

**Comisión Económica para América Latina y el Caribe
Organización de las Naciones Unidas
ONU - CEPAL**

**División de Estadísticas
Unidad de Estadísticas Sociales**



NACIONES UNIDAS



**Guía metodológica sobre el uso de las encuestas de hogares
en la estimación de indicadores educativos**

Convenio CEPAL – UNESCO

Santiago de Chile - septiembre de 2018

Tabla de contenidos

1. Introducción _____	3
2. Elementos básicos de las encuestas de hogares _____	6
3. Estimación y análisis de información utilizando las encuestas de hogares _____	14
4. Algunos indicadores educativos de interés _____	18
4.1. Tasa neta de asistencia _____	20
4.2. Tasa neta de asistencia ajustada _____	21
4.3. Tasa neta de asistencia total _____	23
4.4. Tasa de niños y adolescentes fuera de la escuela _____	24
4.5. Tasa de finalización _____	26
4.6. Tasa de participación en aprendizaje un año antes de la primaria _____	27
5. Conclusiones y recomendaciones _____	30
Referencias _____	33

1. Introducción

La Agenda 2030 para el Desarrollo Sostenible trajo consigo un compromiso internacional de cumplimiento de 17 Objetivos de Desarrollo Sostenible (ODS) y 169 metas adoptadas por los estados miembros de la Organización de las Naciones Unidas. En particular, el objetivo número cuatro (ODS 4) establece “el alcance de una educación inclusiva, equitativa y de calidad y la promoción de oportunidades de aprendizaje durante toda la vida para todos”. Este objetivo está compuesto por siete metas y tres medios de implementación a los cuales se deberá hacer seguimiento continuo y sistemático para garantizar en cada país que se cumpla con lo estipulado en el cuarto objetivo. A su vez, cada una de las metas está compuesta por una serie de indicadores que permiten monitorear el cumplimiento de la Agenda.

En particular, la meta 4.1 implica que “todos los niños y niñas completen la educación primaria y secundaria de forma gratuita, equitativa y de calidad”. Dentro de esta meta se han establecido un conjunto de indicadores que permiten el seguimiento periódico de la meta. Algunos indicadores son la tasa de finalización para la educación primaria, secundaria baja y secundaria alta; la tasa de niños y adolescentes fuera de la escuela para la educación primaria, secundaria baja y secundaria alta. De la misma forma, la meta 4.2 señala que “se debe asegurar que todos los niños y niñas tengan acceso a un desarrollo de calidad en su infancia temprana, cuidado y educación preprimaria para que estén listos para la educación primaria”. Un indicador que se ha planteado para monitorear esta meta es la tasa de participación en el aprendizaje organizado un año antes de la edad oficial de entrada a la primaria. Asimismo, tal como lo plantea UIS (2018), las tasas de asistencia en educación primaria, secundaria baja y secundaria alta son indicadores importantes y fundamentales para el seguimiento de las metas en el ODS 4.

En América Latina y el Caribe, existen tres fuentes de datos que proporcionan la información necesaria para el monitoreo de los indicadores; estas son los censos, los sistemas de información educativa (basados en registros administrativos escolares) y las

encuestas de hogares. En el primer caso, se trata de operaciones estadísticas que, aunque tienen bastante precisión, se presentan en periodos muy alejados de tiempo (comúnmente cada diez años) y que contemplan muy pocas variables educativas. Por otro lado, los sistemas de información educativa, aunque han adquirido un rol preponderante como fuente de datos para la producción de estadísticas dentro del sistema educativo, no cuentan con la cobertura necesaria para captar lo que sucede fuera de la escuela. Por último, las encuestas de hogares se realizan de forma más periódica y pueden incorporar diversas variables de seguimiento educativo.

Las encuestas de hogares son comúnmente implementadas por las oficinas nacionales de estadística (ONE) que son los órganos del gobierno que proveen estadísticas oficiales de tipo social, económico, ambiental o demográfico. A pesar de que pocas encuestas se centran en el tema educativo, sí es posible encontrar que, en la región, la mayoría de ellas contiene información sobre la asistencia al sistema educativo y los años de educación, entre otras variables. Dado que las encuestas de hogares son un instrumento útil para el seguimiento continuo de los indicadores sociales en los países, es posible utilizarlas también para realizar el seguimiento adecuado a algunos de los indicadores de la realidad educativa de los países. Los principios que sustentan la estimación de los indicadores en encuestas de hogares aseguran el insesgamiento de la inferencia, estableciendo un proceso robusto, representativo y consistente.

Dada la facilidad en la implementación de los procesos de estimación en las encuestas de hogares, es posible tener una mayor flexibilidad a la hora de ampliar los constructos que se abordan en las encuestas, además de las variables que componen sus módulos. De esta forma, y teniendo en cuenta la posibilidad de desagregar la información en algunas categorías de interés, las encuestas son un instrumento indispensable en el seguimiento a los ODS, al permitir analizar de manera conjunta los constructos educativos, sociales y económicos en un solo levantamiento. Aun cuando las encuestas de hogares tienen distintas tipologías (por ejemplo, encuestas de empleo, encuestas de ingresos y gastos, encuestas de condiciones de vida, encuestas de demografía y salud, entre otras), la mayoría de los países cuentan con encuestas que levantan información sobre varios constructos y que pueden ser categorizadas como encuestas de propósitos

múltiples. De hecho, es importante resaltar la importancia de este tipo de operaciones estadísticas al momento de estimar la cobertura de los sistemas educativos. Dado que la población de interés de las encuestas de hogares está definida por todos los habitantes de los hogares en cada país, es posible obtener estimaciones directas de la cobertura nacional y por desagregaciones territoriales; mientras que utilizar otras fuentes de información, como los sistemas de información educativa, requiere del involucramiento y análisis de otro tipo de información auxiliar, como proyecciones poblacionales, obteniendo una estimación indirecta cuyas bondades dependerán de la calidad de los registros administrativos y de la información auxiliar.

En resumen, las encuestas de hogares son una fuente de información valiosa para el cálculo de indicadores educativos, la particularidad de este tipo de estudios está en que al seleccionar un subconjunto de la población de interés es posible realizar inferencias precisas y exactas para todo el país, incluyendo algunas desagregaciones geográficas (como regiones, o ciudades importantes) y sociodemográficas (como sexo, edad, condición de actividad o ingresos).

De manera posterior a esta introducción, la sección dos presenta los elementos básicos con los que se diseñan las encuestas de hogares y provee los principios necesarios para que el lector pueda comprender de forma apropiada la naturaleza de las encuestas de hogares y la construcción de las variables de interés en torno al tema educativo. En la sección tres se contempla el análisis de los datos que provienen de las encuestas de hogares, haciendo énfasis en que las estadísticas provistas por este tipo de instrumentos tienen que entenderse en un contexto amplio que involucre, no solo la estimación puntual, sino también una medida de error de muestreo. La sección cuatro presenta un ejemplo detallado de algunos indicadores educativos que pueden ser calculados a partir de la información recopilada en las encuestas de hogares. Utilizando como ejemplo a Panamá, el lector podrá encontrar en detalle los pormenores asociados al cálculo de seis indicadores que permiten hacer seguimiento a las metas asociadas al ODS 4. Finalmente, se presentan algunas recomendaciones y conclusiones con respecto al uso de las encuestas de hogares en relación con los indicadores educativos.

2. Elementos básicos de las encuestas de hogares

2.1. Diseño de las encuestas

Las encuestas de hogares constituyen una investigación parcial de la población y están sujetas a un error de muestreo (a diferencia de los censos que examinan toda la población de interés). En la planeación de las encuestas, este error de muestreo juega un papel fundamental puesto que debe ser minimizado haciendo uso eficiente de los recursos disponibles para la realización de la encuesta. En esta búsqueda de la mejor estrategia para recolectar la información primaria en la encuesta surgen varios elementos que deben ser resaltados para poder analizar de forma apropiada la información obtenida.

