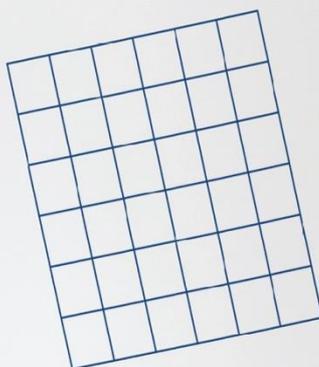


**Diseño Muestral de  
la Encuesta Nacional  
de Empleo,  
Desempleo y  
Subempleo  
(ENEMDU)  
Septiembre 2020**

\_\_\_\_\_  
Octubre, 2020



Instituto Nacional de Estadística y Censos (INEC)

# Diseño Muestral de la Encuesta Nacional de Empleo, Subempleo y Desempleo (ENEMDU) Septiembre 2020

## **Dirección**

Dirección de Infraestructura Estadística y Muestreo

## **Elaborado por:**

William Constante

Jorge Velásquez

Javier Núñez

## **Revisado por:**

Christian Garcés

## **Aprobado por:**

David Sánchez

Quito -Ecuador, 2020

## Contenido

Introducción .....	5
Diseño Muestral ENEMDU Septiembre 2020 .....	6
Determinación del Marco de Muestreo para encuestas de hogares.....	6
Universo de investigación .....	6
Unidad de observación.....	6
Unidad de análisis .....	6
Cobertura geográfica .....	7
Tipo de muestreo.....	7
Dominios de estudio.....	7
Representatividad.....	7
Tamaño y selección de la muestra de la ENEMDU.....	7
Tamaño de la muestra .....	7
Asignación de la muestra .....	9
Selección de la muestra.....	10
Rotación de la muestra por paneles.....	10
Cobertura de viviendas planificada y efectiva.....	11
Cobertura planificada.....	11
Cobertura efectiva .....	11
Cálculo de los factores de expansión .....	12
Factor de expansión teórico .....	13
Factor de expansión ajustado por cobertura.....	14
Calibración de los factores de expansión.....	15
Recorte de los factores de expansión extremos .....	18
Estimación de características de la población.....	23
Estimación de errores .....	23
Métodos de estimación de errores para diseños muestrales complejos .....	24
Referencias .....	26
Anexos .....	27

## Lista de tablas

Tabla 1. Tamaños de muestra ENEMDU septiembre 2020.....	9
Tabla 2. Asignación de la muestra de viviendas y UPM por territorio .....	9
Tabla 3. Rotación de panel ENEMDU septiembre y diciembre 2020 .....	10
Tabla 4. Distribución de la cobertura planificada de viviendas por estrato socioeconómico .....	11
Tabla 5. Distribución de la cobertura efectiva de viviendas por estrato socioeconómico .....	11
Tabla 6. Variables requeridas para declaración del diseño muestral – ENEMDU .....	25

## Lista de gráficos

Gráfico 1. Comparación de los factores de expansión ajustados por cobertura y calibrados por dominio y grupo de edad .....	18
Gráfico 2. Diagramas de caja y bigote de los factores de expansión calibrados de la ENEMDU Septiembre 2020 .....	19
Gráfico 3. Diagramas de caja y bigote del factor de expansión recortados de la ENEMDU Septiembre de 2020 .....	19
Gráfico 4. Comparación de los factores de expansión calibrados y recortados por dominio y grupo de edad .....	20

## 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 de Estadística 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.

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 y errores muestrales.

## Diseño Muestral ENEMDU Septiembre 2020

### Determinación del Marco de Muestreo para encuestas de hogares

El Marco de Muestreo se basa en los resultados definitivos y 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.<sup>1</sup>

### 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, las 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.

---

<sup>1</sup> Para más información de la construcción y estratificación del Marco de Muestreo para encuestas de hogares referirse al siguiente link: [https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Disenio\\_Muestral\\_2018/SIEH%20-MMM.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Disenio_Muestral_2018/SIEH%20-MMM.pdf)

## 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 a un muestreo probabilístico estratificado bietápico de elementos.

## Dominios de estudio

La ENEMDU Septiembre 2020 tiene como sus dominios de diseño Nacional, Urbano-Rural.

## Representatividad

La representatividad de la ENEMDU es Nacional, Urbano-Rural.

## Tamaño y selección de la muestra de la ENEMDU

### Tamaño de la muestra

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

- Estimación: el indicador que se utilizó en el cálculo del tamaño de muestra fue la tasa de desempleo de diciembre de 2019, la cual fue de 3,84%, siendo el valor más actual al momento de realizar el cálculo.
- Margen de error relativo del 15%.
- Nivel de confianza del 95%.
- Tasa de no respuesta del 20%.

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

Para la determinación del tamaño de la muestra se requiere establecer la característica o características a estimar, el nivel de confianza, la precisión requerida y la tasa de no respuesta.

En este sentido, la variable de investigación, en base a la cual se calcula el tamaño de la muestra es el *desempleo*, siendo igual al reportado por la ENEMDU diciembre 2019.

