www.ecuadorencifras.gob.ec

Diseño Muestral

ENEMDU

Septiembre, 2024

INEC | Buenas cifras, mejores vidas

Dirección

Dirección de Infraestructura Estadística y Muestreo

Elaborado por:

Pablo Peñafiel
Felix Encalada

Revisado por:

Christian Garcés

Aprobado por:

Christian Garcés
Julio Muñoz

Quito -Ecuador, 2024

Contenido

Introducción 5
1. Antecedentes 8
Determinación del Marco de Muestreo para encuestas de hogares 8
Agregación de encuestas (trimestral y anual) 8
2. Características del diseño muestral de la ENEMDU 12
Universo de investigación 12
Unidad de observación 12
Unidad de análisis 12
Cobertura geográfica 12
Tipo de muestreo 12
Dominios de estudio y representatividad: 12
3. Tamaño y selección de la muestra de la ENEMDU 14
Tamaño de la muestra 14
Tamaño de muestra de personas 15
Tamaño de muestra de viviendas 16
Tamaño de muestra de UPM 16
Asignación de la muestra 18
Selección de la muestra 19
Rotación de la muestra por paneles 19
Cobertura de viviendas planificada y efectiva 21
Cobertura planificada 21
Cobertura efectiva 22
4. Cálculo de los factores de expansión 24
Factor de expansión teórico: 24
Probabilidad de inclusión de primera etapa (UPM) 24
Probabilidad de inclusión de segunda etapa (viviendas) 25
Factor de expansión teórico o de diseño 25
Factor de expansión ajustado por cobertura 25
Ajuste por cambio de ocupación (no elegibilidad) 25
Ajuste por nadie en casa (elegibilidad desconocida) 26
Ajuste por rechazo (no respondientes) 26
Recorte de los factores de expansión extremos 27
Validación del recorte de los factores de expansión extremos 29
Porcentaje de factores de expansión recortados por celda o post estrato de calibración 29
Error cuadrático medio (MSE) del recorte de los factores de expansión 29
Calibración de los factores de expansión 30

Estimador de calibración 31
Validación de la calibración de los factores de expansión 32
5. Estimaciones de características 35
Estimación de características de la población: 35
Estimación de errores: 35
Métodos de estimación de errores para diseños muestrales complejos: 36
Referencias 38
Anexos 39

Lista de tablas

Tabla 1. Parámetros utilizados para el cálculo del tamaño de muestra ENEMDU 2021-2024 15
Tabla 2. Tamaños de muestra ENEMDU mensual 17
Tabla 3. Tamaños de muestra ENEMDU trimestral 17
Tabla 4. Tamaños de muestra ENEMDU anual 17
Tabla 5. Asignación de la muestra de viviendas y UPM por territorio 19
Tabla 6. Distribución de la cobertura planificada de viviendas por estrato socioeconómico 21
Tabla 7. Variables requeridas para declaración del diseño muestral – ENEMDU 37

Lista de gráficos

Gráfico 1. Esquema de rotación de paneles para la ENEMDU 2021-2024 20
Gráfico 2. Diagramas de caja y bigote de los factores de expansión ajustados por cobertura de la ENEMDU Septiembre 2024. 28
Gráfico 3. Diagramas de caja y bigote del factor de expansión recortados de la ENEMDU Septiembre de 2024 28
Gráfico 4. Comparación de los factores de expansión ajustados por cobertura y recortados por dominio y grupo de edad. 29
Gráfico 5. Comparación de los factores de expansión recortados y calibrados por dominio y grupo de edad. 33

Introducción

El Instituto Nacional de Estadística y Censos (INEC) ejecuta la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) desde 1993. La ENEMDU, es una encuesta de aplicación continua, la cual se realiza los doce meses del año. La información generada de la encuesta sirve de insumo al gobierno para la planificación del desarrollo nacional y su correspondiente monitoreo y evaluación, así como al sector privado y sociedad civil en general para su conocimiento y toma de decisiones.

El Instituto Nacional de Estadística y Censos (INEC), como ente rector del Sistema Estadístico Nacional y dentro de su plan de mejoramiento continuo de las operaciones estadísticas por muestreo, mediante la Dirección de Infraestructura Estadística y Muestreo (DINEM), busca innovar la metodología de los procesos estadísticos que históricamente se han llevado a cabo dentro de la institución.

Así, desde 2018, se plantean algunas mejoras puntuales al diseño muestral de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU). La técnica de muestreo de la ENEMDU es similar a la utilizada en años anteriores y corresponde a un muestreo probabilístico en dos etapas, con estratificación geográfica por dominios de estudio y área urbana y rural. Los estimadores asociados al diseño se calibran por una proyección de población calculada según métodos demográficos. Sin embargo, el nuevo diseño contempla la afinación del marco de muestro que incluye la mejora en los siguientes aspectos:

- Equilibrio de las Unidades Primarias de Muestreo (UPM).
- Estratificación acorde a las UPM equilibradas.
- Optimización de la dispersión de la muestra.

Los cambios implementados tienen el objetivo de mejorar la precisión de los estimadores y oportunidad de la información.

Por otro lado, en el segundo semestre del año 2020, Comisión Económica para América Latina y el Caribe (CEPAL), mediante un trabajo conjunto con el INEC, realizaron un rediseño de la ENEMDU en lo referente a cálculo del tamaño y selección de la muestra, esquema de rotación de paneles y cálculo de los factores de expansión. Estos aspectos se encuentran detallados en el informe de misión de asistencia técnica que fue entregado por CEPAL al INEC en enero de 2021, el cual se titula "Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024)".

En este documento se expone de forma clara y detallada el proceso de diseño muestral de la ENEMDU, referente al cálculo del tamaño y selección de muestra, cálculo de los factores de expansión, estimación de características, errores muestrales, agregación de encuestas, etc.

1. Antecedentes

1. Antecedentes

Determinación del Marco de Muestreo para encuestas de hogares

El Marco de Muestreo se basa en los resultados definitivos de la cartografía del VII Censo de Población y VI de Vivienda del 2010 (CPV-2010); este Marco tiene la característica de ser constituido por áreas geográficas, que tienen límites perfectamente definidos e identificables sobre el terreno.

El Marco ha tenido actualizaciones parciales en función a los cambios presentados en las unidades de observación (viviendas) en determinadas áreas geográficas, debido principalmente a los movimientos y dinámica demográfica propios de la población, así como los cambios en estructuras habitacionales; que en su conjunto hacen necesario un mantenimiento continuo del Marco de Muestreo.

Se han ejecutado, en el periodo 2014-2017, diferentes procesos de actualización cartográfica con diferente alcance, los cuales se detallan a continuación:

- Encuesta Condiciones de Vida 2013-2014: 2.425 sectores censales.
- Actualización ENEMDU 2014: 5.564 sectores censales.
- Proyecto 2015: 548 sectores censales.
- Actualización 2017: 1.779 sectores censales.

El Marco de Muestreo para encuestas de hogares se lo ha dividido por dominios de estudio, y dentro de ellos sus correspondientes UPM a las cuales se asignó un estrato tomando principalmente sus características geográficas, socio-económicas y socio-demográficas con la finalidad de mejorar la precisión y exactitud de los estimadores, minimizando su varianza.

Agregación de encuestas (trimestral y anual)

Tal y como se explica en CEPAL (2021), en base al esquema rotativo 2(2)2 y al balance en el tamaño muestral mensual de la ENEMDU en el periodo 2021-2024, la encuesta puede agregarse de manera trimestral y anual. Con la agregación la ENEMDU gana en dominios de representatividad, permitiendo la desagregación de la información para diferentes dominios geográficos con un nivel de precisión aceptable para cada uno de ellos.

Según CEPAL (2019) estas diferencias en el diseño hacen que se pueda aprovechar al máximo la capacidad inferencial de la encuesta al momento de considerar las siguientes innovaciones metodológicas:

1. Es posible considerar trimestres móviles en la estimación de las cifras concernientes al mercado de trabajo para que mes a mes el Ecuador reciba de forma oportuna cifras confiables del mercado de trabajo en el ámbito nacional, urbana, rural y por las cinco ciudades principales. Son muchos los Institutos Nacionales de Estadística alrededor del mundo que han optado por reportar este tipo de estimaciones basadas en trimestres móviles. En particular:

    a. Las estimaciones serán mucho más suaves y no se verán afectadas significativamente por situaciones coyunturales ajenas al mercado de trabajo; por ejemplo, desastres naturales, oleadas migratorias transitorias, etc.

    b. Además, si un fenómeno exógeno afectara la estructura del mercado de trabajo, la afectación en las cifras de trabajo se daría de forma paulatina en los siguientes dos meses; de tal forma que la dinámica cambiante del mercado de trabajo sería incorporada fácilmente con esta metodología.

    c. La agregación computacional de los trimestres móviles es inmediata en cuanto las muestras dentro cualquier combinación de trimestres móviles son independientes, puesto que el panel es 2(2)2 y la encuesta es continua. De esta forma, los procesos de estimación no tendrían una carga importante en cuanto a procesamiento estadístico.

    d. En cualquier mes, se podrían hacer análisis más profundos del mercado de trabajo, como los referidos a la estimación de cambios brutos.

