



Encuesta Nacional de Ingresos y Gastos de los Hogares Urbanos y Rurales

ENIGHUR - 2024

Marzo · 2026

Diseño Muestral de la Encuesta Nacional de Ingresos y Gastos de los Hogares

Año de investigación 2024



Dirección

Dirección de Infraestructura Estadística y Muestreo

Unidad

Gestión de Diseño Muestral

Elaborado por

Laura Apupalo

Revisado por

Jonathan Rosero

Aprobado por

Jennifer Marcillo

Contenido

Introducción	5
Antecedentes.....	6
1. Planificación.....	7
1.1 Objetivos y delimitación de la operación estadística.....	7
1.1.1 Objetivos.....	7
1.1.2 Delimitación del Área Urbana	8
2.1 Diseñar el marco y la muestra	8
2.1.1 Tipo y etapas de diseño muestral.....	8
2.1.2 Dominios de estudio	9
2.1.3 Marco Muestral	10
2.1.4 Variables de diseño	10
2.1.5 Tamaño muestral.....	10
2.1.5.1 Tamaño de muestra de viviendas	11
2.1.5.2 Efecto de diseño (DEFF).....	11
2.1.5.3 Nivel de confianza.....	12
2.1.5.4 Margen de error	12
2.1.5.5 Tasa de no respondientes (TNR)	12
2.1.5.6 Tamaño de muestra de UPM	12
2.1.5.7 Asignación de la muestra	12
2.1.5.6 Distribución espacial y temporal de la muestra.....	13
2.1.5.7 Resultados del tamaño muestral.....	13
2.1.6 Selección de la muestra	15
2.2 Control de cobertura de campo y muestral	15
2.3 Construcción de factores de expansión.....	16
2.3.1 Probabilidad de primera etapa.....	16
2.3.2 Ajuste de cobertura de primera etapa	17
2.3.3 Probabilidad de segunda etapa.....	18
2.3.4 Factor de expansión teórico	18
2.3.5 Factor de expansión ajustado por cobertura de primera etapa	18
2.3.6 Ajuste por cobertura de viviendas	18
3. Estimaciones y errores muestrales	21
3.1 Intervalos de confianza	21
3.2 Estimador de calibración.....	21
3.3 Recorte de factores de expansión extremos	23
Referencias	27

Introducción

Las Encuestas de Ingresos y Gastos de los hogares representan una de las investigaciones estadísticas más complejas por el número de temáticas investigadas, así como por la profundidad con la que se recolecta la información, que permite conocer las características demográficas, educacionales, ocupacionales de los miembros del hogar, así como el monto de los ingresos generados por los hogares y la manera en que se asignan para la adquisición de bienes y servicios incluidas las transacciones financieras. En el presente documento se detalla el proceso junto con un análisis de cada uno de los pasos involucrados.

Con los resultados de las entrevistas realizadas en las viviendas seleccionadas en la muestra, se analiza la cobertura de la encuesta que interfiere directamente con los procesos relacionados al diseño muestral. La cobertura se refiere a la información que se obtuvo en campo y que ha sido proporcionada por los informantes de los hogares. Se presenta la información recolectada desde el periodo 01 hasta el periodo 13. A lo largo del documento se hace referencia y se presentan los resultados a nivel de vivienda desagregados por provincia y por Coordinación Zonal: Administración Central Campo (Adm.C.Campo), Centro, Litoral y Sur.

El diseño muestral implementado en la Encuesta Nacional de Ingresos de Hogares Urbanos y Rurales (ENIGHUR) es un muestreo probabilístico bietápico estratificado de elementos.

En la primera etapa, se selecciona una muestra estratificada de UPM con probabilidad proporcional al tamaño (PPT), donde la medida de tamaño de cada UPM está dada por el total de viviendas particulares ocupadas. Luego, se enlista la totalidad de viviendas que conforman cada UPM para, en una segunda etapa, seleccionar aleatoriamente un total fijo de 12 viviendas en cada UPM seleccionada.

Antecedentes

El Instituto Nacional de Estadística y Censos ha realizado diversas investigaciones que permiten conocer la composición del presupuesto y el gasto de los hogares del país, entre ellas, las siguientes:

- ❖ Encuesta de Presupuestos Familiares en el Área Urbana (1975)
- ❖ Encuesta de Ingresos y Gastos de los Hogares Rurales (1978-1979)
- ❖ Módulo de Presupuestos Familiares Área Urbana (1991)
- ❖ Encuesta Nacional de Ingresos y Gastos de Hogares Urbanos (1994-1995)
- ❖ Encuesta Nacional de Ingresos y Gastos de Hogares Urbanos - ENIGHU (2003-2004)
- ❖ Encuesta Nacional de Ingresos y Gastos de Hogares Urbanos y Rurales ENIGHUR (2011-2012).

En la ENIGHUR 2011, el diseño muestral fue probabilístico, estratificado, multietápico y por conglomerados, con aproximadamente 5.958 UPM distribuidas en 4 periodos trimestrales, lo que implicaba un levantamiento concentrado en ciertos meses del año y una muestra cercana a 39.600 hogares. En contraste, la ENIGHUR 2024 mantiene el carácter probabilístico bietápico estratificado de elementos, se optimiza la muestra a 3.432 UPM distribuidas de forma continua en 13 periodos, con 264 UPM por periodo y 66 por semana, lo que permite una recolección más homogénea y una mejor captura de variaciones estacionales y coyunturales; además, incorpora tecnologías de captura digital y marcos censales actualizados.

