

Diseño Muestral

Encuesta Nacional sobre
Desnutrición Infantil - ENDI

Septiembre, 2023





Diseño Muestral de la Encuesta Nacional sobre Desnutrición Infantil - ENDI 2022-2023

Quito - Ecuador, 2023

Dirección/Departamento

Dirección de Infraestructura Estadística y Muestreo

Unidad

Gestión de Diseño Muestral

Elaborado por:

Javier Núñez
Angel Gaibor
Giovanella Peña

Revisado por:

Christian Garcés

Aprobado por:

Julio Muñoz



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	10
Enlistamiento de las UPM	10
Selección de viviendas	12
Cobertura	14
Condición de ocupación y resultado de entrevista	14
Tasas de conformidad	15
Población objetivo	16
Lactancia materna	18
Desarrollo infantil	18
Factores de expansión	20
Probabilidad de primera etapa	20
Probabilidad de segunda etapa	20
Factor de expansión teórico	21
Ajuste de cobertura de primera etapa	21
Ajuste de cobertura de segunda etapa	22
Factor de expansión normalizado	25
Factor de expansión lactancia materna	26
Factor de expansión desarrollo infantil	27
Estimaciones de características	28
Error de muestreo	28
Métodos de estimaciones de errores para diseños muestrales complejos	29
Estimación de totales usando fuentes de información auxiliares	30
Anexos	33
Anexo 1	33
Anexo 2	34

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, viviendas y UPM por provincia. . .	7
5.	Distribución de la muestra por mes y área.	9
6.	Tamaño muestral final de UPM.	9
7.	Número de UPM planificadas y enlistadas.	11
8.	Número de UPM por número de viviendas seleccionadas.	13
9.	Condición de elegibilidad de las viviendas.	14
10.	Resumen de distribución de niños menores a 5 años por año y meses. . .	17
11.	Resultado de la entrevista formulario MEF.	18
12.	Resumen de distribución de niños menores a tres años por MEF. . . .	18
13.	Resultado de la entrevista del hogar.	19
14.	Variables requeridas para la declaración del diseño muestral.	30
15.	Condición de elegibilidad y tasa de conformidad (%) por provincia. . .	35

Lista de Figuras

1.	Viviendas ocupadas según marco y enlistamiento.	12
2.	Proporción de viviendas con niños menores a cinco años por UPM. . .	13
3.	Tasas de conformidad por provincia.	16
4.	Distribución de niños menores de 5 años por años y meses.	17
5.	Relación de UPM investigadas y planificadas por estrato.	21
6.	Distribución de viviendas por condición de elegibilidad expandidas por d_1	22
7.	Distribución de viviendas por condición de elegibilidad expandidas por d_2	23
8.	Distribución de viviendas por condición de elegibilidad expandidas por d_3	24
9.	Relación entre el factor d_{1j} y d_{3j} en función de los ajustes de cobertura de vivienda.	25
10.	Distribución de los factores de expansión normalizados.	26



11. Viviendas ocupadas según marco y enlistamiento.	33
12. Proporción de viviendas con niños menores a 5 años por UPM.	34





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 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, clasificándose 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, determinando sus integrantes 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 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 VII Censo de Población y VI de Vivienda del Ecuador (CPV - 2010) junto con las actualizaciones cartográficas realizadas en el periodo 2011-2018.

En la Tabla 1 se presenta la proporción de viviendas catalogadas como ocupadas según el precenso CPV-2010 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	Porcentaje
CPVCENEC	2010	2.685.474	72,2%
ECV20132014	2013	212.549	5,7%
CENSOGAL15	2015	8.520	0,2%
PROYEC2015	2015	52.933	1,4%
ACTENEMDU	2015	611.248	16,4%
ACTUAL2017	2017	147.594	4,0%
Total MMM	2018	3.718.318	100,0%

Fuente: Gestión de Diseño Muestral - DINEM.

Las actualizaciones cartográficas hacen referencia a:

- CPVCENEC: VII Censo de población y VI de vivienda 2010.
- ECV20132014: Actualización cartográfica asociada a la Encuesta de Condiciones de Vida 2013-2014.
- CENSOGAL15: Censo de población y vivienda Galápagos 2015.
- PROYEC2015: Actualización cartográfica asociada a la Encuesta Nacional de Empleo, Desempleo y Subempleo ENEMDU para el periodo 2015.
- ACTENEMDU: Actualización cartográfica asociada a la Encuesta Nacional de Empleo, Desempleo y Subempleo ENEMDU para el periodo 2015.
- ACTUAL2017: Actualización cartográfica asociada a la Encuesta Nacional de Empleo, Desempleo y Subempleo ENEMDU para el periodo 2017.

Unidades Primarias de Muestreo

Cada vivienda del MMM 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, en el caso de la parte

amanzanada se utilizó como límite inferior 100 viviendas ocupadas y en la parte dispersa 60. 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 2018, año en el que se construyó el marco de muestreo actual, permitiéndonos identificar adecuadamente los dominios de diseño geográficos definidos para la ENDI. Además, contiene variables que clasifican las UPM en función al estrato al que pertenecen. Para más información referirse a (INEC, 2022). En la Tabla 2 se presenta el total de UPM por provincia y área presentes en el MMM.

Tabla 2: Distribución de UPM por provincia y área.

Provincia	Rural	Urbano	Total
Azuay	997	927	1.924
Bolívar	386	110	496
Cañar	386	190	576
Carchi	205	169	374
Cotopaxi	856	241	1.097
Chimborazo	825	390	1.215
El Oro	336	1.075	1.411
Esmeraldas	467	530	997
Guayas	1.296	7.292	8.588
Imbabura	368	495	863
Loja	649	430	1.079
Los Ríos	859	901	1.760
Manabí	1.426	1.529	2.955
Morona Santiago	262	85	347
Napo	147	64	211
Pastaza	118	74	192
Pichincha	1.010	5.516	6.526
Tungurahua	915	574	1.489
Zamora Chinchipe	155	56	211
Sucumbíos	235	145	380
Orellana	175	106	281
Santo Domingo	246	658	904
Santa Elena	50	422	472
Total	12.369	21.979	34.348

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 Co-

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, esta 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, utilizando 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 socioeconómicas.
- 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 MMM del INEC, todas las UPM son clasificadas en algún estrato, considerando los siguientes criterios:

- Clasifica a las UPM por provincia y área. Este criterio también considera a las parroquias principales (Quito, Guayaquil, Cuenca, Ambato y Machala).
- 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, no se realiza esta clasificación.

