



**Encuesta
Nacional de
Salud
2016 -2017**

**CÁLCULO DE FACTORES
DE EXPANSIÓN**



SUBSECRETARÍA DE SALUD PÚBLICA
DIVISIÓN DE PLANIFICACIÓN SANITARIA
Departamento de Epidemiología

Financiamiento:

Ministerio de Salud, Chile.

Coordinación Técnica:

Departamento de Epidemiología, División de Planificación Sanitaria, Ministerio de Salud, Chile.

Ejecutor:

Departamento de Salud Pública Pontificia Universidad Católica de Chile.

Contraparte Técnica:

Unidad de Vigilancia de enfermedades no transmisibles, Departamento de epidemiología, División de Planificación Sanitaria, Ministerio de Salud, Chile.

Equipo Vigilancia de Enfermedades No Transmisibles, Departamento de epidemiología:

José Ignacio Alarcón Molina

Sylvina Alvarado Fick

Pía Álvarez Iturriaga

Carmen Antini Irribarra

Carla Cartes Quintana

Nora Díaz Sanzana

Alejandra Dunner Planella

Rocío Martínez Gutiérrez

Jessica Moya Cortés

Rodolfo Pino Aravena

Clelia Vallebuona Stagno

Natalia Vergara Mallega

Contenido

| | |
|---|----|
| 1. CALCULO DE FACTORES DE EXPANSIÓN | 5 |
| 1.1. Cálculo del ponderador de comuna | 5 |
| 1.1.1. Exclusión de comunas de difícil acceso..... | 5 |
| 1.1.2. Conformación de estratos y unidades de primera etapa..... | 5 |
| 1.1.3. Selección de comunas/pseudocomunas | 6 |
| 1.1.4. Cálculo de la probabilidad de selección de comunas/pseudocomunas | 6 |
| 1.1.5. Cálculo del ponderador de selección de comunas | 6 |
| 1.2. Cálculo del ponderador de manzanas/localidades | 8 |
| 1.2.1. Selección de manzanas/localidades | 8 |
| 1.2.2. Cálculo de la probabilidad de selección de manzanas/localidades..... | 8 |
| 1.2.3. Cálculo del ponderador de selección de manzanas/localidades | 9 |
| 1.2.4. Ajuste por elegibilidad de manzanas/localidades | 10 |
| 1.2.5. Ajuste de subselección de manzanas urbanas | 12 |
| 1.2.6. Ponderador de selección de manzanas/localidades ajustado | 13 |
| 1.3. Cálculo del ponderador de viviendas | 15 |
| 1.3.1. Cálculo de la probabilidad de selección de viviendas | 15 |
| 1.3.2. Cálculo del ponderador de selección de viviendas..... | 17 |
| 1.3.3. Ajuste por omisión de localidades rurales | 18 |
| 1.3.4. Ajuste por elegibilidad de viviendas | 19 |
| 1.3.5. Ajuste por no respuesta de viviendas..... | 21 |
| 1.3.6. Modelamiento y estimación de la probabilidad de responder..... | 21 |
| 1.3.7. Conformación de las celdas de ajuste de no respuesta | 22 |
| 1.3.8. Cálculo del factor de ajuste de no respuesta de vivienda | 23 |
| 1.3.9. Cálculo del ponderador de viviendas ajustado | 23 |
| 1.4. Cálculo del ponderador de personas adultas | 28 |
| 1.4.1. Cálculo de la probabilidad de selección de personas | 28 |

| | | |
|--------|---|----|
| 1.4.2. | Cálculo del ponderador de selección de personas | 30 |
| 1.4.3. | Ajuste por submuestreo | 30 |
| 1.4.4. | Ajuste por participación | 32 |
| 1.4.5. | Cálculo del ponderador de personas calibrado | 36 |
| 1.5. | Cálculo del ponderador de niños | 45 |
| 1.5.1. | Cálculo del ponderador base de niños | 46 |
| 1.5.2. | Ajuste por submuestreo | 47 |
| 1.5.3. | Ajuste por participación | 47 |
| 1.5.4. | Cálculo del ponderador de niños calibrado | 48 |
| 5. | REFERENCIAS BIBLIOGRÁFICAS | 52 |

1. CALCULO DE FACTORES DE EXPANSIÓN

La ENS 2016-2017 involucra el estudio de distintos grupos poblacionales de interés, los cuales fueron seleccionados con distintas probabilidades de selección tanto en la muestra principal ENS (ej. adultos mayores) como en las muestras secundarias. Para dar cuenta de una representación adecuada de los grupos poblacionales, y corregir aquellos que se encuentren sub/sobre representados en las muestras, se calculan distintos ponderadores – también conocidos como factores de expansión – para las distintas muestras y submuestras de la ENS 2016-2017. A continuación, se describe el procedimiento de cálculo del factor de expansión de la muestra principal ENS 2016-2017.

1.1. Cálculo del ponderador de comuna

La ENS 2016-2017 posee un diseño probabilístico, multietápico y estratificado según región y área geográfica. La muestra fue seleccionada a partir de dos marcos muestrales, mutuamente excluyentes, que cubren la totalidad del territorio nacional urbano y rural. El 96,7% de las viviendas de las áreas urbanas de las comunas pertenecen a manzanas en el Marco de Manzanas (MM2008) y un 3,3% pertenece a manzanas en el Marco del Censo 2002 (MM2002). Las viviendas en áreas rurales se encuentran registradas en un 100% en localidades del Marco del Censo (MM2002).

La muestra fue seleccionada en cuatro etapas. Las comunas (o partes de comunas), corresponden a las Unidades Primarias de Muestreo (UPM) de la ENS 2016-2017.

1.1.1. Exclusión de comunas de difícil acceso

Antes de realizar la selección de las comunas/pseudocomunas, se procede a descartar del marco de muestreo aquellas localizadas en las Áreas de Dificil Acceso (ADA's) definidas por el Instituto Nacional de Estadísticas (INE) y el Centro UC de Encuestas y Estudios Longitudinales (CEEL). En total, se descartaron 27 áreas que contienen aproximadamente un 0,3% de la población (ver listado en tabla 1 de informe Diseño Muestral)

1.1.2. Conformación de estratos y unidades de primera etapa

Una característica deseable de una muestra es que incluya “con certeza” a unidades para las cuales se tiene planeado la producción de estadísticas. En ENS 2016-2017 las Unidades Primarias de Muestreo (UPM) corresponden a comunas y pseudocomunas

localizadas a lo largo del territorio nacional. A partir de ellas se conformaron 30 estratos de muestreo definidos por el cruce de región y área geográfica.

1.1.3. Selección de comunas/pseudocomunas

Las 99 comunas/pseudocomunas de inclusión forzosa (IF) fueron incluidas en la muestra con probabilidad 1. Las otras comunas/pseudocomunas sin inclusión forzosa (No-IF) fueron organizadas en 30 estratos definidos por región y área y luego fueron seleccionadas en la muestra utilizando un método de selección sistemático con Probabilidad Proporcional al Tamaño (PPT), medido este en términos de la población total (urbana o rural) de 0 y más años, según la Actualización y Proyecciones Cortas de Población a junio de 2015 (INE 2014a, 2014b).

1.1.4. Cálculo de la probabilidad de selección de comunas/pseudocomunas

La probabilidad que una UPM pertenezca a la muestra puede ser calculada como:

$$P_h(i) = \begin{cases} 1 & \text{si UPM es Incl. Forzosa (IF)} \\ \frac{a_h U_{hi}}{U_h} & \text{si UPM No es Incl. Forzosa (No - IF)} \end{cases} \quad (1)$$

Donde,

h : Representa el índice de los estratos de muestreo de primera etapa.

i : Representa el índice de la UPM, correspondiente a una comuna/pseudocomuna de primera etapa.