La ejecución de este tipo de investigaciones permite lograr varios objetivos, lo que implica la adopción de marcos conceptuales que requieren la aplicación de mediciones objetivas y precisas, en concordancia con las recomendaciones más actuales de los Organismos Internacionales especializados, así como a las buenas prácticas en materia de encuestas de hogares desarrolladas por las oficinas de estadísticas a nivel mundial.

1. Planificación

1.1 Objetivos y delimitación de la operación estadística

1.1.1 Objetivos

Objetivo General

La Encuesta Nacional de Ingresos y Gastos de Hogares Urbanos y Rurales – ENIGHUR, tiene como objetivo general proporcionar información sobre la distribución, monto, y estructura del ingreso y el gasto de los hogares, a partir de la caracterización demográfica y socioeconómica de los mismos.

Objetivos Específicos

- ❖ Proporcionar la información estadística para sustentar el Cambio de año base del Índice de Precios al Consumidor (IPC), así como para actualizar la “Canasta de Bienes y Servicios”.
- ❖ Estimar la estructura del consumo de los hogares privados para la adquisición de bienes y servicios, de acuerdo a las diferentes fuentes de ingreso.
- ❖ Viabilizar el conocimiento detallado de la estructura del presupuesto de los hogares, de acuerdo a los distintos orígenes y monto de los ingresos y la forma del gasto, así como los hábitos de consumo de los diferentes bienes y servicios.
- ❖ Conocer el ahorro o endeudamiento de los hogares y la manera en que los hogares lo asignan.
- ❖ Generar información que sirva de insumo para medir la pobreza.
- ❖ Generar información para la estimación global del consumo final y el ingreso personal disponible de los hogares para la conformación del Sistema de Cuentas Nacionales.

Universo de estudio

Todas las viviendas particulares ocupadas del país, los hogares que residen habitualmente en ellas y todas las personas que conforman dichos hogares, tanto en áreas urbanas como rurales, que cumplen con los criterios de inclusión definidos en el diseño muestral de la ENIGHUR para el año de levantamiento.

Unidad de observación y unidad de análisis

Las unidades de observación son las viviendas del país, los hogares que habitan dentro de ellas y los residentes habituales del hogar.

Las unidades de análisis son los hogares y los miembros habituales del hogar.

Cobertura geográfica

Cobertura nacional, que incluye todas las regiones del Ecuador: Costa, Sierra, Amazonía y Región Insular (Galápagos), diferenciando áreas urbanas y rurales.

Población objetivo

Todos los hogares y personas residentes habituales en viviendas particulares ocupadas del territorio ecuatoriano, en áreas urbanas y rurales.

Continuidad

Desde el año 2011 hasta la actualidad.

Periodicidad

Cada 10 años.

1.1.2 Delimitación del Área Urbana

Para los efectos del marco de muestreo, las áreas urbanas se definieron a partir de los centros poblados con una población de 2.000 y más habitantes conforme a las recomendaciones de la Comunidad Andina de Naciones, sin importar si son cabeceras administrativas o no.

2. Diseño y construcción

2.1 Diseñar el marco y la muestra

2.1.1 Tipo y etapas de diseño muestral

El diseño muestral implementado en la ENIGHUR es un muestreo probabilístico bietápico estratificado de elementos. En la primera etapa, se selecciona una muestra estratificada de UPM con probabilidad proporcional al tamaño (PPT), donde la medida de tamaño de cada UPM está dada por el total de viviendas particulares ocupadas. Posteriormente, se enlista la totalidad de viviendas que conforman cada UPM para, en una segunda etapa, seleccionar aleatoriamente un total fijo de 12 viviendas en cada UPM seleccionada

2.1.2 Dominios de estudio

El dominio de diseño de la encuesta consiste en una subpoblación que puede ser identificada en el marco de muestreo y, por tanto, puede ser manejada independientemente en el tamaño de muestra, procedimientos de muestreo y demás.

El levantamiento de información de la ENIGHUR está planificado para 13 periodos. Cada periodo está conformado de 4 semanas, por lo que, el levantamiento está previsto para un total de 52 semanas y la muestra está distribuida espacial y temporalmente respetando los periodos indicados.

La ENIGHUR para cumplir con su objetivo considera dos grupos de dominios, los cuales son:

- ❖ Grupo 1: considera las 24 provincias del Ecuador, las cuales constituyen el primer nivel de desagregación geográfico según el “Clasificador Geográfico Estadístico 2021 - Esquema de codificación de la División Política Administrativa del país”.
- ❖ Grupo 2: considera 9 cantones auto representados que son de interés debido a su comportamiento económico.

Los cantones auto representados se dividen en dos grupos:

- Cantón completo (urbano y rural): Cuenca, Ambato, Quito y Loja.
- Únicamente la parte urbana del cantón: Machala, Santo Domingo de los Tsáchilas, Guayaquil, Manta, Esmeraldas.

Tabla 1. Dominios de estudio

N	Provincia	Cantón
1	Resto Azuay	Cuenca
2	Bolívar	
3	Cañar	
4	Carchi	
5	Cotopaxi	
6	Chimborazo	
7	Resto El Oro	Machala
8	Resto Esmeraldas	Esmeraldas
9	Resto Guayas	Guayaquil
10	Imbabura	
11	Resto Loja	Loja

12	Los Ríos	
13	Resto Manabí	Manta
14	Morona Santiago	
15	Napo	
16	Pastaza	
17	Resto Pichincha	Quito (DMQ)
18	Resto Tungurahua	Ambato
19	Zamora Chinchipe	
20	Galápagos	
21	Sucumbíos	
22	Orellana	
23	Santa Elena	
24	Resto Santo Domingo de los Tsáchilas	Santo Domingo

Fuente: Gestión de Diseño Muestral - DINEM

2.1.3 Marco Muestral

El Marco de muestreo a utilizar para la ENIGHUR es el Marco Maestro de Muestreo para encuestas de hogares del INEC (MMM), el cual se presenta como una lista organizada en forma de base de datos que contiene las viviendas registradas en el precenso del VIII Censo de Población y VII de Vivienda del Ecuador (CPV – 2022).