Para más información acerca de la estratificación del MMM, remitirse al documento (Instituto Nacional de Estadística y Censos, 2019).

Nota: Puesto que la ENDI considera la realización de varias rondas anuales, se ha previsto migrar y ajustar su proceso de diseño muestral al Marco de Muestreo construido desde los resultados finales del Censo de Población y Vivienda 2022.

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 menores a cinco años.

En la segunda etapa de muestreo, se selecciona de manera aleatoria un número fijo de viviendas, siendo 8 el número de viviendas por UPM escogido 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 menor de 3 años por mujer en edad fértil efectiva, mientras que en la segunda, se escoge un niño mejor de 5 años por hogar efectivo en el segundo semestre de levantamiento de información. En los dos casos, la selección se hace mediante el criterio de cumpleaños más próximo. 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 utilizando 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:

$$e_{rel} = 0,15$$

$$z = 1,96$$

$$tnr = 0,20$$

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 en realidad no cumplían con esta característica. Esta decisión mitiga potenciales tasas de no respuesta altas en los dominios de estudio, asegurando alcanzar el total de niños 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 menores a cinco años, el mismo se determinó a partir de la ENSANUT 2018. Además, a partir de la ENSANUT 2018 se determinó el efecto de diseño y el promedio de niños menores a cinco 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 menores a 5 años	Efecto de diseño	Promedio de niños 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 menores a 5 años	Efecto de diseño	Promedio de niños 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

Utilizando los parámetros descritos en la sección anterior y aplicando la fórmula para determinar el tamaño muestral, se obtiene el número de niños 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 para el promedio de niños por vivienda en cada provincia, el mismo que se obtuvo a partir de la ENSANUT 2018. 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, viviendas y UPM por provincia.

Provincia	Número de niños 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

Provincia	Número de niños menores a cinco años	Número de viviendas	Número de UPM
Morona Santiago	717	538	68
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:** Esto está directamente relacionado con la parte operativa, es decir, permite tener control del 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 el número de UPM 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 ordenando las UPM seleccionadas por provincia, área y estrato, y repitiendo el vector formado por los números del 1 al 12 en el siguiente orden 1, 4, 7, 10, 2, 5, 8, 11, 3, 6, 9, 12.

De esta forma, se tiene que el número de UPM por provincia es el mismo en cada uno de los meses, garantizando el mismo número de encuestadores en cada uno de los meses. Para ejemplificar los resultados de la distribución a nivel de provincia-área y provincia-estrato, se escogen las provincias 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 los diferentes ajustes al distribuir la muestra, el tamaño final en número de UPM cambia. En la Tabla 6 se presenta el tamaño final de UPM por provincia y por mes.

Tabla 6: Tamaño muestral final de UPM.

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

Provincia	Número de UPM	Número de UPM por mes
Chimborazo	108	9
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.

Selección de la muestra de UPM

Tras determinar el tamaño final de UPM, se procedió a realizar la selección del primer año de forma proporcional al tamaño dentro de cada estrato (siendo la medida de tamaño de la UPM el número de viviendas ocupadas según el Marco de Muestreo) y así planificar el enlistamiento de manera oportuna, el cual, actualiza el listado de viviendas ocupadas corrigiendo en parte la desactualización del Marco de Muestreo.

Enlistamiento de las UPM

Dos meses 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, diferenciándolas 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 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 los inconvenientes causados por el Paro Nacional del Ecuador entre el 13 y 30 de junio de 2022, además de cierres en la frontera con Colombia y sin accesibilidad a conjuntos habitacionales, que no permitieron el ingreso del equipo de encuestadores. En la Tabla 7 se detalla el número de UPM planificadas y el número de UPM enlistadas.

Tabla 7: Número de UPM planificadas y enlistadas.

Dominio	N° de UPM planificadas	N° de UPM enlistadas
Azuay	144	144
Bolívar	108	108
Cañar	108	108
Carchi	84	84
Cotopaxi	72	72
Chimborazo	108	108
El Oro	192	192
Esmeraldas	144	141
Guayas	156	155
Imbabura	84	84
Loja	96	96
Los Ríos	108	108
Manabí	156	156
Morona Santiago	72	70
Napo	96	96
Pastaza	84	83
Pichincha	312	310
Tungurahua	180	172
Zamora Chinchipe	108	108
Sucumbíos	120	118
Orellana	72	71
Santo Domingo	168	167
Santa Elena	84	84
Total	2856	2835

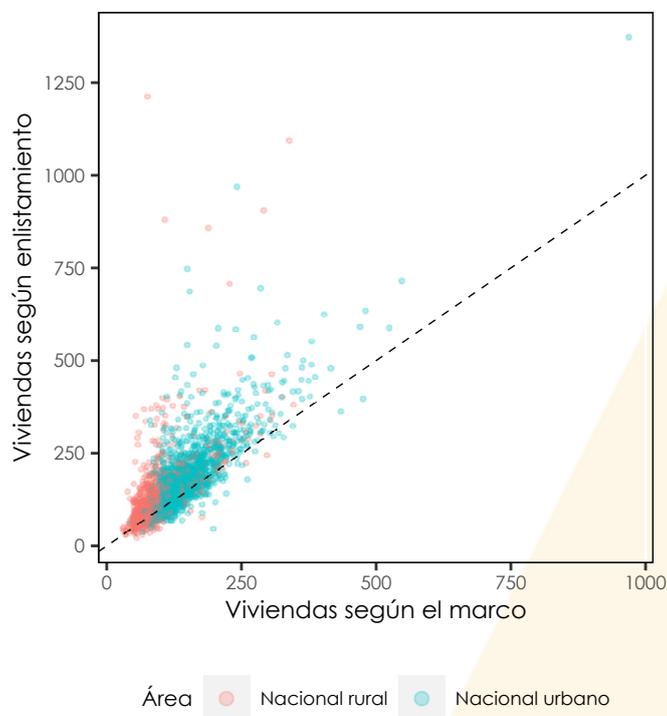
Fuente: Gestión de Diseño Muestral - DINEM.

