



Diseño Muestral

ENDI Galápagos 2026

Octubre · 2026



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

Dirección/Departamento

Dirección de Infraestructura Estadística y Muestreo

Unidad

Gestión de Diseño Muestral

Elaborado por:

Geovanny Gutiérrez

Revisado por:

Jonathan Rosero

Aprobado por:

Jennifer Marcillo

Quito - Ecuador, 2026

Tabla de contenido

Introducción	1
Objetivos de la encuesta y población objetivo	1
Dominios de estudio	1
Marco muestral	2
Unidades Primarias de Muestreo	2
Diseño Muestral	4
Determinación del tamaño muestral	4
Cálculo del tamaño muestral	5
Selección de viviendas	6
Cobertura	7
Condición de ocupación y resultado de entrevista	7
Tasas de conformidad	9
Población objetivo	9
Factores de expansión	11
Probabilidad de primera etapa	11
Probabilidad de segunda etapa	11
Factor de expansión teórico	11
Ajuste de cobertura de primera etapa	12
Ajuste de cobertura de segunda etapa	12
Calibración a proyecciones poblacionales	15
Recorte de factores calibrados	16
Factor de expansión normalizado	16
Factor de expansión Lactancia Materna	18
Factor de expansión Desarrollo Infantil	19
Resumen de los factores por módulo	20
Estimaciones de características	21
Error de muestreo	21
Métodos de estimaciones de errores para diseños muestrales com- plejos	22
Estimación de la varianza con factores calibrados	23
Estimación de totales mediante fuentes de información auxiliares	25

Introducción

La implementación de la Encuesta Nacional sobre Desnutrición Infantil en la provincia de Galápagos presenta retos operativos más complejos que en las otras provincias de Ecuador, por esta razón se realiza una estrategia diferente de muestreo. En el presente documento se describe el marco de muestreo utilizado, la estrategia de muestreo implementada, el análisis de cobertura muestral y el cálculo de factores de expansión para ENDI Galápagos.

Objetivos de la encuesta y población objetivo

La Encuesta Nacional sobre Desnutrición Infantil (ENDI) tiene como principal objetivo medir los datos de desnutrición infantil, sobrepeso y obesidad, controles prenatales, vacunación oportuna de neumococo y rotavirus, anemia, pruebas de calidad de agua, entre otros. La población objetivo de la ENDI son los niños y niñas menores de cinco años de edad.

Dominios de estudio

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

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

En el caso de la ENDI Galápagos el dominio de estudio es la provincia como tal, cuyo operativo se lleva a cabo de julio a octubre de 2026.

Marco muestral

El Marco de Muestreo utilizado en la ENDI Galápagos se construyó a partir de los resultados del Precenso del VIII Censo de Población y VII de Vivienda del Ecuador (CPV 2022), proceso en el cual se llevó a cabo una actualización cartográfica integral de todo el territorio ecuatoriano.

En la provincia de Galápagos a base precensal cuenta con un total de 18.764 registros, los mismo se distribuyen de acuerdo a su condición de ocupación como se muestra en la Tabla 1.

Tabla 1: Distribución de registros base precensal por condición de ocupación.

Condición de ocupación	Número de registros
Colectiva	296
Particular - Desocupada	2.240
Particular - En construcción	1.000
Particular - Ocupada	8.282
Particular - Temporal	1.954
Sin vivienda	4.992
Total	18.764

Fuente: Gestión de Diseño Muestral - DINEM.

Para la construcción del Marco de Muestreo se consideran únicamente las 8.282 viviendas particular - ocupadas de acuerdo al Precenso.

Unidades Primarias de Muestreo

Las Unidades Primarias de Muestreo (UPM) se delimitan con base en las parroquias que conforman Galápagos, mediante la unión de manzanas y sectores en el área amanzanada y dispersa, respectivamente. Las UPM conformadas cuentan con al menos 60 viviendas ocupadas, sin embargo, debido al tamaño de ciertas parroquias en su área amanzanada o dispersa se permitieron las uniones que se describen a continuación:

- La parroquia Isla Santa María Floreana (200152) tiene 42 viviendas en la parte amanzanada y 2 en la parte dispersa, se genera una única UPM con todas las viviendas.
- La parroquia Tomás de Berlanga (200251) tiene 11 viviendas en la parte amanzanada y 80 en la parte dispersa, se genera una única UPM con todas las viviendas.
- La parte dispersa de la parroquia Puerto Ayora (200350) tiene 12 viviendas en total en 1 sector disperso, por lo que se junta con los sectores dispersos de la parroquia Bellavista (200351).