2.1.4 Variables de diseño

El tamaño de la muestra se calcula en función a su variable de diseño gastos en alimentos (Monetarios).

2.1.5 Tamaño muestral

Para poder calcular los tamaños de muestra por estimación de medias, se generó el gasto promedio por hogar para el rubro de gastos en ALIMENTOS - MONETARIOS, con el fin de obtener las estimaciones a partir de la información de la encuesta ENIGHUR 2012. Dichas estimaciones serán el insumo para el cálculo del tamaño de muestra.

Para calcular un tamaño de muestra que responda a todos los requerimientos planteados se sigue el siguiente esquema:

1. Se define como variable de diseño los gastos ALIMENTOS – MONETARIOS

2. Se calcula un tamaño de muestra independiente para cada dominio de diseño.
3. El tamaño de muestra anterior se transforma a tamaño de muestra de UPM.
4. El tamaño de muestra por UPM se distribuye para todos los estratos presentes en el marco. De ser necesario, se aumenta el tamaño de muestra de UPM por estrato a 2, para asegurar el cálculo de los estimadores de varianza resultantes de la encuesta.

2.1.5.1 Tamaño de muestra de viviendas

El tamaño de muestra de la ENIGHUR se obtuvo mediante la siguiente ecuación.

$$n_k = \frac{n_{0,k}}{1 + \frac{n_{0,k}}{N_k}} * \frac{1}{1 - tnr}$$

Donde:

$$n_0 = \frac{Z_{1-\frac{\alpha}{2}}^2 S^2}{\varepsilon^2 \mu^2}$$

$$S^2 = \sigma^2 DEFF$$

n_k : El tamaño de la muestra de viviendas para el dominio de estudio k .

N_k : El tamaño de la población en el dominio k .

μ_k : La media poblacional de la variable de interés en el dominio k .

σ_k^2 : La varianza poblacional de la variable de interés en el dominio k .

$DEFF_k$: El efecto del diseño muestral en el dominio k .

α : Nivel de confianza ($\alpha = 0,95$).

tnr : Tasa de no respuesta.

2.1.5.2 Efecto de diseño (DEFF)

Con el objetivo de poder tener un acercamiento a los parámetros reales relacionados a los efectos de diseño para cada uno de los dominios de estudio, se elaboró una reconstrucción de los estratos en la base de cobertura ENIGHUR-2012. Dichos estratos se elaboraron considerando la parte geográfica subdivida en áreas urbanas (estrato 1) y rurales (estrato 2). Mientras que para las provincias que cuentan con cantones auto-representadas se consideró el cantón (estrato 3), el resto del área urbana (estrato 1) y el resto del área rural (estrato 2).

2.1.5.3 Nivel de confianza

Con una puntuación estándar z_{α}^2 se define como el valor “z” para un nivel de confiabilidad deseada en la distribución normal con probabilidad igual, que corresponde a 1,96.

2.1.5.4 Margen de error

Para el cálculo del margen de error relativo (MER) en cada uno de los dominios de estudio se tomó como referencia el margen de error obtenido en las estimaciones de la ENIGHUR-2012. De este modo, se definió un MER entre el 3% y 4% para los cantones auto-representadas; para los dominios ubicados en la Amazonia, un MER entre 3% y 10% y un valor de 14% para Morona Santiago, siendo este el máximo MER. Para el resto de provincias, el MER varía entre el 2,5% y 10%.

2.1.5.5 Tasa de no respondientes (TNR)

Esta tasa permite conocer la proporción de viviendas elegibles no efectivas para el total de viviendas visitadas. En este sentido, la tasa de no respuesta para la encuesta fue calculada a partir de la base de cobertura de la ENIGHUR-2012 y finalmente definida como el promedio de las TNR's de cada dominio de estudio, cuyo resultado se aproxima al 20%, resultado aproximado de la TNR promedio por provincia.

$$TNR = 1 - \frac{N^{\circ} \text{ viviendas efectivas}}{N^{\circ} \text{ de viviendas seleccionadas}}$$

2.1.5.6 Tamaño de muestra de UPM

Es necesario calcular el número de UPM que deben ser seleccionadas en la primera etapa de muestreo a partir de la relación:

$$n_{UPMk} = \frac{\text{Número de viviendas calculadas}}{\text{Carga técnica operativa}}$$

Para la presente operación estadística se definió operativamente que el número de viviendas a investigarse por UPM será doce (12).

2.1.5.7 Asignación de la muestra

La muestra es distribuida proporcionalmente por cada estrato dentro de cada dominio de estudio.



$$n_{ki} = n_{UPMk} * \frac{N_{ki}}{N_k}$$

Donde:

n_{ki} : Tamaño de la muestra para el estrato i del dominio de estudio k .

n_{UPMk} : Tamaño de la muestra para el dominio de estudio k .

N_{ki} : Total de viviendas ocupadas en el estrato i del dominio de estudio k .

N_k : Total de viviendas ocupadas en el dominio de estudio k .

En los estratos, donde $n_{ki} = 1$ se aproxima a mínimo muestral $n_{ki} = 2$.

