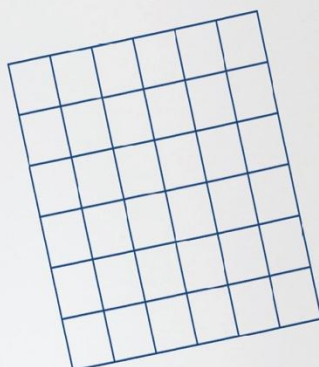


**Diseño Muestral de
la Encuesta
Nacional de Salud
y Nutrición 2018**

METODOLOGÍA



Dirección de Infraestructura Estadística y Muestreo

Metodología de Diseño Muestral de la Encuesta Nacional de Salud y Nutrición (ENSANUT)

2018

Quito - Ecuador, 2019

Dirección

Dirección de Infraestructura Estadística y Muestreo

Unidad

Gestión de Diseño Muestral

Elaborado por:

Javier Núñez

Revisado por:

Christian Garcés

Aprobado por:

Sebastián Carvajal

Tabla de contenido

1	MARCO DE MUESTREO DE LA ENCUESTA NACIONAL DE SALUD Y NUTRICIÓN ENSANUT 2018.....	6
1.1	OBJETIVOS DE LA ENCUESTA Y POBLACIÓN OBJETIVO	6
1.2	DOMINIOS DE LA ENCUESTA.....	6
1.3	MARCO DE MUESTREO	7
1.3.1	<i>Evaluación del marco de muestreo</i>	<i>8</i>
1.4	ESTRATIFICACIÓN.....	9
1.5	TAMAÑO DE MUESTRA.....	9
1.5.1	<i>Tamaño muestral y errores de muestreo.....</i>	<i>10</i>
1.5.2	<i>Determinación del tamaño muestral</i>	<i>11</i>
1.5.3	<i>Muestra “a tomar” por UPM.....</i>	<i>13</i>
1.5.4	<i>Tamaño de muestra de UPM</i>	<i>15</i>
1.5.5	<i>Distribución de la muestra</i>	<i>16</i>
1.6	DISEÑO MUESTRAL DE LA ENSANUT 2018.....	16
1.7	ENLISTAMIENTO DE HOGARES	17
1.7.1	<i>Lectura y tratamiento inicial de la base de datos de enlistamiento.....</i>	<i>18</i>
1.7.2	<i>Generación de bases a nivel de hogar, vivienda y manzana</i>	<i>18</i>
1.8	SELECCIÓN DE VIVIENDAS	20
1.8.1	<i>Muestra inicial ENSANUT 2018.....</i>	<i>21</i>
1.8.2	<i>Muestra complementaria ENSANUT 2018.....</i>	<i>22</i>
1.9	SELECCIÓN DE LAS DIFERENTES POBLACIONES OBJETIVO.....	23
1.10	FACTORES DE EXPANSIÓN DE ENSANUT 2018.....	23
1.10.1	<i>Bases de publicación.....</i>	<i>23</i>
1.10.2	<i>Esquema de ponderación de ENSANUT 2018.....</i>	<i>24</i>
1.10.3	<i>Cobertura.....</i>	<i>26</i>
1.10.4	<i>Probabilidades de inclusión para ENSANUT 2018</i>	<i>29</i>
1.10.5	<i>Factores de expansión ENSANUT 2018</i>	<i>36</i>

1.10.6	Recorte de factores de expansión extremos	37
1.10.7	Calibración de los factores de expansión ENSANUT 2018	39
1.11	ESTIMACIÓN DE CARACTERÍSTICAS	41
1.12	ERROR DE MUESTREO.....	41
1.13	MÉTODOS DE ESTIMACIÓN DE ERRORES PARA DISEÑOS MUESTRALES COMPLEJOS	42
2	BIBLIOGRAFÍA.....	45

Lista de tablas

Tabla 1	Procedencia de las viviendas del MMM por proceso de Actualización cartográfica.	7
Tabla 2	Número de Unidades Primarias de Muestreo por provincia y área.	8
Tabla 3	Prevalencia de las variables de diseño usadas en la ENSANUT 2018.	11
Tabla 4	Prevalencia de las variables de diseño usadas en la ENSANUT 2018.	12
Tabla 5	Tamaño de muestra de personas por provincia de la ENSANUT 2018.	12
Tabla 6	Tamaño de muestra de UPM asociado a los niños menores de 5 años por dominio de diseño para la ENSANUT 2018.	14
Tabla 7	Tamaño de muestra de UPM de la ENSANUT 2018.	15
Tabla 8	Resumen provincial del enlistamiento de ENSANUT 2018.	19
Tabla 9	Fechas de entrega de la muestra de ENSANUT 2018.	22
Tabla 10	Tamaño muestra complementaria por provincia, ENSANUT 2018.	23
Tabla 11	Bases de datos y población objetivo de ENSANUT 2018.	25
Tabla 12	Cobertura de ENSANUT 2018 por provincia y grupo de selección.	27
Tabla 13	Variables requeridas para la declaración del diseño muestral	44

1 MARCO DE MUESTREO DE LA ENCUESTA NACIONAL DE SALUD Y NUTRICIÓN ENSANUT 2018

1.1 Objetivos de la encuesta y población objetivo

Generar indicadores sobre los principales problemas y la situación de salud de la población ecuatoriana con el fin de evaluar y generar las políticas públicas en temas de salud y nutrición.

Las poblaciones objetivo de la ENSANUT 2018 son:

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

1.2 Dominios de la encuesta

En general, uno de los objetivos principales de una encuesta es comparar sus resultados entre diferentes desagregaciones, como por ejemplo urbano 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.

Por un lado, un dominio de diseño consiste en una subpoblación que puede identificarse en el marco de muestreo y, por lo tanto, puede ser manejada independientemente en el tamaño de muestra, procedimientos de muestreo y demás. 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.

La ENSANUT 2018 considera dos grupos de dominios de diseño geográfico en su implementación, 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 2010 - Esquema de codificación de la División Político Administrativa del país”.
- Grupo 2: considera las 9 zonas de planificación determinadas por la Secretaría Nacional de Planificación y Desarrollo - SENPLADES, desde las cuales se coordina estratégicamente a las entidades del sector público, gestionando su planificación en su área de jurisdicción (SENPLADES, 2017).

Para que los estimadores calculados a partir de la ENSANUT 2018 sean confiables a nivel de dominio de diseño, es necesario asegurarse de que el número de casos sea

suficiente en cada uno de ellos, considerando los niveles de precisión deseados para los mismos. Esto se consigue calculando un tamaño de muestra adecuado para cada una de las poblaciones objetivo por dominio de diseño.

Por último, en el Ecuador continental existen territorios denominados Zonas no Delimitadas (ZND) cuya jurisdicción, al momento de delimitar el Marco de Muestreo, estaba en proceso de regulación. Puesto que su extensión territorial es pequeña, no se consideran como dominios de diseño, sin embargo su inclusión es necesaria para garantizar la representatividad nacional de todos los indicadores de ENSANUT 2018, siendo susceptibles de selección todos los territorios del Ecuador.

1.3 Marco de muestreo

El Marco de muestreo a utilizar para la ENSANUT 2018 es el Marco Maestro de Muestreo para encuestas de hogares del INEC (MMM), el cual se presenta como una lista organizada en forma de base de datos que contiene las viviendas registradas en el precenso del VII Censo de Población y VI de Vivienda del Ecuador (CPV - 2010) junto con las actualizaciones cartográficas realizadas en el periodo 2011-2018. En la **Tabla 1** se presenta la proporción de viviendas catalogadas como ocupadas según precenso CPV-2010 y proceso de actualización.

Tabla 1 Procedencia de las viviendas del MMM por proceso de Actualización cartográfica.

Actualización cartográfica	Año de actualización	Total de viviendas	Porcentaje
CPVCENEC	2010	2685474	72,2%
ECV20132014	2013	212549	5,7%
CENSOGAL15	2015	8520	0,2%
PROYEC2015	2015	52933	1,4%
ACTENEMDU	2015	611248	16,4%
ACTUAL2017	2017	147594	4,0%
Total MMM	2018	3718318	100,0%

Fuente: Gestión de Diseño Muestral - DINEM

El Marco de muestreo codifica las Unidades Primarias de Muestreo (UPM) en función a la División Político Administrativa definida en el Clasificador Geográfico Estadístico - DPA, permitiéndonos identificar adecuadamente los dominios de diseño geográficos definidos para ENSANUT 2018. Además, contiene variables que clasifican las UPM en función al estrato al que pertenecen. En la **Tabla 2** se presenta el total de UPM por provincia y área presentes en el MMM.

Tabla 2 Número de Unidades Primarias de Muestreo por provincia y área.

Provincia	Rural	Urbano
Azuay	2051	3015
Bolívar	801	478
Cañar	820	775
Carchi	518	770
Cotopaxi	1785	1022
Chimborazo	1618	1578
El Oro	780	3879
Esmeraldas	1133	2259
Guayas	2847	24127
Imbabura	834	2206
Loja	1401	1793
Los Ríos	1871	3588
Manabí	2939	6132
Morona Santiago	604	371
Napo	409	292
Pastaza	288	329
Pichincha	2166	19003
Tungurahua	1962	1939
Zamora Chinchipe	403	266
Galápagos	176	172
Sucumbíos	604	646
Orellana	476	506
Santo Domingo De Los Tsáchilas	489	2659
Santa Elena	120	1905
Zonas En Estudio	209	0
Total UPM	27304	79710

Fuente: Elaboración propia

Cabe resaltar que la definición de urbano/rural establecida en el Marco Maestro de Muestreo de encuestas dirigidas a hogares del INEC es la de 2000 personas o más, el cual es un estándar normalizado a nivel de la Comunidad Andina de Naciones (CAN) para asegurar comparabilidad internacional. Para más información al respecto, referirse al documento (CAN, 2000).

1.3.1 Evaluación del marco de muestreo

El MMM nos ofrece cobertura geográfica de todo el territorio ecuatoriano, identificando adecuadamente los dominios geográficos menores, permitiendo una distribución apropiada de la muestra. Sin embargo, tal y como se observa en la **Tabla 1**, la mayoría de la información correspondiente a la condición de ocupación de las viviendas presentes en el marco de muestreo proviene del precenso CPV-2010, lo cual dificultaría ubicar las diferentes poblaciones objetivo de la encuesta,

en especial los niños menores a 5 años de edad, puesto que su promedio por vivienda es el más bajo entre todas las poblaciones objetivo.

Puesto que ya han transcurrido 8 años desde la implementación del MMM, seleccionar la muestra de ENSANUT 2018 sin un proceso de actualización previo traería consigo graves problemas de cobertura para la mayoría de las poblaciones objetivo. Sin embargo, una actualización completa es restrictiva económicamente, por lo que se decide actualizarlo de manera parcial.

La actualización parcial del MMM consiste en enlistar un número de UPM adecuadamente distribuidas y presentes en todos los estratos de selección de la encuesta. Para tal fin, se empieza por calcular el tamaño de muestra necesario a nivel de UPM por estrato, para luego seleccionar las UPM a ser enlistadas.

1.4 Estratificación

La estratificación se refiere a la subdivisión de una población determinada en subconjuntos con características propias. Esta acción se lleva a cabo como una etapa previa a la selección de la muestra, utilizando variables que aportan información para todas las unidades de la población.

Los objetivos del proceso de estratificación son los siguientes:

- Agrupar a las unidades del marco en grupos de acuerdo a un conjunto de características socioeconómicas.
- Formar grupos de UPM homogéneos en su interior y heterogéneos entre ellos.
- Mejorar la eficiencia del diseño muestral mediante la disminución de la varianza de los estimadores.

Para más información acerca de la estratificación del MMM, remitirse al documento (INEC, 2019).

1.5 Tamaño de muestra

Como se menciona en las secciones 1.1 y 1.2 del presente documento, los requerimientos de información de la ENSANUT 2018 están planteados para diferentes poblaciones objetivo y dominios de diseño. Para calcular un tamaño de muestra que responda a todos los requerimientos planteados se sigue el siguiente esquema:

1. Se definen las variables de diseño de la encuesta. En el marco de la Subcomisión de Salud y Nutrición, se definieron las siguientes como variables de diseño:
 - a. Desnutrición crónica en menores de 5 años.
 - b. Sobrepeso y obesidad en niños entre 5 y 11 años.
 - c. Proporción de mujeres en edad fértil (MEF).

2. Se calcula un tamaño de muestra independiente para cada combinación de población objetivo y dominio de diseño generado.
3. El tamaño de muestra anterior se transforma a tamaño de muestra de UPM.
4. Se compara el tamaño de muestra de UPM de cada combinación de población objetivo y dominio de diseño, siendo el mayor de estos el tamaño de muestra de UPM por dominio de diseño.
5. 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.

A continuación se detalla el desarrollo en cada etapa.

1.5.1 Tamaño muestral y errores de muestreo

Los estimadores de una encuesta por muestreo están afectados por dos tipos de errores: *errores muestrales* y *errores no muestrales*. Los errores muestrales están presentes debido a que se observa únicamente una parte de la población objetivo, además, están relacionados al tamaño de muestra y variabilidad entre las unidades de muestreo. Por otro lado, los errores no muestrales resultan de problemas al momento de recolectar, procesar, analizar o difundir los datos de la encuesta.

El reto al momento de decidir el tamaño muestral es balancear las demandas de análisis de información, precisión deseada de los estimadores y la capacidad de implementación y presupuesto de la organización a cargo de la encuesta. En el marco de la Subcomisión de Salud y Nutrición, se propusieron, discutieron y acordaron las siguientes como variables de diseño de la ENSANUT 2018:

- Desnutrición crónica en menores de 5 años.
- Sobrepeso y obesidad en niños entre 5 y 11 años.
- Proporción de mujeres en edad fértil (MEF) de 12 a 49 años.

La prevalencia de las variables de diseño se calculó a partir de la Encuesta de Condiciones de Vida 2013 - 2014 (ECV 2013 - 2014), la cual es una encuesta que permitió obtener indicadores sobre los niveles de vida y el bienestar de la población relacionando varios factores como educación, salud, pobreza e inequidad para la aplicación de política pública, espectros de información similares a ENSANUT 2018. En la **Tabla 3** se muestra las prevalencias calculadas.

Tabla 3 Prevalencia de las variables de diseño usadas en la ENSANUT 2018.

Provincia	Desnutrición crónica	Sobrepeso y obesidad	Proporción de MEF
Azuay	0,32	0,33	0,30
Bolívar	0,40	0,27	0,27
Cañar	0,31	0,34	0,30
Carchi	0,35	0,38	0,28
Cotopaxi	0,34	0,23	0,29
Chimborazo	0,44	0,21	0,28
El Oro	0,13	0,34	0,30
Esmeraldas	0,22	0,26	0,28
Guayas	0,17	0,35	0,29
Imbabura	0,29	0,39	0,29
Loja	0,28	0,28	0,27
Los Ríos	0,17	0,25	0,27
Manabí	0,19	0,29	0,29
Morona Santiago	0,44	0,29	0,27
Napo	0,30	0,18	0,29
Pastaza	0,35	0,22	0,29
Pichincha	0,26	0,35	0,30
Tungurahua	0,33	0,31	0,30
Zamora Chinchipe	0,31	0,25	0,28
Galápagos	0,18	0,49	0,32
Sucumbíos	0,26	0,23	0,28
Orellana	0,27	0,25	0,28
Santo Domingo De Los Tsáchilas	0,16	0,31	0,29
Santa Elena	0,41	0,31	0,28

Fuente: ECV 2013-2014

1.5.2 Determinación del tamaño muestral

El tamaño de muestra de la ENSANUT 2018 se determinó fijando un límite en la precisión para el error relativo deseado para cada variable de diseño considerada, calculando el tamaño de muestra mediante la siguiente ecuación:

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

donde:

- n = Tamaño de muestra resultante.
 p = Prevalencia del estimador.
 e_{rel} = Error relativo asociado a la prevalencia p .
 z = Nivel de confianza.
 N = Tamaño de la población objetivo del estimador.
 $deff$ = Efecto de diseño.
 tnr = Tasa de no respuesta esperada.

Se determinó el tamaño de la población objetivo asociada a cada estimador (N) a partir de las proyecciones poblacionales del INEC al mes de Junio 2016. En la **Tabla 4** se encuentra el valor asignado al resto de parámetros usados para el cálculo del tamaño muestral.

Tabla 4 Prevalencia de las variables de diseño usadas en la ENSANUT 2018.