2. Al realizar la agregación de todos los meses del año, es posible realizar la estimación de los fenómenos del mercado de trabajo y pobreza con representatividad, nacional, urbano, rural, por las cinco ciudades principales y además por las 24 provincias del Ecuador.

    a. Las estimaciones ganarán en confiabilidad, puesto que en las agregaciones anuales se tiene una mayor muestra que redunda en que las medidas de precisión de las estimaciones sean menores, haciéndose más angosto su intervalo de confianza.

    b. Al hacer comparaciones entre dos años, la potencia estadística se incrementará sustancialmente, reduciendo la probabilidad de cometer el Error Tipo II (aceptar que no hay diferencias entre las cifras del empleo cuando en realidad sí las hay).

c. La agregación anual trae algunos retos computacionales puesto que, debido al esquema de rotación 2(2)2, aparecerán algunos términos de correlación en las UPM que deben ser considerados a la hora de realizar los cómputos del error de muestreo.

d. Las estimaciones interanuales de los cambios brutos a través de las matrices de transición serán más robustas y consistentes, puesto que la fracción de muestreo en las agregaciones anuales es más grande que en el esquema pasado.

Para el periodo 2021-2024, el INEC tiene planificado producir estadísticas oficiales a partir de la ENEMDU agregada trimestral 4 veces al año, una por cada trimestre, y una vez al año para la ENEMDU agregada anual.

2. Características del diseño muestral de la ENEMDU

2. Características del diseño muestral de la ENEMDU

Universo de investigación

El universo de estudio de la ENEMDU son personas de 5 y más años de edad, residentes en las viviendas del Ecuador, exceptuando la población que reside en viviendas colectivas, viviendas flotantes y población indigente (sin techo).

Unidad de observación

La unidad de observación son todas las viviendas particulares ocupadas que se encuentran en territorio nacional, mismas que tienen ligada su identificación geográfica mediante fuentes cartográficas.

Unidad de análisis

Para el caso de los indicadores laborales, la población de referencia son todas las personas mayores o iguales a 15 años.

Cobertura geográfica

La cobertura geográfica está definida por las viviendas ocupadas que se encuentren ubicadas dentro del territorio ecuatoriano incluyendo la región insular.

Tipo de muestreo

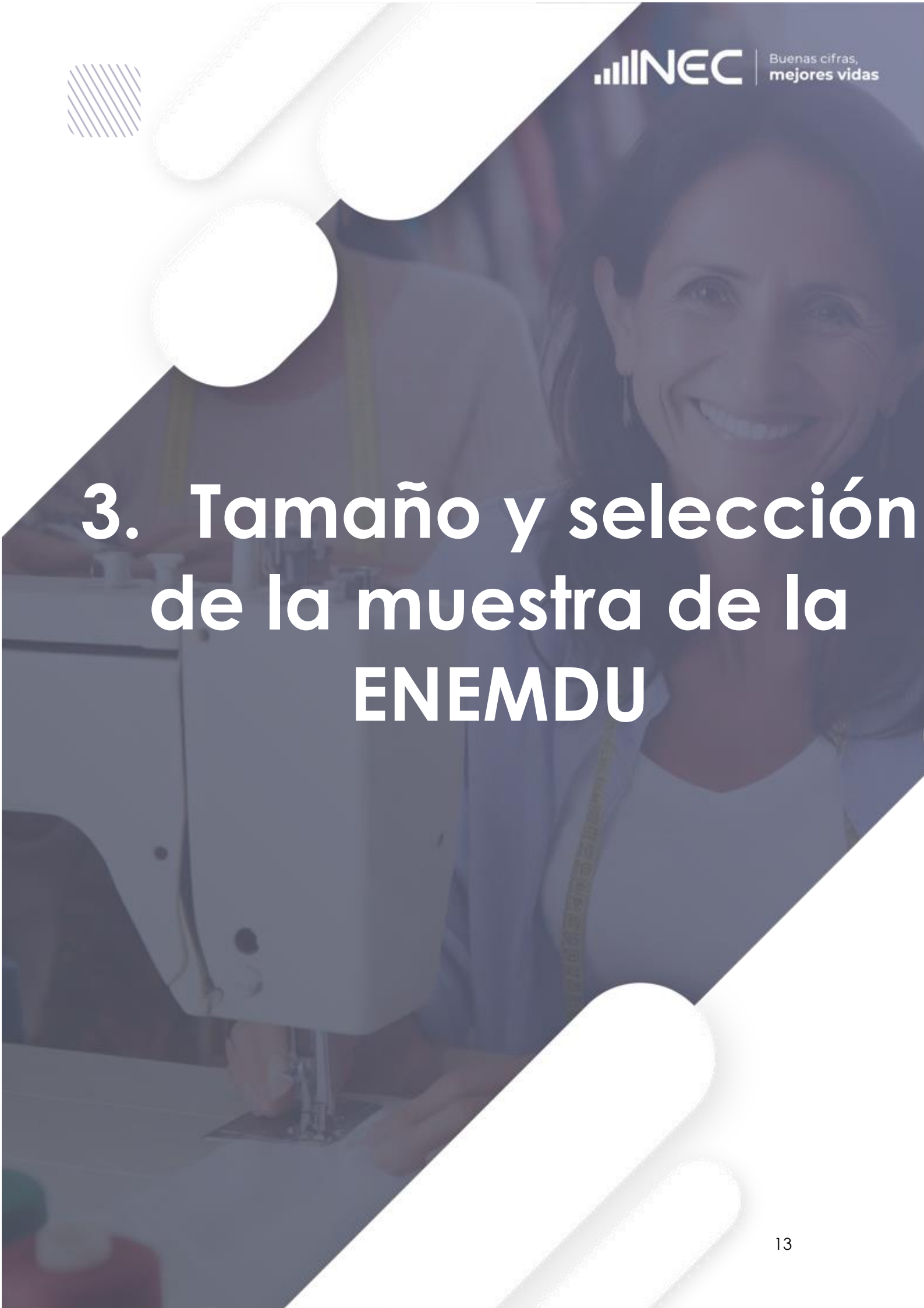
El tipo de muestreo de la ENEMDU corresponde a un muestreo probabilístico estratificado bietápico de elementos.

Dominios de estudio y representatividad:

a) **ENEMDU mensual:** La ENEMDU mensual tiene como sus dominios de diseño y representatividad Nacional, Urbano-Rural.

b) **ENEMDU trimestral:** La ENEMDU trimestral tiene como dominios de diseño y representatividad Nacional, Urbano-Rural y 5 ciudades principales (Quito, Guayaquil, Cuenca, Machala y Ambato).

c) **ENEMDU anual:** La ENEMDU anual tiene como dominios de diseño y representatividad Nacional, Urbano-Rural, 5 ciudades principales (Quito, Guayaquil, Cuenca, Machala y Ambato) y 24 provincias del Ecuador.



3. Tamaño y selección de la muestra de la ENEMDU

3. Tamaño y selección de la muestra de la ENEMDU

Tamaño de la muestra

El tamaño de muestra por dominio de la ENEMDU fue calculado considerando los siguientes parámetros:

- **N:** número de personas en cada dominio en el Marco de Muestreo.
- **M:** número de UPM en cada dominio en el Marco de Muestreo.
- **R:** porcentaje de población económicamente activa (PEA) en cada dominio. Calculado a partir de la ENEMDU anual 2019.
- **B:** promedio de personas por hogar en cada dominio, calculado a partir de la ENEMDU anual 2019.
- **Rho (ρ):** coeficiente de correlación intraclase de la tasa de desempleo para cada uno de los dominios, valor que es calculado a partir de la información de la ENEMDU anual 2019.
- **P:** tasa de desempleo en cada dominio, estimado a partir de la ENEMDU anual 2019.
- **A:** amplitud del intervalo de confianza.
- **Delta (δ):** margen de error relativo asociado a la tasa de desempleo en cada dominio. Este valor se calcula como la mitad del ancho del intervalo (A) de confianza dividido para la tasa de desempleo -estimada a partir de la ENEMDU anual 2019- y este resultado elevado al cuadrado.
- **1-α:** nivel de confianza del 95%.
- **TNR:** tasa de no respuesta del 20%, valor calculado mediante la realización de un análisis histórico de la cobertura de la encuesta en los periodos 2018, 2019 y 2020.

Para cada uno de los cálculos y escenarios de tamaño de muestra, la confiabilidad estadística para la nueva ENEMDU 2021 – 2024 se fija en 95%, además, el número de viviendas seleccionadas dentro de cada una de las UPM se sigue manteniendo en 7. Los parámetros utilizados para el cálculo de muestra en cada uno de los dominios para la encuesta se presentan en la Tabla 1:

Tabla 1. Parámetros utilizados para el cálculo del tamaño de muestra ENEMDU 2021-2024