En total se forman 109 UPM cuya distribución en el número de viviendas por parroquia se describe en la Tabla 2

Tabla 2: Distribución de UPM por número de viviendas ocupadas por parroquia.

Parroquia	Código	Número de UPM	Número de Viviendas		
			Mínimo	Promedio	Máximo
Puerto Baquerizo Moreno	200150	31	60	72,23	106
El Progreso	200151	3	69	72,67	76
Isla Santa María Floreana	200152	1	44	44,00	44
Puerto Villamil	200250	10	60	69,00	94
Tomás de Berlanga	200251	1	91	91,00	91
Puerto Ayora	200350	48	60	74,50	113
Bellavista	200351	13	60	82,69	105
Santa Rosa	200352	2	77	89,00	101

Fuente: Gestión de Diseño Muestral - DINEM.

Diseño Muestral

El diseño muestral implementado es un muestreo probabilístico bietápico de elementos. En la primera etapa, se seleccionan 44 de las 109 UPM en las que se divide Galápagos con probabilidad proporcional al tamaño.

En la segunda etapa de muestreo, se enlistaron por completo todas las viviendas de las UPM seleccionadas, y participaron en la encuesta todas las viviendas con niños/as menores a 5 años identificadas. Además, participaron también las viviendas cuyo resultado de entrevista en el listado fue rechazo.

Determinación del tamaño muestral

El tamaño de muestra se determinó en dos partes:

- * Se calcula el tamaño de muestra de niños/as menores a 5 años de la misma manera que para la ronda 2 de ENDI Galápagos.
- * Este tamaño de muestra se ajusta en función al coeficiente de variación esperado para el Desnutrición Crónica en niños/as menores a 5 años en función a los resultados obtenidos en ENDI Galápagos Ronda 2.

Tamaño de muestra ENDI Galápagos Ronda 1

El tamaño de muestra se determinó mediante la siguiente ecuación:

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

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.
- def = Efecto de diseño.
- tnr = Tasa de no respuesta.

Además, se fijó la tasa de no respuesta en 0,20 ya que en la Encuesta Nacional de Salud y Nutrición 2018 (ENSANUT 2018), a pesar de haber contado con un proceso de enlistamiento previo al levantamiento de información, el número de viviendas identificadas originalmente como viviendas con niños/as en realidad no cumplía con esta característica.

Por otra parte, el indicador de diseño es la prevalencia de desnutrición crónica en niños/as menores a cinco años, el mismo se determinó a partir de la ENSANUT

2018, con un valor de 23,6%. El diseño muestral es muy parecido al muestreo aleatorio simple, por lo que se supone un efecto de diseño igual a 1.

Finalmente, el resto de parámetros involucrados en el cálculo del tamaño se fijan en:

$$e_{rel} = 0,165$$

$$z = 1,96$$

Con los parámetros descritos y la aplicación de la fórmula para determinar el tamaño muestral, se obtiene un tamaño muestral de 364 niños/as menores a cinco años para la ENDI Galápagos Ronda 1.

Ajuste del tamaño muestral en función al Coeficiente de Variación esperado

La Ronda 1 de la ENDI Galápagos reportó una desnutrición crónica en niños/as menores a 5 años del 9,6%, con un coeficiente de variación del 15,6%; sin embargo, el CV esperado por diseño era alrededor del 8,5%. Este no se alcanzó principalmente por dos motivos: baja considerable en la prevalencia del fenómeno (23,6% a 9,6%) y una variabilidad en la condición de ocupación de las viviendas de la provincia más alta de la esperada.

Para mitigar estos efectos, se incluyó un proceso de ajuste del tamaño muestral para ENDI Galápagos en función del coeficiente de variación esperado del 9,0% con base en la siguiente ecuación:

$$n_2 \geq n_1 \cdot \left(\frac{CV_1}{CV_2} \right)^2$$

donde:

n_1 = Tamaño de muestra ENDI Galápagos ronda 1.

CV_1 = Coeficiente de variación de ENDI Galápagos ronda 1.

