

Cálculo del factor de expansión con base a la ENDI 2022 - 2023

ENSANUT 2018

Septiembre, 2023





Documento del Cálculo del factor de expansión con base a la ENDI 2022 - 2023 de la ENSANUT 2018

Quito - Ecuador, 2023

Dirección/Departamento

Dirección de Infraestructura Estadística y Muestreo

Unidad

Gestión de Diseño Muestral

Elaborado por:

Javier Nuñez
Ángel Gaibor
Giovanella Peña

Revisado por:

Christian Garcés

Aprobado por:

Julio Muñoz



Contenido

Introducción	1
Marco muestral	3
Cambio en los dominios geográficos	3
Unidades Primarias de Muestreo	4
Estratificación	5
Tamaño Muestral	7
Cobertura	9
Condición de ocupación y resultado de entrevista	9
Tasas de conformidad	11
Población objetivo	12
Factores de expansión	14
Probabilidad de primera etapa	14
Probabilidad de segunda etapa	14
Factor de expansión teórico	14
Ajuste de cobertura de primera etapa	15
Ajuste de cobertura de segunda etapa	15
Factor de expansión normalizado	18
Factor de expansión Desarrollo Infantil	20
Estimaciones de características	21
Error de muestreo	21
Métodos de estimaciones de errores para diseños muestrales complejos	22
Estimación de totales usando fuentes de información auxiliares	23

Lista de Tablas

1. Diferencias ENDI - ENSANUT.	2
2. Distribución de viviendas de acuerdo al proceso de actualización cartográfica.	3
3. Distribución de UPM por provincia y área.	4
4. Muestra de UPM y viviendas sobre L_C de la ENSANUT.	7
5. Condición de elegibilidad por área.	9
6. Condición de elegibilidad reclasificada por área.	10

7.	Condición de elegibilidad por área.	11
8.	Resumen de distribución de niños menores a cinco años por año y meses.	13
9.	Variables requeridas para la declaración del diseño muestral.	23

Lista de Figuras

1.	Tasa de conformidad por área.	12
2.	Distribución de niños menores de cinco años por años y meses.	13
3.	Distribución de viviendas por condición de elegibilidad expandidas por d_1	16
4.	Distribución de viviendas por condición de elegibilidad expandidas por d_2	17
5.	Distribución de viviendas por condición de elegibilidad expandidas por d_3	18
6.	Relación entre el factor d_{1j} y d_{3j} en función de los ajustes de cobertura de vivienda.	18
7.	Distribución de los factores de expansión normalizados.	20

Introducción

Desde el 2018, el Instituto Nacional de Estadística y Censo ha ejecutado dos encuestas orientadas a la medición de la Desnutrición Crónica Infantil (DCI): la Encuesta Nacional de Salud y Nutrición 2018 (ENSANUT 2018) y la Encuesta Nacional sobre Desnutrición Infantil 2022-2023 (ENDI). Sin embargo, al considerar sus objetivos específicos estas operaciones estadísticas presentan diferencias de diseño, metodológicas, conceptuales, entre otras.

Desde la perspectiva de diseño muestral, puesto que la muestra seleccionada para la ENSANUT 2018 consideró viviendas con y sin infantes menores a 5 años, es posible replicar el proceso de cálculo de la cobertura y factores de expansión de la ENDI.

El presente documento tiene como finalidad describir el cálculo de factores de expansión para la ENSANUT 2018 siguiendo la metodología utilizada para la ENDI, excluyendo la provincia de Galápagos puesto que el diseño muestral de ENDI en esta provincia es diferente. Para esto se realiza una descripción y comparación de sus poblaciones objetivo y marco muestral. Luego, se describe el tamaño muestral de la ENSANUT 2018 considerando únicamente a las viviendas seleccionadas con infantes menores a 5 años. Por último, se presenta la cobertura y factores de expansión calculados para la ENSANUT 2018.

Encuesta Nacional de Salud y Nutrición 2018

La ENSANUT 2018 tuvo como objetivo principal generar indicadores sobre los principales problemas y la situación de salud de la población ecuatoriana con el fin de evaluar y generar las políticas públicas en temas de salud y nutrición.

