



Diseño Muestral

ENDI Galápagos

Octubre, 2023





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:

Javier Núñez
Angel Gaibor
Giovanella Peña

Revisado por:

Christian Garcés

Aprobado por:

Julio Muñoz


Quito - Ecuador, 2023





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
Parámetros muestrales	4
Cálculo del tamaño muestral	5
Selección de viviendas	5
Cobertura	7
Condición de ocupación y resultado de entrevista	7
Tasas de conformidad	8
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
Factor de expansión normalizado	15
Estimaciones de características	17
Error de muestreo	17
Métodos de estimaciones de errores para diseños muestrales complejos	18
Estimación de totales usando fuentes de información auxiliares	19





Introducción

Para la implementación de la Encuesta Nacional sobre Desnutrición Infantil 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 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.

En el caso de la ENDI Galápagos el dominio de estudio es la provincia como tal, cuyo operativo se lleva a cabo de abril a julio de 2023.

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 considerando las parroquias que conforman Galápagos, juntando manzanas y sectores en el área amanzanada y dispersa, respectivamente. Las UPM conformadas cuentan con al menos 80 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 95 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	26	80	86,12	97
El Progreso	200151	3	69	72,67	76
Isla Santa María Floreana	200152	1	44	44,00	44
Puerto Villamil	200250	8	82	86,25	91
Tomás de Berlanga	200251	1	91	91,00	91
Puerto Ayora	200350	42	80	89,21	105
Bellavista	200351	12	82	89,58	98
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 las 95 UPM en las que se divide Galápagos.

En la segunda etapa de muestreo, se selecciona de manera sistemática aleatoria un número fijo de viviendas ocupadas, siendo 24 el número de viviendas seleccionadas por UPM. A continuación se detallan los procesos involucrados en el diseño muestral de la ENDI Galápagos.

Determinación del tamaño muestral

El tamaño de muestra 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

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 el no contar con un proceso de enlistamiento particular para Galápagos.

Por otra parte, 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, siendo 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:

$$e_{rel} = 0,165$$
$$z = 1,96$$
$$tnr = 0,20$$

Cálculo del tamaño muestral

Utilizando los parámetros descritos en la sección anterior y aplicando la fórmula para determinar el tamaño muestral, se obtiene un tamaño muestral de 364 niños menores a cinco años. Dadas las características del marco construido, este tamaño debe transformarse a viviendas ocupadas.

En primer lugar, para transformar el número de niños menores a cinco años a viviendas con niños menores a cinco años se debe multiplicar por un factor que los relacione. De acuerdo a resultados preliminares del CPV 2022 la proporción de niños menores a cinco años en viviendas con niños menores a cinco años es de 1,12, por lo tanto se debe seleccionar 326 viviendas con niños menores a cinco años.

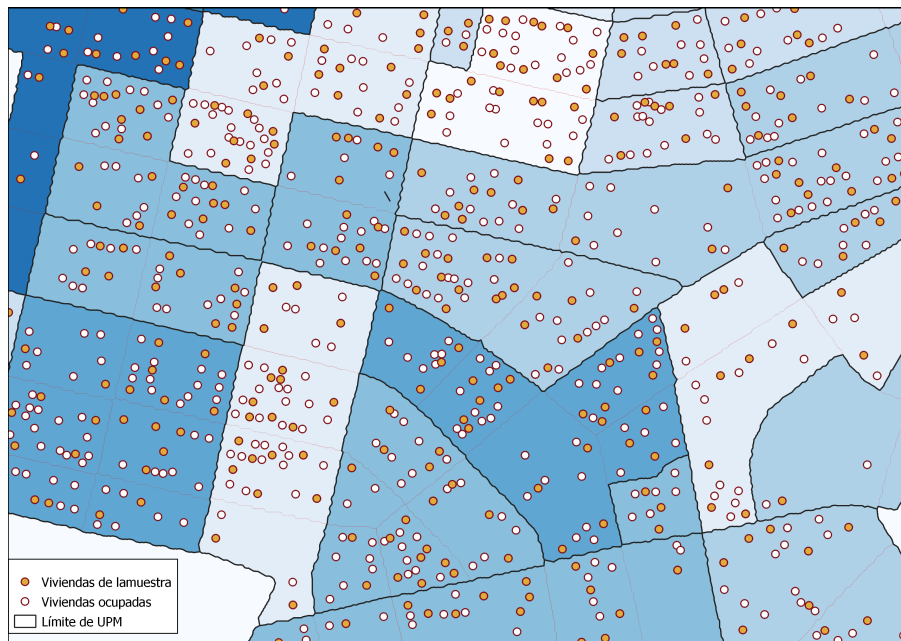
En segundo lugar, puesto que Galápagos no contó con un proceso de enlistamiento no se pueden clasificar las viviendas ocupadas en viviendas con y sin niños menores de cinco años, por lo tanto la muestra se debe escoger del total de viviendas ocupadas. Para tal efecto se divide el tamaño muestral de viviendas con niños menores a cinco años para su proporción sobre el total de viviendas ocupadas, siendo este 0,1812 de acuerdo a información preliminar del CPV 2022, siendo 1.797 el número de viviendas ocupadas necesarias. Bajo estas consideraciones el número de viviendas sin niño va a ser considerablemente más alto que el de viviendas con niños.