CV_2 = Coeficiente de variación esperado para ENDI Galápagos ronda 2.

Cálculo del tamaño muestral

Con los parámetros descritos y la aplicación de la fórmula para determinar el tamaño muestral, se obtiene un tamaño muestral de 766 niños/as menores a cinco años para la ENDI Galápagos Ronda 2. Dadas las características del marco construido y la estrategia de muestreo a utilizar, este tamaño debe transformarse a Unidades Primarias de Muestreo.

Para transformarlo, utilizamos la relación entre el tamaño de muestra necesario y el stock de niños/as menores a 5 años en Galápagos (2,017), la cual es 37,9%, lo cual equivale a seleccionar 37 UPM de las 95 disponibles.

Así mismo, para la obtención del tamaño muestral para el 2026, se toma un CV esperado de 9,0% que resulta en la selección de 44 UPM para el operativo actual.



Selección de viviendas

Las 44 UPM seleccionadas en Galápagos se enlistaron por completo; se actualizó su condición de ocupación y se identificó si algún miembro era parte de la población objetivo de la encuesta. Se diligenciaron todas las viviendas que tenían al menos un niño/a menor a 5 años, junto con las viviendas clasificadas como *rechazo* y *nadie en casa*.

Cobertura

La cobertura analiza la información obtenida en campo de los informantes de las viviendas encuestadas. Para tal efecto, el proceso fue el siguiente:

1. Se enlistaron todas las viviendas de las UPM seleccionadas, las cuales se clasificaron en uno de los siguientes grupos:
 - No es vivienda (viviendas temporales, desocupadas, en construcción, destruidas, negocios).
 - Vivienda ocupada sin niños/as menores a 5 años.
 - Vivienda ocupada con al menos un/a niño/a menor a 5 años.
 - Rechazo.
 - Nadie en casa.
2. Según la condición y resultado de entrevista del enlistamiento, se clasifican las viviendas en dos grupos:
 - Viviendas a ser encuestadas: incluye las viviendas con al menos un niño/a menor a 5 años y Nadie en casa.
 - Viviendas a no ser encuestadas: todas las demás.

A continuación, se presenta el seguimiento de la cobertura, además de los cambios de ocupación y de la población objetivo, los cuales influyen directamente en la fase de ponderación y estimación de características de la ENDI Galápagos.

Condición de ocupación y resultado de entrevista

Como se puede observar en la Tabla 3, se investigaron 632 viviendas en Galápagos, de las cuales el 90,35% fueron efectivas. Además, se tiene que las viviendas catalogadas como *Nadie en casa* fueron la principal fuente de no efectividad, que representaron el 4,75% del total de viviendas visitadas.

Tabla 3: Condición de elegibilidad de las viviendas.

Condición de la vivienda	Galápagos	Porcentaje
Efectiva	571	90,35%
Rechazo	20	3,16%
Nadie en casa	30	4,75%
Vivienda temporal	1	0,16%
Vivienda desocupada	2	0,32%
Vivienda en construcción	1	0,16%
Vivienda destruida	0	0,00%

Negocio	0	0,00 %
Vivienda sin niño/a	7	1,11 %
Otra razón	0	0,0 %
Total	632	100,00 %

Fuente: Gestión de Diseño Muestral - DINEM.

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

1. Elegibilidad conocida

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

2. Elegibilidad desconocida

- Elegibilidad desconocida (ED)
 - Nadie en casa

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

Tabla 4: Condición de elegibilidad por dominio de diseño.

Condición de elegibilidad	Galápagos	Porcentaje
Elegible respondiente (RE)	571	90,35 %
Elegible no respondiente (NR)	20	3,16 %
No elegibles (NE)	11	1,74 %
Elegibilidad desconocida (ED)	30	4,75 %
Total	632	100,0 %

Fuente: Gestión de Diseño Muestral - DINEM.

Del total de viviendas investigadas, 11 viviendas fueron no elegibles en Galápagos, de las cuales 7 pertenecen a la categoría *viviendas sin niños/as menores de 5 años* como se observa en la Tabla 3; además, existe una mínima cantidad de

viviendas que son *elegibles no respondiente* y *elegibilidad desconocida*, como se aprecia en la Tabla 4.