Las poblaciones objetivo de la ENSANUT 2018 son:

- Mujeres en edad fértil - MEF (10-49 años),
- Niños menores a 5 años de edad,
- Niños entre 5 y 11 años de edad,
- Personas de 5 a 17 años,
- Personas mayores a 10 años,
- Hombres mayores a 12 años,

El diseño muestral que sigue la ENSANUT 2018 es bietápico estratificado de elementos teniendo como dominios de diseño las 24 provincias del país. En una primera etapa se seleccionaron Unidades Primarias de Muestreo (UPM), en las mismas se realizó un proceso de enlistamiento de las personas que habitaban en las viviendas ocupadas de las UPM. Posterior a esto se clasificó y seleccionó las viviendas dentro de los siguientes dos grupos:

- L_C : Viviendas con niños menores a cinco años.
- L_S : Viviendas sin niños menores a cinco años.

El proceso de enlistamiento se llevó a cabo entre el 3 de septiembre y el 31 de octubre de 2018, mientras que el levantamiento de información se ejecutó en dos

periodos el primero entre noviembre y diciembre de 2018 y el segundo entre junio y julio de 2019.

Encuesta Nacional sobre Desnutrición Infantil 2022 - 2023

La 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.

El diseño muestral que sigue la ENDI es bietápico estratificado de elementos, cuyos dominios de diseño son las 24 provincias del país. En una primera etapa se seleccionaron Unidades Primarias de Muestreo (UPM), en las mismas se realizó un proceso de enlistamiento con la finalidad de clasificar a las viviendas en los grupos L_C y L_S , realizando la selección de viviendas únicamente sobre el grupo L_C .

La muestra de la ENDI se distribuyó temporal y espacialmente, estando presente durante 12 meses consecutivos (desde el 12 de julio de 2022 hasta el 10 de julio de 2023) y todos los estratos del marco de muestreo, respectivamente. El enlistamiento se llevó a cabo 2 meses antes del levantamiento de cada uno de los periodos empezando el 23 de abril de 2022 y terminando el 14 de abril de 2023.

Comparación entre ENSANUT y ENDI

En la Tabla 1 se presentan las principales diferencias entre ENDI y ENSANUT.

Tabla 1: Diferencias ENDI - ENSANUT.

ENDI	ENSANUT
Muestra distribuida en 12 meses de levantamiento	Dos periodos de levantamiento de dos meses cada uno
Población objetivo niños menores a 5 años	Múltiples poblaciones objetivo entre ellas niños menores a 5 años
Dos meses de diferencia entre enlistamiento y encuesta	De dos hasta ocho meses de diferencia entre enlistamiento y encuesta
Selección de viviendas únicamente con niños menores a 5 años	Selección de viviendas con y sin niños menores a 5 años

Fuente: Gestión de Diseño Muestral - DINEM.

A pesar de las diferencias entre la ENDI y la ENSANUT, se puede replicar el cálculo de cobertura y factores de expansión de la ENDI sobre la ENSANUT considerando únicamente las viviendas que se seleccionaron sobre el conjunto L_C .

Marco muestral

El Marco de muestreo utilizado para la ENSANUT 2018 como 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 2 se presenta la proporción de viviendas catalogadas como ocupadas según el precenso CPV-2010 y los diferentes procesos de actualización.

Tabla 2: 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.

Cambio en los dominios geográficos

El MMM debe tener la capacidad de acoplarse a los cambios estructurales en el Ecuador (por ejemplo, creación de nuevas provincias) a lo largo del tiempo, manteniendo un equilibrio entre esta dinámica y la estabilidad necesaria para la producción de estadísticas subnacionales.

En el año 2020, en el MMM se implementaron los cambios sobre los límites provinciales realizados por la Secretaría Técnica del Comité Nacional de Límites Internos (CONALI) hasta diciembre de 2018, los cuales son:

- cantón La Concordia ,
- parroquia Las Golondrinas (Zona no Delimitada),
- parroquia Manga del Cura (Zona no Delimitada), y
- parroquia El Piedrero (Zona no Delimitada).

Debido a esta actualización, la ENDI no contempla muestra sobre las Zonas no Delimitadas puesto que ya fueron redistribuidas a sus respectivas provincias en el MMM previo a la selección de su muestra. Sin embargo, dado el bajo porcentaje que estas representan (0,5%), el impacto en la comparabilidad es desestimable.

Unidades Primarias de Muestreo

Cada vivienda del MMM está asociada de forma unívoca a una manzana o sector disperso. Tanto en la parte amanzana 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, permitiéndonos identificar adecuadamente los dominios de diseño geográficos definidos para cada encuesta. Además, contiene variables que clasifican las UPM en función al estrato al que pertenecen. Para más información referirse a (INEC, 2022). En la Tabla 3 se presenta el total de UPM por provincia y área presentes en el MMM actual.

