www.ecuadorencifras.gob.ec

ENCUESTA NACIONAL SOBRE DESNUTRICIÓN Infantil



Diseño Muestral

ENDI 2025-2026

Octubre · 2026

INEC | Buenas cifras, mejores vidas

Documento Diseño Muestral de la Encuesta Nacional sobre Desnutrición Infantil - ENDI 2025-2026

Dirección/Departamento
Dirección de Infraestructura Estadística y Muestreo

Unidad
Gestión de Diseño Muestral

Elaborado por:
Geovanny Gutiérrez

Revisado por:
Jonathan Rosero

Aprobado por:
Jennifer Marcillo

Quito - Ecuador, 2026

Contenido

Introducción 1
Objetivos de la encuesta y población objetivo 1
Dominios de estudio 1
Marco muestral 2
Unidades Primarias de Muestreo 2
Estratificación 4
Diseño Muestral 5
Determinación del tamaño muestral 5
Parámetros muestrales 6
Cálculo del tamaño muestral 7
Distribución de la muestra 8
Selección de la muestra de UPM 11
Enlistamiento de las UPM 12
Selección de viviendas 14
Cobertura 16
Condición de ocupación y resultado de entrevista 16
Tasas de conformidad 17
Población objetivo 18
Lactancia materna 20
Desarrollo infantil 20
Paralización del levantamiento 21
Factores de expansión 23
Probabilidad de primera etapa 23
Probabilidad de segunda etapa 23
Factor de expansión teórico 23
Ajuste de cobertura de primera etapa 24
Ajuste de cobertura de segunda etapa 25
Calibración a proyecciones poblacionales 28
Recorte de factores calibrados 29
Factor de expansión normalizado 29
Factor de expansión Lactancia Materna 31
Factor de expansión Desarrollo Infantil 32
Resumen de los factores por módulo 33
Estimaciones de características 35
Error de muestreo con factores de expansión calibrados 35
Métodos de estimaciones de errores para diseños muestrales complejos 37
Estimación de totales mediante fuentes de información auxiliares . . 38
Anexos 40
Anexo 1 40

Anexo 2 41
Anexo 3 42
Anexo 4 43

Lista de Tablas

1. Distribución de viviendas de acuerdo al proceso de actualización cartográfica. 2
2. Distribución de UPM por provincia y área. 3
3. Insumos obtenidos a partir de la ENSANUT 2018. 6
4. Tamaño muestral a nivel de niños/as, viviendas y UPM por provincia. 7
5. Distribución de la muestra por mes y área. 9
6. Tamaño muestral final de UPM ENDI R1 y R2. 9
7. Tamaño muestral final de UPM ENDI R3. 11
8. Número de UPM planificadas y enlistadas. 12
9. Número de UPM por número de viviendas seleccionadas. 14
10. Condición de elegibilidad de las viviendas. 16
11. Resumen de distribución de niños y niñas menores a 5 años por año y meses. 19
12. Resultado de la entrevista formulario MEF. 20
13. Resumen de distribución de niños y niñas menores a tres años por MEF. 20
14. Resultado de la entrevista del hogar. 21
15. Pasos de cálculo de los factores de expansión por módulo. 33
16. Variables requeridas para la declaración del diseño muestral. 38
17. Condición de elegibilidad y tasa de conformidad (%) por provincia. 42

Introducción

Objetivos de la encuesta y población objetivo

La Encuesta Nacional sobre Desnutrición Infantil (ENDI) tiene como principal objetivo medir los datos de desnutrición crónica, desnutrición aguda, desnutrición global, sobrepeso y obesidad, controles prenatales, vacunación oportuna de neumococo y rotavirus, anemia, lactancia materna, consumo de hierro y ácido fólico en el embarazo, entre otros. La población objetivo de la ENDI son los niños y las niñas menores a cinco años de edad.

Dominios de estudio

En general, una de las características principales de una encuesta es comparar sus resultados entre diferentes desagregaciones, como por ejemplo nacional urbano, nacional rural, provincial, quintiles de ingreso, niveles de educación, entre otros. Estas desagregaciones se conocen como dominios de estudio, y se clasifican en dominios de diseño y dominios de análisis.

Un dominio de diseño consiste en una subpoblación que puede identificarse en el marco de muestreo y, por lo tanto, puede ser manejada independientemente desde el diseño muestral. Mientras que los dominios de análisis son subpoblaciones que no pueden ser identificadas en el marco de muestreo, por lo que sus integrantes se determinan según la información levantada en la encuesta.

El levantamiento de información de la ENDI está planificado a lo largo de 12 meses, por lo que la muestra está distribuida temporal y espacialmente. De esta forma, los dominios de diseño son a nivel urbano, rural y provincial con la agregación de los 12 meses de la muestra.

Es importante mencionar que, debido a restricciones operativas, el levantamiento de información en la provincia de Galápagos tendrá otras consideraciones desde el diseño, por lo que los dominios descritos excluyen a esta provincia.

Marco muestral

El Marco de muestreo a utilizar para la ENDI Ronda 3 es el Marco Maestro de Muestreo para encuestas de hogares del INEC (MMM), el cual se presenta como una lista organizada en forma de base de datos que contiene las viviendas registradas en el precenso del VIII Censo de Población y VII de Vivienda del Ecuador (CPV - 2022) junto con las actualizaciones cartográficas realizadas en el periodo 2022-2025.

En la Tabla 1 se presenta la proporción de viviendas catalogadas como ocupadas según el precenso CPV-2022 y los diferentes procesos de actualización.

Tabla 1: Distribución de viviendas de acuerdo al proceso de actualización cartográfica.

Actualización cartográfica	Año de actualización	Total de viviendas
Precenso	2022	4.587.419
ENIGHUR	2025	299.597

Fuente: Gestión de Diseño Muestral - DINEM.

Las actualizaciones cartográficas hacen referencia a:

- ENIGHUR: Encuesta Nacional de Ingresos y Gastos en Hogares Urbanos y Rurales del Ecuador 2025.

Unidades Primarias de Muestreo

Cada vivienda del MMM 2022 está asociada a una y solo una manzana o sector disperso. Tanto en la parte amanzanada como en la parte dispersa se forman conjuntos de manzanas o sectores, respectivamente, a los que denominaremos Unidades Primarias de Muestreo (UPM), de tal forma que el número de viviendas ocupadas de estos conjuntos sea lo más homogéneo posible (desde el punto de vista del total de viviendas ocupadas); el límite inferior de tamaño es de 40 viviendas particulares ocupadas. Además, las UPM son geográficamente conexas.

El Marco de muestreo codifica las UPM en función a la División Político Administrativa definida en el Clasificador Geográfico Estadístico (DPA 2022), año en el que se construyó el marco de muestreo actual, lo que permite identificar adecuadamente los dominios de diseño geográficos definidos para la ENCIET. Además, contiene variables que clasifican las UPM en función al estrato al que pertenecen. Para más información referirse a [Instituto Nacional de Estadística y Censos, 2024]. En la Tabla 2 se presenta el total de UPM por provincia y área presentes en el MM 2022.

Tabla 2: Distribución de UPM por provincia y área.

Provincia	Total de UPM
Azuay	2697
Bolívar	648
Cañar	820
Carchi	604
Cotopaxi	1472
Chimborazo	1626
El Oro	2453
Esmeraldas	1833
Guayas	14405
1Imbabua	1580
Loja	1565
Los Ríos	2986
Manabí	5162
Morona Santiago	567
Napo	357
Pastaza	340
Pichincha	10481
Tungurahua	1998
Zamora Chinchipe	335
Sucumbíos	662
Orellana	549
Santo Domingo	1665
Santa Elena	1130
Nacional	55929

Fuente: Gestión de Diseño Muestral - DINEM.

Cabe resaltar que la definición de urbano/rural establecida en el MMM es la de 2.000 personas o más, el cual es un estándar normalizado a nivel de la Comunidad Andina de Naciones para asegurar comparabilidad internacional. Esta establece que, si la población en el área amanzanada de una parroquia supera los 2.000 habitantes, está se considera urbana, caso contrario será catalogada como rural. Para más información al respecto, referirse al documento [Comunidad Andina de Naciones, 2000].

Estratificación

La estratificación se refiere a la división de una población determinada en subconjuntos con características propias. Esta acción se lleva a cabo como una etapa previa a la selección de la muestra, mediante variables que aportan información para todas las unidades de la población.

Los objetivos del proceso de estratificación son los siguientes:

- Agrupar a las unidades del marco de acuerdo a un conjunto de características socio económicas.
- Formar grupos de UPM homogéneos en su interior y heterogéneos entre ellos.
- Mejorar la eficiencia del diseño muestral mediante la disminución de la varianza de los estimadores.

En el caso del MM 2022 del INEC, todas las UPM son clasificadas en algún estrato, con base en los siguientes criterios:

- Clasifica a las UPM por provincia y área. Este criterio también considera a los diez cantones autorepresentados (Cuenca, Riobamba, Loja, Quito, Ambato, Machala, Esmeraldas, Guayaquil, Manta y Santo Domingo).
- Cada división geográfica se clasifica según sus características sociales y económicas en 3 grupos. Sin embargo, si el número de UPM presentes es muy pequeño, se clasifican en dos o un grupo.

Para más información acerca de la estratificación del MM 2022, remitirse al documento [Instituto Nacional de Estadística y Censos, 2024].

Diseño Muestral

El diseño muestral implementado en la ENDI es un muestreo probabilístico bietápico estratificado de elementos. En la primera etapa, se selecciona una muestra estratificada de UPM con probabilidad proporcional al tamaño (PPT), donde la medida de tamaño de cada UPM está dada por el total de viviendas particulares ocupadas. Luego, se enlista la totalidad de cada UPM seleccionada, con la finalidad de corregir los errores existentes en el marco e identificar aquellas viviendas con niños/as menores a cinco años.

En la segunda etapa de muestreo, se selecciona de manera sistemática aleatoria un número fijo de viviendas: 7 por UPM, cifra escogida tras consideraciones operativas y presupuestales.

Existe una tercera etapa de muestreo únicamente para las secciones de Lactancia Materna y Desarrollo Infantil. En la primera, se escoge un niño o niña menor de 3 años por mujer en edad fértil efectiva, mientras que en la segunda, se escoge un niño o niña menor de 5 años por hogar efectivo en el segundo semestre de levantamiento de información. En el módulo de Lactancia Materna la selección se hace mediante el criterio de cumpleaños más próximo, mientras que en Desarrollo Infantil se realiza la selección de manera aleatoria. A continuación se detallan los procesos involucrados en el diseño muestral de la ENDI.

Determinación del tamaño muestral

El tamaño de muestra de la ENDI se determinó mediante la siguiente ecuación:

$$n \geq \frac{N \cdot p \cdot (1-p) \cdot deff}{N \cdot \left(\frac{p \cdot e_{rel}}{z}\right)^2 + p \cdot (1-p) \cdot deff} \cdot \frac{1}{1-tnr}$$

donde:

n = Tamaño de muestra resultante.
p = Prevalencia del estimador de diseño.
e_{rel} = Error relativo asociado a la prevalencia p.
z = Valor asociado al nivel de confianza.
N = Tamaño de la población objetivo del estimador.
$deff$ = Efecto de diseño.
tnr = Tasa de no respuesta

Parámetros muestrales

Para calcular el tamaño muestral de la ENDI se necesita determinar los diferentes parámetros por dominio de estudio. Los valores determinados para los parámetros que servirán de forma general para todos los dominios de estudio son:

$e_{rel} = 0{,}15$
$z \quad = 1{,}96$
$tnr = 0{,}20$

Se fijó la tasa de no respuesta en 0,20 ya que en la Encuesta Nacional de Salud y Nutrición 2018 (ENSANUT 2018), a pesar de haber contado con un proceso de enlistamiento previo al levantamiento de información, el número de viviendas identificadas originalmente como viviendas con niños/as en realidad no cumplían con esta característica. Esta decisión mitiga potenciales tasas de no respuesta altas en los dominios de estudio y asegura alcanzar el total de niños/as menores a cinco años planificado al momento de calcular el tamaño muestral de la encuesta.

