



Instituto Nacional de Estadística y Censos (INEC)

# Diseño Muestral de la Encuesta Nacional de Empleo, Subempleo y Desempleo Telefónica (ENEMDU – Telefónica)

## **Dirección**

Dirección de Infraestructura Estadística y Muestreo

## **Elaborado por:**

William Constante

Javier Núñez

Laura Tierra

## **Revisado por:**

Christian Garcés

## **Aprobado por:**

David Sánchez

Quito -Ecuador, 2020

## Contenido

Antecedentes.....	4
Objetivos de la operación estadística.....	4
Diseño muestral ENEMDU Telefónica.....	5
Selección y tamaño de la muestra .....	5
Métodos para corrección de sesgo.....	8
Matching entre la muestra original y la ENEMDU Telefónica.....	9
Cálculo de los factores de expansión de la ENEMDU Telefónica.....	12
Comparabilidad con periodos anteriores .....	28
Referencias .....	29
Anexos .....	31

## Lista de tablas

Tabla 1. Distribución de la muestra por periodo .....	6
Tabla 2. Casos presentes entre los identificadores de persona de la ENEMDU y la ENEMDU Telefónica.....	9
Tabla 3. Diferencia de años por rangos de edad .....	11
Tabla 4. Hogares y personas de la muestra original por mes de levantamiento	13
Tabla 5. Total de individuos respondientes y no respondientes de la ENEMDU Telefónica.....	17
Tabla 6. Estadísticos descriptivos de los Propensity Score estimados.....	18
Tabla 7. Distribución de hogares efectivos investigados por semana.....	19
Tabla 8. Deltas (ponderadores de balanceo) para cada semana .....	19

## Lista de gráficos

Gráfico 1. Diagrama de flujo del proceso de cálculo de los factores de expansión de la ENEMDU Telefónica .....	12
Gráfico 2. Soporte común de los individuos respondientes y no respondientes	18
Gráfico 3. Comparación de los factores de expansión ajustados por probabilidad de respuesta y balanceados por dominio y semana de levantamiento de información.....	20
Gráfico 4. Comparación de los factores de expansión balanceados y calibrados por dominio y grupo de edad. ....	24
Gráfico 5. Diagramas de caja y bigote del factor de expansión calibrados de la ENEMDU Telefónica.....	25
Gráfico 6. Diagramas de caja y bigote del factor de expansión recortados de la ENEMDU Telefónica.....	25
Gráfico 7. Comparación de los factores de expansión calibrados y recortados por dominio y grupo de edad .....	26

## Antecedentes

El Instituto Nacional de Estadística y Censos (INEC) ejecuta la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) desde 1993. La ENEMDU, hasta el 2019 fue una encuesta de aplicación continua, la cual se ejecutaba los doce meses del año. La información generada de ella 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.

En 2020, la emergencia sanitaria por el COVID19 alteró las actividades de la encuesta, la cual hasta marzo de 2020 se realizaba de manera totalmente presencial (cara a cara) en toda la muestra de hogares.

Ante las medidas de inmovilidad que impedían continuar con las actividades de la encuesta, el INEC, con asesoramiento internacional planificó el levantamiento de la ENEMDU Telefónica mediante llamadas telefónicas, con la finalidad de continuar proporcionando información al país sobre la situación del mercado laboral y las fuentes de ingreso de la población.

De igual manera, durante el transcurso del levantamiento de información, el INEC mantuvo varias asistencias técnicas proporcionadas por CEPAL (Ver Anexo 1), cuyo objetivo era el fortalecimiento y revisión del proceso de construcción de los factores de expansión de la ENEMDU Telefónica.

En este documento se expone de forma clara y detallada el proceso de diseño muestral de la ENEMDU Telefónica.

## Objetivos de la operación estadística

El objetivo general de la encuesta ENEMDU Telefónica es proveer información del mercado laboral ecuatoriano durante la emergencia sanitaria por el COVID-19.

Los objetivos específicos son:

1. Entregar datos de las principales categorías poblacionales en relación con el mercado de trabajo: población económicamente activa (ocupados, desempleados) e inactiva.
2. Contextualizar el efecto de la pandemia en el mercado laboral ecuatoriano.

## Diseño muestral ENEMDU Telefónica

### Selección y tamaño de la muestra

El INEC ha estado en constante acercamiento tanto con organismos internacionales como con oficinas de estadística de la región, con el fin de identificar las mejores estrategias para enfrentar la actual coyuntura social por la cual están pasando la mayoría de países de la región.

Según CEPAL “es necesario tomar decisiones sobre la muestra de hogares que se utilizará para continuar la recolección de información mientras se mantengan las limitaciones que afectan a los operativos de campo” (CEPAL, 2020).

CEPAL (2020) además manifiesta que la información que se publique de las encuestas que se levanten durante estos meses excepcionales, debe estar basada en la selección probabilística de una muestra de hogares o personas, recomendando que se adopte un cambio temporal en el diseño de las encuestas de hogares y se opte por el seguimiento de un panel en los meses subsiguientes, por cuanto la rotación de paneles y la aplicación de los diseños de muestra usuales deberá detenerse hasta que se recupere la normalidad y se pueda obtener información de forma presencial.

Asimismo, CEPAL considera que este panel de hogares o viviendas “estaría compuesto por las recopilaciones efectivas inmediatamente anteriores; por ejemplo, enero y febrero de 2020 y los últimos meses de 2019. Evidentemente, el mes de marzo de 2020 quedaría descartado al no corresponder a una recopilación usual o completa” (CEPAL, 2020).

Tomando en cuenta las recomendaciones de CEPAL, la muestra de la encuesta ENEMDU Telefónica levantada durante los meses de mayo-junio corresponde a 16.819 hogares, los cuales se seleccionaron de la siguiente manera:

- ✓ La muestra de hogares de las 5 ciudades principales (Quito, Guayaquil, Cuenca, Machala y Ambato) fue seleccionada a partir del levantamiento de información del mes de enero 2020.
- ✓ La muestra de los restos nacionales urbanos y rurales se seleccionó a partir de la ENEMDU levantada en el mes de diciembre del año 2019.

Los hogares que forman parte de la muestra son los hogares efectivos en los levantamientos previos mencionados, es decir, no se toma en cuenta los hogares que presentaban una condición de no elegibilidad (desocupados, convertidos en negocio, temporales, en construcción, etc.), o los que fueron rechazados en el levantamiento de información en campo.

### Distribución de la muestra por periodo de levantamiento de información

La muestra resultante de hogares se distribuyó en 3 periodos de levantamiento de información vía llamadas telefónicas, durante los meses de mayo y junio del presente año. En la tabla que se muestra a continuación se presenta la repartición mencionada:

**Tabla 1.** Distribución de la muestra por periodo

Periodo	Hogares	Porcentaje	Mes de levantamiento
1	5.617	33,40%	Mayo
2	5.613	33,40%	Junio
3	5.589	33,20%	Junio
<b>Total</b>	<b>16.819</b>	<b>100,00%</b>	

Como se ilustra en la Tabla 1, la distribución de la muestra es equitativa en los tres periodos del levantamiento de información, en el que cada uno abarca aproximadamente un 33% de la muestra total. Esta distribución de la muestra consideró tanto la capacidad operativa como la necesidad de contar representación equitativa de todos los periodos de referencia de información.

### Características de la muestra ENEMDU Telefónica

Esta muestra fue construida y seleccionada a partir de los hogares efectivos en levantamientos de meses inmediatamente anteriores.

Las muestras de las ENEMDU usuales son basadas en un tipo de muestreo probabilístico, cuya selección de UPM y de viviendas se realiza aleatoriamente<sup>1</sup>, lo cual implica que los hogares escogidos en la ENEMDU Telefónica, de igual manera, estén basados en una selección probabilística.

### Errores muestrales y no muestrales

Un estadístico es susceptible a diferentes fuentes de error. Särndal (2003) distingue cinco etapas en una encuesta, que van desde la planificación hasta la publicación de resultados, y sus fuentes de error asociadas, como se presenta a continuación.

#### 1. Selección de la muestra.

Esta etapa consiste en la ejecución del diseño muestral preconcebido. En esta, se calcula y selecciona un tamaño de muestra adecuado desde un marco existente o construido específicamente para la encuesta. Los errores asociados a esta etapa son:

- Errores de marco: dentro de las características que debe cumplir un marco muestral se encuentran:
  - o Completitud: dentro del marco deben constar el total de elementos muestrales existentes en la población. De no ser así, existiría un error por omisión.
  - o No duplicidad: cada registro que identifica un elemento muestral debe ser único, de otra manera se afectaría las probabilidades de selección de los mismos.

<sup>1</sup> La metodología correspondiente al diseño muestral de la Encuesta Nacional de Empleo, Desempleo y Subempleo se encuentra en el siguiente link: [https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Diseno\\_Muestral\\_2018/ENEMDU\\_Documento\\_Metodologico\\_de\\_Dise%C3%B1o\\_Muestral.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Diseno_Muestral_2018/ENEMDU_Documento_Metodologico_de_Dise%C3%B1o_Muestral.pdf)

- Errores de muestreo: es el error causado por observar una muestra en vez del total de la población.
2. Recolección de datos.  
Esta etapa engloba el plan de trabajo y la ejecución de un modo específico de recolección (entrevista personal, entrevista telefónica, entrevista vía correo electrónico, entre otros). Los errores asociados a esta etapa son:
    - Errores de medida: cuando el entrevistado da respuestas incorrectas; el entrevistador interpreta mal o influencia la respuesta del entrevistado; las preguntas del cuestionario no están claramente formuladas.
    - Errores debido a la no respuesta: información no recolectada a nivel de individuo o variable. Por ejemplo, rechazo total o parcial.
  3. Procesamiento de datos.  
En esta etapa se preparan los datos recolectados para su estimación y análisis. Los procesos involucrados son: codificación, digitación, edición, validación e imputación. Los errores asociados a esta etapa son:
    - Errores de codificación.
    - Errores de transcripción.
    - Errores introducidos o no corregidos en la edición.
    - Errores en los valores imputados.
  4. Estimación y análisis.  
En esta etapa implica el cálculo de los estadísticos de interés y sus respectivas medidas de precisión (estimador de la varianza, coeficiente de variación e intervalo de confianza). Otros análisis estadísticos pueden ser realizados, tales como comparación de subgrupos, análisis de correlación y regresión, etc. Los errores presentes en las etapas 1 a 3 afectan los estadísticos y, en el mejor de los casos, deberían ser considerados en el cálculo y estimación de sus medidas de precisión.
  5. Disseminación de resultados y evaluación postencuesta  
Esta etapa engloba la publicación de los resultados de la encuesta, incluyendo una declaración general de las condiciones alrededor de la encuesta.

Los errores en los estadísticos están divididos tradicionalmente en dos grandes categorías: error muestral y error no muestral. El error muestral es, como se mencionó, el error causado por observar una muestra en vez del total de la población. El error muestral está sujeto a la variación intra-muestras. Los errores no muestrales incluyen todos los otros errores (Särndal, 2003, p. 16).