Tabla 3: 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

Provincia	Rural	Urbano	Total
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 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 subdivisión 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.

Tamaño Muestral

El diseño muestral que siguió la ENSANUT 2018 fue un muestreo bietápico estratificado de elementos, una vez seleccionadas las UPM se siguió un proceso de enlistamiento para identificar las viviendas con niños menores a cinco años de las que no, definiendo los siguientes conjuntos:

- L_C : Viviendas con niños menores a cinco años.
- L_S : Viviendas sin niños menores a cinco años.

En la segunda etapa de selección, se escogieron viviendas tanto sobre L_C y L_S . El número de viviendas a seleccionar sobre cada conjunto varió en función a cada provincia y sus características demográficas. Para mayores detalles acerca del diseño muestral de ENSANUT 2018 revisar el documento (INEC, 2019).

Para el cálculo de los factores de expansión considerando la metodología aplicada en la ENDI, se consideran únicamente las viviendas seleccionadas sobre L_C . En la Tabla 4 se detalla el número de UPM seleccionadas para cada provincia y el número de viviendas con niños menores a cinco años.

Tabla 4: Muestra de UPM y viviendas sobre L_C de la ENSANUT.

Provincia	Número de UPM	Número de viviendas
Azuay	78	642
Bolívar	83	526
Cañar	71	642
Carchi	78	649
Cotopaxi	102	709
Chimborazo	114	713
El Oro	192	1.784
Esmeraldas	108	1.093
Guayas	210	1.449
Imbabura	77	636
Loja	83	771
Los Ríos	144	1.409
Manabí	133	1.284
Morona Santiago	84	639
Napo	126	748
Pastaza	96	632
Pichincha	102	908
Tungurahua	77	653
Zamora Chinchipe	84	711
Galápagos	111	401
Sucumbíos	96	889



Provincia	Número de UPM	Número de viviendas
Orellana	89	863
Santo Domingo	156	1.530
Santa Elena	78	527
Zonas no Delimitadas	12	109
Total	2.584	20.917

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, 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 este ejercicio.

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 5: Condición de elegibilidad por área.

Condición de la vivienda	Nacional urbano	Nacional rural	Nacional	Porcentaje
Efectiva	12.016	7.259	19.275	94,0%
Rechazo	182	18	200	1,0%
Nadie en casa	91	35	126	0,6%
Vivienda temporal	82	155	237	1,2%
Vivienda desocupada	311	140	451	2,2%
Vivienda en construcción	18	14	32	0,2%
Vivienda destruida	10	17	27	0,1%
Negocio	18	8	26	0,1%
Otra razón	78	64	142	0,7%
Total	12.806	7.710	20.516	100,0%

Fuente: Gestión de Diseño Muestral - DINEM.

Sin embargo, puesto que en la ENSANUT 2018 se investigaron varias poblaciones objetivo, sobre L_C existen viviendas efectivas sin infantes menores a 5 años. Para equiparar el cálculo de cobertura entre esta y la ENDI, estas viviendas se reclasifican a *vivienda efectiva sin niños menores a 5 años*. En la tabla 6 se muestra la condición de elegibilidad por área luego de realizar la reclasificación.

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.

Tabla 6: Condición de elegibilidad reclasificada por área.

Condición de la vivienda	Nacional urbano	Nacional rural	Nacional	Porcentaje
Efectiva	10.189	6.409	16.598	80,9%
Rechazo	182	18	200	1,0%
Nadie en casa	91	35	126	0,6%
Vivienda temporal	82	155	237	1,2%
Vivienda desocupada	311	140	451	2,2%
Vivienda en construcción	18	14	32	0,2%
Vivienda destruida	10	17	27	0,1%
Negocio	18	8	26	0,1%
Vivienda sin niño	1.827	850	2.677	13,0%
Otra razón	78	64	142	0,7%
Total	12.806	7.710	20.516	100,0%

Fuente: Gestión de Diseño Muestral - DINEM.

Para calcular la cobertura por dominio de diseño a nivel de vivienda y población objetivo se 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
 - Otra razón, cuál?
 - Efectiva sin niños

2. Elegibilidad desconocida

- Elegibilidad desconocida (ED)
 - Nadie en casa

A continuación, se presenta el resultado de la condición de elegibilidad de las viviendas investigadas:

Tabla 7: Condición de elegibilidad por área.