Para la ENDI, el indicador de diseño es la prevalencia de desnutrición crónica en niños/as menores a cinco años, el mismo se determinó a partir de la ENSANUT 2018. Además, se determinaron también el efecto de diseño y el promedio de niños/as menores a cinco a años por vivienda. En la Tabla 3 se detalla estos valores por provincia.

Tabla 3: Insumos obtenidos a partir de la ENSANUT 2018.

Provincia	**Desnutrición crónica de niños/as menores a 5 años**	**Efecto de diseño**	**Promedio de niños/as menores a 5 años por vivienda**
Azuay	28,5 %	2,60	1,16
Bolívar	35,6 %	2,46	1,12
Cañar	26,8 %	1,59	1,14
Carchi	30,5 %	1,60	1,14
Cotopaxi	31,3 %	1,38	1,16
Chimborazo	35,0 %	2,36	1,16
El Oro	16,5 %	1,69	1,17
Esmeraldas	18,5 %	1,47	1,19
Guayas	18,1 %	1,57	1,21
Imbabura	28,7 %	1,53	1,15
Loja	23,3 %	1,29	1,19
Los Ríos	19,6 %	1,12	1,20
Manabí	22,7 %	1,95	1,20
Morona Santiago	34,9 %	1,84	1,33
Napo	27,2 %	1,67	1,27

Provincia	Desnutrición crónica de niños/as menores a 5 años	Efecto de diseño	Promedio de niños/as menores a 5 años por vivienda
Pastaza	34,4%	2,32	1,30
Pichincha	21,6%	3,58	1,14
Tungurahua	32,3%	3,79	1,16
Zamora Chinchipe	24,2%	1,60	1,19
Sucumbíos	26,7%	2,05	1,20
Orellana	30,3%	1,49	1,27
Santo Domingo	19,4%	1,82	1,16
Santa Elena	37,0%	2,35	1,22

Fuente: Gestión de Diseño Muestral - DINEM.

Cálculo del tamaño muestral

Con los parámetros descritos en la sección anterior y la aplicación de la fórmula para determinar el tamaño muestral, se obtiene el número de niños/as menores a cinco años necesarios a considerar en el levantamiento de información. Para determinar el número de viviendas a seleccionar, se divide el número de niños/as para el promedio de niños/as por vivienda en cada provincia. Finalmente, para obtener el número de UPM, se divide el número de viviendas para 8. En la Tabla 4 se detalla los valores para cada dominio de diseño de la encuesta.

Tabla 4: Tamaño muestral a nivel de niños/as, viviendas y UPM por provincia.

Provincia	Número de niños/as menores a cinco años	Número de viviendas	Número de UPM
Azuay	1.374	1.182	148
Bolívar	899	800	100
Cañar	899	787	99
Carchi	744	651	82
Cotopaxi	640	552	69
Chimborazo	919	791	99
El Oro	1.780	1.527	191
Esmeraldas	1.357	1.138	143
Guayas	1.517	1.250	157
Imbabura	799	698	88
Loja	890	752	94
Los Ríos	973	811	102
Manabí	1.410	1.176	147
Morona Santiago	717	538	68

Provincia	Número de niños/as menores a cinco años	Número de viviendas	Número de UPM
Napo	903	712	89
Pastaza	893	688	86
Pichincha	2.748	2.420	303
Tungurahua	1.660	1.429	179
Zamora Chinchipe	983	827	104
Sucumbíos	1.147	958	120
Orellana	713	562	71
Santo Domingo	1.570	1.357	170
Santa Elena	837	684	86
Total	26.372	22.290	2.795

Fuente: Gestión de Diseño Muestral - DINEM.

Distribución de la muestra

Una vez determinado el tamaño muestral en cada dominio de estudio, se procede a su distribución espacial y temporal. Para garantizar la distribución espacial se toma en cuenta los estratos que conforman cada una de las provincias. Por otro lado, puesto que la ENDI recabará información a lo largo de un año, la distribución temporal debe considerar los 12 meses del año. La importancia de distribuir la muestra por mes y estrato se detalla a continuación:

- **Mes:** Está directamente relacionado con la parte operativa, puesto que fija el número de personas contratadas por zonales, vehículos, estipendios, etc.; tanto para el enlistamiento como para el levantamiento de información, por lo que es preferible que el número de UPM por provincia sea el mismo en cada uno de los meses.

- **Estrato:** Permite responder a la representatividad requerida de la encuesta con la agregación de los 12 meses, es decir, a nivel nacional urbano, nacional rural y provincial.

La forma más sencilla de realizar esta distribución es dividir de forma proporcional el tamaño muestral de cada provincia entre sus estratos y aproximar el valor obtenido al múltiplo de 12 más cercano, sin embargo esto haría que crezca el tamaño muestral de forma innecesaria, por esta razón, para optimizar los recursos disponibles, la distribución de la muestra siguió el siguiente esquema:

1. Se distribuyó de forma proporcional al número de viviendas ocupadas entre urbano y rural en cada provincia.

2. El valor obtenido se lo redondeó al múltiplo de seis más cercano.

3. Se sumaron los valores obtenidos en cada provincia, si el valor no era múltiplo de 12 se sumó 3 en lo urbano y 3 en lo rural.

4. En cada subdominio provincia - área se distribuyó la muestra entre los estratos que lo componen.

5. De ser necesario, se aumentó el valor obtenido por estrato a cuatro, para garantizar al menos 2 UPM por semestre.

6. En los casos que se debió aumentar a cuatro, se restó de los estratos del subdominio provincial - área al que pertenece el estrato en el que se aumentó la muestra, para que el total no supere el fijado anteriormente.

7. Para garantizar la distribución temporal, se identificó a cada UPM seleccionada con un mes diferente, esto se consigue al ordenar las UPM seleccionadas por provincia, área y estrato, y repetir el vector formado por los números del 1 al 12 en el siguiente orden 1, 7, 4, 10, 2, 8, 5, 11, 3, 9, 6, 12.

De esta forma, se tiene que el número de UPM por provincia es el mismo en cada uno de los meses, con lo cual se consideran las restricciones operativas. En la Tabla 5 se muestra la distribución a nivel de provincia - área para los dominios Guayas, Bolívar y Napo.

En la Tabla 5 se presenta la distribución de la muestra por urbano y rural, donde se aprecia que su diferencia es a lo más 1 entre dos meses consecutivos.

Tabla 5: Distribución de la muestra por mes y área.

Provincia	Área	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
Bolívar	Urbano	4	3	3	4	3	3	4	3	3	3	3	3
	Rural	5	6	6	5	6	6	5	6	6	6	6	6
Guayas	Urbano	12	12	11	12	12	11	12	11	11	12	11	11
	Rural	1	1	2	1	1	2	1	2	2	1	2	2
Napo	Urbano	4	3	3	4	3	3	4	3	3	3	3	3
	Rural	4	5	5	4	5	5	4	5	5	5	5	5

Fuente: Gestión de Diseño Muestral - DINEM.

Tamaño final a nivel de UPM

Después de realizar la distribución espacial y temporal de la muestra de UPM, esta presenta diferencias con la inicialmente calculada. En la Tabla 6 se presenta el tamaño final de UPM por provincia y por mes.

Tabla 6: Tamaño muestral final de UPM ENDI R1 y R2.

Provincia	Número de UPM	Número de UPM por mes
Azuay	144	12
Bolívar	108	9
Cañar	108	9
Carchi	84	7
Cotopaxi	72	6
Chimborazo	108	9

Provincia	Número de UPM	Número de UPM por mes
El Oro	192	16
Esmeraldas	144	12
Guayas	156	13
Imbabura	84	7
Loja	96	8
Los Ríos	108	9
Manabí	156	13
Morona Santiago	72	6
Napo	96	8
Pastaza	84	7
Pichincha	312	26
Tungurahua	180	15
Zamora Chinchipe	108	9
Sucumbíos	120	10
Orellana	72	6
Santo Domingo	168	14
Santa Elena	84	7
Total	2.856	238

Fuente: Gestión de Diseño Muestral - DINEM.

El tamaño muestral previamente expuesto se utilizó para las rondas 1 y 2 de la ENDI, para la ronda 3 al contar con datos actualizados se realizó una redistribución de la muestra mediante el siguiente algoritmo:

$$n_{r3} = n_{r2} \cdot \left(\frac{cv_{r2}}{cv_{r3}}\right)^2$$

n_{r3} = Tamaño de muestra ronda 3.
n_{r2} = Tamaño de muestra ronda 2.
cv_{r2} = Coeficiente de variación ronda 2.
cv_{r3} = Coeficiente de variación esperado ronda 3.

Además, dado que las UPM del MM 2022 son en promedio más pequeñas se decidió pasar de 8 a 7 viviendas por UPM y obtener un coeficiente de variación meta de entre 9 y 10 %, así en algunas provincias se redujo el número de viviendas a visitar y en otras se aumentó, en la siguiente tabla se presenta el tamaño final:

Tabla 7: Tamaño muestral final de UPM ENDI R3.

Provincia	Número de UPM	Número de UPM por mes
Azuay	108	9
Bolívar	120	10
Cañar	120	10
Carchi	96	8
Cotopaxi	96	8
Chimborazo	96	8
El Oro	204	17
Esmeraldas	228	19
Guayas	192	16
Imbabura	108	9
Loja	168	14
Los Ríos	180	15
Manabí	168	14
Morona Santiago	96	8
Napo	120	10
Pastaza	156	13
Pichincha	168	14
Tungurahua	96	8
Zamora Chinchipe	120	10
Sucumbíos	228	19
Orellana	156	13
Santo Domingo	168	14
Santa Elena	96	8
Total	3.288	274

Fuente: Gestión de Diseño Muestral - DINEM.

Selección de la muestra de UPM

Tras determinar el tamaño final de UPM, se procedió a realizar la selección de las UPM de forma proporcional al tamaño dentro de cada estrato (la medida de tamaño de la UPM es el número de viviendas ocupadas según el Marco de Muestreo) y así planificar el enlistamiento de manera oportuna, con lo cual se actualizó el listado de viviendas ocupadas y se corrigió en parte la desactualización del Marco de Muestreo.

Enlistamiento de las UPM

Un mes y medio antes del levantamiento de información, la Dirección de Cartografía y Operaciones de Campo (DICA) realiza el proceso de enlistamiento de las UPM, el cual consiste en actualizar el listado de viviendas de cada UPM y diferenciarlas por su condición de ocupación (ocupadas, desocupadas, en construcción, temporal, colectivas, edificios sin vivienda, lotes vacíos), además de identificar las viviendas en las que habita la población objetivo, es decir niños/as menores a cinco años. Además, el enlistamiento produce los materiales cartográficos necesarios para ubicar a las viviendas seleccionadas en la encuesta.

Un conjunto de UPM no fueron parte del enlistamiento debido a diferentes inconvenientes como problemas de accesibilidad debido a condiciones climatológicas, rechazo por parte de conjuntos residenciales, dificultades debido a los problemas de seguridad presentes en todo el territorio ecuatoriano. En la Tabla 8 se detalla el número de UPM planificadas y el número de UPM enlistadas.

Tabla 8: Número de UPM planificadas y enlistadas.