Los fundamentos teóricos de los procesos de estimación en encuestas por muestreo pueden ser encontrados en Kish (1965) y en Gutiérrez (2016). Para que la inferencia que se realiza desde la encuesta sea precisa y exacta, los hogares en la muestra deben ser seleccionados de forma aleatoria para evadir cualquier sesgo de selección. Para ello, es necesario tener acceso a un dispositivo que permita identificar y ubicar a todos y cada uno de los hogares del país. De esta forma cada hogar tendrá una probabilidad de inclusión en la muestra diferente de cero, permitiendo de esta forma que todos los hogares puedan conformar la muestra y ninguno se quede por fuera de la participación en la selección. Si bien no todos los hogares serán parte de la muestra, todos tendrán un chance de participación en la selección.

Este dispositivo que enlista los hogares recibe el nombre de marco muestral y es un elemento indispensable para poder realizar cualquier encuesta probabilística. Ahora bien, dada la dinámica misma de los hogares, no existe una mecánica de enlistamiento exhaustivo de todos los hogares de un país y, por lo tanto, la consecución del marco de muestreo debe ser hecha por etapas, de forma tal que el principio de exhaustividad en la selección de los hogares se cumpla y ninguno sea excluido del sorteo probabilístico. Utilizando la información que viene de los censos, es común que las oficinas nacionales de estadística establezcan esquemas de enumeración en sus territorios (rurales y urbanos) que contemplen la agregación cartográfica de viviendas. Estas unidades de

enumeración se conocen como sectores censales y, en la mayoría de los casos, corresponden a bloques o manzanas de viviendas en el sector urbano y agrupaciones de viviendas o unidades productivas, en el sector rural. Debido a que estas enumeraciones se realizan de forma exhaustiva (puesto que vienen de los censos), es posible usarlas para la selección de los hogares.

Si se consideran todos los sectores censales, en una primera etapa de muestreo, para posteriormente seleccionar hogares dentro de cada sector incluido en esa primera etapa, entonces se configura un esquema de selección (en dos etapas) que le asigna un chance de inclusión a todos los hogares del país en el sorteo probabilístico. La anterior es una manera de sortear las dificultades encontradas al no tener un marco de hogares de forma explícita. En adelante denominaremos a los sectores censales como Unidades Primarias de Muestreo (UPM), puesto que conforman la primera etapa de muestreo. Nótese que, una vez se ha realizado la selección de UPMs, es necesario realizar un proceso de empadronamiento de hogares dentro de aquellas UPMs que fueron seleccionadas en la muestra. Este proceso permitirá enumerar y enlistar a todos los hogares de las UPMs para posteriormente realizar la selección de los hogares en una segunda etapa de muestreo.

Otra consideración importante acerca del esquema del levantamiento de información en encuestas de hogares es que en la mayoría de los casos el muestreo de hogares es auto-ponderado y estratificado, lo que significa que la selección de hogares se hace de manera independiente en la población en subgrupos de interés, que tienen una connotación geográfica en la mayoría de los casos; por ejemplo, una estratificación común es la partición de la población en regiones urbanas y rurales. Dentro de estos estratos a cada UPM se le asigna una probabilidad de inclusión proporcional al número de viviendas o habitantes. Es decir, aquellas UPM más grandes, en las que existe una mayor densidad poblacional, tendrán un mayor chance de ser seleccionadas.

Una vez que las UPM son seleccionadas en la primera etapa de muestreo, es posible considerar una variedad de posibilidades al momento de ejecutar la selección de hogares en la segunda etapa de muestreo. Por ejemplo, podría ser posible realizar una medición a todos los hogares de las UPMs seleccionadas, o incluso seleccionar un sólo hogar en

cada una de las UPMs. Lo que se acostumbra a usar en los levantamientos de la segunda etapa de muestreo es seleccionar una submuestra de hogares en cada UPM. Es más, para que sea más expedito supervisar y sistematizar la logística del operativo de campo en cada UPM seleccionada se incluye un número fijo de hogares, sin importar el tamaño de la UPM. La selección de los hogares se realiza asignándole la misma probabilidad a todos los hogares dentro de cada UPM.

Este esquema de selección en las encuestas de hogares crea un escenario adecuado para minimizar el error de muestreo que en conjunción con el cálculo de un tamaño de muestra apropiado de UPMs y hogares hacen que la inferencia en las encuestas de hogares sea insesgada y precisa y se puedan tomar las mejores decisiones con base en las cifras resultantes de su análisis. El tamaño de muestra en encuestas de hogares suele ser suficiente para realizar inferencias en los niveles de desagregación planeados. Cada dominio de estudio requiere de un análisis detallado en cuanto a su tamaño de muestra y la agregación de estos cálculos redundante en una cifra que por lo general reúne a miles de hogares en un despliegue logístico de grandes proporciones.

Por último, un aspecto importante que es necesario resaltar es que, ante la ausencia de un marco de muestreo exhaustivo sobre todos los hogares, el tamaño de muestra se debe aumentar para que la precisión del análisis estadístico no se vea afectada por el efecto de homogeneidad que se puede encontrar en algunas variables de interés en poblaciones aglomeradas. Por ejemplo, si el interés se centra en la tasa de asistencia escolar, es muy probable que en una misma UPM los hogares que la componen tengan un comportamiento estructural homogéneo y que todos los niños de la población tengan hábitos de asistencia escolar similares. Lo anterior se da porque existe un efecto de aglomeración sociodemográfica que sucede de forma natural; de esta manera, las familias con condiciones socioeconómicas más favorables se aglomeran entre ellas en las UPMs y es posible encontrar bajas tasas de niños y adolescentes fuera de la escuela en todos aquellos hogares. De forma similar ocurre con las familias vulnerables menos favorecidas cuyos niños trabajan; estas familias se aglomeran en las UPMs y de esta forma es posible encontrar altas tasas de inasistencia escolar.

Como se mencionó anteriormente, ante el panorama de homogeneidad dentro de las UPMs, un camino parsimonioso en la asignación de los tamaños de muestra a los estratos consiste en seleccionar bastantes UPMs para asegurar una mejor dispersión de la realidad de la variable de interés y en cada UPM seleccionada hacer una medición de unos pocos hogares. Este efecto de aglomeración es diferente para cada variable del cuestionario, pues la estructura de homogeneidad no es idéntica para indicadores de pobreza, que para indicadores de inasistencia escolar. De esta forma, y ante la realidad de medición multivariada de las encuestas de hogares, que intentan indagar por varios constructos de investigación a la vez, es posible calcular de forma directa el impacto de la homogeneidad sobre el tamaño de muestra. Este cálculo se realiza de forma directa por los softwares computacionales y se conoce con el nombre de DEFF o efecto de diseño, el cual representa el coeficiente de deflación sobre el tamaño de muestra efectivo en la variable de interés. La siguiente ecuación representa este comportamiento

$$n_{eff} = n/DEFF$$

Por lo anterior, si el efecto de diseño DEFF es mayor que uno, lo cual sucede en la mayoría de los casos en los que la estructura de homogeneidad es alta dentro de las UPMs, entonces el tamaño de muestra necesario n_{eff} para realizar inferencias precisas para la variable de interés se ve reducido en relación con el tamaño de muestra original n . Por ejemplo, es posible que para una desagregación de interés el tamaño de muestra de la encuesta sea de $n = 200$ personas y que al indagar por una característica de interés en particular se haya calculado un efecto de diseño de $DEFF = 4$. Esto quiere decir que el tamaño de muestra efectivo para realizar algún tipo de inferencia será de tan sólo $n_{eff} = 50$ unidades, lo que podría ser muy bajo para sustentar una inferencia precisa y confiable.