Es importante mencionar que, antes de la distribución proporcional por cada estrato en cada dominio de estudio, se realiza un ajuste a los tamaños calculados. Este tipo de ajuste responde estrictamente a solventar requerimientos referentes a la parte operativa de la encuesta, por lo que, se busca mantener un tamaño muestral por zonal para garantizar que el número de encuestadores y supervisores estén acordes a lo planificado y presupuestado por el equipo encargado, sin afectar la calidad del diseño muestral.

2.1.5.6 Distribución espacial y temporal de la muestra

Con el objetivo de definir claramente el levantamiento de información de cada UPM en cada dominio de estudio, se realiza una distribución en espacio y tiempo para definir los periodos y semanas en la que se debe visitar cada UPM.

Se considera levantar la misma cantidad de UPM en cada periodo. La asignación de la muestra queda definida con un total de 264 UPM por periodo, y en cada dominio de estudio se busca mantener una cantidad similar de UPM a levantar entre cada periodo. Esto garantiza que la encuesta no se concentre en determinados periodos y esté distribuida equitativamente a lo largo de toda la línea de tiempo que dura la operación estadística.

2.1.5.7 Resultados del tamaño muestral

Una vez realizado el proceso de asignación y ajuste muestral mencionados en el aparatado anterior, los resultados del tamaño muestral para cada dominio son los siguientes:

Tabla 2. Tamaño muestral por dominio

Dominio	Nombre del dominio	Tamaño (UPM)
01	Azuay	69
02	Bolívar	95
03	Cañar	92
04	Carchi	83
05	Cotopaxi	97
06	Chimborazo	104
07	El Oro	65
08	Esmeraldas	77
09	Guayas	126
10	Imbabura	95
11	Loja	74
12	Los Ríos	111
13	Manabí	150
14	Morona Santiago	98
15	Napo	66
16	Pastaza	83
17	Pichincha	144
18	Tungurahua	81
19	Zamora Chinchipe	94
20	Galápagos	104
21	Sucumbíos	76
22	Orellana	67
23	Santo Domingo de los Tsachilas	27
24	Santa Elena	60
33	Quito	164
42	Guayaquil	210
30	Cuenca	131
40	Machala	124
34	Ambato	135
41	Esmeraldas Ciudad	141
44	Santo Domingo	120
43	Manta	132

32	Loja Ciudad	137
-----------	-------------	-----

Fuente: Gestión de Diseño Muestral - DINEM

2.1.6 Selección de la muestra

Para la selección de la muestra, los sectores censales se definieron como las unidades primarias de muestreo (UPM); en tanto, que las unidades de segunda etapa están representadas por las viviendas. Las UPM'S fueron seleccionadas con probabilidad proporcional al tamaño (PPT) utilizando como medida de tamaño el total de las viviendas reportadas en la cartografía censal. Por su parte, las viviendas se seleccionaron asignando la misma probabilidad de selección a todas las viviendas que conforman una unidad primaria.

2.2 Control de cobertura de campo y muestral

La cobertura hace referencia a la información obtenida en campo y proporcionada por los informantes de las viviendas seleccionadas. En el presente documento, se analiza la cobertura de forma agregada de los 13 periodos de investigación. En este sentido, la cobertura a nivel de UPM, vivienda y población objetivo toma en cuenta tanto el resultado de la entrevista como la condición de ocupación de las viviendas visitadas, las cuales se clasifican en dos grandes grupos: elegibilidad conocida y elegibilidad desconocida, con la siguiente subclasificación:

❖ **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?

❖ **Elegibilidad desconocida**

- Elegibilidad desconocida (ED)

* Nadie en casa

En la Tabla 3 podemos apreciar el detalle de la condición de ocupación y el resultado de la entrevista. Esta información constituye el insumo para el cálculo de las tasas de conformidad. La información se presenta desagregada a nivel de coordinación zonal y condición de ocupación.

Tabla 3: Condición de ocupación y resultado de entrevista

CONDICIÓN OCUPACIÓN	ADM. C. CAMPO	CENTRO	LITORAL	SUR
Completa	7852	8048	10053	9296
Nadie en casa	318	166	493	319
Otra razón	42	10	180	88
Rechazo	827	247	1142	592
Vivienda convertida en negocio	11	1	18	8
Vivienda desocupada	168	78	264	142
Vivienda en construcción	7	9	30	13
Vivienda inhabitada o destruida	5	6	20	5
Vivienda temporal	130	171	280	145

Fuente: Gestión de Diseño Muestral – DINEM

2.3 Construcción de factores de expansión

2.3.1 Probabilidad de primera etapa

En la primera etapa, se seleccionan UPM con una probabilidad proporcional al tamaño definido por el número de viviendas ocupadas. Por lo que 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:

$ppe_{i,h}$: es la probabilidad de inclusión de primera etapa de la UPM i en el estrato h .

m_h : es el número de UPM seleccionadas en el estrato h .