Dominio	**N° de UPM planificadas**	**N° de UPM enlistadas**
Azuay	108	107
Bolívar	120	118
Cañar	120	120
Carchi	96	96
Cotopaxi	96	96
Chimborazo	96	96
El Oro	204	204
Esmeraldas	228	226
Guayas	192	191
Imbabura	108	106
Loja	168	167
Los Ríos	180	180
Manabí	168	168
Morona Santiago	96	93
Napo	120	120
Pastaza	156	153
Pichincha	168	168
Tungurahua	96	96
Zamora Chinchipe	120	120
Sucumbíos	228	227
Orellana	156	154

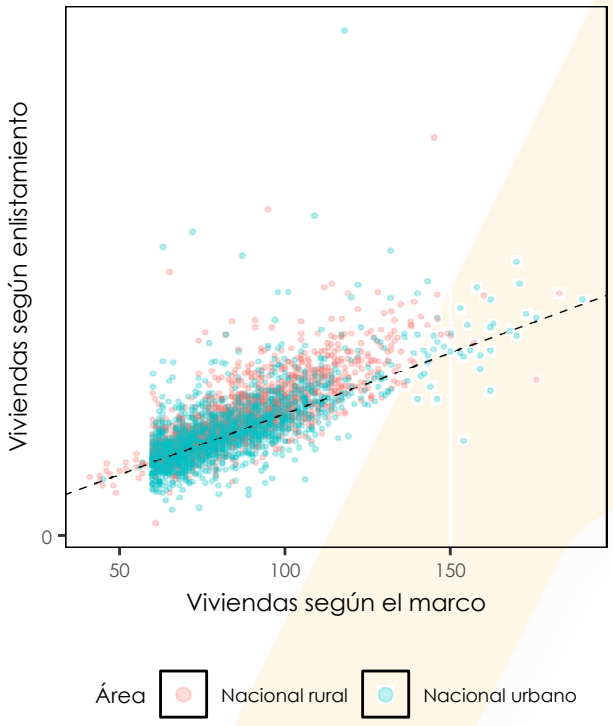
Santo Domingo	168	168
Santa Elena	96	96
Total	3288	3270

Fuente: Gestión de Diseño Muestral - DINEM.

En total, no se enlistaron 18 UPM que representa el 0,55 % de las UPM planificadas; 3 en Morona Santiago y Pastaza, 2 en Bolívar, Esmeraldas, Imbabura, Orellana y 4 provincias con diferencia de 1 UPM entre lo planificado y enlistado, en 13 provincias se enlisto el total de UPM planificadas.

En el Figura 1 se muestra la diferencia entre la medida de tamaño (número de viviendas ocupadas) de las UPM según el MMM y la resultante del proceso de enlistamiento.

Ilustración 1: Viviendas ocupadas según marco y enlistamiento.



Fuente: Gestión de Diseño Muestral - DINEM.

A nivel nacional, la medida de tamaño crece en un 7.68 % (4.33 % en nacional urbano y 12.95 % en nacional rural). La variación máxima fue del 247.22 %, mientras que la variación mínima fue del -83.61 %. En el Anexo 1 se presenta la diferencia para cada una de las provincias.

Selección de viviendas

Una vez realizado el proceso de enlistamiento, el último paso es la selección de viviendas con niños/as menores a cinco años, sin embargo, no todas las UPM tuvieron al menos 7 viviendas elegibles. Para compensar el déficit del número de viviendas, se escogió una vivienda adicional en otras UPM del mismo estrato por cada UPM con menos de 7 viviendas seleccionadas. En la Tabla 9 se muestra el número de UPM de acuerdo al número de viviendas seleccionadas.

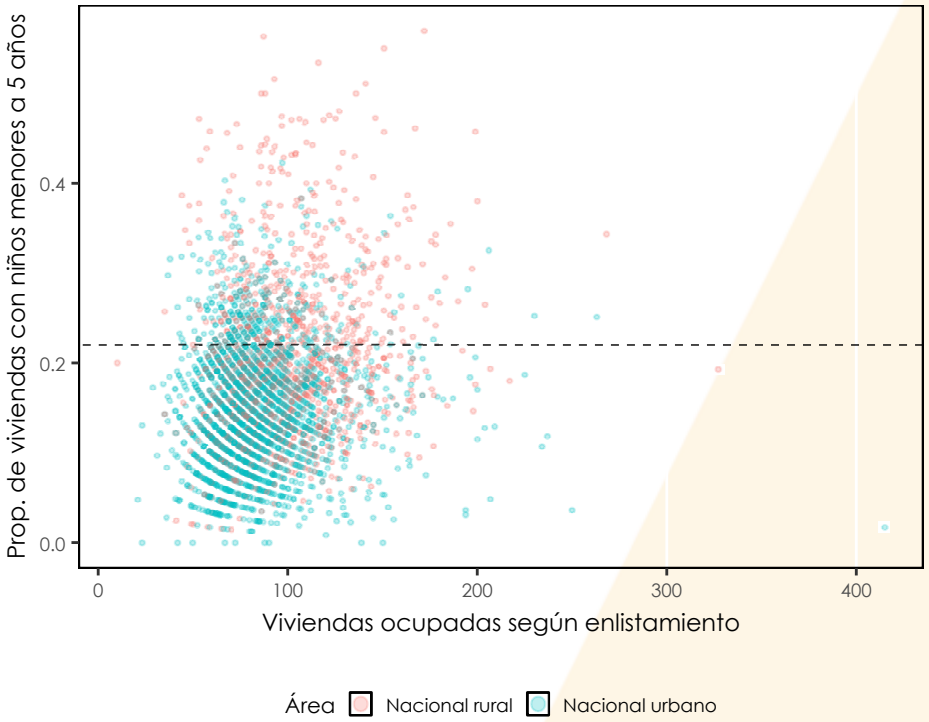
Tabla 9: Número de UPM por número de viviendas seleccionadas.

N° viviendas seleccionadas	Número de UPM		
	Nacional urbano	Nacional rural	Total
0	10	17	27
1	10	6	16
2	42	3	45
3	57	7	64
4	84	8	92
5	91	18	109
6	107	28	135
7	1.276	1.036	2.312
8	385	103	488
Total	2.062	1.226	3.288

Fuente: Gestión de Diseño Muestral - DINEM.

En la Figura 2 se tiene la proporción de viviendas con niños/as menores a cinco años según el número de viviendas ocupadas, la línea entrecortada representa la proporción nacional (0,22). Se puede apreciar claramente que en el nacional rural hay una proporción más alta. En el Anexo 2 se presenta la proporción para cada una de las provincias.

Ilustración 2: Proporción de viviendas con niños/as menores a cinco años por UPM.



Fuente: Gestión de Diseño Muestral - DINEM.

Cobertura

La cobertura analiza la información obtenida en campo de los informantes de las viviendas seleccionadas. A continuación, se presenta el seguimiento de la cobertura de la tercera ronda de la encuesta, además de los cambios de ocupación[1] y de la población objetivo, los cuales influyen directamente en la fase de ponderación y estimación de características de la ENDI.

Condición de ocupación y resultado de entrevista

Las condiciones que presentan las viviendas visitadas en la encuesta a nivel nacional, nacional urbano y nacional rural son las siguientes:

Tabla 10: Condición de elegibilidad de las viviendas.

Condición de la vivienda	Nacional urbano	Nacional rural	Nacional	Porcentaje
Completa	12.499	7.833	20.332	92,0%
Rechazo	158	24	182	0,8%
Nadie en casa	153	78	231	1,0%
Vivienda temporal	35	55	90	0,4%
Vivienda desocupada	235	68	303	1,4%
Vivienda en construcción	11	4	15	0,1%
Vivienda destruida	1	4	5	0,0%
Negocio	1	0	1	0%
Vivienda sin niño/a	577	295	872	3,9%
Otra razón	40	38	78	0,4%
Total	13.710	8.399	22.109	100,0%

Fuente: Gestión de Diseño Muestral - DINEM.

Como se puede observar en la Tabla 10, se investigaron 22.109 viviendas a nivel nacional (92,0% efectivas), 13.710 viviendas en el nacional urbano y 8.399 viviendas en el nacional rural. Adicionalmente, puesto que la cobertura efectiva supera el 80% (tasa de no respuesta del 20%), se asegura alcanzar el total de niños/as menores a cinco años planificado al momento de diseñar la encuesta.

Al analizar las otras categorías, se puede observar que el 3,9% de viviendas no contaban con niños/as menores de 5 años, estas viviendas se categorizan como no elegibles como se explica a continuación.

[1] Las viviendas seleccionadas por parte de diseño muestral, son viviendas ocupadas con hogares que tienen la población objetivo, sin embargo no necesariamente cumplen con esta característica al momento de realizar la investigación.

En este sentido, la cobertura a nivel de vivienda y población objetivo toma en cuenta la condición de ocupación y el resultado de la entrevista de las viviendas visitadas, las cuales se clasifican en dos grandes grupos: elegibilidad conocida y elegibilidad desconocida, con la siguiente subclasificación:

1. Elegibilidad conocida
    - Elegible respondiente (RE)
        - Efectiva
    - Elegible no respondiente (NR)
        - Rechazo
    - No elegibles (NE)
        - Temporal
        - Desocupada
        - En construcción
        - Inhabitable o destruida
        - Convertida en negocio
        - Sin niños/as menores a 5 años
        - Otra razón, cuál?
2. Elegibilidad desconocida
    - Elegibilidad desconocida (ED)
        - Nadie en casa

Posteriormente se define las tasas de conformidad de las viviendas investigadas.

Tasas de conformidad

Al tomar en cuenta la condición de elegibilidad de las viviendas investigadas, se calculan las siguientes tasas de conformidad:

- Tasa de respondientes (T_{RE}): esta tasa permite conocer la proporción de viviendas efectivas para el total de viviendas visitadas.

$$T_{RE} = \frac{RE}{RE + NR + NE + ED}$$

- Tasa de no respondientes (T_{NR}): esta tasa permite conocer la proporción de viviendas que fueron rechazo para el total de viviendas visitadas.

$$T_{NR} = \frac{NR}{RE + NR + NE + ED}$$

- Tasa de no elegibles (T_{NE}): esta tasa indica la proporción de viviendas no elegibles que formaron parte de la muestra y que, en un principio, no deberían haber formado parte del marco de muestreo.

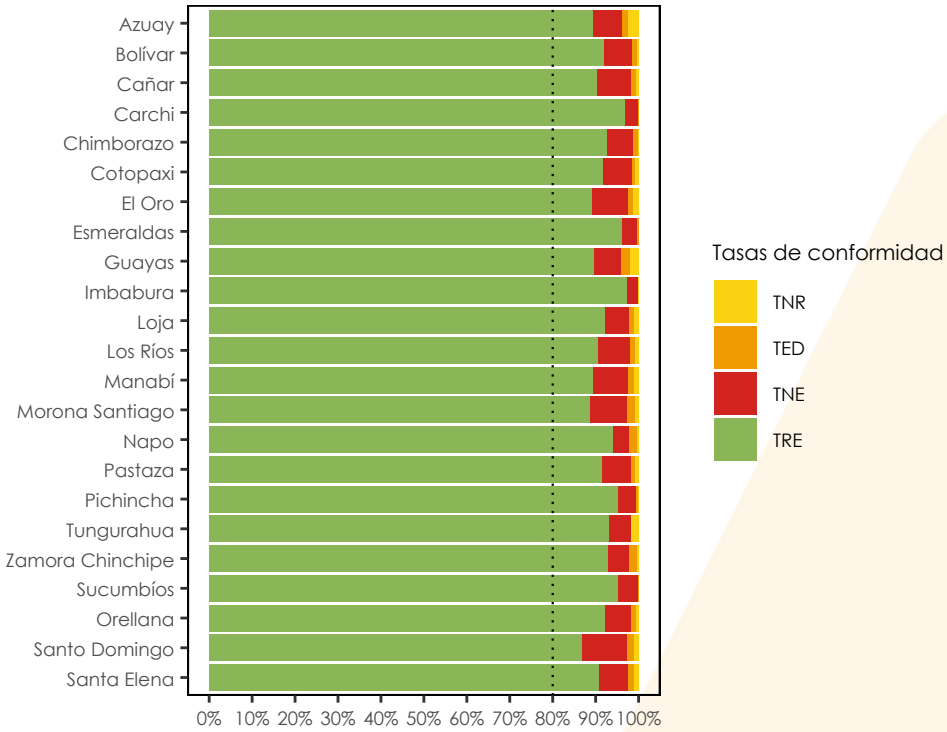
$$T_{NE} = \frac{NE}{RE + NR + NE + ED}$$

- Tasa de elegibilidad desconocida (T_{ED}): esta tasa permite conocer la proporción de viviendas cuya condición de elegibilidad no pudo ser determinada en campo.

$$T_{ED} = \frac{ED}{RE + NR + NE + ED}$$

A continuación, se realiza un análisis por dominio de diseño de las tasas de conformidad.