2.2. Construcción de variables de interés en las encuestas de hogares

A pesar de que las encuestas de hogares en la región son disímiles en su estructura y contenido, es posible estandarizar la metodología de estimación de indicadores sociales

y educativos, puesto que existen elementos en común en los instrumentos que se utilizan para levantar la información. Es así como los cuestionarios de las encuestas de hogares contienen algunas variables fundamentales a la hora de estimar indicadores educativos, como por ejemplo la asistencia al sistema educativo y sus niveles, el número de años de estudio completados por los individuos, entre otras.

Con respecto a la asistencia escolar, la mayoría de los cuestionarios indagan acerca de este evento por medio de una pregunta simple. Por ejemplo, *¿está matriculado para recibir clases este año en un centro educativo?*, o *¿asiste actualmente a un centro educativo?*, o *¿en qué nivel se matriculó este año?*, *¿asiste a la escuela actualmente?* Las anteriores preguntas están encaminadas a determinar si el individuo es parte o no del sistema educativo del país. Por otro lado, el nivel y grado de asistencia al sistema educativo puede ser captada de forma altamente desagregada; por ejemplo, la siguiente ilustración presenta las posibilidades de respuesta que se consideran en la Encuesta de Mercado Laboral (EML) de Panamá, levantada por el Instituto Nacional de Estadística y Censo de Panamá.

<i>¿Qué nivel y qué grado o año escolar más alto aprobó?</i>	
Ningún grado	A
Pre-kínder o prejardín	B
Kínder o jardín	C
Enseñanza especial	D
Primaria	E
Vocacional	F
Secundaria	G
Superior no universitaria	H
Superior universitaria	I
Especialidad (Postgrado)	J
Maestría	K
Doctorado	L

Ilustración 1. Nivel y grado escolar más alto aprobado en la EML de Panamá.

Los cuestionarios de las encuestas de hogares de los países permiten indagar acerca del último nivel educativo alcanzado por el respondiente. A través de preguntas tales como *¿cuál es el último nivel educativo alcanzado?*, o *¿cuál es el último año o grado aprobado?*, o *¿cuál es el nivel educativo más alto que alcanzó?*, o *¿a qué nivel corresponde ese último año o curso que aprobó?*, es posible determinar la cantidad de años de estudio que la persona ha acumulado a lo largo de su vida. En algunos casos, la variable de asistencia se combina junto con las posibles categorías del sistema educativo del país. Luego, para hacer comparable esta variable con otros cuestionarios de la región, se puede colapsar las categorías apropiadas de la siguiente forma: a) *No aplicable*, b) *Asiste*, c) *No asiste*. Esta transformación se utiliza sistemáticamente para todos los países de la región.

A pesar de que los cuestionarios de las encuestas de hogares no han estandarizado la forma de preguntar por la asistencia escolar o el último nivel alcanzado, en todas las encuestas se registra el último grado de educación alcanzada en determinado nivel. De esta manera, el encuestador registra, no sólo el nivel (primaria, secundaria, terciaria o posgrados), sino el grado o año dentro de ese nivel. Luego, con la información recolectada recuperar el último nivel educativo alcanzado, es posible combinar la variable colapsada con los años de estudio y con la información que se obtenga en cada país acerca de la edad de entrada al sistema educativo y la duración de los ciclos propedéuticos.

En particular, el conteo de los años de estudio empieza a partir de la edad inicial de entrada a primaria en cada país. Como ejemplo particular, si la primaria tiene una duración de seis años en un país, entonces la combinación de nivel y años dará como resultado que la variable de años de estudio en este nivel varíe de 1 a 6 años; si la secundaria tiene una duración de cinco años, la variable de años de estudio en este nivel variará de 7 a 11 años; dependiendo del nivel de educación terciaria, la variable de años de estudio podrá variar de 12 a 19 años, y así sucesivamente. Nótese que, para este ejemplo particular, la variable de años de estudio tomará un valor nulo para todos los individuos menores de 6 años o para los individuos mayores de seis años que nunca hayan asistido al sistema educativo. Por supuesto, en cada país existen diferentes

edades de entrada a la educación primaria, diferentes niveles y ciclos educativos, y diferentes duraciones por nivel; por ende, la construcción de esta variable se debe realizar de forma particular.

En resumen, un insumo fundamental para el cálculo de los indicadores educativos tiene que ver con la información, particular para cada país, de la edad en la que los niños comienzan su formación educativa y la duración de cada ciclo propedéutico. La siguiente tabla resume la duración de los ciclos y las edades para los países miembros del Sistema de la Integración Centroamericana (SICA), según la Clasificación Internacional Normalizada de la Educación (CINE 2011). Por ejemplo, para el caso específico de Panamá, la edad en que los niños deben empezar a estudiar es de seis años y la duración del ciclo primario es de seis años, por ende, la edad en la que deben empezar el ciclo de la secundaria baja, cuya duración es de tres años, es a los doce años. La edad en la que deben comenzar el ciclo de secundaria alta, que también tiene una duración de tres años, será a los 15 años.

Tabla 1. Edades de entrada y duración de los ciclos educativos en los países del SICA

País	Edad inicial	Duración de la primaria	Edad de entrada - secundaria baja	Duración de la secundaria baja	Edad de entrada - secundaria alta	Duración de la secundaria alta
Costa Rica	6	6	12	3	15	2
República Dominicana	6	6	12	2	14	4
El Salvador	7	6	13	3	16	2
Guatemala	7	6	13	3	16	2
Honduras	6	6	12	3	15	2
Nicaragua	6	6	12	3	15	2
Panamá	6	6	12	3	15	3
Belice	5	6	11	4	15	2

Por último, vale pena mencionar que, además de estas variables relacionadas con la asistencia al sistema educativo, también se pregunta por otras variables sociodemográficas que serán utilizadas para realizar desagregaciones de interés. Por ejemplo, la variable edad (en número de años cumplidos) es medida en todas las encuestas de la región, así como la variable área (urbano o rural) y sexo que también está presente en las encuestas. Por otro lado, para poder hacer desagregaciones más complejas; por ejemplo, a nivel de quintiles de ingreso, es común que la encuesta tenga un módulo de medición de los ingresos personales y del hogar, para luego crear una variable derivada que represente el ingreso per cápita de los individuos del hogar. Esta variable final se define como el promedio por persona de la suma de todos los ingresos personales (sueldos y salarios, ganancias de independientes e ingresos del empleo, jubilaciones, pensiones y transferencias) y del hogar (ingresos de negocios, alquileres y rentas).

3. Estimación y análisis de información utilizando las encuestas de hogares

Sin tener en cuenta los principios fundamentales del diseño de muestreo de las encuestas sería imposible realizar un análisis apropiado de los datos provenientes de las encuestas de hogares puesto que la forma en la que son recolectadas las observaciones dista mucho de los supuestos básicos de la teoría estadística inferencial clásica. En primer lugar, no existe independencia entre las observaciones puesto que, al utilizar los esquemas de estratificación y aglomeración mediante UPMs, la probabilidad de selección de un individuo en un hogar depende de la inclusión de los otros individuos del hogar. De hecho, si un hogar es seleccionado en la muestra, se realiza una medición exhaustiva de todos los miembros del hogar. En segundo lugar, la forma de selección de los hogares en la muestra hace que la probabilidad de inclusión de los hogares en las UPMs no sea idéntica. Con esto en mente, es necesario incorporar algunos elementos de estimación que contemplan estas dificultades.