En total, no se enlistaron 21 UPM que representa el 0,74% de las UPM planificadas, 8 UPM en Tungurahua, 3 UPM en Esmeraldas, 3 provincias con diferencia de 2 UPM y 4 provincias con diferencia de 1 UPM entre lo planificado y enlistado, en

14 provincias se enlistó 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 28,85% (24,31% en nacional urbano y 40,35% en nacional rural). La variación máxima fue del 1.496,05%, mientras que la variación mínima fue del -76,77%. 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 menores a cinco años, sin embargo, no todas las UPM tuvieron al menos 8 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 8 viviendas seleccionadas. En la Tabla 8 se muestra el número de UPM de acuerdo al número de viviendas seleccionadas.

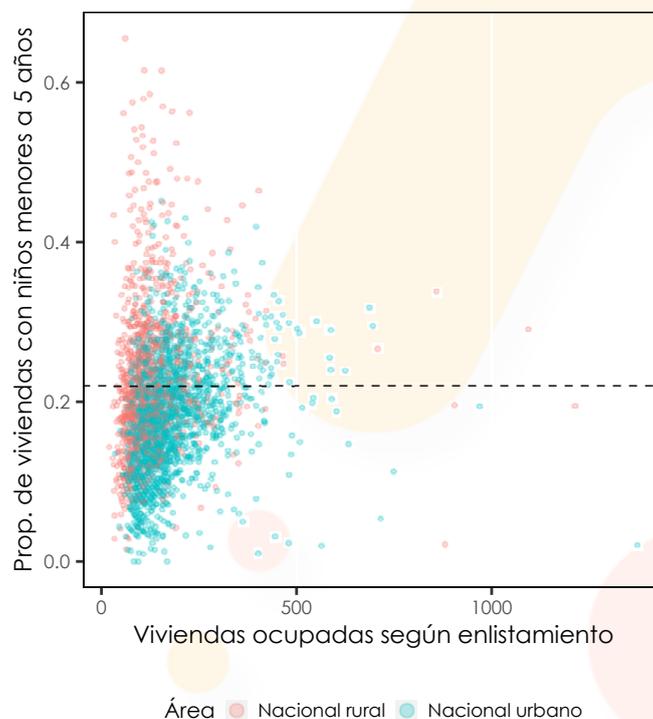
Tabla 8: Número de UPM por número de viviendas seleccionadas.

N° viviendas seleccionadas	Número de UPM		
	Nacional urbano	Nacional rural	Total
1	4	2	6
2	5	4	9
3	9	3	12
4	16	11	27
5	29	10	39
6	23	15	38
7	16	23	39
8	1.570	918	2.488
9	99	75	174
Total	1.771	1.061	2.832

Fuente: Gestión de Diseño Muestral - DINEM.

En la Figura 2 se tiene la proporción de viviendas con niños 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 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 del primer año de la encuesta, además de los cambios de ocupación¹ 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 9: Condición de elegibilidad de las viviendas.

Condición de la vivienda	Nacional urbano	Nacional rural	Nacional	Porcentaje
Efectiva	11.745	7.404	19.149	85,7%
Rechazo	217	35	252	1,1%
Nadie en casa	221	107	328	1,5%
Vivienda temporal	78	137	215	1,0%
Vivienda desocupada	435	106	541	2,4%
Vivienda en construcción	10	6	16	0,1%
Vivienda destruida	4	14	18	0,1%
Negocio	5	1	6	0,0%
Vivienda sin niño	1.161	541	1.702	7,6%
Otra razón	75	32	107	0,5%
Total	13.951	8.383	22.334	100%

Fuente: Gestión de Diseño Muestral - DINEM.

Como se puede observar en la Tabla 9, se investigaron 22.334 viviendas a nivel nacional, siendo el 85,7% efectivas, 13.951 viviendas en el nacional urbano y 8.383 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 menores a cinco años planificado al momento de diseñar la encuesta.

Analizando las otras categorías, se puede observar que el 7,6% de viviendas no contaban con niños menores de 5 años, estas viviendas se categorizan como no

¹Las 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.

elegibles como se explica a continuación.

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 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

Tomando 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.

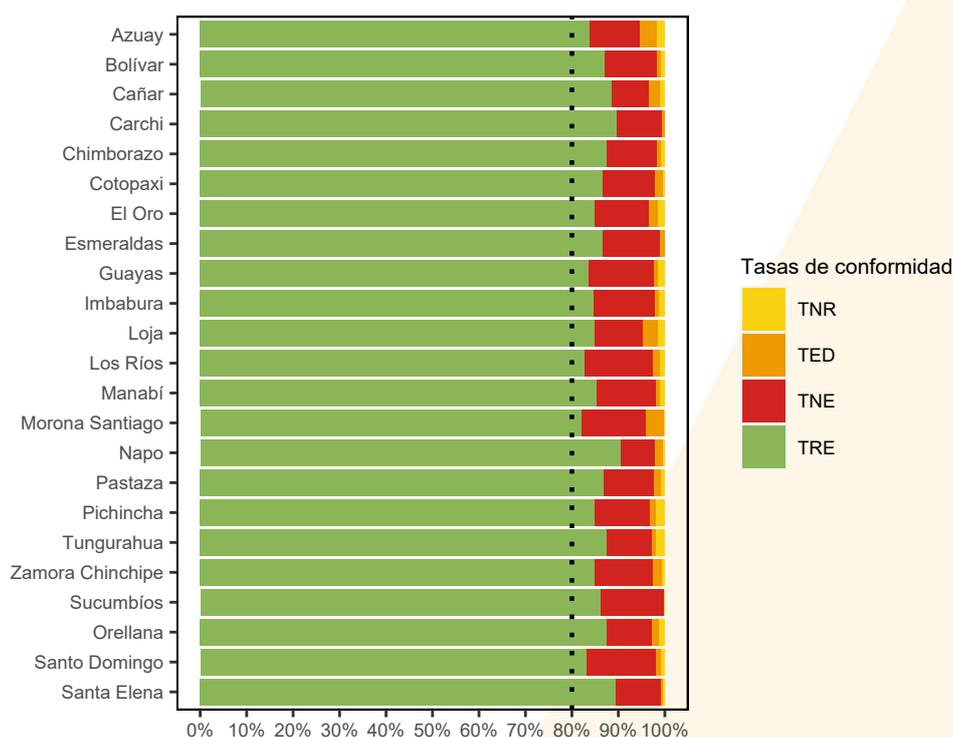
$$T_{NE} = \frac{NE}{RE + NR + NE + ED}$$

- 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.