Atendiendo a lo que señala CEPAL (2020), “cambiar el modo de recolección de información de una encuesta de hogares, de un modo presencial a un modo telefónico o a través de la web, puede traer consigo consecuencias indeseadas y, en particular, generar sesgos (de selección, de cobertura y por ausencia de respuesta) de quienes responden a la encuesta”.

En una encuesta ENEMDU tradicional, se presentan errores muestrales y no muestrales, estos últimos son tratados en el proceso de cálculo de los factores

de expansión, mediante ajustes y técnicas<sup>2</sup> que permite reducir el sesgo generado por la no respuesta o por la no cobertura.

Adicionalmente el INEC, al igual que las Organizaciones Nacionales de Estadística (ONE) de la región, evalúa previamente la proporción de la muestra original que cuenta con información de contacto válida, con miras a anticipar el posible sesgo de cobertura que se puede presentar en su realización. Sin embargo, es posible que muchas veces anticipar el sesgo o minimizarlo, no refleje la realidad una vez que se haya completado el levantamiento de la información, dado que el sesgo puede encontrarse vinculado con características que no se conocen o no se presentan en el marco de muestreo (OIT, 2020).

### **Cobertura geográfica**

La cobertura de la muestra es a nivel nacional, incluyendo todo el territorio ecuatoriano.

### **Métodos para corrección de sesgo**

Como se indicó previamente, al cambiar el modo de recolección de información de una encuesta, puede traer consigo sesgos indeseados.

El mecanismo para evitar el sesgo de no respuesta implementado por los países de la región se encuentra principalmente orientado en la ejecución de procesos mixtos, en los que se levantan encuestas vía telefónica para la mayor parte de la muestra, mientras que para los individuos de difícil contacto se utiliza encuestas cara a cara (OIT, 2020). Asimismo, es importante notar que los recursos insuficientes para cubrir completamente la muestra dan lugar a sesgos de cobertura, esto debido a que aquellas personas que no pudieron ser contactadas pueden tener altas probabilidades de encontrarse empleadas o desempleadas lo cual incide sobre los indicadores laborales (OIT, 2020).

En ese sentido, CEPAL (2020) señala varias técnicas que permiten corregir el sesgo generado por este levantamiento de información: ajuste por probabilidad de respuesta, calibración de los factores de expansión en dos etapas y postestratificación basada en modelos multinivel.<sup>3</sup>

Adicionalmente, es importante señalar que los ajustes requieren de información sobre los hogares que respondieron y no respondieron a la encuesta, para lo cual varios países se insumen de información auxiliar, como es el caso de fuentes administrativas (OIT, 2020).

---

<sup>2</sup> Para una mayor información acerca de la construcción de los factores de expansión, revisar la metodología del diseño muestral de la ENEMDU.

<sup>3</sup> Estas técnicas mencionadas se explicarán de manera más detallada en el documento metodológico que se publicará junto con los resultados de esta operación estadística.



## Matching entre la muestra original y la ENEMDU Telefónica

### Antecedentes

El proceso matching o emparejamiento de la información consiste en cotejar los formularios de un mismo hogar, es decir, se juntan los datos de los mismos individuos, encuestados en dos periodos de tiempo diferentes.

La consideración principal para la construcción de la base matching ENEMDU Telefónica, consiste en el emparejamiento de los miembros del hogar que fueron investigados en la Encuesta Nacional de Empleo, Desempleo y Subempleo – ENEMDU Telefónica y la Encuesta Nacional de Empleo, Desempleo y Subempleo – ENEMDU, ronda de diciembre 2019 (Restos) y ronda de enero 2020 (ciudades principales).

### Construcción

Todos los casos investigados en la ronda ENEMDU Telefónica, ingresan al proceso matching. Para los dominios Restos de la ENEMDU Telefónica, se realiza el emparejamiento de los hogares investigados con la ronda ENEMDU Diciembre 2019, mientras que para los dominios de las ciudades principales (Quito, Guayaquil, Cuenca, Machala y Ambato) con la ronda ENEMDU Enero 2020.

La clave única de matching que permite realizar esta actividad por medios manuales o automáticos es el identificador de persona, este identificador es la unión del identificador de Unidad Primaria de Muestreo (UPM), vivienda, hogar y código de persona; el mismo que se genera tanto en la base ENEMDU Telefónica como en las bases ENEMDU de los meses de diciembre 2019 y enero 2020, con la finalidad de unir la información de los miembros del hogar que respondieron en las 2 rondas investigadas.

Cabe resaltar, que los identificadores de personas, en las 2 rondas emparejadas no siempre son los mismos, se presentan casos como los que se puede observar en las siguientes tablas:

**Tabla 2.** Casos presentes entre los identificadores de persona de la ENEMDU y la ENEMDU Telefónica

ENEMDU Telefónica		ENEMDU Diciembre 2019	
Código Persona	Persona	Código Persona	Persona
1	A	1	B
2	B	2	C
3	C	3	A
4	D		

ENEMDU Telefónica		ENEMDU Diciembre 2019	
Código Persona	Persona	Código Persona	Persona
1	A	1	B
2	B	2	C
3	C	3	A
4	D	4	E
		5	F
		6	G

### Etapas de emparejamiento

Etapa 1.- Se realiza el emparejamiento entre las dos rondas, a través del identificador de personas, la mayoría de casos son coincidentes.

Etapa 2.- Para los casos no coincidentes entre rondas se procede a realizar un emparejamiento manual caso por caso.

Una vez realizadas estas 2 etapas de emparejamiento, el resultado es la base matching, la misma que pasa a un proceso de validación.

### Validación de la base matching

La validación de la base matching, se realiza comparando las variables nombre y apellidos completos, sexo, edad y cédula de identidad/ciudadanía. De este procedimiento es importante mencionar las respectivas validaciones de consistencia de la información que se realizan:

- Nombres y Apellidos: se presentan casos en que los nombres y/o apellidos cambian de posición entre una ronda y otra, la validación para confirmar si es la misma persona se realiza principalmente por medio del número de cédula de identidad o ciudadanía, edad, sexo y para algunos casos con la relación de parentesco.
- Escritura de Nombres y Apellidos: otro caso consiste en la escritura de los nombres y/o apellidos diferentes entre una ronda y otra, como por ejemplo: Jenny María en la ronda 1 y María Jeni en la ronda 2, sin embargo es la misma persona. Por tal razón esta persona forma parte de la base matching debido a que pertenece a las dos rondas.
- Sexo: se valida esta variable, considerando que en las 2 rondas, el sexo debe ser el mismo, es decir sexo masculino con código "1" y sexo femenino con código "2". Generalmente la inconsistencia de esta variable es por la digitación.
- Edad: tomando en cuenta la subjetividad de esta variable, donde el verdadero valor depende de varios factores externos, como el tipo de Informante sea Directo o Calificado, es una de las razones para casos no coincidentes en la edad, por lo cual se consideran diferencias en su valor

por rangos de edad, los cuales se han establecido de acuerdo a análisis realizados. La diferencia de años permitida por rangos de edad se describen en la siguiente tabla:

**Tabla 3.** Diferencia de años por rangos de edad

Rangos de edad	Diferencia en años
0 a 10	2
11 a 20	3
21 a 50	5
51 y más	10

- Cédula de Identidad/ciudadanía: para la validación de esta variable se compara el número de cédula de identidad o ciudadanía entre las 2 rondas, en el cual existen casos en que los números son diferentes entre una ronda y otra, para la validación de estos casos se verifica con las variables nombres y apellidos completos, sexo, edad, los mismos se presentan por falta de información por parte del Informante o por error en la digitación.

### Observaciones del proceso matching

- El proceso matching se inicia una vez finalizados los procesos de validación y verificación de la información recopilada en la encuesta, conformando las bases de datos definitivas de cada ronda.
- A partir de la generación del identificador de persona, se realiza el respectivo emparejamiento entre los individuos que permanecen en las dos rondas.
- Se procede a la respectiva validación a través de las variables: nombres y apellidos completos, sexo, edad y cédula de identidad o ciudadanía.
- Como resultado de lo expuesto se obtiene una base matching que garantiza una base horizontal a partir de la cual se generan estadísticas de dinámica laboral.

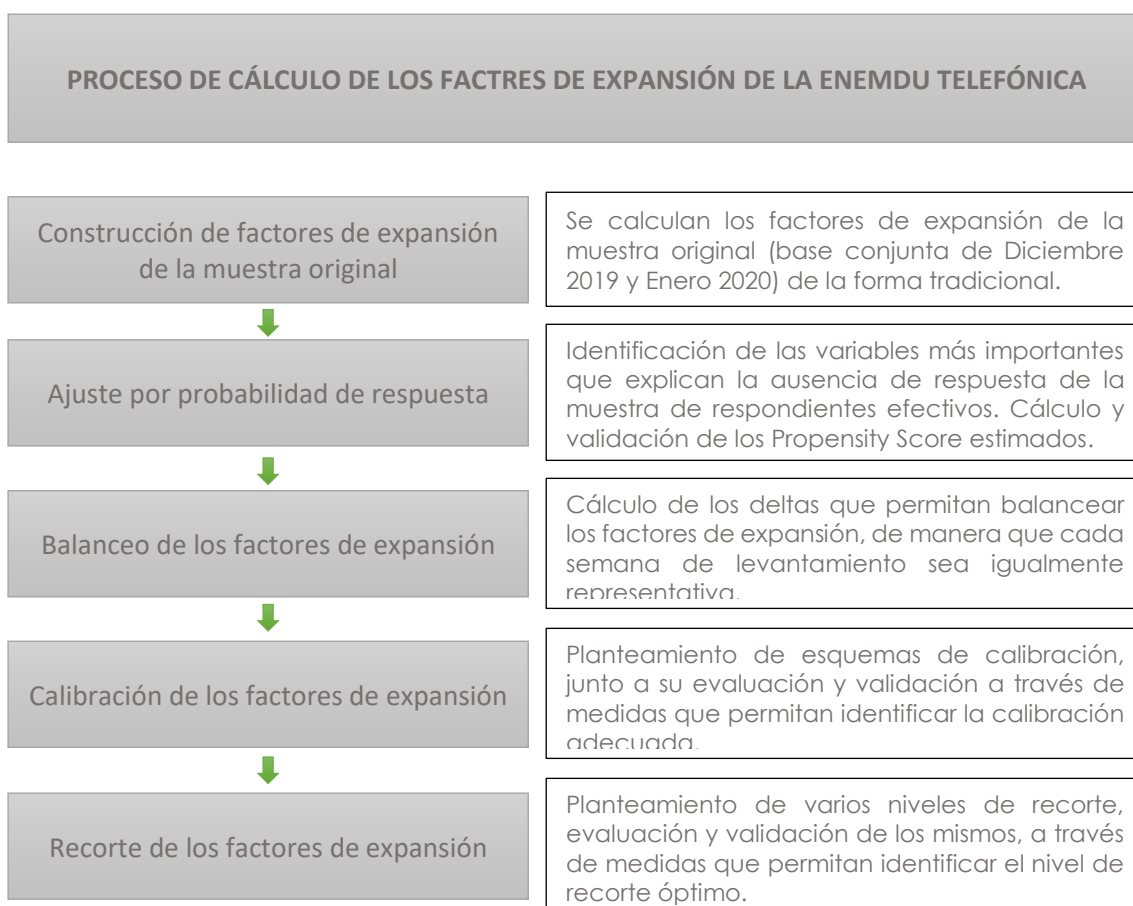
## Cálculo de los factores de expansión de la ENEMDU Telefónica

### Introducción

El proceso de cálculo de los factores de expansión de la ENEMDU Telefónica involucra varias actividades interrelacionadas, como la construcción de los pesos de muestreo de la muestra original, ajuste por probabilidad de respuesta, balanceo, calibración y recorte de los factores de expansión de la ENEMDU Telefónica. Cabe mencionar que en cada actividad de este proceso existe la validación correspondiente.

