la muestra.

Feed the Future (FtF o Alimentar el Futuro) recomienda el uso de PPT en su guía de muestreo dirigida a los implementadores. Al encuestar a individuos a través de grupos de productores, recomienda encuestar a todos los productores del grupo seleccionado; esto evitaría los problemas descritos en los puntos 1) y 2) anteriores. Con respecto al punto 2) al realizar encuestas de hogares, se recomienda utilizar una lista actualizada de todos los participantes individuales del proyecto como la medida del tamaño. A continuación, utiliza el mayor tamaño de muestra calculado en todos los indicadores del proyecto para encuestar a los participantes seleccionados, al tiempo que reconoce que existe el riesgo de no muestrear un número adecuado de encuestados por indicador individual (cuando el indicador solo es aplicable a una subpoblación específica) (Stukel 2018a, 2018b).

---

*La selección de grupos de PPT solo se auto pondera si se utiliza la selección de muestras aleatorias simple o sistemática en la segunda etapa.*

La guía de monitoreo y evaluación de actividades de emergencia de BHA proporciona tres ejemplos del uso de PPT para seleccionar grupos con múltiples indicadores. El hogar es siempre la unidad de muestreo, y en la segunda etapa se utiliza una muestra aleatoria simple o sistemática. Los ejemplos demuestran la complejidad del uso de PPT como metodología de muestreo y deben ser revisados por cualquier personal de CRS interesado en usar PPT para la recopilación de datos de múltiples indicadores ((*Technical Guidance for Monitoring, Evaluation, and Reporting for Emergency Activities 2022*)).

Dada la complejidad del análisis, las preocupaciones sobre las medidas del tamaño de los grupos y el aumento de los costos, el personal de CRS debe utilizar el PPT con cuidado. En lugar del PPT, los grupos y los individuos pueden seleccionarse mediante otras formas de muestreo aleatorio y las ponderaciones de la muestra pueden utilizarse en el análisis. Tenga en cuenta, sin embargo, que si se decide utilizar el PPT y los analistas van a realizar regresiones sobre los datos recopilados, las ponderaciones muestrales pueden reducir la precisión de las estimaciones de los coeficientes (Lee and Solon 2011). En la sección 6.2 se ofrece un análisis más detallado de las ponderaciones muestrales utilizadas en los análisis de regresión.

### 6.1.3 Resumen

Es importante seleccionar muestras o encuestados aleatorios. Siempre que esto no sea posible, asegúrese de documentar cualquier sesgo de selección de la muestra lo mejor posible. Si utiliza la selección de clústeres proporcional a la población (PPT), asegúrese de comprender su complejidad y limitaciones antes de su uso. Si el PPT no es una buena opción, consulte la sección 6.2 sobre pesos de muestra.

## 6.2 Uso de tamaños de muestra en el análisis de datos

En esta sección se describe por qué, cuándo y cómo utilizar los pesos de muestra. También proporciona una visión general de las ponderaciones de la falta de respuesta y de las ponderaciones de la muestra en los análisis de regresión. Se cierra con una sección sobre cómo calcular los intervalos de confianza y la aplicabilidad del factor de corrección de la población finita en la etapa de análisis.

### 6.2.1 Ponderación de la muestra: cálculo

Los pesos de las muestras solo son aplicables a algunas muestras agrupadas o estratificadas (al igual que con el muestreo aleatorio simple o sistemático, todos tienen la misma probabilidad de selección de muestra). Se aplican a muestras estratificadas que se aumentaron de tamaño para permitir el análisis entre estratos, o a muestras que se estratificaron con fines organizativos (véase la sección 6.2.1).

Las ponderaciones de la muestra se utilizan para reflejar el número de personas de la población que representa un individuo. Por ejemplo, si las escuelas se seleccionan al azar en la primera etapa de un diseño agrupado, y en la segunda etapa se seleccionan al azar 10 niñas de la única clase de 2º grado de la escuela<sup>22</sup> con 30 niñas en la clase, cada niña muestreada representa a otras 3 niñas.

Estadísticamente, los pesos de la muestra son la probabilidad inversa de ser seleccionado en la muestra y se definen como  $w_i = 1/\pi_i$ , donde  $\pi_i$  es la probabilidad de que el individuo  $i$  sea seleccionado en la muestra. En el ejemplo anterior,  $\pi_i$  para cada niña es 10/30 y, por lo tanto, su  $w_i$  es 3.

<sup>22</sup> Este ejemplo hace referencia a un indicador de educación de USAID que especifica a los estudiantes de 2º grado. Por lo tanto, todas las aulas que no sean de 2º grado serían excluidas de la encuesta, y no es necesario tener en cuenta las ponderaciones de la muestra.

---

*Los pesos de las muestras solo son aplicables a algunas muestras agrupadas o estratificadas.*

Un punto adicional, para facilitar los cálculos. En este ejemplo, si en la primera etapa se seleccionaron al azar 50 escuelas de 200, entonces cada escuela representa otras 4 escuelas. Dado que todas las escuelas tienen el mismo peso de muestra, los pesos de las escuelas se pueden ignorar en los cálculos.

Si las 50 escuelas se estratificaron entre dos regiones geográficas, lo que dio como resultado diferentes ponderaciones escolares, entonces se deben incluir sus ponderaciones. Por ejemplo, la Región A tiene 110 escuelas y la Región B tiene 90. Veintiocho escuelas fueron seleccionadas al azar de la Región A y 25 de la Región B. Así, las escuelas de la Región A tienen un peso de 110/25 y las de la Región B tienen un peso de 90/25.

### 6.2.2 Pesos de la muestra - uso

Los pesos de la muestra siempre deben utilizarse cuando se proporcionan estadísticas descriptivas univariadas para indicadores individuales, como medias/proporciones, totales, medianas, etc.

Sin embargo, lo ideal sería que los resultados de los análisis de regresión multivariante informaran de los resultados ponderados y no ponderados y, cuando haya diferencias, incluyeran un análisis de las razones subyacentes. Por ejemplo, las observaciones de una escuela que tiene 90 alumnos de segundo grado, frente a las que tienen 30, tendrán 3 veces más peso; si hay efectos heterogéneos del proyecto para las escuelas grandes frente a las pequeñas (por ejemplo, las escuelas más grandes tienen una mayor proporción de profesores por alumno; esta falta de atención de los estudiantes da lugar a peores resultados educativos, etc.), entonces las medias condicionales podrían ser diferentes para los análisis ponderados frente a los no ponderados (Solon, Haider, and Wooldridge 2015).