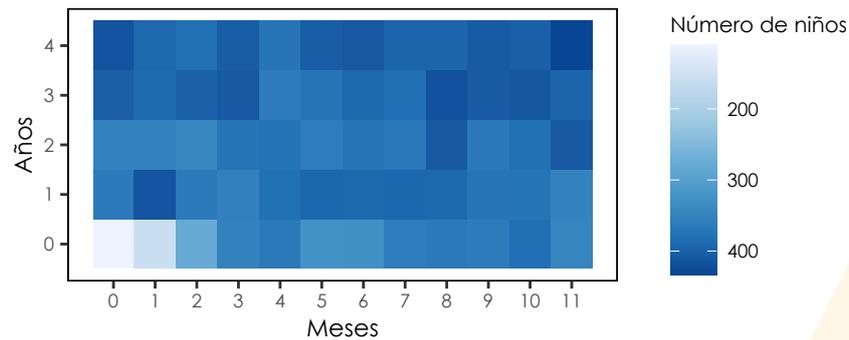
Todas las provincias cuenta con una tasa de respondientes mayor al 80%, tasa de no elegibles desde 7,54% que pertenece a Napo hasta 14,98% en Santo Domingo, además, cuentan con tasa de elegibilidad desconocida entre 0,11% y 3,75% en Sucumbíos y Morona Santiago respectivamente. Y a excepción de la provincia de Carchi, la tasa de no respondientes esta entre 0,09% y 2,06% en Esmeraldas y Tungurahua respectivamente.

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

Población objetivo

En los 12 meses de levantamiento de información, la ENDI entrevistó a 22.260 niños menores a cinco años. En el siguiente gráfico se presenta su distribución por su edad en años y meses:

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



Fuente: Gestión de Diseño Muestral - DINEM.

Puesto que las viviendas con niños menores de cinco años se seleccionan de manera aleatoria, se espera que el total de niños por año cumplido sea uniforme, lo cual se aprecia en la tabla 10. Con respecto al total de niños 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 4 años se encuentra la mayor concentración de niños con un 21,6% del total de niños.

Tabla 10: Resumen de distribución de niños menores a 5 años por año y meses.

Años	Meses con menor concentración	Meses con mayor concentración	Población	Porcentaje
0	0	10	3.739	16,8%
1	11	1	4.535	20,4%
2	2	8 - 11	4.458	20,0%
3	4	8	4.729	21,2%
4	4	11	4.799	21,6%

Fuente: Gestión de Diseño Muestral - DINEM.

Lactancia materna

La sección Lactancia Materna se encuentra en el formulario de Mujeres en Edad Fertil (MEF). En la Tabla 11 se presenta el resultado de entrevista del formulario.

Tabla 11: Resultado de la entrevista formulario MEF.

Resultado de la entrevista	Nacional urbano	Nacional rural	Nacional	Porcentaje
Completa	11.632	7.323	18.955	64,45 %
Rechazo	60	13	73	0,25 %
Ausente	154	115	269	0,91 %
No tiene hijos menores a 5 años	5.824	4.265	10.089	34,30 %
Otra razón	11	13	24	0,08 %
Total	17.681	11.729	29.410	100,00 %

Fuente: Gestión de Diseño Muestral - DINEM.

De las 18.955 MEF que participaron a nivel nacional en el formulario, 11.736 tienen niños menores a 3 años. A continuación, se presenta la distribución de niños menores a tres años por MEF a nivel nacional, nacional urbano y nacional rural.

Tabla 12: Resumen de distribución de niños menores a tres años por MEF.

Distribución	MEF con 1 hijo	MEF con 2 hijos	MEF con 3 hijos	Total MEF	Total hijos
Nacional urbano	6.718	305	5	7.028	7.343
Nacional rural	4.464	239	5	4.708	4.957
Nacional	11.182	544	10	11.736	12.300

Fuente: Gestión de Diseño Muestral - DINEM.

Las 11.736 MEF tienen un total de 12.300 niños menores a tres años. En la sección de lactancia materna participa un niño por cada MEF, es decir 11.736 niños, 7.028 niños están en el nacional urbano y 4.708 niños en el nacional rural como se observa en la Tabla 12.

Desarrollo infantil

En el formulario de desarrollo infantil participa un niño menor a 5 años por hogar, seleccionado mediante el método del cumpleaños más próximo. Dependiendo de la edad del niño seleccionado se aplican diferentes secciones del mismo. En la tabla 13 se presenta la cobertura del formulario de desarrollo infantil a nivel nacional, nacional urbano y nacional rural.

Tabla 13: Resultado de la entrevista del hogar.

Condición por hogar	Nacional urbano	Nacional rural	Nacional	Porcentaje
Efectiva	5.985	3.859	9.844	98,97 %
Rechazo	25	5	30	0,30 %
Ausente	35	33	68	0,68 %
Otra razón	2	2	4	0,04 %
Total	6.047	3.899	9.946	100 %

Fuente: Gestión de Diseño Muestral - DINEM.

Fueron seleccionados 9.946 niños menores a 5 años, siendo el 98,97 % efectivos a nivel nacional, 5.985 en el nacional urbano y 3.859 en rural.

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, siendo su medida de tamaño 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 .

Es importante mencionar que la probabilidad de inclusión se calculó en base al total de UPM planificadas para todo el año de levantamiento.