La expresión matemática utilizada para el cálculo de tamaño de muestra de personas es:

$$n_k = \frac{\frac{z^2}{(rme * P)^2} * S^2}{1 + \frac{\frac{z^2}{(rme * P)^2} * S^2}{N}} * \frac{1}{1 - tnr}$$

Donde:

- $z$  = el nivel de confianza
- $rme$  = es el margen de error relativo
- $P$  = estimación de la variable de diseño (desempleo)
- $N$  = tamaño de la población
- $tnr$  = tasa de no respuesta

A su vez, la varianza  $S^2$  es calculada a través del siguiente algoritmo:

$$S^2 = P(1 - P) * deff$$

donde  $deff$  es el efecto de diseño calculado a partir del coeficiente de correlación intraclase entre las UPM y la variable de diseño (desempleo), y el número promedio de viviendas levantadas por UPM.

$$deff = 1 + (\bar{n} - 1) * \rho$$

### Tamaño de muestra de viviendas

El número de viviendas que deben ser seleccionadas estará determinado por la muestra de personas ( $n_k$ ), 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:

$$n_{vk} = \frac{n_k}{r * b}$$

### 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:

$$n_{UPMk} = \frac{n_{vk}}{\text{Carga técnica operativa}}$$

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 siete (7).

A través de la aplicación de los algoritmos de cálculo, y tomando en cuenta que se levantan 7 viviendas por UPM junto con el presupuesto disponible de la operación estadística, se obtuvo un tamaño de muestra de 9.058 viviendas, cuya distribución por dominio se muestra en la Tabla 1:

**Tabla 1.** Tamaños de muestra ENEMDU septiembre 2020.

Subpoblación	Viviendas del marco	UPM del marco	Muestra UPM	Muestra Viviendas
Urbano	2.715.812	79.647	838	5.866
Rural	1.002.506	27.352	456	3.192
<b>Nacional</b>	<b>3.718.318</b>	<b>106.999</b>	<b>1.294</b>	<b>9.058</b>

### Asignación de la muestra

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

$$n_{ki} = n_k \cdot \frac{N_{ki}}{N_k}$$

Donde:

- $n_{ki}$  = Tamaño de la muestra para el estrato  $i$  del dominio de estudio  $k$ .
- $n_k$  = Tamaño de la muestra para el dominio de estudio  $k$ .
- $N_{ki}$  = Total de viviendas ocupadas en el estrato  $i$  del dominio de estudio  $k$ .
- $N_k$  = Total de viviendas ocupadas en el dominio de estudio  $k$ .

**Tabla 2.** Asignación de la muestra de viviendas y UPM por territorio

Dominio	Viviendas del marco	UPM del marco	Muestra UPM	Muestra Viviendas
Quito	473.957	13.811	82	574
Guayaquil	589.772	16.907	204	1.428
Cuenca	84.623	2.423	128	896
Machala	62.658	1.797	104	728
Ambato	47.817	1.376	120	840
Resto Sierra Urbano	592.918	17.565	72	504
Resto Costa Urbano	780.055	23.186	78	546
Amazonía Urbano	79.861	2.410	46	322
Sierra Rural	538.194	14.530	180	1.260
Costa Rural	360.130	9.862	192	1.344
Amazonía Rural	99.813	2.784	80	560
Región Insular	8.520	348	8	56
<b>Total</b>	<b>3.718.318</b>	<b>106.999</b>	<b>1.294</b>	<b>9.058</b>

En la Tabla 2 se presenta la asignación de la muestra por dominio, señalando que estos no son dominios de diseño de esta encuesta, por cuanto la representatividad de la ENEMDU 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. Cada UPM, junto con las viviendas seleccionadas dentro de la misma, pertenecen a un único panel, por lo tanto, su seguimiento es a nivel de UPM.

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.

**Tabla 3.** Rotación de panel ENEMDU septiembre y diciembre 2020

Septiembre			Diciembre		
Panel	Número	UPM	Panel	Número	UPM
E	005	647	E	005	647
F	006	647	I	009	647
<b>Total</b>		<b>1.294</b>	<b>Total</b>		<b>1.294</b>

Como se ilustra en la Tabla 3, en los meses de septiembre y diciembre de 2020 se levantará la información de la ENEMDU a partir de un tamaño de muestra igual para cada mes (1.294 UPM).

Esta estructura de rotación de la encuesta considera dos paneles cada mes, cada uno formado por 647 UPM. En los meses de septiembre y diciembre, existe un 50% de traslape de la muestra de viviendas, debido a que se repite un panel, por ejemplo, las viviendas que se investigarán en el mes de septiembre presentan un 50% de traslape con las de diciembre, ya que pertenecen al mismo panel E.