En el Gráfico 1 se muestra el diagrama de flujo del proceso de construcción de los factores de expansión de la ENEMDU Telefónica, junto con una explicación relacionada a cada actividad:

**Gráfico 1.** Diagrama de flujo del proceso de cálculo de los factores de expansión de la ENEMDU Telefónica.



## Cálculo del factor de expansión de la muestra original

Como se mencionó en la sección de diseño muestral de este documento, la muestra de la ENEMDU Telefónica fue seleccionada a partir de levantamientos anteriores. La muestra original está conformada por 58.773 personas, que forman parte de 16.819 hogares, contenidos en 2.438 UPM.

Del total de registros de la muestra original 18.556 individuos provienen de la ENEMDU Enero 2020 y 40.217 observaciones de la ENEMDU Diciembre 2019, como se presenta en la Tabla 4:

**Tabla 4.** Hogares y personas de la muestra original por mes de levantamiento

Mes	Año	Hogares	Personas
Enero	2020	5.564	18.556
Diciembre	2019	11.255	40.217
<b>Total</b>		<b>16.819</b>	<b>58.773</b>

El tipo de muestreo de la muestra original, correspondió a un muestreo probabilístico estratificado bietápico de elementos, en cuya primera etapa se seleccionaron UPM del Marco Maestro de Muestreo para encuestas de hogares del INEC, mientras que en la segunda etapa se escogen viviendas (Unidades Secundarias de Muestreo) en las UPM seleccionadas en la primera etapa.

El cálculo del factor de expansión teórico y los ajustes de cobertura correspondientes de la muestra original debe tomar en cuenta la estrategia de diseño muestral de la ENEMDU tradicional.<sup>4</sup>

Una vez calculados los factores de expansión teóricos y ajustados por cobertura de la muestra original, estos son transferidos a encuesta ENEMDU Telefónica.

## Ajuste por probabilidad de respuesta

Rosenbaum y Rubin (1983), propusieron una técnica útil para dilucidar la estructura de la ausencia de respuesta y, por consiguiente, corregir el sesgo de cobertura y el sesgo por ausencia de respuesta (Lensvelt-Mulders, Lugtig y Hubregtse, 2009).

Para el manejo efectivo de la ausencia de respuesta se consideran las variables dicotómicas  $I_k$  y  $D_k$ , que indican si el individuo pertenece a la muestra original y si ha respondido a la ENEMDU Telefónica, respectivamente. Suponiendo que la distribución de las respuestas efectivas puede ser estimada, la probabilidad de respuesta (Propensity Score) de una persona en la muestra está dada por:

$$\phi_k = \Pr(D_k = 1 \mid I_k = 1).$$

Esta probabilidad es distinta para cada persona y puede ser estimada usando los datos del panel (muestra original).

4 Para más información del diseño muestral de la ENEMDU tradicional recurrir al siguiente link: [https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Diseño\\_Muestral\\_2018/ENEMDU\\_Documento\\_Metodologico\\_de\\_Dise%C3%B1o\\_Muestral.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Diseño_Muestral_2018/ENEMDU_Documento_Metodologico_de_Dise%C3%B1o_Muestral.pdf)

Contar con la muestra original, para la cual se obtuvo toda la información del cuestionario en un período anterior, constituye un excelente punto de partida para tratar de eliminar el sesgo, puesto que se cuenta con un conjunto robusto de covariables para determinar el mejor modelo a fin de estimar el patrón de ausencia de respuesta en la muestra de respondientes efectivos (CEPAL, 2020).

### **Algoritmo de clasificación**

Otro método para formar clases para ajustes por no respuesta es a través de un algoritmo de clasificación. El objetivo es clasificar las unidades como respondientes o no respondientes en función de las covariables disponibles para todos los casos de muestra. Por lo tanto, los datos de entrada son los mismos que para el ajuste por probabilidad de respuesta (Propensity Score).

Algunas de las aplicaciones principales de los algoritmos de clasificación están en la construcción de árboles de decisión, en el cual se clasificará los individuos utilizando covariables disponibles en clases que están relacionadas con su probabilidad de ser respondientes (Valliant, 2013).

La técnica que se utilizará para este objetivo es el algoritmo de clasificación CART (Classification and Regression Trees), el cual presenta las siguientes ventajas (Valliant, 2013):

1. Las interacciones de covariables se manejan automáticamente.
2. La forma en que las covariables ingresan al modelo no tiene que hacerse explícita.
3. La selección de qué covariables e interacciones asociadas deben incluirse se realiza automáticamente.
4. Los valores de las variables, ya sean categóricos o continuos, se combinan (agrupan) automáticamente.

El algoritmo CART, Breiman et al (1984) corresponde a una técnica de aprendizaje supervisado, en el cual se crea una jerarquía de particiones que se puede representar en forma de un árbol (binario), es decir, el árbol de particiones se construye a partir de condiciones binarias respecto de las coordenadas originales de los datos.

En este caso, la variable objetivo (dependiente) es una variable dicotómica que identifica si la persona respondió o no respondió la ENEMDU Telefónica. La meta es obtener una función que permita predecir, a partir de variables predictoras (independientes), el valor de la variable objetivo para casos desconocidos.

## Identificación de las variables más relevantes que explique la respuesta o ausencia de respuesta

CEPAL (2020) manifiesta que “se debe prestar especial atención a la elección de predictores en el modelo de regresión logística, que debería funcionar bien si las variables de información auxiliar disponibles son relevantes y explicativas de la respuesta telefónica; de otra forma, esta metodología no tendrá ningún beneficio para la reducción del sesgo (y posiblemente lo exacerbará) y dará como resultado errores estándares más grandes”.

El algoritmo de clasificación CART permitirá obtener las variables más relevantes que expliquen la respuesta o ausencia de respuesta en la muestra de respondientes efectivos. En este caso, se realizó 2 modelos de clasificación, con diferentes variables explicativas, de manera que las variables importantes resultantes obtenidas del mejor modelo sean las que se utilizarán en el modelo de probabilidad de respuesta (Propensity Score).

Se debe tener en cuenta que la variable dependiente es una variable dicotómica: 1 si la persona respondió (34.935 individuos) y 0 si la persona no respondió a la ENEMDU Telefónica (23.838 individuos), dando el total de 58.773 registros de la muestra original. A continuación se detallarán las variables explicativas usadas en cada uno de los 2 algoritmos de clasificación realizados:

Variables explicativas utilizadas en el algoritmo de clasificación número 1

- Sexo.
- Sexo del jefe de hogar.
- Grupo de edad (menores de 15 años, de 15 a 65 años y mayores a 65 años) (edad).
- Mes de levantamiento (mes).
- Dominio de diseño.
- Estrato.
- Semana de levantamiento.
- Relación de parentesco.
- Nivel de instrucción.
- Autoidentificación étnica.
- Condición de actividad.
- Ingreso per cápita.
- Tasa de pobreza (Dic 2019).
- Tasa de extrema pobreza (Dic 2019).
- Empleo en el sector formal.
- Empleo en el sector informal.
- Tipo de empleador (público y privado).
- Tipo de trabajo (asalariado e independiente).
- Total de personas del hogar.
- Razón de dependencia por hogar.

VARIABLES explicativas utilizadas en el algoritmo de clasificación número 2

- Las variables utilizadas en el algoritmo de clasificación número 2 corresponden a las variables del algoritmo de clasificación número 1 dicotomizadas.

A través de la aplicación de los dos algoritmos de clasificación, se obtuvo como resultado las variables más importantes que explican la respuesta o no respuesta de los individuos pertenecientes a la muestra de respondientes efectivos (Ver Anexos 2 y 3); con este insumo se procedió a calcular los modelos de probabilidad de respuesta.

CEPAL (2020) señala que si se asume que la probabilidad de respuesta depende de alguna combinación lineal de las covariables disponibles en la muestra original, es posible ajustar un modelo en que la variable dependiente es  $D_k$  y un vector de covariables independientes.

Kim y Riddles (2012) muestran que es posible utilizar un modelo basado en el ajuste de la probabilidad de respuesta de la muestra telefónica mediante la siguiente expresión:

$$\text{logit}(\hat{\varphi}_k) = X_k \hat{\beta},$$

Donde  $\hat{\beta}$  es el vector de coeficientes estimado de la regresión logística y  $X_k$  representa la matriz de covariables más importantes.

### Modelos logísticos realizados

Con la finalidad de obtener la probabilidad estimada de respuesta o no respuesta, se efectuaron 4 modelos regresión logística, en los cuales la variable dependiente es dicotómica e identifica si el individuo respondió a la muestra telefónica (1) o no (0).

A continuación se detallan las variables independientes utilizadas en los 4 modelos efectuados:

- Modelo 1: Sexo y grupo de edad.
- Modelo 2: Las variables descritas en el Anexo 2 de este documento.
- Modelo 3: Las variables descritas en el Anexo 2 de este documento más la variable grupo de edad.
- Modelo 4: Las variables descritas en el Anexo 3 de este documento.

Los resultados de los 4 modelos planteados sugieren que el mejor modelo logístico es el segundo o tercer modelo, por cuanto presentan el menor criterio de información de Akaike (Ver Anexo 4); sin embargo, la variable grupo de edad añadida en el tercer modelo no es una variable importante para explicar la respuesta en la ENEMDU Telefónica, por cuanto los puntajes de propensión estimados únicamente se encuentran en el rango de 0,55 hasta 0,61 (Ver Anexo 5); llegando a la conclusión que el mejor modelo para explicar la respuesta o ausencia de la misma en la muestra de respondientes efectivos es el Modelo 2. Con este modelo, se obtienen las probabilidades estimadas  $\hat{\varphi}_k$  (Propensity



Score) para respondientes y no respondientes de la muestra de respondientes efectivos, y a su vez estas probabilidades son transferidas a la muestra telefónica, entonces el factor de expansión ajustado por probabilidad de respuesta toma la siguiente forma:

$$w_k = \frac{d_k}{\hat{\phi}_k}$$

Donde  $d_k$  corresponde a los factores de expansión de la muestra original que se transfieren a la ENEMDU Telefónica, y  $\hat{\phi}_k$  son los Propensity Score estimados, de igual manera, transferidos a la ENEMDU Telefónica.