Como lo menciona Bethlehem (2009), el esquema de inferencia de las encuestas de hogares se basa en la inferencia inductiva, que va de lo particular a lo general, y en el principio de representatividad que afirma que un hogar que es seleccionado en una muestra se representa a sí mismo y a otros más en la población. El principio de representatividad está plasmado en el factor de expansión de la encuesta. Esta es una cantidad asociada a cada unidad observada en la muestra que es utilizada para ponderar las observaciones y realizar una inferencia precisa a nivel poblacional que describa la realidad del país o de las regiones y desagregaciones de interés que lo componen.

El factor de expansión de la encuesta provee los elementos necesarios para calcular las estimaciones requeridas de manera insesgada puesto que se define de manera inversamente proporcional a la probabilidad de inclusión de la unidad en la muestra. Sin embargo, este factor de expansión también contempla algunos ajustes que se deben realizar debido a la desactualización del marco de muestreo. Por ejemplo, se contempla la exclusión de aquellas estructuras que no contienen hogares y que no deberían ser

elegibles (como edificios de oficinas, etc.); también el ajuste sobre las unidades cuya elegibilidad se desconoce (como aquellas casas en las que no se encontró a ningún respondiente); la reponderación causada por la ausencia de respuesta como una realidad cada vez más frecuente en este tipo de investigaciones y por último; la calibración sobre proyecciones poblacionales demográficas actualizadas. Una vez ajustado el factor de expansión, este es utilizado directamente para calcular las estimaciones en la población.

El cómputo de las estimaciones da como resultado un número real. Sin embargo, este número no debe ser considerado para efectos de publicación, o de toma de decisiones de política pública, si no se ciñe a los estándares de precisión con los cuales la encuesta fue diseñada. Para evaluar la calidad, precisión y confiabilidad de las estimaciones es necesario acompañar la estimación con su correspondiente error de muestreo inducido por la varianza del estimador. Debido a las particularidades de los diseños de muestreo en encuestas de hogares, la varianza de este promedio ponderado está relacionada con el efecto de diseño, que actúa como un factor de inflación en la dispersión de los estimadores por muestreo. Como usualmente el efecto de diseño DEFF es mayor que uno, entonces la varianza asociada con el estimador por muestreo será más grande y por consiguiente la inferencia será menos precisa puesto que el error de muestreo será más grande y los intervalos de confianza serán más amplios.

Debido a que las estimaciones provenientes de las encuestas están sujetas al error de muestreo por las diferencias inherentes entre la población real y la expansión de la muestra, es necesario tener una estimación de este potencial error de muestreo sobre el indicador específico. Una forma de estimar este error es mediante el error estándar que es utilizado para construir intervalos de confianza. Por ejemplo, bajo condiciones de regularidad en la encuesta, la estimación del promedio del indicador más o menos 1.96 veces el error estándar conlleva a un intervalo de confianza del 95%, que implica que el valor real del indicador (hipotéticamente calculado con todos los datos de la población) estará contenido dentro de estos límites el 95% de las veces. Si la amplitud del intervalo de confianza es muy grande, la estimación del indicador perderá toda precisión puesto

que se genera más incertidumbre. De la misma forma, si el intervalo de confianza es angosto, entonces esta estimación se considera precisa y con menos incertidumbre.

Por ejemplo, considere que el indicador de interés es la tasa de finalización de la primaria básica y que mediante el uso de los datos provenientes de encuestas de hogares se ha generado una estimación puntual del 65%. Si el intervalo de confianza de esta estimación estuviera entre 64% y 66%, diríamos que este es un escenario que genera una muy alta confianza y precisión. En cambio, si el intervalo de confianza de la estimación estuviera entre 40% y 90%, no podríamos concluir con certeza qué está ocurriendo en el sistema educativo del país, puesto que el indicador podría ser bajo (cercano al 40%) o alto (cercano al 90%) y no tenemos suficiente precisión. Es por esto por lo que la amplitud del intervalo de confianza debe ser considerada como una medida de la calidad de las estimaciones. Nótese que entre más crezca el DEFF, más amplio será el intervalo de confianza puesto que se tendrán menos observaciones efectivas y la incertidumbre crecerá. Lo anterior también ocurre cuando se intentan obtener estimaciones más desagregadas, puesto a medida que aumenta el nivel de profundidad en la inferencia, se pierden más observaciones y aumenta la incertidumbre en la cifra puntual.

Una medida de variabilidad absoluta que resume la precisión y exactitud de una estimación es el coeficiente de variación, cuyos insumos principales son el error de estándar y la estimación puntual. Esta medida se acostumbra a utilizar en encuestas de hogares para decidir acerca de la publicación o no de una cifra. Para los indicadores educativos de tipo proporción o tasas, el coeficiente de variación tiene un valor mínimo de cero. Aunque no existe un consenso generalizado, es muy frecuente encontrar que las oficinas nacionales de estadística supriman la publicación de las estimaciones puntuales cuyo coeficiente de variación sea mayor al 20%.

Debido a que los indicadores educativos que contempla esta guía son de tipo tasa o proporción, es necesario considerar algunos aspectos técnicos que definen la precisión de las estimaciones puntuales. En primer lugar, los intervalos de confianza no pueden contener valores que sean negativos o mayores a 100%. La razón de lo anterior es que, en ningún caso, una proporción puede tomar valores negativos ni mayores a 100%. Además de lo anterior, el cálculo del coeficiente de variación se realizó teniendo en

cuenta una medida robusta y simétrica que no se deja afectar por valores de proporciones extremas (de incidencia nula o total). De esta forma, las proporciones cercanas a cero o 100% tendrán un coeficiente de variación más alto, mientras que las cercanas al 50% tienen un coeficiente de variación menor. En general, se necesita de este tipo de medidas robustas en investigaciones que producen estimaciones de proporciones muy bajas o cercanas al 100%. Por ejemplo, el análisis de encuestas de hogares de demografía y salud que tratan constructos de baja incidencia, como dependencia a sustancias psicoactivas o incidencias de enfermedades raras, utiliza este tipo de medidas. El lector interesado puede encontrar la base teórica de estas medidas de variación en los trabajos de Choy et al. (1998), Barnett-Walker et al. (2003) y Parker, Talih, y Malec (2017).

4. Algunos indicadores educativos de interés

En esta sección se describe la forma de estimación de algunos indicadores de asistencia escolar, finalización de los ciclos educativos y participación en el sistema. Para fines de demostración, los cálculos se realizan utilizando datos de la Encuesta de Mercado Laboral de Panamá, de agosto de 2014, levantada por el Instituto Nacional de Estadística y Censo de Panamá. La principal finalidad de esta encuesta es la medición de los cambios del mercado laboral, por lo que se indaga por la condición de actividad económica, ocupación, lugar de trabajo e ingresos, aunque también se abordan algunas preguntas de investigación que están ligadas a la estimación de los indicadores educativos. La representatividad de la encuesta es nacional y permite desagregaciones por provincias y áreas (urbana y rural).

Como se mencionó anteriormente, en las encuestas de hogares se pregunta por la edad, la asistencia escolar y el nivel de instrucción. Además de lo anterior, la encuesta reporta el sexo, el área y, los ingresos. Adicionalmente, la base de datos de la encuesta presenta el factor de expansión de los individuos. Con estas variables es posible configurar las desagregaciones de interés que son requeridas para el cálculo de los indicadores.

A continuación, se presenta de manera detallada el proceso de estimación que se debe surtir para obtener estimaciones de seis indicadores educativos que pueden ser estimados a partir del uso apropiado de las encuestas de hogares. De manera general, todos los indicadores que se abordan en este documento tienen la forma de un promedio ponderado, de la siguiente manera:

$$\hat{R}_d = 100 * \frac{\sum_{k \in S_d} w_k y_k}{\sum_{k \in S_d} w_k}$$

En donde \hat{R}_d hace referencia al estimador calculado sobre el dominio de interés d , la suma $\sum_{k \in S_d}$ abarca a todos los individuos k en la muestra del dominio de interés S_d , w_k se refiere al peso de muestreo del individuo k y, finalmente y_k es la variable de interés del individuo k .