Probabilidad de segunda etapa

Dado que el enlistamiento permitió identificar las viviendas con niños menores a 5 años, la selección de segunda etapa se realizó únicamente sobre este grupo, siendo:

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

donde:

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

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

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

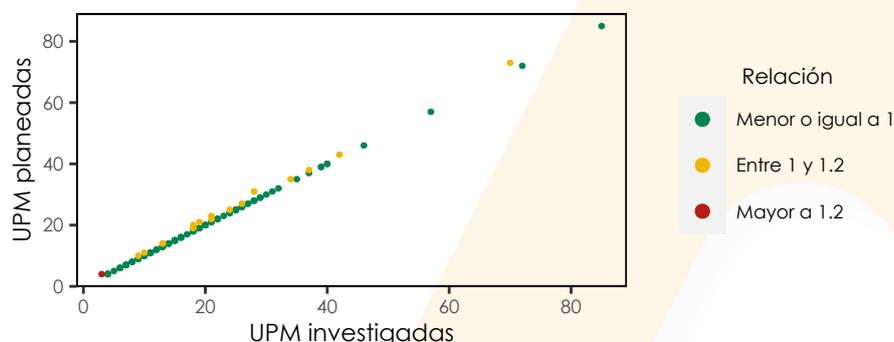
Ajuste de cobertura de primera etapa

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:

$$a_{1_h} := \frac{m_h}{m'_h},$$

donde m'_h es el número de UPM investigadas en el estrato h . En la siguiente figura se aprecia la relación entre la distribución de UPM investigadas y la distribución de UPM planeadas por estrato.

Ilustración 5: Relación de UPM investigadas y planeadas por estrato.



Fuente: Gestión de Diseño Muestral - DINEM.

Esta relación nos indica que los factores de expansión esperados se multiplican por un ajuste menor o igual a 1 en 130 estratos, un ajuste entre 1 y 1,2 en 16 estratos y un estrato con un ajuste superior a 1,2.

Factor de expansión ajustado por cobertura de primera etapa

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

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

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 .

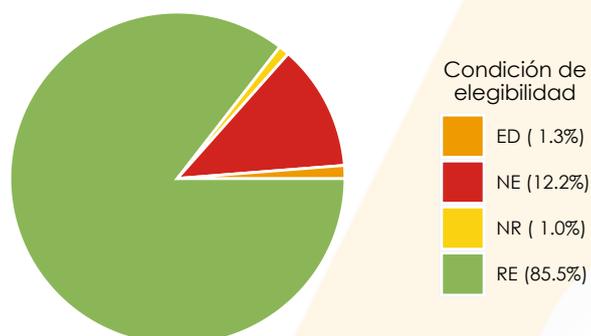
Ajuste de cobertura de segunda etapa

Al momento de la entrevista en campo, cada vivienda muestreada 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, dependiendo de 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 todas las viviendas con niños menores de cinco años por condición de elegibilidad considerando el factor ajustado por cobertura de primera etapa $(d_{1j})^2$, la cual se aprecia en la Figura 6.

Ilustración 6: Distribución de viviendas 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 va cambiando su distribución.

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 en parte porque la falta de respuesta se considera un problema más relevante que debe recibir atención.

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

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, basándonos 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:

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

donde:

a_{2h} = ajuste por elegibilidad desconocida en el estrato h .

s_h = muestra total de viviendas en el estrato h .

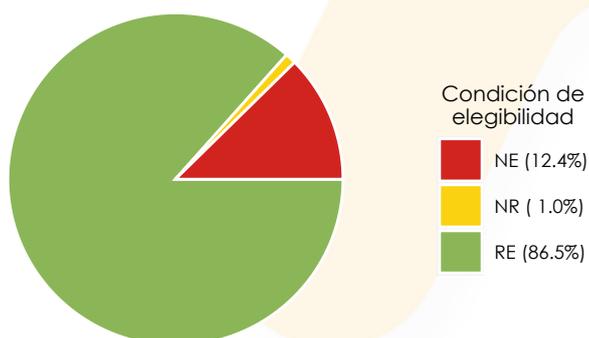
d_{1j} = factor de expansión de la vivienda j .

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

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

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

Ilustración 7: Distribución de viviendas 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, dependiendo de cuánto se conoce de los no respondientes. Para la ENDI, la información disponible de los no respondientes es escasa y limitada, por lo tanto, se utilizará un enfoque similar a la sección anterior.

El ajuste por no respuesta para la vivienda j del estrato h se calcula utilizando los factores de expansión ajustados por elegibilidad desconocida, quedando determinado por:

$$a_{3h} := \frac{\sum_{j \in s_{h,R}} d_{2j}}{\sum_{j \in RE} d_{2j}},$$

donde:

a_{3h} = ajuste por no respuesta en el estrato h .

$s_{h,R} = s_h \cap (RE \cup NR)$.

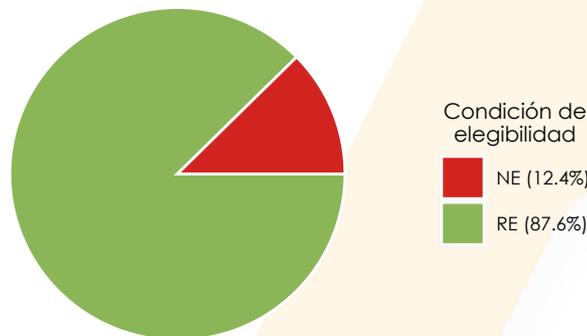
d_{2j} = factor de expansión ajustado por elegibilidad desconocida de la vivienda j .

Por lo tanto, el factor de expansión ajustado por no respuesta (d_{3j}) queda definido de la siguiente manera:

$$d_{3j} := \begin{cases} a_{3h} \cdot d_{2j} & j \in s_h \cap RE, \\ d_{2j} & j \in s_h \cap NE, \\ 0 & j \in s_h \cap NR. \end{cases}$$

El ajuste de no respuesta considera que las viviendas no respondientes tienen características similares a las respondientes. Así, considerando los ajustes por cobertura de vivienda, la distribución final es:

Ilustración 8: Distribución de viviendas por condición de elegibilidad expandidas por d_3 .