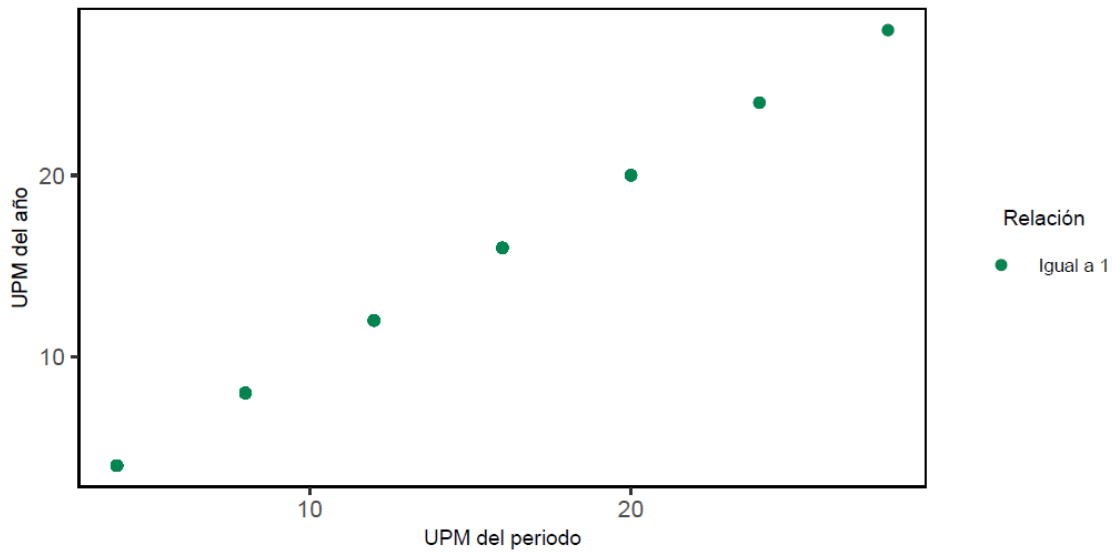
N_i : el número de viviendas de la UPM i según el marco de muestreo.

N_h : el número de viviendas en el estrato h según el marco de muestreo.

Es importante mencionar que la probabilidad de inclusión se calcula con el número de UPM planificadas para cada mes, por este motivo la probabilidad de primera etapa se debe ajustar tomando en cuenta la muestra diligenciada en el periodo de referencia, es decir:

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

Figura 1. Distribución de UPM anual por periodo



Fuente: Gestión de Diseño Muestral - DINEM

Nota: Por facilidad en la notación, se mantendrá $ppe_{i,h}$ y m_h para referirnos a la probabilidad de inclusión de primera etapa y el número de UPM planificadas del periodo p.

2.3.2 Ajuste de cobertura de primera etapa

Debido a diferentes motivos de no investigar una UPM, se debe realizar un ajuste en el que se tomen en cuenta únicamente las UPM que fueron investigadas, el ajuste sería:

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

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

2.3.3 Probabilidad de segunda etapa

La selección de segunda etapa se realiza sobre el conjunto de viviendas particulares ocupadas según el precenso 2022, siendo:

$$pse_i = \frac{k_i}{N_i}$$

donde:

pse_i : probabilidad de segunda etapa de las viviendas en la UPM i .

k_i : número de viviendas particulares ocupadas seleccionadas en la UPM i .

N_i : número total de viviendas particulares ocupadas en la UPM i .

2.3.4 Factor de expansión teórico

El factor de expansión teórico es el inverso del producto de las probabilidades de selección:

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

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

2.3.5 Factor de expansión ajustado por cobertura de primera etapa

El primer ajuste que se realiza para los factores de expansión es el de cobertura de primera etapa, el factor ajustado sería:

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

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

2.3.6 Ajuste por cobertura de viviendas

Al momento de la entrevista en campo, cada vivienda muestreada se clasifica 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, se pueden tomar diferentes decisiones cómo manejar las unidades en ED.

Tanto los marcos como las muestras seleccionadas a partir de ellos pueden contener unidades cuya elegibilidad no puede ser determinada. Además, 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.

2.3.6.1 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 aborda a través de métodos simples, en parte porque se sabe poco de los casos con elegibilidad desconocida y en parte porque la falta de respuesta se considera un problema más serio que requiere más 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 ENIGHUR se realizará 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

2.3.6.2 Ajuste por no respuesta

El ajuste por no respuesta puede ser sencillo o muy complejo, dependiendo de cuánto se conoce de los no respondientes. Para la ENIGHUR, dado que la información disponible de los no respondientes es muy limitada, por lo tanto, se utilizará un enfoque similar al de 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:

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

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:

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

El ajuste de no respuesta considera que las viviendas no respondientes tienen características similares a las respondientes. En consecuencia, el estimador de no elegibilidad sobre el marco de muestreo es del 13,6%, lo que significa que aproximadamente una de cada 8 viviendas identificadas como viviendas particulares ocupadas no lo son.

Al analizar la relación entre los factores d_{3j} y d_{1j} en 105 estratos se tiene un incremento menor a 1,1, mientras que hay una relación entre 1,1 y 1,2 en 18 y en 9 estratos hay un incremento mayor a 1,2.

3. Estimaciones y errores muestrales

Los criterios de calidad que se mencionan en esta sección se pueden tomar en cuenta para determinar si una estadística se debe considerar precisa y confiable según la recomendación de Gutiérrez et al., (2020).

3.1 Intervalos de confianza

La precisión de una estadística se debe estudiar a la luz del intervalo de confianza generado por la medida de probabilidad asociada al diseño de muestreo de la encuesta (Heeringa, West & Berglund, 2010)

$$(\hat{\theta} - t_{0,975,gl} * se(\hat{\theta}), \hat{\theta} + t_{0,975,gl} * se(\hat{\theta}))$$

En donde $\hat{\theta}$ es un estimador por muestreo para el parámetro de interés θ , $t_{0,975,gl}$ es el percentil 0,975 de una distribución t-student con gl grados de libertad, que están daos por la resta entre el número de UPM seleccionadas menos el número de estratos de muestreo considerados y $se(\hat{\theta})$ es el error estándar de la estimación, definido por la raíz cuadrada de la varianza del estimador; es decir:

$$se(\hat{\theta}) = \sqrt{\widehat{Var}(\hat{\theta})}$$

3.2 Estimador de calibración

La proyección poblacional utilizada corresponde al punto medio del periodo de levantamiento. Dado que la encuesta abarca un total de 52 semanas, se considera la semana 26 como referencia para la estimación poblacional, la cual corresponde al mes de mayo de 2025. El estimador de calibración se define de la siguiente manera:

Considere una función de distancia G con argumentos $x = \frac{w_k}{d_k}$ con las siguientes propiedades:

- ❖ G es positiva y estrictamente convexa,
- ❖ $G(1) = G'(1) = 0$, y
- ❖ $G''(1) = 1$

Bajo esta definición $G\left(\frac{w_k}{d_k}\right)$ mide la distancia de los factores de expansión originales d_k a los nuevos factores de expansión w_k , siendo $\sum_s d_k G\left(\frac{w_k}{d_k}\right)$ la medida de distancia para toda la muestra s . Por lo tanto, el problema de optimización es:

Minimizar $\sum_s d_k G\left(\frac{w_k}{d_k}\right) - \lambda'(\sum_s w_k x_k - \sum_U x_k)$, donde U hace referencia a la población, $x_k = (x_{k_1}, \dots, x_{k_j} \dots x_{k_J})'$ es un vector de valores auxiliares y $\lambda = (\lambda_1, \dots, \lambda_j \dots \lambda_J)'$ es un J - vector de multiplicadores de Lagrange. Para calcular los nuevos factores de expansión, primero se debe determinar el valor de λ , el cual se obtiene resolviendo las ecuaciones de calibración:

$$\sum_s d_k F(x'_k \lambda) x_k = \sum_U x_k$$

Luego, el estimador de calibración queda definido por:

$$\hat{t}_{yc} = \sum_s w_k y_k = \sum_s d_k F(x'_k \lambda) y_k$$

para los (y_k, x_k) datos observados en la muestra ($k \in s$) y una función de distancia G dada. Nótese que $w_k = d_k F(x'_k \lambda)$ es el nuevo factor de expansión calibrado.

En este contexto, notaremos $g_k = F(x'_k \lambda)$.

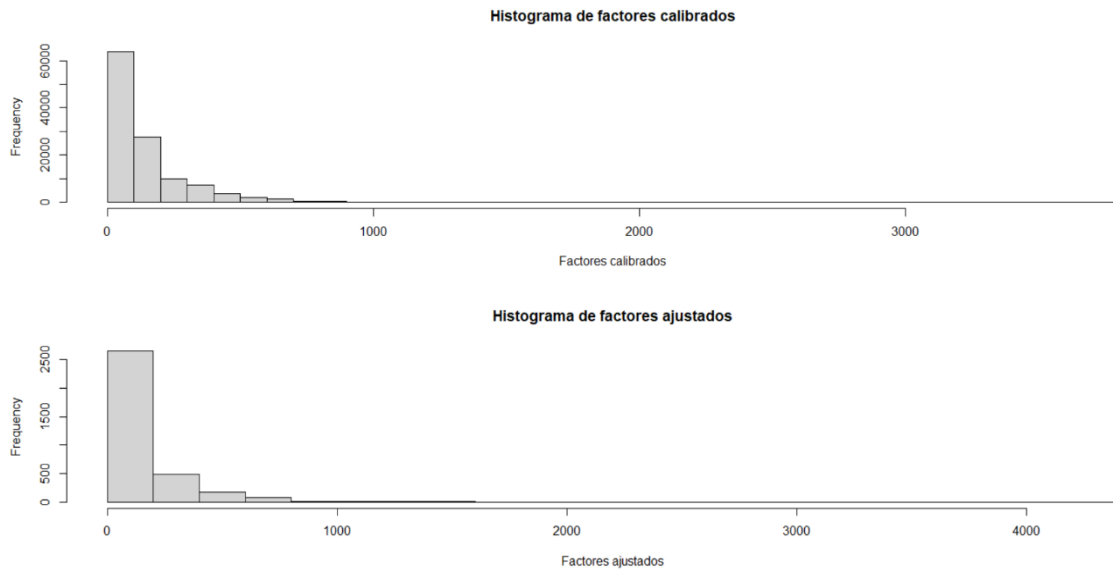
Cabe mencionar que, en este proceso de construcción de factores de expansión para la ENIGHUR, se calibra los pesos de muestreo recortados, por tanto, los ponderadores calibrados son calculados con la siguiente expresión:

$$w_{k_c} = w_{k_r} * g_k$$

Donde w_{k_c} son los factores de expansión calibrados, mientras que w_{k_r} son los pesos de muestreo recortados y g_k los pesos de calibración.

Cabe señalar que la calibración de los factores de expansión se realizó a nivel de UPM, es decir, todos los individuos de una UPM presentan un mismo ponderador independientemente de sus características demográficas como dominio y área.

Figura 2. Distribución de factores calibrados y ajustados



Fuente: Gestión de Diseño Muestral - DINEM

3.3 Recorte de factores de expansión extremos

Debido a los ajustes por no respuesta presente para algunas poblaciones objetivo, la variabilidad de los factores de expansión se ha visto incrementada. Tal y como se expone en Potter (1990), la variación extrema en los factores de expansión puede resultar en varianzas muestrales excesivamente grandes.