Tasas de conformidad

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

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

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

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

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

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

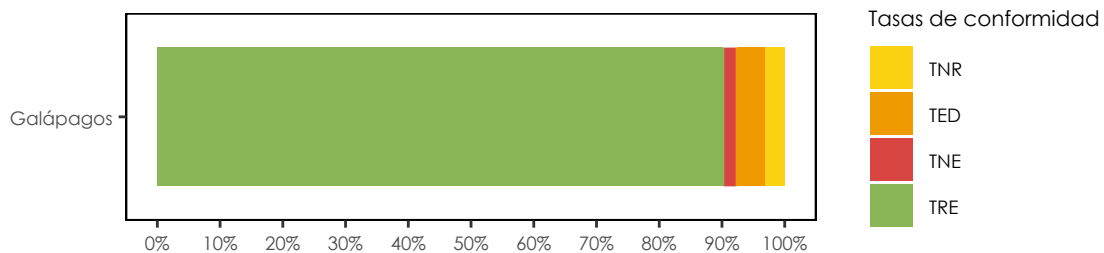
$$T_{NE} = \frac{NE}{RE + NR + NE + ED}$$

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

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

A continuación, se presentan las tasas de conformidad en Galápagos.

Figura 1: Tasas de conformidad para ENDI Galápagos.

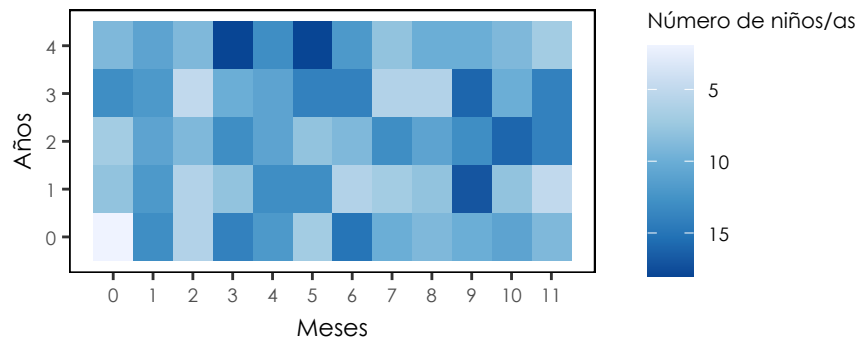


Fuente: Gestión de Diseño Muestral - DINEM.

Población objetivo

En el periodo de levantamiento de información, la ENDI Galápagos entrevistó a 629 niños/as menores de cinco años. En el siguiente gráfico se presenta su distribución por su edad en años y meses:

Figura 2: Distribución de niños/as menores de 5 años por años y meses.



Fuente: Gestión de Diseño Muestral - DINEM.

Puesto que las viviendas se seleccionan de manera sistemática aleatoria, se espera que el total de niños/as por año cumplido sea uniforme. En la tabla 5 se muestra su proporción por edad simple.

Tabla 5: Resumen de distribución de niños/as menores a cinco años por año y meses.

Años	Meses con menor concentración	Meses con mayor concentración	Población	Porcentaje
0	0	6	118	18,76 %
1	11	9	111	17,65 %
2	0	10	135	21,46 %
3	2	9	131	20,82 %
4	11	3, 5	134	21,30 %

Fuente: Gestión de Diseño Muestral - DINEM.

Factores de expansión

El diseño muestral de la ENDI Galápagos es probabilístico bietápico 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.

Probabilidad de primera etapa

En la primera etapa, se seleccionaron 44 UPM. La estratificación se definió a nivel de cantón, por lo que los estratos h corresponden a los tres cantones de la provincia. Por lo tanto, la probabilidad de inclusión de primera etapa de la UPM i es:

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

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 .
 M_h = número de UPM en el estrato h .

Probabilidad de segunda etapa

Dado que en Galápagos se realizó el enlistamiento y el levantamiento de información de manera integrada, se actualizó la condición de ocupación y elegibilidad de todas las viviendas de las UPM seleccionadas, y se entrevistaron todas las viviendas particulares con personas presentes con al menos un niño/a menor de 5 años, es decir:

$$pse_i := \frac{n_i}{N_i},$$

donde:

pse_i = probabilidad de segunda etapa de las viviendas en la UPM i .
 n_i = viviendas ocupadas seleccionadas en la UPM i .
 N_i = viviendas ocupadas en la UPM i .