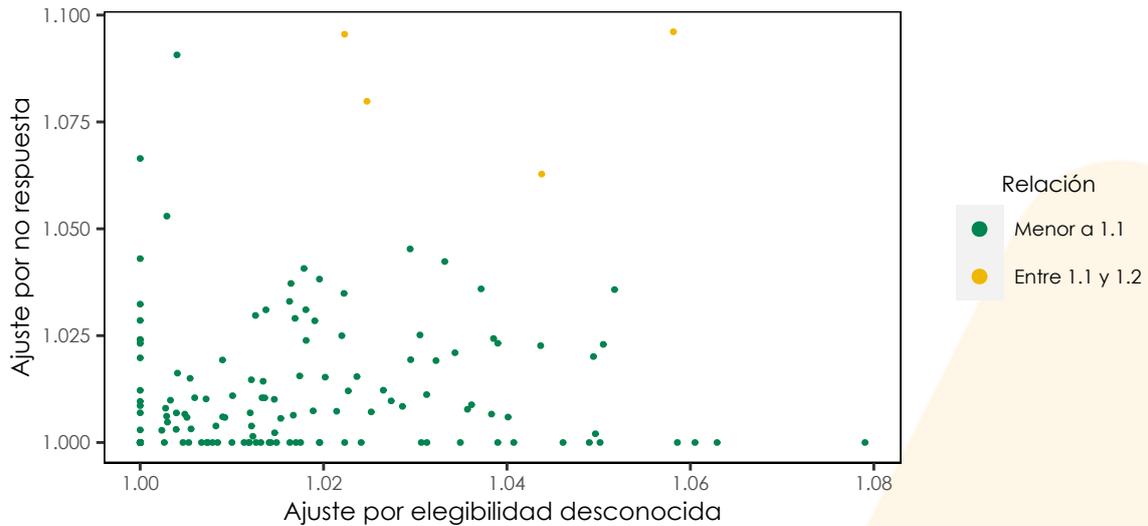


Fuente: Gestión de Diseño Muestral - DINEM.

En conclusión, el estimador de no elegibilidad sobre el marco de muestreo es del 12.4%, lo que significa que aproximadamente una de cada 9 viviendas identificadas como viviendas con niños 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_{3j} y d_{1j} , siendo:

Ilustración 9: Relación entre el factor d_{1j} y d_{3j} en función de los ajustes de cobertura de vivienda.



Fuente: Gestión de Diseño Muestral - DINEM.

Se observa que en 143 estratos tiene un incremento menor a 1,1 y existe una relación entre 1,1 y 1,2 en 4 estratos, lo que implica que en los 147 estratos de diseño, el ajuste total de cobertura de segunda etapa es menor al 20%.

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 dependiendo de 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 considerando 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 multiplicando el 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 usando fuentes de información auxiliares tal y como se describe en la siguiente sección.

El factor de normalización se define como:

$$f_n := \frac{\sum_{j \in s} n_j}{\sum_{j \in s} d_{3j} \cdot n_j},$$

donde:

f_n = factor de normalización.

n_j = número de personas en la vivienda j .

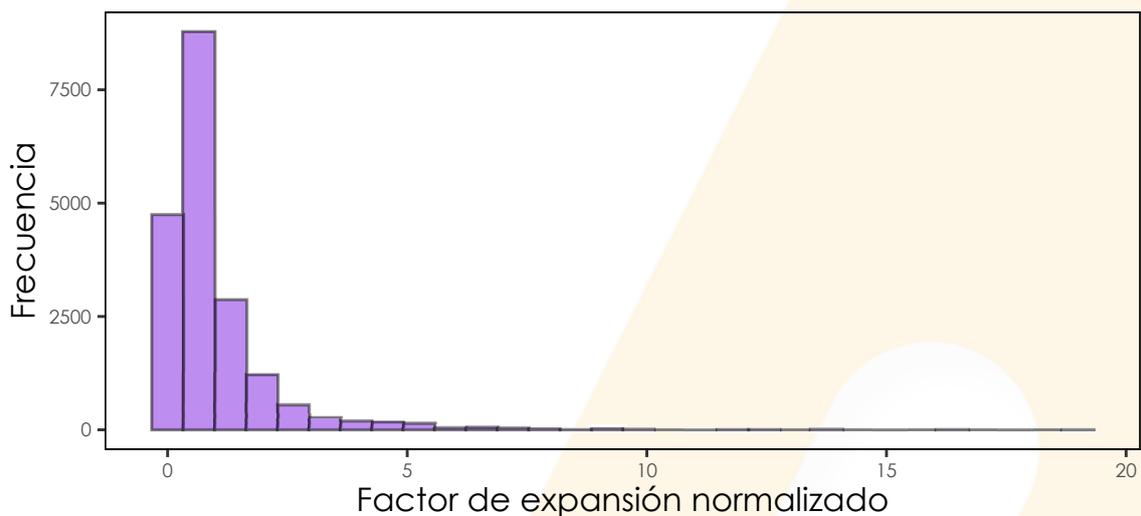
d_{3j} = factor de expansión ajustado por no respuesta de la vivienda j .

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

$$d_{n_j} := f_n \cdot d_{3j}.$$

En la Figura 10 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 10: 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 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. Por lo tanto, la probabilidad de selección de tercera etapa es:

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

donde L_k es el número total de niños menores a tres años de la MEF k encuestada en la vivienda j . Por lo tanto, el factor de expansión queda definido por:

$$d_{4j} := d_{3j} \cdot \frac{1}{pte_{j,k}},$$

Siguiendo un proceso equivalente al mostrado en la sección anterior, el factor de expansión de lactancia materna normalizado $dln_{n_{j,k}}$ es:

$$dln_{n_{j,k}} = fn_{lm} \cdot d_{4j}.$$

donde fn_{lm} es el factor de normalización de lactancia materna.

Factor de expansión desarrollo infantil

Para la sección de Desarrollo Infantil, se selecciona un niño menor a 5 años por cada hogar mediante el método de cumpleaños más próximo. 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:

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

donde D_k es el número total de niños menores a 5 años en el hogar k encuestado en la vivienda j . Por lo tanto, el factor de expansión queda definido por:

$$d_{5j} := d_{3j}^* \cdot \frac{1}{pte_{j,k}},^3$$

Siguiendo un proceso equivalente al mostrado en la sección anterior, el factor de expansión de desarrollo infantil normalizado $ddein_{n_{j,k}}$ es:

$$ddein_{n_{j,k}} = fn_{df} \cdot d_{5j}.$$

donde fn_{df} es el factor de normalización de desarrollo infantil.

³Nota: d_{3j}^* hace referencia al factor de expansión ajustado por no respuesta considerando únicamente las viviendas investigadas en el segundo semestre de levantamiento

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 determinada⁴. Está dado por (Särndal et al., 1992):

$$\hat{Y}_{m_{HT}} = \sum_i \sum_j d_{n_k} \cdot y_k,$$

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 .

Considerando la definición del factor de normalización utilizado en la ENDI, se obtiene que $\hat{\bar{Y}}_m = \hat{Y}_{m_{HT}}/n_i$, obteniendo así estimadores de tipo media.