La anterior expresión contiene algunos elementos que se deben considerar detenidamente. En primer lugar, nótese la presencia constante del factor de expansión w_k en cada componente del estimador, que además está supeditado a una desagregación de interés d . Es decir, es específico para un subgrupo poblacional, que puede estar constituido de manera geográfica, como por ejemplo a nivel de una ciudad, región o país. De la misma manera, esta desagregación puede también estar constituida por algún cruce sociodemográfico de interés. En general las desagregaciones que se consideran son quintiles de ingreso, sexo, área y todos sus posibles cruces. Como los diversos indicadores de interés están supeditados a diferentes niveles educativos, entonces los subgrupos de interés estarán restringidos también por la edad de los individuos, según como sea medida cada nivel educativo.

Por último, se aclara que la variable de interés y_k es observada únicamente para los individuos que pertenecen a la muestra s_d ; es decir, para los individuos que pertenecen a la desagregación de interés, en el grupo de edad de interés y que hacen parte de los hogares de la muestra de UPMs en cada uno de los estratos. Para los indicadores considerados en este documento, esta variable es siempre dicotómica (presencia o ausencia de una característica educativa sobre un individuo del hogar) y su construcción depende de cada indicador que se quiera estimar.

Como se mencionó anteriormente, un insumo fundamental para la estimación de algunos indicadores educativos es la edad del individuo. En algunas ocasiones, esta medición debe contrastarse con la fecha del año de ingreso al sistema escolar que el país ha definido. Sin embargo, este contraste constituye un reto metodológico a la luz de la variedad de periodos de estudio permitidos en la región. Es así como para algunos países el inicio del año escolar coincide con el inicio del año calendario, pero en otros no es así. Por lo anterior, siguiendo las recomendaciones de UIS (2016a), es necesario realizar un tratamiento a los datos de las encuestas de hogares para poder contrastar la edad real del individuo con el inicio del año escolar. Por ejemplo, si el periodo de recolección de información de la encuesta tuvo lugar en septiembre y se reporta que un niño tiene 6 años en ese momento, existe una muy alta probabilidad de que la edad del niño al iniciar el año escolar, es decir en febrero, fuera de cinco años; de esta forma este niño no sería

elegible para comenzar la educación primaria en ese año. Si la encuesta pregunta por la fecha exacta de nacimiento de los individuos, esta variable podría ser usada para determinar la edad real. Una discusión más detallada acerca de las diferentes variantes que se pueden abordar para disminuir el efecto de la edad reportada al momento del levantamiento de la encuesta y su posible disparidad con el inicio y duración de los años escolares, puede encontrarse en UIS (2016b).

4.1. Tasa neta de asistencia (*net attendance rate*)

La tasa neta de asistencia representa el porcentaje de niños, en un grupo específico de edad para un determinado nivel educativo, que asisten a ese nivel durante el año académico de referencia. Su forma de cálculo es la de un promedio ponderado que divide el número estimado de individuos en un grupo de edad que atienden un determinado nivel de educación, por el número estimado de individuos en ese grupo de edad.

En este caso, existe un filtro natural para cada uno de los indicadores de los niveles educativos. De esta forma, la edad será considerada para formar el grupo de interés con respecto al nivel educativo específico. De esta forma, para calcular la tasa neta de asistencia en educación primaria, se establece el grupo de interés como aquellos individuos que tienen entre 6 y 11 años; en educación secundaria baja, entre 12 y 14 años; en educación secundaria alta, entre 15 y 17 años. De la misma forma, la variable de interés toma valores de acuerdo con el nivel de educación y a la asistencia del individuo al sistema educativo. Para primaria, secundaria baja y secundaria alta la variable de interés toma el valor de uno, si asiste a estos niveles educativos, respectivamente.

La siguiente tabla muestra la estimación de la asistencia neta para primaria en Panamá, calculada para los niños que tienen entre 6 y 11 años, como porcentaje de aquellos que sí asisten a la escuela. Este indicador se estima para los niveles de desagregación mostrados: sexo, área urbana y quintiles de ingreso. Además, la tabla presenta el tamaño de muestra observado, como conteo de individuos en el grupo de interés, el tamaño de muestra efectivo, que está corregido por efecto de diseño (DEFF), el límite inferior del

intervalo de confianza (Li), el límite superior del intervalo de confianza (Ls) y el coeficiente de variación (cv). Las demás tablas que se mostrarán en este documento tienen esta misma estructura. Nótese que, las estimaciones marcadas con un asterisco se consideran poco precisas y no deberían ser tenidas en cuenta para propósitos de toma de decisiones. En particular, no son confiables las estimaciones de la tasa de asistencia para el primer quintil de los hombres en el área urbana, ni para el quinto quintil de las mujeres en el área urbana.

Tabla 2. Estimaciones de la tasa neta de asistencia para primaria en Panamá.

Sexo	Área	Quintil	Tamaño de muestra	Tamaño de muestra efectivo	DEFF	Tasa	Li	Ls	cv	Marca
Hombre	Urbana	Primero	456	54	8.45	0.96	0.86	0.99	0.20	*
Hombre	Urbana	Segundo	329	316	1.04	0.98	0.96	0.99	0.10	
Hombre	Urbana	Tercero	237	212	1.12	0.98	0.95	0.99	0.13	
Hombre	Urbana	Cuarto	170	178	0.96	0.98	0.95	0.99	0.13	
Hombre	Urbana	Quinto	104	105	0.99	0.97	0.92	0.99	0.16	
Mujer	Urbana	Primero	425	165	2.57	0.96	0.92	0.98	0.12	
Mujer	Urbana	Segundo	270	310	0.87	0.99	0.98	1.00	0.15	
Mujer	Urbana	Tercero	210	143	1.47	0.98	0.93	0.99	0.14	
Mujer	Urbana	Cuarto	172	-	-	1.00	-	-	-	
Mujer	Urbana	Quinto	86	-	-	1.00	-	-	-	*

4.2. Tasa neta de asistencia ajustada (*adjusted net attendance rate*)

La tasa neta de asistencia ajustada representa el porcentaje de individuos, en un grupo específico de edad que asistieron a la educación primaria o secundaria durante el año académico de referencia. Su forma de cálculo es la de un promedio ponderado que divide el número estimado de individuos, en un grupo de edad específico, que asistieron a la educación primaria o secundaria, por el número estimado de individuos en ese grupo de edad.

Al igual que en el caso anterior, la edad define el grupo de interés con respecto al nivel educativo específico. Para primaria, se establece el grupo de interés entre 6 y 11 años; en educación secundaria baja, entre 12 y 14 años; en educación secundaria alta, entre 15 y 17 años. La variable de interés toma valores de acuerdo con el nivel de educación alcanzado y a su asistencia en el sistema educativo. Para primaria, secundaria baja y secundaria alta la variable de interés toma el valor de uno, si asiste a la escuela y ha atendido los niveles educativos descritos anteriormente, respectivamente.

La diferencia con la tasa neta de asistencia es que este indicador ajustado incluye a aquellos individuos que están en el grupo de interés y asisten al sistema educativo en cualquiera de sus ciclos. Por tanto, un niño que tenga 11 años y que este cursando secundaria baja, sería incluido en el cálculo, a diferencia de la tasa neta de asistencia que lo excluiría de su cálculo. En este contexto, la tasa neta de asistencia ajustada extiende la medición a aquellos individuos de la edad oficial en primaria que asisten a la educación secundaria porque podrían acceder a la educación primaria antes de la entrada oficial o podrían saltarse algunos grados debido a su buen desempeño. El aumento de este indicador podría reflejar una mejora en la participación de los niños en la edad escolar oficial, la disminución de la población objetivo o ambos (UIS, 2017).