Parámetro	Valor
Error relativo (e_{rel})	0.15
Nivel de confianza (z)	0.95
Efecto de diseño ($deff$)	2.00
Tasa de no respuesta (tnr)	0.05

Fuente: Elaboración propia

Se fijó el efecto de diseño en 2 debido a que no se pudo aproximar su valor desde la ENSANUT 2011. Además, puesto que la ECV 2013 - 2014 tenía un diseño muestral distinto a lo requerido por ENSANUT 2018, los efectos de diseño de las dos encuestas son distintos, razón por la cual su utilización no hubiera sido adecuada.

La tasa de no respuesta se fijó en 5% puesto que el levantamiento en campo consideraba un proceso de actualización previo. En la **Tabla 5** se observa el tamaño de muestra asociado a cada población objetivo considerando los dominios y variables de diseño de la ENSANUT 2018.

Tabla 5 Tamaño de muestra de personas por provincia de la ENSANUT 2018.

Provincia	Niños menores de 5 años	Niños de 5 a 11 años	Mujeres de 12 a 49 años
Azuay	778	731	734
Bolívar	525	956	827
Cañar	795	671	733
Carchi	657	580	781
Cotopaxi	695	1176	759
Chimborazo	461	1348	804
El Oro	2292	677	744
Esmeraldas	1285	1032	785
Guayas	1786	656	767
Imbabura	860	553	778
Loja	938	904	861

Los Ríos	1705	1057	841
Manabí	1518	873	774
Morona Santiago	455	872	832
Napo	830	1465	767
Pastaza	650	1138	745
Pichincha	1028	681	744
Tungurahua	715	779	729
Zamora Chinchipe	778	996	800
Galápagos	1342	338	591
Sucumbíos	1024	1150	786
Orellana	950	1043	800
Santo Domingo De Los Tsáchilas	1821	798	750
Santa Elena	521	778	794

Fuente: Elaboración propia

1.5.3 Muestra “a tomar” por UPM

El tamaño de muestra expuesto en la sección 1.5.2 hace referencia al total de individuos por población objetivo a ser entrevistados en cada dominio de diseño. El siguiente paso es convertirlo a un tamaño de muestra de UPM, definiendo el número de individuos “a tomar” por UPM de cada población objetivo.

Sin embargo, puesto que las poblaciones objetivo son muy distintas, el tamaño de muestra de UPM variaba de manera significativa para cada una de ellas. Para solventar este inconveniente se realizaron varias simulaciones de selección a partir de la información del CPV - 2010 con la finalidad de identificar a la población más difícil de encontrar, siendo esta los niños menores de 5 años. Este análisis se hizo por provincia, debido a diferencias en las características demográficas entre ellas.

Con estos antecedentes, para calcular el tamaño de muestra de UPM asociado a los niños menores de 5 años, se fijó un número de viviendas con niños menores a 5 años por UPM de manera diferenciada para cada dominio de diseño considerando: sus características demográficas y la capacidad operativa del INEC, la cual fijaba en 18 el número máximo de viviendas a visitar por UPM.

Como se observa en la **Tabla 5**, para ciertos dominios de diseño, el número de niños menores a 5 años esperados en muestra es menor al necesario. Esto se debe a que el número de viviendas con niños menores a 5 años “a tomar” por UPM se disminuyó para no afectar de sobremanera el efecto de diseño de los demás estimadores a calcular desde la ENSANUT 2018, resguardando el cálculo adecuado de los errores asociados a los mismos (varianza). Entre dominios de diseño, la diferencia fluctúa entre el 3 y 15 por ciento.

Tabla 6 Tamaño de muestra de UPM asociado a los niños menores de 5 años por dominio de diseño para la ENSANUT 2018.

Provincia	Niños menores de 5 años	Viviendas con niños menores a 5 años por UPM	Tamaño de muestra UPM	Niños esperados en muestra
Azuay	778	9,8	75	736
Bolívar	525	7,1	83	591
Cañar	795	10,7	71	758
Carchi	657	9,4	78	734
Cotopaxi	695	8,0	102	820
Chimborazo	461	7,3	114	833
El Oro	2292	11,0	192	2112
Esmeraldas	1285	12,4	108	1341
Guayas	1786	8,5	209	1774
Imbabura	860	9,5	77	733
Loja	938	11,1	83	924
Los Ríos	1705	11,6	144	1670
Manabí	1518	11,5	133	1532
Morona Santiago	455	10,0	84	836
Napo	830	9,3	126	1166
Pastaza	650	8,6	96	821
Pichincha	1028	9,7	104	1011
Tungurahua	715	9,7	74	719
Zamora Chinchipe	778	10,2	84	856
Galápagos	1342	4,9	115	563
Sucumbíos	1024	10,8	96	1038
Orellana	950	12,0	89	1072
Santo Domingo De Los Tsáchilas	1821	11,3	156	1760
Santa Elena	521	8,5	78	665
Zonas no Delimitadas	---	---	12	132

Fuente: Elaboración propia

A continuación, se fijó el tamaño de muestra "a tomar" por UPM para las poblaciones objetivo *niños entre 5 a 11 años y mujeres de 12 a 49 años* en 12 observaciones, el cual era suficiente considerando el tamaño de muestra de UPM calculado para niños menores de 5 años.

Con respecto a las Zonas no Delimitadas, se toma un tamaño 12 UPM considerando los mínimos muestrales y su distribución en territorio. En cuanto, la provincia de Galápagos es un dominio de diseño especial, puesto que por su tamaño poblacional y sus características demográficas (mortalidad, fecundidad y migración) diferentes a las del Ecuador continental, se hace necesario aplica un

diseño especial para encontrar un balance entre lo necesario por muestra y stock poblacional disponible en la provincia.

1.5.4 Tamaño de muestra de UPM

En la **Tabla 7** se muestra el tamaño de muestra de UPM para la ENSANUT 2018, considerando, los valores “a tomar” de cada población objetivo. Para cada dominio de diseño, el tamaño de muestra de UPM es el máximo del tamaño de muestra de UPM asociado a cada población objetivo. Así, el tamaño de muestra de UPM de la ENSANUT 2018 es de 2583 UPM, correspondiente a la suma del tamaño de muestra de UPM de todos los dominios de diseño.

Tabla 7 Tamaño de muestra de UPM de la ENSANUT 2018.

Provincia	Niños menores de 5 años	Niños de 5 a 11 años	Mujeres de 12 a 49 años
Azuay	75	61	70
Bolívar	83	80	79
Cañar	71	56	70
Carchi	78	49	75
Cotopaxi	102	98	73
Chimborazo	114	113	77
El Oro	192	57	71
Esmeraldas	108	86	75
Guayas	209	55	73
Imbabura	77	47	75
Loja	83	76	82
Los Ríos	144	89	81
Manabí	133	73	74
Morona Santiago	84	73	80
Napo	126	123	74
Pastaza	96	95	71
Pichincha	104	57	71
Tungurahua	74	65	70
Zamora Chinchipe	84	83	77
Galápagos	115	29	57
Sucumbíos	96	96	75
Orellana	89	87	77
Santo Domingo De Los Tsáchilas	156	67	72
Santa Elena	78	65	76
Zonas no Delimitadas	12	12	12

Fuente: Elaboración propia

1.5.5 Distribución de la muestra

Considerando que la ENSANUT 2018 es una encuesta que cubre un amplio espectro de información, la muestra se distribuye de manera proporcional entre los estratos de cada dominio de diseño. La distribución proporcional depende únicamente del tamaño de los estratos, los cuales están determinados por el total viviendas particulares ocupadas en cada uno. Así,

$$n_{i_h} = n_i \cdot \frac{N_{i_h}}{\sum_{h=1}^{H_i} N_{i_h}}$$

donde:

n_{i_h} = Muestra para el estrato h del dominio de diseño i .

n_i = Tamaño de muestra del dominio de diseño i .

N_{i_h} = Tamaño del estrato h del dominio de diseño i .

H_i = Número de estratos en el dominio de diseño i .

Luego de la repartición, si a algún estrato le corresponde menos de 2 UPM, este número se ajusta a 2 para asegurar el cálculo apropiado de errores muestrales.

1.6 Diseño muestral de la ENSANUT 2018

El diseño muestral implementado en la ENSANUT 2018 es un muestreo probabilístico bietápico estratificado de elementos.

En la primera etapa, se selecciona una muestra estratificada de UPM con probabilidad proporcional al tamaño (PPT), donde la medida de tamaño de cada UPM está dada por el total de viviendas particulares ocupadas. Luego, se enlista la totalidad de cada UPM seleccionada, con la finalidad de corregir los errores existentes en el marco y proveer de un marco de selección de viviendas adecuado.

En la segunda etapa de muestreo, se selecciona de manera aleatoria un número variable de viviendas, siendo 18 el promedio de viviendas por UPM. El número variable de viviendas se debe a la necesidad de seleccionar viviendas con niños menores a 5 años. Todos los hogares de las viviendas seleccionadas son investigados, en los que se identifican los grupos poblacionales de interés tales como niños menores a 5 años, las mujeres entre 12 y 49 años, los hombres mayores a 12 años, personas entre 5 y 17 años, entre otros, con la finalidad de diligenciar los cuestionarios específicos a cada población objetivo, considerando:

- Todos los niños menores a 5 años son entrevistados.
- Todas las mujeres entre 10 y 49 años son entrevistadas.
- Un hombre mayor o igual a 12 años es seleccionado y entrevistado.
- Una persona entre 5 a 17 años es seleccionada y entrevistada.
- Una persona mayor o igual a 10 años es seleccionada y entrevistada.

- Un niño menor a 5 años es seleccionado y entrevistado (Desarrollo Infantil).

En la sección 1.7 se detalla el proceso de enlistamiento realizado y en la sección 1.8 se detalla el proceso de selección de viviendas.

1.7 Enlistamiento de hogares

Después de seleccionar las UPM para la encuesta, se realiza un enlistamiento completo previo a la selección de viviendas. El enlistamiento consistió en visitar todas las viviendas que conforman cada una de las UPM con tres grandes objetivos: conocer su condición de ocupación, la composición de los hogares que la habitan y generar material cartográfico de apoyo para el levantamiento de información.

Para la implementación del enlistamiento se entregó un listado de las manzanas, o localidades en el caso de ser una UPM dispersa, que integran cada UPM en función a la información del marco de muestreo, junto con el total de viviendas y total de viviendas ocupadas de cada manzana o localidad¹. Esto con la finalidad de estimar costos y planificar la logística de campo del enlistamiento. A continuación se detallan las preguntas incluidas en el formulario de enlistamiento.

- Ubicación geográfica y muestral: identificación estadística de la UPM a enlistar.
- Identificación de la vivienda: dirección de la vivienda (localidad, barrio, calles, y demás).
- Datos del informante: persona que provee la información.
- Condición de ocupación de la vivienda: ocupada, desocupada, en construcción, temporal, colectiva, destruida.
- Resultado de la entrevista: completa, rechazo, nadie en casa, etc.
- Fecha de enlistamiento.
- Datos de los miembros del hogar: incluye
 - o nombres completos,
 - o sexo,
 - o edad,
 - o fecha de nacimiento,
 - o parentesco con el jefe de hogar,
- Información auxiliar: incluye limitaciones físicas y la fecha de probable parto en caso de estar embarazada.

Como parte del material cartográfico de apoyo, se prepararon y actualizaron croquis de las UPM a participar para poder ubicar las viviendas seleccionadas y realizar adecuadamente los procesos de revisita y supervisión del levantamiento de información. Entre los formularios de ENSANUT - 2018 se incluye el formulario de

¹ Información calculada a partir del precenso de viviendas del CPV 2010.

enlistamiento. Para más información del procedimiento de enlistamiento en campo dirigirse a la sección 3.1 de la Metodología de ENSANUT - 2018.

El enlistamiento se realizó desde el 03 de septiembre de 2018 hasta el 31 de octubre de 2018. Para procurar que la diferencia entre la fecha de enlistamiento y la fecha de entrevista sea la menor posible, se planificó realizarlo por jornadas, recorriendo sucesivamente todas las UPM seleccionadas de cada provincia, generando cortes sucesivos de la base de datos de enlistamiento, y así seleccionar sucesivamente la muestra de viviendas para las provincias enlistadas.

Sobre cada corte de la base de enlistamiento se realizaron controles y validaciones, con la finalidad de generar el marco de muestreo de viviendas a partir de la base de enlistamiento. A continuación se detallan los diferentes procesos realizados.

1.7.1 Lectura y tratamiento inicial de la base de datos de enlistamiento

En esta primera etapa, se realizó un tratamiento general sobre las bases de datos y la generación de variables de apoyo para la selección de viviendas.

- Verificación de formato de las variables de enlistamiento.
- Códigos correspondientes al Clasificador Geográfico Estadístico - DPA.
- Creación de identificadores para control y validación del enlistamiento (id_sector, id_man, id_man_div, id_viv, id_hog)².
- Verificación del sexo de cada persona enlistada.
- Verificación de la fecha de nacimiento de cada persona enlistada. La edad en años meses y días de cada persona se calcula como la diferencia entre la fecha de enlistamiento y la fecha de nacimiento de cada persona. Esta información es crucial para identificar las viviendas que tienen niños menores a 5 años, MEF, hombres mayores a 12 años y demás poblaciones de interés.
- Tratamiento de las variables de ubicación de las viviendas enlistadas (calle principal, calle secundaria, número de la vivienda, ciudadela, lote, número de departamento, piso y patio).
- Tratamiento de los nombres de las personas y su relación con el jefe de hogar.

1.7.2 Generación de bases a nivel de hogar, vivienda y manzana

A partir de la base de personas trabajada, se verifica que:

- No existan registros e identificadores duplicados en la base de personas.
- De existir duplicados se envía a revisar y corregir la información.

² Id_sector: Identificador de sector censal.
Id_man: identificador de manzana.
Id_man_div: identificación de manzana dividida.
Id_viv: identificador de vivienda.
Id_hog: identificador de hogar.

- Puesto que la unidad secundaria de muestreo es la vivienda, se generan agregados a este nivel con la siguiente información:
 - o Condición de ocupación de la vivienda.
 - o Número de hogares en la vivienda.
 - o Número total de personas.
 - o Número total de mujeres entre 10 y 11 años de edad.
 - o Número total de mujeres entre 12 y 49 años de edad.
 - o Número total de personas entre 5 y 17 años de edad.
 - o Número total de niños entre 13 y 42 meses de edad.
 - o Número total de niños entre 43 y 59 meses de edad.
 - o Número total de niños menores a 5 años de edad.
- Se generan agregados a nivel de manzana con la finalidad de comparar el total de viviendas y total de viviendas ocupadas de las manzanas enviadas.

Además, puesto que a lo largo del tiempo se crean nuevas manzanas y/o localidades o las existentes sufren cambios (por ejemplo, la división de una manzana o dos localidades se juntan para formar una nueva), es necesario que estos cambios estén contrastados y justificados, con la finalidad de asegurar que el marco de selección de viviendas sea completo, sin intrusos y sin duplicidad³. En la **Tabla 8** se resume el resultado del enlistamiento de ENSANUT 2018, posterior a todas las validaciones realizadas.

Tabla 8 Resumen provincial del enlistamiento de ENSANUT 2018.

Provincia	Manzanas/ localidades a enlistar	Manzanas/ localidades enlistadas	Mujeres de 12 a 49 años	Niños menores a 5 años	Hombres de 12 o más años	Cobertura
Azuay	393	400	8171	2020	9998	100,0%
Bolívar	629	658	8376	2090	11188	100,0%
Cañar	722	740	9179	2650	10807	100,0%
Carchi	826	873	9589	2393	13157	100,0%
Cotopaxi	605	629	13151	3523	16290	100,0%
Chimborazo	767	796	12458	3056	15746	100,0%
El Oro	1996	2104	24562	6739	32761	100,0%
Esmeraldas	917	965	16902	5518	20515	100,0%
Guayas	1392	1430	24226	7032	31993	98,8%
Imbabura	680	722	10688	2503	13725	100,0%
Loja	824	846	9731	2489	13009	100,0%
Los Ríos	1336	1422	21244	6344	27825	100,0%
Manabí	1226	1383	20313	5788	27533	100,0%
Morona Santiago	907	965	10863	4321	12671	100,0%

³ En total se registraron 449 uniones y 1785 divisiones de manzanas/localidades.