Ilustración 3: Tasas de conformidad por provincia.



Fuente: Gestión de Diseño Muestral - DINEM.

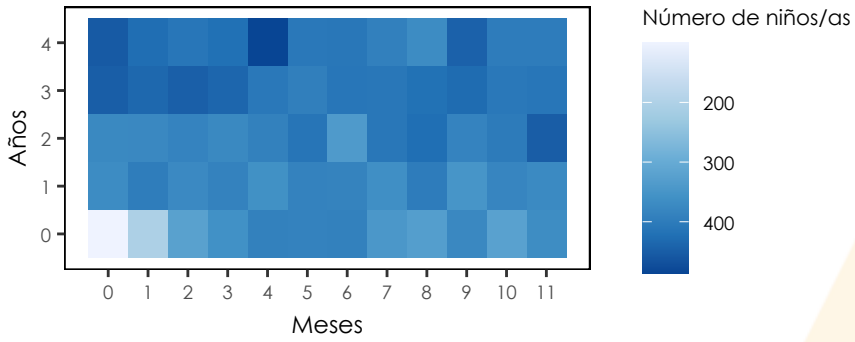
Todas las provincias cuenta con una tasa de respondientes mayor al 80 %, tasa de no elegibles desde 2.59 % que pertenece a Imbabura hasta 10.56 % en Santo Domingo, además, cuentan con tasa de elegibilidad desconocida entre 0.15 % y 2.02 % en Carchi y Guayas respectivamente. Por último, a excepción de las provincias de Esmeraldas y Carchi, la tasa de no respondientes está entre 0.06 % y 2.51 % en Sucumbíos y Azuay respectivamente.

En el Anexo 3, se presenta el resultado de la condición de elegibilidad y la tasa de conformidad para cada provincia.

Población objetivo

En los doce meses operativos de levantamiento de información, la ENDI entrevistó a 23.247 niños/as menores a cinco años. En el siguiente gráfico se presenta su distribución por su edad en años y meses:

Ilustración 4: Distribución de niños y niñas menores de 5 años por años y meses.



Fuente: Gestión de Diseño Muestral - DINEM.

Puesto que las viviendas con niños/as menores de cinco años se seleccionan de manera aleatoria, se espera que el total de niños/as por año cumplido sea uniforme, lo cual se aprecia en la tabla 11. Con respecto al total de niños/as con cero años de edad, su diferencia se debe al desfase temporal entre el enlistamiento y el levantamiento de la información, mientras que en la edad de 3 años se encuentra la mayor concentración de niños/as con un 21,8% del total de niños/as.

Tabla 11: Resumen de distribución de niños y niñas menores a 5 años por año y meses.

Años	Meses con menor concentración	Meses con mayor concentración	Población	Porcentaje
0	0	4	3901	16,8%
1	9	8	4537	19,5%
2	6	11	4727	20,3%
3	5	0	5058	21,8%
4	8	4	5024	21,6%

Fuente: Gestión de Diseño Muestral - DINEM.

Lactancia materna

La sección Lactancia Materna se encuentra en el formulario de Mujeres en Edad Fértil (MEF). En la Tabla 12 se presenta el resultado de entrevista del formulario.

Tabla 12: Resultado de la entrevista formulario MEF.

Resultado de la entrevista	Nacional urbano	Nacional rural	Nacional	Porcentaje
Efectiva	12.217	7.716	19.933	64,9 %
Rechazo	85	19	104	0,3 %
Ausente	238	188	426	1,4 %
No tiene hijos/as menores a 5 años	6.055	4.150	10.205	33,2 %
Otra razón	24	15	39	0,1 %
Total	18.619	12.088	30.707	100,0 %

Fuente: Gestión de Diseño Muestral - DINEM.

De las 19.933 MEF que participaron a nivel nacional en el formulario, 12.129 tienen niños/as menores a 3 años. A continuación, se presenta la distribución de niños/as menores a tres años por MEF a nivel nacional, nacional urbano y nacional rural.

Tabla 13: Resumen de distribución de niños y niñas menores a tres años por MEF.

Distribución	MEF con 1 hijo/a	MEF con 2 hijos/as	MEF con 3 hijos/as	Total MEF	Total hijos/as
Nacional Urbano	7.109	303	5	7.417	7.730
Nacional Rural	4.496	215	1	4.712	4.929
Total	11605	518	6	12.129	12.659

Fuente: Gestión de Diseño Muestral - DINEM.

Las 12.129 MEF tienen un total de 12.659 niños/as menores a tres años. En la sección de lactancia materna participa un niño/a por cada MEF seleccionado mediante el método del cumpleaños más próximo, es decir 12.129 niños/as, 7.417 niños/as están en el nacional urbano y 4.929 niños/as en el nacional rural como se observa en la Tabla 13.

Desarrollo infantil

En el formulario de desarrollo infantil participa un niño/a menor a 5 años por hogar, seleccionado de manera aleatoria. Según la edad del niño/a seleccionado/a se aplican diferentes secciones del mismo. En la tabla 14 se presenta la cobertura del formulario de desarrollo infantil a nivel nacional, nacional urbano y nacional rural.

Tabla 14: Resultado de la entrevista del hogar.

Condición por hogar	Nacional urbano	Nacional rural	Nacional	Porcentaje
Efectiva	6.140	3.939	10.079	98,6%
Rechazo	24	7	31	0,3%
Ausente	55	50	105	1,0%
Otra razón	2	2	4	0,0%
Total	6.221	3.998	10.219	100,0%

Fuente: Gestión de Diseño Muestral - DINEM.

Fueron seleccionados 10.219 niños/as menores a 5 años a nivel nacional (98,6% efectivos), 6.140 en el nacional urbano y 3.939 en rural.

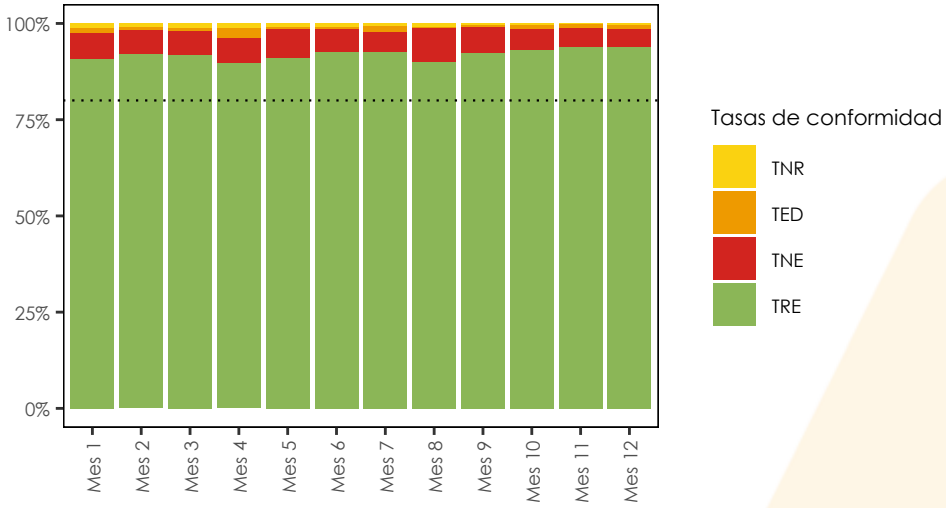
Paralización del levantamiento

Del 1 al 14 de diciembre del 2025 todos los procesos relacionados a la ENDI fueron paralizados debido a temas presupuestarios, lo que paralizó el levantamiento de información a mitad del mes 4.

La encuesta sufrió además, una paralización en el levantamiento del octavo mes operativo cuando el enlistamiento ya estaba en curso, por lo que se decidió postergar el enlistamiento de la primera jornada del noveno mes operativo.

Esta paralización modificó el tiempo entre el enlistamiento y encuesta, en comparación a los demás periodos, sin embargo al analizar la cobertura por mes, esta afectación es mínima, por lo tanto, no se evidencia un impacto significativo en la ejecución del diseño muestral planificado para la ENDI. En el siguiente gráfico se presenta la evolución de las tasas de conformidad de la ENDI por mes de levantamiento.

Ilustración 5: Evolución de las tasas de conformidad de la ENDI por mes.



Fuente: Gestión de Diseño Muestral - DINEM.

Como se puede observar en la ilustración anterior, en el cuarto y octavo mes operativo existe una reducción en la TRE en detrimento de la TNE, lo cual era esperado puesto que al aumentar el tiempo entre el enlistamiento y la encuesta, la condición de ocupación de las viviendas y su característica de tener a la población objetivo presenta más cambios. En el Anexo 4 la evolución de las tasas de conformidad de la ENDI por provincia y mes.

Factores de expansión

El diseño muestral de la ENDI es bietápico estratificado por conglomerados, por lo que para el cálculo de los factores de expansión se consideran las probabilidades de inclusión en cada una de las etapas, así como los respectivos ajustes por cobertura.

Probabilidad de primera etapa

En la primera etapa, se seleccionaron UPM con probabilidad proporcional al tamaño, cuya medida de tamaño es el número total de viviendas ocupadas. Por lo tanto, la probabilidad de inclusión de primera etapa de la UPM i en el estrato h es:

$$ppe_{i,h} := m_h \frac{N_i}{N_h},$$

donde:

$ppe_{i,h}$ = probabilidad de inclusión de primera etapa de la UPM i en el estrato h.
m_h = número de UPM seleccionadas en el estrato h.
N_i = medida de tamaño de la UPM i.
N_h = número total de viviendas en el estrato h.

Probabilidad de segunda etapa

Dado que el enlistamiento permitió identificar las viviendas con niños/as menores a 5 años, la selección de segunda etapa se realizó únicamente sobre este grupo, por lo que la probabilidad de segunda etapa es:

$$pse_i := \frac{k_i}{K_i},$$

donde:

pse_i = probabilidad de segunda etapa de las viviendas en la UPM i.
k_i = viviendas con niños/as menores a cinco años seleccionadas en la UPM i.
K_i = número total de viviendas con niños/as menores a cinco años en la UPM i.

Factor de expansión teórico

El factor de expansión teórico se define como la multiplicación de los inversos de las probabilidades de selección, es decir:

$$d_{0_{i,h}} := \frac{1}{ppe_{i,h} \cdot pse_i},$$

donde $d_{0_{i,h}}$ es el factor de expansión teórico de las viviendas en la UPM i del estrato h.

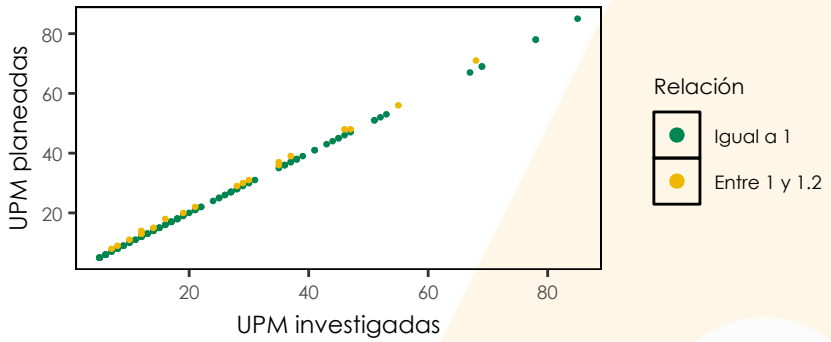
Ajuste de cobertura de primera etapa

Debido a diferentes motivos, se puede tomar la decisión de no investigar una UPM, por esta razón se debe realizar un ajuste a la probabilidad de selección de las UPM investigadas, el cual se define por:

$$a_1 := \frac{m}{m'},$$

donde $m = \sum_h m_h$ es el número total de UPM seleccionadas y m' es el número total de UPM investigadas. El ajuste se realiza de forma global, es decir, es el mismo para todas las UPM investigadas. En la siguiente figura se aprecia la relación entre la distribución de UPM investigadas y la distribución de UPM planificadas por estrato.

Ilustración 6: Relación de UPM investigadas y planificadas por estrato.