## 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 2020:

**Tabla 4.** Distribución de la cobertura planificada de viviendas por estrato socioeconómico

Dominio	Estrato Socioeconómico			Total
	Alto (1)	Medio (2)	Bajo (3)	
Quito	210	238	126	574
Guayaquil	364	812	252	1.428
Cuenca	252	392	252	896
Machala	224	280	224	728
Ambato	280	280	280	840
Resto Sierra Urbano	168	168	168	504
Resto Costa Urbano	196	196	154	546
Amazonía Urbano	140	98	84	322
Sierra Rural	462	504	294	1.260
Costa Rural	476	574	294	1.344
Amazonía Rural	210	224	126	560
Región Insular	28	14	14	56
<b>Total</b>	<b>3.010</b>	<b>3.780</b>	<b>2.268</b>	<b>9.058</b>

### Cobertura efectiva

Por otro lado, la cobertura efectiva de viviendas levantadas en la ENEMDU de septiembre 2020, se muestra a continuación:

**Tabla 5.** Distribución de la cobertura efectiva de viviendas por estrato socioeconómico

Dominio	Estrato Socioeconómico			Total
	Alto (1)	Medio (2)	Bajo (3)	
Quito	190	216	117	523
Guayaquil	337	789	253	1.379
Cuenca	233	378	241	852
Machala	202	275	220	697
Ambato	273	261	266	800
Resto Sierra Urbano	160	157	159	476
Resto Costa Urbano	179	186	143	508

Amazonía Urbano	133	89	75	297
Sierra Rural	426	488	285	1.199
Costa Rural	453	546	288	1.287
Amazonía Rural	186	205	123	514
Región Insular	29	12	14	55
<b>Total</b>	<b>2.801</b>	<b>3.602</b>	<b>2.184</b>	<b>8.587</b>

### 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 razones 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, la calibración y el recorte de los pesos de muestreo extremos, los cuales se describen 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 la primera etapa (UPM) es calculada a partir del siguiente algoritmo:

$$\pi_i = \frac{n_{i_h}}{N_{i_h}}$$

Donde:

$n_{i_h}$  = es el número de UPM investigadas en el estrato  $h$  en la muestra.

$N_{i_h}$  = es el 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:

$$w_i = \frac{1}{\pi_i}$$

### 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:

$$\pi_{j|i} = \frac{n_{j_i}}{N_{j_i}}$$

Donde:

$n_{j_i}$  = es el número de viviendas investigadas en la UPM  $i$  en la muestra.

$N_{j_i}$  = es el 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:

$$w_j = \frac{1}{\pi_{j|i}}$$

### Factor de expansión teórico o de diseño

El factor de expansión teórico o de diseño 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:

$$w_k = \frac{1}{\pi_i * \pi_{j|i}}$$

## 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), se realiza el siguiente ajuste:

$$a_1 = 1 - \frac{\# \text{ de viviendas no elegibles}}{\# \text{ de viviendas seleccionadas}}$$

Por lo tanto, el factor de expansión ajustado por cambio de ocupación es:

$$w_{k1}^{ne} = w_k * a_1,$$

donde:

- $w_{k1}^{ne}$  = es el factor de expansión ajustado por no elegibilidad.
- $w_k$  = es el factor de expansión teórico.
- $a_1$  = ajuste por cambio de ocupación (no elegibles).

### Ajuste por nadie en casa

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

$$a_2 = \frac{\# \text{ de viviendas investigadas en la UMPs}}{\# \text{ de viviendas completas en la UMPs}}$$

Por lo tanto, el factor de expansión ajustado por cambio de ocupación es:

$$w_{k2}^{ed} = w_{k1}^{ne} * a_2,$$

donde:

- $w_{k2}^{ed}$  = es el factor de expansión ajustado por elegibilidad desconocida.
- $w_{k1}^{ne}$  = es el factor de expansión ajustado por no elegibilidad.
- $a_2$  = ajuste por nadie en casa (elegibilidad desconocida).

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:

$$a_3 = \frac{\# \text{ de viviendas investigadas en la UMPs}}{\# \text{ de viviendas completas en la UMPs}}$$

Cabe recalcar, que si una vivienda fue rechazada 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 cobertura viene dado por la siguiente multiplicación:

$$w_k^{(*)} = w_{k2}^{ed} * a_3$$

donde:

- $w_k^{(*)}$  = es el factor de expansión ajustado por elegibilidad conocida.
- $w_{k2}^{ed}$  = es el factor de expansión ajustado por elegibilidad desconocida.
- $a_3$  = ajuste por rechazo (no respondientes).

### 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 estándares 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  $\lambda$ , 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  $g = F(x_k' \lambda)$ .

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

$$w_k^{(**)} = w_k^{(*)} * g_k$$

Donde  $w_k^{(**)}$  son los factores de expansión calibrados, mientras que  $w_k^{(*)}$  son los pesos de muestreo ajustados por cobertura y  $g_k$  son los pesos de calibración.

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 propuestas por Silva (2004) (Ver Anexo 2).