Napo	1071	1134	15854	5393	19539	100,0%
Pastaza	685	696	11728	4014	13937	97,2%
Pichincha	538	573	12385	2748	15923	100,0%
Tungurahua	411	431	8163	2055	10736	100,0%
Zamora Chinchiipe	966	1030	10535	3225	12893	100,0%
Galápagos	205	215	2691	662	3512	100,0%
Sucumbíos	1011	1086	12962	4063	16213	100,0%
Orellana	1098	1146	13591	5073	16396	100,0%
Santo Domingo De Los Tsáchilas	1363	1480	22820	6668	28256	100,0%
Santa Elena	1439	1610	15422	4712	20073	100,0%
Zonas no Delimitadas	103	112	1416	460	1951	100,0%
Total Nacional	22110	23446	327020	95539	416647	99,84%

Fuente: Elaboración propia

A nivel nacional, la cobertura del enlistamiento fue del **99,84%**, alcanzando el 100% en 22 de las 24 provincias (23 dominios geográficos si contamos las Zonas no Delimitadas) a excepción de las provincias de:

- *Guayas*: rechazo completo de 6 manzanas con urbanizaciones o conjuntos habitacionales; 11 manzanas sin información por desalojo en la zona.
- *Pastaza*: no se trabajaron 19 localidades por falta de recursos logísticos para vuelos de ingreso a la zona a enlistar.

A pesar de la falta de completitud en estas dos provincias, el faltante no afecta significativamente para el desarrollo de la encuesta, teniendo suficientes UPM/viviendas por estrato a ser seleccionadas.

Por último, cabe resaltar que el proceso de enlistamiento puso especial atención en identificar a los niños menores de 5 años, puesto que su presencia o no en una vivienda condiciona su probabilidad de selección.

1.8 Selección de viviendas

Las viviendas cuya condición de ocupación en el enlistamiento de ENSANUT 2018 fue *ocupada* (la cual incluye las viviendas rechazadas y nadie en casa) fueron consideradas para la selección de viviendas de segunda etapa. En primer lugar, las viviendas fueron clasificadas en dos conjuntos (estratos implícitos dentro de la UPM):

- Conjunto L_C : vivienda con al menos un niño menor a 5 años, y
- Conjunto L_S : vivienda sin un niño menor a 5 años.

Puesto que no se conoce la composición del hogar de las viviendas rechazadas o nadie en casa, estas fueron ubicadas en el conjunto *viviendas sin niño menor a 5 años*. Luego, dentro de cada conjunto se seleccionaron viviendas de manera aleatoria sin reposición en función a los tamaños calculados en la sección 1.5.3. Por ejemplo, en la provincia de Cotopaxi se seleccionaron 8 viviendas con niños menores de 5 años y 10 viviendas sin niños menores a 5 años por UPM.

En el caso de que una UPM tuviera menos del número necesario de viviendas con niños menores de 5 años, se seleccionaban todas las existentes y las faltantes se tomaban de otra UPM del mismo estrato. Siguiendo el ejemplo anterior, supongamos que la UPM "1" tenía únicamente 6 viviendas con niños menores de 5 años, entonces el total de viviendas seleccionadas en ella fue de 16 (6 con niños y 10 sin niños menores a 5 años). Luego, en la UPM "2" se seleccionaron 20 viviendas en total (10 con niños y 10 sin niños menores de 5 años). Esto sucedió en el 14% de las UPM, donde el 97.67% de las UPM tuvieron entre 16 y 20 viviendas seleccionadas.

Para una próxima ronda de ENSANUT, se sugiere considerar este procedimiento sujeto a la condición de que el tamaño total de viviendas por UPM sea el mismo entre todas las UPM. Para una próxima ronda de ENSANUT, se sugiere considerar este procedimiento sujeto a la condición de que el tamaño total de viviendas por UPM sea el mismo entre todas las UPM.

Cabe resaltar que, a diferencia de otras encuestas dirigidas a hogares que organiza el INEC, ENSANUT 2018 no contempla una metodología de reemplazos a nivel de vivienda.

1.8.1 Muestra inicial ENSANUT 2018

La muestra a nivel de viviendas fue entregada en formularios de Muestra y Cobertura (MyC) específicamente desarrollados para ENSANUT 2018, con la finalidad de identificar adecuadamente los formularios a diligenciar en cada una de las viviendas a visitar por UPM. Entre los formularios disponibles se incluye un ejemplo de formulario MyC usado en ENSANUT 2018.

Entre la información contenida en los formularios MyC consta:

- Variables de ubicación cartográfica y geográfica de la vivienda.
- Listado completo de hogares y personas de las viviendas seleccionadas.
- Nombre, sexo, edad en años y edad en meses de los miembros del hogar.

Puesto que el enlistamiento se realizó por jornadas, los formularios MyC fueron entregados acorde al cierre de las jornadas de enlistamiento y la validación realizada sobre su información. En la **Tabla 9** se detalla las fechas de entrega de los formularios MyC.

Tabla 9 Fechas de entrega de la muestra de ENSANUT 2018.

Provincia	Fecha de corte de enlistamiento	Fecha de entrega de la muestra
Azuay	15/11/2018	15/11/2018
Cotopaxi		
Manabí		
Cañar		
Carchi		
Imbabura		
Tungurahua		
Santo Domingo de los Tsáchilas	16/11/2018	19/11/2018
Chimborazo	20/11/2018	20/11/2018
Morona Santiago		
Zamora Chinchipe		
Zonas no Delimitadas		21/11/2018
Pichincha		
Galápagos		
Los Ríos	22/11/2018	22/11/2018
Esmeraldas	28/11/2018	29/11/2018
Guayas		
Loja		
Napo		
Sucumbíos		
Orellana		
Santa Elena		
Bolívar	05/12/2018	06/12/2018
El Oro		
Pastaza		

Fuente: Elaboración propia

1.8.2 Muestra complementaria ENSANUT 2018

La ENSANUT 2018 fue planificada para ser levantada en dos meses del 2018, sin embargo, como el periodo de tiempo era corto para la ejecución de todas las actividades, se solicitó la certificación de rubros de manera plurianual para así poder culminar las actividades de recolección en 2019. El INEC diligenció aproximadamente el 80% de la muestra total entre el 15 de noviembre de 2018 y el 6 de enero del 2019. El 20% restante fue levantado en los meses de junio y julio de 2019.

En el periodo enero - mayo 2019 se realizó un análisis de cobertura muestral con la información levantada, identificando ciertos estratos a nivel nacional con un rendimiento muy por debajo de lo esperado (5%).

Por tal razón, se decidió incrementar el tamaño de muestra de UPM en 8, con la finalidad de salvaguardar la cobertura y cálculo adecuado de los errores muestrales a nivel de estrato. Estas UPM fueron seleccionadas del enlistamiento de ENSANUT 2018. En la **Tabla 10** se muestra el incremento de muestra a nivel de UPM por provincia.

Tabla 10 Tamaño muestra complementaria por provincia, ENSANUT 2018.

Provincia	Número de UPM
Azuay	3
Guayas	1
Pichincha	1
Tungurahua	3

Fuente: Elaboración propia

Con este incremento, el tamaño de muestra de ENSANUT 2018 es de 46638 viviendas, repartidas en 2591 UPM.

1.9 Selección de las diferentes poblaciones objetivo

Todos los hogares de las viviendas seleccionadas son investigados. Además, en cada hogar se identifican los grupos poblacionales de interés (tales como niños menores a 5 años), con la finalidad de diligenciar los cuestionarios específicos a cada población objetivo, considerando:

- Todos los niños menores a 5 años son entrevistados.
- Todas las mujeres entre 10 y 49 años son entrevistadas.
- Un hombre mayor o igual a 12 años es seleccionado y entrevistado.
- Una persona entre 5 a 17 años es seleccionada y entrevistada.
- Una persona mayor o igual a 10 años es seleccionada y entrevistada.
- Un niño menor a 5 años es seleccionado y entrevistado (Desarrollo Infantil).

El método de selección de personas utilizado es el de *cumpleaños más próximo*, el cual consiste en ordenar a las personas seleccionables en función a su día y mes de nacimiento desde el más cercano al más lejano, considerando como fecha de referencia la fecha de la entrevista. Si el día y mes de cumpleaños de una persona coincide con el día y mes de la entrevista, esta persona se ubica al final de la lista ordenada. La persona seleccionada es la primera de la lista ordenada.

1.10 Factores de expansión de Ensanut 2018

1.10.1 Bases de publicación

La información levantada en campo de ENSANUT 2018 se agrupa en 9 bases de datos, dependiendo de la población objetivo a la que está dirigida, teniendo cada una su propio factor de expansión. A continuación se enlistan las diferentes bases de datos a ser publicadas para la ENSANUT 2018.

- **Base 1 – f1_personas:** en esta base de datos encontrarán información referente a las características sociodemográficas de cada miembro del hogar, así como también información de educación, trabajo, ingresos, salud, peso, talla, entre otra, todos los miembros del hogar.
- **Base 2 – f1_hogar:** en esta base de datos encontrarán información referente a las características de las viviendas, hogares y la información correspondiente a Seguridad Alimentaria. Adicionalmente se encontrara variables de desagregación a nivel de hogar.
- **Base 3 – f1_etiquetado:** En esta base de datos encontrarán información referente a las características de las viviendas, hogares e información correspondiente a Seguridad Alimentaria. Adicionalmente se encontrara variables de desagregación a nivel de hogar.
- **Base 4 – f2_mef:** en esta base de datos encontrarán información de mujeres de 10 a 49 años de edad referente a historia de embarazos, salud materna, actividad sexual, planificación familiar, nupcialidad y enfermedades de transmisión sexual.
- **Base 5 – f2_lactancia:** en esta base de datos encontrarán información referente a lactancia materna en el último hijo nacido y actualmente vivo, menor a 3 años.
- **Base 6 – f2_salud_niñez:** en esta base de datos encontrarán información referente a los 3 últimos hijos nacidos vivos de la MEF, menores de 5 años.
- **Base 7 – f3_ssr_hombres:** en esta base de datos encontrarán información de un hombre seleccionado del hogar mayor a 12 años de edad, y contiene temas como, actividad sexual y salud reproductiva, planificación familiar y enfermedades de transmisión sexual.
- **Base 8 – f4_fact_riesgo:** en esta base de datos encontrarán información de un niño de 5 a 17 años, con información referente a salud oral, actividad física, alimentación y nutrición, consumo de alcohol y tabaco.
- **Base 9 – f5_desarrollo_infantil:** En esta base de datos encontrarán información de un niño menor de 5 años del hogar, con temas como: programas de primera infancia, oportunidad de juego, disciplina infantil, desarrollo, aprendizaje y educación, lenguaje, motricidad e Inventario Home.

1.10.2 Esquema de ponderación de ENSANUT 2018

Como se explica en la Sección 1.6, la selección de tercera etapa es distinta para cada población objetivo, haciendo necesario una ponderación de tercera etapa

para algunas de ellas. En la **Tabla 11** se detalla la población objetivo para cada una de las bases ENSANUT 2018, y si existe o no una selección de tercera etapa o no.

Tabla 11 Bases de datos y población objetivo de ENSANUT 2018.

Base de datos	Población objetivo	Selección de tercera etapa
f1_personas	Todas las personas del hogar	No
f1_hogar	Jefe de hogar	No aplica
f1_etiquetado	Una persona mayor o igual a 10 años	Si
f2_mef	Todas las mujeres entre 10 y 49 años	No
f2_lactancia	Último hijo nacido y actualmente vivo menor a 3 años de cada MEF	No aplica
f2_salud_niñez	Últimos 3 hijos nacidos vivos de la MEF, menores de 5 años	No aplica
f3_ssr_hombres	Un hombre mayor o igual a 12 años	Si
f4_fact_riesgo	Una persona de 5 a 17 años	Si
f5_desarrollo_infantil	Un niño menor de 5 años	Si

Fuente: Elaboración propia

El procedimiento de ponderación general para la Encuesta Nacional de Salud y Nutrición ENSANUT 2018 requiere dos o tres etapas, dependiendo de la población objetivo. La primera etapa consiste en calcular una ponderación de UPM. De ser necesario, se aplica un ajuste por no respuesta a este nivel.

La segunda etapa consiste en calcular una ponderación de vivienda por UPM. Puesto que las viviendas fueron seleccionadas independientemente dentro de cada estrato implícito de cada UPM (vivienda con o sin un niño menor de 5 años), se tienen dos ponderaciones de vivienda distintas por UPM. De ser necesario, se utiliza un ajuste por no respuesta y no elegibilidad independiente para cada conjunto de selección a nivel de estrato.

El producto de la ponderación de primera y segunda etapa da como resultado el factor de expansión de hogar donde, de no haber tercera etapa de selección, todas las personas del mismo hogar tienen el mismo factor de expansión.

Por otro lado, de haber una tercera etapa de selección, se calcula una ponderación de persona. De ser necesario, se aplica un ajuste por no respuesta a nivel de conjunto de selección por estrato a la ponderación de persona.

De donde, el factor de expansión de persona correspondiente al producto de las ponderaciones intermedias: la ponderación de primera etapa (UPM), la ponderación de segunda etapa (vivienda) y la ponderación de tercera etapa (persona).

1.10.3 Cobertura

La información de cobertura de una encuesta nos permite evaluar su rendimiento muestral. Un seguimiento oportuno y adecuado de la misma permite generar estrategias para reducir los errores no muestrales presentes en la etapa de recolección de datos⁴.

Esta información se convierte en el insumo principal al momento de ajustar los factores de expansión de una encuesta en función a la relación entre entrevistas efectivas y entrevistas realizadas para cada una de las poblaciones objetivo investigadas.

La ENSANUT 2018 es una encuesta con tres etapas de selección, a continuación se presenta el reporte de cobertura para cada una de ellas, junto con la cobertura para cada una de las poblaciones objetivo a nivel nacional por conjunto de selección.

1.10.3.1 Cobertura de UPM

En la ENSANUT 2018, la cobertura a nivel de UPM fue del 99,96%, es decir 2590 de las 2591 UPM seleccionadas tuvieron al menos una vivienda efectiva. La UPM no efectiva se ubicó en la provincia de Morona Santiago.

1.10.3.2 Cobertura de viviendas

En la ENSANUT 2018 se seleccionaron y visitaron 46638 viviendas, de las cuales 43097 fueron efectivas, para una cobertura nacional del 92,41%. Puesto que la cobertura debe evaluarse de manera acorde a la selección de viviendas, en la **Tabla 12** se muestra la cobertura de la ENSANUT 2018 a nivel de provincia y grupo de selección.

⁴ Este error no muestral se conoce como error debido a la no respuesta.

Tabla 12 Cobertura de ENSANUT 2018 por provincia y grupo de selección.

Provincia	Conjunto de selección	Viviendas completas	Viviendas rechazo	Viviendas nadie en casa	Viviendas no elegibles
Azúay	L_C	580	14	11	37
	L_S	654	23	12	73
Bolívar	L_C	487	7	4	28
	L_S	890	1	9	68
Cañar	L_C	606	0	0	36
	L_S	570	1	0	65
Carchi	L_C	627	8	5	9
	L_S	706	6	5	38
Cotopaxi	L_C	661	9	3	36
	L_S	1016	20	5	86
Chimborazo	L_C	657	14	2	40
	L_S	1197	26	5	108
El Oro	L_C	1693	8	11	72
	L_S	1589	11	9	63
Esmeraldas	L_C	1002	15	2	74
	L_S	755	11	20	65
Guayas	L_C	1384	12	5	48
	L_S	2129	37	14	150
Imbabura	L_C	589	11	13	23
	L_S	660	27	20	43
Loja	L_C	755	4	0	12
	L_S	698	2	2	21
Los Ríos	L_C	1377	0	2	30
	L_S	1158	2	0	23
Manabí	L_C	1253	3	2	26
	L_S	1082	3	2	23
Morona Santiago	L_C	565	3	8	63
	L_S	759	2	4	108
Napó	L_C	691	3	5	49
	L_S	1372	18	23	107
Pastaza	L_C	589	4	4	35
	L_S	1000	19	1	76
Pichincha	L_C	803	50	23	32
	L_S	817	69	31	55
Tungurahua	L_C	591	20	6	36
	L_S	636	31	11	55
Zamora Chinchipe	L_C	645	1	1	64
	L_S	708	0	5	88
Galápagos	L_C	363	5	7	26
	L_S	1440	25	48	156

Sucumbíos	L_C	810	3	13	63
	L_S	764	3	20	52
Orellana	L_C	813	4	3	43
	L_S	683	3	4	49
Santo Domingo De Los Tsáchilas	L_C	1480	6	2	42
	L_S	1220	9	2	47
Santa Elena	L_C	518	0	1	8
	L_S	843	3	3	28
Zonas no Delimitadas	L_C	106	0	0	3
	L_S	106	0	0	1
Total Nacional	L_C	19645	204	133	935
	L_S	23452	352	255	1648

Fuente: Elaboración propia.