Para solventar este inconveniente se han desarrollado algunos procedimientos para limitar o reducir en tamaño de factores de expansión extremos, algunos de los cuales están descritos en Potter (1990, 1993). Para ENIGHUR 2024 se utiliza el método de recorte y redistribución de factores de expansión más común, descrito en CEPAL (2021), el cual propone el siguiente método iterativo:

- (1) Fijar una cota superior para los factores de expansión.
- (2) Todo factor de expansión mayor a la cota superior se fija en el valor de la cota, definiendo:

$$f_{exptrim} = \begin{cases} U & \text{si } w_k \geq U \\ w_k & \text{si } w_k < U \end{cases}$$

Donde $f_{exptrim}$ es el conjunto de los factores de expansión recortados.

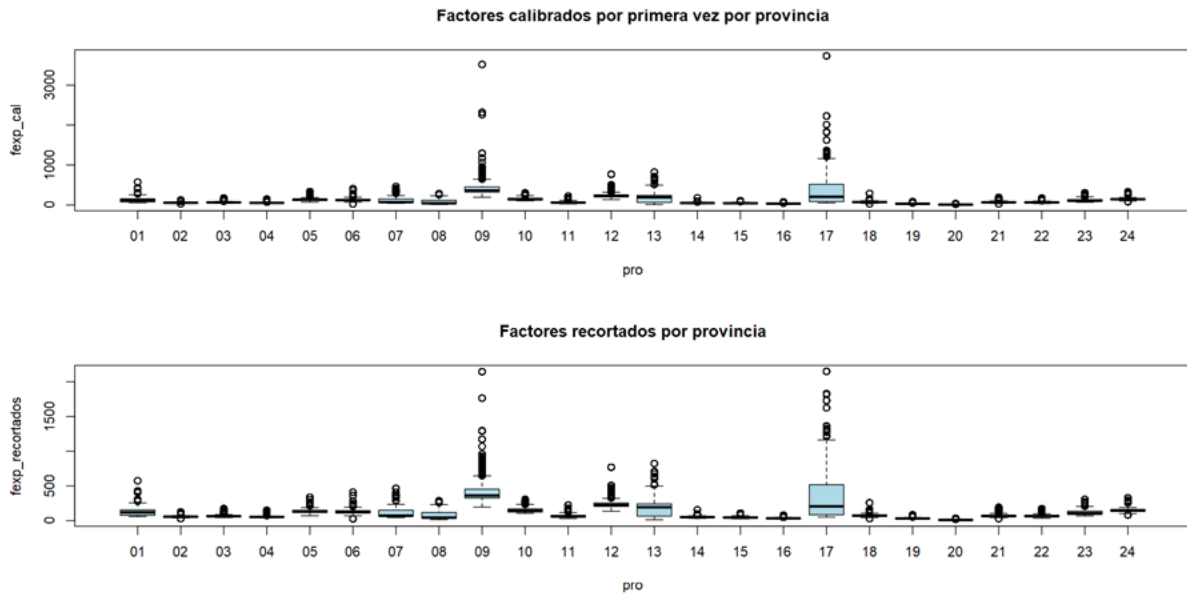
- (3) Calcular $K = \sum |f_{exp} - f_{exptrim}|$.
- (4) Distribuir K equitativamente entre todos los pesos no recortados.
- (5) Repetir los pasos (1)-(4) hasta que $K = 0$.

Puesto que ENIGHUR 2024 presenta factores de expansión extremos únicamente en



la cola derecha de su distribución, se fija U en 3.5 veces la mediana para realizar el recorte a nivel de estrato.

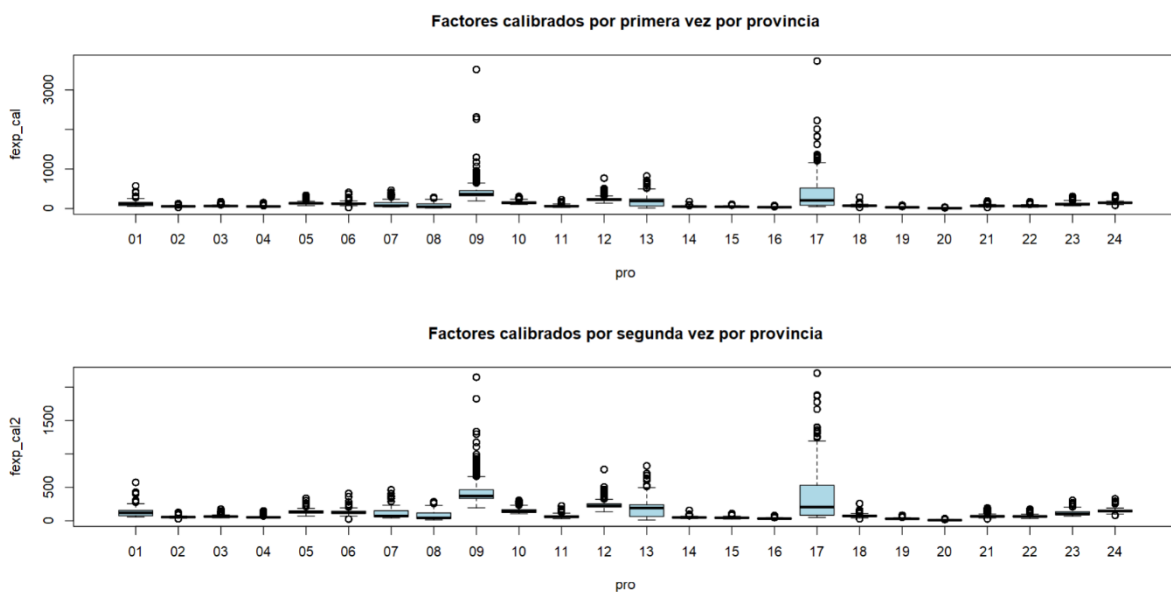
Figura 3. Factores calibrados y recortados por provincia



Fuente: Gestión de Diseño Muestral - DINEM

Lo cual al obtener los recortes realizados se calibra por segunda vez los factores con el fin de proporcionar la cantidad de proyección por cada id de calibración correspondiente.