U_{hi} : Representa el total de personas de 0 y más años que posee la UPM i , del estrato h , según la Actualización y Proyecciones Cortas de Población a junio de 2015 (INE 2014a, 2014b).

U_h : Representa el total de personas de 0 y más años que posee el estrato h , según la Actualización y Proyecciones Cortas de Población a junio de 2015 (INE 2014a, 2014b).

a_h : Número de comunas/pseudocomunas a seleccionar en el estrato h .

1.1.5. Cálculo del ponderador de selección de comunas

El ponderador de selección de comunas viene dado por el inverso de la probabilidad de selección de las comunas:

$$w_{hi} = \frac{1}{P_h(i)} \quad (2)$$

Este ponderador sirve de base para la construcción del resto de los factores de expansión y, en general, no es utilizado para desarrollar estimaciones en forma directa. Sin embargo, es importante calcularlo y chequearlo ya que ciertas técnicas de análisis multinivel requieren disponer del ponderador asociado a las UPM para el correcto cálculo de los errores de muestreo.

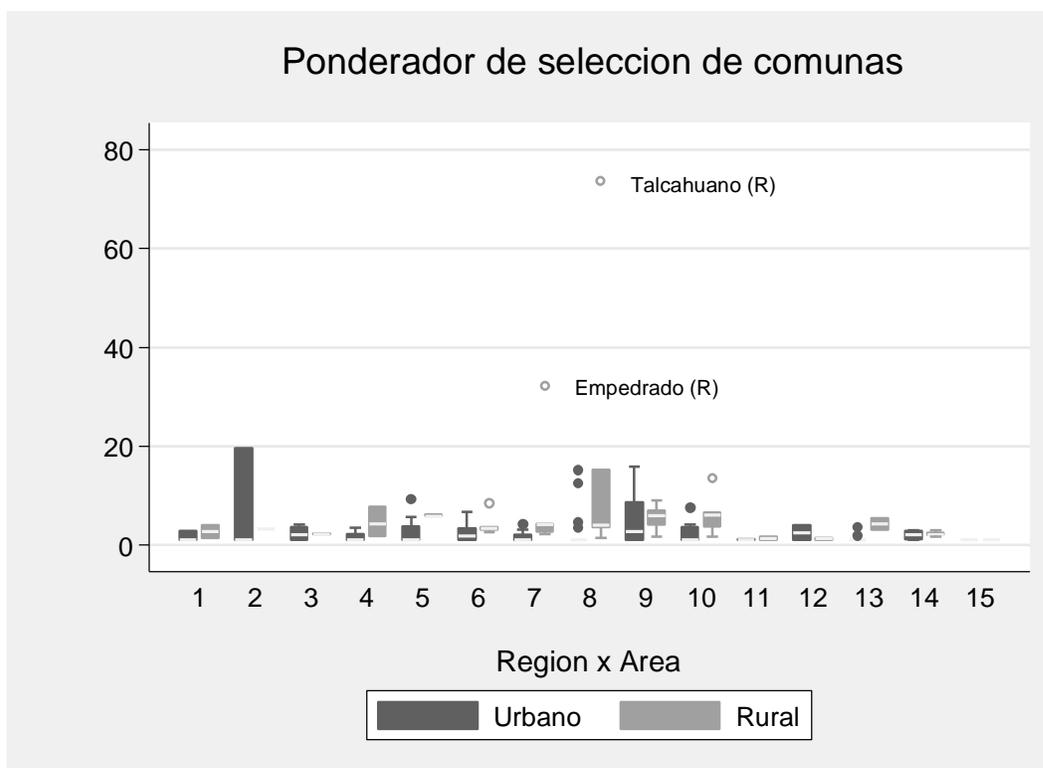
La Tabla 1 muestra los cálculos de la probabilidad de selección de comunas y del ponderador de selección de comunas. Las probabilidades de selección van desde 0.013 (comunas No-IF) a 1 (comunas IF). Las comunas No-IF de Coyhaique y Punta Arenas presentaron probabilidades ligeramente mayores a uno y fueron truncadas a 1. La suma del ponderador de comunas es 586, lo cual está relativamente cerca del total de 638 partes de comunas (comunas según área urbano/rural).

La Figura 1 muestra la distribución del ponderador de selección de comunas por región y área (urbano, rural). En regiones como la 2, 4 y 8 la distribución de los ponderadores es bien distinta entre el área urbana y rural. En la región 7 y 8, las partes rurales de las comunas de Empedrado y Talcahuano presentan los ponderadores más altos.

Tabla 1. Estadísticas descriptivas del ponderador de selección de comunas y la probabilidad de selección de comunas. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|-----------------|------------|--------------|----------------------|------------|------------|
| P_hi | 183 | 0,692 | 0,368 | 0,136 | 1,000 |
| W_hi | 183 | 3,203 | 6,435 | 1,000 | 73,597 |

Figura 1. Distribución del ponderador de selección de comunas, por región y área. ENS 2016-2017.



1.2. Cálculo del ponderador de manzanas/localidades

1.2.1. Selección de manzanas/localidades

Para la selección de las USM (manzanas/localidades) se utilizó un método de selección sistemático con Probabilidad Proporcional al Tamaño (PPT), medido éste como el total de viviendas particulares de la manzana o localidad, según el Marco Muestral Actualizado de Viviendas 2008 o Censo de Población y Viviendas de 2002.

A lo largo de este documento se denomina “manzana urbana” a las manzanas registradas en el MM2008, “manzanas rurales” a las manzanas registradas en el MM2002 y “localidades rurales” a las localidades registradas en el MM2002.

1.2.2. Cálculo de la probabilidad de selección de manzanas/localidades

La probabilidad condicional de seleccionar la USM j , al interior de la UPM i , puede ser calculada como:

$$P_{hi}(j|i) = \begin{cases} \frac{b_{hi} M_{hij}}{M_{hi}} & \text{si USM es manzana (MM2008 o MM2002)} \\ \frac{b_{hi} M_{hij}}{M_{hi}^{50v}} & \text{si USM es localidad (MM2002)} \end{cases} \quad (3)$$

Donde,

j : Representa el índice de la USM, correspondiente a una manzana o localidad.

b_{hi} : Representa el total de USM a seleccionar en la UPM i , del estrato h .

M_{hi} : Representa el total de viviendas que posee la UPM i , del estrato h , según el marco de muestreo.

M_{hi}^{50v} : Representa el total de viviendas que posee la UPM i , del estrato h , según el marco de muestreo. Considera solo las USM con más de 50 viviendas.

M_{hij} : Representa el total de viviendas que posee la USM j , en la UPM i , del estrato h , según el marco de muestreo.

Sin embargo, para poder determinar el número de manzanas/localidades que cada USM representan en la población, es necesario determinar la probabilidad incondicional de selección de las USMs, las que se obtienen a partir del producto de la probabilidad de selección de las UPM's y la probabilidad condicional de selección de las USM's, tal como se señala a continuación:

$$P_{hi}(j) = P_h(i) \cdot P_{hi}(j|i) \quad (4.1)$$

$$P_{hi}(j) = \begin{cases} \frac{a_h U_{hi}}{U_h} \cdot \frac{b_{hi} M_{hij}}{M_{hi}} & \text{si USM es manzana (MM2008 o MM2002)} \\ \frac{a_h U_{hi}}{U_h} \cdot \frac{b_{hi} M_{hij}}{M_{hi}^{50v}} & \text{si USM es localidad (MM2002)} \end{cases} \quad (4.2)$$

1.2.3. Cálculo del ponderador de selección de manzanas/localidades

Una vez determinadas las probabilidades de selección de las comunas y la probabilidad de selección de las manzanas/localidades, se procede a calcular el ponderador de selección de manzanas y localidades. Éste se define como el inverso de la probabilidad de selección de las USM (manzanas y localidades), es decir:

$$w_{hij} = \frac{1}{P_{hi}(j)} \quad (5.1)$$

$$w_{hij} = \begin{cases} \frac{U_h}{a_h U_{hi}} \cdot \frac{M_{hi}}{b_{hi} M_{hij}} & \text{si USM es manzana (MM2008 o MM2002)} \\ \frac{U_h}{a_h U_{hi}} \cdot \frac{M_{hi}^{50v}}{b_{hi} M_{hij}} & \text{si USM es localidad (MM2002)} \end{cases} \quad (5.2)$$

Así, el ponderador w_{hij} se puede interpretar como el número de manzanas/localidades en la población de manzanas/localidades de la comuna i , que representa la manzana/localidad seleccionada en la muestra.