Factor de expansión teórico

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

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

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

Ajuste de cobertura de primera etapa

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

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

donde m es el número total de UPM seleccionadas y m' es el número total de UPM investigadas en la ENDI. El ajuste se realiza de forma global, es decir, es el mismo para todas las UPM investigadas, incluidas las de Galápagos, de manera consistente con los factores nacionales.

Factor de expansión ajustado por cobertura de primera etapa

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

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

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

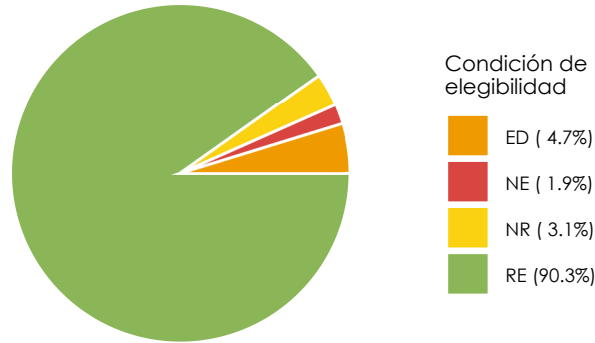
Al momento de la entrevista en campo, cada vivienda muestreada se clasificó en uno de los cuatro conjuntos anteriormente descritos (RE , NR , NE y ED). Si existen unidades que son no elegibles en la muestra, esto es evidencia de que hay unidades no elegibles tanto sobre ED como sobre lo no muestreado.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

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

Para tal efecto, es necesario considerar la distribución inicial estimada de todas las viviendas por condición de elegibilidad con base en el factor ajustado por cobertura de primera etapa (d_{1j})¹, la cual se aprecia en la Figura 3.

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

Figura 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 la ENDI Galápagos se realiza el ajuste a nivel de estrato, con base en el análisis de la cobertura a nivel de UPM y viviendas, así como por la naturaleza de la población objetivo de la misma. Por lo tanto, el ajuste por elegibilidad desconocida se define como:

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

donde:

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

s_h = muestra total de viviendas en el estrato h .

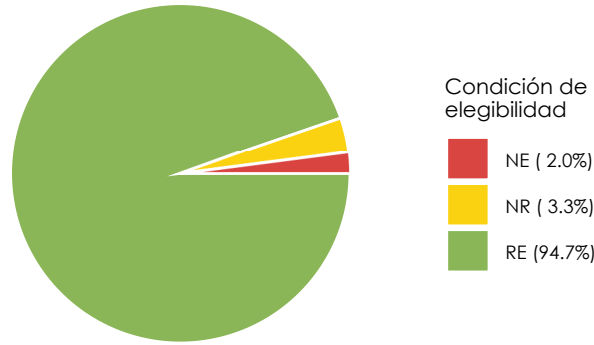
d_{1j} = factor de expansión de la vivienda j .

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, lo que da como resultado la distribución presentada a continuación:

Figura 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, según cuánto se conoce de los no respondientes. Para la ENDI Galápagos, la información disponible de los no respondientes es escasa y limitada, por lo tanto, la propensión a responder se modela únicamente con el estrato (cantón) como variable explicativa.

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

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

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

donde:

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

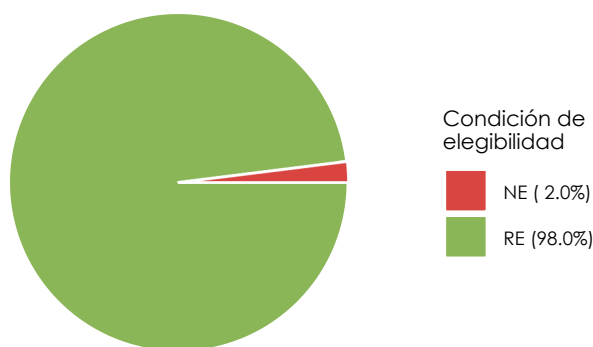
Al ser el estrato la única variable explicativa, la propensión estimada $\hat{\phi}_h$ es constante dentro de cada estrato. El factor de expansión ajustado por no respuesta (d_{3j}) queda definido de la siguiente manera:

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

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

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

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



Fuente: Gestión de Diseño Muestral - DINEM.

El impacto de aplicar los ajustes de elegibilidad desconocida y no respuesta sobre los respondientes elegibles (*RE*) puede cuantificarse mediante los factores d_{3j} y d_{1j} .