Figura 4. Factores calibrados por primera y segunda vez



Fuente: Gestión de Diseño Muestral – DINEM



A continuación, se presenta la comparación gráfica entre los factores de expansión recortados y calibrados por segunda vez. Además, se conoce cuantos estratos fueron recortados a nivel de hogar.

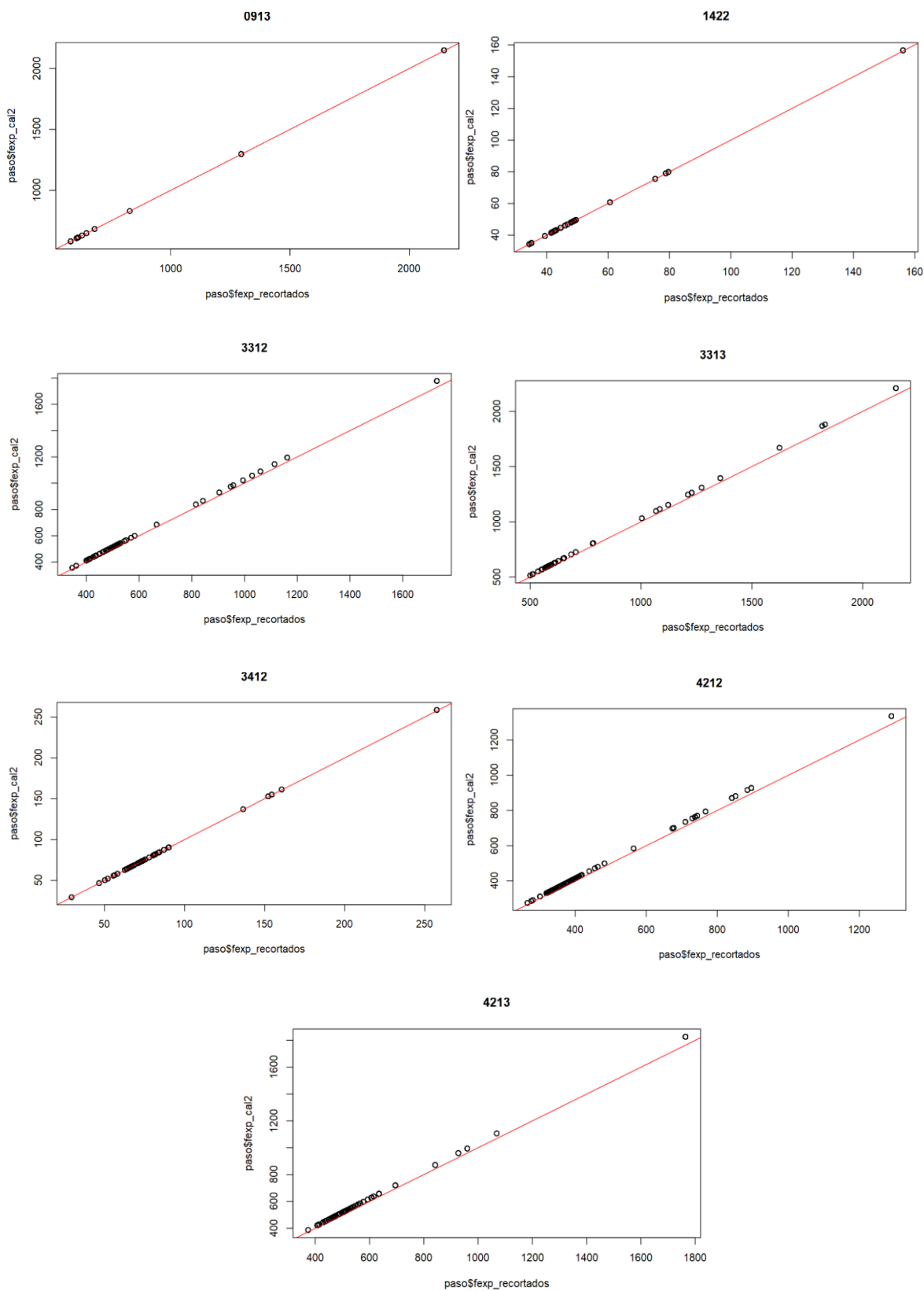
Tabla 4. Recorte de estratos

Dominio	Estrato	Recortados
09	0913	9
14	1422	8
20	2099	10
33	3312	16
33	3313	9
34	3412	11
42	4212	10
42	4213	8

Fuente: Gestión de Diseño Muestral - DINEM



Figura 5. Factores calibrados posterior al recorte



Fuente: Gestión de Diseño Muestral - DINEM



Referencias

- [Potter, B. N. (1990)]. *Adjusting survey weights to reduce extreme variance in expansion factors. Journal of Survey Methodology.*
- [Potter, B. N. (1993)]. *Methods for limiting extreme sampling weights in population surveys. Survey Research Methods.*
- [Heeringa, S. G., West, B. T., & Berglund, P. A. (2010)]. *Applied survey data analysis. CRC Press.*
- [Comisión Económica para América Latina y el Caribe (CEPAL). (2021)]. *Encuesta de hogares en América Latina: Diseño y análisis estadístico*

50 ^{Medio siglo} contando historias

INEC

Buenas cifras,
mejores vidas



@InecEcuador



@ecuadorencifras



@ecuadorencifras



INECEcuador