### 6.2.3 Medias/proporciones ponderadas y totales

La ecuación para calcular una media ponderada univariante es:

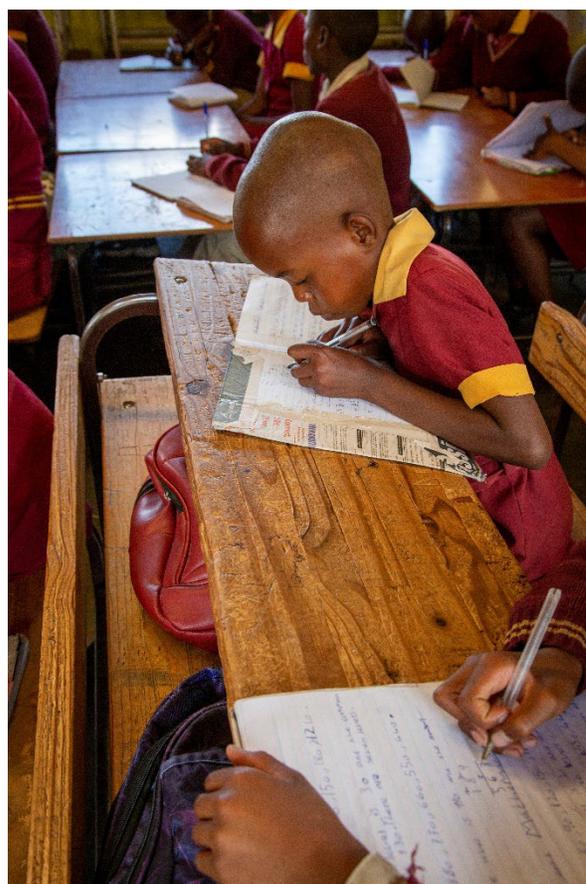
$$\bar{y} = \frac{\sum_i^I y_i * w_i}{\sum_i^I w_i} \quad (10)$$

donde  $y$  es nuestro indicador de interés para el individuo  $i$  con peso  $w$ .

Como ejemplo simple, use un indicador binario (aprobado = 1 / reprobado = 0) para 5 estudiantes de 2 aulas cada uno (Tabla 4). Dado que cada aula tiene un número diferente de alumnos, los del aula A (30 alumnos) tienen un peso de 30/5 = 6 y los del aula B (40 alumnos) tienen un peso de 40/5 = 8.

**TABLA 4. DATOS DE EJEMPLO DE PESO DE MUESTRA**

ID de estudiante	$y_i$	$w_i$	$y_i * w_i$
1	1	6	6
2	1	6	6
3	1	6	6
4	1	6	6
5	0	6	0
6	1	8	8
7	0	8	0
8	1	8	8
9	0	8	0
10	0	8	0
<b>SUMA</b>	<b>6</b>	<b>70</b>	<b>40</b>



Estudiantes de primaria en Lesotho. [Dooshima Tsee]

Por lo tanto, la proporción ponderada de estudiantes que aprobaron es la siguiente ecuación (10):

$$\hat{y}_w = 40/70 = 57 \%$$

mientras que la proporción simple es  $\hat{y} = \sum_i^I y_i/n = 6/10 = 60 \%$ .<sup>23</sup>

Cuando utilice una muestra representativa para extrapolar a una población más grande, simplemente multiplique  $\hat{y}_w$  por la población total. En este ejemplo, se estima que  $0,57*(30+40) = 40$  estudiantes han aprobado el examen en comparación con el  $0,60*(30+40) = 42$  estudiantes que se habrían estimado utilizando la muestra no ponderada.

### 6.2.4 Ponderaciones de la falta de respuesta

Las ponderaciones de la falta de respuesta se utilizan cuando una parte de la muestra prevista no respondió a la encuesta. Esta falta de respuesta podría deberse a que los participantes se negaron a participar, no pudieron ser localizados o los datos no se podían medir (por ejemplo, el rendimiento de un cultivo que nunca se cosechó).

Por lo tanto, las muestras se ponderan en función de las características observables de los encuestados y los no encuestados (como el sexo, la edad, la ubicación geográfica, etc.), además de su probabilidad de selección, como se describe en la sección anterior. Por ejemplo, las mujeres encuestadas se ponderarían para representar a las mujeres no encuestadas, o las encuestadas

<sup>23</sup> En notación estadística, el  $\hat{\phantom{y}}$  por encima de  $y$  se usa para las proporciones, mientras que  $\bar{\phantom{y}}$  por encima de  $y$  se usa para las medias.

*Utilice con precaución las ponderaciones que no responden. No asuma que los encuestados pueden reemplazar adecuadamente a los que no respondieron.*

mayores para representar a las personas mayores que no respondieron. Esta guía no proporciona detalles sobre cómo calcular y utilizar las ponderaciones de no respuesta, pero Raab (2009) y Consejo Nacional de Investigaciones Científicas (2002) explican la aplicación práctica en detalle.

Tenga en cuenta que las ponderaciones de falta de respuesta deben usarse con precaución. Cualquier falta de respuesta a la encuesta que no sea aleatoria crea una muestra sesgada, y ese sesgo debe documentarse (véase la sección 6.1.1). Los analistas no pueden asumir que los encuestados reales pueden reemplazar adecuadamente a los no encuestados.

Continuando con el ejemplo de la sección 6.1.1, ¿por qué sólo se puede medir el rendimiento de algunos agricultores? ¿La razón principal se debe a que algunos agricultores no cosecharon y, por lo tanto, no hubo rendimiento para medir? ¿No cosecharon porque no podían regar en un año de sequía o utilizar insumos agrícolas resistentes a la sequía? En este caso, el uso de ponderaciones de falta de respuesta para dar mayor peso a los agricultores encuestados que cosecharon (y suponiendo que puedan representar a los agricultores que no cosecharon) exacerbaría el sesgo de la muestra.

### 6.2.5 Muestras agrupadas o estratificadas y análisis de regresión

Al informar sobre las medias condicionales ponderadas de los análisis de regresión, los valores ponderados deben utilizar la contraparte ponderada adecuada (por ejemplo, mínimos cuadrados ponderados, verosimilitud máxima ponderada, etc.).