Calibración a proyecciones poblacionales

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

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

$$\mathbf{x}_j = \left(x_j^{(20)}, x_j^h, x_j^{(0)}, \dots, x_j^{(3)} \right)^T,$$

donde:

- $x_j^{(20)}$ = número de niños/as menores de 5 años del hogar j (Galápagos constituye una sola provincia).
 x_j^h = número de niños (sexo masculino) menores de 5 años en el hogar j .
 $x_j^{(a)}$ = número de niños/as de a años cumplidos en el hogar j .

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

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

con

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

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

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

Recorte de factores calibrados

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

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

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

Factor de expansión normalizado

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

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

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

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

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

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

donde:

f_n = factor de normalización.

n_R = número de hogares efectivos, es decir $n_R = |s_R|$.

d_{5j} = factor de expansión calibrado y recortado del hogar j .

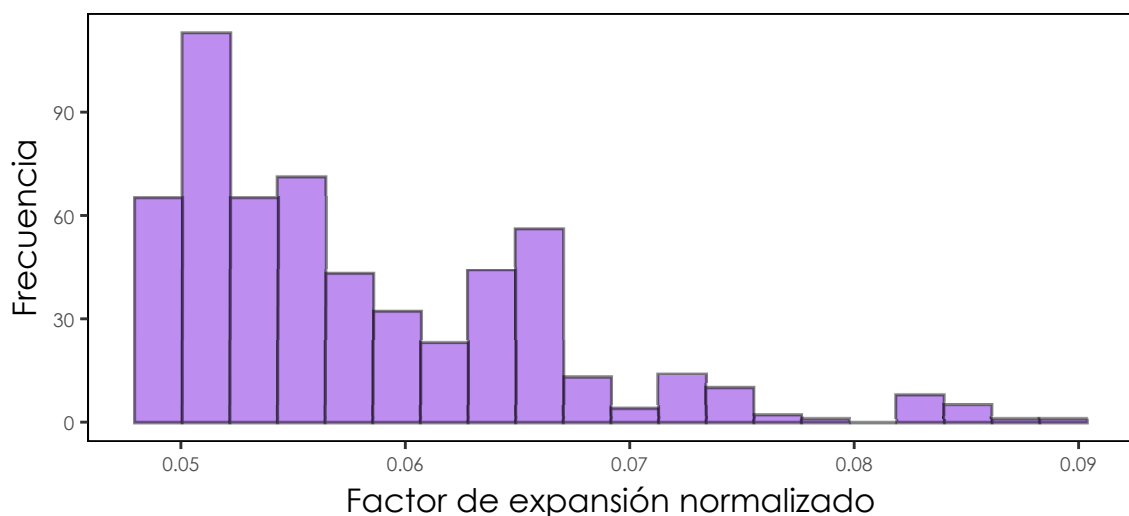
De esta forma, el factor de expansión normalizado d_{nj} es:

$$d_{nj} := f_n \cdot d_{5j},$$

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

En la Figura 6 se muestra la distribución de los factores de expansión normalizados para ENDI Galápagos. Se puede apreciar que la distribución es la esperada para el diseño aplicado, la misma se explica por el número de viviendas por UPM.

Figura 6: Distribución de los factores de expansión normalizados.



Fuente: Gestión de Diseño Muestral - DINEM.

Factor de expansión Lactancia Materna

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

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

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

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

Ajuste de cobertura por no respuesta

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

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

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

$$d_{6j,k}^{lm} := \frac{d_{5j}}{\phi_h^{lm}} \quad \text{si } r_{j,k} = 1.$$

Calibración, recorte y normalización

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

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

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

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

$$d_{lm_{n,j,k}} = f_{n_{lm}} \cdot d_{8,j,k}^{lm}, \quad f_{n_{lm}} := \frac{n_R^{lm}}{\sum_{(j,k) \in s_R^{lm}} d_{8,j,k}^{lm}},$$

donde $f_{n_{lm}}$ es el factor de normalización de Lactancia Materna y $n_R^{lm} = |s_R^{lm}|$ es el número de niños/as efectivos del módulo.

Factor de expansión Desarrollo Infantil

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

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

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

El módulo de Desarrollo Infantil se levantó durante todo el operativo de Galápagos, por lo que no requiere ajuste por levantamiento en el segundo semestre y se parte directamente del factor de hogares calibrado y recortado $d_{5,j}$, antes de su normalización.