1.10.3.3 Cobertura por población objetivo

A continuación se presenta la información de cobertura para las poblaciones objetivo de la ENSANUT 2018 a nivel nacional por conjunto de selección.

Personas de 10 años y más

Conjunto de selección	Total de hogares	Seleccionadas	Rechazo / No en casa	Porcentaje
L_C	19673	18579	1094	94.4%
L_S	23638	22359	1279	94.6%

Fuente: Elaboración propia

Mujeres de 10 a 49 años

Conjunto de selección	Total de hogares	Seleccionadas	Rechazo / No en casa	Porcentaje
L_C	18741	26985	1948	93.3%
L_S	14813	21715	1761	92.5%

Fuente: Elaboración propia

Hombres de 12 años y más

Conjunto de selección	Total de hogares	Seleccionados	Rechazo / No en casa	Porcentaje
L_C	17808	15930	1878	89.5%
L_S	20248	18310	1938	90.4%

Fuente: Elaboración propia

Personas de 5 a 17 años

Conjunto de selección	Total de hogares	Seleccionadas	Rechazo / No en casa	Porcentaje
L_C	12957	12337	620	95.2%
L_S	11850	11284	566	95.2%

Fuente: Elaboración propia

Niños menores de 5 años (Desarrollo Infantil)

Conjunto de selección	Total de hogares	Seleccionados	Rechazo / No en casa	Porcentaje
L_C	16982	16430	552	96.7%
L_S	1368	1299	69	95.0%

Fuente: Elaboración propia

1.10.4 Probabilidades de inclusión para ENSANUT 2018

Probabilidad de inclusión de primera etapa (UPM)

El diseño muestral de la ENSANUT 2018 requirió que la probabilidad de selección de cada UPM sea proporcional al tamaño de la misma, siendo el número de viviendas ocupadas con personas presentes el tamaño de la UPM. La *probabilidad de inclusión de primera etapa para la i-ésima UPM muestreada dentro del estrato h* queda definida por:

$$PPE_{hi} = m_h \cdot \frac{m_{hi}}{M_h}$$

Donde,

$$M_h = \sum_{i=1}^{N_h} m_{hi},$$

PPE_{hi} = Probabilidad de selección de primera etapa para la UPM i del estrato h .

m_{hi} = Medida de tamaño de la UPM i del estrato h .

m_h = Número de UPM de la muestra en el estrato h .

N_h = Número total de UPM en el estrato h .

M_h = Total de viviendas ocupadas en el estrato h .

Ajuste por no respuesta a nivel de UPM

Para realizar este ajuste, es necesario definir la tasa de no respuesta a nivel de UPM, la cual mide la proporción de UPM seleccionadas que efectivamente participaron en la encuesta. Este ajuste se calcula separadamente para cada dominio de diseño y estrato, y es:

$$TNR_h = \frac{m_{h_efec}}{m_h}$$

Donde,

- TNR_h = Tasa de no respuesta de UPM en el estrato h.
 m_{h_efec} = Número de UPM efectivas de la muestra en el estrato h.
 m_h = Número de UPM de la muestra en el estrato h.

Por lo tanto, la probabilidad de inclusión final de primera etapa para la i-ésima UPM en el estrato h es:

$$PPE_{Fhi} = PPE_{hi} \cdot TNR_h$$

Probabilidad de inclusión de segunda etapa (vivienda)

Las viviendas de cada UPM fueron divididas en dos conjuntos según la información del enlistamiento:

- L_{shi} : viviendas sin un niño menor de 5 años (de manera general L_S),
- L_{chi} : viviendas con al menos un niño menor de 5 años (de manera general L_C).

Donde,

$$m'_{hi} = L_{shi} + L_{chi}$$

es el total de viviendas ocupadas con personas presentes en la i-ésima UPM en el estrato h según el enlistamiento. Luego, las viviendas fueron seleccionadas dentro de cada conjunto de manera independiente con igual probabilidad, por lo que las probabilidades de inclusión de segunda etapa están dadas por:

$$PSE_{shij} = \frac{n_{sj}}{L_{shi}}$$

Dónde:

- PSE_{shij} = Probabilidad de selección de segunda etapa para la vivienda sin niño menor a 5 años S_j en la UPM i del estrato h.
 n_{sj} = Número de viviendas seleccionadas sin niños menores de 5 años en la UPM i del estrato h.
 L_{shi} = Total de viviendas sin niños menores de 5 años en la UPM i del estrato h.

Y, equivalentemente para las viviendas de L_C , se tiene también:

$$PSE_{chij} = \frac{n_{cj}}{L_{chi}}$$

Dónde:

- PSE_{Chij} = Probabilidad de selección de segunda etapa para la vivienda con niño menor a 5 años C_j en la UPM i del estrato h .
- n_{Cj} = Número de viviendas seleccionadas con niños menores de 5 años en la UPM i del estrato h .
- L_{Chi} = Total de viviendas con niños menores de 5 años en la UPM i del estrato h .

Adicionalmente es necesario ajustar las probabilidades de inclusión mediante la proporción de viviendas seleccionadas que efectivamente participaron en la encuesta (efectivas), las viviendas rechazo y nadie en casa (no respuesta) y por cambio de ocupación⁵ (no elegibilidad).

En la ENSANUT 2018 se realiza este ajuste a nivel de estrato y conjunto de selección L_S o L_C , por lo que las viviendas pertenecientes al mismo estrato y conjunto de selección comparten el mismo ajuste.

Ajuste por no respuesta a nivel de vivienda

El cálculo del ajuste por no respuesta se realizó por estrato y conjunto de selección. Para las viviendas del conjunto L_S está dada por la siguiente expresión:

$$Aj_{nr_sh} = \frac{n_{sh}}{n_{sh} - n_{sh_nr} - n_{sh_nc} - n_{sh_ex}}$$

Dónde:

- Aj_{nr_sh} = Ajuste por no respuesta para las viviendas sin niños menores de 5 años del estrato h .
- n_{sh} = Total de viviendas sin niños menores de 5 años visitadas en el estrato h .
- n_{sh_nr} = Total de viviendas sin niños menores de 5 años 'rechazo' en el estrato h .
- n_{sh_nc} = Total de viviendas sin niños menores de 5 años 'nadie en casa' en el estrato h .
- n_{sh_ex} = Total de viviendas sin niños menores de 5 años 'no elegibles' en el estrato h .

De manera equivalente para las viviendas del conjunto L_C , el ajuste por no respuesta está dado por:

$$Aj_{nr_ch} = \frac{n_{ch}}{n_{ch} - n_{ch_nr} - n_{ch_nc} - n_{ch_ex}}$$

Dónde:

- Aj_{nr_ch} = Ajuste por no respuesta para las viviendas con niños menores de 5 años del estrato h .
- n_{ch} = Total de viviendas con niños menores de 5 años visitadas en el estrato h .
- n_{ch_nr} = Total de viviendas con niños menores de 5 años 'rechazo' en el estrato h .
- n_{ch_nc} = Total de viviendas con niños menores de 5 años 'nadie en casa' en el estrato h .
- n_{ch_ex} = Total de viviendas con niños menores de 5 años 'no elegibles' en el estrato h .

⁵ La categoría cambio de ocupación hace referencia a las viviendas cuyo resultado de entrevista fue vivienda temporal, desocupada, en construcción, inhabitable o destruida, convertida en negocio y otra razón.

Ajuste de no elegibilidad a nivel de vivienda

Para las viviendas del conjunto L_S , el ajuste de no elegibilidad está dado por la siguiente expresión:

$$Aj_{ne_sh} = \frac{n_{sh} - n_{sh_ex}}{n_{sh}}$$

Dónde:

Aj_{ne_sh} = Ajuste por no elegibilidad para las viviendas sin niños menores de 5 años en el estrato h.

n_{sh} = Total de viviendas sin niños menores de 5 años visitadas en el estrato h.

n_{sh_ex} = Total de viviendas sin niños menores de 5 años 'no elegibles' en el estrato h.

Equivalentemente para las viviendas del conjunto L_C , el ajuste de no elegibilidad está dado por:

$$Aj_{ne_ch} = \frac{n_{ch} - n_{ch_ex}}{n_{ch}}$$

Dónde:

Aj_{ne_ch} = Ajuste por no elegibilidad para las viviendas con niños menores de 5 años en el estrato h.

n_{ch} = Total de viviendas con niños menores de 5 años visitadas en el estrato h.

n_{ch_ex} = Total de viviendas con niños menores de 5 años 'no elegibles' en el estrato h.

Al igual que los cálculos anteriores, la probabilidad de inclusión final de segunda etapa se calcula de manera independiente para las viviendas de los conjuntos de selección L_S y L_C .

Por lo tanto, la probabilidad de inclusión final de segunda etapa para la j-ésima vivienda sin niño menor de 5 años dentro de la i-ésima UPM en el estrato h es:

$$PSE_F_{shij} = \frac{PSE_{shij}}{Aj_{nr_sh} \cdot Aj_{ne_sh}}$$

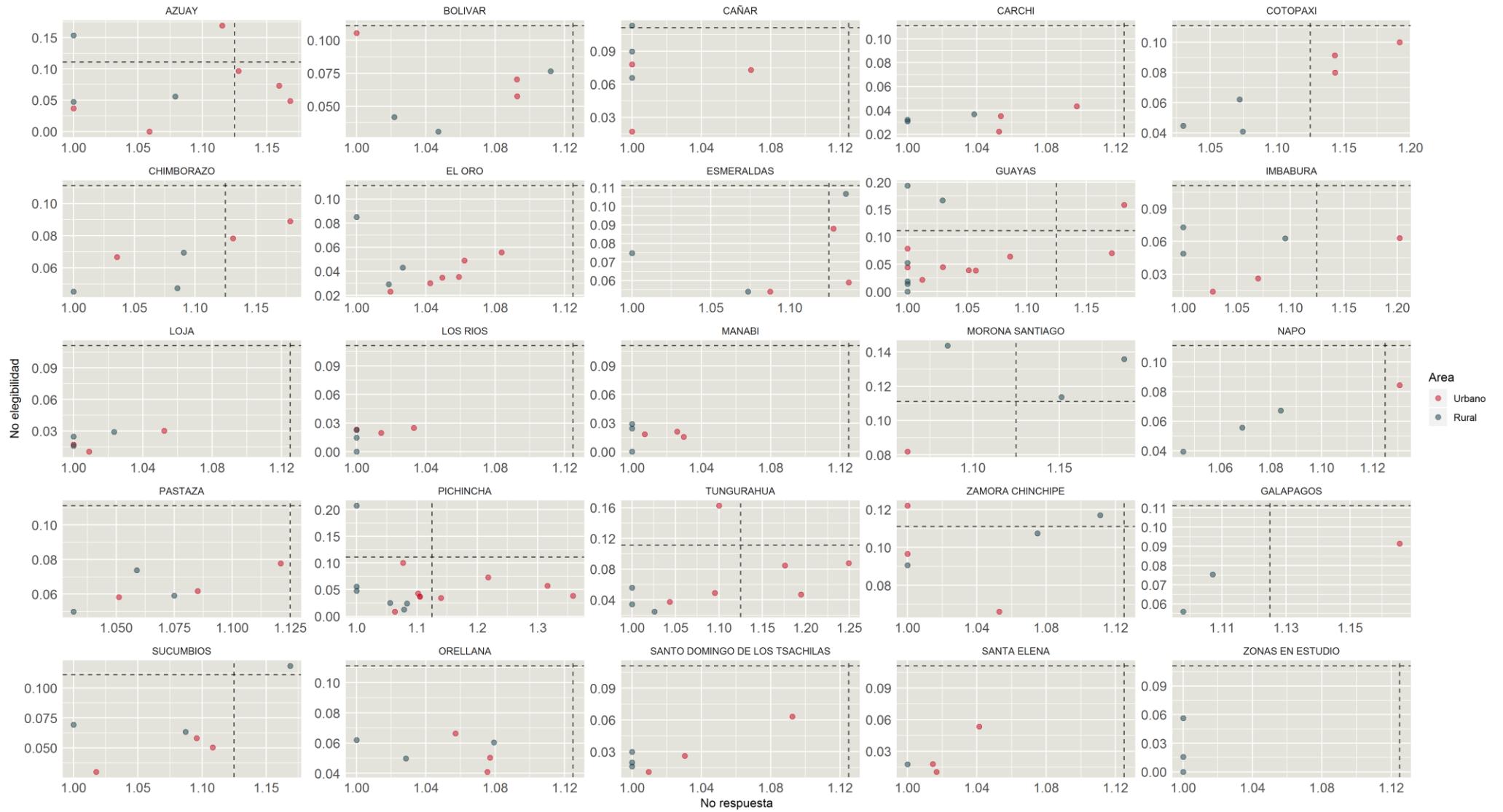
Equivalentemente, la probabilidad de inclusión final de segunda etapa para la j-ésima vivienda con al menos un niño menor de 5 años dentro de la i-ésima UPM en el estrato h es:

$$PSE_F_{chij} = \frac{PSE_{chij}}{Aj_{nr_ch} \cdot Aj_{ne_ch}}$$

A continuación se presenta un gráfico comparativo de los ajustes de no respuesta y no elegibilidad a los factores de expansión ENSANUT 2018, en el cual se puede comparar el rendimiento de la encuesta a nivel provincial, donde se puede apreciar que:

- Los ajustes por no respuesta y no elegibilidad de las provincias de Azuay, Guayas, Morona Santiago, Pichincha y Tungurahua alcanzan un máximo de 30% a nivel de estrato, siendo las provincias con más problemas de no respuesta y no elegibilidad desde el enlistamiento.
- El ajuste por no respuesta de las provincias de Cotopaxi, Chimborazo y Galápagos alcanza un máximo del 20% a nivel de estrato.
- El rendimiento del resto de provincias se ubica en los rangos esperados desde diseño.

Corrección realizada a los factores de expansión de ENSANUT 2018: no respuesta vs no elegibilidad, b1_personas



Probabilidad de inclusión de tercera etapa (personas)

La probabilidad de inclusión de tercera etapa se calcula en función a las poblaciones objetivo presentadas en la **Tabla 11**. Puesto que dentro de cada hogar de la vivienda j de la UPM i en el estrato h se selecciona únicamente a una persona por población objetivo, se simplifica la notación con subíndice k . Por lo tanto, la probabilidad de inclusión de tercera etapa para la persona k es:

$$PTE_k = \frac{1}{M_k},$$

Dónde:

PTE_k = Probabilidad de selección de tercera etapa para la persona seleccionada del hogar k .
 M_k = Total de personas de la población objetivo en el hogar k .

Ajuste por no respuesta a nivel de población objetivo

Para el cálculo del ajuste por no respuesta a nivel de población objetivo se debe considerar que existieron hogares con personas seleccionadas que no facilitaron información, por lo que se debe recalcular su probabilidad de selección⁶.

El cálculo del ajuste por no respuesta se realizó por estrato y conjunto de selección L_S o L_C , siendo el mismo esquema de cálculo para todas las poblaciones objetivo. Para las personas seleccionadas dentro de los hogares del conjunto L_S está dada por la siguiente expresión:

$$Aj_{nr_shk} = \frac{me_{sh}}{TPO_{sh}}$$

Dónde:

$$TPO_{sh} = \sum_h \left(I_{Sk} \cdot \frac{1}{PTE_k} \right),$$

Aj_{nr_shk} = Ajuste por no respuesta a la persona k del conjunto L_S del estrato h .
 me_{sh} = Total de personas elegibles dentro del conjunto L_S en el estrato h .
 I_{Sk} = Función indicatriz de la persona k sobre el conjunto L_S .
 TPO_{sh} = Total de personas elegibles del conjunto L_S en los hogares con una encuesta efectiva en el estrato h .

Equivalentemente, para las personas seleccionadas dentro de los hogares del conjunto L_C está dada por la siguiente expresión:

⁶ Esto es necesario puesto que si la corrección se hace a nivel de hogar, el factor de expansión de tercera etapa sería $M_k/0$, lo que es indeterminado.

$$Aj_{nr_chk} = \frac{me_{ch}}{TPO_{ch}}$$

Dónde:

$$TPO_{ch} = \sum_h \left(I_{ck} \cdot \frac{1}{PTE_k} \right),$$

Aj_{nr_chk} = Ajuste por no respuesta a la persona k del conjunto L_C del estrato h .

me_{ch} = Total de personas elegibles dentro del conjunto L_C en el estrato h .

I_{ck} = Función indicatriz de la persona k sobre el conjunto L_C .

TPO_{ch} = Total de personas elegibles del conjunto L_C en los hogares con una encuesta efectiva en el estrato h .