Además, debido a que es probable que las observaciones dentro de un grupo estén correlacionadas, los errores estándar de los coeficientes siempre deben agruparse (Cameron and Miller 2015). Los paquetes estadísticos tienen funciones para esto; la función adecuada variará en función del método de análisis.

Controle cualquier estratificación o aleatorización de la muestra en los análisis de regresión mediante el uso de variables binarias para cada estrato o unidad de aleatorización (excluyendo una para evitar la trampa de variables ficticias).

### 6.2.6 Intervalos de confianza

Como referencia, a continuación, se muestran las ecuaciones para calcular los intervalos de confianza (CI) alrededor de una media muestral. Para cualquier muestra aleatoria, se aplica la siguiente ecuación:

$$\text{Intervalo de confianza} = \text{media} \pm SE * t_{\alpha/2} \quad (11)$$

Donde SE es el error estándar de la media muestral y se define en el siguiente párrafo. Tenga en cuenta que, para los indicadores binarios,  $z_{\alpha/2}$  debe usarse en lugar de  $t_{\alpha/2}$ , y ambos se describen en la sección 2.2.1. Típicamente, en ciencias sociales, cuando se informan intervalos de confianza,  $\alpha = 0.05$ . El lado derecho de la ecuación (11) también se conoce como margen de error.

Obsérvese que los errores estándar de la media muestral o proporción serán diferentes para los indicadores continuos (ecuación (12)) y binarios (ecuación (13)).

$$SE_{\text{continuo}} = \frac{SD}{\sqrt{n}} \quad (12)$$

Véase el apéndice I si se necesita la ecuación de la desviación estándar ( $SD$ ). En las ecuaciones (12) y (13),  $n$  es el tamaño final de la muestra después de la recolección de datos.

---

*El lado derecho de la ecuación (11) también se conoce como margen de error.*

$$SE_{\text{Binario}} = \sqrt{\frac{p(1-p)}{n}} \quad (13)$$

En el caso de las muestras agrupadas, el SE debe multiplicarse por  $\sqrt{\frac{1+\rho}{1-\rho}}$  donde  $\rho$  es la CCI (Bence 1995). Esto debe hacerse antes de insertar el SE en la ecuación (11). Esto es cierto tanto para los indicadores continuos como para los binarios.

### 6.2.7 FPC

Si el factor de corrección por población finita (FPC) se aplica a la muestra (véase la sección 4.1), el SE también debe multiplicarse por él. Como recordatorio, el FPC es  $1 - \left(\frac{n}{N}\right)$ .

### 6.2.8 Resumen

En la sección 6.2 se explica que las ponderaciones de las muestras permiten esencialmente que las observaciones de un solo grupo representen a otros que no fueron encuestados en el mismo grupo; de este modo, se dará mayor importancia a las observaciones de grupos más grandes. Las ponderaciones de las muestras deben utilizarse en los estadísticos de resumen, pero las regresiones deben informar de los resultados de las muestras ponderadas y no ponderadas.

En la sección 6.2 también se ofrece una visión general de las ponderaciones de la falta de respuesta y de cómo utilizar muestras ponderadas en los análisis de regresión; ambas subsecciones incluyen enlaces a otras referencias para obtener más información. Se cerró con las ecuaciones necesarias para calcular los intervalos de confianza para los indicadores binarios y continuos y se señaló la modificación necesaria para los indicadores recogidos de muestras agrupadas. Si se utiliza en los cálculos del tamaño de la muestra, el factor de corrección de la población finita también debe utilizarse en el cálculo de los intervalos de confianza en la etapa de análisis.

# Apéndice 1. Coeficiente de correlación intraclase

El coeficiente de correlación intraclase (CCI) solo se aplica a los diseños agrupados e indica qué parte de la variabilidad de los datos se debe a las diferencias entre grupos frente a los individuos dentro de grupos. Para obtener una imagen real de la población, si los individuos dentro de los grupos son iguales entre sí, es mejor encuestar a menos individuos dentro de cada grupo y a un mayor número de grupos. De lo contrario, la similitud de las respuestas dentro de un grupo magnificará las diferencias en las respuestas de los individuos entre grupos, lo que dará lugar a mayores desviaciones estándar (Killip, Mahfoud, and Pearce 2004).

## A1.1 Efecto de diseño frente a CCI

Siempre que se informe de un efecto de diseño para un estudio, o cuando se utilice un efecto de diseño en una ecuación, los autores deben aclarar la ecuación subyacente utilizada. En la mayoría de los casos, cuando se trabaja con ecuaciones de tamaño de muestra, la ecuación subyacente es:

$$\text{Efecto de diseño} = 1 + ((m - 1)\rho) \quad (14)$$

donde  $\rho$  y  $m$  son el CCI y el número de individuos por grupo, respectivamente. Tenga en cuenta que el efecto de diseño depende de ambas variables, por lo que las ecuaciones que utilizan un efecto de diseño sin especificar  $m$  excluyen una información crítica.

Sin embargo, el CCI no depende del número de personas encuestadas, lo que lo hace más portátil en todos los diseños de encuestas (Stukel 2018a). Consulte a continuación para obtener más información sobre cómo calcular un CCI a partir de los datos de la encuesta. En el Apéndice 1.4 se presenta una lista de los valores de CCI ya calculados para determinados indicadores estándar de los principales donantes de proyectos de CRS.

Normalmente, se utilizan valores de referencia del CCI, pero es probable que el CCI aumente con el tiempo. Esto se debe a que las actividades de los proyectos suelen realizarse a nivel de grupo (por ejemplo, la formación de los maestros se realiza para todos los profesores de cada escuela, todos los agricultores de un grupo de productores, etc.). Por lo tanto, es probable que los resultados individuales se correlacionen más dentro de los grupos con el tiempo, después de recibir las intervenciones del proyecto. Dado que CRS ha estado operando en la mayoría de sus programas nacionales durante muchos años, es probable que un proyecto sea una continuación de otro, o una continuación de otra ONG, potencialmente en los mismos grupos. Por esta razón, esta guía fomenta el auto cálculo de los CCI a partir de datos recientes para indicadores similares en la misma área operativa. Además, Handa et al. (2018) han mostrado las diferencias de los CCI para el mismo indicador en todas las regiones geográficas, lo que pone aún más de relieve la importancia de utilizar los CCI específicos de un proyecto o país.