Factor de expansión

Al incorporar la tercera etapa de selección se obtiene:

$$d_{6,j}^{dein} := d_{5,j} \cdot \frac{1}{pte_j} = d_{5,j} \cdot D_j.$$

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

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

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

$$d_{8,j}^{dein} := g_j^{dein} \cdot d_{7,j}^{dein}, \quad d_{dein_{n_j}} = f_{n_{dein}} \cdot d_{9,j}^{dein}, \quad f_{n_{dein}} := \frac{n_R^{dein}}{\sum_{j \in s_R^{dein}} d_{9,j}^{dein}},$$

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

Resumen de los factores por módulo

La Tabla 6 resume los pasos aplicados en cada uno de los factores de expansión.

Tabla 6: Pasos de cálculo de los factores de expansión por módulo.

Paso	Hogares	Lactancia Materna	Desarrollo Infantil
Unidad de análisis	Hogar	Niño/a menor de 3 años	Niño/a menor de 5 años seleccionado
Factor de partida	d_0 con cobertura de UPM, elegibilidad desconocida y no respuesta (modelo logístico por estrato)	Factor de hogares calibrado y recortado d_5	d_5 ajustado por tercera etapa
Ajuste de cobertura por no respuesta	Incluido en el factor de partida	Modelo logístico por estrato	Modelo logístico por estrato
Calibración	Provincia, sexo y edad simple (menores de 5 años)	Provincia, sexo y edad simple (menores de 3 años)	Provincia, sexo y edad simple (menores de 5 años)
Recorte	3.5 veces la mediana por estrato	3.5 veces la mediana por estrato	3.5 veces la mediana por estrato
Normalización	Hogares efectivos	Niños/as efectivos/as	Niños/as efectivos/as

Fuente: Gestión de Diseño Muestral - DINEM.

Estimaciones de características

Una vez obtenidos los factores de expansión normalizados, para calcular las estimaciones de la ENDI nos apoyamos en el estimador de Horvitz-Thompson, el cual se utiliza para estimar el total de una característica determinada². 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 .

A partir de la definición del factor de normalización utilizado en la ENDI, se obtiene que $\hat{\bar{Y}}_m = \hat{Y}_{m_{HT}}/n$; así, se obtienen 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:

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

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

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

donde:

²Nótese 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:

- Métodos exactos.
- Métodos de la última UPM.
- Aproximaciones por linealización.
- Técnicas de replicación.

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

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

que existen estudios [Kish and Frankel, 1974] que concluyen que las diferencias presentadas no son significativas cuando se trata de grandes muestras.

El INEC y sus usuarios utilizan diversos programas estadísticos para la estimación de los parámetros de interés, entre los cuales los más utilizados son SPSS, Stata y R. En virtud de las características de cada uno de los métodos, es la técnica de la última UPM en combinación con la linealización de Taylor, la cual induce a una adecuada aproximación del error muestral sobre los indicadores más importantes de las encuestas dirigidas a hogares, pero sobre todo resalta por su facilidad de cálculo y replica. En este sentido, será esta la técnica utilizada para la estimación de los errores muestrales en la ENDI.

Las variables requeridas para declarar el diseño muestral en los programas estadísticos (SPSS, Stata y R) y que permiten calcular los errores de muestreo bajo el esquema anterior son presentadas en la tabla 7, donde se describe las etiquetas de las variables identificadoras de las UPM, estratos y factor de expansión.

Tabla 7: 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 (cantón)
Factor de expansión	fexp	Factor de expansión normalizado

Fuente: Gestión de Diseño Muestral - DINEM.

Estimación de la varianza con factores calibrados

Los factores de expansión finales de la ENDI Galápagos están calibrados a las Estimaciones y Proyecciones de Población 2026 por provincia, sexo y edad simple, de modo que el estimador de total $\hat{Y}_{cal}$ no coincide con el estimador de Horvitz-Thompson construido con los factores originales. Por esta razón, la varianza de los estimadores no puede calcularse directamente sobre la variable de interés y_k : hacerlo ignora la información auxiliar utilizada en la calibración y sobrestima el error de muestreo de las variables relacionadas con ella. De acuerdo con [Särndal et al., 1992], la varianza del estimador calibrado se aproxima mediante la varianza del estimador de Horvitz-Thompson aplicado a los residuos de la regresión de y_k sobre las variables de calibración, con la técnica de la última UPM descrita anteriormente. El procedimiento es el mismo para los factores de Hogares, Lactancia Materna y Desarrollo Infantil; solo cambian la unidad de análisis k y el vector de variables de calibración $\mathbf{x}_k$ de cada módulo.