Utilizar el factor de expansión ajustado por probabilidad de respuesta en el cálculo de los estimadores deseados minimizaría el sesgo de selección que se generó por el cambio de modo de recolección de la información (CEPAL, 2020).

### Validación de los Propensity Score estimados

De los 37.406 individuos de los cuales se levantó información en la ENEMDU Telefónica, se presentan 34.935 personas que hicieron el match entre las 2 rondas, es decir, entre la muestra original y la ENEMDU Telefónica. Por tanto, de los 37.406 respondientes efectivos solo se utilizan los 34.935 registros en la transferencia del Propensity Score de la muestra original a la ENEMDU Telefónica, debido al matching. La diferencia de 2.471 registros, corresponden a individuos que actualmente viven en los hogares de la muestra original, sin embargo, no estuvieron presentes al momento de realizarse la encuesta en Diciembre 2019 o Enero 2020.

En la Tabla 5 se aprecia que la muestra original está conformada por 58.773 individuos, de los cuales 34.935 personas respondieron, mientras que 23.838 individuos no respondieron a la ENEMDU Telefónica.

**Tabla 5.** Total de individuos respondientes y no respondientes de la ENEMDU Telefónica

<b>Muestra original</b>	<b>Individuos</b>	<b>Porcentaje</b>
Respondientes	34.935	59,44%
No respondientes	23.838	40,56%
<b>Total</b>	<b>58.773</b>	<b>100,00%</b>

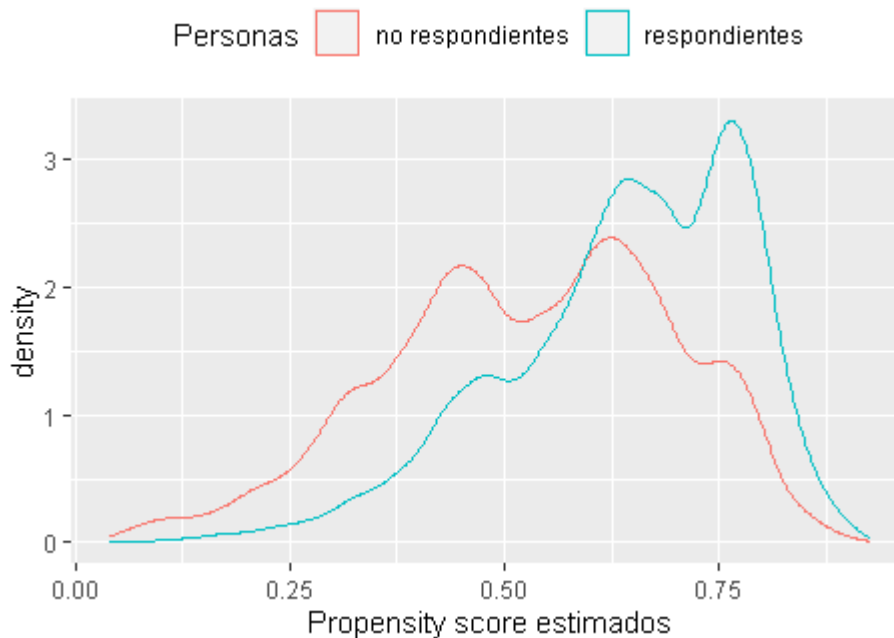
Los puntajes de propensión estimados a partir de un modelo logístico, deben estar en un rango entre 0 y 1, además, la sumatoria de estos puntajes debe ser igual al número de respondientes efectivos de la ENEMDU Telefónica, que en este caso son 34.935 individuos.

**Tabla 6.** Estadísticos descriptivos de los Propensity Score estimados

Estadístico	Valor
Mínimo	0,0406
Primer cuartil	0,4801
Mediana	0,6199
Media	0,5944
Tercer cuartil	0,7246
Máximo	0,9246

Por otra parte, otra validación que se debe realizar corresponde a la verificación de soporte común, este ejercicio compara las funciones de densidad de los individuos de la muestra original que respondieron o no respondieron la ENEMDU Telefónica, gráfico que se muestra a continuación:

**Gráfico 2.** Soporte común de los individuos respondientes y no respondientes



Como se ilustra en el Gráfico 2, tanto las personas de la muestra original que respondieron como los que no respondieron la ENEMDU Telefónica comparten las mismas cotas o extremos inferiores o superiores de los Propensity Score estimados.

### Balanceo de los factores de expansión

La recolección de información de la ENEMDU Telefónica se realizó durante 9 semanas, durante el 5 de Mayo al 30 de Junio del año 2020.

**Tabla 7.** Distribución de hogares efectivos investigados por semana

Semana	Hogares	Hogares (%)
1	813	7,82%
2	867	8,34%
3	942	9,06%
4	705	6,78%
5	524	5,04%
6	752	7,23%
7	590	5,67%
8	4.036	38,82%
9	1.168	11,23%
<b>Total</b>	<b>10.397</b>	<b>100,00%</b>

Como se puede evidenciar en la Tabla 7, el levantamiento de información no fue homogéneo, es decir, el número de hogares investigados en cada semana no fue el mismo, por ejemplo, en la octava semana se levantó la información de un 38,82% de hogares, mientras que en la quinta semana se recolectó los datos de un 5,04% de hogares. Por tal motivo, se vuelve necesario balancear los factores de expansión, con la finalidad de representar equitativamente a cada periodo de referencia de información.

El balanceo de los pesos de muestreo se realizó con el siguiente algoritmo:

$$\delta_s = \frac{\hat{N}}{9 * \sum_s w_{k_s}}$$

Donde:

- $\delta_s$  = *ponderador de balanceo de los factores de expansión.*
- $\hat{N}$  = *total de individuos estimado con el factor de expansión ajustado por probabilidad de respuesta.*
- $w_{k_s}$  = *factor de expansión ajustado por probabilidad de respuesta en la semana de levantamiento s.*

**Tabla 8.** Deltas (ponderadores de balanceo) para cada semana

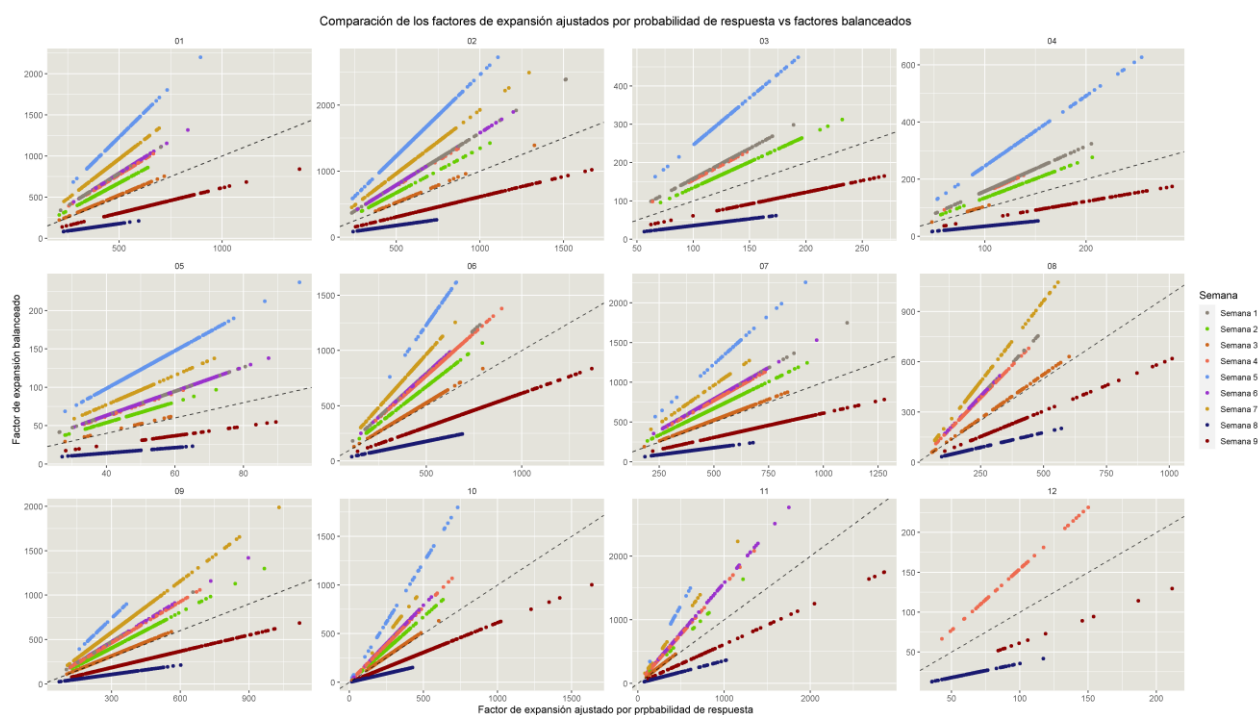
Semana	Delta
1	1,210
2	1,419
3	1,097
4	1,782
5	2,087
6	1,709
7	1,794
8	0,304
9	0,926

En la Tabla 8 se muestra los deltas (ponderadores de balanceo) que se utilizarán para balancear los factores de expansión. Con esto se logra que cada semana de levantamiento de información tenga la misma representatividad a pesar de que la recolección de información no fue homogénea; por ejemplo, la octava semana al tener una mayor muestra levantada, tiene el menor delta (0,304); mientras que la quinta semana al haber presentado la menor proporción de hogares investigados, tendrá el mayor delta (2,087).

Los factores de expansión balanceados resultan de la multiplicación del factor de expansión ajustado por probabilidad de respuesta por el ponderador que balancea la muestra para cada semana (delta), de la siguiente forma:

$$w_k^{(*)} = w_k * \delta_k.$$

**Gráfico 3.** Comparación de los factores de expansión ajustados por probabilidad de respuesta y balanceados por dominio y semana de levantamiento de información.



Como se ilustra en el Gráfico 3, cada cuadrado representa un dominio de diseño, en el cual los cinco primeros cuadros corresponden a las ciudades principales, los seis siguientes son los restos urbanos y rurales por región (Sierra, Costa y Amazonía) y el último es la Región Insular. Por otra parte, la línea entrecortada representa una línea de 45 grados.

Se identifica un comportamiento similar en todos los dominios de diseño, en los cuales el balanceo de los ponderadores en la octava y novena semana de levantamiento reduce el valor de los factores de expansión con respecto a los pesos de muestreo ajustados por probabilidad de respuesta; mientras que en el resto de semanas los factores de expansión balanceados son mayores a los ajustados por probabilidad de respuesta.

## 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.
5. Al calcular los factores de expansión, es necesario realizar ciertos ajustes, los cuales vienen dados por problemas de cobertura o de rechazo; es por tanto que la calibración es usualmente el último ajuste que se efectúa a los factores de expansión, haciendo uso de información auxiliar que reduce la varianza y corrige los problemas de cobertura que no pudieron ser corregidos en los ajustes previos (Gutiérrez, 2018).