Al igual que los cálculos del ajuste por no respuesta anteriores, la probabilidad de inclusión final de segunda etapa se calcula de manera independiente para las viviendas de los conjuntos de selección L_S y L_C .

Por lo tanto, la probabilidad de inclusión final de tercera etapa para la k -ésima persona seleccionada de la población objetivo sobre el conjunto de hogares L_S es:

$$PTE_{F_{Sk}} = \frac{PTE_k}{Aj_{nr_Shk}}$$

De igual manera, la probabilidad de inclusión final de tercera etapa para la k -ésima persona seleccionada de la población objetivo sobre el conjunto de hogares L_C es:

$$PTE_{F_{Ck}} = \frac{PTE_k}{Aj_{nr_chk}}$$

1.10.5 Factores de expansión ENSANUT 2018

Como se describe en la **Tabla 11**, la información recolectada por ENSANUT 2018 está ordenada en 9 bases de datos, en función a cada población objetivo específica. Además, como las viviendas fueron seleccionadas a partir de una estratificación explícita de dentro de cada UPM (conjuntos de selección L_S y L_C), viviendas de la misma UPM pero seleccionadas sobre conjuntos diferentes tienen un diferente factor de expansión.

Considerando todo lo expuesto anteriormente, los factores de expansión de ENSANUT 2018 quedan definidos de la siguiente manera:

– Bases f1_personas:

$$f_{exp_{hogar}} = \begin{cases} \frac{1}{PPE_{F_{hi}}} \cdot \frac{1}{PSE_{F_{Shij}}}, & \text{si la viviendas es seleccionada sobre } L_S \\ \frac{1}{PPE_{F_{hi}}} \cdot \frac{1}{PSE_{F_{Chij}}}, & \text{si la viviendas es seleccionada sobre } L_C \end{cases}.$$

- Base $f1_hogar$: se usa el factor de expansión de la base $f1_personas$ correspondiente al jefe de hogar de cada uno de los hogares investigados.
- Bases $f1_etiquetado$, $f2_mef$, $f3_ssr_hombres$, $f4_fact_riesgo$ y $f5_desarrollo_infantil$, el factor de expansión es:

$$f_{exp\,persona} = \begin{cases} \frac{1}{PPE_{F_{hi}}} \cdot \frac{1}{PSE_{F_{shij}}} \cdot \frac{1}{PTE_{F_{sk}}}, & \text{si la viviendas es seleccionada sobre } L_S \\ \frac{1}{PPE_{F_{hi}}} \cdot \frac{1}{PSE_{F_{chij}}} \cdot \frac{1}{PTE_{F_{ck}}}, & \text{si la viviendas es seleccionada sobre } L_C \end{cases}$$

donde k varía en función a la población objetivo de cada base de datos.

- Para las bases $f2_lactancia$ y $f2_salud_niñez$ se usa el factor de expansión de la MEF madre de cada niño investigado. Esto debido a:
 1. Los niños investigados dependen exclusivamente de la si la MEF madre acepto dar la información o no.
 2. La información de Lactancia Materna se recaba para el último nacido vivo y actualmente vivo menor de 3 años de cada MEF.
 3. La información de salud en la niñez se recaba para los últimos 3 nacidos vivos de cada MEF.

1.10.6 Recorte de factores de expansión extremos

Debido a los ajustes por no respuesta, no elegibilidad y a la selección de tercera etapa presente para algunas poblaciones objetivo, la variabilidad de los factores de expansión de ENSANUT 2018 se ha visto sumamente incrementada. Tal y como se expone el 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). Sin embargo, para Valliant (2018) son *ad hoc* y sumamente teóricos. Para ENSANUT 2018 se utiliza el método de recorte y redistribución de factores de expansión⁷ más común, descrito en Valliant (2018), el cual propone el siguiente método iterativo:

⁷ En inglés, Weight Trimming and Redistribution.

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

$$fexp_{trim} = \begin{cases} U & \text{si } w_k \geq U, \\ w_k & \text{si } L < w_k < U, \\ L & \text{si } w_k \leq L. \end{cases}$$

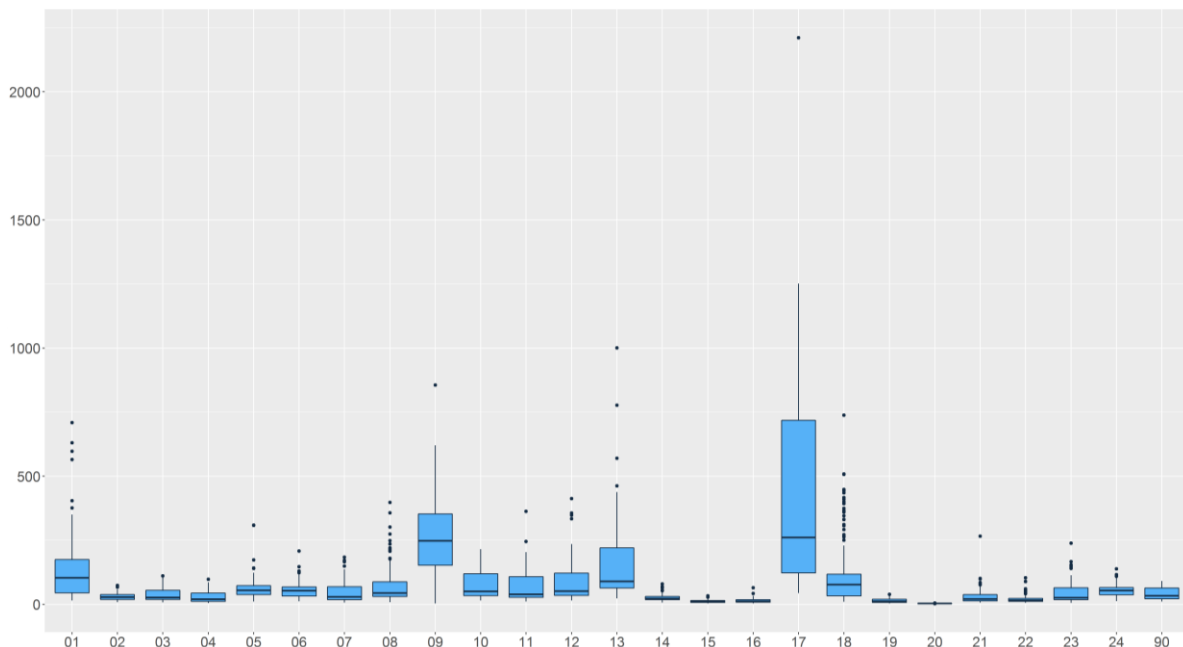
Donde $fexp_{trim}$ es el conjunto de los factores de expansión recortados.

- (3) Calcular $K = \sum_k |fexp - fexp_{trim}|$.
- (4) Distribuir K equitativamente entre todos los pesos no recortados.
- (5) Repetir los pasos (1)-(4) hasta que $K = 0$.

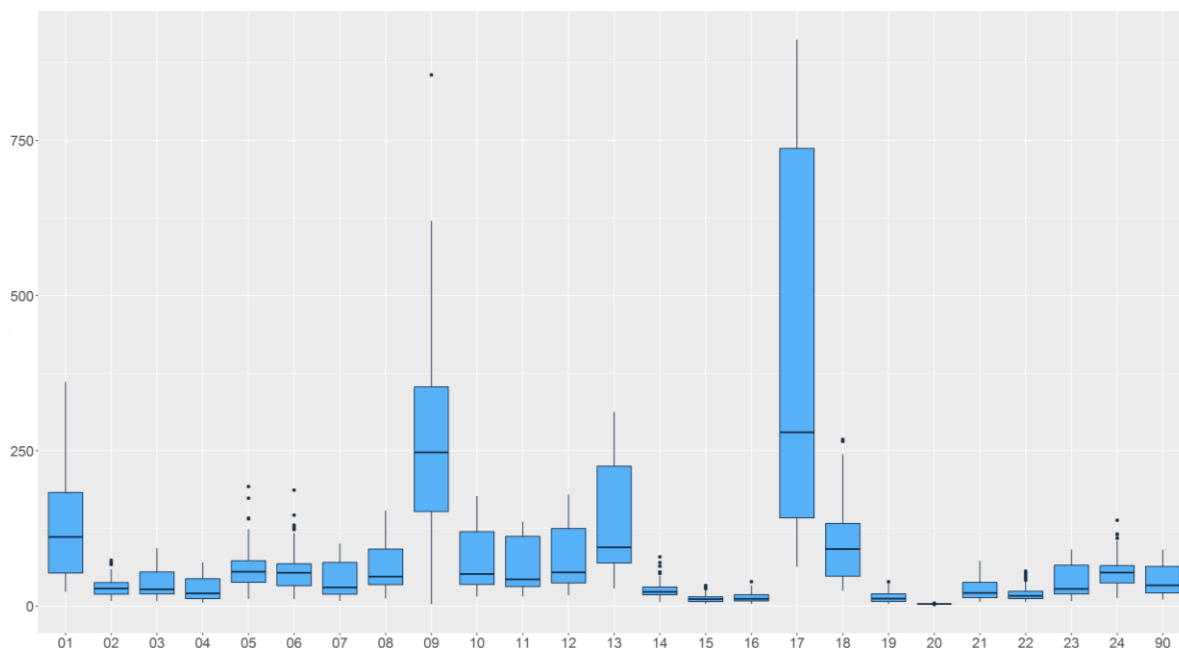
Puesto que ENSANUT 2018 presenta factores de expansión extremos únicamente en la cola derecha de su distribución, no se fija L y se fija U en 3.5 veces la mediana. Además, el recorte y redistribución de los factores de ENSANUT 2018 se hizo de manera independiente para cada provincia y base de datos, con la finalidad de diferenciar los dominios de diseño y métodos de selección de cada población objetivo.

A continuación se presentan los diagramas de caja y bigote para los factores de expansión finales $fexp_{hogar}$ y recortados $fexp_{hogar_trim}$ para la base $fl_personas$ por provincia.

Factores de expansión finales ENSANUT 2018



Factores de expansión recortados ENSANUT 2018



Para revisar más información respecto al recorte y redistribución de los Factores de expansión de ENSANUT 2018, incluidos los diagramas de caja y bigote correspondientes a las demás bases de datos, dirigirse al Anexo 1.

1.10.7 Calibración de los factores de expansión ENSANUT 2018

Considerando que la muestra fue seleccionada a partir del MMM para encuestas dirigidas a hogares del INEC, el cual está constituido en su mayoría por información que data del 2010 y que el periodo de levantamiento de información es 2018, existe una diferencia sustancial en la constitución de los dominios de estudio. Con el fin de solventar este inconveniente, se decide calibrar los factores de expansión utilizando como fuente de información auxiliar las proyecciones poblacionales vigentes del INEC.

Para identificar cual es el mejor método de calibración a utilizar de entre todos los considerados, se contrastaron los *Criterios para evaluar el éxito de la calibración* introducidos en (Silva A., 2004). Para conocer el detalle y los resultados de estos criterios dirigirse al Anexo 1.

1.10.7.1 Esquema de calibración

Puesto que cada base de ENSANUT 2018 responde a su propia población objetivo, se calibra cada una de ellas de manera separada. A continuación se describe el procedimiento de calibración realizado para las bases de ENSANUT 2018.

Base f1_personas

Para la base f1_personas se aplicó el método de calibración denominado Raking Generalizado, término acuñado en (Deville J.C., Särndal C.E. y Sautory O., 1993), el cual hace referencia a los procesos de estimación en encuestas con información auxiliar de la forma de totales marginales conocidos.

Con este método se asegura que la suma de los factores de expansión calibrados en cada subconjunto sea igual a los totales marginales de los mismos provistos desde la información auxiliar. Los totales marginales utilizados para la calibración de la base f1_personas se dividieron en dos grupos:

1. Región natural por área: Se calculan los totales marginales de cada región natural por área (a excepción de la región Insular la cual no se diferencia por área) para los siguientes subconjuntos:
 - a. (0,5] años: niños menores de 5 años,
 - b. [5,12) años: niños de 5 a 11 años,
 - c. [12,50) años por sexo: personas de 12 a 49 años por sexo, y
 - d. [50, 100+) años por sexo: personas mayores o iguales a 50 años por sexo.En total existen 42 subconjuntos de calibración para los totales marginales 1 (Las Zonas no Delimitadas aportan su población al rural de la región natural Costa).
2. Provincial: se calculan los totales marginales de cada provincia para los siguientes subconjuntos:
 - a. (0,5] años: niños menores de 5 años,
 - b. [5,12) años: niños de 5 a 11 años,
 - c. [12,50) años por sexo: personas de 12 a 49 años por sexo, y
 - d. [50, 100+) años por sexo: personas mayores o iguales a 50 años por sexo.En total se definen 145 subconjuntos de calibración para los totales marginales 2: 144 para las 24 provincias del Ecuador 1 uno para las Zonas no Delimitadas (para las cuales se define solo un subconjunto, su total poblacional).

Bases f1_etiquetado, f2_mef, f3_ssr_hombres, f4_fact_riesgo y f5_desarrollo_infantil

Para las bases *f1_etiquetado*, *f2_mef*, *f3_ssr_hombres*, *f4_fact_riesgo* y *f5_desarrollo_infantil* se utilizó un método de calibración conocido como Post-estratificación completa. En este método se clasifica la muestra en postestratos, definidos por el cruce de una a más variables categóricas, ajustando el total poblacional de cada postestrato al indicado por la fuente de información auxiliar.

A continuación se presenta la definición de cada postestrato para las bases en cuestión.

- Los postestratos para la base f1_etiquetado se definen por el cruce a nivel nacional de las variables área, sexo y un grupo de edad, personas de 10 o más años.
- Los postestratos para la base f2_mef se definen por el cruce a nivel nacional de las variables área y grupo de edad, la cual consta de 8 grupos quinquenales [10, 15) años, [15, 20) años, [20, 25) años, [25, 30) años, [30, 35) años, [35, 40) años, [40, 45) años y [45, 50) años.
- Los postestratos para la base f3_ssr_hombres se definen por el cruce a nivel nacional de las variables área y un grupo de edad, hombres de 12 años y más.
- Los postestratos para la base f4_fact_riesgo se definen por el cruce a nivel nacional de las variables área, sexo y un grupo de edad, personas de 5 a 17 años.
- Los postestratos para la base f5_desarrollo_infantil se definen por el cruce a nivel nacional de las variables área y un grupo de edad, niños menores de 5 años de edad.

La calibración de los factores de expansión ENSANUT 2018 se realizó mediante las funciones especializadas del paquete “*sampling*” del software estadístico R. Para más información dirigirse a (Tillé Y. y Matei A., 2016).

1.11 Estimación de características

Una vez obtenidos los factores de expansión calibrados, se calculan los estimadores de ENSANUT 2018 mediante el estimador de Horvitz-Thompson, el cual sirve para estimar el valor total de una característica determinada. Está dado por (Särndal C.E., Swensson B. y Wretman J., 1992).

$$\hat{Y}_{mHT} = \sum_i \sum_j f_{exp_{kp}} \cdot y_k$$

Dónde:

- \hat{Y}_{mHT} = *estimador para el total Y_k de la variable y en el dominio de estudio m*
- i = *número de UPM investigadas en el dominio de estudio m .*
- j = *número de viviendas investigadas en la UPM i del dominio de estudio m*
- y_k = *valor de la variable y para la unidad de análisis k .*

1.12 Error de muestreo

Luego de realizar la estimación respectiva para la variable de interés a nivel de dominio de estudio, el error de muestreo podrá ser calculado a partir de la estimación de la varianza del estimador del total \hat{Y}_{mHT} . Para calcular los errores de

muestreo de cada estimador, se debe tomar en cuenta los diferentes aspectos del diseño muestral, es decir, las dos etapas de muestreo, la estratificación presente en los dominios de estudio y los procesos de selección en cada una de las etapas.

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

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

Un estimador insesgado para la varianza está dado por:

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

En el cual el \hat{V}_i apropiado es

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

Dónde:

- i, j = Índice que recorre los UPM i, j en el dominio de estudio m .
- q, r = Índices que recorren las viviendas de la UPM i en el dominio de estudio m .
- π_{Ii} = Probabilidad de selección de la Etapa I para la i – ésima UPM en el dominio dominio de estudio m .
- $\pi_{IIq|i}$ = Probabilidad de selección de la Etapa II para la q – ésima vivienda, dada la i – ésima UPM.
- $\check{\Delta}_{Iij}$ = Cantidad Δ expandida asociada a las UPM i, j .
- $\check{\Delta}_{IIqr|i}$ = Cantidad Δ expandida asociada a las viviendas q, r dada la selección de la i – ésima UPM.