La siguiente tabla muestra la estimación de la asistencia neta ajustada para primaria en Panamá, calculada para los niños que tienen entre 6 y 11 años, como porcentaje de aquellos que sí asisten a la escuela y han atendido al menos un nivel de primaria. No son confiables las estimaciones de la tasa neta de asistencia ajustada para el primer y quinto quintil de los hombres en el área urbana, ni para el quinto quintil de las mujeres en el área urbana. Nótese que la estimación para el cuarto quintil de las mujeres en el área urbana no tiene un efecto de diseño estimado, puesto que todas las observaciones en esta desagregación son equivalentes; a pesar de lo anterior, la estimación es confiable pues sobrepasa el umbral de observaciones mínimas.

Tabla 3. Estimaciones de la tasa neta de asistencia ajustada para primaria en Panamá.

Sexo	Área	Quintil	Tamaño de muestra	Tamaño de muestra efectivo	DEFF	Tasa	Li	Ls	cv	Marca
Hombre	Urbana	Primero	456	53	8.62	0.96	0.86	0.99	0.21	*
Hombre	Urbana	Segundo	329	284	1.16	0.99	0.97	1.00	0.13	
Hombre	Urbana	Tercero	237	199	1.19	0.99	0.96	1.00	0.15	
Hombre	Urbana	Cuarto	170	220	0.77	1.00	0.97	1.00	0.19	
Hombre	Urbana	Quinto	104	86	1.21	0.99	0.92	1.00	0.22	*
Mujer	Urbana	Primero	425	164	2.59	0.97	0.93	0.99	0.13	
Mujer	Urbana	Segundo	270	277	0.97	1.00	0.97	1.00	0.16	
Mujer	Urbana	Tercero	210	144	1.46	0.98	0.94	1.00	0.16	
Mujer	Urbana	Cuarto	172	-	-	1.00	-	-	-	
Mujer	Urbana	Quinto	86	-	-	1.00	-	-	-	*

4.3. Tasa neta de asistencia total (*total net attendance rate*)

La tasa neta de asistencia total representa el porcentaje de individuos, en un grupo específico de edad que han atendido algún nivel de educación durante el año académico de referencia. Su forma de cálculo es la de un promedio ponderado que divide el número estimado de individuos en un grupo de edad que han atendido algún nivel de educación, por el número estimado de individuos en ese grupo de edad. La diferencia entre la tasa neta de asistencia total y la tasa neta de asistencia ajustada define una medida de la proporción de niños, en el grupo oficial de edad escolar relevante, que asisten a niveles de educación inferiores al previsto para su edad. Por ejemplo, en la educación primaria, la diferencia entre estos dos indicadores se debe a la asistencia a la educación preescolar, mientras que la diferencia para la educación secundaria baja se debe a la asistencia a la educación preescolar o primaria (UIS, 2017).

Al igual que en los casos anteriores, la edad define el grupo de interés con respecto al nivel educativo específico. Además, la variable de interés toma valores de acuerdo con su asistencia al sistema educativo de primaria. Por tanto, para todos los niveles (primaria,

secundaria baja y secundaria alta) la variable de interés toma el valor de uno si asiste a la escuela y ha atendido al menos un nivel de primaria.

La siguiente tabla muestra la estimación de la asistencia neta total en Panamá, calculada para los individuos que tienen entre 15 y 17 años, como porcentaje de aquellos que sí asisten a la escuela y han atendido al menos un nivel de primaria. Nótese que no son confiables las estimaciones de la tasa neta de asistencia total para el quinto quintil de los hombres en el área urbana, ni para el cuarto y quinto quintil de las mujeres en el área urbana. A pesar de que todas las observaciones del quinto quintil de las mujeres en el área urbana son homogéneas, la estimación no es confiable pues no cuenta con el número mínimo de observaciones.

Tabla 4. Estimaciones de la tasa neta de asistencia total para individuos entre 15 y 18 años en Panamá.

Sexo	Área	Quintil	Tamaño de Muestra	Tamaño de muestra efectivo	DEFF	Tasa	Li	Ls	cv	Marca
Hombre	Urbana	Primero	206	97	2.1	0.76	0.67	0.84	0.13	
Hombre	Urbana	Segundo	150	107	1.4	0.83	0.75	0.89	0.12	
Hombre	Urbana	Tercero	146	87	1.7	0.86	0.77	0.92	0.14	
Hombre	Urbana	Cuarto	111	96	1.1	0.89	0.81	0.94	0.13	
Hombre	Urbana	Quinto	67	62	1.1	0.97	0.88	0.99	0.20	*
Mujer	Urbana	Primero	230	126	1.8	0.84	0.77	0.90	0.11	
Mujer	Urbana	Segundo	172	142	1.2	0.90	0.84	0.94	0.11	
Mujer	Urbana	Tercero	121	103	1.2	0.93	0.86	0.97	0.14	
Mujer	Urbana	Cuarto	99	79	1.2	0.96	0.88	0.98	0.17	*
Mujer	Urbana	Quinto	60	-	-	1.00	-	-	-	*

4.4. Tasa de niños y adolescentes fuera de la escuela (*out-of-school rate*)

La tasa de niños y adolescentes fuera de la escuela representa el porcentaje de individuos, en un grupo de edad específico de edad que no han asistido a la escuela o no

han adquirido el nivel de educación básica. Su forma de cálculo es la de un promedio ponderado que divide el número estimado de individuos en un grupo de edad que no asisten a la escuela o no han atendido el nivel de educación primaria, por el número estimado de individuos en ese grupo de edad. Este indicador también se puede calcular como el complemento de la tasa neta de asistencia ajustada.

Para este indicador, la edad define el grupo de interés con respecto al nivel educativo específico. Además, la variable de interés toma valores de acuerdo con el nivel educativo alcanzado y a su exclusión o inasistencia al sistema educativo. La siguiente tabla muestra la estimación de la tasa de niños y adolescentes fuera de la escuela en Panamá, calculada para los niños que tienen entre 6 y 11 años, como porcentaje de aquellos que no asisten a la escuela y no han completado algún nivel de primaria. Nótese que no son confiables las estimaciones de la tasa de niños y adolescentes fuera de la escuela para el primer y quinto quintil de los hombres en el área urbana, ni para el quinto quintil de las mujeres en el área urbana.

Tabla 5. Estimaciones de la tasa de niños y adolescentes fuera de la escuela para primaria en Panamá.

Sexo	Área	Quintil	Tamaño de muestra	Tamaño de muestra efectivo	DEFF	Tasa	Li	Ls	cv	Marca
Hombre	Urbana	Primero	456	53	8.62	0.04	0.01	0.14	0.21	*
Hombre	Urbana	Segundo	329	284	1.16	0.01	0.00	0.03	0.13	
Hombre	Urbana	Tercero	237	199	1.19	0.01	0.00	0.04	0.15	
Hombre	Urbana	Cuarto	170	220	0.77	0.00	0.00	0.03	0.19	
Hombre	Urbana	Quinto	104	86	1.21	0.01	0.00	0.08	0.22	*
Mujer	Urbana	Primero	425	164	2.59	0.03	0.01	0.07	0.13	
Mujer	Urbana	Segundo	270	277	0.97	0.00	0.00	0.03	0.16	
Mujer	Urbana	Tercero	210	144	1.46	0.02	0.00	0.06	0.16	
Mujer	Urbana	Cuarto	172	-	-	0.00	-	-	-	
Mujer	Urbana	Quinto	86	-	-	0.00	-	-	-	*

4.5. Tasa de finalización (*completion rate*)

La tasa de finalización representa el porcentaje de individuos, en un grupo de tres a cinco años más a la edad de referencia que atendieron un nivel de educación específico en el año académico de referencia. Su forma de cálculo es la de un promedio ponderado que divide el número estimado de individuos un grupo de tres a cinco años más a la edad de referencia que atendieron un nivel de educación específico en el año académico de referencia, por el número estimado de individuos en ese grupo de tres a cinco años más a la edad de referencia.