Dominio	N	M	R	B	Rho	P	A
Azuay	450.652	2.643	54,90%	4,92	0,0440	2,25%	1,60%
Bolívar	224.270	1.279	51,30%	3,97	0,0170	1,31%	1,60%
Cañar	300.749	1.596	51,90%	4,33	0,0300	3,67%	1,60%
Carchi	192.573	1.288	48,60%	3,72	0,0440	4,60%	1,80%
Cotopaxi	514.752	2.807	58,10%	4,25	0,0190	1,83%	1,60%
Chimborazo	546.499	3.196	56,10%	4,52	0,0240	1,59%	1,60%
El Oro	428.880	2.862	48,60%	4,16	0,0200	4,35%	1,60%
Esmeraldas	627.896	3.392	40,00%	4,73	0,0700	10,17%	1,80%
Guayas	1.674.002	10.112	43,60%	5,51	0,0510	3,38%	1,60%
Imbabura	496.033	3.076	46,10%	4,26	0,0780	6,11%	1,60%
Loja	547.507	3.194	53,20%	4,54	0,0780	3,85%	1,60%
Los Ríos	898.647	5.459	44,30%	4,85	0,0410	2,89%	1,60%
Manabí	1.554.229	9.198	45,60%	5,40	0,0730	2,47%	1,60%
Morona Santiago	195.141	975	47,40%	4,54	0,0510	1,83%	1,60%
Napo	132.566	701	49,40%	4,36	0,0370	2,92%	1,80%
Pastaza	111.690	617	47,00%	4,05	0,0110	3,07%	1,80%
Pichincha	1.180.351	7.358	48,00%	4,78	0,0480	6,19%	1,60%
Tungurahua	399.561	2.525	60,10%	4,68	0,0130	1,15%	1,60%
Zamora Chinchipe	119.508	669	50,70%	4,13	0,0260	3,32%	1,80%
Galápagos	32.395	348	53,50%	3,22	0,0400	1,82%	1,80%
Sucumbíos	227.509	1.250	43,70%	4,21	0,0330	5,42%	1,80%
Orellana	161.472	982	46,90%	4,32	0,0420	2,69%	1,60%
Sto Domingo de los Tsáchilas	494.320	3.133	44,50%	4,21	0,0620	2,53%	1,60%
Santa Elena	386.346	2.025	40,50%	4,89	0,2000	3,20%	1,80%
Quito	1.950.476	13.811	46,40%	5,66	0,0260	9,12%	1,80%
Guayaquil	2.610.712	16.907	45,40%	6,71	0,0190	3,07%	1,00%
Cuenca	402.447	2.423	44,50%	6,08	0,0380	5,35%	1,80%
Machala	270.241	1.797	43,50%	6,38	0,0380	6,13%	1,80%
Ambato	192.509	1.376	49,50%	5,59	0,0420	4,79%	1,80%

Fuente: CEPAL. (2021). Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024).

Los algoritmos que se utilizaron para el cálculo de los tamaños de muestra tanto de personas, como de viviendas y UPM, se detallan a continuación.

Tamaño de muestra de personas

La expresión matemática utilizada para el cálculo de tamaño de muestra de personas en cada dominio de diseño D es:

$$n_D \geq \frac{P(1-P)\,Deff}{\frac{\delta^2 P^2}{z_\alpha^2} + \frac{P(1-P)\,Deff}{N_D}} * \frac{1}{1-T_{NR}}$$

donde:

z = $percentil\ de\ la\ distribución\ normal\ estándar\ asociado\ al\ nivel\ de\ confianza\ 1-\alpha$

δ = $margen\ de\ error\ relativo\ máximo.$

P = $estimación\ de\ la\ variable\ de\ diseño\ (tasa\ de\ desempleo).$

N_D = $tamaño\ de\ la\ población\ en\ cada\ dominio\ de\ diseño\ D.$

T_{NR} = $tasa\ de\ no\ respuesta.$

El efecto de diseño $Deff$, definido como una función de la correlación existente entre la variable de interés (desempleo) y la conformación de las UPM, está dado por la siguiente expresión:

$$Deff \approx 1 + (\bar{n} - 1) * \rho$$

Donde, $\bar{n}$ es el número promedio de personas de la población económicamente activa (PEA) que serán encuestadas y ρ es la correlación intraclase entre el desempleo y la conformación de las UPM.

A su vez, $\bar{n}$ es calculada a través del siguiente algoritmo:

$$\bar{n} = 7 * r * b$$

Donde, el número 7 se refiere al número de viviendas investigadas en cada UPM, r es el porcentaje de población económicamente activa (PEA) y b es el promedio de personas por hogar.

Por último, cabe señalar que la tasa de desempleo, la cual varía dependiendo del dominio de diseño, con la finalidad de controlar el ancho del intervalo de confianza generado a partir de la encuesta. Es por esta razón que se fijó el error máximo relativo como función de la amplitud del intervalo de confianza A. Por lo tanto, se tiene que:

$$\delta = \left(\frac{A}{2}\right)^2$$

Tamaño de muestra de viviendas

El número de viviendas que deben ser seleccionadas estará determinado por la muestra de personas (n_D), número promedio de personas por vivienda (b) y el porcentaje de personas que presentan la característica de interés ($r = Proporción\ de\ la\ PEA$), de la siguiente forma:

$$n_{vD} = \frac{n_D}{r * b}$$

Tamaño de muestra de UPM

Las viviendas y las personas que participan en la encuesta forman parte de UPM previamente seleccionadas. En este paso final, es necesario calcular el número de UPM que deben ser seleccionadas en la primera etapa de muestreo a partir de la relación:

$$n_{UPM_D} = \frac{n_{vD}}{Carga\ técnica\ operativa}$$

La carga técnica operativa se refiere al número de viviendas asignadas a cada encuestador como carga de trabajo; que fue el resultado de un previo análisis

de correlación intraclase donde se pudo verificar, mediante simulaciones matemáticas, el número de observaciones necesarias para minimizar la varianza dentro de cada UPM. Como resultado de este procedimiento se definió tanto operativa como técnicamente que el número de viviendas a investigarse por UPM será 7.

Considerando las restricciones presupuestarias presentes, y luego de aplicar los algoritmos de cálculo correspondientes, se obtiene un tamaño de muestra de 9.016 viviendas mensuales. Además, al fijar en 7 el número de viviendas levantadas por UPM, el tamaño de muestra es de 1.288 UPM a ser visitadas cada mes en la ENEMDU. La distribución por dominio se muestra a continuación:

Tabla 2. Tamaños de muestra ENEMDU mensual

Dominio	Viviendas del marco	UPM del marco	Muestra UPM	Muestra Viviendas
Urbano	2.715.812	79.647	952	6.664
Rural	1.002.506	27.352	336	2.352
Nacional	**3.718.318**	**106.999**	**1.288**	**9.016**

Fuente: CEPAL. (2021). Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024).

Tabla 3. Tamaños de muestra ENEMDU trimestral

Dominio	Viviendas del marco	UPM del marco	Muestra UPM	Muestra Viviendas
Urbano	2.715.812	79.647	2.856	19.992
Rural	1.002.506	27.352	1.008	7.056
Quito	473.957	13.811	408	2.856
Guayaquil	589.772	16.907	384	2.688
Cuenca	84.623	2.423	276	1.932
Machala	62.658	1.797	312	2.184
Ambato	47.817	1.376	252	1.764

Fuente: CEPAL. (2021). Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024).

Tabla 4. Tamaños de muestra ENEMDU anual

Dominio	Viviendas del marco	UPM del marco	Muestra UPM	Muestra Viviendas
Urbano	2.715.812	79.647	11.424	79.968
Rural	1.002.506	27.352	4.032	28.224
Quito	473.957	13.811	1.632	11.424
Guayaquil	589.772	16.907	1.536	10.752
Cuenca	84.623	2.423	1.104	7.728
Machala	62.658	1.797	1.248	8.736
Ambato	47.817	1.376	1.008	7.056
Azuay	96.736	2.643	288	2.016
Bolívar	46.122	1.279	288	2.016
Cañar	56.745	1.596	288	2.016

Dominio	Viviendas del marco	UPM del marco	Muestra UPM	Muestra Viviendas
Carchi	43.923	1.288	336	2.352
Cotopaxi	100.202	2.807	288	2.016
Chimborazo	114.597	3.196	288	2.016
El Oro	97.861	2.862	384	2.688
Esmeraldas	116.307	3.392	912	6.384
Guayas	348.635	10.112	336	2.352
Imbabura	104.598	3.076	720	5.040
Loja	110.928	3.194	528	3.696
Los Ríos	189.694	5.459	288	2.016
Manabí	320.021	9.198	336	2.352
Morona Santiago	34.563	975	192	1.344
Napo	24.091	701	240	1.680
Pastaza	21.361	617	288	2.016
Pichincha	258.888	7.358	624	4.368
Tungurahua	90.743	2.525	288	2.016
Zamora Chinchipe	23.108	669	288	2.016
Galápagos	8.520	348	192	1.344
Sucumbíos	43.234	1.250	384	2.688
Orellana	33.317	982	288	2.016
Sto. Domingo de los Tsáchilas	107.630	3.133	336	2.352
Santa Elena	67.667	2.025	528	3.696

Fuente: CEPAL. (2021). Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024).

Mediante la agregación trimestral, la ENEMDU alcanza representatividad para las cinco ciudades principales, al establecer intervalos de confianza para la tasa de desocupación muy precisos para las mismas, con márgenes de error absolutos menores al 1%. Mientras que, al considerar su agregación anual, la ENEMDU puede obtener estimaciones para la tasa de desempleo con una precisión aceptable para cada una de las 24 provincias del Ecuador, con márgenes de error absoluto de a lo más 1,8%.

Asignación de la muestra

La asignación del tamaño de muestra en cada uno de los 150 estratos de muestreo se llevó a cabo usando la asignación de Kish (Kish, 1987; Maligalig & Martinez, 2013) que permite distribuir en cada uno de los estratos un tamaño de muestra óptimo, a través del siguiente algoritmo:

$$n_h = n \cdot \frac{\sqrt{\frac{1}{H^2}\left(\frac{N_h}{N}\right)^2}}{\sum_{h=1}^{H}\sqrt{\frac{1}{H^2}\left(\frac{N_h}{N}\right)^2}}$$

donde:

n_h = $Tamaño\ de\ muestra\ para\ el\ estrato\ h.$
N_h = $Número\ de\ UPM\ del\ estrato\ h.$
N = $Número\ de\ UPM\ en\ el\ Marco\ de\ Muestreo.$
n = $Número\ de\ UPM\ seleccionadas\ en\ la\ muestra\ en\ la\ primera\ etapa\ de\ muestreo.$
H = $Número\ de\ estratos\ en\ el\ dominio.$

Tabla 5. Asignación de la muestra de viviendas y UPM por territorio

Subpoblación	Viviendas del marco	UPM del marco	Muestra UPM	Muestra Viviendas
Quito	473.957	13.811	136	952
Guayaquil	589.772	16.907	128	896
Cuenca	84.623	2.423	92	644
Machala	62.658	1.797	104	728
Ambato	47.817	1.376	84	588
Resto Sierra Urbano	592.918	17.565	192	1.344
Resto Costa Urbano	780.055	23.186	144	1.008
Amazonía Urbano	79.861	2.410	64	448
Sierra Rural	538.194	14.530	164	1.148
Costa Rural	360.130	9.862	88	616
Amazonía Rural	99.813	2.784	76	532
Región Insular	8.520	348	16	112
Total	**3.718.318**	**106.999**	**1.288**	**9.016**

Fuente: CEPAL. (2021). Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024).