1.13 Métodos de estimación 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 y Frankel (1974), Wolter (1985) y Lehtonen y 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, tamaños y proporciones.
- 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, es que no generan resultados idénticos del error de muestreo, pero hay que señalar que existen estudios (Kish y 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 ENSANUT 2018.

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

Tabla 13 Variables requeridas para la declaración del diseño muestral


Característica	Variable	Descripción
Unidad primaria de muestreo	upm	Unidad primaria de muestreo de ENSANUT 2018
Estratos	estrato	Identificación de estrato muestral (aproximación clasificación socio-económica).
Factor de expansión	Fexp	Factor de expansión calibrado


Fuente: Elaboración propia

2 BIBLIOGRAFÍA

- CAN. (2000). Seminario: Censo 2000 de Población y Vivienda de los Países Andinos. Lima: Comunidad Andina de Naciones.
- Deville J.C., Särndal C.E. y Sautory O. (1993). *Generalized Raking Procedures in Survey Sampling*. Journal of the American Statistical Association.
- INEC. (2019). *Marco Maestro de Muestreo para encuestas de hogares*. Quito-Ecuador: Instituto Nacional de Estadística y Censos.
- Kish L. (1972). *Muestreo de Encuestas*. Ciudad de México: Trillas.
- Kish L. y Frankel M.R. (1974). Inference from Complex Samples. Journal of the Royal Statistical Society: Series B (Methodological) 36(1), 1-22.
- Lehtonen R. y Pahkinen E.J. (1995). Practical methods for design and analysis of complex surveys. Chichester: John Wiley.
- Potter F.J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the section on survey research methods*.
- Potter F.J. (1993). The effect of weight trimming on nonlinear survey estimates. American Statistical Association.
- Särndal C.E., Swensson B. y Wretman J. (1992). *Model Assisted Survey Sampling*. Nueva York: Springer.
- SENPLADES. (2017). *Plan Nacional de Desarrollo 2017-2021. Toda una Vida*. Quito: Secretaría Nacional de Planificación y Desarrollo.
- Silva A. (2004). *Calibration Estimation: When and Why, How much and How*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística - IBGE.
- Tillé Y. y Matei A. (2016). *sampling: Survey Sampling*. R package version 2.8.
- Valliant R., Dever J.A. y Kreuter F. (2018). *Practical Tools for Designing and Weighting Survey Samples*. Springer International Publishing.
- Wolter K.M. (1985). *Introduction to variance estimation*. New York: Springer-Verlag.


**CADA
HECHO
DE TU
VIDA**
Cuenta

 @ecuadorencifras

 INEC/Ecuador

 @InecEcuador

 INECEcuador

 t.me/equadorencifras

 INEC Ecuador

Anexo 1

**Factores de
expansión**

ENSANUT 2018

Tabla de contenido

1	INTRODUCCIÓN	3
2	Recorte y redistribución de los factores de expansión ENSANUT 2018	3
2.1	CÓDIGO UTILIZADO PARA EL RECORTE Y REDISTRIBUCIÓN	8
3	Calibración de los factores de expansión ENSANUT 2018	8
3.1	BASE 1 F1_PERSONAS	11
3.2	BASE 3 F1_ETIQUETADO	16
3.3	BASE 4 F2_MEF	17
3.4	BASE 7 F3_SSR_HOMBRES	20
3.5	BASE 8 F4_FACT_RIESGO	21
3.6	BASE 9 F5_DESARROLLO_INFANTIL	22
4	BIBLIOGRAFÍA	23

Lista de tablas

Tabla 1	Descriptivos pesos g, base f1_personas.	12
Tabla 2	Descriptivos factores de expansión iniciales y finales, base f1_personas.	12
Tabla 3	Comparativa criterios de evaluación de la calibración, base f1_personas.	13
Tabla 4	Descriptivos pesos g, base f2_mef.	18
Tabla 5	Descriptivos factores de expansión iniciales y finales, base f2_mef.	18
Tabla 6	Comparativa criterios de evaluación de la calibración, base f2_mef.	18

1 INTRODUCCIÓN

En este anexo se presenta a detalle más información referente a los procesos de recorte y calibración realizados a los factores de expansión de ENSANUT 2018.

2 Recorte y redistribución de los factores de expansión ENSANUT 2018

En esa sección se presentan el total de factores de expansión recortados por provincia para cada una de las bases de datos ENSANUT 2018. Para más información acerca de este procedimiento, dirigirse a la sección 3.10.6 de la Metodología de la ENSANUT 2018.

Como se puede observar en las siguientes tablas, el porcentaje de factores de expansión recortados es más alto en las bases *f1_etiquetado*, *f3_ssr_hombres* y *f4_fact_riesgo* que en las bases de datos *f1_personas* y *f2_mef*. Esto se debe a que las primeras consideran una tercera etapa de selección para captar la información de sus respectivas poblaciones objetivos, lo que incrementa la posibilidad de encontrar factores de expansión considerablemente altos.

Base 1 *f1_personas*

Provincia	No recortados	Recortados	Porcentaje
Azuay	4622	186	3,9%
Bolívar	5152	0	0,0%
Cañar	4406	182	4,0%
Carchi	4839	58	1,2%
Cotopaxi	6618	22	0,3%
Chimborazo	6900	39	0,6%
El Oro	11603	919	7,3%
Esmeraldas	6842	588	7,9%
Guayas	13792	0	0,0%
Imbabura	4753	239	4,8%
Loja	4994	711	12,5%
Los Ríos	9653	502	4,9%
Manabí	9026	597	6,2%
Morona Santiago	5417	0	0,0%
Napo	8643	0	0,0%
Pastaza	6215	73	1,2%
Pichincha	5517	499	8,3%
Tungurahua	4114	494	10,7%
Zamora Chinchipe	5449	0	0,0%
Galápagos	5221	0	0,0%
Sucumbíos	6020	145	2,4%

Orellana	6348	63	1,0%
Santo Domingo De Los Tsáchilas	9906	591	5,6%
Santa Elena	5948	0	0,0%
Zonas no Delimitadas	841	0	0,0%

Base 3 f1_etiquetado

Provincia	No recortados	Recortados	Porcentaje
Azuay	1089	92	7,8%
Bolívar	1272	36	2,8%
Cañar	1059	68	6,0%
Carchi	1224	76	5,8%
Cotopaxi	1587	26	1,6%
Chimborazo	1703	48	2,7%
El Oro	2803	262	8,5%
Esmeraldas	1473	156	9,6%
Guayas	3305	60	1,8%
Imbabura	1076	66	5,8%
Loja	1197	136	10,2%
Los Ríos	2223	204	8,4%
Manabí	2134	145	6,4%
Morona Santiago	1178	50	4,1%
Napo	1882	77	3,9%
Pastaza	1461	68	4,4%
Pichincha	1348	189	12,3%
Tungurahua	1020	135	11,7%
Zamora Chinchipe	1140	74	6,1%
Galápagos	1765	2	0,1%
Sucumbíos	1391	111	7,4%
Orellana	1356	77	5,4%
Santo Domingo De Los Tsáchilas	2294	282	10,9%
Santa Elena	1296	19	1,4%
Zonas no Delimitadas	182	21	10,3%

Base 4 f2_mef

Provincia	No recortados	Recortados	Porcentaje
Azuay	1354	50	3,6%
Bolívar	1446	0	0,0%
Cañar	1285	59	4,4%
Carchi	1380	55	3,8%
Cotopaxi	1895	7	0,4%
Chimborazo	1942	7	0,4%
El Oro	3068	333	9,8%
Esmeraldas	1985	167	7,8%
Guayas	3973	7	0,2%
Imbabura	1365	96	6,6%
Loja	1354	194	12,5%
Los Ríos	2827	177	5,9%
Manabí	2638	151	5,4%
Morona Santiago	1480	0	0,0%
Napo	2459	0	0,0%
Pastaza	1884	25	1,3%
Pichincha	1502	240	13,8%
Tungurahua	1175	135	10,3%
Zamora Chinchipe	1565	0	0,0%
Galápagos	1617	0	0,0%
Sucumbíos	1825	45	2,4%
Orellana	1817	17	0,9%
Santo Domingo De Los Tsáchilas	2896	190	6,2%
Santa Elena	1769	0	0,0%
Zonas no Delimitadas	244	0	0,0%

Base 7 f3_ssr_hombres

Provincia	No recortados	Recortados	Porcentaje
Azuay	879	94	9,7%
Bolívar	1024	47	4,4%
Cañar	828	61	6,9%
Carchi	1023	49	4,6%
Cotopaxi	1267	54	4,1%
Chimborazo	1326	71	5,1%
El Oro	2318	225	8,8%
Esmeraldas	1221	152	11,1%

Guayas	2776	98	3,4%
Imbabura	888	45	4,8%
Loja	995	78	7,3%
Los Ríos	1964	125	6,0%
Manabí	1811	142	7,3%
Morona Santiago	902	43	4,6%
Napo	1634	88	5,1%
Pastaza	1193	70	5,5%
Pichincha	1135	137	10,8%
Tungurahua	835	139	14,3%
Zamora Chinchipe	902	45	4,8%
Galápagos	1502	16	1,1%
Sucumbíos	1204	90	7,0%
Orellana	1141	76	6,2%
Santo Domingo De Los Tsáchilas	1972	215	9,8%
Santa Elena	1124	33	2,9%
Zonas no Delimitadas	164	19	10,4%

Base 8 f4_fact_riesgo

Provincia	No recortados	Recortados	Porcentaje
Azuay	565	64	10,2%
Bolívar	684	32	4,5%
Cañar	567	52	8,4%
Carchi	607	64	9,5%
Cotopaxi	922	34	3,6%
Chimborazo	910	31	3,3%
El Oro	1550	207	11,8%
Esmeraldas	981	98	9,1%
Guayas	1815	77	4,1%
Imbabura	624	59	8,6%
Loja	656	85	11,5%
Los Ríos	1289	127	9,0%
Manabí	1114	124	10,0%
Morona Santiago	715	36	4,8%
Napo	1194	73	5,8%
Pastaza	876	59	6,3%
Pichincha	697	118	14,5%
Tungurahua	539	84	13,5%
Zamora Chinchipe	727	55	7,0%
Galápagos	804	8	1,0%

Sucumbíos	827	80	8,8%
Orellana	899	48	5,1%
Santo Domingo De Los Tsáchilas	1327	140	9,5%
Santa Elena	820	33	3,9%
Zonas no Delimitadas	116	8	6,5%

Base 9 f5_desarrollo_infantil

Provincia	No recortados	Recortados	Porcentaje
Azuay	472	28	5,6%
Bolívar	427	0	0,0%
Cañar	527	13	2,4%
Carchi	550	16	2,8%
Cotopaxi	590	7	1,2%
Chimborazo	586	11	1,8%
El Oro	1315	64	4,6%
Esmeraldas	812	31	3,7%
Guayas	1264	27	2,1%
Imbabura	522	19	3,5%
Loja	621	31	4,8%
Los Ríos	1249	27	2,1%
Manabí	1105	36	3,2%
Morona Santiago	536	12	2,2%
Napo	640	11	1,7%
Pastaza	542	32	5,6%
Pichincha	625	59	8,6%
Tungurahua	486	43	8,1%
Zamora Chinchipe	576	19	3,2%
Galápagos	361	0	0,0%
Sucumbíos	721	31	4,1%
Orellana	709	22	3,0%
Santo Domingo De Los Tsáchilas	1304	39	2,9%
Santa Elena	512	11	2,1%
Zonas no Delimitadas	88	0	0,0%

2.1 Código utilizado para el recorte y redistribución

Para realizar el recorte y redistribución de los factores de expansión, utilizamos el algoritmo descrito en (Valliant R., Dever J.A. y Kreuter F., 2018), el cual fue implementado en el software estadístico R. El valor $p = 3.5$ es el recomendado para su implementación. A continuación se presenta la función utilizada.

```
trimming <- function(x, p=3.5){
  # cota superior
  U <- p*median(x)

  # factores recortados
  i <- x>U
  sum(i)
  xt <- ifelse(x>U, U, x)

  # cantidad s
  s <- sum(x-xt)
  j <- 0

  # lazo while
  while(s>0){
    x <- xt + (s/(length(x)-sum(i)))*(1-i)

    # nuevos factores recortados
    i <- x>U
    sum(i)

    xt <- ifelse(x>U, U, x)

    # nueva cantidad s
    s <- sum(x-xt)
    j <- j+1
  }
  return(xt)
}
```

3 Calibración de los factores de expansión ENSANUT 2018

La calibración de los factores de expansión se hizo en función a los procedimientos desarrollados en (Deville J.C., Särndal C.E. y Sautory O., 1993). Para ENSANUT 2018, se desarrollaron diferentes esquemas de calibración, para los cuales se verificó su consistencia, factibilidad e idoneidad según los criterios para evaluar el éxito de la calibración, descritos en (Silva A., 2004).

Estimador de calibración

El estimador de calibración se define de la siguiente manera:

Considere una función de distancia G con argumentos $x = w_k/d_k$ con las siguientes propiedades:

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

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

Minimizar $\sum_s d_k G(w_k/d_k) - \lambda'(\sum_s w_k \mathbf{x}_k - \sum_U \mathbf{x}_k)$, donde U hace referencia a la población, $\mathbf{x}_k = (x_{k1}, \dots, x_{kj}, \dots, x_{kJ})'$ 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:

$$\sum_s d_k F(\mathbf{x}_k' \lambda) \mathbf{x}_k = \sum_U \mathbf{x}_k.$$

Luego, el estimador de calibración queda definido por:

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

Donde: los (y_k, x_k) son datos observados en la muestra ($k \in s$) y una función de distancia G dada. Nótese que $w_k = d_k F(\mathbf{x}_k' \lambda)$ es el nuevo factor de expansión calibrado. En este contexto, notaremos $g = F(\mathbf{x}_k' \lambda)$.

Para la calibración de los factores de expansión de ENSANUT 2018 se exploraron los resultados obtenidos con las distancias Lineal, Multiplicativa (Raking) y Logit, propuestas en (Deville J.C., Särndal C.E. y Sautory O., 1993). Su implementación se realizó mediante las funciones especializadas del paquete "sampling" del software estadístico R. Para más información dirigirse a (Tillé Y. y Matei A., 2016).

Cabe mencionar también que los esquemas de calibración realizados para cada una de las bases de ENSANUT 2018 consideraron la población objetivo de cada una de ellas para delimitar los grupos de edad a ser considerados en su implementación.

Evaluación de la calibración

Silva (2004) propone 7 criterios de evaluación, para que el proceso sea adecuado. Los criterios planteados son:

- Error relativo promedio sobre las variables auxiliares

$$M1 = \frac{1}{p} \sum_{j=1}^p \frac{|\hat{t}_{xc} - t_x|}{t_x}.$$

- Coeficiente de variación HT relativo promedio

$$M2 = \frac{1}{p} \sum_{j=1}^p \frac{(Var(\hat{t}_{x\pi}))^{1/2}}{t_x}.$$

- Proporción de pesos extremos (límite inferior)

$$M3 = \frac{1}{n} \sum_{k \in S} I(g_k < L).$$

- Proporción de pesos extremos (límite superior)

$$M4 = \frac{1}{n} \sum_{k \in S} I(g_k > U).$$

- Coeficiente de variación de los g_k

$$M5 = \frac{\sigma(g)}{\bar{g}}.$$

- Distancia entre los pesos de calibración y los pesos originales

$$M6 = \frac{1}{n} \sum_{k \in S} \frac{(w_k - d_k)^2}{d_k} = \frac{1}{n} \sum_{k \in S} d_k (g_k - 1)^2.$$

- Eficiencia de los estimadores de calibración sobre la estimación del diseño de muestreo

$$M7 = \frac{1}{J} \sum_{j=1}^J \frac{\hat{V}_g(\hat{T}_{y_j c})}{\hat{V}(\hat{T}_{y_j})}.$$

Los esquemas de calibración considerados para ENSANUT 2018 fueron comparados a través de los criterios propuestos por Silva (2004), eligiendo el más adecuado.

3.1 Base 1 f1_personas

Esquema de calibración

Para la base f1_personas se consideraron dos esquemas de calibración:

1. El primero consistió en introducir la información auxiliar en forma de totales conocidos en cada uno de los cruces de las variables provincia, área, sexo y grupo de edad. Este método se conoce también como Postestratificación Completa.

Al ser una Postestratificación Completa los g resultantes son los mismos independientemente de la distancia G que se use. Por defecto tomamos la distancia multiplicativa.

2. El segundo consideró el método de calibración denominado Raking Generalizado, término acuñado en (Deville J.C., Särndal C.E. y Sautory O., 1993), el cual hace referencia a los procesos de estimación en encuestas con información auxiliar de la forma de totales marginales conocidos.