Sean $\tilde{d}_k$ el factor de expansión previo a la calibración y d_k^c el factor calibrado y recortado de la unidad k (por ejemplo, d_{3k} y d_{5k} en el caso de hogares). El

estimador de total calibrado es:

$$\hat{Y}_{cal} = \sum_{k \in s_R} d_k^c \cdot y_k.$$

El cálculo de su varianza se realiza en los siguientes pasos:

- 1) Se estiman los coeficientes de la regresión de y_k sobre las variables de calibración:

$$\hat{\mathbf{B}} = \left(\sum_{k \in s_R} \tilde{d}_k \mathbf{x}_k \mathbf{x}_k^\top \right)^{-1} \sum_{k \in s_R} \tilde{d}_k \mathbf{x}_k y_k.$$

- 2) Se calcula el residuo de cada unidad:

$$\hat{e}_k = y_k - \mathbf{x}_k^\top \hat{\mathbf{B}}.$$

- 3) Se obtiene el total ponderado de los residuos de cada UPM i del estrato h :

$$\hat{z}_{hi} = \sum_{k \in s_{hi}} d_k^c \cdot \hat{e}_k.$$

- 4) Se estima la varianza con la técnica de la última UPM:

$$\hat{V}(\hat{Y}_{cal}) = \sum_h \frac{m_h}{m_h - 1} \sum_{i \in s_h} (\hat{z}_{hi} - \bar{z}_h)^2, \quad \bar{z}_h = \frac{1}{m_h} \sum_{i \in s_h} \hat{z}_{hi},$$

donde:

$\mathbf{x}_k$ = vector de variables de calibración de la unidad k (provincia, sexo y edad simple de los niños/as de la población de referencia del módulo).

s_{hi} = unidades de análisis de la UPM i del estrato h .

m_h = número de UPM seleccionadas en el estrato h (cantón).

El recorte de factores se considera parte del factor final, por lo que d_k^c incluye su efecto.

Para medias y proporciones, estimadas como razón $\hat{\bar{Y}} = \hat{Y}_{cal} / \hat{N}$ con $\hat{N} = \sum_{k \in s_R} d_k^c$ (equivalente a utilizar los factores normalizados), se aplica el mismo procedimiento sobre la variable linealizada:

$$u_k = \frac{y_k - \hat{\bar{Y}}}{\hat{N}},$$

que sustituye a y_k en los pasos 1) al 4). Para dominios de estudio, y_k toma el valor cero fuera del dominio, y se mantienen en el cálculo todas las UPM de la muestra. Finalmente, el coeficiente de variación se calcula como $CV = \sqrt{\hat{V}(\cdot)} / \hat{Y}_{cal}$, o su equivalente para la media. Las estimaciones de total pueden obtenerse directamente con el factor calibrado d_k^c , que reproduce los totales poblacionales utilizados en la calibración.

Estimación de totales mediante fuentes de información auxiliares

Como se menciona en las secciones anteriores, el factor de expansión normalizado puede ser usado para estimar medias, proporciones, tasas y razones de los principales indicadores de la ENDI. Sin embargo, si es necesario contar con estimaciones de totales para los mismos, estas se pueden obtener con el apoyo de información auxiliar disponible para la población objetivo de los mismos³. Por ejemplo:

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

se tiene que:

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

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 al estimador de total auxiliar, se reemplaza $\hat{Y}_m$ por $LI(\hat{Y}_m)$ y $LS(\hat{Y}_m)$, respectivamente.

³El total auxiliar puede ser obtenido de diferentes fuentes de información, entre ellas: censo de población, proyecciones poblacionales y registros administrativos

Referencias

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



Buenas cifras,
mejores vidas

**ENCUESTA
NACIONAL SOBRE
DESNUTRICIÓN
*Infantil***



@InecEcuador



@ecuadorencifras



@ecuadorencifras



INECEcuador