Condición de elegibilidad	Nacional urbano	Nacional rural	Nacional	Porcentaje
Elegible respondiente (RE)	10.189	6.409	16.598	80,9%
Elegible no respondiente (NR)	182	18	200	1,0%
No elegibles (NE)	2.344	1.248	3.592	17,5%
Elegibilidad desconocida (ED)	91	35	126	0,6%
Total	12.806	7.710	20.516	100,0%

Fuente: Gestión de Diseño Muestral - DINEM.

Del total de viviendas investigadas, 3.592 viviendas fueron no elegibles a nivel nacional (2.344 viviendas en nacional urbano y 1.248 en nacional rural), de las cuales la mayoría pertenecen a la categoría *viviendas sin niños menores de 5 años* como se observa en la Tabla 6. Además, existe una mínima cantidad de viviendas por área clasificadas como que son *elegibles no respondiente* y *elegibilidad desconocida*, como se aprecia en la Tabla 7.

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.

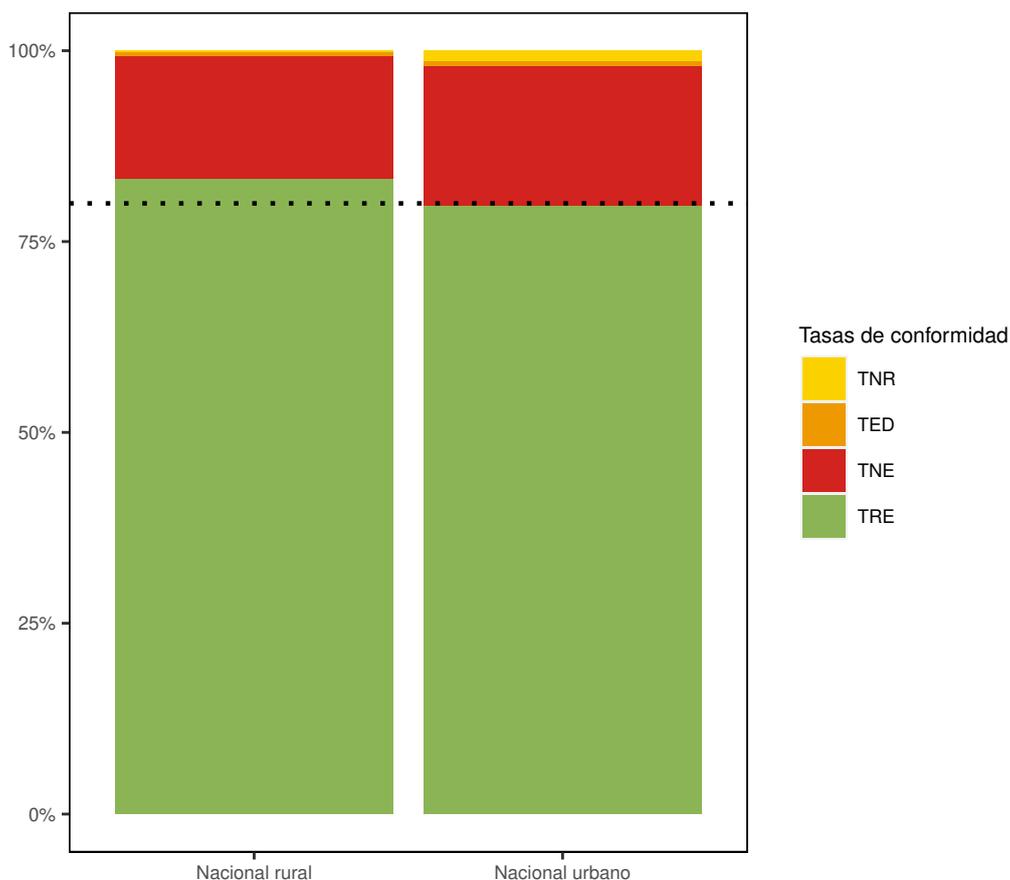
$$T_{NE} = \frac{NE}{RE + NR + NE + ED}$$

- 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 1: Tasa de conformidad por área.