Con este método se asegura que la suma de los factores de expansión calibrados en cada subconjunto sea igual a los totales marginales de los mismos provistos desde la información auxiliar. Los totales marginales utilizados para la calibración de la base f1_personas se dividieron en dos grupos:

- a. Región natural por área: Se calculan los totales marginales de cada región natural por área, a excepción de la región Insular la cual no se diferencia por área, para los siguientes subconjuntos:
 - (0,5] años: niños menores de 5 años,
 - [5,12) años: niños de 5 a 11 años,
 - [12,50) años por sexo: personas de 12 a 49 años por sexo, y
 - [50, 100+) años por sexo: personas mayores o iguales a 50 años por sexo.

En total existen 42 subconjuntos de calibración para los totales marginales 1 (Las Zonas no Delimitadas aportan su población al rural de la región natural Costa. Para identificar los 42 subconjuntos de calibración se genera una variable, llamada id_calib_v1, la cual relaciona cada observación con su subconjunto de calibración.

- b. Provincial: se calculan los totales marginales de cada provincia para los siguientes subconjuntos:
 - (0,5] años: niños menores de 5 años,
 - [5,12) años: niños de 5 a 11 años,
 - [12,50) años por sexo: personas de 12 a 49 años por sexo, y
 - [50, 100+) años por sexo: personas mayores o iguales a 50 años por sexo.

En total se definen 145 subconjuntos de calibración para los totales marginales 2: 144 para las 24 provincias del Ecuador 1 uno subconjunto para la totalidad de las Zonas no Delimitadas. Para identificar los 145 subconjuntos de calibración se genera una variable, llamada *id_calib_v2*, la cual relaciona cada observación con su subconjunto de calibración.

Evaluación de la calibración

En primer lugar, se muestra un análisis descriptivo de los pesos *g* y de los factores de expansión iniciales y finales para los esquemas de calibración considerados.

Tabla 1 Descriptivos pesos *g*, base *f1_personas*.

Esquema calibración	Distancia	Mínimo	Primer cuartil	Mediana	Media	Tercer cuartil	Máximo
1	Raking	0,4954	1,1145	1,2439	1,3050	1,4024	3,5175
2	Raking	0,8521	1,1576	1,2675	1,3046	1,3918	2,5399
2	Lineal	0,7492	1,1597	1,2739	1,3059	1,3957	2,4903
2	Logit	0,8534	1,1575	1,2665	1,3046	1,3917	2,4932

Fuente: Elaboración propia

Tabla 2 Descriptivos factores de expansión iniciales y finales, base *f1_personas*.

Esquema calibración	Factor de expansión	Mínimo	Primer cuartil	Mediana	Media	Tercer cuartil	Máximo
Iniciales	d_k	2,18	18,93	39,91	81,72	89,81	912,60
1	w_{k_raking}	2,77	24,16	50,96	101,87	110,00	2068,21
2	w_{k_raking}	3,44	24,37	51,50	101,61	107,79	1942,61
2	w_{k_lineal}	3,45	24,38	51,49	101,72	108,28	1873,84
2	w_{k_logit}	3,44	24,38	51,49	101,61	107,78	1926,16

Fuente: Elaboración propia

Como se puede observar, los pesos *g* del esquema de calibración 1 tienen valores más dispersos alrededor de 1, en especial su valor máximo, el cual es alrededor de un 40% más grande que para el esquema de calibración 2.

A continuación se presenta el resultado de los criterios propuestos por Silva (2004) para los esquemas de calibración de la base *f1_personas*. M^*_1 hace referencia al esquema de calibración región natural por área, mientras que M^*_2 hace referencia al esquema de calibración provincial.

Tabla 3 Comparativa criterios de evaluación de la calibración, base f1_personas.

Criterio	Esquema de calibración 1		Criterio	Esquema de calibración 2			
	n	Raking		n	Raking	Lineal	Logit
M1	288	5,97E-9	M1_1	42	5,23E-04	717,2	6,46E-05
			M1_2	150	1,75E-04	363,5	1,89E-05
M2	288	7,26%	M2_1	42	3,28%		
			M2_2	145	4,93%		
M3	---	---	M3	---	---	---	---
M4	200	0,12%	M4	0	0%	0%	0%
M5	167906	2,56E-14	M5	16874 7	2,03E-15	7,23E-15	7,02E-15
M6	167906	8,5089	M6	16874 7	7,2679	7,3211	7,2674

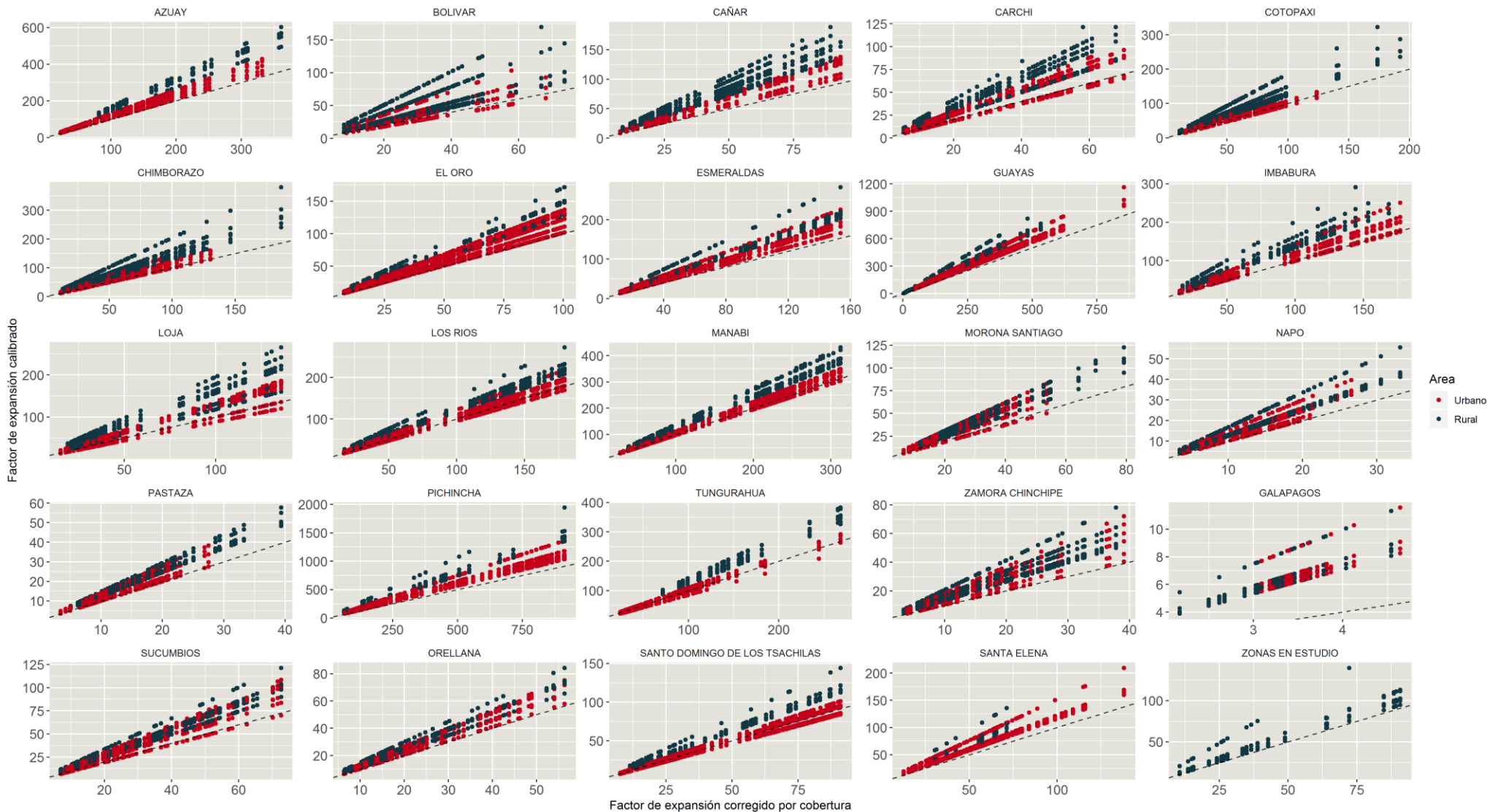
Fuente: Elaboración propia

Como se observa en la Tabla 3, el esquema de calibración 2 es más eficiente que el esquema de calibración 1 pues tiene un coeficiente de variación del estimador de HT relativo promedio más bajo (M2), 0% de g extremos superiores (M4) y menor distancia entre los pesos de calibración y los pesos originales (M6).

Con respecto a las diferentes funciones de distancia evaluadas en el esquema de calibración 2, sus resultados son muy parecidos, sin existir una verdadera ganancia entre una y otra, por lo que se decide utilizar la distancia Multiplicativa o Raking.

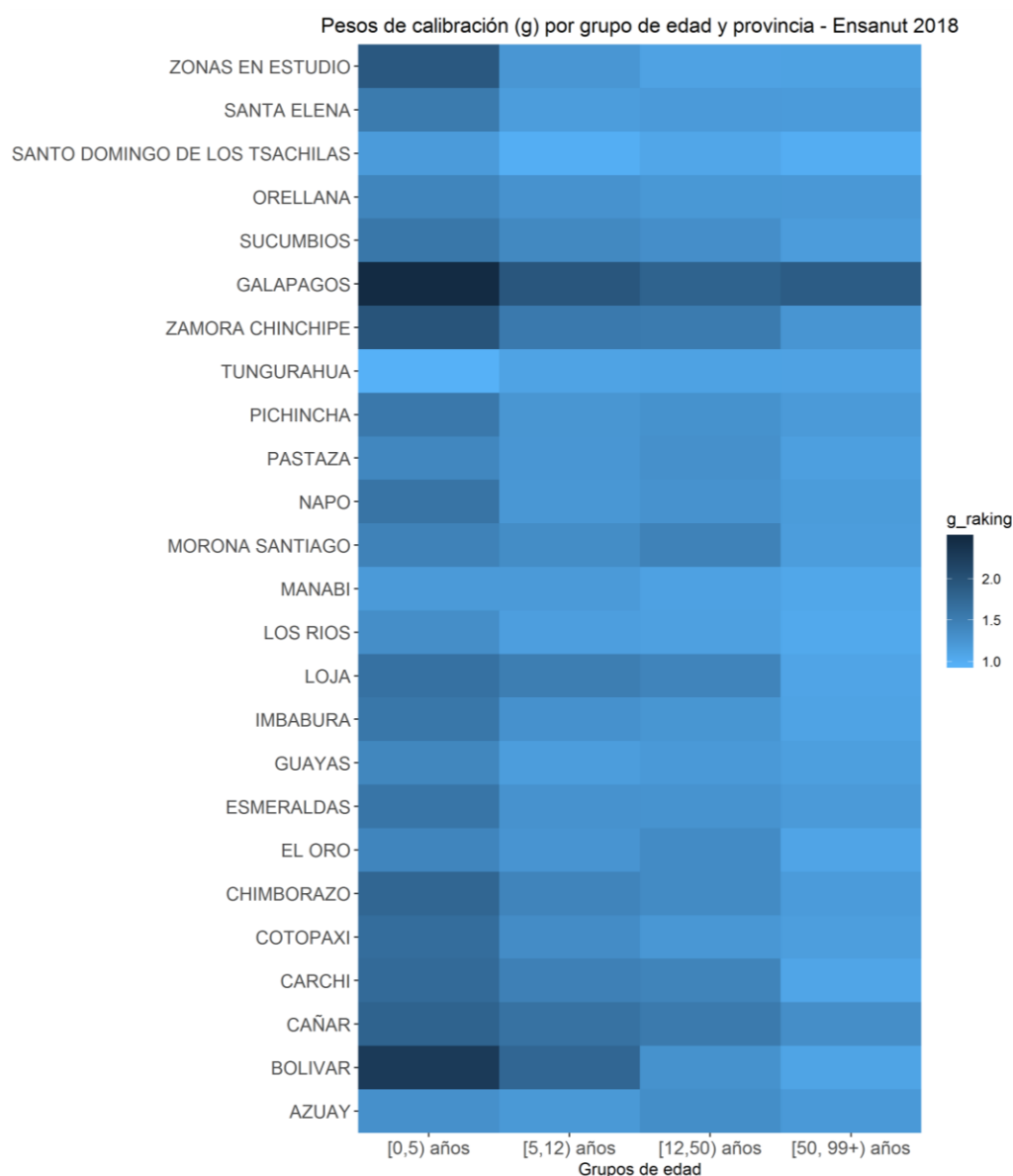
A continuación se presentan algunos gráficos descriptivos acerca de la calibración escogida para la base f1_personas.

Comparación de los factores de expansión corregidos por cobertura vs calibrados de ENSANUT 2018, b1_personas



Distribución de los pesos de calibración (g) - ENSANUT 2018, base f1_personas





Como se aprecia en el gráfico anterior, los pesos de calibración más altos se encuentran en el grupo de edad [0, 5) años, siendo especialmente alto en las provincias de Bolívar, Galápagos y Zamora Chinchipe. Esto sugiere un distanciamiento significativo entre los totales poblacionales estimados desde ENSANUT 2018 y las proyecciones poblacionales vigentes (Este distanciamiento es pronunciado para todos los grupos de edad de la provincia de Galápagos).

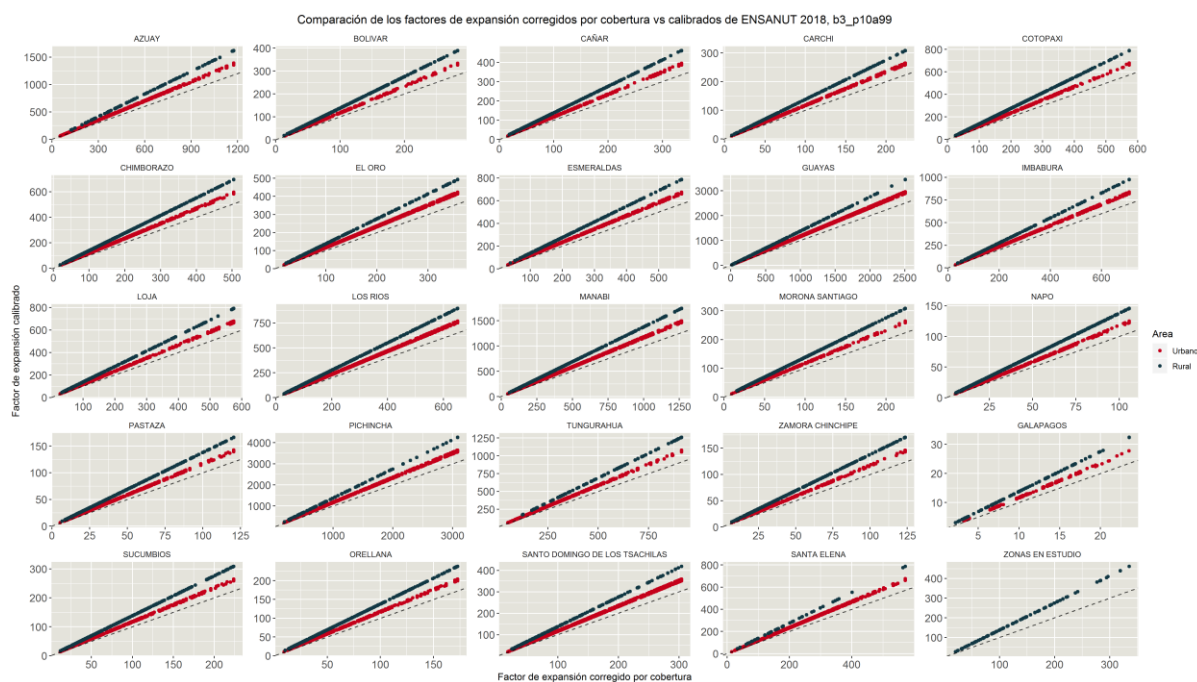
3.2 Base 3 f1_etiquetado

Esquema de calibración

Para la *f1_etiquetado* se consideró el siguiente esquema de calibración.

- Nacional por área: se calculan los totales marginales a nivel nacional por área, sexo y un grupo de edad, [12,100+) años.

A continuación se presenta la comparación gráfica entre los factores de expansión corregidos por cobertura versus los factores de expansión calibrados para la base *f1_etiquetado*.