1.2.4. Ajuste por elegibilidad de manzanas/localidades

A nivel de manzanas y localidades, los ponderadores debieran representar a las unidades elegibles de la población para el estudio. Para esta encuesta, las manzanas/localidades “elegibles” corresponden a aquellas en que se puede verificar la existencia de al menos una vivienda particular ocupada al momento de realizar el empadronamiento. Existen también, sin embargo, manzanas/localidades “no elegibles” y otras cuya elegibilidad no se puede determinar durante el periodo de trabajo de campo - denominadas manzanas/localidades de “elegibilidad desconocida”.

Las manzanas/localidades “no elegibles” son aquellas donde se puede verificar en el empadronamiento que no hay viviendas particulares ocupadas. Por ejemplo, una manzana donde la mayoría de las edificaciones correspondían a viviendas particulares ocupadas en el año 2008, pero durante el empadronamiento de 2016 se observa que todas las edificaciones corresponden ahora a uso comercial, serían clasificadas como manzana/localidades “no elegible” en la muestra ENS 2016-2017. Las manzanas/localidades de “elegibilidad desconocida” son aquellas que, por diversos motivos, no es posible determinar su elegibilidad. Por ejemplo, aquellas que no se enviaron a terreno por problemas de acceso, por considerarla un área peligrosa, o simplemente por no haber sido enviada al campo antes del término del trabajo en terreno.

1.2.4.1. Ajuste por elegibilidad desconocida de manzanas urbanas

El ajuste por elegibilidad desconocida de manzanas/localidades se realiza en dos pasos. Primero, se identifican las manzanas/localidades de “elegibilidad desconocida” durante el empadronamiento de ENS 2016-2017, y segundo, se redistribuye el peso de estas manzanas/localidades entre las manzanas/localidades de “elegibilidad conocida”. Debido al pequeño número de unidades en esta categoría, el ajuste se realiza al interior de cada región r . Los pesos de redistribuyen utilizando la siguiente expresión:

$$R_{r,known}^{MZ} = \begin{cases} \frac{\sum_{h \in r} \sum_{i \in \Omega_h} \sum_{j \in \Theta_i} w_{hij}}{\sum_{h \in r} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,known}} w_{hij}} & \text{si USM es eleg. conocida} \\ 0 & \text{si USM es eleg. desconocida} \end{cases} \quad (6)$$

Donde,

r : Indicador de región, que toma valores de 1 a 15, que representan las quince regiones del país.

w_{hij} : Ponderador de selección de manzanas/localidades asociado a la manzana/localidad j , seleccionada al interior de la comuna/pseudocomuna i en el estrato h .

Ω_h : Conjunto de comunas/pseudocomunas seleccionados en el estrato h

Θ_i : Conjunto de manzanas/localidades seleccionadas – de elegibilidad conocida y desconocida - en la comuna/pseudocomuna i .

$\Theta_{i,known}$: Conjunto de manzanas/localidades seleccionadas y clasificadas de elegibilidad conocida pertenecientes a la comuna/pseudocomuna i .

Una interpretación más intuitiva de la expresión (6) viene dada al observar que el numerador corresponde a la estimación del total de USM's estimadas en la región r y el denominador corresponde a la estimación del total de USM's con elegibilidad conocida en la región r .

El ajuste por elegibilidad desconocida se aplica a las manzanas/localidades de elegibilidad conocida. Las manzanas/localidades de elegibilidad desconocida son descartadas del proceso de construcción de los ponderadores a partir de esta etapa. Sin embargo, el peso poblacional asociado a estas manzanas/localidades se conserva ya que se redistribuyó entre las manzanas/localidades de elegibilidad conocida.

1.2.4.2. Ajuste por no elegibilidad de manzanas/localidades

Tras haber realizado el ajuste por elegibilidad desconocida, es posible descartar las observaciones asociadas a las unidades clasificadas como “No Elegibles”. Es importante resaltar que esta eliminación sólo se puede realizar una vez realizado el ajuste por elegibilidad desconocida, ya que estas observaciones se contemplan en el cálculo del

factor de ajuste, el que considera todas las unidades clasificadas con elegibilidad conocida (elegibles y no elegibles).

$$R_{eleg}^{MZ} = \begin{cases} 1 & \text{si USM es elegible} \\ 0 & \text{si USM es no elegible} \end{cases} \quad (7)$$

La Tabla 2 muestra el total de manzanas/localidades en la muestra que fueron clasificadas como elegibles, no elegibles y de elegibilidad desconocida, según marco de muestreo y tipo de USM.

1.2.5. Ajuste de subselección de manzanas urbanas

Las manzanas/localidades muy grandes y las muy pequeñas generan problemas operativos durante el trabajo de campo. Las primeras porque producen una carga excesiva de trabajo al personal encargado del empadronamiento de viviendas, y las segundas porque imposibilitan la selección de los mínimos de viviendas requeridos por manzana/localidad. En la ENS 2016 se desarrollaron procedimientos distintos para lidiar con estos problemas en las áreas urbanas y rurales.

Para reducir problemas operativos en las áreas urbanas se procedió a subdividir las manzanas grandes y subseleccionar solo una parte de ella para el proceso de empadronamiento. Las manzanas pequeñas, por otra parte, se ligaron (anexaron) con otras manzanas cercanas en el plano. La expresión (8) presenta el ajuste que se aplica a manzanas urbanas grandes (101 o más viviendas) para dar cuenta del proceso de subdivisión de manzanas. Las manzanas urbanas pequeñas (1-10 viviendas) y medianas (11-100 viviendas) no se ajustan, lo que se ve reflejado en la fórmula por un factor igual a 1. Las localidades rurales tampoco se ajustan - es decir, se ajustan por un factor igual a 1.

$$R_{hij,subm}^{MZ} = \begin{cases} 1 & \text{si USM es localidad (MM2002)} \\ 1 & \text{si USM es manzana pequeña o mediana (MM2008 o MM2002)} \\ \frac{1}{SM_{hij}} & \text{si USM es manzana grande (MM2008 o MM2002)} \end{cases} \quad (8)$$

Donde SM_{hij} corresponde al número de submanzanas en que se dividió la manzana j seleccionada originalmente. La Tabla 3 presenta el total de manzanas pequeñas, medianas y grandes seleccionadas en la muestra, según marco muestral de procedencia. La Tabla 4 presenta las estadísticas descriptivas del ajuste por submanzanas.

1.2.6. Ponderador de selección de manzanas/localidades ajustado

Finalmente, el ponderador de selección de manzanas/localidades ajustado por elegibilidad desconocida ($R_{r, know}^{MZ}$), no elegibilidad (R_{eleg}^{MZ}) y división de manzanas ($R_{hij, subm}^{MZ}$) viene dado por la expresión (9.2):

$$w'_{hij} = w_{hij} \cdot R_{r, know}^{MZ} \cdot R_{eleg}^{MZ} \quad (9.1)$$

$$w''_{hij} = w'_{hij} \cdot R_{hij, subm}^{MZ} = [P_{hi}^{AJ}(j)]^{-1} \quad (9.2)$$

Donde el término $P_{hi}^{AJ}(j)$ representa la probabilidad de selección ajustada de la manzana j en la comuna/pseudocomuna i en el estrato h .