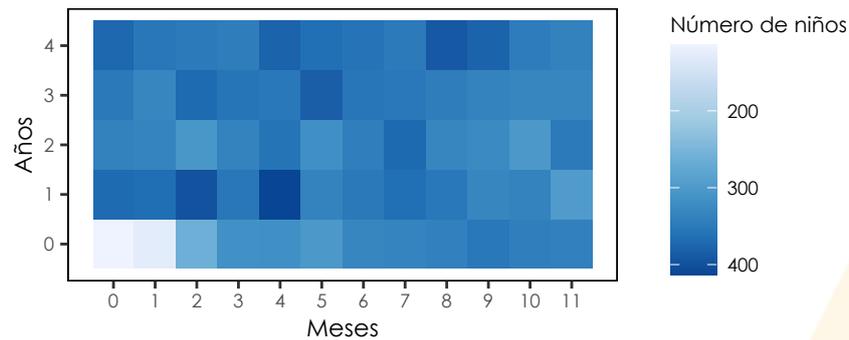
Fuente: Gestión de Diseño Muestral - DINEM.

La tasa de respondientes es del 79,6 % para nacional urbano y 83,1 % para nacional rural, mientras que la tasa de no elegibles es del 18,3 % y del 16,2 %, respectivamente. Por último, las demás tasas se encuentra por debajo del 2,0 %.

Población objetivo

Considerando únicamente las viviendas seleccionadas sobre L_C , la ENSANUT 2018 entrevistó a 20.301 niños menores a 5 años. En el siguiente gráfico se presenta su distribución por su edad en años y meses:

Ilustración 2: Distribución de niños menores de cinco años por años y meses.



Fuente: Gestión de Diseño Muestral - DINEM.

Puesto que las viviendas con niños menores de 5 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 8. Con respecto al total de infantes con 0 años de edad, su diferencia se debe al desfase temporal entre el enlistamiento y el levantamiento de la información.

Tabla 8: Resumen de distribución de niños menores a cinco años por año y meses.

Años	Meses con menor concentración	Meses con mayor concentración	Población	Porcentaje
0	0	9	3.432	17,2%
1	11	4	4.168	21,0%
2	2, 10	7	3.950	19,8%
3	1, 11	5	4.122	20,7%
4	11	8	4.250	21,3%

Fuente: Gestión de Diseño Muestral - DINEM.

Factores de expansión

El diseño muestral de la ENSANUT es bietápico estratificado de elementos, 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, los cuales se desarrollaron tomando como referencia lo expuesto en (Valliant et al., 2013).

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 .

Probabilidad de segunda etapa

Dado que el enlistamiento permitió identificar las viviendas sobre L_C y L_S , la selección de segunda etapa se realizó únicamente sobre L_C , 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 $d_{0i,h}$ 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:

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

donde m'_h es el número de UPM investigadas en el estrato h .

Factor de expansión ajustado por cobertura de primera etapa

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

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

donde $d_{1i,h}$ 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

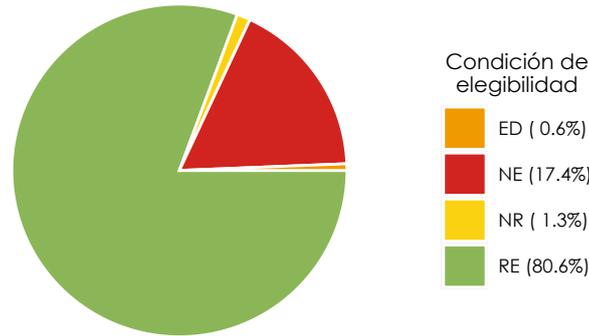
Al momento de la entrevista en campo, cada vivienda muestreada se clasificó en uno de los cuatro conjuntos anteriormente descritos (*RE*, *NR*, *NE* y *ED*). Si existen unidades que son no elegibles en la muestra, esto es evidencia de que hay unidades no elegibles tanto sobre *ED* como sobre lo no muestreado. 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 5 años por condición de elegibilidad considerando el factor ajustado por cobertura de primera etapa $(d_{1j})^2$, la cual se aprecia en la Figura 3.

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

Ilustración 3: Distribución de viviendas por condición de elegibilidad expandidas por d_1 .



Fuente: Gestión de Diseño Muestral - DINEM.

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.

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 este ejercicio, 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 .