## 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  $g_k$

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

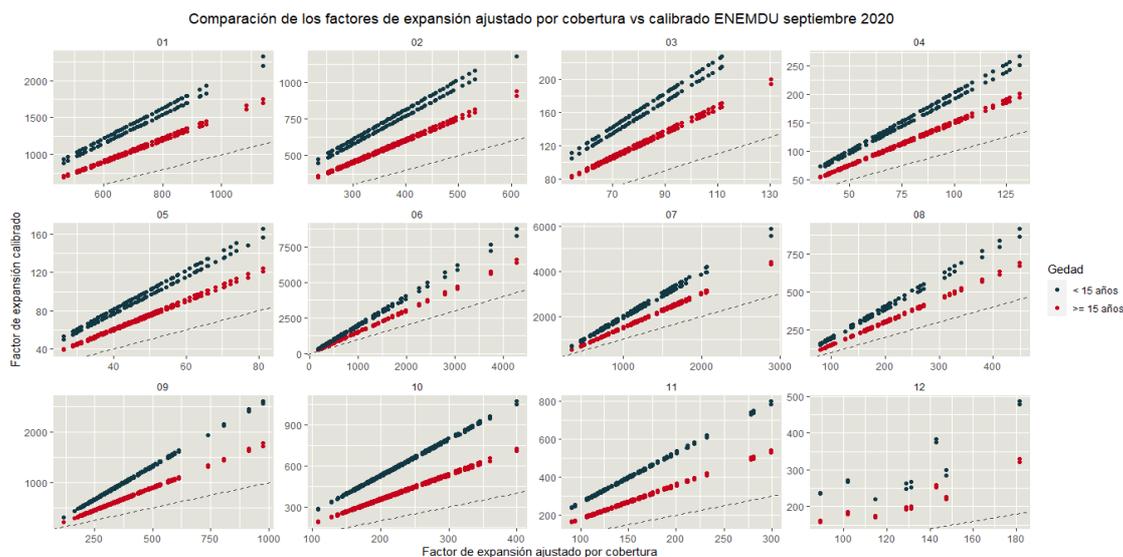
- Distancia entre los pesos de calibración y los pesos originales

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

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

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

**Gráfico 1.** Comparación de los factores de expansión ajustados por cobertura y calibrados por dominio y grupo de edad



Según el Gráfico 1, en la mayoría de dominios, los factores de expansión calibrados son más grandes que los ponderadores ajustados por cobertura, debido a la condición de reproducir con exactitud los totales poblacionales por sexo y grupo de edad (menores a 15 años, y mayores o iguales a 15 años) en los diferentes posts estratos de calibración.

### 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:

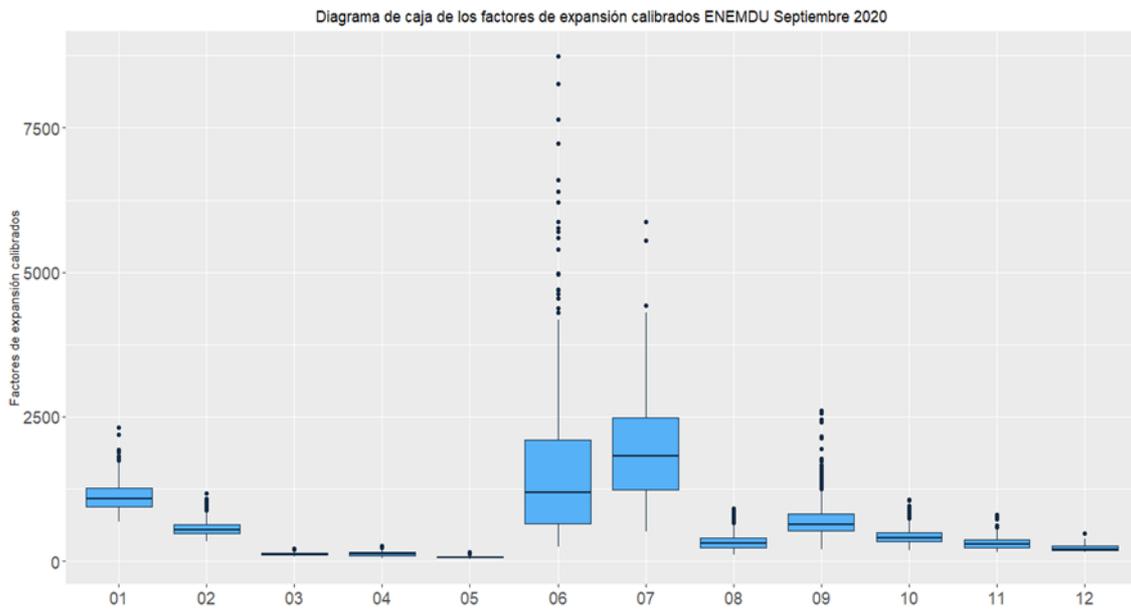
$$fexp_{trim} = \begin{cases} U & \text{si } w_k^{(**)} \geq U, \\ w_k^{(**)} & \text{si } L < w_k^{(**)} < U, \\ L & \text{si } w_k^{(**)} \leq L. \end{cases}$$