Error de muestreo

Luego de realizar la estimación respectiva para la variable de interés a nivel de dominio de estudio, el error de muestreo podrá ser calculado a partir de la estimación de la varianza del estimador del total $\hat{Y}_{m_{HT}}$. Para calcularlos adecuadamente, se debe tomar en cuenta los diferentes aspectos del diseño muestral, es decir, las etapas de muestreo, la estratificación presente en los dominios de estudio y los procesos de selección en cada una de las etapas.

Con todos estos elementos, el coeficiente de variación para el estimador $\hat{Y}_{k_{HT}}$ viene dado por la siguiente expresión:

$$CV(\hat{Y}_{k_{HT}}) = \frac{\sqrt{\hat{V}_{2st}(\hat{Y}_{k_{HT}})}}{\hat{Y}_{k_{HT}}}.$$

Un estimador insesgado para la varianza está dado por:

$$\hat{V}_{2st}(\hat{t}_\pi) = \sum_{s_I} \sum_{I_{ij}} \check{\Delta}_{I_{ij}} \frac{\hat{t}_{i\pi}}{\pi_{I_i}} \frac{\hat{t}_{i\pi}}{\pi_{I_j}} + \sum_{s_I} \frac{\hat{V}_i}{\pi_{I_i}},$$

en el cual, el \hat{V}_i apropiado es:

$$\hat{V}_i = \sum_{s_{II_i}} \sum_{II_{qr|i}} \check{\Delta}_{II_{qr|i}} \frac{\hat{t}_{iq\pi}}{\pi_{II_{q|i}}} \frac{\hat{t}_{ir\pi}}{\pi_{II_{r|i}}},$$

donde:

⁴Nótese que, al utilizar factores normalizados, el estimador de total calculado aproxima el total muestral, más no el total poblacional.

- i, j = índices que recorren las UPM en el dominio m .
- q, r = índices que recorren las viviendas en la UPM i del dominio m .
- π_{Ii} = Probabilidad de selección de la Etapa I para la UPM i del dominio m .
- $\pi_{IIq|i}$ = Probabilidad de selección de la Etapa II para la vivienda q de n la UPM i .

Por último, al considerar la relación $\hat{Y}_m = \hat{Y}_{mHT}/n$ y las propiedades de la varianza, se obtiene que:

$$CV(\hat{Y}_k) = \frac{\sqrt{\hat{V}_{2st}(\hat{Y}_k)}}{\hat{Y}_k}.$$

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, siendo los más utilizados 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 14, donde se describe las etiquetas de las variables identificadoras de las UPM, estratos y factor de expansión.

Tabla 14: 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 usando 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 apoyándonos en información auxiliar disponible para la población objetivo de los mismos⁵. Por ejemplo:

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

se tiene que:

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

⁵El total auxiliar puede ser obtenido de diferentes fuentes de información incluyendo: censo de población, proyecciones poblacionales y registros administrativos



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

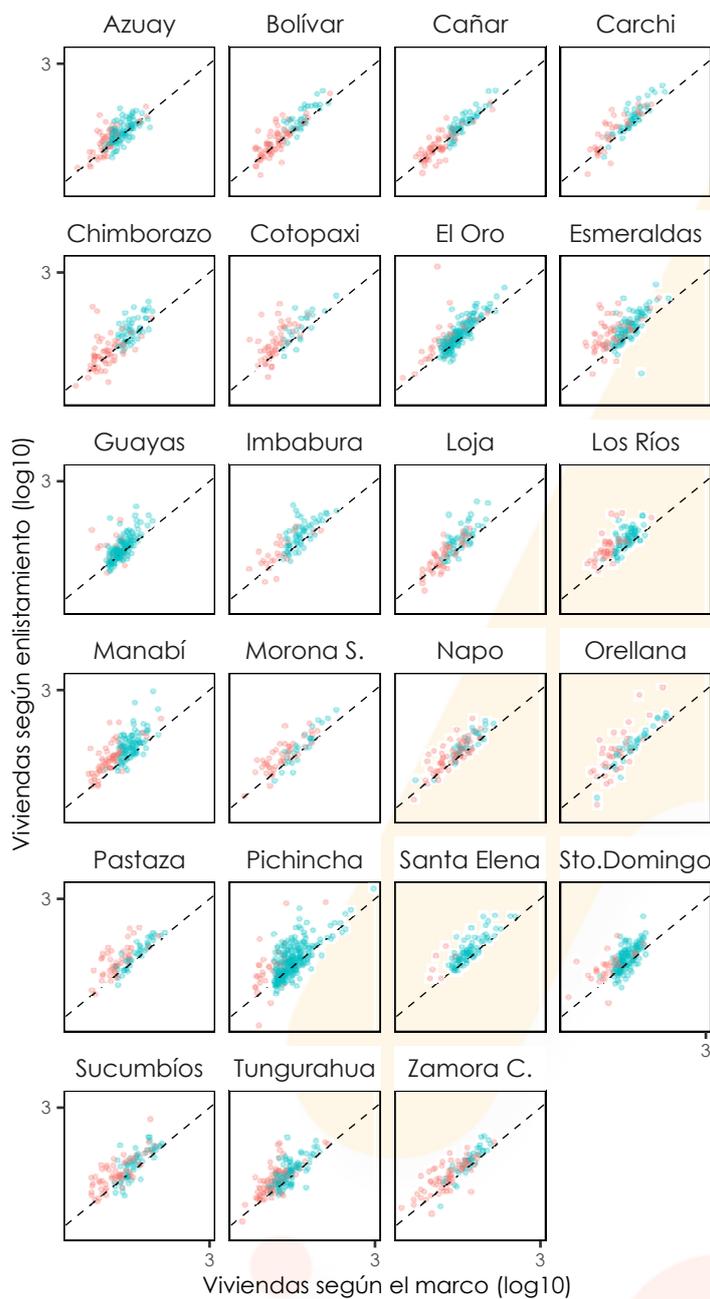
Referencias

- (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.
- (INEC, 2022) INEC (2022). Geografía estadística. https://www.ecuadorencifras.gob.ec/documentos/web-inec/Geografia_Estadistica/Micrositio_geoportal/index.html#clasificador-geog-dpa. Recuperado el 2023-02-02.
- (Instituto Nacional de Estadística y Censos, 2019) Instituto Nacional de Estadística y Censos (2019). *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