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 determina 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 Telefónica, se calibra los pesos de muestreo balanceados, 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 balanceados y  $g_k$  son los pesos de calibración.

Para la ENEMDU Telefónica, se aplicó dos esquemas de calibración (Ver Anexos 6 y 7), y se evaluó cada uno a través de criterios para validar la calibración propuestas por Silva (2004).

## 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 peses extremos (límite inferior)

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

- Proporción de peses 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 peses de calibración y los peses originales

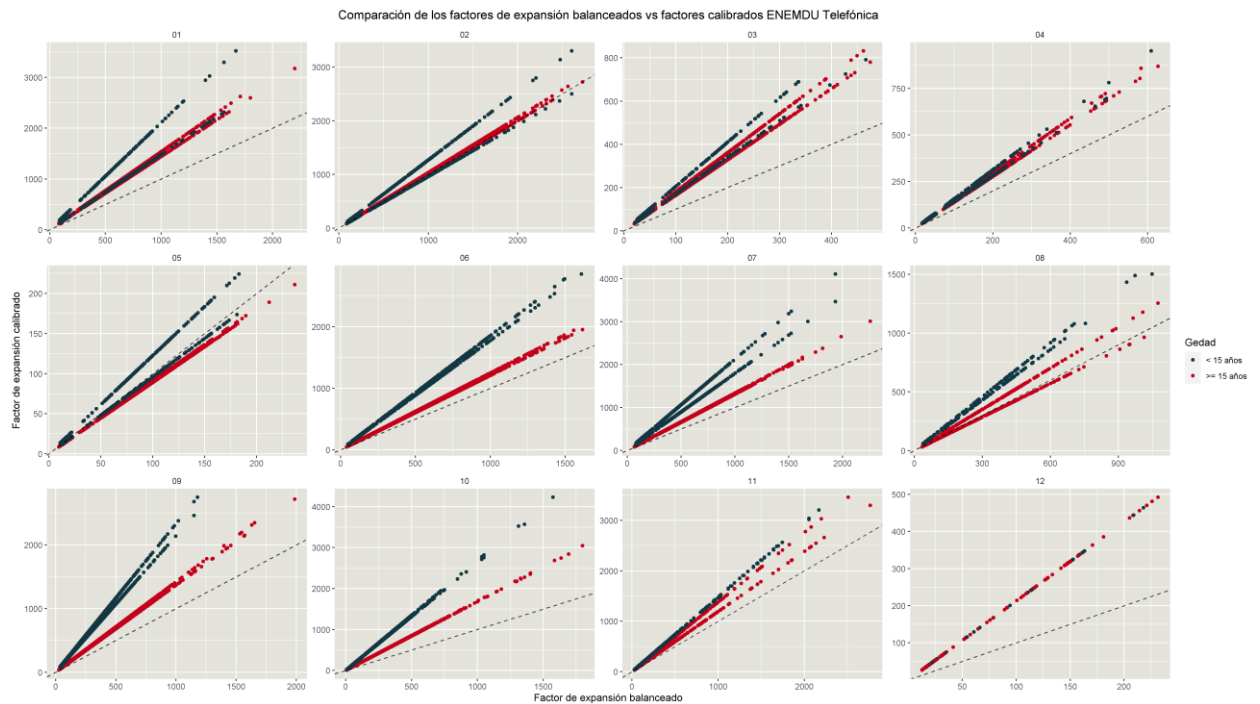
$$M6 = \frac{1}{n} \sum_{k \in S} \frac{(w_k - d_k)^2}{d_k} = \frac{1}{n} \sum_{k \in S} d_k (g_k - 1)^2.$$

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

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

Los dos esquemas de calibración considerados para la ENEMDU Telefónica fueron comparados a través de los criterios propuestos por Silva (2004), eligiendo el más adecuado, obteniendo como resultado que el esquema de calibración más adecuado es el segundo, en cuyo caso se planteaban 45 celdas o post estratos de calibración (Ver Anexos 8 y 9).

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



Según el Gráfico 4, en la mayoría de dominios, los factores de expansión calibrados son mayores a los balanceados, por cuanto, la calibración reproduce 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 dominios de diseño.

### Recorte de los factores de expansión extremos

Debido a los ajustes por probabilidad de respuesta, balanceo y calibración, la variabilidad de los factores de expansión de la ENEMDU Telefónica 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 Telefónica 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:

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

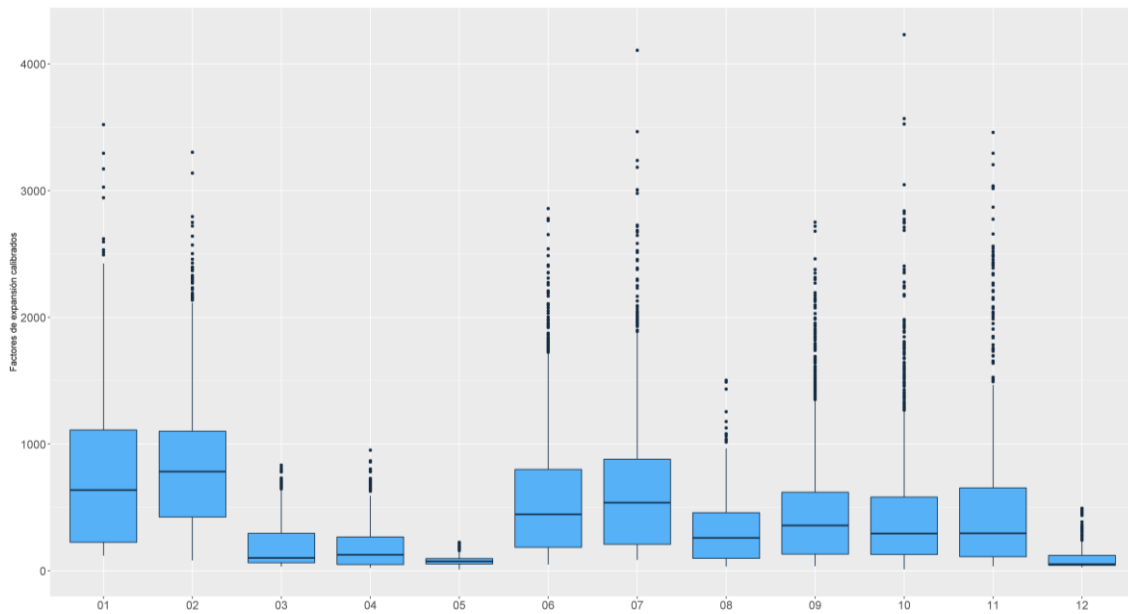
Donde  $fexp_{trim}$  es el conjunto de los factores de expansión recortados.



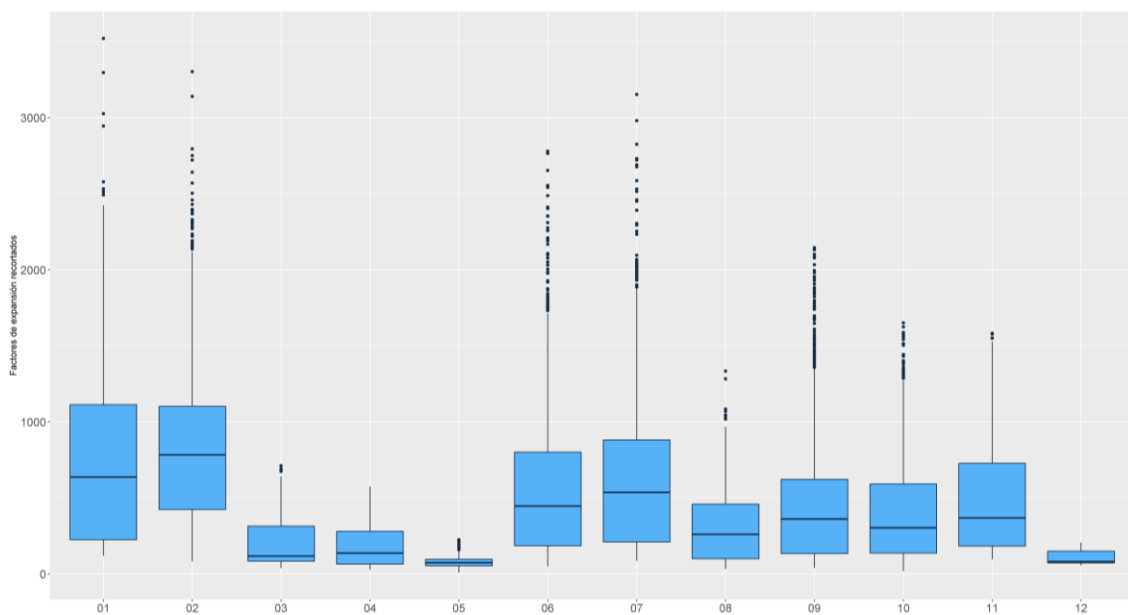
- (3) Calcular  $K = \sum_k |fexp - fexp_{trim}|$ .
- (4) Distribuir  $K$  equitativamente entre todos los pesos no recortados.
- (5) Repetir los pasos (1)-(4) hasta que  $K = 0$ .

Puesto que la ENEMDU Telefónica presenta factores de expansión extremos únicamente en la cola derecha de su distribución, no se fija  $L$  y se fija  $U$  en 4 veces la mediana. Cabe señalar que, el recorte y redistribución de los factores de expansión se realizó de manera independiente para cada celda o post estrato de calibración (detallados en el segundo esquema de calibración).