La edad sigue siendo la característica que se considera para formar el grupo de interés con respecto al nivel educativo específico. De esta forma, en el caso particular de Panamá en donde el rango de edad para primaria es de 6 a 11 años, entonces el grupo de interés para calcular la tasa de finalización de primaria se define por aquellos individuos que tienen entre 14 y 16 años; para secundaria baja, entre 17 y 19 años; en educación secundaria alta, entre 20 y 22 años. Nótese que los anteriores rangos de edad vienen de la definición propia del indicador; por ejemplo, para primaria la edad de finalización es 11 años; por tanto, el grupo de interés está dado por la suma de tres y cinco años a esta edad de referencia ($11 + 3 = 14$; $11 + 5 = 16$). Similarmente ocurre para los demás rangos de edad. De la misma forma, la variable de interés toma valores de acuerdo con el último nivel de educación alcanzado. Para primaria, secundaria baja y secundaria alta la variable de interés toma el valor de uno si ha completado estos niveles educativos, respectivamente.

La siguiente tabla muestra la estimación de la tasa de finalización para secundaria baja en Panamá, calculada para los individuos que tienen entre 18 y 20 años, como porcentaje de aquellos que han completado este nivel educativo. En particular, no son confiables las estimaciones de la tasa de asistencia para los quintos quintiles de hombres y mujeres en el área urbana.

Tabla 6. Estimaciones de la tasa de finalización para secundaria baja en Panamá.

Sexo	Área	Quintil	Tamaño de muestra	Tamaño de muestra efectivo	DEFF	Tasa	Li	Ls	cv	Marca
Hombre	Urbana	Primero	181	113	1.6	0.62	0.53	0.70	0.12	
Hombre	Urbana	Segundo	128	99	1.3	0.78	0.68	0.85	0.12	
Hombre	Urbana	Tercero	138	105	1.3	0.86	0.78	0.91	0.12	
Hombre	Urbana	Cuarto	109	98	1.1	0.84	0.76	0.90	0.13	
Hombre	Urbana	Quinto	59	-	-	1.00	-	-	-	*
Mujer	Urbana	Primero	183	126	1.5	0.78	0.70	0.85	0.11	
Mujer	Urbana	Segundo	163	91	1.8	0.85	0.76	0.91	0.13	
Mujer	Urbana	Tercero	134	128	1.1	0.92	0.85	0.95	0.12	
Mujer	Urbana	Cuarto	105	99	1.1	0.97	0.91	0.99	0.17	
Mujer	Urbana	Quinto	72	60	1.2	0.98	0.89	1.00	0.24	*

4.6. Tasa de participación en aprendizaje organizado un año antes de la edad de ingreso a primaria (*participation rate in organized learning one year before primary entry age*)

Esta tasa equivale al porcentaje de niños, en el grupo de edad de entrada a primaria, que atendieron un nivel de educación específico en el año académico de referencia. Su forma de cálculo es la de un promedio ponderado que divide el número estimado de individuos un año menores a la edad de ingreso a la primaria que asistieron al sistema educativo, por el número estimado de individuos en ese grupo a nivel poblacional. Por ejemplo, para el caso específico de Panamá, en donde la edad de ingreso a primaria se ha definido a los seis años, el grupo de interés lo conforman los niños con cinco años, un año menores a la edad de iniciación en primaria, mientras que la variable de interés toma valores de acuerdo con el reporte de asistencia al sistema educativo. Por lo tanto, la variable de interés tomaría el valor de uno para un niño de 5 años que reporte asistir al sistema educativo.

La siguiente tabla muestra la estimación de la tasa de participación en aprendizaje organizado un año antes de la edad de ingreso a primaria en Panamá, calculada para los niños que tienen cinco años, como porcentaje de aquellos que asisten al sistema educativo. En particular, debido el alto nivel de desagregación, ninguna estimación resulta ser confiable.

Tabla 7. Estimaciones de la tasa de participación aprendizaje organizado un año antes de la edad de ingreso a primaria en Panamá.

Sexo	Área	Quintil	Tamaño de muestra	Tamaño de muestra efectivo	DEFF	Tasa	Li	Ls	cv	Marca
Hombre	Urbana	Primero	73	48	1.53	0.84	0.70	0.92	0.18	*
Hombre	Urbana	Segundo	45	64	0.71	0.94	0.85	0.98	0.18	*
Hombre	Urbana	Tercero	38	21	1.84	0.87	0.64	0.96	0.28	*
Hombre	Urbana	Cuarto	25	-	-	1.00	-	-	-	*
Hombre	Urbana	Quinto	20	-	-	1.00	-	-	-	*
Mujer	Urbana	Primero	58	43	1.35	0.87	0.73	0.94	0.19	*
Mujer	Urbana	Segundo	55	45	1.21	0.94	0.82	0.98	0.21	*
Mujer	Urbana	Tercero	25	20	1.27	0.86	0.63	0.96	0.29	*
Mujer	Urbana	Cuarto	25	21	1.19	0.95	0.72	0.99	0.32	*
Mujer	Urbana	Quinto	19	-	-	1.00	-	-	-	*

Sin embargo, cuando se flexibiliza la desagregación es posible encontrar estimaciones precisas y confiables. Por ejemplo, la siguiente tabla muestra la estimación de la tasa de participación en aprendizaje organizado un año antes de la edad de ingreso a primaria en Panamá, únicamente para los quintiles de ingreso. En este caso, la estimación de los tres primeros quintiles resulta ser confiable. Es posible que al realizar la estimación de este indicador por sexo o por área, de forma separada y no anidada, se encuentren también estimaciones confiables, puesto que el tamaño de muestra sería más elevado y, por ende, suficiente para obtener estimaciones confiables.

Tabla 8. Estimaciones por quintil de ingreso de la tasa de participación en aprendizaje organizado un año antes de la edad de ingreso a primaria en Panamá.

Quintil	Tamaño de muestra	Tamaño de muestra efectivo	DEFF	Tasa	Li	Ls	cv	Marca
Primero	345	178	1.9	0.71	0.64	0.77	0.09	
Segundo	187	137	1.4	0.88	0.81	0.92	0.11	
Tercero	145	104	1.4	0.90	0.83	0.95	0.13	
Cuarto	87	52	1.7	0.97	0.86	0.99	0.22	*
Quinto	60	52	1.1	0.98	0.88	1.00	0.25	*

4.7. Índices de paridad (*parity indices*)

Es posible relacionar la brecha existente en el acceso escolar entre niños y niñas a través de los índices de paridad. Esta es una medida que se calcula como el cociente del indicador estimado para las niñas entre el indicador estimado para los niños. De esta forma, por ejemplo, para la tasa neta de asistencia en primaria, el índice de paridad sería la división entre la tasa neta de asistencia estimada en primaria para las niñas entre la tasa neta de asistencia estimada en primaria para los niños. Para otros niveles educativos (secundaria baja, secundaria alta, etc.) este índice se estima de manera similar. Un índice de paridad de uno implica que no hay diferencias en el acceso al sistema educativo de niños y niñas. Sin embargo, si el índice de paridad es menor a uno implica que existe un menor acceso de las niñas con respecto a los niños; de la misma manera, un índice de paridad mayor a uno implica que existe un mayor acceso de las niñas con respecto a los niños.