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

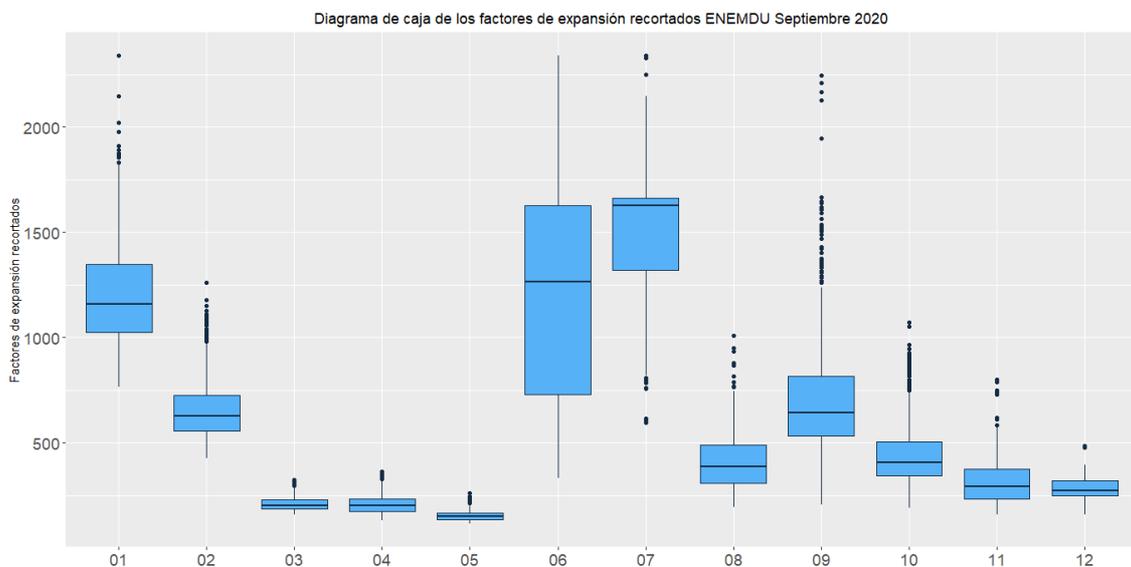
- (3) Calcular  $K = \sum_k |fexp - fexp_{trim}|$ .
- (4) Distribuir  $K$  equitativamente entre todos los pesos no recortados.
- (5) Repetir los pasos (1)-(4) hasta que  $K = 0$ .

Puesto que 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 veces la mediana. Cabe señalar que, el recorte y redistribución de los factores de expansión se realizó de manera independiente para cada una de las ocho celdas o post estratos de calibración.

**Gráfico 2.** Diagramas de caja y bigote de los factores de expansión calibrados de la ENEMDU Septiembre 2020

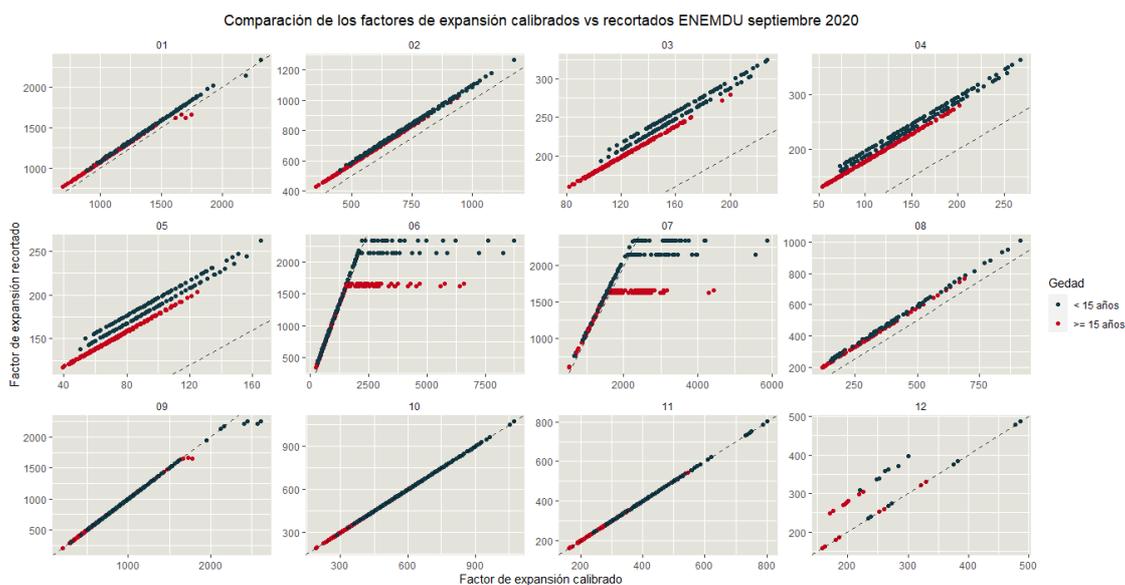


**Gráfico 3.** Diagramas de caja y bigote del factor de expansión recortados de la ENEMDU Septiembre de 2020



En los gráficos 2 y 3 se presentan los diagramas de caja y bigote de los factores de expansión calibrados y recortados, respectivamente, a nivel de los 12 dominios usuales<sup>2</sup>. Al realizar una comparación entre estos se aprecia que los factores de expansión calibrados presentan una mayor variabilidad que los factores recortados, teniendo valores extremos más grandes en comparación con la función de distribución de los ponderadores recortados, al realizar la identificación del efecto del recorte y redistribución se eliminan los valores atípicos o extremos de pesos de muestreo.

**Gráfico 4.** Comparación de los factores de expansión calibrados y recortados por dominio y grupo de edad



Según el Gráfico 4, existen algunos dominios donde se presenta una mayor cantidad de factores de expansión recortados, tales como en el Resto Sierra y Resto Costa Urbano.

### 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 al recortar con una cota superior de 3 o 3,5 veces la mediana, existen celdas que

<sup>2</sup> 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.

poseen un gran porcentaje de ponderadores que se recortan, sin embargo, a partir del recorte de 4 veces la mediana, el porcentaje de pesos de muestreo recortados no supera el 10% (Ver Anexo 3).