**Gráfico 5.** Diagramas de caja y bigote del factor de expansión calibrados de la ENEMDU Telefónica.



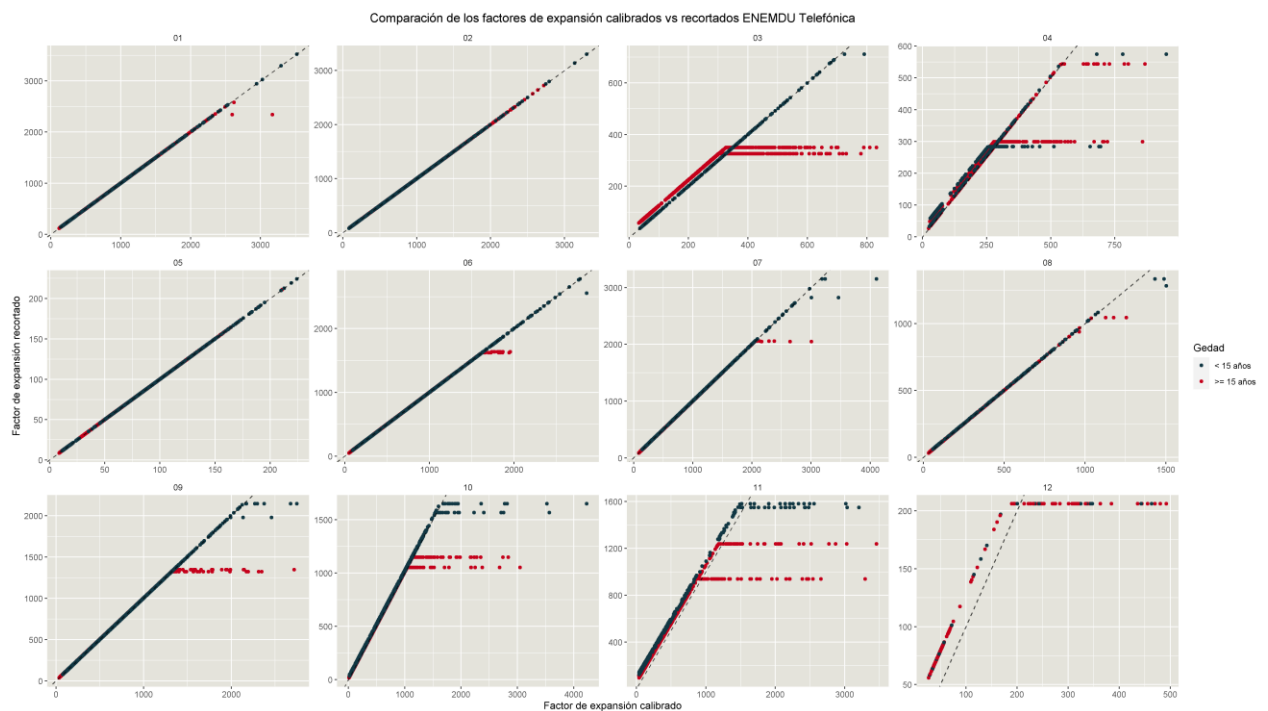
**Gráfico 6.** Diagramas de caja y bigote del factor de expansión recortados de la ENEMDU Telefónica.



Por un lado, en los Gráficos 5 y 6 se evidencia que los factores de expansión calibrados presentan una mayor variabilidad que los recortados, de igual manera, la función de distribución de los ponderadores calibrados tienen valores extremos más grandes en comparación con la función de distribución de los ponderadores recortados, identificando el efecto del recorte y redistribución en la eliminación de valores atípicos o extremos de pesos de muestreo.

Según el Gráfico 7, existen algunos dominios donde se presenta una mayor cantidad de factores de expansión recortados, como por ejemplo Cuenca, Machala, Sierra, Costa, Amazonía Rural y la Región Insular, es decir, existían algunos individuos que presentaban ponderadores extremos que fueron recortados y redistribuidos, con la finalidad de reducir la variabilidad de los factores de expansión.

**Gráfico 7.** Comparación de los factores de expansión calibrados y recortados por dominio y grupo de edad.



### Validación del recorte de los factores de expansión extremos

Para la ENEMDU Telefónica 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 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 20% (Ver Anexo 10).

## 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 2 indicadores principales: PEA y desempleo, 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 11).

Según los resultados presentados en el Anexo 11, se aprecia que en su mayoría, los niveles de recorte óptimos se encuentran en 3 y 3,5 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 4, 4,5 o 5 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 una revisión de la cobertura de la ENEMDU Telefónica, 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 12), el recorte y redistribución reduce la variabilidad de los factores de expansión. Además, según se evidencia en las tablas que se muestran a continuación, 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 13).

### **Estimación de características y cálculo de errores**

Para la estimación de características y cálculo de errores, referirse al documento de diseño muestral de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) tradicional.

### **Comparabilidad con periodos anteriores**

La modificación en la metodología de levantamiento de información, supone una pérdida de comparabilidad con los indicadores del mercado laboral de periodos anteriores, no obstante, esto puede solucionarse aplicando metodologías que permitan un empalme de la serie, así lo señala CEPAL (2020), "este tipo de modificaciones plantean un escenario en que las autoridades nacionales deben asumir la pérdida de comparabilidad de la serie de indicadores de empleo e ingresos. Esta pérdida de comparabilidad puede remediarse de manera parcial más adelante, aplicando instrumentos que recojan la información retrospectivamente. No obstante, ello no implica que se deba interrumpir la recolección de información sobre el empleo y las demás estadísticas sociales."

## Referencias

Breiman L., Friedman J., Stone C., Olshen R. (1993). *Classification and Regression Trees*. Chapman & Hall, London

CEPAL. (2020). *Recomendaciones para la publicación de estadísticas oficiales a partir de encuestas de hogares frente a la coyuntura de la enfermedad por coronavirus (COVID-19)*. Obtenido de [https://rtc-cea.cepal.org/sites/default/files/rtc\\_connected/files/recomendaciones-encuestas-hogares-covid-final\\_1.pdf](https://rtc-cea.cepal.org/sites/default/files/rtc_connected/files/recomendaciones-encuestas-hogares-covid-final_1.pdf)

CEPAL. (2020). *Recomendaciones para eliminar el sesgo de selección en las encuestas de hogares en la coyuntura de la enfermedad por coronavirus (COVID-19)*. Obtenido de [https://repositorio.cepal.org/bitstream/handle/11362/45552/1/S2000316\\_es.pdf](https://repositorio.cepal.org/bitstream/handle/11362/45552/1/S2000316_es.pdf)

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.

Kim, J. K. y M. K. Riddles (2012), "Some theory for propensity-score-adjustment estimators in survey sampling", *Survey Methodology*, vol. 38, N° 2.

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

Lensvelt-Mulders, G., P. Lugtig y M. Hubregtse (2009), "Separating selection bias and non-coverage in Internet panels using propensity matching", *Survey Practice*, 2, N° 6.

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

Rosenbaum, P. R. y D. B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, vol. 70, N° 1.

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: Asistencias técnicas proporcionadas por CEPAL.

Fecha	Temas tratados
24/06/2020	Presentación del diseño muestral de la ENEMDU Telefónica a CEPAL y OIT, concerniente a la selección y tamaño de muestra, levantamiento de información en campo, prueba piloto, estructura del formulario, etc.
01/07/2020	Presentación de la cobertura efectiva de la ENEMDU Telefónica a CEPAL y OIT, además de algunos avances en el proceso de cálculo de los factores de expansión.
06/07/2020	Formulación definitiva del proceso de cálculo de los factores de expansión de la ENEMDU Telefónica, realización el diagrama de flujo del proceso junto con CEPAL y OIT.
13/07/2020	Presentación a CEPAL de los avances del proceso de cálculo de los factores de expansión, en lo concerniente a cálculo de los ponderadores de la muestra original, balanceo y calibración (estadísticos descriptivos, tablas y gráficos)
20/07/2020	Presentación a CEPAL de los avances del proceso de cálculo de los factores de expansión, en lo concerniente al matching entre la muestra original y la ENEMDU Telefónica, evaluación de la calibración y recorte de los ponderadores
27/07/2020	Revisión del modelo de ajuste por probabilidad de respuesta, algoritmos de clasificación y validación del Propensity Score.
30/07/2020	Presentación a CEPAL de los resultados del modelo de ajuste por probabilidad de respuesta y de la validación de los Propensity score estimados.
03/08/2020	Presentación de los resultados de los indicadores de mercado laboral ante CEPAL y OIT.

### Anexo 2: Importancia de las variables utilizadas en el algoritmo de clasificación número 1.

Variable	Importancia
Estrato	2166,56
Semana de levantamiento	1447,65
Dominio de diseño	584,56
Ingreso per cápita	157,11
Mes de levantamiento	120,80
Autoidentificación étnica	104,11
Tasa de pobreza (Dic 2019)	78,18
Tasa de extrema pobreza (Dic 2019)	29,32
Nivel de instrucción	18,17
Total de personas del hogar	8,30
Condición de actividad	5,66
Razón de dependencia	0,13

**Anexo 3:** Importancia de las variables utilizadas en el algoritmo de clasificación número 2.

<b>Variable</b>	<b>Importancia</b>
Semana 9 de levantamiento	989,88
Semana 8 de levantamiento	515,12
Dominio Amazonía Rural	287,01
Dominio Ambato	160,39
Estrato 2912	70,23
Estrato 2913	55,59
Nivel de instrucción Ninguno	49,09
Estrato 2611	45,60
Ingreso per cápita	44,12
Estrato 1421	39,05
Estrato 2911	34,57
Tasa de pobreza (Dic 2019)	30,89
Otros parientes	23,28
Estrato 1422	22,84
Estrato 2221	16,07
Estrato 1522	14,25
Estrato 1621	12,77
Estrato 2122	12,77
Estrato 2222	12,43
Estrato 2121	12,15
Mestizo	11,03
Tasa de extrema pobreza (Dic 2019)	9,24
Dominio Costa Rural	8,93
Estrato 2812	6,03
Estrato 0621	5,63
Indígena	4,81
Estrato 1311	4,73
Estrato 1321	4,46
Estrato 1322	3,46
Etnia perdida	3,17
Nivel de instrucción perdido	3,17
Montubio	2,45
Estrato 0822	1,77
Estrato 0923	1,39
Estrato 0922	1,07
Estrato 0821	0,89
Estrato 1121	0,48

**Anexo 4:** Criterios de información de Akaike de los modelos logísticos realizados.

<b>Modelo</b>	<b>Akaike</b>
Modelo 1	79.313
Modelo 2	73.189
Modelo 3	73.189
Modelo 4	74.137



**Anexo 5:** Estadísticos descriptivos de los Propensity Score estimados del primer modelo realizado.

Estadístico	Valor
Mínimo	0,5554
Primer cuartil	0,5850
Mediana	0,6002
Media	0,5944
Tercer cuartil	0,6093
Máximo	0,6093

**Anexo 6:** Primer esquema de calibración realizado.

Reproducción de totales poblacionales a nivel nacional: hombres, mujeres, individuos del área urbana, área rural, personas menores de 15 años y personas mayores o iguales a 15 años:

Desagregación Población	
Hombres	8.602.490
Mujeres	8.973.160
Urbano	11.965.963
Rural	5.609.687
Menores	5.087.042
Mayores	12.488.608

**Anexo 7:** Segundo esquema de calibración realizado.

Reproducción de 45 totales poblacionales desagregados por dominio de diseño, sexo (hombre y mujer) y grupo de edad (menores de 15 años, y mayores o iguales a 15 años). Cabe mencionar que para la Región Insular la calibración está realizada únicamente para el total poblacional (32.848 personas), es decir, sin desagregación por sexo o grupo de edad. Estas desagregaciones se denominan celdas o post estratos de calibración.