En la Tabla 5 se presenta la asignación de la muestra por subpoblación en la ENEMDU mensual, cabe recalcar que estos no son dominios de diseño de la encuesta mensual, por cuanto la representatividad de la ENEMDU mensual es Nacional, Urbano-Rural.

Selección de la muestra

La selección de la muestra se realiza en forma aleatoria, en dos etapas:

- **Primera etapa:** selección de Unidades Primarias de Muestreo (UPM) por estrato.
- **Segunda etapa:** selección de viviendas ocupadas dentro de cada una de las UPM seleccionadas en la primera etapa.

La selección de las UPM que forman parte de la muestra se realiza de manera independiente en cada uno de los dominios de forma aleatoria, asignando a cada UPM igual probabilidad de ser seleccionada. De la misma forma, la selección de viviendas es aleatoria dentro de cada UPM seleccionada.

Rotación de la muestra por paneles

El objetivo de realizar una encuesta tipo panel es medir los cambios en los indicadores entre trimestres y años consecutivos, así como la estacionalidad en el tiempo. Un panel representa un conjunto de viviendas dentro de una UPM, las cuales se investigan de acuerdo a una rotación específica durante cierto periodo de tiempo.

El esquema de rotación de paneles para la ENEMDU 2021-2024 presenta una rotación tanto de UPM como de viviendas dentro de las UPM, según CEPAL (2021) esto permitirá que se afiance el conocimiento de la construcción de pesos de muestreo longitudinales y el análisis estadístico de cambios brutos y netos, puesto que es posible establecer una dinámica poblacional de flujos brutos entre trimestres y años. Este esquema de rotación se presenta a continuación:

Gráfico 1. Esquema de rotación de paneles para la ENEMDU 2021-2024

ROTACIÓN DE PANELES ENEMDU 2021-2024													
Año	Trimestre	Mes											
		M1				M2				M3			
2021	T1	a11	b11	c11	d11	e11	f11	g11	h11	i11	j11	k11	l11
	T2	a11	b12	c11	d12	e11	f12	g11	h12	i11	j12	k11	l12
	T3	a12	b12	c12	d12	e12	f12	g12	h12	i12	j12	k12	l12
	T4	a12	b11	c12	d21	e12	f11	g12	h21	i12	j11	k12	l21
2022	T1	a11	b11	c21	d21	e11	f11	g21	h21	i11	j11	k21	l21
	T2	a11	b12	c21	d22	e11	f12	g21	h22	i11	j12	k21	l22
	T3	a12	b12	c12	d22	e12	f12	g12	h22	i12	j12	k12	l22
	T4	a12	b21	c12	d21	e12	f21	g12	h21	i12	j21	k12	l21
2023	T1	a21	b21	c21	d21	e21	f21	g21	h21	i21	j21	k21	l21
	T2	a21	b22	c21	d22	e21	f22	g21	h22	i21	j22	k21	l22
	T3	a22	b22	c31	d22	e22	f22	g31	h22	i22	j22	k31	l22
	T4	a22	b21	c31	d31	e22	f21	g31	h31	i22	j21	k31	l31
2024	T1	a21	b21	c32	d31	e21	f21	g32	h31	i21	j21	k32	l31
	T2	a21	b22	c32	d32	e21	f22	g32	h32	i21	j22	k32	l32
	T3	a22	b22	c31	d32	e22	f22	g31	h32	i22	j22	k31	l32
	T4	a22	b31	c31	d31	e22	f31	g31	h31	i22	j31	k31	l31

Fuente: CEPAL. (2021). Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024).

Como se puede observar en el Gráfico 1 por el lado de las filas se encuentran los años y los trimestres, mientras que las columnas hacen referencia a los meses de cada trimestre, además, es preciso observar que en cada mes de levantamiento la muestra está distribuida en cuatro paneles de rotación. Por ejemplo, en el año 2021, la intersección entre T1 y M1 se refiere al primer mes del primer trimestre del año 2021, en este caso es el mes de enero de 2021, cuya muestra está distribuida en los paneles a11, b11, c11 y d11; en el año 2022, la intersección entre T3 y M2 se refiere al segundo mes del tercer trimestre del año 2022, en este caso es el mes de agosto de 2022, cuya muestra está distribuida en los paneles e12, f12, g12 y h22.

Como se mencionó anteriormente, cada grupo rotacional presenta una rotación tanto de UPM como de viviendas dentro de las UPM, en este sentido, está conformado por un código alfanumérico de un carácter y 2 dígitos, los dos primeros se refieren a la rotación de UPM, mientras que el tercer dígito representa la rotación de viviendas dentro de la UPM. Por ejemplo, el panel de rotación a11 está diseñado de la siguiente manera: "a1" (dos primeras posiciones) es la UPM y "1" (tercer dígito) es el grupo de viviendas dentro de la UPM. Cabe señalar que, para la conformación de la rotación de viviendas, se particionaron las UPM en dos grupos de viviendas.

El esquema de rotación de paneles para la ENEMDU 2021-2024 es un esquema 2(2)2, esto significa que cada grupo rotativo permanece dos períodos consecutivos, sale del estudio por otros dos periodos consecutivos, para volver a aparecer finalmente en los siguientes dos meses consecutivos. Por ejemplo, las viviendas del grupo rotacional "a11" son investigadas en los meses de enero y abril del año 2021, salen del estudio los meses de julio y octubre del año 2021 (debido a que entran las viviendas del panel a12) y vuelven a ser investigadas en los meses de julio y octubre del año 2021.

De igual forma, cabe señalar que en este esquema de rotación de paneles se establece un traslape del 50% de la muestra entre trimestres y años y que el seguimiento se realiza a nivel de viviendas (unidad de observación). Asimismo, es preciso indicar que el tamaño de muestra por cada grupo rotacional dentro de un mes y trimestre específico es de 322 UPM, ejemplificando, los paneles del mes de enero de 2021 (a11, b11, c11 y d11) tienen 322 UPM cada uno, obteniendo un total de 1.288 UPM a ser investigadas cada mes.

Una de las bondades de esta propuesta de rotación es que es compatible con un proceso de actualización cartográfica permanente, donde un panel ingresa con actualización de la condición de ocupación de las viviendas que lo componen, con la finalidad de garantizar la cobertura muestral para la investigación.

Cobertura de viviendas planificada y efectiva

Cobertura planificada

A continuación, se presenta la cobertura planificada de viviendas de la ENEMDU del mes de Septiembre de 2024:

Tabla 6. Distribución de la cobertura planificada de viviendas por estrato socioeconómico

Subpoblación	Estrato Socioeconómico				Total
	Alto (1)	Medio (2)	Bajo (3)	Sin clasificar	
Quito	336	364	252	0	952
Guayaquil	252	420	224	0	896
Cuenca	196	252	196	0	644
Machala	196	308	224	0	728
Ambato	168	224	196	0	588
Resto Sierra Urbano	392	476	476	0	1344
Resto Costa Urbano	308	364	336	0	1008
Amazonía Urbano	112	140	112	84	448
Sierra Rural	392	392	364	0	1.148
Costa Rural	196	196	168	56	616

Amazonía Rural	168	196	168	0	532
Región Insular	56	28	28	0	112
Total	**2.772**	**3.360**	**2.744**	**140**	**9.016**

Fuente: CEPAL. (2021). Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024).

Según la Tabla 6 se puede apreciar que existen viviendas que no están clasificadas en ningún estrato socioeconómico, ya sea en el estrato alto, medio o bajo en la Amazonía Urbana y en la Costa Rural. Este hecho sucede debido a que existen ciertos estratos geográficos, definidos por el cruce de las variables provincia y área que tienen muy pocas UPM, razón por la cual no se clasifican por sus características socioeconómicas. Estos dominios son: Morona Santiago urbana (371 UPM), Napo urbana (292 UPM) y Santa Elena rural (120 UPM).