### 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  $MSE(\hat{Y}_t)$  es estimado por:

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

Donde:

- $\hat{Y}$  = estimación de la media usando el factor de expansión no recortado,
- $\hat{Y}_t$  = estimación de la media usando el factor de expansión recortado,
- $\widehat{Var}(\hat{Y})$  = varianza estimada de  $\hat{Y}$ , y
- $\widehat{Var}(\hat{Y}_t)$  = varianza estimada de  $\hat{Y}_t$ .

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.

Para evaluar los diferentes niveles de recorte aplicados, se estimó los MSE de 3 indicadores principales: PEA, desempleo y pobreza, bajo diferentes desagregaciones como nivel nacional, área, dominio de diseño y post estrato de calibración, determinando los niveles "óptimos" de recorte para cada indicador investigado y para cada nivel de desagregación analizado (Ver Anexo 4).

En los resultados presentados en el Anexo 4 se aprecia que, en su mayoría, los niveles de recorte óptimos se encuentran en 3, 3.5 y 4 veces la mediana como cota superior, sin embargo, en el área rural, en algunos dominios de diseño o celdas de calibración el nivel óptimo de recorte se obtiene cuando la cota superior es de 3 o 4 veces la mediana.

A través de la revisión de la literatura, Valliant (2013) sugiere 3,5 como umbral de recorte. Sin embargo, al haber efectuado el análisis del porcentaje de factores de expansión recortados, la estimación del MSE del recorte y la validación del recorte de los pesos de la ENEMDU, se decidió que la cota superior óptima corresponde a 4 veces la mediana.

Como se puede evidenciar en los diagramas de caja y bigote graficados (Ver Anexo 5), el recorte y redistribución reduce la variabilidad de los factores de expansión. Además, entre más alta sea la cota superior del recorte, menos proporción de factores de expansión serán recortados en cada celda o post estrato de calibración.

Por otra parte, el valor mínimo y la media de los factores de expansión son los mismos para todos los niveles de recorte efectuados, a su vez, el valor máximo es más grande a medida que aumente la cota superior de recorte (Ver Anexo 6).

## Estimación de características de la población

Una vez contruidos los factores de expansión, se calcula los estimadores provenientes de la ENEMDU, se utiliza el estimador de Horvitz-Thompson, el cual sirve para estimar el valor total de una característica determinada (tasa de empleo adecuado, desempleo, subempleo, etc.). Está dado por (Carl-Erik Särndal, 1992).

$$\hat{Y}_{mHT} = \sum_{i=1}^{C_m} \sum_{j=1}^{v_i} P_T^j y_{ij},$$

Donde:

- $\hat{Y}_{mHT}$  = estimador insesgado para el total  $Y_k$  de una variable y en el dominio de estudio  $m$   
 $C_m$  = número de conglomerados investigadas en el dominio de estudio  $m$ .  
 $v_i$  = número de viviendas investigadas en el conglomerado  $i$  del dominio de estudio  $m$   
 $y_{ij}$  = valor de la variable y para la vivienda  $j$  del conglomerado  $i$  del dominio de estudio  $m$

## 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  $\hat{Y}_{mHT}$ . 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  $\hat{Y}_{kHT}$  viene dado por la siguiente expresión:

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

Donde:

- $\hat{V}_{2st}(\hat{Y}_{kHT})$  estimación de la varianza de segunda etapa del estimador del total del individuo  $k$   
 $\hat{Y}_{kHT}$  estimador del total de la característica de interés del individuo  $k$

Un estimador insesgado para la varianza está dado por:

$$\hat{V}_{2st}(\hat{t}_\pi) = \sum \sum_{s_l} \Delta_{lij} \frac{\hat{t}_{i\pi} \hat{t}_{j\pi}}{\pi_{li} \pi_{lj}} + \sum_{s_l} \frac{\hat{V}_i}{\pi_{li}}$$

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

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

Donde:

- $i, j$  = Índice que recorre las 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  $m$ .
- $\pi_{Ii}$  = Probabilidad de selección de la Etapa I para la  $i$  – ésima UPM en el 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  $\Delta$  expandida asociada a las UPM  $i, j$ .
- $\check{\Delta}_{IIqr|i}$  = 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”**, 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:

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

**Tabla 6.** 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 (aproximación clasificación socio-económica)
Ponderación	fexp	Factor de expansión calculado y ajustado (no cobertura)

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

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.

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

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	1.589.365
00_1_1_2	4.223.780
00_1_2_1	1.699.982
00_1_2_2	4.509.635
00_2_1_1	913.259
00_2_1_2	1.905.377
00_2_2_1	901.757
00_2_2_2	1.892.335

### Anexo 2: Medidas para evaluar el esquema de calibración planteado (Esquema 1)

Medida 1:

Dominio	Esquema 1
00_1_1_1	1,5E-14
00_1_1_2	1,3E-15
00_1_2_1	1,9E-13
00_1_2_2	9,1E-15
00_2_1_1	6,4E-09
00_2_1_2	1,4E-14
00_2_2_1	1,2E-08
00_2_2_2	5E-15

Medida 2:

Dominio	Esquema 1
Total	0,02810

Medida 3:

<b>Dominio</b>	<b>L</b>	<b>n</b>	<b>Esquema 1</b>
00_1_1_1	1,93	2.162	-
00_1_1_2	1,54	7.189	100
00_1_2_1	2,04	2.127	-
00_1_2_2	1,49	7.852	-
00_2_1_1	2,63	1.443	-
00_2_1_2	1,82	4.101	-
00_2_2_1	2,68	1.400	-
00_2_2_2	1,77	4.043	-

Medida 4:

<b>Dominio</b>	<b>U</b>	<b>n</b>	<b>Esquema 1</b>
00_1_1_1	1,93	2.162	-
00_1_1_2	1,54	7.189	-
00_1_2_1	2,04	2.127	-
00_1_2_2	1,49	7.852	-
00_2_1_1	2,63	1.443	-
00_2_1_2	1,82	4.101	-
00_2_2_1	2,68	1.400	-
00_2_2_2	1,77	4.043	-

Medida 5:

<b>Dominio</b>	<b>Media peso g</b>	<b>CV del peso g</b>
	<b>Esquema 1</b>	<b>Esquema 1</b>
00_1_1_1	1,925217315	2,48E-16
00_1_1_2	1,536306132	2,54E-16
00_1_2_1	2,03564254	0,00E+00
00_1_2_2	1,490137476	3,45E-16
00_2_1_1	2,631423639	4,44E-17
00_2_1_2	1,820471833	1,49E-16
00_2_2_1	2,678093886	1,06E-16
00_2_2_2	1,76859272	7,70156E-17

Medida 6:

<b>Dominio</b>	<b>Esquema 1</b>
00_1_1_1	326,87
00_1_1_2	110,00
00_1_2_1	421,11
00_1_2_2	92,59
00_2_1_1	640,13
00_2_1_2	171,80
00_2_2_1	677,28
00_2_2_2	156,34

**Anexo 3:** Porcentaje de factores de expansión recortados por celda o post estrato de calibración.

Porcentaje de pesos recortados por celda de calibración (3 veces la mediana):

<b>id_calib</b>	<b>No recortados</b>	<b>Recortados</b>	<b>Porcentaje</b>
00_1_1_2	6292	897	12,48%
00_1_2_2	6933	919	11,70%
00_1_1_1	1913	249	11,52%
00_1_2_1	1883	244	11,47%
00_2_2_2	3999	44	1,09%
00_2_1_2	4065	36	0,88%
00_2_1_1	1431	12	0,83%
00_2_2_1	1390	10	0,71%

Porcentaje de pesos recortados por celda de calibración (3,5 veces la mediana):

<b>id_calib</b>	<b>No recortados</b>	<b>Recortados</b>	<b>Porcentaje</b>
00_1_1_2	6524	665	9,25%
00_1_1_1	1973	189	8,74%
00_1_2_1	1944	183	8,60%
00_1_2_2	7187	665	8,47%
00_2_1_1	1431	12	0,83%
00_2_1_2	4069	32	0,78%
00_2_2_1	1390	10	0,71%
00_2_2_2	4021	22	0,54%

Porcentaje de pesos recortados por celda de calibración (4 veces la mediana):

id_calib	No recortados	Recortados	Porcentaje
00_1_1_2	6626	563	7,83%
00_1_1_1	2000	162	7,49%
00_1_2_1	1968	159	7,48%
00_1_2_2	7277	575	7,32%
00_2_1_2	4080	21	0,51%
00_2_2_1	1394	6	0,43%
00_2_2_2	4032	11	0,27%
00_2_1_1	1440	3	0,21%

Porcentaje de pesos recortados por celda de calibración (4,5 veces la mediana):

id_calib	No recortados	Recortados	Porcentaje
00_1_1_2	6729	460	6,40%
00_1_2_1	1997	130	6,11%
00_1_2_2	7394	458	5,83%
00_1_1_1	2038	124	5,74%
00_2_2_1	1398	2	0,14%
00_2_1_1	1442	1	0,07%
00_2_1_2	4101	0	0,00%
00_2_2_2	4043	0	0,00%

Porcentaje de pesos recortados por celda de calibración (5 veces la mediana):

id_calib	No recortados	Recortados	Porcentaje
00_1_1_2	6801	388	5,40%
00_1_2_1	2014	113	5,31%
00_1_2_2	7461	391	4,98%
00_1_1_1	2057	105	4,86%
00_2_1_1	1443	0	0,00%
00_2_1_2	4101	0	0,00%
00_2_2_1	1400	0	0,00%
00_2_2_2	4043	0	0,00%

**Anexo 4:** Niveles “óptimos” de recorte para cada indicador investigado y para cada estimador y nivel de desagregación analizado.