La Tabla 5 muestra estadísticas descriptivas del ponderador manzanas/localidades ajustado, según marco de muestreo y tipo de USM. La Figura 2 muestra la distribución del ponderador de manzanas/localidades ajustado, según región y área (urbano, rural). La diferencia en la distribución de los ponderadores entre el área urbana y rural se intensifica y se extiende a casi todas las regiones. Las regiones 5, 6, 8 y 13 presentan los ponderadores relativamente más altos.

Tabla 2. Disposición final de casos para manzanas/localidades, según tipo de USM y marco de muestreo. ENS 2016-2017.

| Tipo USM | Elegibles | No elegibles | Elegibilidad desconocida | Total |
|------------------|-----------|--------------|--------------------------|-------|
| Manz 2008 | 965 | 3 | 10 | 978 |
| Manz 2002 | 53 | 2 | 0 | 55 |
| Loc 2002 | 95 | 0 | 0 | 95 |
| Totales | 1.113 | 5 | 10 | 1.128 |

Tabla 3. Total de manzanas según tamaño y marco de muestreo, sobre el total de manzanas inicialmente seleccionadas (n=1.113). ENS 2016-2017.

| Tipo USM | Pequeñas (1-10 viv) | Medianas (11-100 viv) | Grandes (101+ viv) | Total |
|------------------|--------------------------------|----------------------------------|-------------------------------|--------------|
| Manz 2008 | 24 | 799 | 155 | 978 |
| Manz 2002 | 7 | 48 | 4 | 55 |
| Total | 31 | 847 | 159 | 1.033 |

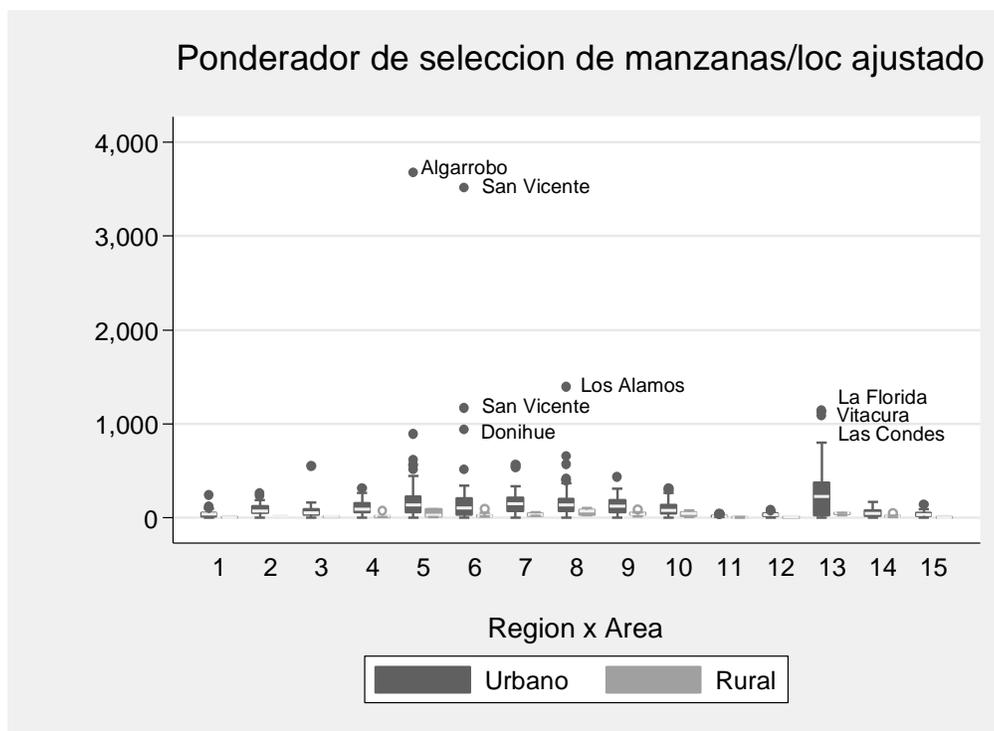
Tabla 4. Ajuste por submanzanas, sobre el total de manzanas elegibles (n=1.018). ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|------------------|------------|--------------|----------------------|------------|------------|
| Manz 2008 | 965 | 0,903 | 0,686 | 0,783 | 1,000 |
| Manz 2002 | 53 | 0,953 | 0,782 | 0,740 | 1,000 |

Tabla 5. Estadísticas descriptivas del ponderador de selección de manzanas/localidades ajustado, según tipo de USM. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|------------------|------------|--------------|----------------------|------------|------------|
| Manz 2008 | 965 | 129.248 | 189,755 | 0,233 | 3.515,783 |
| Manz 2002 | 53 | 183.319 | 510,974 | 3,775 | 3.675,472 |
| Loc 2002 | 95 | 32.603 | 26,253 | 1,000 | 104,963 |
| Total | 1.113 | 123,574 | 210,683 | 0,233 | 3.675,472 |

Figura 2. Distribución del ponderador de manzanas/localidades ajustado (w''_{hij}), por región y área. ENS 2016-2017.



1.3. Cálculo del ponderador de viviendas

1.3.1. Cálculo de la probabilidad de selección de viviendas

Al interior de cada USM (manzana/localidad) se seleccionan viviendas - las unidades de tercera etapa de muestreo (UTM). Esto se realiza posterior al proceso de empadronamiento, donde todas las viviendas particulares ocupadas son identificadas y listadas, para posteriormente seleccionar, utilizando un método de selección aleatorio con igual probabilidad, una cantidad fija de viviendas en cada manzana/localidad - de 6 a 9 viviendas en las zonas urbanas y 10 y más viviendas en las zonas rurales respectivamente.

Así, la probabilidad condicional de incluir la k -ésima vivienda, dado que la USM j -ésima fue seleccionada en la UPM i , puede ser calculada como:

$$P_{hij}(k|j) = \begin{cases} \frac{m_{hij}^{SEL}}{M_{hij}^{EMP}} & \text{si vivienda en manzana (MM2008 o MM2002)} \\ \frac{m_{hij}^{OBS}}{M_{hij}} & \text{si vivienda en localidad (MM2002)} \end{cases} \quad (10)$$

Donde,

K : Representa el índice de la UTM, correspondiente a una vivienda.

m_{hij}^{SEL} : Total de viviendas seleccionadas en la USM j, de la UPM i, del estrato h.

m_{hij}^{OBS} : Total de viviendas observadas en la USM j, de la UPM i, del estrato h.

M_{Hij} : Total de viviendas particulares ocupadas que posee la USM j, en la UPM i, del estrato h, de acuerdo a lo registrado en el marco muestral dado por el Censo 2002 que tiene solo USMs de 50 viviendas y más. Si la cantidad de viviendas observadas en terreno es mayor a la cantidad de viviendas ocupadas registradas en el marco muestral Censo 2002, entonces se debe actualizar el dato y utilizar en M_{hij}^* el número de viviendas observadas que se registre en terreno en reemplazo del número de viviendas ocupadas registradas en el Censo 2002.

M_{Hij}^{EMP} : Total de viviendas particulares ocupadas que posee la USM j, en la UPM i, del estrato h, de acuerdo a lo registrado en el proceso de empadronamiento. Si la USM es grande (101 y más viviendas) y fue submanzanada, entonces aquí corresponde el total de viviendas particulares ocupadas que posee la submanzana empadronada en la USM j, en la UPM i, del estrato h.