3.3 Base 4 *f2_mef*

Esquema de calibración

Para la base *f2_mef* se consideraron dos esquemas de calibración:

- Provincial por área: se calculan los totales marginales de mujeres por provincia, área y un grupo de edad, [12,50) años.
- Nacional por área: se calculan los totales marginales de mujeres a nivel nacional por área y 8 grupos de edad quinquenales: [10, 15) años, [15, 20) años, [20, 25) años, [25, 30) años, [30, 35) años, [35, 40) años, [40, 45) años y [45, 50) años.

Para los dos esquemas se realizó una Postestratificación Completa, es decir, se introdujo la información auxiliar en forma de totales conocidos en cada uno de los cruces de las variables provincia, área, sexo y grupo de edad.

Evaluación de la calibración

Se empezará mostrando un análisis descriptivo de los pesos *g* y de los factores de expansión iniciales y finales para los esquemas de calibración considerados.

Tabla 4 Descriptivos pesos g, base f2_mef.

Esquema calibración	Mínimo	Primer cuartil	Mediana	Media	Tercer cuartil	Máximo
1	0,7315	1,1052	1,2328	1,2689	1,3141	2,3009
2	1,0980	1,2150	1,2490	1,2550	1,3020	1,4960

Fuente: Elaboración propia

Tabla 5 Descriptivos factores de expansión iniciales y finales, base f2_mef.

Esquema calibración	Mínimo	Primer cuartil	Mediana	Media	Tercer cuartil	Máximo
Iniciales	2,20	20,16	42,43	88,08	96,60	1021,56
1	2,89	24,90	52,73	108,62	118,31	1375,09
2	2,86	25,19	53,35	108,32	117,96	1479,36

Fuente: Elaboración propia

Como se puede observar, los pesos g del esquema de calibración 1 tienen valores más dispersos alrededor de 1, en especial sus valor máximo, el cual es alrededor de un 54% más grande que para el esquema de calibración 2. Además, el esquema de calibración 2 no tiene ningún g menor a 1.

A continuación se presenta el resultado de los criterios propuestos por Silva (2004) para los esquemas de calibración de la base f2_mef.

Tabla 6 Comparativa criterios de evaluación de la calibración, base f2_mef.

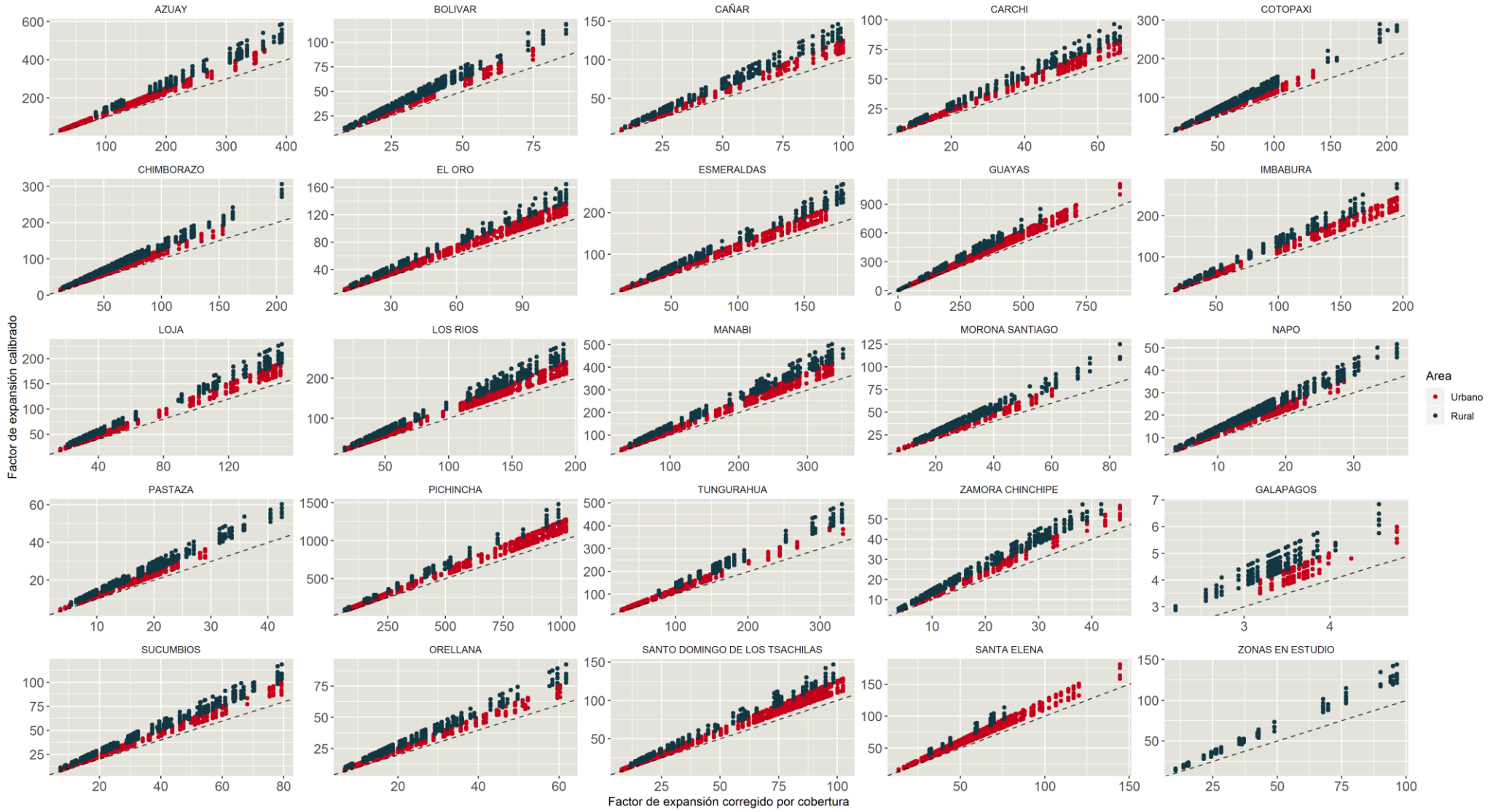
Criterio	Esquema de calibración 1		Esquema de calibración 2	
	n	Raking	n	Raking
M1	48	1,32E-08	16	5,28E-07
M2	48	6.10%	16	3.41%
M3	---	---	---	---
M4	0	0%	0	0%
M5	48456	3,88E-15	48700	1,43E-14
M6	48456	11.73	48700	10.01

Fuente: Elaboración propia

Como se puede observar en la Tabla 6, ninguno de los esquemas de calibración presentan pesos g extremos inferiores o superiores (M3 y M4) y la diferencia en M1 es marginal. Sin embargo, el esquema de calibración 2 es más eficiente que el esquema de calibración 1 pues tiene un coeficiente de variación del estimador de HT relativo promedio más bajo (M2) y menor distancia entre los pesos de calibración y los pesos originales (M6).

A continuación se presenta la comparación gráfica entre los factores de expansión corregidos por cobertura versus los factores de expansión calibrados para la base f2_mef.

Comparación de los factores de expansión corregidos por cobertura vs calibrados de ENSANUT 2018, b4_m10a49



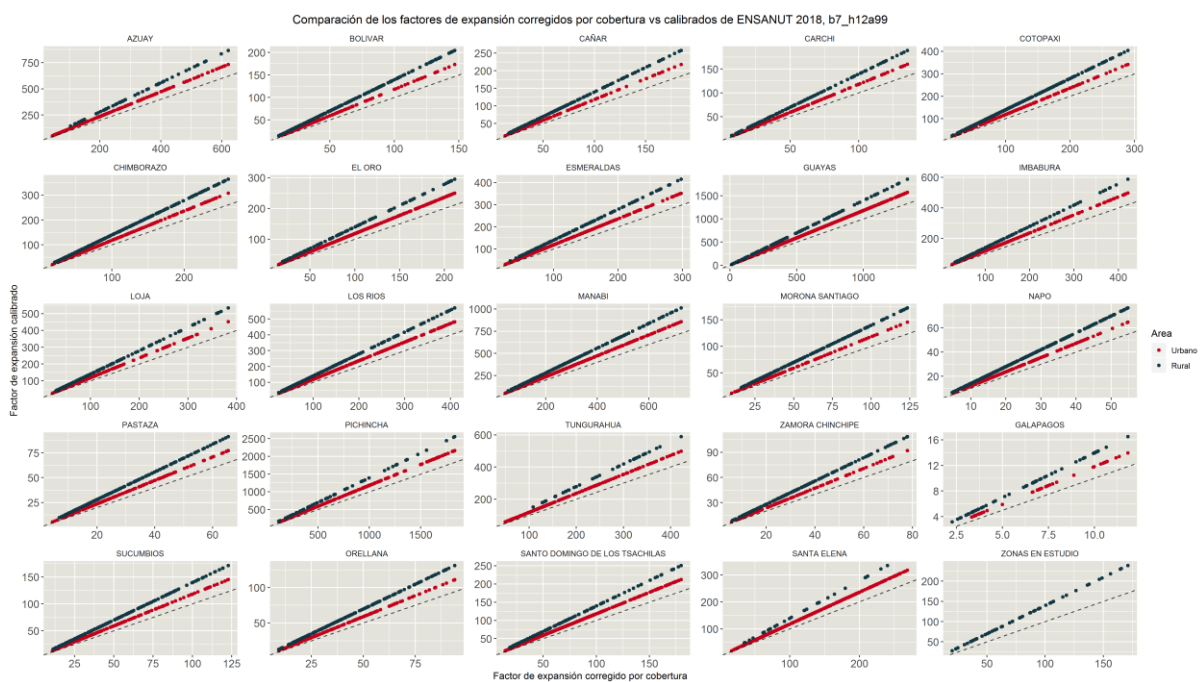
3.4 Base 7 f3_ssr_hombres

Esquema de calibración

Para la *f3_ssr_hombres* se consideró el siguiente esquema de calibración.

- Nacional por área: se calculan los totales marginales de hombres a nivel nacional por área y un grupo de edad, [12,100+) años.

A continuación se presenta la comparación gráfica entre los factores de expansión corregidos por cobertura versus los factores de expansión calibrados para la base *f3_ssr_hombres*.



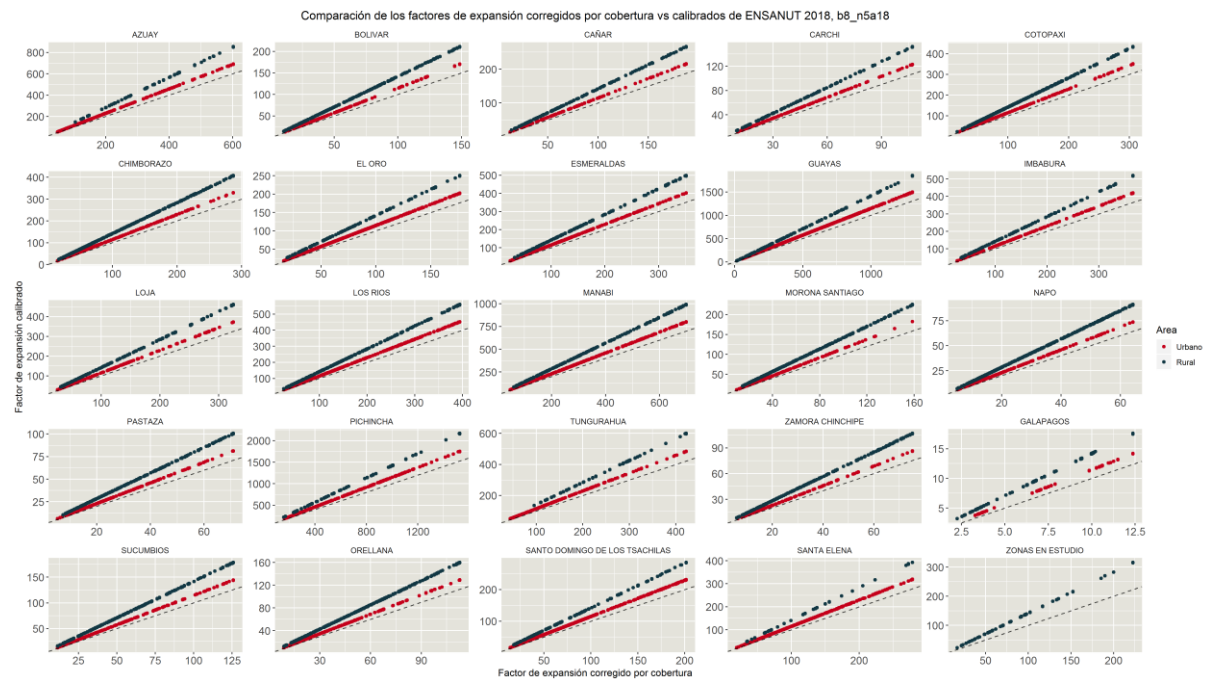
3.5 Base 8 f4_fact_riesgo

Esquema de calibración

Para la *f1_etiquetado* se consideró el siguiente esquema de calibración.

- Nacional por área: se calculan los totales marginales a nivel nacional por área, sexo y un grupo de edad, [15,18] años.

A continuación se presenta la comparación gráfica entre los factores de expansión corregidos por cobertura versus los factores de expansión calibrados para la base *f4_fact_riesgo*.



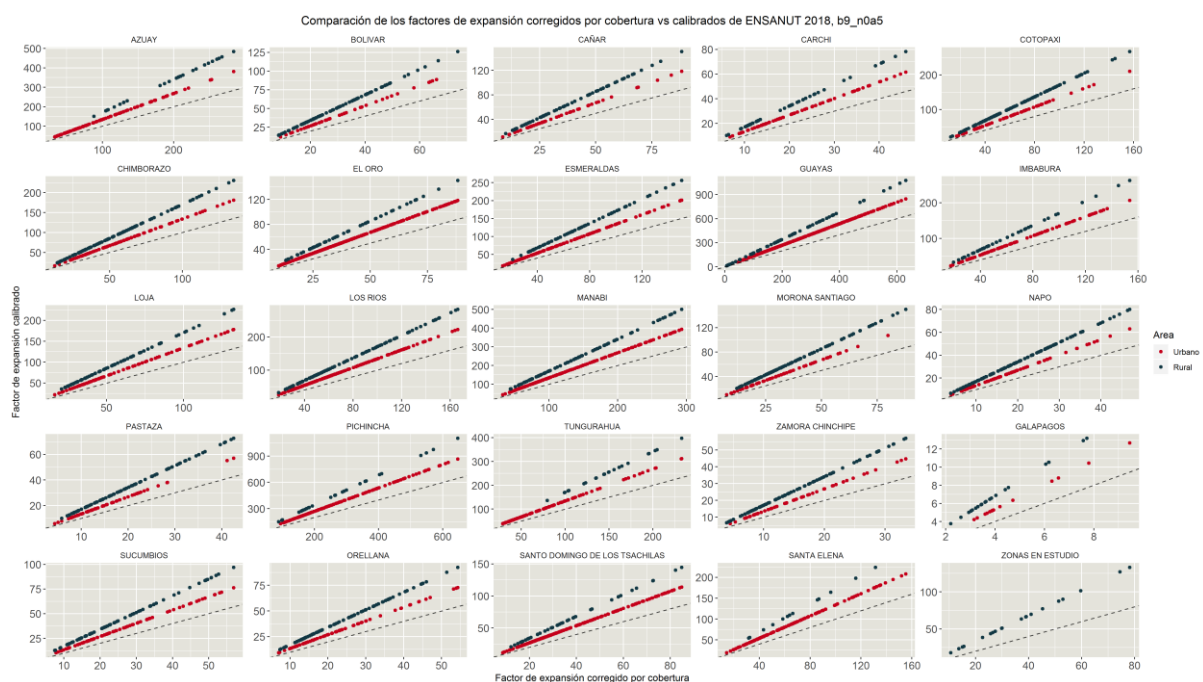
3.6 Base 9 f5_desarrollo_infantil

Esquema de calibración

Para la $f1_etiquetado$ se consideró el siguiente esquema de calibración.

- Nacional por área: se calculan los totales marginales a nivel nacional por área y un grupo de edad, [0, 5) años.

A continuación se presenta la comparación gráfica entre los factores de expansión corregidos por cobertura versus los factores de expansión calibrados para la base $f5_desarrollo_infantil$.



4 BIBLIOGRAFÍA

Deville J.C., Särndal C.E. y Sautory O. (1993). *Generalized Raking Procedures in Survey Sampling*. Journal of the American Statistical Association.

Silva A. (2004). *Calibration Estimation: When and Why, How much and How*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística - IBGE.

Tillé Y. y Matei A. (2016). *sampling: Survey Sampling*. R package version 2.8.

Valliant R., Dever J.A. y Kreuter F. (2018). *Practical Tools for Designing and Weighting Survey Samples*. Springer International Publishing.