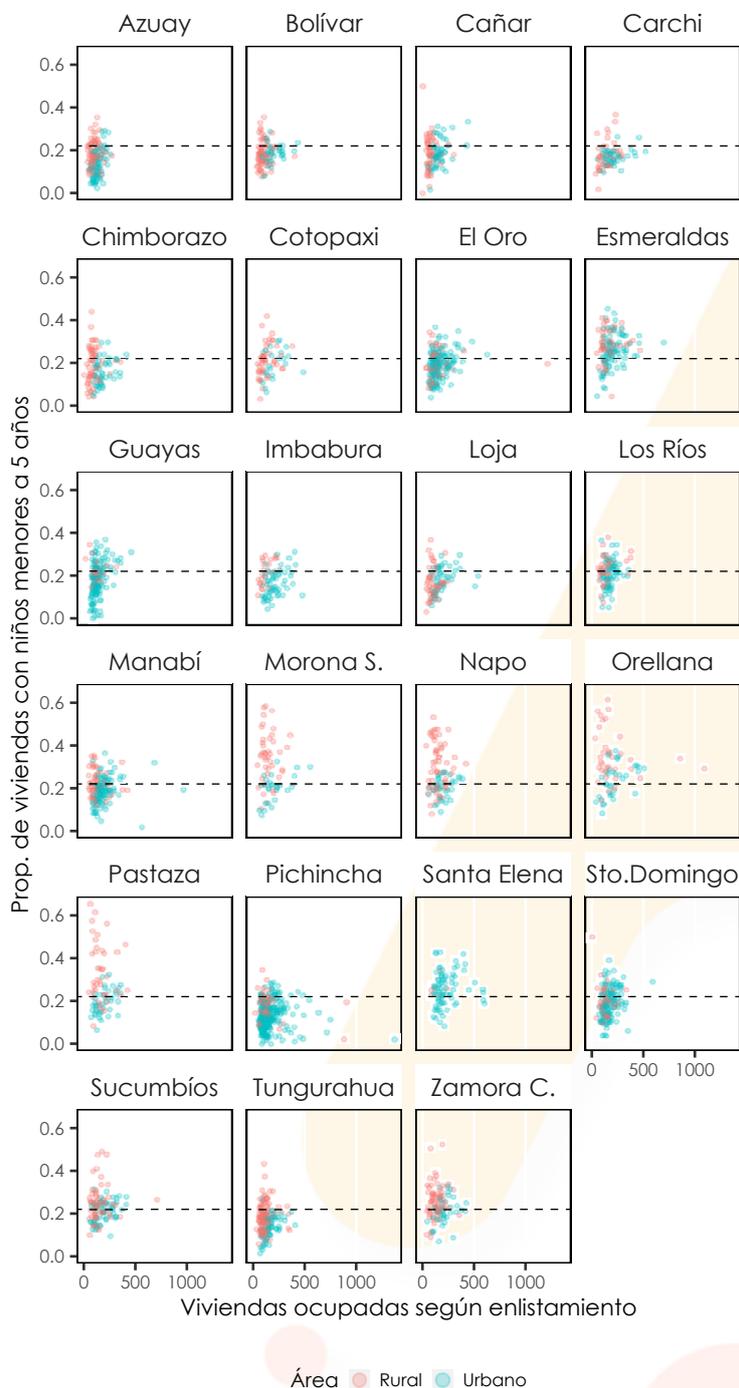
Ilustración 11: Viviendas ocupadas según marco y enlistamiento.



Fuente: Gestión de Diseño Muestral - DINEM.

Anexo 2

Ilustración 12: Proporción de viviendas con niños menores a 5 años por UPM.



Fuente: Gestión de Diseño Muestral - DINEM.

Anexo 3

Tabla 15: Condición de elegibilidad y tasa de conformidad (%) por provincia.

Provincia	NR	ED	NE	RE	Total	TNR	TED	TNE	TRE
Azuay	20	41	121	937	1.119	1,8	3,7	10,8	83,7
Bolívar	7	9	95	744	855	0,8	1,1	11,1	87,0
Cañar	9	20	68	742	839	1,1	2,4	8,1	88,4
Carchi	0	5	63	592	660	0,0	0,8	9,5	89,7
Cotopaxi	2	11	63	494	570	0,4	1,9	11,1	86,7
Chimborazo	8	7	89	727	831	1,0	0,8	10,7	87,5
El Oro	23	31	177	1.298	1.529	1,5	2	11,6	84,9
Esmeraldas	1	11	139	975	1.126	0,1	1	12,3	86,6
Guayas	19	11	164	986	1.180	1,6	0,9	13,9	83,6
Imbabura	9	5	89	566	669	1,3	0,7	13,3	84,6
Loja	12	25	77	638	752	1,6	3,3	10,2	84,8
Los Ríos	10	14	125	713	862	1,2	1,6	14,5	82,7
Manabí	13	12	159	1.064	1.248	1	1	12,7	85,3
Morona Santiago	2	21	78	459	560	0,4	3,8	13,9	82
Napo	4	12	58	695	769	0,5	1,6	7,5	90,4
Pastaza	6	10	72	577	665	0,9	1,5	10,8	86,8
Pichincha	48	28	282	2.013	2.371	2	1,2	11,9	84,9
Tungurahua	28	10	133	1.185	1.356	2,1	0,7	9,8	87,4
Zamora Chinchipe	5	18	107	730	860	0,6	2,1	12,4	84,9
Sucumbíos	2	1	129	814	946	0,2	0,1	13,6	86
Orellana	8	9	54	496	567	1,4	1,6	9,5	87,5
Santo Domingo	13	13	199	1.103	1.328	1	1	15	83,1
Santa Elena	3	4	64	601	672	0,4	0,6	9,5	89,4
Nacional	252	328	2.605	19.149	22.334	1,1	1,5	11,7	85,7

Fuente: Gestión de Diseño Muestral - DINEM.

**ENCUESTA
NACIONAL SOBRE
DESNUTRICIÓN
*Infantil***

INEC

Buenas cifras,
mejores vidas



@Ecuadorencifras



@Ecuadorencifras



@InecEcuador



INECEcuador