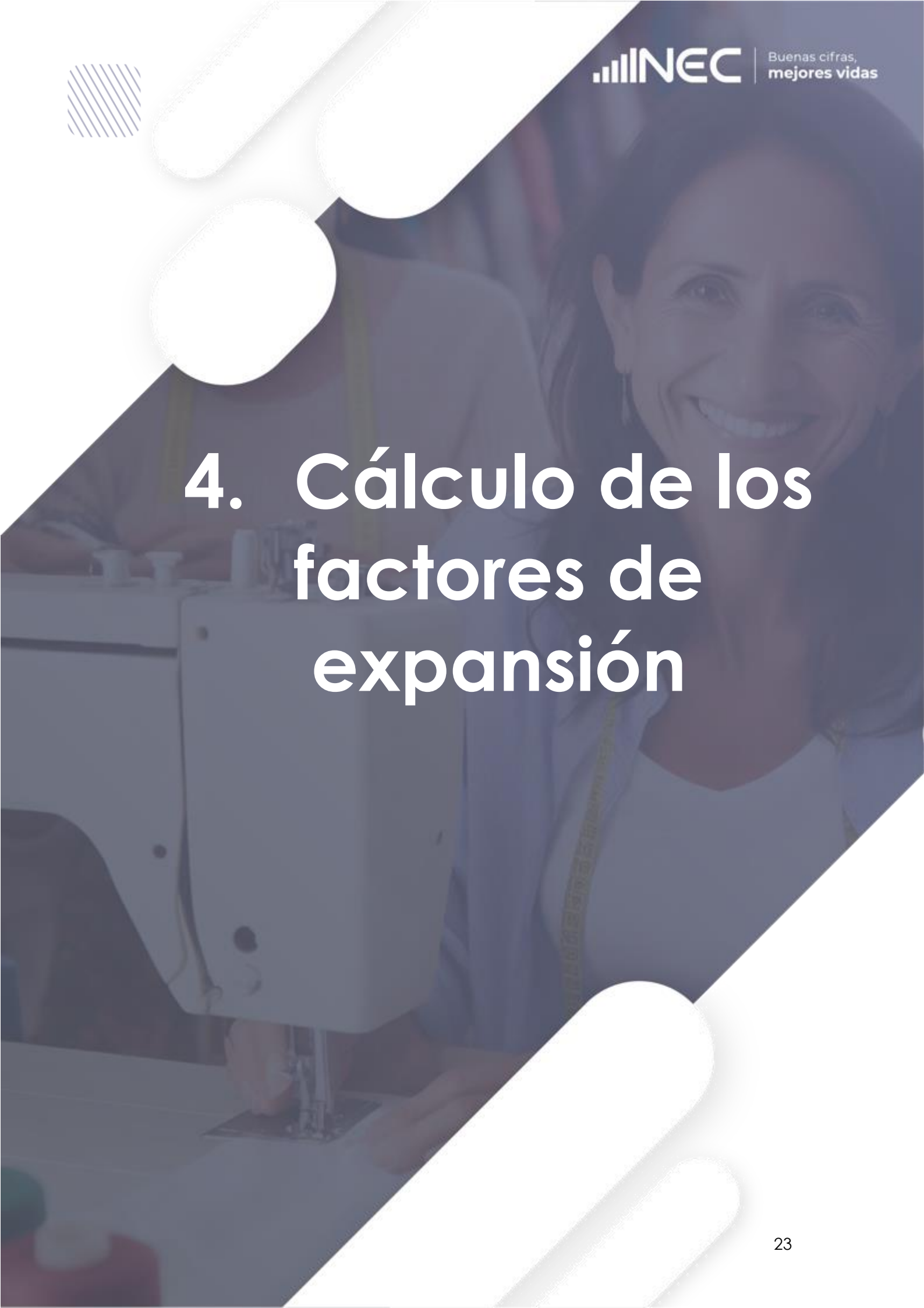
Cobertura efectiva

Por otro lado, la cobertura efectiva de viviendas levantadas en la ENEMDU de Septiembre 2024.

Tabla 7. Distribución de la cobertura efectiva de viviendas por estrato socioeconómico de la ENEMDU de Septiembre 2024.

Dominio	Estrato Socioeconómico				Total
	Alto (1)	Medio (2)	Bajo (3)	Sin clasificar	
Quito	326	352	250	0	928
Guayaquil	236	417	221	0	874
Cuenca	194	252	192	0	638
Machala	190	304	224	0	718
Ambato	165	221	191	0	577
Resto Sierra Urbano	389	473	468	0	1.330
Resto Costa Urbano	305	362	322	0	989
Amazonía Urbano	110	134	109	81	434
Sierra Rural	389	392	358	0	1.139
Costa Rural	191	196	166	56	609
Amazonía Rural	162	194	168	0	524
Región Insular	54	28	28	0	110
Total	**2.711**	**3.325**	**2.697**	**137**	**8.870**

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU). INEC - Ecuador.



4. Cálculo de los factores de expansión

4. Cálculo de los factores de expansión

El procedimiento de ponderación general para la ENEMDU requiere de dos etapas.

- La **primera etapa** consiste en calcular una ponderación de UPM. De ser necesario se aplica un ajuste por no respuesta a nivel de UPM.
- La **segunda etapa** consiste en calcular una ponderación de viviendas dentro de cada UPM. De ser necesario se aplica un ajuste por no respuesta a nivel de vivienda.

La "*falta de respuesta*" se produce cuando no se llega a obtener respuesta de algunas unidades de la muestra. Resulta útil pensar en la población de la muestra dividida en dos conjuntos, el primero, formado por todas las unidades de la muestra de las que se ha obtenido respuesta; y el segundo, por todas las unidades de la muestra de las que no ha podido obtenerse respuesta.

La tasa de falta de respuesta puede estimarse con exactitud si se llevan a cabo recuentos de todos los elementos que cumplen los requisitos que se incluyen en la muestra. La tasa de respuesta en una encuesta se define como el coeficiente entre el número de cuestionarios completados por las unidades y el número total de unidades de la muestra que cumplen los requisitos. La falta de respuesta puede deberse a la ausencia del domicilio de las personas seleccionadas, a los cambios en la condición de ocupación de la vivienda, a la negación de dichas personas a participar o a su incapacidad para responder a las preguntas, entre otras. Otra posible causa de la falta de respuesta puede ser la imposibilidad de llevar a cabo la encuesta en determinadas zonas por razonas climatológicas, dificultades del terreno o cuestiones de seguridad.

El factor de expansión final es el resultado de un procedimiento que involucra el cálculo del factor de expansión de diseño, el ajuste por cobertura, el recorte de los pesos de muestreo extremos y la calibración de los pesos de muestreo; lo cual se describe a continuación:

Factor de expansión teórico:

Probabilidad de inclusión de primera etapa (UPM)

Las UPM fueron seleccionadas con igual probabilidad en los estratos de muestreo h, por tanto, la probabilidad de inclusión de primera etapa (UPM) es calculada a partir del siguiente algoritmo:

$$\pi_{h_i} = \frac{n_h}{N_h}$$

donde:

π_{h_i} = $probabilidad\ de\ inclusión\ de\ primera\ etapa\ de\ la\ UPM\ i\ en\ el\ estrato\ h.$
n_h = $número\ de\ UPM\ investigadas\ en\ el\ estrato\ h\ en\ la\ muestra.$
N_h = $número\ de\ UPM\ del\ estrato\ h\ en\ la\ población.$

Por tanto, la ponderación básica de primera etapa para la i-ésima UPM muestreada en el estrato h es la inversa de la probabilidad de inclusión de la primera etapa y queda definida por:

$$w_{h_i} = \frac{1}{\pi_{h_i}}$$

Probabilidad de inclusión de segunda etapa (viviendas)

Las viviendas fueron escogidas con igual probabilidad de selección en cada UPM levantada en la primera etapa, por tanto, la probabilidad de inclusión de la segunda etapa (vivienda) es calculada a partir del siguiente algoritmo:

$$\pi_{i_j} = \frac{n_i}{N_i}$$

donde:

π_{i_j} = $probabilidad\ de\ inclusión\ de\ segunda\ etapa\ de\ la\ vivienda\ j\ en\ la\ UPM\ i.$
n_i = $número\ de\ viviendas\ investigadas\ en\ la\ UPM\ i\ en\ la\ muestra.$
N_i = $número\ de\ viviendas\ de\ la\ UPM\ i\ en\ la\ población.$

Por tanto, la ponderación básica de segunda etapa para la j-ésima vivienda en la i-ésima UPM viene dada por:

$$w_{i_j} = \frac{1}{\pi_{i_j}}$$

Factor de expansión teórico o de diseño

El factor de expansión teórico o de diseño ($w_{h_{i_j}}$) es el inverso de la multiplicación de las probabilidades de inclusión de la primera y de la segunda etapa, tal como se expresa a continuación:

$$w_{h_{i_j}} = \frac{1}{\pi_{h_i} * \pi_{i_j}}.$$

Con la finalidad de relajar la notación, se notara como w_k al factor de expansión teórico ($w_{h_{i_j}}$).

Factor de expansión ajustado por cobertura

Ajuste por cambio de ocupación (no elegibilidad)

Si las viviendas seleccionadas presentan cambio en su estado de ocupación (viviendas temporales, desocupadas, en construcción, destruidas, convertidas en negocio y otra razón), se realiza el siguiente ajuste:

$$a_1 = 1 - \frac{\#\ de\ viviendas\ no\ elegibles\ en\ la\ UPM\ de\ la\ vivienda\ k}{\#\ \ de\ viviendas\ investigadas\ en\ la\ UPM\ de\ la\ viivenda\ k}$$

Por lo tanto, el factor de expansión ajustado por cambio de ocupación (no elegibilidad) es:

$$w_k^{ne} = w_k * a_1,$$

donde:

w_k^{ne} = $factor\ de\ expansion\ ajustado\ por\ no\ eligibilidad\ de\ la\ vivienda\ k.$
w_k = $factor\ de\ expansión\ teorico\ de\ la\ vivienda\ k.$
a_1 = $ajuste\ por\ cambio\ de\ ocupacion\ (no\ elegibles)\ en\ la\ UPM\ de\ la\ vivienda\ k.$

Ajuste por nadie en casa (elegibilidad desconocida)

El siguiente ajuste consiste en redistribuir los pesos de las unidades secundarias de muestreo con elegibilidad desconocida (nadie en casa) utilizando la siguiente expresión:

$$a_2 = \frac{\#\ \ de\ \ viviendas\ investigadas\ en\ la\ UMP\ de\ la\ vivienda\ k}{\#\ \ de\ viviendas\ completas\ en\ la\ \ UMP\ de\ la\ vivienda\ k}$$

Por lo tanto, el factor de expansión ajustado por elegibilidad desconocida es:

$$w_k^{ed} = w_k^{ne} * a_2,$$

donde:

w_k^{ed} = $factor\ de\ expansion\ ajustado\ por\ eligibilidad\ desconocida\ \ de\ la\ vivienda\ k.$
w_k^{ne} = $factor\ de\ expansión\ ajustado\ por\ no\ elegibilidad\ de\ la\ vivienda\ k.$
a_2 = $ajuste\ por\ nadie\ en\ casa\ (elegibilidad\ desconocida)\ en\ la\ UPM\ de\ la\ vivienda\ k.$

Cabe recalcar que este ajuste se realiza únicamente en las UPM que cuentan con al menos una vivienda con elegibilidad desconocida.