Luego, la probabilidad incondicional de que la k-ésima vivienda pertenezca a la muestra de viviendas de la encuesta ENS 2016-2017 viene dada por la expresión:

$$P_{hij}(k) = P_{hi}^{AJ}(j) \cdot P_{hij}(k|j) \quad (11.1)$$

$$\approx \left[\frac{1}{w'_{hij}} \right] \cdot P_{hij}(k|j) \quad (11.2)$$

Donde,

$P_{Hi}^{AJ}(i)$: Es la probabilidad de selección incondicional “ajustada” de seleccionar una manzana/localidad, en la UPM i , en el estrato h . La probabilidad incluye ajustes por elegibilidad desconocida ($R_{r,known}^{MZ}$), no elegibilidad (R_{eleg}^{MZ}) y división de manzanas ($R_{hij,subm}^{MZ}$).

$P_{Hij}(K|j)$: Es la probabilidad condicional de selección de la k -ésima vivienda, dado que la j -ésima manzana/localidad, de la i -ésima comuna/pseudocomuna seleccionada en el estrato h fue seleccionada.

$P_{Hij}(K)$: Es la probabilidad incondicional de selección de la k -ésima vivienda, perteneciente a la j -ésima manzana/localidad, de la i -ésima comuna/pseudocomuna seleccionada en el estrato h .

1.3.2. Cálculo del ponderador de selección de viviendas

Al igual que el ponderador de selección de manzanas/localidades, el ponderador de selección de viviendas se calcula como el inverso de la probabilidad incondicional de las viviendas, es decir,

$$w_{hijk} = \frac{1}{P_{hij}(k)} \quad (12.1)$$

$$= w''_{hij} \cdot \frac{1}{P_{hij}(k|j)} \quad (12.2)$$

El ponderador en (12.2) puede ser utilizado para estimar el total de viviendas que existen en la población al momento de aplicar la encuesta ENS 2016-2017. Esto es de gran utilidad, ya que es importante recordar que los marcos de muestreo utilizados para la selección de la muestra ENS 2016-2017 se encuentra actualizados al año 2008 (la mayor parte del área urbana) y 2002 (algunas partes de resto de área urbana y la totalidad del área rural). La estimación del total de viviendas que se obtiene a partir del ponderador en (12.2) y los totales en el marco muestral van a diferir a medida que transcurre el tiempo, debido a la desactualización del marco.

1.3.3. Ajuste por omisión de localidades rurales

Para reducir problemas operativos se excluyeron, previo a la selección de las USMs, todas aquellas localidades rurales que contenían menos de 50 viviendas en el marco Censo 2002. En este contexto, utilizar el ponderador de la expresión (12.2) para estimar totales asociados a fenómenos de interés implica subestimar dichos totales. Para corregir este problema, se realizó un ajuste que consiste en aumentar el ponderador en (12.2) según la cantidad de viviendas excluidas en cada celda de ajuste, con la finalidad que el total de viviendas estimadas con este nuevo ponderador sea consistente con el total de viviendas de la población rural.

Para realizar el ajuste en las USM rurales se obtuvo la razón entre el total de viviendas según el marco muestral y el total de viviendas estimadas con el ponderador en (12.2) en cada celda de ajuste. La expresión en (13) refleja este ajuste aplicado utilizando como celdas de ajuste las 15 regiones del país. No se excluyeron manzanas previo a la selección por lo tanto no se ajustan, lo que se ve reflejado en la fórmula que sigue por el factor igual a 1.

$$R_{r,ajom}^{MZ} = \begin{cases} 1 & \text{si USM es manzana (MM2008 o MM2002)} \\ \frac{M_{hi}}{\sum_{j \in \Omega_{hi}} w_{hijk} * M_{hij}} & \text{si USM es localidad (MM2002)} \end{cases} \quad (13)$$

Donde,

r : Indicador de región, que toma valores de 1 a 15.

j : Representa el índice de la USM, correspondiente a una manzana o localidad.

Ω_{hi} : Es el conjunto de manzanas/localidades seleccionadas en la UPM i , del estrato h .

M_{hi} : Representa el total de viviendas que posee la UPM i , del estrato h , según el marco de muestreo. En las localidades rurales, aquí se consideran todas las USM, incluyendo aquellas con menos de 50 viviendas.

M_{hij} : Representa el total de viviendas que posee la USM j , en la UPM i , del estrato h , según el marco de muestreo.

1.3.4. Ajuste por elegibilidad de viviendas

A nivel de viviendas, los factores de expansión de la ENS 2016-2017 debieran representar a la población elegible para el estudio. Para la ENS 2016-2017, las viviendas “elegibles” corresponden a aquellas en que se puede verificar la existencia de al menos una persona de 15 o más años de edad. Existen también, sin embargo, viviendas “no elegibles” y otras cuya elegibilidad no se puede determinar durante el periodo de trabajo de campo - denominadas de “elegibilidad desconocida”.

Las viviendas “no elegibles” son aquellas donde se puede verificar que no reside ninguna persona de 15 años o más. Por ejemplo, una vivienda donde solo residen personas menores de 15 años (“menores emancipados”) sería clasificada como “no elegible” en la muestra ENS 2016-2017. Las viviendas de “elegibilidad desconocida” son aquellas que, por diversos motivos, no fue posible determinar su elegibilidad. Por ejemplo, aquellas que no se enviaron a terreno (ej. por problemas de acceso, por considerarla un área peligrosa, o no se envió a terreno) o aquellas donde no se logró entrevistar a los residentes para establecer la elegibilidad de la vivienda (ej. no contactos y rechazos).

1.3.4.1. Ajuste por elegibilidad desconocida de viviendas

Siguiendo la metodología empleada en las encuestas Casen 2011¹, EANNA 2012² y ENCAVI 2015, el ajuste por elegibilidad desconocida de viviendas se realizó en dos pasos. Primero, se identificaron las viviendas de “elegibilidad desconocida”, y segundo, se redistribuye el peso de estas viviendas entre las viviendas de “elegibilidad conocida”. Este ajuste se realiza al interior de cada región r . Los ponderadores se redistribuyeron utilizando la siguiente expresión:

$$R_{r,known}^{VIV} = \begin{cases} \frac{\sum_{h \in r} \sum_{i \in \Omega_h} \sum_{j \in \Theta_i} \sum_{k \in \Phi_j} w_{hijk}}{\sum_{h \in r} \sum_{i \in \Omega_h} \sum_{j \in \Theta_i} \sum_{k \in \Phi_{j,known}} w_{hijk}} & \text{si UTM es eleg. conocida} \\ 0 & \text{si UTM es eleg. desconocida} \end{cases} \quad (14)$$

Donde,

r : Indicador de región, que toma valores de 1 a 15.

w_{hijk} : Ponderador de selección asociado a la vivienda k , de la manzana/localidad j , seleccionada al interior de la comuna/pseudocomuna i en el estrato h .

- Ω_h : Conjunto de comunas/pseudocomunas seleccionados en el estrato h .
- Θ_i : Conjunto de manzanas/localidades seleccionadas en la comuna/pseudocomuna i .
- φ_j : Conjunto de viviendas seleccionadas –de elegibilidad conocida y desconocida - pertenecientes a la manzana/localidad j seleccionada.
- $\varphi_{j,known}$: Conjunto de viviendas seleccionadas clasificadas de elegibilidad conocida pertenecientes a la manzana/localidad j seleccionada.

De forma intuitiva, se puede observar que el numerador corresponde a la estimación del total de viviendas en la región r y el denominador corresponde a la estimación del total de viviendas con elegibilidad conocida (i.e. viviendas elegibles y no elegibles) en la región r .