## A1.2 Cálculo del CCI

En esta sección se muestra cómo calcular un CCI de muestra incondicional (sin covariables).

El CCI de muestra es el cociente entre y dentro de la varianza del grupo:

---

*El efecto del diseño depende del CCI y del número de individuos por grupo.*

---

*Esta guía fomenta el auto cálculo de los CCI a partir de datos recientes para indicadores similares en la misma área operativa.*

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \quad (15)$$

Un valor CCI de 1 significaría que toda la variación en los datos podría explicarse por las diferencias entre los grupos, por lo que los topógrafos tendrían que visitar muchos grupos, pero solo un individuo por grupo. Un CCI cercano a 0 significaría que toda la variación de los datos podría explicarse por las diferencias entre los individuos, por lo que los encuestadores visitarían menos grupos, pero encuestarían a todos los individuos de cada uno.

### A1.2.1 Varianza general.

Al igual que en el caso anterior, las ecuaciones para calcular los CCI difieren entre indicadores continuos y binarios. Tenga en cuenta que la varianza para un indicador aleatorio discreto (continuo) se define como

$$\sigma_{\text{Continuo}} = \frac{\sum (y - \bar{y})^2}{n} \quad (16)$$

donde  $y$  es una observación individual,  $\bar{y}$  es la media de todas las observaciones, y  $n$  es el número de todas las observaciones. La varianza de un indicador binario es:

$$\sigma_{\text{Binario}} = p(1 - p) \quad (17)$$

Donde  $p$  es la proporción de la población para la que se cumple la condición. Tenga en cuenta que la varianza de un indicador binario es mayor cuando el indicador es verdadero para exactamente el 50 % de la población. Para ver esto, ingrese 0,35, 0,50 y 0,70 para  $p$ . Tenga en cuenta que las varianzas respectivas son 0,23, 0,25 y 0,21.

### A1.2.2 CCI para indicadores continuos

En el caso de los datos continuos, la varianza entre grupos se calcula de la siguiente manera:

$$\sigma_b^2 = \frac{\sum_c n_c (\bar{y}_c - \bar{y})^2}{C - 1} \quad (18)$$

Donde  $\bar{y}_c$  es la media de los datos a nivel individual del  $c^{\text{ésimo}}$  grupo.  $\bar{y}$  es la media de todas las observaciones (DataCamp 2019).<sup>24</sup>  $C$  es el número total de grupos. Por lo tanto, el numerador suma la varianza de cada grupo, ponderándola por el número de observaciones por grupo. El denominador se divide por el número de grupos, lo que proporciona la media entre la varianza de los grupos.

Del mismo modo, dentro del grupo la varianza se calcula como

$$\sigma_w^2 = \frac{\sum_{i1}^{I1} (y_{i1} - \bar{y}_1)^2 + \dots + \sum_{ic}^{Ic} (y_{ic} - \bar{y}_c)^2}{n - C} \quad (19)$$

Donde  $y_{i1}$  denota la observación para el individuo  $i$  en el grupo 1, y  $y_{ic}$  denota la observación para el individuo  $i$  en el grupo  $c^{\text{ésimo}}$ . Por lo tanto, el numerador adiciona la suma de las varianzas dentro de cada grupo. El denominador se divide por el número de observaciones (reducido por el número de grupos), proporcionando así la media dentro de la varianza de grupos.

<sup>24</sup> Para la varianza entre clústeres, utilice las observaciones ponderadas. Si las observaciones dentro de un grupo tienen diferentes pesos (p. ej., la muestra se estratificó por sexo), utilice las observaciones ponderadas. Tenga en cuenta que, si todas las observaciones dentro de un grupo tienen el mismo peso, el uso de los pesos de muestra no tiene ningún efecto sobre la varianza dentro del grupo.

Obsérvese que en el apéndice 2, la pestaña "CCI\_example" contiene un ejemplo elaborado para un indicador continuo (puntuaciones de alfabetización de segundo grado - Koinadugu, Sierra Leona), donde  $\rho = 0,92$ . Este mismo ejemplo puede encontrarse en el Apéndice 3 si los equipos de proyecto prefieren utilizar el software estadístico R.

Para calcular el CCI de los indicadores binarios, utilice un software estadístico más robusto que Excel.

### A1.2.3 CCI para indicadores binarios

Para los indicadores binarios que siguen la distribución de Bernoulli, se han propuesto una variedad de métodos (Schochet 2013; Goldstein, Browne, and Rasbash 2002 (GBR)). Para los indicadores binarios, se recomienda utilizar un software estadístico más robusto que Excel. El software estadístico R tiene una función incorporada, ICCbin()<sup>25</sup> en el paquete "aod" que calcula los CCI para los métodos A-C como se describe en Goldstein. El método A utiliza la distribución logística, que se usa a menudo cuando se analizan indicadores binarios, por lo que se recomienda. En el Apéndice 3 se presenta un ejemplo de código R, utilizando las mismas puntuaciones de alfabetización que para el ejemplo continuo, pero convirtiéndolas en un indicador binario (aprobado/reprobado). En la versión binaria,  $\rho = 0,60$ .

### A1.3 Cálculo de la desviación estándar

Como referencia, la ecuación para calcular la desviación estándar de una muestra (que solo se usa con indicadores continuos) es:

$$SD = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n - 1}} \quad (20)$$

La función "DESVEST.M" en Excel también se puede utilizar.<sup>26</sup>

Si calcula la desviación estándar de una muestra agrupada o estratificada, utilice la ecuación para una desviación estándar ponderada (The Statistical Engineering Division 1996):

$$SD_{\text{ponderada}} = \sqrt{\frac{\sum_i w_i (y_i - \bar{y}_w)^2}{(n' - 1) \frac{\sum_i w_i}{n'}}} \quad (21)$$

donde  $n'$  es el número de ponderaciones distintas de cero. En la práctica, todos los pesos serán  $\geq 1$ , por lo que  $n' = n$ .

Secciones 6.5.2.2 y 6.5.2.3 en Higgins, Li y Deeks (2019) proporciona orientación sobre el cálculo de las desviaciones estándar de los artículos de revistas que no informan de las desviaciones estándar, pero sí de las medias, los intervalos de confianza, los errores estándar y los valores p.