Fuente: Gestión de Diseño Muestral - DINEM.

Esta relación nos indican que los factores de expansión esperados se multiplican por un ajuste igual a 1 en 112 estratos, un ajuste entre 1 y 1.2 en otros 20, y ningún ajuste mayor a 1.2.

Factor de expansión de primera etapa

El inverso de la probabilidad de selección de primera etapa ajustada por cobertura es el factor de expansión de primera etapa.

$$w_{i,h} = \frac{1}{ppe_{i,h}} \cdot a_1$$

Factor de expansión ajustado por cobertura de primera etapa

El factor de expansión ajustado por cobertura de primera etapa queda definido por:

$$d_{1_{i,h}} := a_1 \cdot d_{0_{i,h}},$$

donde $d_{1_{i,h}}$ es el factor de expansión ajustado por cobertura de primera etapa de las viviendas en la UPM i que pertenece al estrato h.

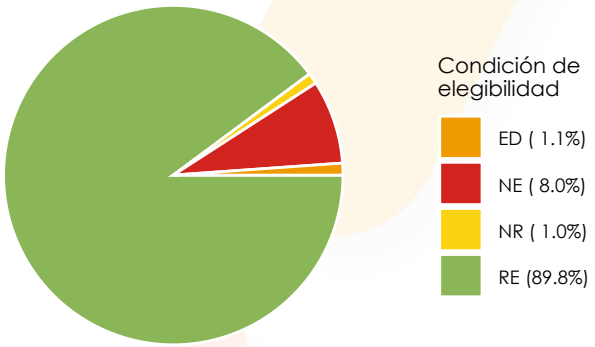
Ajuste de cobertura de segunda etapa

Al momento de la entrevista en campo, cada hogar de las viviendas muestreadas se clasificó en uno de los cuatro conjuntos anteriormente descritos (RE, NR, NE y ED). Si existen unidades que son no elegibles en la muestra, esto es evidencia de que hay unidades no elegibles tanto sobre ED como sobre lo no muestreado. Sin embargo, de acuerdo con la naturaleza de la encuesta, diferentes decisiones se pueden realizar sobre cómo manejar las unidades en ED.

Además, tanto los marcos como las muestras seleccionadas de ellos pueden contener unidades cuya elegibilidad no puede ser determinada. Por otro lado, en la mayoría de las encuestas, no todas las unidades elegibles responderán. De manera usual, se realizan ajustes por elegibilidad desconocida y no respuesta sobre las unidades elegibles para que estas se equiparen al total de la población elegible.

Para tal efecto, es necesario considerar la distribución inicial estimada de todos los hogares con niños/as menores de cinco años por condición de elegibilidad con base en el factor ajustado por cobertura de primera etapa (d_{1j})[2], la cual se aprecia en la Figura 7.

Ilustración 7: Distribución de hogares por condición de elegibilidad expandidas por d_1.



Fuente: Gestión de Diseño Muestral - DINEM.

Luego de aplicar los ajustes al cálculo del factor de expansión presentados a continuación, se observará como cambia progresivamente su distribución.

[2]De aquí en adelante se notará como d_{1_j} al factor de expansión ajustado por cobertura de primera etapa en la vivienda j.

Ajuste por elegibilidad desconocida

Una forma de manejar las unidades en ED es distribuir el peso total de la muestra entre aquellas unidades cuyo estado de elegibilidad se conoce (es decir, sobre las unidades en RE, NR y NE). Por lo general, la elegibilidad desconocida se trata mediante métodos simples en parte porque se sabe poco sobre los casos con elegibilidad desconocida y además porque la falta de respuesta se considera un problema más relevante que debe recibir atención.

La distribución de su peso utiliza un enfoque basado en clases, es decir, el ajuste se puede realizar a nivel de UPM, estrato o cualquier otro conjunto definido para tal efecto. Para la ENDI se realiza el ajuste a nivel de estrato, con base en el análisis de la cobertura a nivel de UPM y viviendas, así como por la naturaleza de la población objetivo de la misma. Por lo tanto, el ajuste por elegibilidad desconocida se define como:

$$a_{2_h} := \frac{\sum_{j \in s_h} d_{1_j}}{\sum_{j \in s_h} d_{1_j} - \sum_{j \in s_h \cap ED} d_{1_j}},$$

donde:

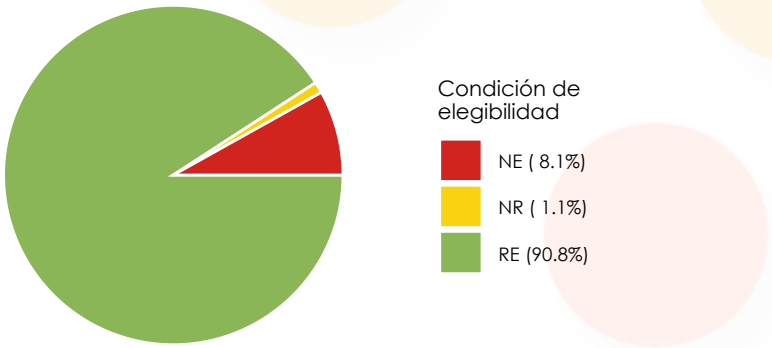
a_{2_h} = ajuste por elegibilidad desconocida en el estrato h.
s_h = muestra total de hogares en el estrato h.
d_{1_j} = factor de expansión de la vivienda j.

Por lo tanto, el factor de expansión ajustado por elegibilidad desconocida (d_{2_j}) queda definido de la siguiente manera:

$$d_{2_j} := \begin{cases} a_{2_h} \cdot d_{1_j} & j \in s_h \cap RE, \\ a_{2_h} \cdot d_{1_j} & j \in s_h \cap NR, \\ a_{2_h} \cdot d_{1_j} & j \in s_h \cap NE, \\ 0 & j \in s_h \cap ED. \end{cases}$$

Luego de aplicar este ajuste, las unidades pertenecientes a ED se distribuyen de manera proporcional entre las otras categorías, lo que da como resultado la distribución presentada a continuación:

Ilustración 8: Distribución de hogares por condición de elegibilidad expandidas por d_2.



Fuente: Gestión de Diseño Muestral - DINEM.

Ajuste por no respuesta

Ajustar por no respuesta puede ser o muy simple o muy elaborado, según cuánto se conoce de los no respondientes. Para la ENDI, la información disponible de los no respondientes es escasa y limitada, por lo tanto, la propensión a responder se modela únicamente con el estrato como variable explicativa.

Los factores se calculan a nivel de hogar, con el fin de no perder información en los ajustes de cobertura y en la calibración. Una vez realizado el ajuste por elegibilidad desconocida, las unidades NE se excluyen de la base, de manera que el ajuste por no respuesta se estima sobre $s_{h,R} = s_h \cap (RE \cup NR)$.

Sea r_j el indicador de respuesta, igual a 1 si el hogar $j \in RE$ y 0 si $j \in NR$. La propensión a responder ϕ_j se modela mediante una regresión logística cuya única covariable es el estrato:

$$\phi_j := P(r_j = 1 \mid j \in s_h), \qquad \text{logit}(\phi_j) = \ln\left(\frac{\phi_j}{1-\phi_j}\right) = \mathbf{z}_j^\top \boldsymbol{\beta},$$

donde:

ϕ_j = propensión a responder del hogar j.
$\mathbf{z}_j$ = vector de variables indicadoras del estrato al que pertenece el hogar j.
$\boldsymbol{\beta}$ = vector de parámetros del modelo, estimados por máxima verosimilitud.

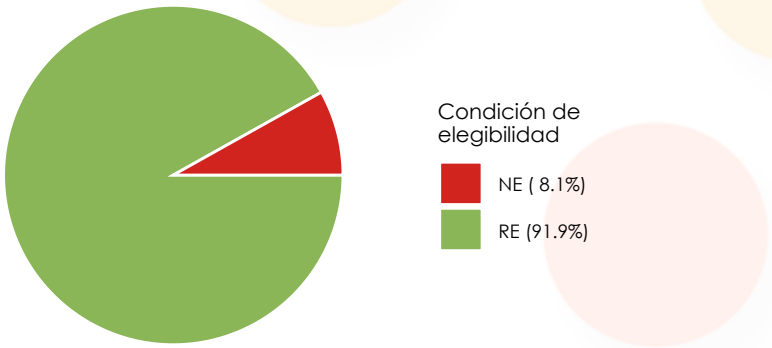
Al ser el estrato la única variable explicativa, la propensión estimada $\widehat{\phi}_h$ es constante dentro de cada estrato. El factor de expansión ajustado por no respuesta (d_{3_j}) queda definido de la siguiente manera:

$$d_{3_j} := \begin{cases} \dfrac{d_{2_j}}{\widehat{\phi}_h} & j \in s_h \cap RE, \\ 0 & j \in s_h \cap NR, \end{cases}$$

donde d_{2_j} es el factor de expansión ajustado por elegibilidad desconocida del hogar j.

El ajuste de no respuesta considera que, dentro de cada estrato, los hogares no respondientes tienen características similares a los respondientes. Así, al considerar los ajustes por cobertura, la distribución final es:

Ilustración 9: Distribución de hogares por condición de elegibilidad expandidas por d_3.

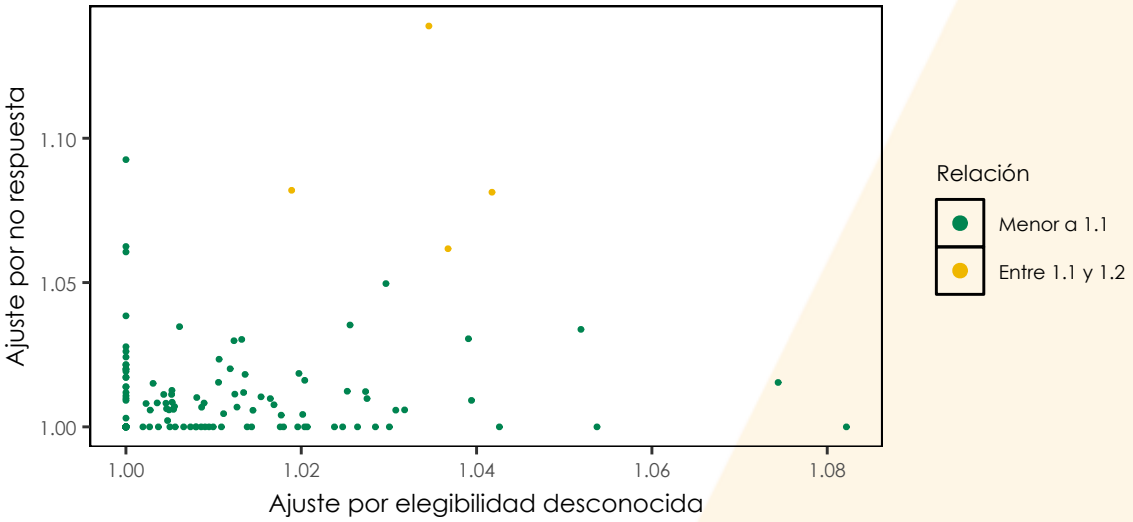


Fuente: Gestión de Diseño Muestral - DINEM.

En conclusión, el estimador de no elegibilidad sobre el marco de muestreo es del 8.1 %, lo que significa que aproximadamente una de cada 12 unidades identificadas como hogares con niños/as menores de 5 años en el enlistamiento no cuentan con la población objetivo.

En el siguiente gráfico se observa el impacto de aplicar los ajustes de elegibilidad desconocida y no respuesta aplicados sobre los respondientes elegibles (RE), a través de la relación entre los factores d_{3_j} y d_{1_j}, que se muestra a continuación:

Ilustración 10: Relación entre el factor d_{1_j} y d_{3_j} en función de los ajustes de cobertura de hogares.



Fuente: Gestión de Diseño Muestral - DINEM.

Se observa que en 129 estratos tiene un incremento menor a 1.1 y una relación entre 1.1 y 1.2 en 4 estratos.