Ajuste por rechazo (no respondientes)

El último ajuste en los pesos de muestreo se realiza si el resultado de la entrevista de las viviendas, al momento de realizar la investigación, fue rechazo. Si este es el caso, el ajuste a realizar es:

$$a_3 = \frac{\#\ \ de\ \ viviendas\ investigadas\ en\ la\ UPM\ de\ la\ vivienda\ k}{\#\ \ de\ viviendas\ completas\ en\ la\ \ UPM\ de\ la\ vivienda\ k}$$

Cabe recalcar, que si una vivienda fue rechazo esta sigue siendo parte del conjunto de viviendas elegibles. Además, este ajuste no se realiza si $a_2 \neq 1$. Por tanto, el factor de expansión ajustado por rechazo viene dado por la siguiente multiplicación:

$$w_{k_a} = w_k^{ed} * a_3$$

donde:

$w_{k_a} = factor\ de\ expansion\ ajustado\ por\ rechazo\ en\ la\ vivienda\ k.$
$w_k^{ed} = factor\ de\ expansión\ ajustado\ por\ elegibilidad\ desconocida\ en\ la\ vivienda\ k.$
$a_3 = ajuste\ por\ rechazo\ (no\ respondientes)\ en\ la\ UPM\ de\ la\ vivienda\ k.$

Recorte de los factores de expansión extremos

Debido a los ajustes por cobertura y calibración la variabilidad de los factores de expansión se ha visto incrementada. Potter (1990) señala que 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 el tamaño de los factores de expansión extremos, algunos de los cuales están descritos en Potter (1990, 1993). Para la ENEMDU se utiliza el método de recorte y redistribución de factores de expansión más común, descrito en Valliant (2013), el cual propone el siguiente método iterativo:

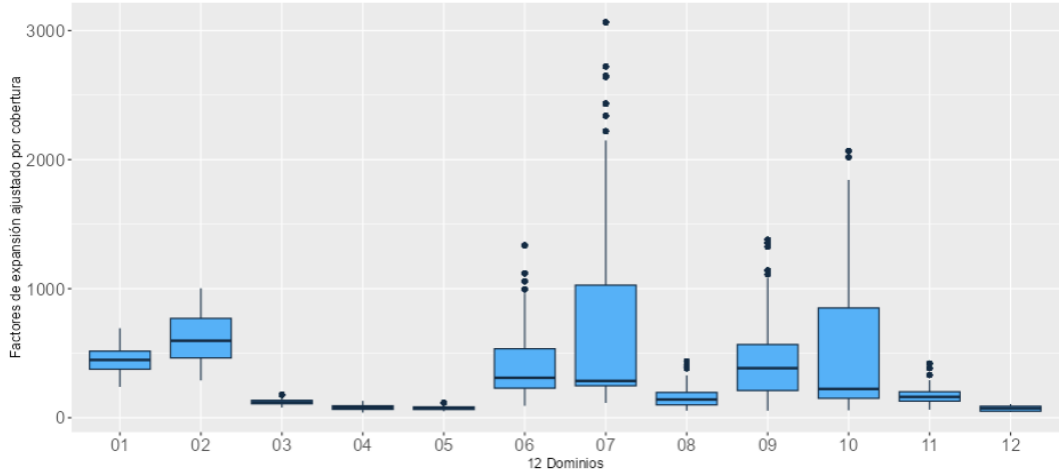
(1) Fijar una cota inferior y superior para los factores de expansión.
(2) Todo factor de expansión mayor a la cota superior (o menor a la cota inferior) se fija en el valor de la cota, definiendo:

$$w_{k_r} = \begin{cases} \mathrm{U} & \mathrm{si} \quad w_{k_a} \geq \mathrm{U}, \\ w_{k_a} & \mathrm{si} \quad \mathrm{L} < w_{k_a} < \mathrm{U}, \\ \mathrm{L} & \mathrm{si} \quad w_{k_a} \leq \mathrm{L}. \end{cases}$$

Donde w_{k_r} es el conjunto de los factores de expansión recortados.
(3) Calcular $R = \sum_{\mathrm{k}} \left| w_{k_a} - w_{k_r} \right|$.
(4) Distribuir R equitativamente entre todos los pesos no recortados.
(5) Repetir los pasos (1)-(4) hasta que $R = 0$.

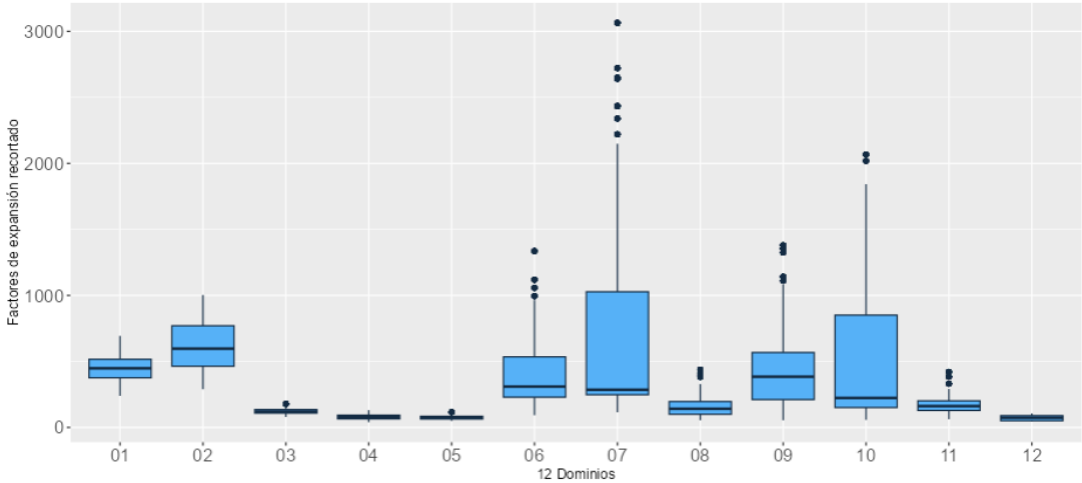
Puesto que la ENEMDU presenta factores de expansión extremos únicamente en la cola derecha de su distribución, L es igual a 1 y U se fija en 4,5 veces la mediana. Cabe señalar que, el recorte y redistribución de los factores de expansión se realizó de manera independiente en los estratos de muestreo, los cuales están formados por la provincia o ciudad principal, el área y el estrato socioeconómico.

Gráfico 2. Diagramas de caja y bigote de los factores de expansión ajustados por cobertura de la ENEMDU Septiembre 2024[1].



Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU). INEC - Ecuador.

Gráfico 3. Diagramas de caja y bigote del factor de expansión recortados de la ENEMDU Septiembre de 2024



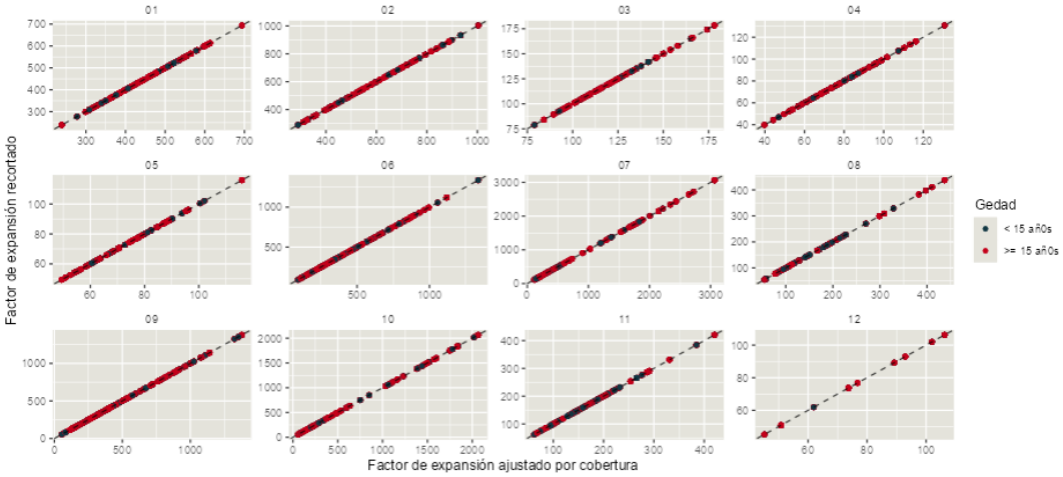
Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU). INEC - Ecuador.

En los gráficos 2 y 3 se presentan los diagramas de caja y bigote de los factores de expansión ajustados por cobertura y recortados, respectivamente, a nivel de las 12 subpoblaciones usuales[2]. Al realizar una comparación entre estos se aprecia que los factores de expansión ajustados por cobertura son iguales a los pesos de muestreo recortados, es decir, no existieron ponderadores extremos que necesitaron ser recortados.