El ajuste por elegibilidad desconocida se aplica a las viviendas de elegibilidad conocida. Las viviendas de elegibilidad desconocida son descartadas del proceso de construcción de los ponderadores a partir de esta etapa. Sin embargo, el peso poblacional asociado a estas viviendas se conserva ya que se redistribuye entre las viviendas de elegibilidad conocida.

1.3.4.2. Ajuste por no elegibilidad de viviendas

La ENS 2016-2017 busca representar a las viviendas con personas en la población objetivo del estudio, por lo tanto, siguiendo la metodología descrita en Valliant *et al.* 2014³ (aplicada en Casen 2011 y EANNA 2012) en esta etapa se procede a descartar los ponderadores asociados a las unidades clasificadas como “No Elegibles”.

$$R_{eleg}^{VIV} = \begin{cases} 1 & \text{si la vivienda } k \text{ es elegible} \\ 0 & \text{si la vivienda } k \text{ es no elegible} \end{cases} \quad (15)$$

Finalmente, el ponderador de selección de viviendas ajustado por elegibilidad desconocida ($R_{r,known}^{VIV}$), no elegibilidad ($R_{r,eleg}^{VIV}$) y omisión de localidades pequeñas ($R_{r,ajom}^{VIV}$) viene dado por:

$$w'_{hijk} = w_{hijk} \cdot R_{r,ajom}^{VIV} \quad (16.1)$$

$$w''_{hijk} = w'_{hijk} \cdot R_{r,known}^{VIV} \cdot R_{eleg}^{VIV} \quad (16.2)$$

El ponderador (16.2) se asigna solo a las viviendas elegibles, es decir, a partir de esta etapa se descartan las viviendas no elegibles y solo se retienen los ponderadores asociados a las viviendas elegibles. La Tabla 6 presenta estadísticas descriptivas para el ponderador en (16.2).

1.3.5. Ajuste por no respuesta de viviendas

La ENS 2016-2017 utiliza la técnica conocida como *propensity stratification* para identificar grupos homogéneos en la probabilidad de participar en la encuesta y así implementar un ajuste que trate de corregir potenciales sesgos de no respuesta en la muestra lograda.

1.3.6. Modelamiento y estimación de la probabilidad de responder

El primer paso para la implementación del ajuste de no respuesta consiste en estimar, para cada vivienda elegible en la muestra, una probabilidad de responder. En la práctica, esto se consigue a través del uso de modelos. Para la ENS 2016-2017 se utilizarán modelos de regresión lineal multivariada para modelar la probabilidad de responder a nivel de manzanaⁱ.

La variable dependiente en el modelo corresponde a la tasa de respuesta en la manzana j , la cual se calcula como el cociente no ponderado entre el total de las viviendas seleccionadas elegibles que responden y el total de las viviendas seleccionadas elegibles:

$$tr_{hij} = \begin{cases} \frac{r_{hij}^{F1}}{m_{hij}^{SEL}} & \text{si USM es manzana (MM2008 o MM2002)} \\ \frac{r_{hij}^{F1}}{m_{hij}^{OBS}} & \text{si USM es localidad (MM2002)} \end{cases} \quad (17)$$

Donde,

r_{hij}^{F1} : total de viviendas seleccionadas que son elegibles y que responden F1 en la USM j de la UPM i del estrato h .

N_{hij} : total de viviendas seleccionadas que son elegibles en la USM j de la UPM i del estrato h .

ⁱ Los modelos de propensión a responder se desarrollaron a nivel de manzana, y no a nivel de vivienda, debido a la falta de predictores a nivel de vivienda.

m_{hij}^{OBS} : total de viviendas seleccionadas contactadas en la USM j, de la UPM i, del estrato h.

La Figura 8 presenta las variables independientes finalmente retenidas en el modelo de regresión implementado. Las variables de nivel comunal incluidas en el modelo se extrajeron a partir de la Tercera Publicación de Estadísticas Comunales (2014) del Ministerio de Desarrollo Socialⁱⁱ. Ver descripción de las variables incluidas en el Anexo 16.

1.3.7. Conformación de las celdas de ajuste de no respuesta

Las estimaciones de la probabilidad de responder de cada vivienda elegible en la muestra (φ_k), derivadas de las predicciones del modelo de probabilidad lineal a estimar a nivel de manzanas, pueden ser utilizadas para ajustes de no respuesta ya sea en forma individual o agrupándolas en “clases”. Para la ENS 2016-2017 se ha optado por utilizarlas “agrupadas” utilizando el método de clases denominado *propensity stratification*⁴ el cual consiste en crear clases que permitan agrupar unidades que tengan las mismas (o similares) probabilidades de responder la encuesta. Las viviendas en una misma clase tienen la misma configuración de predictores, o al menos una probabilidad de responder muy cercana (la cual resume el efecto de los predictores).

El procedimiento de conformación de las clases se realiza en 2 pasos. Primero, se ordenan las predicciones en forma creciente. Luego, se divide el vector de predicciones en 20 partes iguales (o ventiles) en donde el ventíl #1 agrupa a las manzanas (y sus respectivas viviendas) con la menor predicción para la probabilidad de responder y el ventíl #20 agrupa a las manzanas (y viviendas) con la mayor probabilidad de respuesta predicha. Los $g = 20$ grupos conformados de esta manera vienen a definir las “clases” que se utilizarán para implementar el ajuste de No Respuesta que se describe a continuación.

La Figura 3 muestra la distribución de las probabilidades predichas por el modelo para cada ventíl. Se puede ver en la figura que las distintas clases (los ventiles) permiten agrupar casos homogéneos al interior y heterogéneos entre ellos. Propiedad deseable de variables para ajustes de no respuesta. A modo de referencia, la Figura 4 muestra clases formadas por las 15 regiones del país, donde claramente no se consigue la heterogeneidad entre las clases en relación a la tasa de respuesta observada.

ⁱⁱ Descargar base de datos desde sitio web del Ministerio de Desarrollo Social:
http://observatorio.ministeriodesarrollosocial.gob.cl/indicadores/reportes_com1_2.php

1.3.8. Cálculo del factor de ajuste de no respuesta de vivienda

El ajuste de no respuesta que se aplica en la ENS 2016-2017 busca reducir el riesgo de sesgo por no respuesta. Este ajuste consiste redistribuir el peso de los casos que no respondieron la encuesta entre los casos que sí respondieron, al interior de cada una de las $g = 20$ celdas de ajuste (ventiles). Para estos efectos, se calcula una razón de ajuste al interior de cada ventil utilizando el ponderador de selección de viviendas corregido por elegibilidad. En el numerador se incluye la estimación del total de viviendas seleccionadas elegibles, y en el denominador se incluye la estimación del total de viviendas seleccionadas elegibles que responde, lo que puede ser expresado como:

$$R_{g,resp}^{VIV} = \begin{cases} \frac{\sum_{k \in \varphi_{g,eleg}} w''_{hijk}}{\sum_{k \in \varphi_{g,eleg,resp}} w''_{hijk}} & \text{si vivienda } k \text{ responde} \\ 0 & \text{si vivienda } k \text{ no responde} \end{cases} \quad (18)$$

Donde,

w''_{hijk} : Ponderador de selección ajustado por elegibilidad y omisión de conglomerados, asociado a la vivienda k , de la manzana/localidad j , seleccionada al interior de la pseudocomuna i en el estrato h .

g : Índice de las 20 clases (ventiles) creadas para el ajuste de no respuesta.

$\varphi_{g,eleg}$: Conjunto de viviendas seleccionadas elegibles pertenecientes al ventil g .

$\varphi_{g,eleg,resp}$: Conjunto de viviendas seleccionadas elegibles que respondieron la encuesta, pertenecientes al ventil g .