Calibración a proyecciones poblacionales

Con el fin de que los factores de expansión reproduzcan la composición de la población objetivo, los factores ajustados por no respuesta se calibran a las Estimaciones y Proyecciones de Población 2026 de niños/as menores de 5 años a nivel de provincia, sexo y edad simple. Este paso se aplica a todos los factores de la ENDI (Hogares, Lactancia Materna y Desarrollo Infantil); solo cambia la población de referencia.

Para cada hogar j efectivo se define el vector de variables de calibración $\mathbf{x}_j$, cuyos componentes son conteos de niños/as menores de 5 años en el hogar:

$$\mathbf{x}_j = \left(x_j^{(1)}, \ldots, x_j^{(P)},\ x_j^{h},\ x_j^{(0)}, \ldots, x_j^{(3)}\right)^{\top},$$

donde:

$x_j^{(p)}$ = número de niños/as menores de 5 años del hogar j si este pertenece a la provincia p, y 0 en otro caso.
x_j^h = número de niños (sexo masculino) menores de 5 años en el hogar j.
$x_j^{(a)}$ = número de niños/as de a años cumplidos en el hogar j.
P = número de provincias.

Se omiten la categoría de sexo femenino y la edad de 4 años, ya que son linealmente dependientes del resto de variables (la suma de los totales provinciales es igual al total de niños/as menores de 5 años). Sea $\mathbf{t}_x$ el vector de totales poblacionales proyectados de estas variables. El factor calibrado (d_{4_j}) se obtiene mediante calibración lineal:

$$d_{4_j} := g_j \cdot d_{3_j}, \qquad g_j = 1 + \mathbf{x}_j^\top \boldsymbol{\lambda},$$

con

$$\boldsymbol{\lambda} = \left(\sum_{j \in s_R} d_{3_j}\, \mathbf{x}_j \mathbf{x}_j^\top \right)^{-1} \left(\mathbf{t}_x - \sum_{j \in s_R} d_{3_j}\, \mathbf{x}_j \right),$$

donde s_R es el conjunto de hogares efectivos, de modo que se cumple:

$$\sum_{j \in s_R} d_{4_j}\, \mathbf{x}_j = \mathbf{t}_x.$$

Recorte de factores calibrados

La calibración puede generar factores de expansión muy elevados en algunos hogares. Para controlar su dispersión se establece, en cada estrato h, una cota superior igual a 3.5 veces la mediana de los factores calibrados:

$$m_h := 3{,}5 \cdot \text{mediana} \left\{ d_{4_j} : j \in s_h \cap s_R \right\}.$$

Los factores que superan esta cota se fijan en m_h y el excedente se reparte entre el resto de factores del mismo estrato, de manera que se mantienen las propiedades de la calibración y la reproducción de los totales poblacionales proyectados. Se denota por d_{5_j} al factor calibrado y recortado del hogar j.

Factor de expansión normalizado

Debido a la naturaleza y flujos de los que dependen las preguntas presentes en una encuesta, esta puede presentar entrevistas parciales entre los informantes efectivos. Esto dificulta comparar las estimaciones de totales de la misma población objetivo obtenidas entre las secciones/variables, puesto que estas varían según la magnitud de la no respuesta parcial. Esta es una particularidad presente en encuestas como la ENDI, debido en gran medida a sus poblaciones objetivo.

Existen diferentes métodos para solventar este inconveniente que puede ser la imputación de los valores perdidos de las variables necesarias para los cálculos

de los indicadores, calcular un factor de expansión para cada una de los indicadores que considere la cobertura de las preguntas que construyen el indicador. En el [ICF International, 2012] se recomienda utilizar factores normalizados.

Al trabajar con factores normalizados, el número de casos no ponderados coincide con el número de casos ponderados y se calcula como el producto del factor de expansión de cada caso por un factor de normalización. El factor de expansión normalizado puede ser usado para estimar medias, proporciones, tasas y razones. Las estimaciones de total se pueden obtener a partir del factor calibrado y recortado d_{5_j}, que reproduce los totales poblacionales proyectados.

El factor de normalización se define en términos de los hogares efectivos:

$$f_n := \frac{n_R}{\sum_{j \in s_R} d_{5_j}},$$

donde:

f_n = factor de normalización.
n_R = número de hogares efectivos, es decir $n_R = |s_R|$.
d_{5_j} = factor de expansión calibrado y recortado del hogar j.

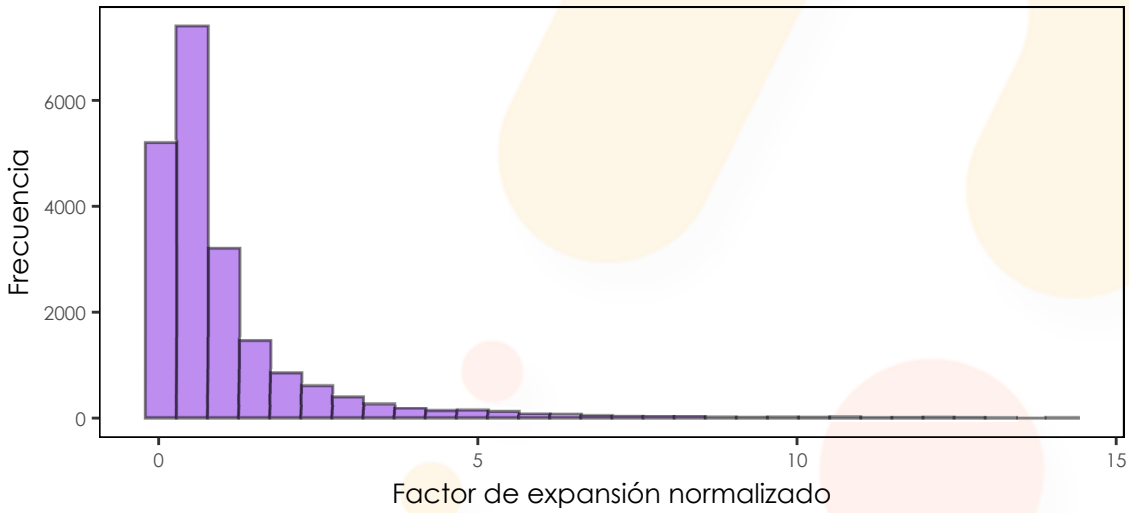
De esta forma, el factor de expansión normalizado d_{n_j} es:

$$d_{n_j} := f_n \cdot d_{5_j},$$

y se cumple que $\sum_{j \in s_R} d_{n_j} = n_R$.

En la Figura 11 se muestra la distribución de los factores de expansión normalizados a nivel nacional. Se puede apreciar que la distribución es la esperada para el diseño aplicado, sin embargo, tiene una cola derecha larga debido a que los factores de expansión corresponden al periodo de análisis.

Ilustración 11: Distribución de los factores de expansión normalizados.



Fuente: Gestión de Diseño Muestral - DINEM.

Factor de expansión Lactancia Materna

Para la sección de Lactancia Materna, se selecciona un niño/a menor a tres años por cada MEF encuestada mediante el método de cumpleaños más próximo. Esto implica la existencia de una tercera etapa de muestreo para esta población objetivo, cuya probabilidad de selección es:

$$pte_{j,k} := \frac{1}{L_k},$$

donde L_k es el número total de niños/as menores a tres años de la MEF k encuestada en el hogar j. Esta probabilidad no se incorpora de forma explícita en el factor, sino que se compensa, junto con la no respuesta, mediante el ajuste de cobertura que se describe a continuación.

El factor de partida es el factor de hogares calibrado y recortado d_{5_j}, antes de su normalización. Se considera como universo a los niños/as menores de 3 años de los hogares efectivos, y sea $r_{j,k}$ el indicador de que el niño/a k del hogar j cuenta con información del módulo de Lactancia Materna.

Ajuste de cobertura por no respuesta

Con el mismo enfoque de la sección de no respuesta, la propensión a contar con información de lactancia materna se modela mediante una regresión logística con el estrato como única variable explicativa:

$$\text{logit}(\phi_{j,k}^{lm}) = \mathbf{z}_j^\top \boldsymbol{\beta}^{lm}, \qquad \phi_{j,k}^{lm} := P(r_{j,k} = 1 \mid j \in s_h).$$

El factor ajustado por cobertura ($d_{6_{j,k}}^{lm}$) es:

$$d_{6_{j,k}}^{lm} := \frac{d_{5_j}}{\widehat{\phi}_h^{lm}} \quad \text{si } r_{j,k} = 1.$$

Calibración, recorte y normalización

Los factores $d_{6_{j,k}}^{lm}$ se calibran a las Estimaciones y Proyecciones de Población 2026 de niños/as menores de 3 años por provincia, sexo y edad simple, conforme al procedimiento de calibración lineal descrito anteriormente. En este caso, $\mathbf{x}_{j,k}$ es el vector de variables indicadoras de la provincia, el sexo masculino y la edad del niño/a (0 y 1 años; se omite la de 2 años por colinealidad), y $\mathbf{t}_x^{lm}$ los totales proyectados correspondientes:

$$d_{7_{j,k}}^{lm} := g_{j,k}^{lm} \cdot d_{6_{j,k}}^{lm}, \qquad \sum_{(j,k)\in s_R^{lm}} d_{7_{j,k}}^{lm}\, \mathbf{x}_{j,k} = \mathbf{t}_x^{lm},$$

donde s_R^{lm} es el conjunto de niños/as con información de lactancia materna. Luego se aplica el recorte con cota superior igual a 3.5 veces la mediana de los factores calibrados en cada estrato, sin alterar la calibración, para obtener $d_{8_{j,k}}^{lm}$.

Finalmente, el factor de expansión de Lactancia Materna normalizado $dlm_{n_{j,k}}$ se calcula en términos de niños/as efectivos/as:

$$dlm_{n_{j,k}} = fn_{lm} \cdot d^{lm}_{8_{j,k}}, \qquad fn_{lm} := \frac{n^{lm}_R}{\sum_{(j,k)\in s^{lm}_R} d^{lm}_{8_{j,k}}},$$

donde fn_{lm} es el factor de normalización de Lactancia Materna y $n^{lm}_R = |s^{lm}_R|$ es el número de niños/as efectivos/as del módulo.

Factor de expansión Desarrollo Infantil

Para la sección de Desarrollo Infantil, se selecciona un niño/a menor a 5 años por cada hogar mediante el aplicativo de ingreso de información de manera aleatoria. Esto implica la existencia de una tercera etapa de muestreo para esta población objetivo. Por lo tanto, la probabilidad de selección de tercera etapa es:

$$pte_j := \frac{1}{D_j},$$

donde D_j es el número total de niños/as menores a 5 años en el hogar j. Solo los niños/as seleccionados/as forman parte del análisis; los no seleccionados intervienen únicamente en el conteo D_j.

Ajuste por levantamiento en el segundo semestre

El módulo de Desarrollo Infantil se recoge únicamente a partir del segundo semestre de levantamiento (UPM investigadas desde el mes 7), por lo que los hogares donde se aplica representan solo una parte de la muestra efectiva. Para que representen a la población en su conjunto, antes de calcular los factores del módulo se ajusta el factor de hogares calibrado y recortado d_{5_j} por la razón entre el número de hogares efectivos de todo el operativo y el número de hogares efectivos con levantamiento de Desarrollo Infantil:

$$a_{sem} := \frac{n_R}{n^{dein}_R}, \qquad d^*_{5_j} := a_{sem} \cdot d_{5_j},$$

donde n_R es el número de hogares efectivos en todo el operativo y n^{dein}_R el número de hogares efectivos en el universo de Desarrollo Infantil. Con los valores de esta edición, $a_{sem} = 20\,950/10\,790 \approx 1{,}94$.