### A1.4 Valores CCI y desviación típica para determinados indicadores

En esta sección del apéndice se proporcionan los valores de la CCI y las desviaciones estándar de determinados indicadores de uso común o específicos de donantes cuyos datos subyacentes se recogen normalmente a través de una muestra representativa. Se mantiene separada de la guía principal, para permitir una actualización frecuente. Consulte la portada del Apéndice 1.4 para conocer la fecha de su actualización más reciente. Los lectores que deseen contribuir con los

En el Apéndice 1.4 se presentan los ICC y las desviaciones típicas de la mayoría de los indicadores estándar del Gobierno de los Estados Unidos.

<sup>25</sup> Asegúrese de usar la función ICCbin() del paquete "aod", y no el paquete "ICBin", ya que este último no parece funcionar.

<sup>26</sup> Si su interfaz Excel está en inglés, utilice la función STDEV.S.

valores CCI a las tablas del Apéndice 1.4, o que deseen ayuda para calcularlos, deben ponerse en contacto con el autor.

# Apéndice 2. Calculadora de tamaño de muestra – Excel

Este apéndice es una hoja de cálculo de Excel, separada de la guía principal. Contiene tablas listas para usar para las ecuaciones (1-8) de esta guía, además de un ejemplo de cálculo de un CCI continuo a partir de datos de muestra.

# Apéndice 3. Calculadora de tamaño de muestra – R

*Este apéndice contiene un código de R listo para usar para el cálculo del CCI de un indicador binario, así como un indicador binario con una covariable.*

```
## Binary Indicator ICC

#####
#Sets up file and reads in data                                     #
#####

#Clears all objects from memory
rm(list=ls())

#Sets the working directory
setwd()

library("readxl")
library("aod")

#Reads Excel file and names it
data <- read_xlsx("CRS_Samples_Annex2. Excel.xlsx", sheet = "ICC_Example", range = "A1:B475",
                 col_names = T)

#####
#Binary ICC - using built-in binary methods with no covariates     #
#####
#Probability of passing - from final evaluation data for 2nd graders

#Create the binary variable from the scores. In this context, a score of 4 or better was
#considered passing
Score_pass = data$LiteracyScore
Score_pass[Score_pass<4] = 0
Score_pass[Score_pass>0] = 1
data = cbind(data,Score_pass)

#To use the ICCbin function, the data must be grouped by numerator and denominator per
#each cluster, and not as individual-level observations
numerator.SchoolPass = aggregate(data$Score_pass,list(data$SchoolCode),sum)
numerator.SchoolPass = numerator.SchoolPass[, 'x']
denominator.SchoolPass = aggregate(data$Score_pass,list(data$SchoolCode),length)
denominator.SchoolPass = denominator.SchoolPass[, 'x']
data.SchoolPass = as.data.frame(cbind(numerator.SchoolPass,denominator.SchoolPass))

#Now, we can compute the ICC using Method A
Pass_Fail.A = iccbin(n = denominator.SchoolPass, y = numerator.SchoolPass,
                    data = data.SchoolPass, method = "A")

#end###
```

```
#####
#Binary sample size with one binary covariate - McConnell Equations 24-25      #
#####

#Calculate sample size with one with one covariate (gender) and treatment group only
#Convert data to binaries
data$PupilsGender[data$PupilsGender=='Boy'] = 0
data$PupilsGender[data$PupilsGender=='Girl'] = 1

#Calculate residual variance
fit = lm(Score_pass ~ PupilsGender,data = data)
resid = as.data.frame(residuals(fit))
data.cov = cbind(resid,data[, 'SchoolCode'])
names(data.cov) = c('resid', 'SchoolCode')

#Calculate total within group variance with gender covariate
y.cov_i_bar = aggregate(data.cov[, 'resid'],list(data.cov[, 'SchoolCode']),mean)
data.cov = merge(data.cov,y.cov_i_bar,by.x="SchoolCode",by.y = "Group.1")
colnames(data.cov)[ncol(data.cov)] = c('SchoolMeanRes')
data.cov = cbind(data.cov,(data.cov[, 'resid']-data.cov[, 'SchoolMeanRes'])^2)
colnames(data.cov)[ncol(data.cov)] = c('DevianceMeanRes_Sqrd')
Sum_Var.cov_withinCluster =
aggregate(data.cov$DevianceMeanRes_Sqrd,list(data.cov$SchoolCode),sum)
Total_Var.cov_withinCluster = sum(Sum_Var.cov_withinCluster[,2])/(n-g)

#Calculate between group variance with gender covariate
y.cov_bar = mean(data.cov$resid)
Var.cov_betweenCluster = sum(n_i[,2]*((y.cov_i_bar[,2] - y.cov_bar)^2))/(n-1)

#Calculate rho
rho.gender = Var.cov_betweenCluster/(Var.cov_betweenCluster+Total_Var.cov_withinCluster)

#Calculate sample size
#Ratio of treatment to control
pi = .5

#Normal cdf values
z_alpha = qnorm(0.05/2)
z_beta = qnorm(1-0.8)

#Values analyzed
x.0 = 0
x.1 = 1

#Probability of being a boy or girl in sample
theta.0 = 253/(221+253)
theta.1 = 221/(221+253)

#Probability of passing at baseline for boys (61%)/ girls (56%)
p0.0 = 154/253
p0.1 = 124/221

#Probability of passing at endline for boys (69%)/ girls (69%)
p1.0 = 175/253
```

```

p1.1 = 152/221

#Effect size for each of the covariate's values, where delta_q = p1q - p0q
delta.0 = p1.0 - p0.0
delta.1 = p1.1 - p0.1

#Overall effect size
delta = theta.0*(delta.0) + theta.1*(delta.1)

#M matrix
m1 = (pi*theta.0*p1.0*(1-p1.0) + (1-pi)*theta.0*p0.0*(1-p0.0)) +
      (pi*theta.1*p1.1*(1-p1.1) + (1-pi)*theta.1*p0.1*(1-p0.1))
m2 = pi*theta.0*p1.0*(1-p1.0) +
      pi*theta.1*p1.1*(1-p1.1)
m3 = x.0*(pi*theta.0*p1.0*(1-p1.0) + (1-pi)*theta.0*p0.0*(1-p0.0)) +
      x.1*(pi*theta.1*p1.1*(1-p1.1) + (1-pi)*theta.1*p0.1*(1-p0.1))
m4 = x.0*(pi*theta.0*p1.0*(1-p1.0)) +
      x.1*(pi*theta.1*p1.1*(1-p1.1))
m5 = x.0^2*(pi*theta.0*p1.0*(1-p1.0) + (1-pi)*theta.0*p0.0*(1-p0.0)) +
      x.1^2*(pi*theta.1*p1.1*(1-p1.1) + (1-pi)*theta.1*p0.1*(1-p0.1))
M = matrix(c(m1, m2, m3, m2, m2, m4, m3, m4, m5), nrow = 3, ncol = 3, byrow = T)

#g vector
g1 = theta.0*(p1.0*(1-p1.0) - p0.0*(1-p0.0)) + theta.1*(p1.1*(1-p1.1) - p0.1*(1-p0.1))
g2 = theta.0*(p1.0*(1-p1.0)) + theta.1*(p1.1*(1-p1.1))
g3 = x.0*theta.0*(p1.0*(1-p1.0) - p0.0*(1-p0.0)) + x.1*theta.1*(p1.1*(1-p1.1) - p0.1*(1-p0.1))
g = matrix(c(g1, g2, g3), nrow = 1, ncol = 3, byrow = T)

N_m.5 = ((g%%solve(M)%%t(g))*((z_beta + z_alpha)^2)/delta^2)*(1 + (5 - 1)*rho.gender))/2
A = (g%%solve(M)%%t(g))
m_range = matrix(c(5, 10, 15, 20, 15, 30))

deff.5 = (1 + (5 - 1)*rho.gender)
k.5 = ((deff.5/5)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.10 = (1 + (10 - 1)*rho.gender)
k.10 = ((deff.10/10)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.15 = (1 + (15 - 1)*rho.gender)
k.15 = ((deff.15/15)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.20 = (1 + (20 - 1)*rho.gender)
k.20 = ((deff.20/20)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.25 = (1 + (25 - 1)*rho.gender)
k.25 = ((deff.25/25)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.30 = (1 + (30 - 1)*rho.gender)
k.30 = ((deff.30/30)*A*(((z_beta + z_alpha)^2)/delta^2))/2

k_range = matrix(c(k.5, k.10, k.15, k.20, k.25, k.30))
Final_range = cbind(m_range,round(k_range,0),round(m_range*k_range,0))
colnames(Final_range) = c('# per cluster','# clusters','Total N')
#end###