Finalmente, suponiendo una tasa de no respuesta del 0,2 se tendría una muestra de 2.247 viviendas ocupadas, al distribuir ese valor entre las 95 UPM que conforman Galápagos, se aproxima a 24 viviendas por UPM teniendo un total de 2.280 viviendas ocupadas a seleccionar.

Selección de viviendas

En cada una de las 95 UPM de Galápagos se seleccionan 24 viviendas ocupadas de forma sistemática aleatoria. En la Figura 1 se puede apreciar una parte de la parroquia Puerto Ayora, en la que se presenta la distribución de las viviendas seleccionadas en la muestra.

Figura 1: Distribución de la muestra en Puerto Ayora.



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 la ENDI Galápagos.

Condición de ocupación y resultado de entrevista

Como se puede observar en la Tabla 3, se investigaron 2.280 viviendas en Galápagos, siendo el 12,6% efectivas. Además, se puede determinar que el porcentaje de viviendas con niños menores a 5 años, efectivas y rechazo, con respecto a las viviendas ocupadas² tras el levantamiento de información representa el 15,0%.

Tabla 3: Condición de elegibilidad de las viviendas.

Condición de la vivienda	Galápagos	Porcentaje
Efectiva	288	12,6%
Rechazo	10	0,4%
Nadie en casa	58	2,5%
Vivienda temporal	69	3,0%
Vivienda desocupada	111	4,9%
Vivienda en construcción	14	0,6%
Vivienda destruida	12	0,5%
Negocio	54	2,4%
Vivienda sin niño	1.637	71,8%
Otra razón	27	1,2%
Total	2.280	100,0%

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

¹Las viviendas seleccionadas por parte de diseño muestral, son viviendas ocupadas de acuerdo a la actualización cartográfica del CPV 2022.

²Viviendas ocupadas: Efectiva, rechazo, nadie en casa y vivienda sin niño

- 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

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)	288	12,6%
Elegible no respondiente (NR)	10	0,4%
No elegibles (NE)	1.924	84,5%
Elegibilidad desconocida (ED)	58	2,5%
Total	2.280	100,0%

Fuente: Gestión de Diseño Muestral - DINEM.

Del total de viviendas investigadas, 1.924 viviendas fueron no elegibles en Galápagos, de las cuales la mayoría pertenecen a la categoría *viviendas sin niños 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

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 rechazado 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.

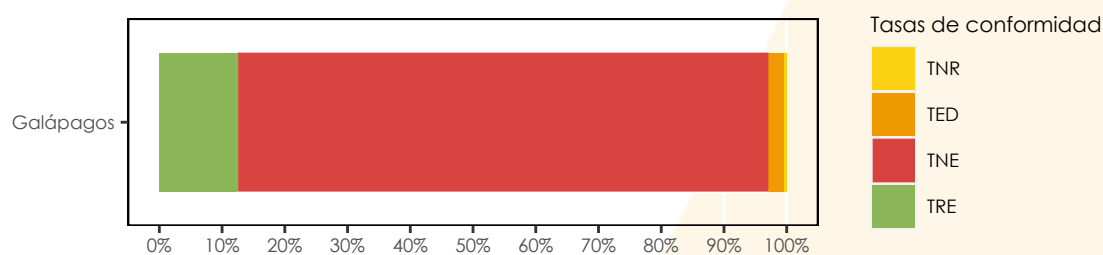
$$T_{NE} = \frac{NE}{RE + NR + NE + ED}$$

- 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 presenta las tasas de conformidad en Galápagos.