Factor de expansión

Al incorporar la tercera etapa de selección se obtiene:

$$d^{dein}_{6_j} := d^*_{5_j} \cdot \frac{1}{pte_j} = d^*_{5_j} \cdot D_j.$$

Sea r_j^{dein} el indicador de que la entrevista del niño/a seleccionado del hogar j fue efectiva. De manera análoga al ajuste de no respuesta de hogares, se realiza un ajuste de cobertura por no respuesta, en el que la propensión se modela mediante una regresión logística con el estrato como única variable explicativa:

$$\text{logit}(\phi_j^{dein}) = \mathbf{z}_j^\top \boldsymbol{\beta}^{dein}, \qquad d_{7_j}^{dein} := \frac{d_{6_j}^{dein}}{\widehat{\phi}_h^{dein}} \quad \text{si } r_j^{dein} = 1.$$

Posteriormente, los factores se calibran a las Estimaciones y Proyecciones de Población 2026 de niños/as menores de 5 años por provincia, sexo y edad simple (los mismos totales $\mathbf{t}_x$ utilizados para hogares), se recortan con una cota igual a 3.5 veces la mediana por estrato sin alterar la calibración, y se normalizan en términos de niños/as efectivos/as:

$$d_{8_j}^{dein} := g_j^{dein} \cdot d_{7_j}^{dein}, \qquad ddein_{n_j} = fn_{dein} \cdot d_{9_j}^{dein}, \qquad fn_{dein} := \frac{n_R^{dein}}{\sum_{j \in s_R^{dein}} d_{9_j}^{dein}},$$

donde $d_{9_j}^{dein}$ es el factor calibrado y recortado, fn_{dein} es el factor de normalización de Desarrollo Infantil, s_R^{dein} es el conjunto de niños/as con entrevista efectiva de Desarrollo Infantil y $n_R^{dein} = |s_R^{dein}|$.

Resumen de los factores por módulo

La Tabla 15 resume los pasos aplicados en cada uno de los factores de expansión.

Tabla 15: Pasos de cálculo de los factores de expansión por módulo.

Paso	**Hogares**	**Lactancia materna**	**Desarrollo Infantil**
Unidad de análisis	Hogar	Niño/a menor de 3 años	Niño/a menor de 5 años seleccionado
Factor de partida	d_0 con cobertura de UPM, elegibilidad desconocida y no respuesta (modelo logístico por estrato)	Factor de hogares calibrado y recortado d_5	d_5 ajustado por levantamiento en el segundo semestre y por tercera etapa
Ajuste de cobertura por no respuesta	Incluido en el factor de partida	Modelo logístico por estrato	Modelo logístico por estrato
Calibración	Provincia, sexo y edad simple (menores de 5 años)	Provincia, sexo y edad simple (menores de 3 años)	Provincia, sexo y edad simple (menores de 5 años)

Recorte	3.5 veces la mediana por estrato	3.5 veces la mediana por estrato	3.5 veces la mediana por estrato
Normalización	Hogares efectivos	Niños/as efectivos/as	Niños/as efectivos/as

Fuente: Gestión de Diseño Muestral - DINEM.

Estimaciones de características

Una vez obtenidos los factores de expansión normalizados, para calcular las estimaciones de la ENDI nos apoyamos en el estimador de Horvitz-Thompson, el cual se utiliza para estimar el total de una característica determinada[3]. Está dado por [Särndal et al., 1992]:

$$\hat{Y}_{m_{HT}} = \sum_{k \in s_m} d_k y_k,$$

donde:

$\hat{Y}_{m_{HT}}$ = estimador HT de la variable y sobre el dominio m.
i = índice que recorre las UPM en el dominio m.
j = índice que recorre las viviendas en la UPM i del dominio m.
y_k = valor de la variable y para la unidad de análisis k.

A partir de la definición del factor de normalización utilizado en la ENDI, se obtiene que $\widehat{\overline{Y}}_m = \hat{Y}_{m_{HT}}/n$; así, se obtienen estimadores de tipo media.

Error de muestreo con factores de expansión calibrados

La estimación de la varianza debe ser coherente con el estimador puntual y con el diseño muestral. Cuando se utilizan factores de expansión calibrados, no es recomendable aplicar sin ajustes una expresión de varianza derivada para los factores de expansión originales: la calibración incorpora restricciones sobre variables auxiliares y estas deben reflejarse en la aproximación de la varianza.

Sea w_k el factor de expansión calibrado de la unidad k, y_k la variable de interés y $\mathbf{x}_k$ el vector de variables auxiliares utilizadas en la calibración, cuyos totales poblacionales $\mathbf{t}_x$ son conocidos. El estimador calibrado de un total es:

$$\hat{t}_{y,\text{cal}} = \sum_{k \in s} w_k y_k, \qquad \sum_{k \in s} w_k \mathbf{x}_k = \mathbf{t}_x.$$

Para una calibración de tipo lineal, una aproximación usual basada en linealización consiste en estimar la varianza mediante los residuos de una regresión auxiliar. Se define:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{k \in s} w_k q_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \sum_{k \in s} w_k q_k \mathbf{x}_k y_k, \qquad \hat{e}_k = y_k - \mathbf{x}_k' \hat{\boldsymbol{\beta}},$$

donde q_k representa la constante asociada al método de calibración. En consecuencia, la varianza aproximada del estimador calibrado se obtiene al estimar

[3] Nótese que, al utilizar factores normalizados, el estimador de total calculado aproxima el total muestral, más no el total poblacional.

la varianza de diseño del total de residuos:

$$\widehat{\text{Var}}(\hat{t}_{y,\text{cal}}) \;\approx\; \widehat{\text{Var}}_d\left(\sum_{k\in s} d_k \hat{e}_k\right),$$

donde d_k denota el factor de expansión previo a la calibración y $\widehat{\text{Var}}_d$ debe respetar el diseño de selección, incluidos la estratificación y el muestreo por etapas. Esta expresión es una aproximación de linealización, no una identidad exacta válida para cualquier procedimiento de calibración. Si el diseño o el algoritmo de ponderación tiene características adicionales, por ejemplo: recorte de pesos, ajustes por no respuesta o calibración no lineal, estas deben evaluarse al seleccionar el método de varianza.

Aproximación por conglomerados de última etapa. Cuando se utiliza el método de la última UPM, se construye primero la contribución linealizada de cada UPM i del estrato h:

$$\hat{E}_{hi} = \sum_{k\in s_{hi}} d_k \hat{e}_k, \qquad \overline{E}_h = \frac{1}{m_h}\sum_{i=1}^{m_h} \hat{E}_{hi},$$

donde m_h es el número de UPM muestrales del estrato h. Bajo la aproximación usual de conglomerados de última etapa, una forma del estimador de varianza es:

$$\widehat{\text{Var}}(\hat{t}_{y,\text{cal}}) \approx \sum_h \frac{m_h}{m_h - 1}\sum_{i=1}^{m_h}(\hat{E}_{hi} - \overline{E}_h)^2.$$

Esta expresión requiere al menos dos UPM muestrales por estrato y corresponde a una aproximación por conglomerados; los factores de escala y las correcciones deben ajustarse a la especificación efectiva del diseño. Si las probabilidades de selección, los factores previos a la calibración o la estructura de estratos no están disponibles, no debe asumirse que esta fórmula reproduce por sí sola la varianza correcta.

En la ENDI, la variable `fexp` descrita en la tabla 16 corresponde a un factor normalizado. Por tanto, para estimar medias o proporciones debe mantenerse la coherencia entre el estimador puntual y su varianza. Si la estimación es una razón ponderada,

$$\hat{\overline{Y}} = \frac{\sum_{k\in s} w_k y_k}{\sum_{k\in s} w_k},$$

la variable linealizada puede expresarse como

$$\hat{e}_k = y_k - \hat{\overline{Y}},$$

o, en una formulación general de razón, como el residuo del numerador respecto al denominador. La varianza de la media o proporción se estima a partir de la varianza de diseño de la suma ponderada de esos residuos, con base en la estructura de estratos y UPM. Si los factores normalizados suman exactamente el tamaño muestral, la media ponderada puede escribirse como un total ponderado dividido por dicho tamaño; aun así, debe verificarse que la normalización y la calibración se hayan aplicado de forma compatible.

Finalmente, el error estándar, el coeficiente de variación y el intervalo de confianza se calculan como:

$$EE(\hat{\theta}) = \sqrt{\widehat{\text{Var}}(\hat{\theta})}, \qquad CV(\hat{\theta}) = \frac{EE(\hat{\theta})}{|\hat{\theta}|} \times 100.$$

Los grados de libertad y el valor crítico para el intervalo de confianza deben ser consistentes con el diseño complejo utilizado.

Métodos de estimaciones de errores para diseños muestrales complejos

Aunque la selección del diseño de muestreo y el estimador sean de libre elección para los investigadores, no lo es el cálculo de las medidas de confiabilidad y precisión. Dado que la base científica sobre la cual descansa el muestreo es la inferencia estadística, se deben respetar las normas básicas para la asignación y posterior cálculo de los errores estándar, el cual cuantifica la incertidumbre acerca de las estimaciones en una encuesta.

Los métodos de estimación de los errores muestrales pueden clasificarse en cuatro categorías:

a) Métodos exactos.
b) Métodos de la última UPM.
c) Aproximaciones por linealización.
d) Técnicas de replicación.

Para la descripción de los métodos se ha tomado como referencia los textos de [Kish and Frankel, 1974], [Wolter, 1985] y [Lehtonen and Pahkinen, 1995]. A continuación, se describe las principales características de cada uno de los métodos de estimación de errores para el muestreo complejo:

- Los métodos exactos pueden ser utilizadas para estimar totales, medias, proporciones y tamaños.
- La linealización de Taylor debe ser utilizada para estimar parámetros no lineales como razones, medias dentro de dominios, cuartiles o funciones de distribución.
- La técnica de la última UPM junto con la linealización de Taylor puede ser utilizada para estimar la varianza de los indicadores de interés de las encuestas dirigidas a hogares que tengan diseños muestrales complejos. Esta es la técnica que por defecto utiliza el software SPSS.
- Las técnicas de replicación pueden ser usadas para estimar eficientemente todos los parámetros de interés, sin importar su forma funcional.
- La comparación general entre los métodos de linealización y replicación no generan resultados idénticos del error de muestreo, pero hay que señalar que existen estudios [Kish and Frankel, 1974] que concluyen que las diferencias presentadas no son significativas cuando se trata de grandes muestras.

El INEC y sus usuarios utilizan diversos programas estadísticos para la estimación de los parámetros de interés, entre los cuales los más utilizados son SPSS, Stata

y R. En virtud de las características de cada uno de los métodos, es la técnica de la última UPM en combinación con la linealización de Taylor, la cual induce a una adecuada aproximación del error muestral sobre los indicadores más importantes de las encuestas dirigidas a hogares, pero sobre todo resalta por su facilidad de cálculo y replica. En este sentido, será esta la técnica utilizada para la estimación de los errores muestrales en la ENDI.

Las variables requeridas para declarar el diseño muestral en los programas estadísticos (SPSS, Stata y R) y que permiten calcular los errores de muestreo bajo el esquema anterior son presentadas en la tabla 16, donde se describe las etiquetas de las variables identificadoras de las UPM, estratos y factor de expansión.

Tabla 16: Variables requeridas para la declaración del diseño muestral.

Característica	Variable	Descripción
Unidad primaria de muestreo	id_upm	Unidad primaria de muestreo de ENDI
Estratos	estrato	Identificación de estrato muestral (aproximación de clasificación socio-demográfica)
Factor de expansión	fexp	Factor de expansión normalizado

Fuente: Gestión de Diseño Muestral - DINEM.

Estimación de totales mediante fuentes de información auxiliares

Como se menciona en las secciones anteriores, el factor de expansión normalizado puede ser usado para estimar medias, proporciones, tasas y razones de los principales indicadores de la ENDI. Sin embargo, si es necesario contar con estimaciones de totales para los mismos, estas se pueden obtener con el apoyo de información auxiliar disponible para la población objetivo de los mismos[4]. Por ejemplo:

- Sea $\hat{\overline{Y}}_m$ un estimador de proporción asociado a la población objetivo niños/as menores a 5 años en el dominio m,
- Sea N_{m0a5} el total auxiliar referente al total poblacional de niños/as menores a 5 años en el dominio m,

se tiene que:

$$\hat{Y}_{m_{aux}} = \hat{\overline{Y}}_m \cdot N_{m0a5}.$$

es el estimador de total auxiliar asociado a la población N_{m0a5} que cuenta con la característica $\hat{\overline{Y}}_m$. Para obtener los límites del intervalo de confianza asociados al estimador de total auxiliar, se reemplaza $\hat{\overline{Y}}_m$ por $LI(\hat{\overline{Y}}_m)$ y $LS(\hat{\overline{Y}}_m)$, respectivamente.