```

# Apéndice 4. Descripciones formales de los cálculos del tamaño de la muestra

Esencialmente, al proporcionar estimaciones del tamaño de la muestra en cualquier documento, el lector debe tener suficiente información para poder recrear los cálculos si es necesario. Por lo tanto, tenga en cuenta si el proyecto utilizará un diseño agrupado, así como el tamaño de la muestra calculado para cada marco muestral que se encuestará (es decir, estudiantes y escuelas individuales; maestros; productores, etc.). Tenga en cuenta también el número de grupos y el número de individuos por grupo. Haga referencia a la ecuación utilizada de un documento de acceso público (una vez que esta guía esté disponible fuera de CRS, se puede citar). También aclare los valores de todos los parámetros utilizados ( $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\rho$ ,  $p$ ,  $SD$ , etc.) Puede ser más fácil presentar parte de esta información en una tabla. Vea a continuación un ejemplo de una propuesta reciente.

"Se utilizará un enfoque de muestreo por grupos en dos etapas para seleccionar a todos los encuestados para las encuestas cuantitativas (Tabla A4.1). En la primera etapa, las escuelas se seleccionarán al azar como grupos y luego se seleccionarán estudiantes, maestros, cocineros, cuidadores y madres (dentro de las respectivas comunidades que alimentan a las escuelas) en la segunda etapa. Dentro de cada escuela, también se entrevistará al director de la escuela. Las ecuaciones utilizadas para determinar el tamaño de la muestra generan el tamaño mínimo de la muestra necesario para detectar una diferencia estadística en los indicadores clave de resultados a lo largo del tiempo. Todas las muestras se incrementarán en al menos un 5 %, en caso de errores en la recogida de datos o en la transcripción.

Los tamaños de las muestras se calcularon utilizando las ecuaciones (6), (19) y (22) para resultados continuos agrupados, binarios no agrupados y binarios agrupados, respectivamente, en McConnell y Vera-Hernandez (2015), utilizando el nivel de potencia estándar del 80 % y el nivel de significación del 5 %. Los detalles específicos de los indicadores se indican en las notas a pie de página individuales. Algunos indicadores se convirtieron en porcentajes para que se puedan detectar cambios en las diferencias de medias a lo largo de la vida del proyecto."

**TABLA 5. EJEMPLO DE PRESENTACIÓN DE TAMAÑO DE MUESTRA**

INDICADOR	ENCUESTADO INDIVIDUAL	VALOR BASAL (ESTIMADO)	META (VIDA ÚTIL DEL PROYECTO)	CORRELACIÓN INTRACLASE (CCI)	GRUPOS	INDIVIDUALES POR GRUPO
Tasa promedio de asistencia de estudiantes <sup>27</sup>	Aulas	93 %	97 %	0,74	1.500	5
Porcentaje de estudiantes que demuestran que saben leer	Estudiantes	21 %	41 %	0,43 (Duflo, Glennerster, and Kremer 2007)	44	5
Porcentaje de cuidadores(as) que informan que dedican tiempo a actividades de alfabetización	Cuidadores	42 %	62 %	0,20 <sup>28</sup>	34	5
Porcentaje de niños(as) de 6 a 23 meses que reciben una dieta mínima aceptable	Madres de niños de 6 a 23 meses	67 %	79 %	0,08 (Moss et al. 2018)	56	5

<sup>27</sup> Este es un indicador nuevo para el USDA, y es difícil encontrar valores ICC publicados para él. Sin embargo, sobre la base de los cálculos del ICC a partir de los datos oficiales de asistencia del proyecto McGovern-Dole (Fase 4) de Sierra Leona, implementado por CRS, la variabilidad esperada se basa en gran medida en las escuelas, y no en las aulas dentro de las escuelas, de ahí el gran ICC. La desviación estándar relevante fue de 0.44. Dada la gran cantidad de muestras necesarias, STARS simplemente utilizará un censo de todas las aulas de todas las escuelas para este indicador para el estudio de referencia y revisará estos cálculos utilizando los datos de línea de base después de la recopilación.