Dominio de diseño	Población menor a 15 años		Población mayor o igual a 15 años	
	Hombre	Mujer	Hombre	Mujer
Quito	200.646	223.146	746.739	807.251
Guayaquil	310.334	329.590	972.596	1.034.741
Cuenca	45.376	52.351	144.162	166.193
Machala	32.488	33.250	102.971	105.315
Ambato	21.220	23.677	71.040	79.268
Resto Sierra Urbano	334.244	363.566	825.399	892.719
Resto Costa Urbano	575.242	605.498	1.222.414	1.291.299
Amazonía Urbano	61.921	60.702	118.407	112.198
Sierra Rural	434.816	472.839	959.141	1.043.069
Costa Rural	348.010	313.676	734.928	662.426
Amazonía Rural	124.700	109.620	198.824	174.790

**Anexo 8:** Estadísticos descriptivos de los pesos (g) de calibración y de los factores de expansión calibrados.

Estadístico descriptivo	Peso g de calibración		Factor calibrado	
	Esquema 1	Esquema 2	Esquema 1	Esquema 2
Mínimo	1,20	9,20	0,89	8,70
Primer cuartil	1,23	139,20	1,25	137,35
Mediana	1,23	279,71	1,38	299,71
Media	1,44	469,86	1,47	469,86
Tercer cuartil	1,54	691,73	1,68	705,26
Máximo	2,11	4.473,95	2,69	4.230,94

**Anexo 9:** Medidas para evaluar los esquemas de calibración planteados

Medida 1:

Dominio	Esquema 1	Esquema 2
Quito	0,1691	0,0000
Guayaquil	0,3590	0,0000
Cuenca	0,1984	0,0000
Machala	0,1351	0,0000
Ambato	0,4493	0,0000
Resto Sierra Urbano	0,0546	0,0000
Resto Costa Urbano	0,1122	0,0000
Amazonía Urbano	0,1421	0,0000
Sierra Rural	0,0805	0,0000
Costa Rural	0,1557	0,0000
Amazonía Rural	0,3053	0,0000
Región Insular	0,2994	0,0000

Medida 2:

Dominio	Esquema 1	Esquema 2
Quito	0,0787	0,0787
Guayaquil	0,0612	0,0612
Cuenca	0,0962	0,0962
Machala	0,1076	0,1076
Ambato	0,0674	0,0674
Resto Sierra Urbano	0,0628	0,0628
Resto Costa Urbano	0,0535	0,0535
Amazonía Urbano	0,1044	0,1044
Sierra Rural	0,0516	0,0516
Costa Rural	0,0584	0,0584
Amazonía Rural	0,1451	0,1451
Región Insular	0,1805	0,1810

Medida 3:

<b>Dominio</b>	<b>L</b>	<b>n</b>	<b>Esquema 1</b>	<b>Esquema 2</b>
Quito	1,44	2679	2183	0
Guayaquil	0,96	3327	0	0
Cuenca	1,64	2122	1666	921
Machala	1,38	1620	1234	0
Ambato	0,89	2661	0	0
Resto Sierra Urbano	1,21	4414	1663	0
Resto Costa Urbano	1,31	6141	4674	0
Amazonía Urbano	0,95	1149	0	444
Sierra Rural	1,37	6668	0	0
Costa Rural	1,68	5126	3953	2038
Amazonía Rural	1,19	1197	0	0
Región Insular	2,13	302	302	302

Medida 4:

<b>Dominio</b>	<b>U</b>	<b>n</b>	<b>Esquema 1</b>	<b>Esquema 2</b>
Quito	2,11	2679	0	0
Guayaquil	1,27	3327	823	0
Cuenca	2,05	2122	0	0
Machala	1,56	1620	386	187
Ambato	1,22	2661	1663	239
Resto Sierra Urbano	1,85	4414	0	454
Resto Costa Urbano	2,12	6141	0	702
Amazonía Urbano	1,53	1149	313	0
Sierra Rural	2,33	6668	0	749
Costa Rural	2,69	5126	0	0
Amazonía Rural	1,48	1197	1197	190
Región Insular	2,13	302	0	0

Medida 5:

Dominio	Media peso g		CV del peso g	
	Esquema 1	Esquema 2	Esquema 1	Esquema 2
Quito	1,30	1,53	1,7334E-15	1,3453E-15
Guayaquil	1,33	1,04	2,7738E-15	3,9152E-15
Cuenca	1,31	1,75	5,8001E-16	8,608E-16
Machala	1,32	1,44	2,5238E-15	1,3325E-15
Ambato	1,30	0,93	3,2205E-15	4,9808E-16
Resto Sierra Urbano	1,31	1,35	1,2463E-15	5,258E-15
Resto Costa Urbano	1,32	1,47	4,5724E-15	1,2941E-15
Amazonía Urbano	1,34	1,17	2,7323E-15	1,7883E-15
Sierra Rural	1,65	1,58	2,529E-15	4,1634E-15
Costa Rural	1,65	1,91	2,6791E-15	3,9833E-16
Amazonía Rural	1,69	1,34	8,1193E-16	3,1125E-16
Región Insular	1,48	2,13	7,3261E-16	0

Medida 6:

Dominio	Esquema 1	Esquema 2
Quito	57,9441	152,2062
Guayaquil	104,7556	6,0337
Cuenca	14,8998	63,3196
Machala	16,2531	23,0866
Ambato	9,8119	1,0352
Resto Sierra Urbano	53,3391	74,6153
Resto Costa Urbano	57,8754	123,4275
Amazonía Urbano	40,7021	19,5336
Sierra Rural	131,2838	126,9698
Costa Rural	101,4727	209,3562
Amazonía Rural	223,1448	51,3567
Región Insular	15,7478	64,9458

**Anexo 10:** Porcentaje de factores de expansión recortados por celda o post estrato de calibración.

Se evaluó el porcentaje de factores de expansión recortados por celda o post estrato de calibración, este post estrato está conformado por 5 dígitos, por ejemplo, el post estrato "04\_1\_1\_1" está construido de acuerdo al siguiente detalle:

Dígito	Descripción
04	Dominio de diseño (Machala)
1	Área (Urbana)
1	Grupo de edad (menores a 15 años)
1	Sexo (Hombres)

En este caso, el post estrato "04\_1\_1\_1" identifica a los individuos hombres, menores a 15 años, ubicados en la ciudad de Machala<sup>5</sup>.

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

id_calib	No recortados	Recortados	Porcentaje
04_1_1_1	124	75	37,69%
04_1_2_2	395	234	37,20%
03_1_2_1	481	264	35,44%
03_1_2_2	602	319	34,64%
12_0_0_0	230	72	23,84%
11_2_1_2	130	39	23,08%
11_2_1_1	147	43	22,63%
11_2_2_2	344	74	17,70%
11_2_2_1	360	60	14,29%
10_2_1_2	482	56	10,41%
10_2_1_1	574	61	9,61%
10_2_2_2	1792	123	6,42%
10_2_2_1	1917	121	5,94%
04_1_2_1	575	30	4,96%
08_1_1_2	139	7	4,79%
01_1_1_1	263	12	4,36%
06_1_2_1	1592	71	4,27%
09_2_2_1	2366	104	4,21%
03_1_1_2	215	9	4,02%
09_2_1_2	720	29	3,87%
04_1_1_2	180	7	3,74%
09_2_1_1	730	26	3,44%
09_2_2_2	2602	91	3,38%
06_1_1_1	439	15	3,30%
06_1_1_2	439	15	3,30%
01_1_2_2	1140	34	2,90%
08_1_2_1	381	11	2,81%
06_1_2_2	1793	50	2,71%
01_1_2_1	984	25	2,48%
03_1_1_1	227	5	2,16%
07_1_1_2	689	13	1,85%
01_1_1_2	217	4	1,81%
08_1_2_2	437	7	1,58%
07_1_1_1	755	10	1,31%
07_1_2_2	2360	30	1,26%
08_1_1_1	165	2	1,20%
02_1_1_1	444	5	1,11%
02_1_1_2	370	4	1,07%
07_1_2_1	2267	17	0,74%
02_1_2_1	1188	3	0,25%
02_1_2_2	1310	3	0,23%
05_1_2_2	1155	1	0,09%
05_1_1_1	268	0	0,00%
05_1_1_2	239	0	0,00%
05_1_2_1	998	0	0,00%

<sup>5</sup> Cabe indicar que los dominios de las 5 ciudades principales: Quito, Guayaquil, Cuenca, Machala y Ambato forman parte únicamente del área Urbana.

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

id_calib	No recortados	Recortados	Porcentaje
04_1_2_2	452	177	28,14%
04_1_1_1	149	50	25,13%
03_1_2_1	562	183	24,56%
03_1_2_2	719	202	21,93%
12_0_0_0	237	65	21,52%
11_2_1_2	140	29	17,16%
11_2_1_1	161	29	15,26%
11_2_2_2	356	62	14,83%
11_2_2_1	371	49	11,67%
10_2_1_2	505	33	6,13%
10_2_1_1	602	33	5,20%
10_2_2_1	1967	71	3,48%
10_2_2_2	1849	66	3,45%
04_1_2_1	586	19	3,14%
04_1_1_2	182	5	2,67%
09_2_1_2	731	18	2,40%
09_2_1_1	738	18	2,38%
06_1_2_1	1625	38	2,29%
09_2_2_1	2418	52	2,11%
08_1_1_2	143	3	2,05%
08_1_2_1	384	8	2,04%
06_1_2_2	1811	32	1,74%
09_2_2_2	2647	46	1,71%
01_1_1_1	271	4	1,45%
06_1_1_1	448	6	1,32%
03_1_1_1	229	3	1,29%
01_1_2_2	1160	14	1,19%
08_1_2_2	439	5	1,13%
07_1_1_1	757	8	1,05%
01_1_1_2	219	2	0,90%
06_1_1_2	450	4	0,88%
07_1_1_2	697	5	0,71%
07_1_2_2	2373	17	0,71%
01_1_2_1	1002	7	0,69%
08_1_1_1	166	1	0,60%
02_1_1_2	372	2	0,53%
07_1_2_1	2272	12	0,53%
02_1_1_1	447	2	0,45%
02_1_2_1	1191	0	0,00%
02_1_2_2	1313	0	0,00%
03_1_1_2	224	0	0,00%
05_1_1_1	268	0	0,00%
05_1_1_2	239	0	0,00%
05_1_2_1	998	0	0,00%
05_1_2_2	1156	0	0,00%

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

id_calib	No recortados	Recortados	Porcentaje
12_0_0_0	242	60	19,87%
04_1_1_1	163	36	18,09%
03_1_2_1	613	132	17,72%
04_1_2_2	523	106	16,85%
03_1_2_2	782	139	15,09%
11_2_2_2	362	56	13,40%
11_2_1_1	165	25	13,16%
11_2_1_2	150	19	11,24%
11_2_2_1	384	36	8,57%
10_2_1_2	517	21	3,90%
10_2_1_1	612	23	3,62%
04_1_2_1	591	14	2,31%
04_1_1_2	183	4	2,14%
10_2_2_1	1995	43	2,11%
08_1_1_2	143	3	2,05%
10_2_2_2	1881	34	1,78%
09_2_1_2	740	9	1,20%
09_2_2_1	2445	25	1,01%
09_2_2_2	2667	26	0,97%
06_1_2_1	1648	15	0,90%
03_1_1_1	230	2	0,86%
08_1_2_1	389	3	0,77%
09_2_1_1	751	5	0,66%
06_1_2_2	1831	12	0,65%
08_1_1_1	166	1	0,60%
07_1_1_2	698	4	0,57%
07_1_1_1	761	4	0,52%
08_1_2_2	443	1	0,23%
06_1_1_1	453	1	0,22%
07_1_2_1	2280	4	0,18%
01_1_2_2	1172	2	0,17%
07_1_2_2	2387	3	0,13%
01_1_2_1	1008	1	0,10%
01_1_1_1	275	0	0,00%
01_1_1_2	221	0	0,00%
02_1_1_1	449	0	0,00%
02_1_1_2	374	0	0,00%
02_1_2_1	1191	0	0,00%
02_1_2_2	1313	0	0,00%
03_1_1_2	224	0	0,00%
05_1_1_1	268	0	0,00%
05_1_1_2	239	0	0,00%
05_1_2_1	998	0	0,00%
05_1_2_2	1156	0	0,00%
06_1_1_2	454	0	0,00%