Figura 2: Tasas de conformidad para ENDI Galápagos.

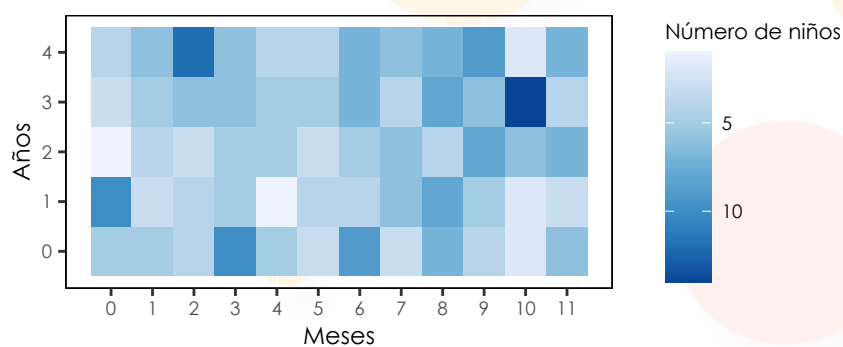


Fuente: Gestión de Diseño Muestral - DINEM.

Población objetivo

En el periodo de levantamiento de información, la ENDI Galápagos entrevistó a 322 niños menores a cinco años. En el siguiente gráfico se presenta su distribución por su edad en años y meses:

Figura 3: Distribución de niños 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 por año cumplido sea uniforme, lo cual se aprecia en la tabla 5.

Tabla 5: 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	10	3	63	19,6%
1	4	0	55	17,1%
2	0	9	56	17,5%
3	0	10	74	22,9%
4	10	2	74	22,9%

Fuente: Gestión de Diseño Muestral - DINEM.

Factores de expansión

El diseño muestral de la ENDI Galápagos 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 las 95 UPM que conforman todo Galápagos. 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 no hubo enlistamiento se seleccionaron 24 viviendas ocupadas en cada una de las UPM:

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

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:

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

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:

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

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

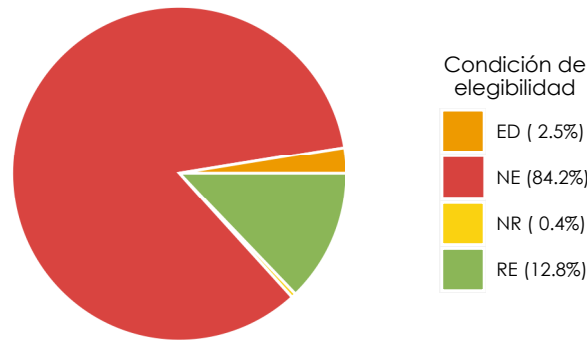
Al momento de la entrevista en campo, cada vivienda muestreada se clasificó en uno de los cuatro conjuntos anteriormente descritos (RE , NR , NE y ED). Si existen unidades que son no elegibles en la muestra, esto es evidencia de que hay unidades no elegibles tanto sobre ED como sobre lo no muestreado. 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 por condición de elegibilidad considerando el factor ajustado por cobertura de primera etapa (d_{1j})³, la cual se aprecia en la Figura 4.

³De aquí en adelante se notará como d_{1j} al factor de expansión ajustado por cobertura de primera etapa en la vivienda j .

Figura 4: 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, 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 .

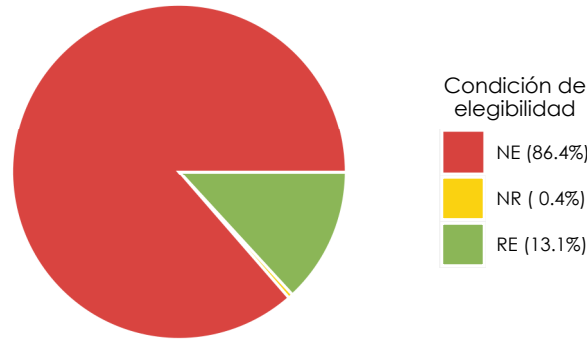
d_{1j} = factor de expansión de la vivienda j .