[1] Los 12 dominios son los siguientes: 1 Quito, 2 Guayaquil, 3 Cuenca, 4 Machala, 5 Ambato, 6 Resto Sierra Urbano, 7 Resto Costa Urbano, 8 Amazonía Urbano, 9 Sierra Rural, 10 Costa Rural, 11 Amazonía Rural y 12 Región Insular.

2 Quito, Guayaquil, Cuenca, Machala, Ambato, Resto Sierra Urbano, Resto Costa Urbano, Resto Amazonía Urbano, Resto Sierra Rural, Resto Costa Rural, Resto Amazonía Rural, Región Insular.

Gráfico 4. Comparación de los factores de expansión ajustados por cobertura y recortados por dominio y grupo de edad.



Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU). INEC - Ecuador.

En el Gráfico 4 se puede evidenciar que, dado que no se presentan factores de expansión recortados en los diferentes dominios, los pesos de muestreo están sobre la línea de 45 grados (línea punteada negra), esto significa que los ponderadores recortados son iguales a los ajustados por cobertura.

Validación del recorte de los factores de expansión extremos

Se evaluaron 5 niveles de recorte de factores de expansión tomando 5 cotas superiores diferentes (3, 3.5, 4, 4.5 y 5 veces la mediana). Esta evaluación del nivel de recorte se efectuó, por un lado, tomando en cuenta el porcentaje de factores de expansión recortados por celda o post estrato de calibración y por otro una estimación del Error Cuadrático Medio (MSE) del recorte.

Porcentaje de factores de expansión recortados por celda o post estrato de calibración

A través de la identificación del porcentaje de factores de expansión recortados por celda o post estrato de calibración, se puede identificar que no existieron porcentajes de pesos de muestreo recortados en los diferentes estratos de muestreo.

Error cuadrático medio (MSE) del recorte de los factores de expansión

Otro criterio que permite evaluar la calidad del recorte de los ponderadores, es evaluar una estimación del Error Cuadrático Medio (MSE) para los elementos de datos seleccionados en varios niveles de recorte para determinar empíricamente el nivel de recorte (Cox y McGrath, 1981). El supuesto subyacente a este procedimiento es que para un conjunto de pesos y datos existe un punto en el que la reducción en la varianza de muestreo resultante del

recorte se compensa con el aumento en el cuadrado del sesgo introducido en la estimación.

En esta técnica, el $\mathrm{MSE}(\hat{\hat{Y}}_t)$ es estimado por:

$$\widehat{\mathrm{MSE}}\left(\hat{\hat{Y}}_t\right) = (\hat{\hat{Y}}_t - \hat{\hat{Y}})^2 - \widehat{\mathrm{Var}}\left(\hat{\hat{Y}}\right) + 2\left[\widehat{\mathrm{Var}}(\hat{\hat{Y}}_t)\widehat{\mathrm{Var}}(\hat{\hat{Y}})\right]^{1/2}$$

donde:

$\hat{\hat{Y}}$ = estimación de la media usando el factor de expansión no recortado,

$\hat{\hat{Y}}_t$ = estimación de la media usando el factor de expansión recortado,

$\widehat{\mathrm{Var}}(\hat{\hat{Y}})$ = varianza estimada de $\hat{\hat{Y}}$,

$\widehat{\mathrm{Var}}(\hat{\hat{Y}}_t)$ = varianza estimada de $\hat{\hat{Y}}_t$.

Asimismo, Potter (1990) sostiene que este procedimiento se implementa calculando repetidamente la estimación del MSE para elementos de datos seleccionados a diferentes niveles de recorte de pesos de muestreo. El nivel 'óptimo' de recorte es el punto que minimiza el MSE estimado (es decir, minimiza la varianza muestral y el sesgo cuadrático estimado) para el conjunto de datos.

Calibración de los factores de expansión

La calibración de los factores de expansión (Deville J.C., Särndal C.E. y Sautory O., 1993) es un ajuste que se realiza a los ponderadores con el propósito de que las estimaciones de algunas variables de control reproduzcan con exactitud los totales poblacionales de dichas variables.

Cuando los estudios por muestreo están afectados por la ausencia de respuesta, es deseable tener las siguientes propiedades en la estructura inferencial que sustenta el muestreo:

1. Sesgo pequeño o nulo.
2. Errores estándar pequeños.
3. Un sistema de ponderación que reproduzca la información auxiliar disponible.
4. Un sistema de ponderación que sea eficiente al momento de estimar cualquier característica de interés en un estudio multipropósito.

Heredia (2010) manifiesta que para la calibración de los factores de expansión es necesario tomar en cuenta la siguiente información:

- Considerar una fuente de información auxiliar que se utiliza como "población" que puede ser censos, registros administrativos u otras encuestas en las cuales se conocen los totales por variables o características que se desee estudiar.
- Otra fuente de información es la "muestra" de la cual procederán los estadísticos que infieren a los parámetros poblacionales.

- Identificar las variables de interés.
- Estimación de los totales de las variables de interés de la muestra.

El objetivo de la calibración es obtener un nuevo sistema de factores de expansión w_k que se encuentren cerca de los ponderadores de diseño d_k, de tal forma que cuando los ponderadores sean usados para estimar los totales de las variables auxiliares, dichos totales sean reproducidos con exactitud de manera que los nuevos factores conserven cualquier propiedad buena de estimación de los pesos básicos.

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

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

Luego, el estimador de calibración queda definido por:

$$\hat{t}_{yc} = \sum_s w_k y_{k=} \sum_s d_k F(x_k'\lambda) y_k$$

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

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

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

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

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

Para la ENEMDU se aplicó un esquema de calibración en el cual se presentan 8 celdas o post estratos de calibración (Ver Anexo 1) y se evaluó cada uno a través de criterios para validar la calibración (Silva, 2004). (Ver Anexo 2).

Es necesario mencionar que las proyecciones poblacionales son elaboradas en un área diferente a la Gestión de Diseño Muestral. Los criterios demográficos aplicados en la generación de las proyecciones poblacionales son explícitamente responsabilidad de los funcionarios que las elaboraron y quienes en la actualidad forman parte del equipo del Censo de Población y Vivienda 2021-2023 (Equipo de Proyecciones de Población del CPV).

Validación de la calibración de los factores de expansión

Silva (2004) propone 7 medidas para evaluar la calidad de la calibración de los factores de expansión, las cuales se detallan a continuación:

- 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

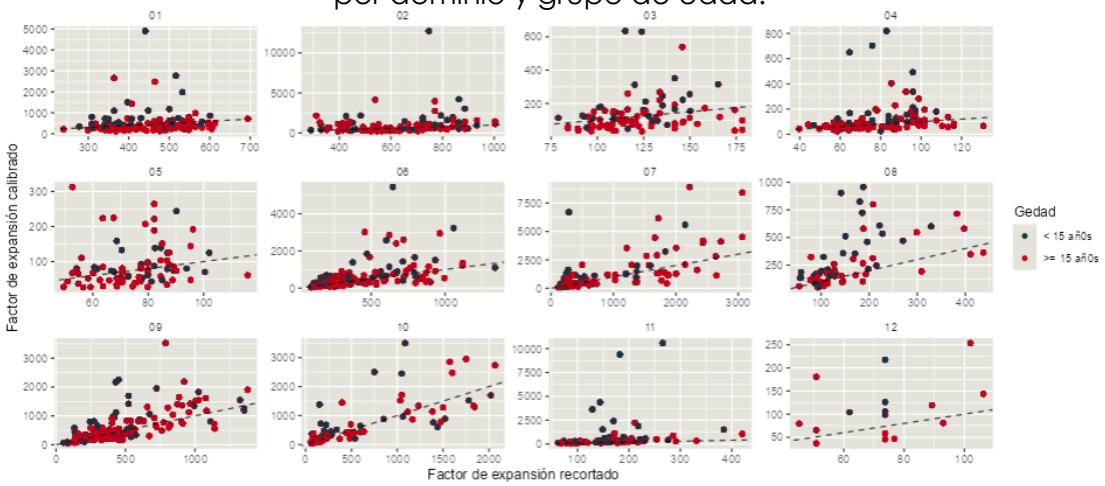
$$M6 = \frac{1}{n}\sum_{k\in S}\frac{(w_{k_c} - w_{k_r})^2}{w_{k_r}} = \frac{1}{n}\sum_{k\in S} w_{k_r}(g_k - 1)^2.$$

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

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

En el siguiente gráfico se evidencia que, en la mayoría de dominios, los factores de expansión calibrados son más grandes que los ponderadores recortados debido a la condición de reproducir con exactitud los totales poblacionales por sexo y grupo de edad[3] en las diferentes celdas o post estratos de calibración.

Gráfico 5. Comparación de los factores de expansión recortados y calibrados por dominio y grupo de edad.



Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU). INEC - Ecuador.

[3] Menores a 15 años y mayores o iguales a 15 años.