1.3.9. Cálculo del ponderador de viviendas ajustado

Una vez estimados los factores de ajuste, el Ponderador de viviendas ajustado por No Respuesta puede ser expresado como,

$$w'''_{hijk} = w''_{hijk} \cdot R_{g,resp}^{VIV} = \left[P_{hij}^A(k) \right]^{-1} \quad (19)$$

El ponderador en (19) se asigna solo a las viviendas elegibles que respondieron la encuesta, es decir, a partir de esta etapa se descartan los casos que no respondieron la encuesta y sólo se retienen los ponderadores asociados a los casos que respondieron la encuesta.

Tabla 6. Ponderador de viviendas ajustado por omisión de conglomerados y por elegibilidad, según tipo de USM. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|------------------|------------|--------------|----------------------|------------|------------|
| Manz 2008 | 7.684 | 402,54 | 468,82 | 1,61 | 3.750,13 |
| Manz 2002 | 338 | 527,83 | 1.496,68 | 31,93 | 12.209,49 |
| Loc 2002 | 1.290 | 371,60 | 241,09 | 28,31 | 1.425,83 |

Tabla 7. Estadísticas descriptivas de las tasas de respuesta observadas (trUSM_viv) y predichas (trUSM_pred) en modelo de prob. de responder F1 a nivel de manzana/localidad. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|-------------------|------------|--------------|----------------------|------------|------------|
| trUSM_pred | 1.107 | 0,679 | 0,110 | 0,176 | 0,977 |
| trUSM_viv | 1.107 | 0,680 | 0,247 | 0,000 | 1,000 |

Tabla 8. Modelo de la probabilidad de responder F1 a nivel de manzana/localidad en base a características a nivel de vivienda. ENS 2016-2017.

| | | | Source | SS | df | MS |
|----------------------|---|--------|---------------|-----------|-----------|-----------|
| Number of obs | = | 1,088 | | | | |
| F(21 1066) | = | 11,550 | Model | 11,993 | 21 | 0,571 |
| Prob > F | = | 0,000 | Residual | 52,729 | 1066 | 0,049 |
| R-squared | = | 0,185 | Total | 64,721 | 1087 | 0,060 |
| Adj R-squared | = | 0,169 | | | | |
| Root MSE | = | 0,222 | | | | |

| Variable | Coef. | Std. Err. | t | P> t | IC95% Inf. | IC95% Sup. |
|-----------------|--------------|------------------|----------|-----------------|-----------------------|-----------------------|
| salud16 | -0,005 | 0,002 | -2,240 | 0,025 | -0,009 | -0,001 |
| educacion1 | 0,002 | 0,001 | 3,130 | 0,002 | 0,001 | 0,003 |
| prestaciones2 | -0,034 | 0,017 | -1,980 | 0,048 | -0,069 | 0,000 |
| ingreso1 | 0,000 | 0,000 | -3,370 | 0,001 | 0,000 | 0,000 |
| prestaciones7 | 0,001 | 0,000 | 4,400 | 0,000 | 0,000 | 0,001 |
| prestaciones1 | 0,107 | 0,054 | 1,980 | 0,048 | 0,001 | 0,213 |
| prestaciones17 | 0,000 | 0,000 | 2,580 | 0,010 | 0,000 | 0,000 |
| prestaciones4 | 0,000 | 0,000 | -2,000 | 0,045 | 0,000 | 0,000 |
| entorno2 | 0,000 | 0,000 | -2,380 | 0,017 | 0,000 | 0,000 |
| educacion14 | -0,003 | 0,001 | -4,550 | 0,000 | -0,004 | -0,002 |
| educacion6 | 0,004 | 0,001 | 3,420 | 0,001 | 0,002 | 0,007 |
| prestaciones3 | 0,000 | 0,000 | 2,120 | 0,034 | 0,000 | 0,000 |
| prestaciones8 | 0,000 | 0,000 | -3,880 | 0,000 | 0,000 | 0,000 |
| prestaciones10 | 0,000 | 0,000 | 2,270 | 0,023 | 0,000 | 0,000 |
| prestaciones9 | 0,000 | 0,000 | -2,240 | 0,025 | -0,001 | 0,000 |
| salud9 | 0,012 | 0,005 | 2,680 | 0,008 | 0,003 | 0,022 |
| salud11 | -0,012 | 0,009 | -1,380 | 0,167 | -0,029 | 0,005 |
| vivienda1 | 0,008 | 0,003 | 2,890 | 0,004 | 0,003 | 0,014 |
| educacion4 | 0,003 | 0,001 | 1,960 | 0,050 | 0,000 | 0,006 |
| salud14 | -0,007 | 0,003 | -1,910 | 0,056 | -0,013 | 0,000 |
| prestaciones15 | 0,000 | 0,000 | -2,710 | 0,007 | 0,000 | 0,000 |
| _cons | 0,213 | 0,515 | 0,410 | 0,678 | -0,796 | 1,223 |

Figura 3. Distribución de la probabilidad predicha de responder F1 a nivel de manzana/localidad, según ventíl. ENS 2016-2017.

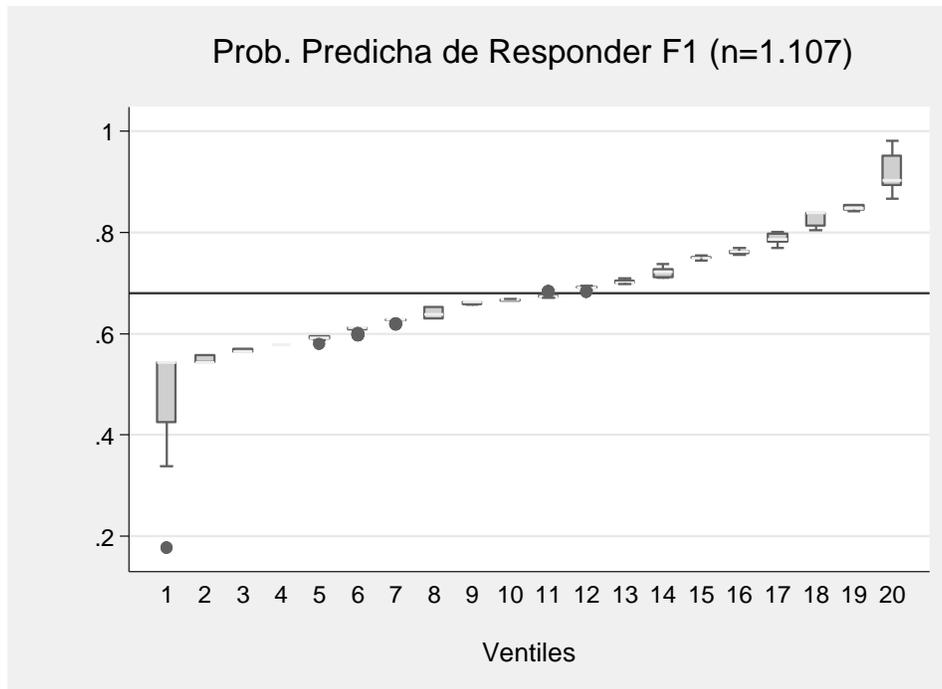


Figura 4. Distribución de las tasas de respuesta observadas de responder F1 a nivel de manzana/localidad, según región. ENS 2016-2017.