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:

Figura 5: 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 Galápagos, 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 .

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

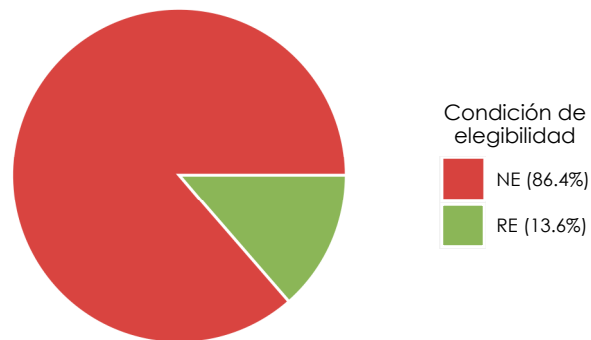
d_{2j} = factor de expansión ajustado por elegibilidad desconocida de la vivienda j .

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:

Figura 6: Distribución de viviendas por condición de elegibilidad expandidas por d_3 .



Fuente: Gestión de Diseño Muestral - DINEM.

El impacto de aplicar el ajuste de elegibilidad desconocida y no respuesta aplicados sobre los respondientes elegibles (RE), puede cuantificarse mediante los factores d_{3j} y d_{1j} , aumentando en un 7% la magnitud de los factores iniciales para todas las viviendas en RE de Galápagos.

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 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 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 .

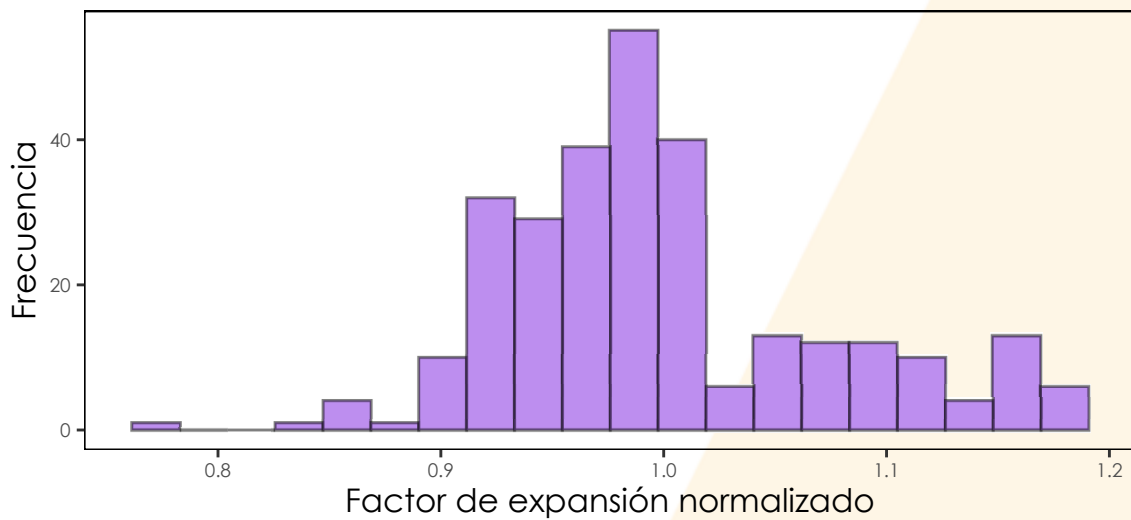
d_{3_j} = factor de expansión ajustado por no respuesta de la vivienda j .

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

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

En la Figura 7 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 7: Distribución de los factores de expansión normalizados.



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 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:

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

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:

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

Tabla 6: Variables requeridas para la declaración del diseño muestral.

Característica	Variable	Descripción
Unidad primaria de muestreo	id_upm	Unidad primaria de muestreo de ENDI
Unidad secundaria de muestreo	id_viv	Unidad secundaria de muestreo de ENDI
Probabilidad de selección de primera etapa	ppe	Probabilidad de selección de primera etapa
Probabilidad de selección de segunda etapa	pse	Probabilidad de selección de segunda etapa
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 ,



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



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





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.

**ENCUESTA
NACIONAL SOBRE
DESNUTRICIÓN
*Infantil***

INEC

Buenas cifras,
mejores vidas



@Ecuadorencifras



@Ecuadorencifras



@InecEcuador



INECEcuador