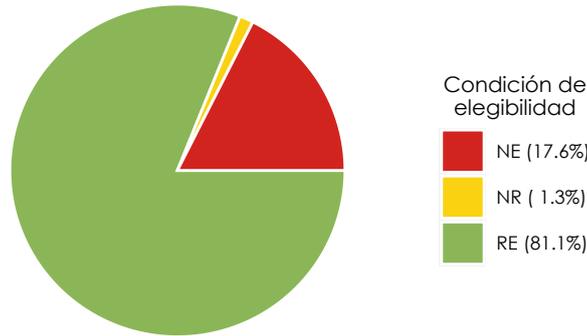
d_{1j} = factor de expansión de la vivienda j .

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 4: 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 este ejercicio, 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 .

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

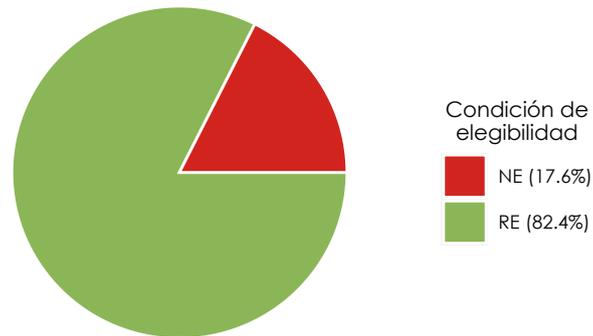
d_{2j} = factor de expansión ajustado por elegibilidad desconocida de la vivienda j .

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:

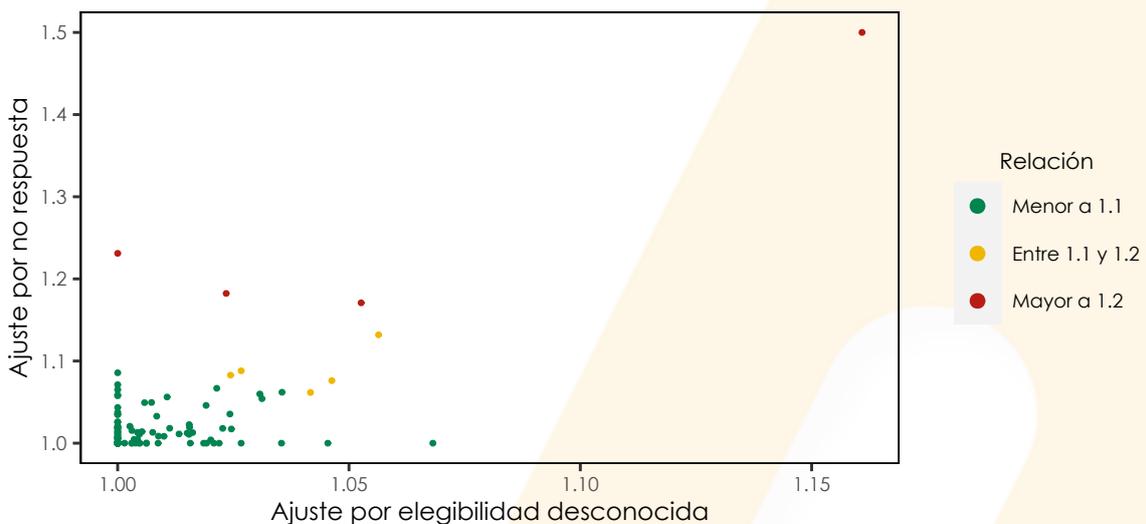
Ilustración 5: Distribución de viviendas por condición de elegibilidad expandidas por d_3 .



Fuente: Gestión de Diseño Muestral - DINEM.

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 6: 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 puede observar que en 153 estratos se tuvo un incremento menor a 1,1, mientras que existe una relación entre 1,1 y 1,2 en 5 estratos y mayor a 1,2 en 4 estratos de diseño.

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 ENSANUT y 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 .