<sup>28</sup> Dada la naturaleza personalizada de este indicador, es difícil encontrar un valor ICC publicado para él. Dado que se tratará de una práctica doméstica, se asume un valor de ICC de 0.20, que se encuentra entre los ICC identificados a nivel de la escuela y de la madre anteriores.

# Apéndice 5. Guía de referencia rápida

---

*Lo mejor es utilizar este apéndice después de familiarizarse con la guía principal.*

*Esta guía de referencia rápida pretende ser un recordatorio rápido de los pasos básicos en el proceso de cálculo del tamaño de la muestra. Es mejor usarlo después de revisar la guía completa y no cubre tantos escenarios posibles.*

**Paso 1:** Compilar una lista de indicadores que se medirán (la Tabla 6 proporciona un ejemplo). Para cada indicador, anote si es continuo o binario; el demandado; y el grupo del encuestado (si la recopilación de datos se agrupará). *Sección 1 de la guía.*

**Paso 2:** Añadir a la lista, para cada indicador, el cambio a medir. A menudo, este es el objetivo de la vida útil del proyecto menos el valor de referencia esperado, pero también se pueden esperar diferencias entre los grupos de control y tratamiento al final de un estudio, o diferencias entre los grupos al final de un proyecto.

Si no se van a realizar comparaciones, se debe tener en cuenta el margen de error dentro del cual se medirá el indicador. *Sección 2.2.2 de la guía.*

**Paso 3:** Añadir a la lista, para cada indicador cuya recopilación de datos se agrupará, el coeficiente de correlación intraclase (CCI). *Apéndice 1 de la guía.*

**Paso 4:** Agregue a la lista, para cada indicador continuo, la desviación estándar relevante. *Apéndice 1.3 de la guía.*

**Paso 5:** Para cada indicador, use la Figura 1 para determinar qué ecuación usar. Agregue esto a la lista. Tenga en cuenta que, si en el Paso 2 se determinó que no se medirá ningún cambio a lo largo del tiempo, utilice las ecuaciones (5-8) en lugar de las ecuaciones (1-4).

**Paso 6:** Utilizando la información anterior y la hoja de cálculo de Excel en el Apéndice 2, calcule el número mínimo de grupos (si corresponde) y encuestados por grupo. Agregue esto a la lista.

**Paso 7:** Si los encuestados (marcos muestrales) se superponen a los indicadores, utilice el tamaño de muestra más grande recomendado por tipo de encuestado.

**Paso 8:** Finalice los cálculos aumentando el tamaño de la muestra entre un 5 y un 20 % para tener en cuenta los errores de recopilación de datos, la deserción, etc. *Sección 3 de la guía.*

**TABLA 6. EJEMPLO DE LISTA PREPARATORIA PARA CÁLCULOS**

INDICADOR	CONTINUO O BINARIO	TIPO DE ENCUESTADO	TIPO DE GRUPO	CAMBIAR PARA DETECTAR	CCI	SD	ECUACIÓN	GRUPOS NECESARIOS	ENCUESTADOS NECESARIOS POR GRUPO	TAMAÑO FINAL DEL GRUPO (ENCUESTADO)
Tasa promedio de asistencia de estudiantes	Continuo	Aulas	Escuela	4 %	0,74	0,44	2	1.500	5	Censo <sup>29</sup>
Porcentaje de estudiantes que demuestran que pueden leer texto a nivel de grado	Binario	Estudiantes	Escuela	20 %	0,43	N/A	4	44	5	45 (6)
Porcentaje de personas que demuestran el uso de nuevas prácticas seguras de preparación de alimentos	Binario	Cocineros	Escuela	35 %	0,90	N/A	4	14	2	16 (2)
Porcentaje de cuidadores que informan haber dedicado tiempo a actividades de alfabetización con sus hijos en edad escolar en la semana anterior	Binario	Cuidadores	Escuela	20 %	0,20	N/A	4	34	5	35 (3)
Porcentaje de niños de 6 a 23 meses que reciben una dieta mínima aceptable	Binario	Madres de niños de 6 a 23 meses	Escuela	12%	0.08	N/A	4	56	5	56 (6)

<sup>29</sup> Tenga en cuenta que el proyecto atendió a menos de 1.500 escuelas, por lo que deberán hacer un censo de todas las aulas en todas las escuelas.

# Confidencial - Apéndice 6. Otras guías de tamaño de muestra

**El Apéndice 6 no debe compartirse fuera del CRS. Esto incluye a las organizaciones socias de CRS.**

Este apéndice se centra en otras guías de tamaño de muestra que no son de CRS. Se mantiene como un apéndice separado de la guía principal, ya que contiene información que es propiedad de CRS.