5. Estimaciones de características

5. Estimaciones de características

Estimación de características de la población:

Una vez construidos los factores de expansión se calculan los estimadores provenientes de la ENEMDU, para ello se utiliza el estimador de Horvitz-Thompson el cual sirve para estimar el valor total de una característica determinada. Está dado por (Carl-Erik Särndal, 1992):

$$\hat{Y}_{HT} = \sum_{k \in D} \sum_{l \in k} w_{k_c} y_{k_l}$$

donde:

$\hat{Y}_{HT}$ = $estimador\ HT\ para\ el\ total\ de\ la\ característica\ de\ interes\ Y\ de\ la\ variable\ y$

w_{k_c} = $factor\ de\ expansión\ calibrado\ de\ la\ vivienda\ k.$

y_{k_l} = $valor\ de\ la\ variable\ y\ para\ la\ persona\ l\ de\ la\ vivienda\ k.$

Estimación de errores:

Una vez realizada la estimación respectiva para la variable de interés a nivel de dominio de estudio el error de muestreo es calculado a partir de la estimación de la varianza del estimador del total $\hat{Y}_{HT}$. Para calcular adecuadamente 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 $\hat{Y}_{HT}$ viene dado por la siguiente expresión:

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

donde:

$\hat{V}_{2st}(\hat{Y}_{HT})$ = $estimación\ de\ la\ varianza\ de\ dos\ etapas\ del\ estimador\ HT\ del\ total\ de\ la\ variable\ y.$

$\hat{Y}_{HT}$ = $estimador\ \ HT\ del\ total\ de\ la\ característica\ de\ interés\ Y.$

Un estimador insesgado para la varianza está dado por:

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

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

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

donde:

i, j = $Índice\ que\ recorre\ las\ UPM\ i, j\ en\ el\ dominio\ de\ estudio\ D.$

q, r = $Índices\ que\ recorren\ las\ viviendas\ de\ la\ UPM\ i\ en\ el\ dominio\ de\ estudio\ D.$

π_{Ii} = $Probabilidad\ de\ selección\ de\ la\ Etapa\ I\ para\ la\ i - ésima\ UPM\ en\ el\ dominio\ de\ estudio\ D.$

$\pi_{IIq|i}$ = $Probabilidad\ de\ selección\ de\ la\ Etapa\ II\ para\ la\ q - ésima\ vivienda, dada\ la\ i - ésima\ UPM.$

$\breve{\Delta}_{Iij}$ = $Cantidad\ \Delta\ expandida\ asociada\ a\ las\ UPM\ i, j.$

$\breve{\Delta}_{IIqr|i}$ = $Cantidad\ \Delta\ expandida\ asociada\ a\ las\ viviendas\ q, r\ dada\ la\ selección\ de\ la\ i - ésima\ UPM.$

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 del margen de error que constituye una medida unificada del error total de muestreo 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 del último conglomerado
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) que se encuentran descritos en el documento "ENEMDU: Cálculo de errores estándar y declaración de muestras complejas[4]" donde se realiza una breve descripción de los métodos convencionales para estimar varianzas o errores muestrales para estimaciones basados en muestreo complejo, que es una característica de la ENEMDU.

A continuación, se describirá las principales características de cada uno de los métodos de estimación de errores para el muestreo complejo:

[4] El documento se encuentra disponible: Cálculo del error de muestreo y declaración de muestras complejas en la (ENEMDU), INEC (2019).

- Los métodos exactos pueden ser utilizados 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 del último conglomerado 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 utiliza para la estimación de los parámetros de interés y sus correspondientes errores de muestreo diversos programas estadísticos tales como SPSS, Stata y R. En virtud de las características de cada uno de los métodos es la técnica del último conglomerado en combinación con la linealización de Taylor, la cual induce a una muy buena aproximación del error muestral sobre los indicadores más importantes de las encuestas dirigidas a hogares, además de su facilidad de cálculo y replica. En este sentido, será esta la técnica la utilizada para la estimación de los errores muestrales en la ENEMDU. Las variables requeridas para declarar el diseño muestral en los programas estadísticos (SPSS, Stata y R) y ejecutar el cálculo de los errores de muestreo son presentadas en la Tabla 7, donde se describe las etiquetas de las variables identificadoras de las UPM, estratos y factores de expansión.

Tabla 7. Variables requeridas para declaración del diseño muestral – ENEMDU

Característica	Variable	Descripción
UPM	Upm	Agrupación de viviendas ocupadas en un número entre 30 a 60, próximas entre sí y con límites definidos.
Estratos	Estrato	Identificación de estrato muestral
Ponderación	Fexp	Factor de expansión calibrado y ajustado (no cobertura)

Fuente: CEPAL. (2021). Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024).

Es importante indicar que los estratos de muestreo están definidos por el cruce entre Provincia (25 grupos) + Área (2 grupos) + estrato socioeconómico de la UPM (3 grupos). Además, las UPM deben tener identificadores únicos dentro de cada estrato y a través del tiempo. Por último, los hogares deben estar unívocamente identificados, así como su pertenencia a las UPM, a los estratos de muestreo y a las rondas del panel correspondiente.

Referencias

CEPAL. (2021). *Recomendaciones Metodológicas para el Rediseño de la Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU 2021 – 2024)*. Informe de misión al Instituto Nacional de Estadística y Censos (INEC) de Ecuador.
CEPAL. (2019). *Revisión del esquema de agregación y análisis de la Encuesta Nacional de Empleo, Desempleo y Subempleo – ENEMDU.* Misión de Asistencia Técnica.

Cox, B.G. and McGrath, D.S. (1981), "*An Examination of the Effect of Sample Weight Truncation on the Mean Square Error of Survey Estimates,*" presented at Biometrics Society ENAR meeting, Richmond, VA. March 1981.

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

Gutiérrez, A. (2018). *Revisión del diseño de muestreo y esquema de análisis de la Encuesta Nacional de Empleo, Desempleo y Subempleo-ENEMDU.* Misión de Asistencia Técnica. Quito.

Heredia, O. (2010). *El problema de la Calibración.* Aguascalientes: Dirección de diseño y marcos estadísticos - Instituto Nacional de Estadística y Geografía – INEGI.

Hidiroglu, M., Särndal, C., & Binder, D. (1995). Weighting and Estimation in Business Surveys. En Cox, Binder, Chinnappa, Christianson, Colledge, & Kott, Business Survey Methods (págs. 477-502). Jhon Wiley & Sons.

Lemaitre & Dufour (1987). An integrated method for weighting persons and families. *Survey Methodology*, 13(2).

Kish, L. (1987). Statistical Design for Research. New York: Wiley.

Kish, L. (1972). *Muestreo de Encuestas.* México: Trillas.

Maligalig & Martinez (2013). *Developing a Master Sample Design for Households Surveys in Developing Countries: A Case Study In Bangladesh.* Survey Methods: Insights from the Field.

Potter F.J. (1990). A study of procedures to identify and trim extreme sampling weights. *American Statistical Association*, pp 225–230.

Särndal, C.E. y Swensson, B. y. W. J. (2003). *Model Assisted Survey Sampling.* Springer Series in Statistics. Springer: New York.

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

Valliant R., Dever J.A. y Kreuter F. (2013). *Practical Tools for Designing and Weighting Survey Samples.* Springer International Publishing.

Anexos

Anexo 1: Esquema de calibración propuesto.

El post estrato al que pertenece cada observación se puede identificar mediante un id de 5 dígitos, por ejemplo "00_1_1_1", que hacen referencia a los cruces utilizados para definir cada post estrato de acuerdo al siguiente detalle:

Dígito	Descripción y valores posibles
00	Dominio geográfico: nacional
1	Área: urbana (1) o rural (2)
1	Sexo: hombre (1) mujer (2)
1	Grupo de edad: menor a 15 años (1) 15 o más años (2)

En el ejemplo citado anteriormente, el post estrato "00_1_1_1" identifica a los individuos hombres menores a 15 años de la parte urbana a nivel nacional.

id_calib	Población
00_1_1_1	1.681.205
00_1_1_2	4.470.932
00_1_2_1	1.798.093
00_1_2_2	4.773.519
00_2_1_1	966.091
00_2_1_2	2.016.518
00_2_2_1	954.171
00_2_2_2	2.003.137

Anexo 2: Medidas para evaluar el esquema de calibración planteado

Medida 1:

id_calib	er_upm_c
00_1_1_1	5,54E-16
00_1_1_2	0,00E+00
00_1_2_1	6,47E-16
00_1_2_2	1,95E-16
00_2_1_1	9,20E-07
00_2_1_2	3,15E-07
00_2_2_1	4,86E-07
00_2_2_2	2,34E-07

Medida 2:

El valor de esta medida es del 10,99%.

Medida 3:

id_calib	L	n	M3_upm_prop_c
00_1_1_1	1	2.224	28,87%
00_1_1_2	1	7.719	46,30%
00_1_2_1	1	2.038	22,57%
00_1_2_2	1	8.942	50,04%
00_2_1_1	1	960	21,46%
00_2_1_2	1	3.000	32,63%
00_2_2_1	1	877	14,03%
00_2_2_2	1	3.100	37,65%

Medida 4:

id_calib	n	U	U3	M4_upm_U_prop_c	M4_upm_U3_prop_c
00_1_1_1	2.224	2,05%	3	30,08%	18,03%
00_1_1_2	7.719	1,48%	3	29,64%	9,12%
00_1_2_1	2.038	2,34%	3	31,45%	22,18%
00_1_2_2	8.942	1,40%	3	30,31%	8,51%
00_2_1_1	960	2,77%	3	25,31%	24,27%
00_2_1_2	3.000	1,71%	3	28,73%	12,87%
00_2_2_1	877	2,99%	3	28,28%	28,28%
00_2_2_2	3.100	1,56%	3	30,84%	11,65%

Medida 5:

id_calib	cv_g_upm_c
00_1_1_1	1,10%
00_1_1_2	1,17%
00_1_2_1	1,11%
00_1_2_2	1,17%
00_2_1_1	2,14%
00_2_1_2	2,11%
00_2_2_1	1,90%
00_2_2_2	2,07%

Medida 6:

id_calib	dist_g_upm_c
00_1_1_1	2.240,46
00_1_1_2	1185,8308
00_1_2_1	3.270,83
00_1_2_2	1074,492
00_2_1_1	16.254,12
00_2_1_2	4.707,72
00_2_2_1	15.637,11
00_2_2_2	3.845,17

INEC Buenas cifras, mejores vidas