Niveles de recorte óptimo a nivel nacional:

<b>Nacional</b>	<b>PEA</b>	<b>Desempleo</b>	<b>Pobreza</b>
Nivel de recorte	3,5	4	4,5

Niveles de recorte óptimo por área:

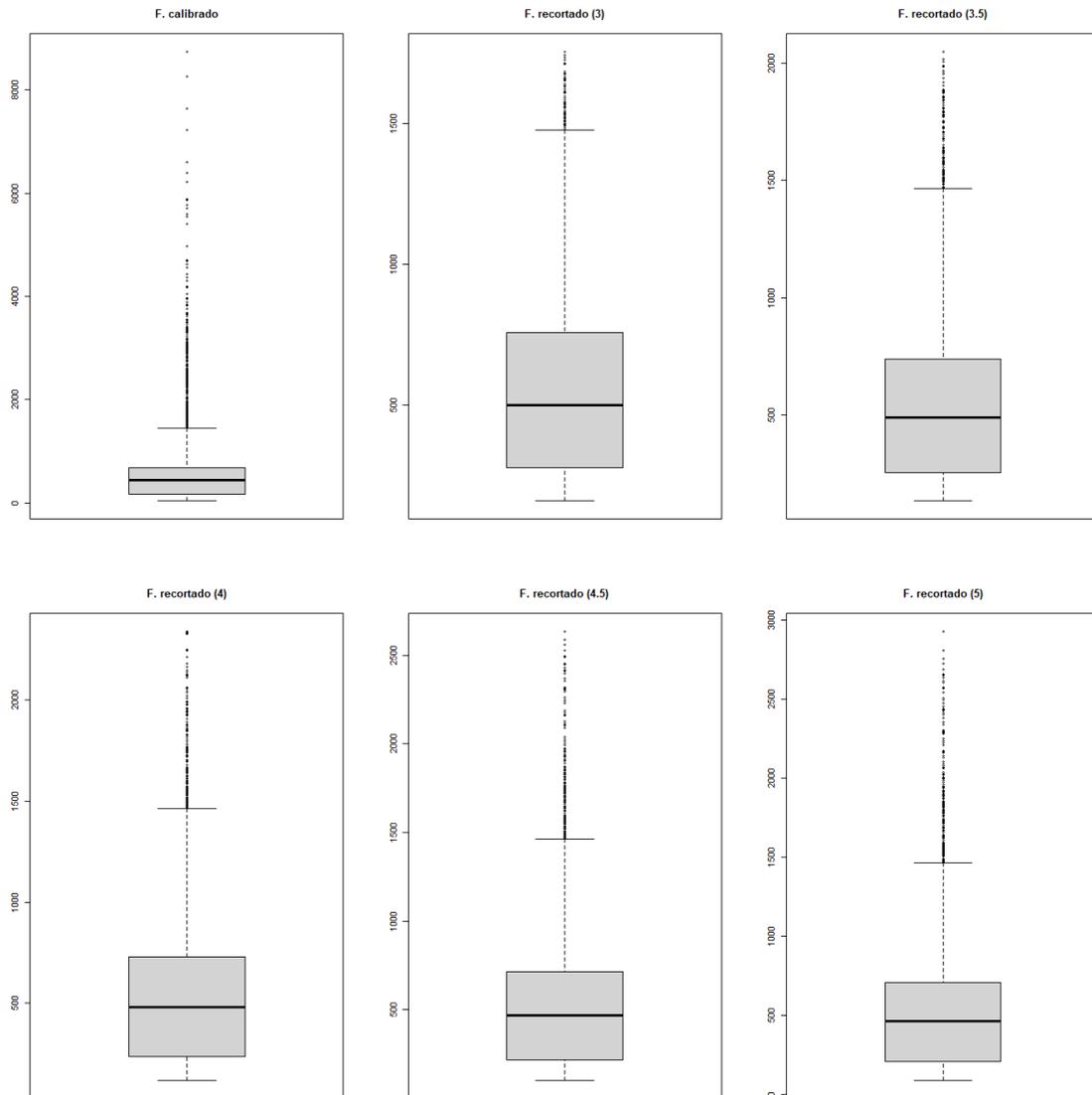
<b>Área</b>	<b>PEA</b>	<b>Desempleo</b>	<b>Pobreza</b>
Urbano	3,5	4	4
Rural	4	3	3

Niveles de recorte óptimo por celda o post estrato de calibración:

<b>Post estrato</b>	<b>PEA</b>	<b>Desempleo</b>	<b>Pobreza</b>
00_1_1_1	-	-	4
00_1_1_2	3	5	4
00_1_2_1	-	-	5
00_1_2_2	4	3	4
00_2_1_1	-	-	4
00_2_1_2	4	3	4
00_2_2_1	-	-	4
00_2_2_2	4	4	4

Nota: Los indicadores de PEA y desempleo no cuentan con niveles de recorte óptimos para todas las celdas o post estratos de calibración, debido a que estas estimaciones están calculadas por definición para todas las personas mayores o iguales a 15 años, no para los menores a 15 años.

**Anexo 5:** Diagramas de caja y bigote del factor de expansión calibrado y los recortados.



**Anexo 6:** Estadísticos descriptivos de los factores de expansión recortados según cota superior de recorte.

Estadístico descriptivo	Cota superior de recorte (veces la mediana)				
	3	3,5	4	4,5	5
Mínimo	160.5	135.6	116.8	101.6	89.18
Primer cuartil	276.3	253.6	234.9	219.9	207.75
Mediana	498.9	489.2	480.6	470.2	464.19
Media	581.7	581.7	581.7	581.7	581.70
Tercer cuartil	756.9	737.9	727.4	716.9	709.78
Máximo	1754.3	2046.7	2339.1	2631.5	2923.90

**CADA  
HECHO  
DE TU  
VIDA**  
*Cuenta*

 @ecuadorencifras

 INEC/Ecuador

 @InecEcuador

 INECEcuador

 t.me/equadorencifras

 INEC Ecuador