# Bibliografía

- Bence, James R. 1995. "Analysis of Short Time Series: Correcting for Autocorrelation." *Ecology* 76 (2): 628-639. <https://doi.org/10.2307/1941218>.
- Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317-372. <https://doi.org/10.3368/jhr.50.2.317>.
- Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge University Press.
- Catholic Relief Services. 2023. "Monitoring, Evaluation, Accountability and Learning (MEAL) Policies & Procedures." <https://www.crs.org/our-work-overseas/research-publications/monitoring-evaluation-accountability-and-learning-policies-procedures>.
- Chowa, Gina, David Ansong, and Mathieu R. Despard. 2014. "Financial Capabilities: Multilevel Modeling of the Impact of Internal and External Capabilities of Rural Households." *Social Work Research* 38 (1): 19-35. <https://doi.org/10.1093/swr/svu002>.
- Culligan, Mike, Leslie Sherriff, Clara Hagens, Guy Sharrock, and Roger Steele. 2019. *A Guide to the MEAL DPro: Monitoring, Evaluation, Accountability and Learning for Development Professionals*. Catholic Relief Services, Humentum, and the Humanitarian Leadership Academy (Downloaded from <http://mealdpro.org/> on July 25, 2019).
- DataCamp. 2019. "Inferential Statistics Course." Accessed July 9, 2019. <https://www.datacamp.com/community/open-courses/inferential-statistics>.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. *Using Randomization in Development Economics Research: A Toolkit*. Vol. 6059. *Discussion Paper Series*. London: Centre for Economic Policy Research.
- Frost, Jim. 2020. "Comparing Hypothesis Tests for Continuous, Binary, and Count Data." *Statistics By Jim*. Accessed 28 October 2020. <https://statisticsbyjim.com/hypothesis-testing/comparing-hypothesis-tests-data-types/#:~:text=Additionally%2C%20the%20samples%20sizes%20are,sizes%20can%20become%20quite%20large>.
- Goldstein, Harvey, William Browne, and Jon Rasbash. 2002. "Partitioning Variation in Multilevel Models." *Understanding Statistics* 1 (4): 223-231. [https://doi.org/10.1207/S15328031US0104\\_02](https://doi.org/10.1207/S15328031US0104_02).
- Handa, Sudhanshu, Thomas de Hoop, Mitchell Morey, and David Seidenfeld. 2018. "ICC Values in International Development: Evidence across Many Domains in sub-Saharan Africa." Centre for the Study of African Economics conference, United Kingdom.
- Hansen, Morris H., and Willimam N. Hurwitz. 1942. "On the Theory of Sampling from Finite Populations." *The Annals of Mathematical Statistics* 14 (4): 333-362. <https://doi.org/https://www.jstor.org/stable/2235923>.
- Higgins, JPT, T Li, and JJ Deeks, eds. 2019. *Cochrane Handbook for Systematic Reviews of Interventions*. Edited by JPT Higgins, J Thomas, J Chandler, M Cumpston, T Li, MJ Page and VA Welch. Version 6.0 ed, *Cochrane Handbook for Systematic Reviews of Interventions*: Available from [www.handbook.cochrane.org](http://www.handbook.cochrane.org).
- Horvitz, Daniel G., and Donovan J. Thompson. 1952. "A generalization of sampling without replacement from a finite universe." *Journal of the American Statistical Association* 47 (260): 663-685. <https://doi.org/10.2307/2280784>.
- Killip, Shersten, Ziyad Mahfoud, and Kevin Pearce. 2004. "What is an intracluster correlation coefficient? Crucial concepts for primary care researchers." *Annals of Family Medicine* 2 (3): 204-208.
- Lana, Milza M. 2012. "The effects of line spacing and harvest time on processing yield and root size of carrot for Cenourete® production." *Horticultura Brasileira* 30 (2): 7.
- Lee, Jin Young, and Gary Solon. 2011. "The fragility of estimated effects of unilateral divorce laws on divorce rates." *The BE Journal of Economic Analysis & Policy* 11 (1).
- McConnell, Brendon, and Marcos Vera-Hernandez. 2015. *Going beyond simple sample size calculations: a practitioner's guide*. Institute for Fiscal Studies.
- Moss, Cami, Tesfaye Hailu Bekele, Mihretab Melesse Salasibew, Joanna Sturgess, Girmay Ayana, Desalegn Kuche, Solomon Eshetu, Andinet Abera, Elizabeth Allen, and Alan D Dangour. 2018. "Sustainable Undernutrition

- Reduction in Ethiopia (SURE) evaluation study: a protocol to evaluate impact, process and context of a large-scale integrated health and agriculture programme to improve complementary feeding in Ethiopia." *BMJ Open* 8 (7). <https://doi.org/10.1136/bmjopen-2018-022028>.
- National Research Council. 2002. *Studies of Welfare Populations: Data Collection and Research Issues*. Edited by Michele Ver Ploeg, Robert A. Moffitt and Constance F. Citro. Washington, DC: The National Academies Press.
- Noggle, Eric. 2017. *The SILC Financial Diaries*. Catholic Relief Services (Baltimore, MD).
- Raab, Gillian, and Susan Purdon. 2009. "5.2.1 How the weights are calculated using population data." *Practical Exemplars on the Analysis of Surveys*. ReStore Project, National Centre for Research Methods (NCRM). Last Modified 5 July 2009. Accessed 22 April 2020. <http://www.restore.ac.uk/PEAS/nonresponse.php>.
- Schochet, Peter Z. 2013. "Statistical Power for School-Based RCTs With Binary Outcomes." *Journal of Research on Educational Effectiveness* 6 (3): 263-294. <https://doi.org/10.1080/19345747.2012.725803>.
- Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. "What Are We Weighting For?" *Journal of Human Resources* 50 (2): 301-316. <https://doi.org/10.3368/jhr.50.2.301>.
- Stukel, Diana Maria. 2018a. *Feed the Future Population-Based Survey Sampling Guide*. Food and Nutrition Technical Assistance Project, FHI 360 (Washington, DC. Downloaded from <https://www.fantaproject.org/sites/default/files/resources/FTF-PBS-Sampling%20Guide-Apr2018.pdf> on July 8, 2019; FHI 360).
- . 2018b. *Participant-Based Survey Sampling Guide for Feed the Future Annual Monitoring Indicators*. Food and Nutrition Technical Assistance Project, FHI 360 (Washington, DC. Downloaded from [https://www.fantaproject.org/sites/default/files/resources/Sampling-Guide-Participant-Based-Surveys-Sep2018\\_0.pdf](https://www.fantaproject.org/sites/default/files/resources/Sampling-Guide-Participant-Based-Surveys-Sep2018_0.pdf) on July 8, 2019; FHI 360).
- Sullivan, Lisa. 2019. "Power and Sample Size Determination." Accessed July 19, 2019. [http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704\\_Power/BS704\\_Power\\_print.html](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Power/BS704_Power_print.html).
- Technical Guidance for Monitoring, Evaluation, and Reporting for Emergency Activities*. 2022. USAID Bureau for Humanitarian Assistance (Washington, DC. Downloaded from [https://www.usaid.gov/sites/default/files/2022-05/BHA\\_Emergency\\_ME\\_Guidance\\_February\\_2022.pdf](https://www.usaid.gov/sites/default/files/2022-05/BHA_Emergency_ME_Guidance_February_2022.pdf) on December 18, 2023).
- The Statistical Engineering Division. 1996. DATAPLOT Reference Manual. Downloaded from <https://www.itl.nist.gov/div898/software/dataplot/refman2/ch2/weightsd.pdf> on July 25, 2019: National Institutes of Standards and Technology, Information Technology Laboratory.
- Thompson, Steven K. 2012. *Sampling*. Edited by Walter A. Shewhart and Samuel S. Wilks. Third ed. *Wiley Series in Probability and Statistics*. Hoboken, New Jersey: Wiley.
- Thomsen, Ib, Dinke Tesfu, and David A. Binder. 1986. "Estimation of Design Effects and Intraclass Correlations When Using Outdated Measures of Size." *International Statistical Review / Revue Internationale De Statistique* 54 (3): 343-349. <https://doi.org/10.2307/1403063>