[4]El total auxiliar puede ser obtenido de diferentes fuentes de información, entre ellas: censo de población, proyecciones poblacionales y registros administrativos

Referencias

[Comunidad Andina de Naciones, 2000] Comunidad Andina de Naciones (2000). *Seminario: Censo 2000 de Población y Vivienda de los Países Andinos*. CAN, Lima.

[ICF International, 2012] ICF International (2012). *Demographic and Health Survey Sampling and Household Listing Manual. MEASURE DHS*. ICF International, Calverton, Maryland.

[Instituto Nacional de Estadística y Censos, 2024] Instituto Nacional de Estadística y Censos (2024). *Marco Maestro de Muestreo para encuestas de hogares*. INEC, Quito.

[Kish and Frankel, 1974] Kish, L. and Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1):1–22.

[Lehtonen and Pahkinen, 1995] Lehtonen, R. and Pahkinen, E. J. (1995). *Practical Methods for Design and Analysis of Complex Surveys*. John Wiley, Chichester.

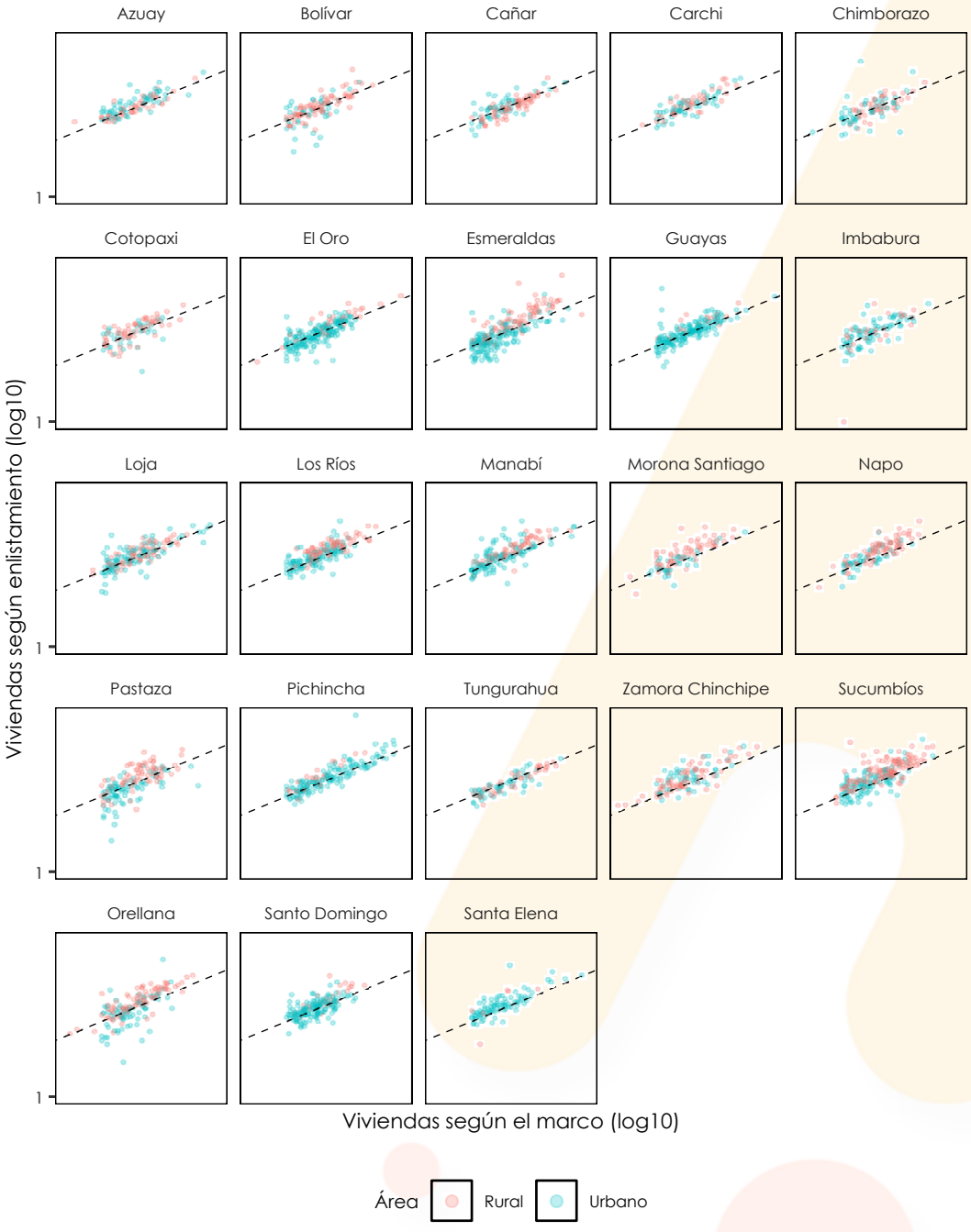
[Särndal et al., 1992] Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer, Nueva York.

[Wolter, 1985] Wolter, K. M. (1985). *Introduction to variance estimation*. Springer-Verlag, Nueva York.

Anexos

Anexo 1

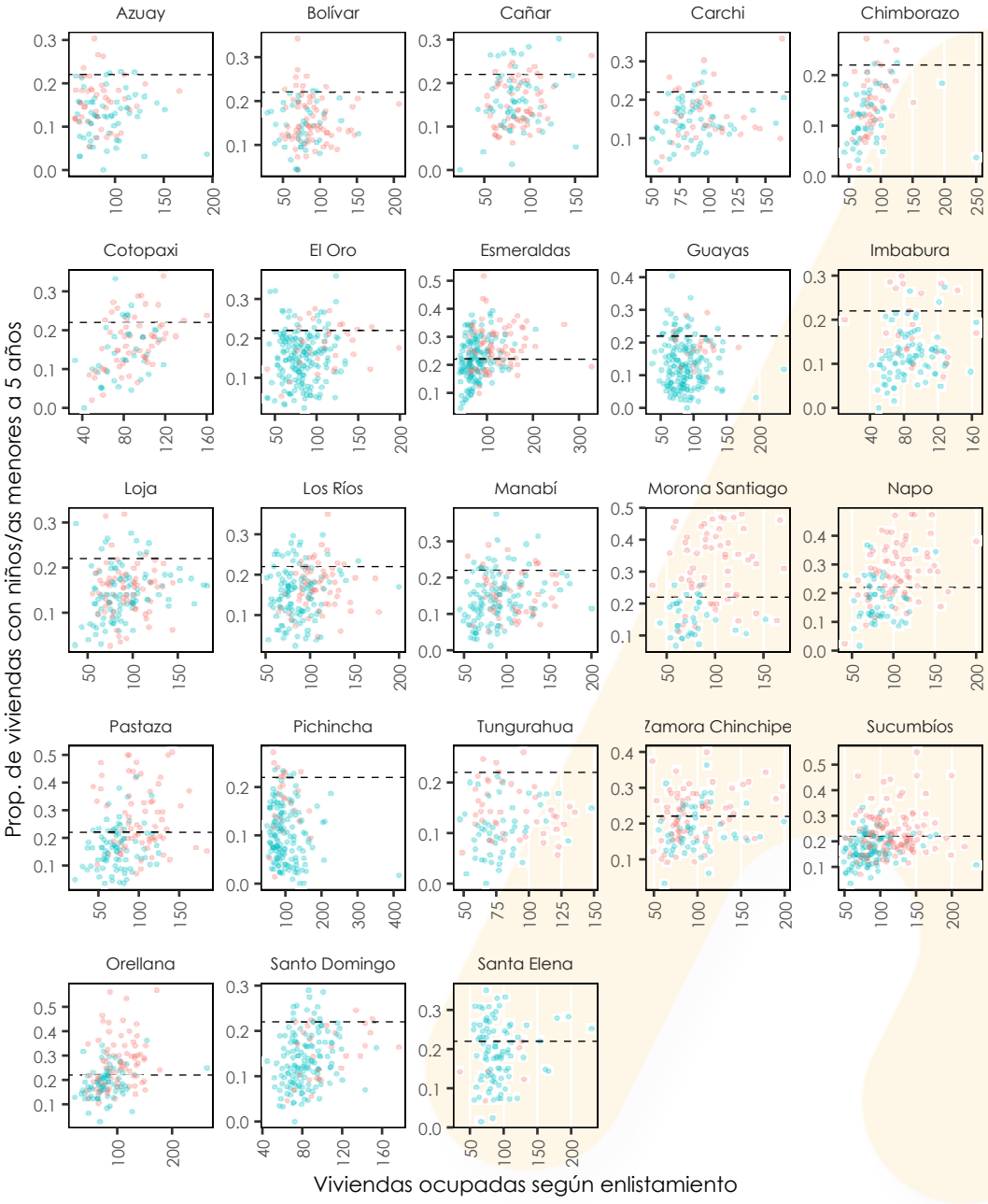
Ilustración 12: Viviendas ocupadas según marco y enlistamiento.



Fuente: Gestión de Diseño Muestral - DINEM.

Anexo 2

Ilustración 13: Proporción de viviendas con niños/as menores a 5 años por UPM.



Fuente: Gestión de Diseño Muestral - DINEM.

Anexo 3

Tabla 17: Condición de elegibilidad y tasa de conformidad (%) por provincia.

Provincia	NR	ED	NE	RE	Total	TNR	TED	TNE	TRE
Azuay	18	10	48	640	716	2,5	1,4	6,7	89,4
Bolívar	3	9	54	744	810	0,4	1,1	6,7	91,9
Cañar	6	9	65	745	825	0,7	1,1	7,9	90,3
Carchi	0	1	20	636	657	0	0,2	3	96,8
Cotopaxi	5	5	43	572	625	0,8	0,8	6,9	91,5
Chimborazo	2	6	38	570	616	0,3	1	6,2	92,5
El Oro	20	13	117	1.223	1.373	1,5	0,9	8,5	89,1
Esmeraldas	0	7	54	1.516	1.577	0	0,4	3,4	96,1
Guayas	26	25	79	1.108	1.238	2,1	2	6,4	89,5
Imbabura	2	0	18	674	694	0,3	0	2,6	97,1
Loja	12	14	65	1.047	1.138	1,1	1,2	5,7	92
Los Ríos	11	15	90	1.114	1.230	0,9	1,2	7,3	90,6
Manabí	12	17	91	992	1.112	1,1	1,5	8,2	89,2
Morona Santiago	6	12	57	578	653	0,9	1,8	8,7	88,5
Napo	4	15	32	785	836	0,5	1,8	3,8	93,9
Pastaza	9	10	71	969	1.059	0,8	0,9	6,7	91,5
Pichincha	3	3	46	993	1.045	0,3	0,3	4,4	95
Tungurahua	11	1	31	576	619	1,8	0,2	5	93,1
Zamora Chinchipe	4	15	41	778	838	0,5	1,8	4,9	92,8
Sucumbíos	1	4	74	1.508	1.587	0,1	0,3	4,7	95
Orellana	7	13	66	993	1.079	0,6	1,2	6,1	92
Santo Domingo	13	18	119	977	1.127	1,2	1,6	10,6	86,7
Santa Elena	7	9	45	594	655	1,1	1,4	6,9	90,7
Nacional	182	231	1.364	20.332	22.109	0,8	1	6,2	92

Fuente: Gestión de Diseño Muestral - DINEM.

Anexo 4

Ilustración 14: Tasas de conformidad por provincia y mes



Fuente: Gestión de Diseño Muestral - DINEM.

INEC | Buenas cifras, **mejores vidas**

ENCUESTA NACIONAL SOBRE DESNUTRICIÓN *Infantil*



@InecEcuador @ecuadorencifras @ecuadorencifras INECEcuador