Sin embargo, estas estimaciones están sujetas al error de muestreo, puesto que se trata de estimaciones provenientes de encuestas de hogares. Por lo tanto, es necesario calcular medidas de incertidumbre que permitan hacer una inferencia acertada con respecto a la estimación puntual. En particular, es muy útil calcular el intervalo de confianza de las estimaciones de los índices de paridad. De esta forma, si el intervalo de confianza contiene al uno, implica que no hay diferencias significativas entre niños y niñas. Por otro lado, si ambos límites del intervalo de confianza son menores que uno,

implica que hay evidencia de que existe un menor acceso de las niñas con respecto a los niños; si ambos límites del intervalo de confianza son mayores que uno, implica que hay evidencia de que existe un mayor acceso de las niñas con respecto a los niños.

La siguiente tabla muestra los índices de paridad de género para la tasa neta de asistencia en primaria para algunos subgrupos de interés en Panamá. Nótese que, además de presentar las estimaciones puntuales del índice, también se presentan los límites inferior (Li) y superior (Ls) del intervalo de confianza, así como el coeficiente de variación (cv) y marca que implica que sí hay una diferencia significativa entre niños y niñas. En este caso particular, esta diferencia resulta ser significativa únicamente en el cuarto quintil de ingresos.

Tabla 9. Índices de paridad de género para la tasa neta de asistencia en primaria en Panamá.

Subgrupo	Índice	Li	Ls	cv	Diferencia significativa
Nacional	1.001	0.985	1.02	0.846	
Urbano	1.008	0.987	1.03	1.097	
Rural	0.991	0.966	1.02	1.289	
Primer quintil	0.993	0.946	1.04	2.394	
Segundo quintil	0.996	0.975	1.02	1.074	
Tercer quintil	1.002	0.98	1.02	1.12	
Cuarto quintil	1.015	1.002	1.03	0.654	+++
Quinto quintil	1.025	1	1.05	1.233	

Por último, se resalta que además de estimar los índices de paridad de género, también es posible estimar índices de paridad por área (dividiendo el indicador de la zona rural entre el indicador de la zona urbana). Así, es posible medir el progreso de los países en su meta de cerrar las brechas de acceso al sistema educativo entre niños y niñas, y entre las áreas urbana y rural.

5. Conclusiones y recomendaciones

Esta guía ha discutido acerca de la importancia de las encuestas de hogares a la hora de realizar un monitoreo continuo a algunos indicadores relacionados con el ODS 4. Mucho se ha mencionado acerca de la pertinencia de este tipo de instrumentos y de las bondades que las encuestas proveen en términos de precisión. Sin embargo, esta última sección está dedicada a hacer un llamado a los investigadores que analizan estos conjuntos de datos para que tengan en cuenta que, de la misma forma en que las encuestas tienen ventajas, también tienen limitaciones y no se deben usar de forma indiscriminada. Con la transparencia que exige el cumplimiento de la Agenda 2030 para el desarrollo sostenible, cada vez es más común encontrar los conjuntos de datos de las encuestas de hogares disponibles para los investigadores en la región. Antes de pretender analizar las encuestas es necesario comprender que se trata de un conjunto de datos con algunas características que deben ser tenidas en cuenta al momento del análisis. Todos los indicadores estimados a partir de las encuestas de hogares deben estar compuestos de dos partes: la primera correspondiente a la estimación puntual del indicador que es procesado rápidamente por el software computacional, y la segunda corresponde al nivel de incertidumbre asociado a esta estimación puntual. Con estos dos elementos es necesario decidir acerca de la pertinencia de la cifra para tomar decisiones de política pública.

Para lograr un análisis apropiado utilizando los datos de las encuestas de hogares es necesario familiarizarse con la estructura de la encuesta, como las variables que se utilizaron para la estratificación, la definición de las UPMs y el factor de expansión como catalizador del principio de representatividad que permite estimar indicadores de interés a nivel poblacional y a nivel de las desagregaciones que el diseño de la encuesta permita. Estos elementos son esenciales para el tratamiento de los datos en cuanto a la determinación de las medidas de incertidumbre para decidir si la cifra es precisa y confiable o si, por el contrario, es necesario descartarla. Estas medidas de incertidumbre vienen en la forma de los intervalos de confianza y de los coeficientes de variación. Un efecto de la incertidumbre en la estimación es que la amplitud de los intervalos de

confianza crece y por consiguiente el coeficiente de variación sobrepasará el umbral permitido de publicación. En el caso de que el investigador pase por alto los principios de diseño de muestreo de la encuesta, es muy probable que se subestime la variación real de los estimadores conllevando a un sesgo que es dañino para los propósitos de inferencia, porque al subestimar la varianza real se pueden tomar decisiones equivocadas que pueden afectar negativamente a decenas de miles de niños dentro de los sistemas educativos de los países de la región.

También es necesario tener en cuenta el tamaño de muestra de los individuos que componen el subgrupo de interés. Ya se mostró cómo, para algunos indicadores educativos, como por ejemplo la tasa de participación en aprendizaje organizado un año antes de la edad de ingreso a primaria, el tamaño de muestra se reduce tanto que no es posible obtener estimaciones confiables o precisas en algunos casos en que se combinan simultáneamente todas las desagregaciones de interés. En algunos casos, aunque el tamaño de muestra observado parezca ser suficiente, el efecto de la homogeneidad de la variable de interés en la estructura de muestreo propia de la encuesta, definida por las UPMs dentro de los estratos, hace que el efecto de diseño sea alto y que, por ende, el tamaño de muestra efectivo sea mucho menor al observado. Con estas restricciones, es necesario decidir acerca de la viabilidad de la cifra como una estimación sobre la cual se pueda confiar.

Con estas consideraciones en mente, el investigador puede estar confiado en que las encuestas de hogares le proveerán los insumos necesarios para realizar el seguimiento adecuado a los indicadores de interés y, por consiguiente, hacer un monitoreo preciso que, con la ayuda de los gobiernos y su decisión de cumplir con los compromisos adquiridos en la Agenda 2030, redundará en la consecución exitosa de todas las metas asociadas al ODS 4.

Referencias

Barnett-Walker, Kortnee C., James R. Chromy, Teresa R. Davis, Steven L. Emrich, Dawn M. Odom, y Lisa E. Packer. 2003. «2001 National Household Survey on Drug Abuse».

Bethlehem, Jelke. 2009. *The rise of survey sampling*. Statistics Netherlands.

Choy, Wai, Dean R. Gerstein, Felicia Gray, y Rashna Ghadialy. 1998. *National Household Survey on Drug Abuse: Main Findings, (1993)*. DIANE Publishing.

Gutiérrez, Andrés. 2016. *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*. Segunda edición. Ediciones de la U.

Kish, Leslie. 1965. *Survey Sampling*. John Wiley; Sons.

Parker, JD, M Talih, y DJ Malec. 2017. «National Center for Health Statistics Data Presentation Standards for Proportions». *Vital Health Stat 2* (175).

UIS. 2016a. «Estimation of the numbers and rates of out-of-school children and adolescents using administrative and household survey data. Information paper; 35», octubre. UNESCO Institute for Statistics. doi:10.15220/978-92-9189-207-5-en.

———. 2016b. «Age adjustment techniques in the use of household survey data. Information paper; 34», noviembre. UNESCO Institute for Statistics. doi:10.15220/978-92-9189-206-8-en.

———. 2017. «Glosary» Recuperado 19 de octubre de 2018, de <http://uis.unesco.org/en/glossary>

———. 2018. «Guía Abreviada de Indicadores de Educación para el ODS 4». UNESCO Institute for Statistics.