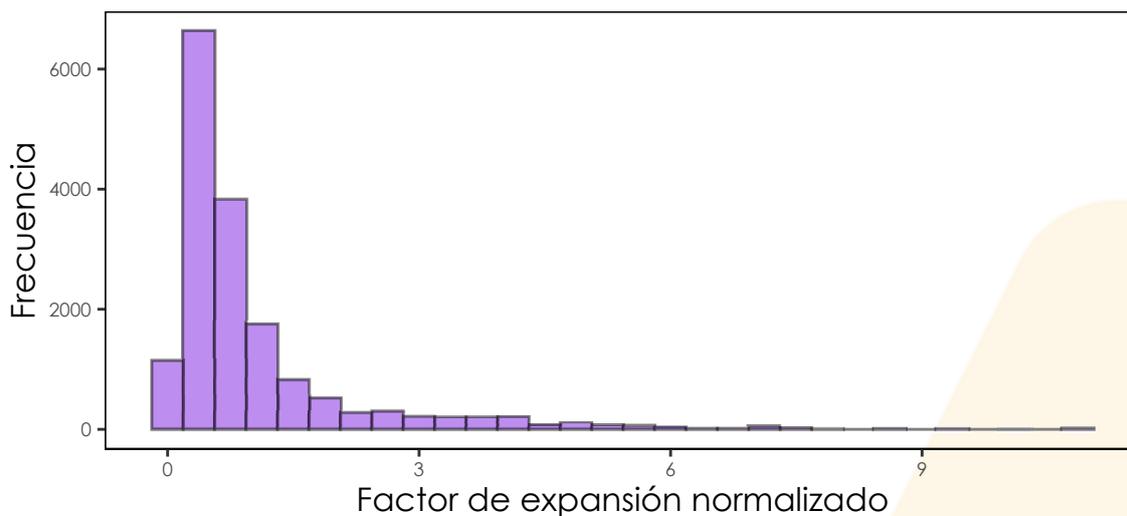
d_{3j} = factor de expansión ajustado por no respuesta de la vivienda j .

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

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

En la Figura 7 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 principalmente al incremento significativo en el tamaño de medida de algunas UPM y en menor medida a los ajustes de no respuesta y elegibilidad desconocida.

Ilustración 7: Distribución de los factores de expansión normalizados.



Fuente: Gestión de Diseño Muestral - DINEM.

Factor de expansión Desarrollo Infantil

En las 16.598 viviendas efectivas de este ejercicio, existen 16.623 hogares, de los cuales 16.078 participaron en la sección de Desarrollo infantil, para la cual se selecciona un niño menor a 5 años por 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:

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

donde L_k es el número total de infantes menores a 5 años en el hogar k . Por lo tanto, el factor de expansión queda definido por:

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

Siguiendo un proceso equivalente al mostrado en la sección anterior, el factor de expansión de Desarrollo Infantil normalizado $ddi_{n_j,k}$ es:

$$ddi_{n_j,k} = fn_{di} \cdot d_{4_j}.$$

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

Estimaciones de características

Una vez obtenidos los factores de expansión normalizados, para calcular las estimaciones nos apoyamos en el estimador de Horvitz-Thompson. Este estimador se utiliza para estimar el total de una característica determinada³, el cual está dado por (Särndal et al., 1992):

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

donde:

- \hat{Y}_{mHT} = 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, se obtiene que $\hat{Y}_m = \hat{Y}_{mHT}/n$, 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}_{mHT} . 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}_{kHT} viene dado por la siguiente expresión:

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

Un estimador insesgado para la varianza está dado por:

$$\hat{V}_{2st}(\hat{t}_\pi) = \sum_{s_I} \sum \Delta_{Iij} \frac{\hat{t}_{i\pi}}{\pi_{Ii}} \frac{\hat{t}_{j\pi}}{\pi_{Ij}} + \sum_{s_I} \frac{\hat{V}_i}{\pi_{Ii}},$$

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

$$\hat{V}_i = \sum_{s_{IIi}} \sum \Delta_{IIqr|i} \frac{\hat{t}_{iq\pi}}{\pi_{IIq|i}} \frac{\hat{t}_{ir\pi}}{\pi_{IIr|i}},$$

donde:

³Nótese 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.

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 9, donde se describe las etiquetas de las variables identificadoras de las UPM, estratos y factor de expansión.

Tabla 9: 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. 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 mismos⁴. 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:

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

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

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



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, 2019) INEC (2019). *Metodología de Diseño Muestra de la Encuesta Nacional de Salud y Nutrición 2018*. INEC, Quito.
- (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.
- (Valliant et al., 2013) Valliant, R., Dever, J. A., and Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, Nueva York.
- (Wolter, 1985) Wolter, K. M. (1985). *Introduction to variance estimation*. Springer-Verlag, Nueva York.

**ENCUESTA
NACIONAL SOBRE
DESNUTRICIÓN
*Infantil***

INEC

Buenas cifras,
mejores vidas



@Ecuadorencifras



@Ecuadorencifras



@InecEcuador



INECEcuador