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

<b>id_calib</b>	<b>No recortados</b>	<b>Recortados</b>	<b>Porcentaje</b>
12_0_0_0	245	57	18,87%
03_1_2_2	809	112	12,16%
11_2_1_1	167	23	12,11%
04_1_1_1	175	24	12,06%
03_1_2_1	663	82	11,01%
04_1_2_2	565	64	10,17%
11_2_1_2	152	17	10,06%
11_2_2_2	378	40	9,57%
11_2_2_1	395	25	5,95%
10_2_1_2	525	13	2,42%
04_1_1_2	183	4	2,14%
04_1_2_1	593	12	1,98%
10_2_2_1	2009	29	1,42%
10_2_1_1	626	9	1,42%
10_2_2_2	1894	21	1,10%
09_2_2_2	2674	19	0,71%
08_1_1_1	166	1	0,60%
09_2_2_1	2456	14	0,57%
09_2_1_2	745	4	0,53%
08_1_2_1	390	2	0,51%
07_1_1_1	762	3	0,39%
07_1_1_2	700	2	0,28%
06_1_2_1	1659	4	0,24%
09_2_1_1	755	1	0,13%
06_1_2_2	1841	2	0,11%
01_1_2_2	1173	1	0,09%
07_1_2_2	2388	2	0,08%
07_1_2_1	2283	1	0,04%
01_1_1_1	275	0	0,00%
01_1_1_2	221	0	0,00%
01_1_2_1	1009	0	0,00%
02_1_1_1	449	0	0,00%
02_1_1_2	374	0	0,00%
02_1_2_1	1191	0	0,00%
02_1_2_2	1313	0	0,00%
03_1_1_1	232	0	0,00%
03_1_1_2	224	0	0,00%
05_1_1_1	268	0	0,00%
05_1_1_2	239	0	0,00%
05_1_2_1	998	0	0,00%
05_1_2_2	1156	0	0,00%
06_1_1_1	454	0	0,00%
06_1_1_2	454	0	0,00%
08_1_1_2	146	0	0,00%
08_1_2_2	444	0	0,00%



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

id_calib	No recortados	Recortados	Porcentaje
12_0_0_0	257	45	14,90%
03_1_2_2	825	96	10,42%
11_2_1_1	171	19	10,00%
11_2_1_2	154	15	8,88%
03_1_2_1	680	65	8,72%
04_1_1_1	183	16	8,04%
11_2_2_2	388	30	7,18%
04_1_2_2	590	39	6,20%
11_2_2_1	402	18	4,29%
04_1_1_2	184	3	1,60%
10_2_1_2	530	8	1,49%
04_1_2_1	598	7	1,16%
10_2_2_1	2016	22	1,08%
10_2_1_1	629	6	0,94%
10_2_2_2	1901	14	0,73%
09_2_2_2	2683	10	0,37%
09_2_2_1	2461	9	0,36%
07_1_1_2	700	2	0,28%
09_2_1_2	748	1	0,13%
01_1_2_2	1173	1	0,09%
07_1_2_2	2388	2	0,08%
01_1_1_1	275	0	0,00%
01_1_1_2	221	0	0,00%
01_1_2_1	1009	0	0,00%
02_1_1_1	449	0	0,00%
02_1_1_2	374	0	0,00%
02_1_2_1	1191	0	0,00%
02_1_2_2	1313	0	0,00%
03_1_1_1	232	0	0,00%
03_1_1_2	224	0	0,00%
05_1_1_1	268	0	0,00%
05_1_1_2	239	0	0,00%
05_1_2_1	998	0	0,00%
05_1_2_2	1156	0	0,00%
06_1_1_1	454	0	0,00%
06_1_1_2	454	0	0,00%
06_1_2_1	1663	0	0,00%
06_1_2_2	1843	0	0,00%
07_1_1_1	765	0	0,00%
07_1_2_1	2284	0	0,00%
08_1_1_1	167	0	0,00%
08_1_1_2	146	0	0,00%
08_1_2_1	392	0	0,00%
08_1_2_2	444	0	0,00%
09_2_1_1	756	0	0,00%

**Anexo 11:** 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>
Nivel de recorte	3	3

Niveles de recorte óptimo por área:

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

Niveles de recorte óptimo por dominio de diseño:

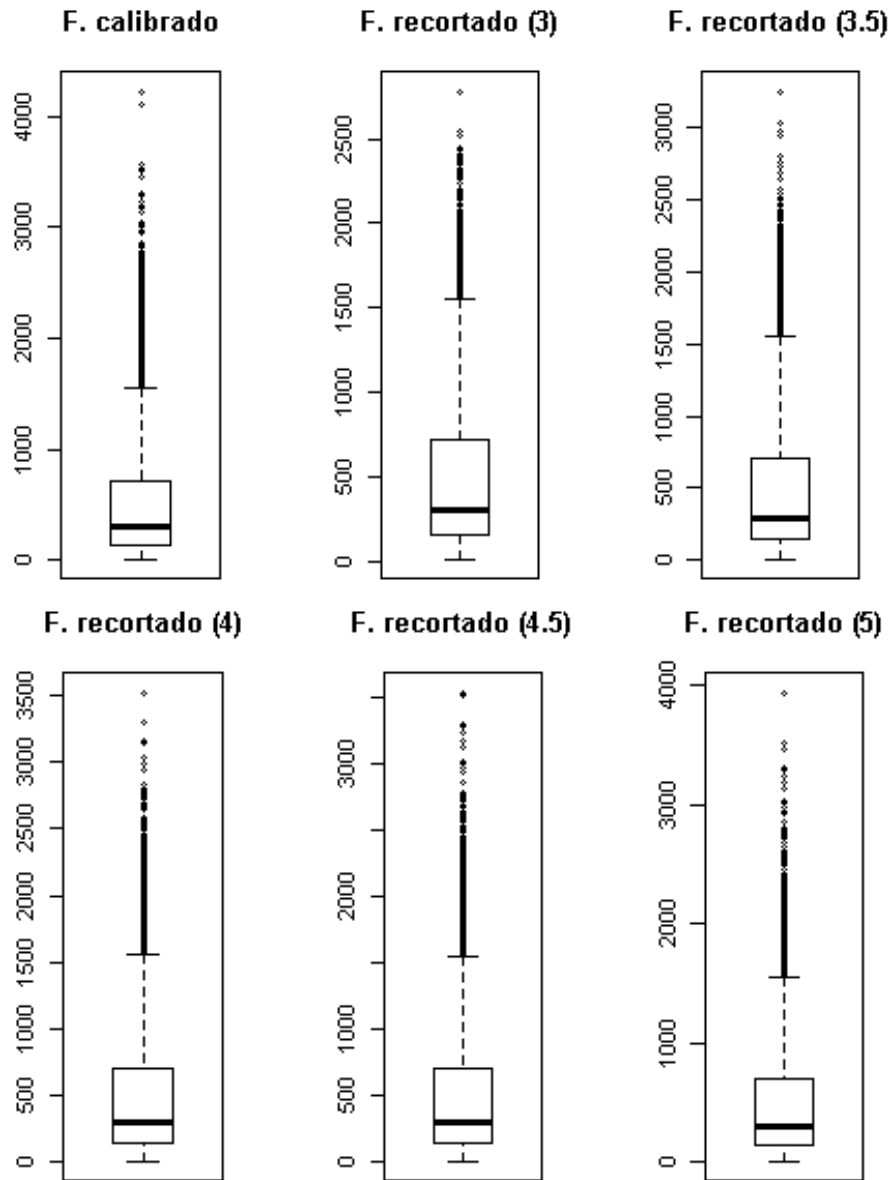
<b>Dominio</b>	<b>PEA</b>	<b>Desempleo</b>
Quito	3	3
Guayaquil	3,5	3
Cuenca	3	3
Machala	3	4
Ambato	3,5	3
Resto Sierra Urbano	3	4
Resto Costa Urbano	5	3
Amazonía Urbano	3	3
Sierra Rural	3	4,5
Costa Rural	5	3
Amazonía Rural	3	5
Región Insular	5	3

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

Post estrato	PEA	Desempleo
01_1_1_1	-	-
01_1_1_2	-	-
01_1_2_1	3	3
01_1_2_2	3	3
02_1_1_1	-	-
02_1_1_2	-	-
02_1_2_1	3	3
02_1_2_2	3	3
03_1_1_1	-	-
03_1_1_2	-	-
03_1_2_1	3,5	3,5
03_1_2_2	3	5
04_1_1_1	-	-
04_1_1_2	-	-
04_1_2_1	3	3
04_1_2_2	3	4
05_1_1_1	-	-
05_1_1_2	-	-
05_1_2_1	3	3
05_1_2_2	3,5	3
06_1_1_1	-	-
06_1_1_2	-	-
06_1_2_1	3	3,5
06_1_2_2	3	3
07_1_1_1	-	-
07_1_1_2	-	-
07_1_2_1	4,5	3
07_1_2_2	3	3
08_1_1_1	-	-
08_1_1_2	-	-
08_1_2_1	3	3
08_1_2_2	3	3
09_2_1_1	-	-
09_2_1_2	-	-
09_2_2_1	3	5
09_2_2_2	3	5
10_2_1_1	-	-
10_2_1_2	-	-
10_2_2_1	3	3
10_2_2_2	5	5
11_2_1_1	-	-
11_2_1_2	-	-
11_2_2_1	4,5	5
11_2_2_2	3	5
12_0_0_0	5	3

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



**Anexo 13:** 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	8,70	8,70	8,70	8,70	8,70
Primer cuartil	153,68	147,36	143,90	141,69	140,43
Mediana	295,88	299,25	302,85	304,58	302,85
Media	469,86	469,86	469,86	469,86	469,86
Tercer cuartil	711,99	710,52	708,24	706,77	706,21
Máximo	2.783,72	3.247,68	3.520,90	3.546,45	3.940,50


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

 @ecuadorencifras

 INEC/Ecuador

 @InecEcuador

 INECEcuador

 t.me/equadorencifras

 INEC Ecuador