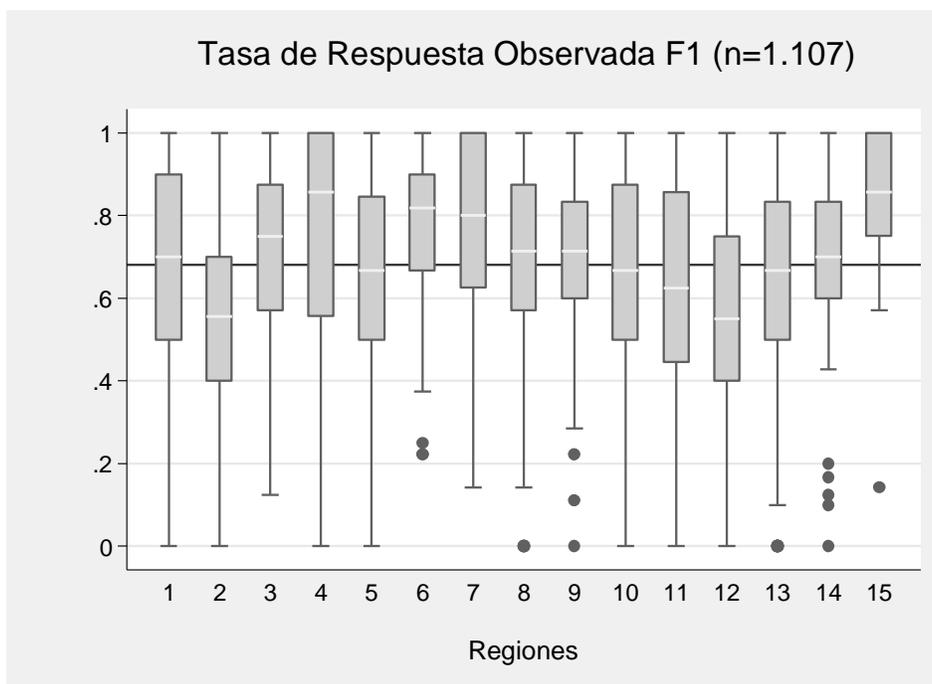
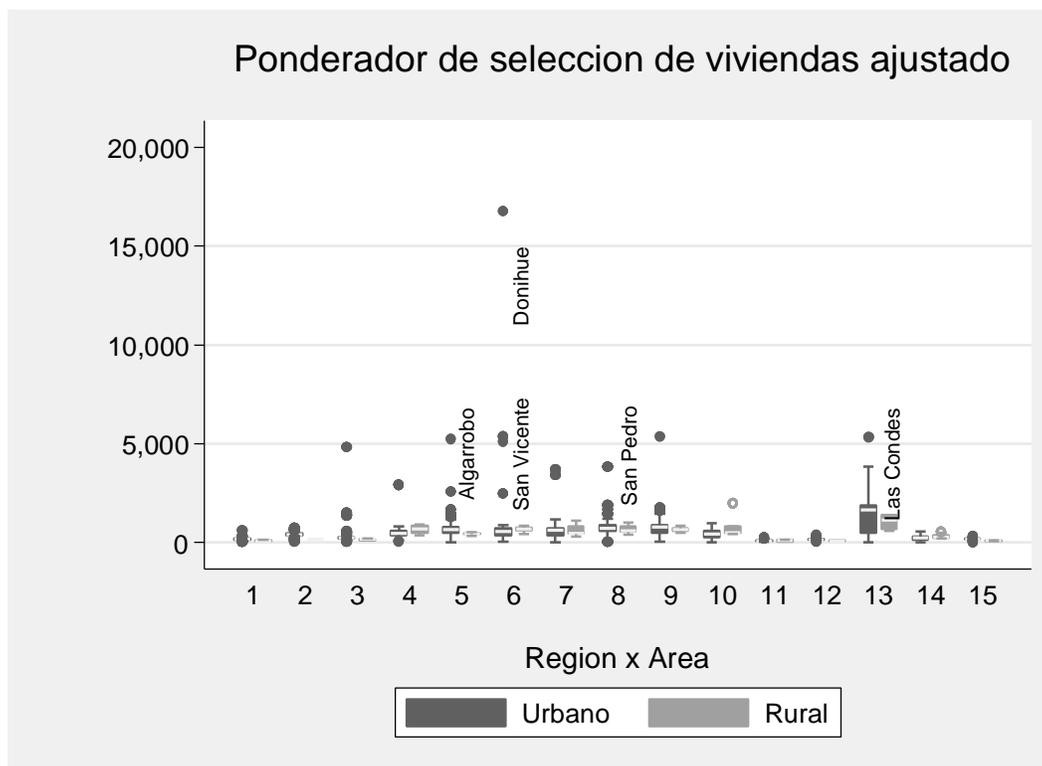


Tabla 9. Estadísticas descriptivas del ponderador de viviendas ajustado por elegibilidad y no respuesta, según tipo de USM. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|------------------|------|--------|---------------|-------|-----------|
| Manz 2008 | 4977 | 617,42 | 692,35 | 1,90 | 5.909,10 |
| Manz 2002 | 265 | 644,08 | 1.801,29 | 43,86 | 16.774,29 |
| Loc 2001 | 991 | 511,96 | 325,49 | 43,83 | 1.985,90 |

Figura 5. Distribución del ponderador de viviendas ajustado por omisión de conglomerados, elegibilidad y no respuesta, según región y área. ENS 2016-2017.



1.4. Cálculo del ponderador de personas adultas

En la ENS 2016-2017 las unidades de muestreo corresponden a las comunas, manzanas, viviendas y personas. A diferencia de las comunas, manzanas y viviendas, las personas corresponden, además, a unidades de interés analítico del estudio. Por esta razón, se aplica en esta etapa un último ajuste denominado “ajuste a población externa” o “calibración”⁵, que permite corregir el ponderador de personas de manera que pueda reflejar las mejores estimaciones disponibles de la población objetivo de estudio – las personas de 15 o más años que residen en viviendas particulares ocupadas a lo largo del territorio nacional.

1.4.1. Cálculo de la probabilidad de selección de personas

En atención a la población objetivo de la ENS 2016-2017, en cada vivienda de la muestra se selecciona para entrevistar a sólo una persona de 15 o más años de edadⁱⁱⁱ. Existen diversos métodos para seleccionar al entrevistado. Para ENS 2016-2017 se utilizó un método de enumeración completa.

El método de enumeración completa es probabilístico, es decir, permite asociar una probabilidad de selección a todas las personas residentes elegibles en la vivienda seleccionada. Para la ENS 2016-2017, se optó por asignar a las personas de 65 años y más dos veces las chances de selección que a las personas de 15-64 años de edad. Por lo tanto, la probabilidad condicional que la persona l sea seleccionada en la vivienda k de la manzana/localidad j de la comuna/pseudocomuna i del estrato h , está dada por:

$$P_{hijk}(l|k) = \frac{q_l}{\sum q_l} \quad (20)$$

$$q_l = \begin{cases} 1 & \text{si la persona tiene 15 – 64 años} \\ 2 & \text{si la persona tiene 65 años y más} \end{cases} \quad (21)$$

ⁱⁱⁱ Encuestas como la Encuesta de Caracterización Socioeconómica Nacional (Casen) y la Encuesta Nacional de Empleo (ENE), por otra parte, no seleccionan un adulto al azar, sino que solicitan completar la entrevista con el jefe de hogar o bien con cualquier persona de 18 o más años presente al momento de la visita.

Donde,

$\sum q_l$: Suma ponderada del total de personas elegibles en la vivienda k . El ponderador asociado a las personas de 15-64 años es igual a “1” y el ponderador asociado a las personas de 65 años y más es igual a “2”. Para el cálculo de la probabilidad de selección se utiliza la edad reportada durante el proceso de Empadronamiento^{iv}.

De esta forma, la probabilidad de que la l -ésima persona pertenezca a la muestra viene dada por:

$$P_{hijk}(l) = P_{hij}^{AJ}(k) \cdot P_{hijk}(l|k) \quad (22.1)$$

$$= \left[\frac{1}{w_{hijk}'''} \right] \cdot P_{hijk}(l|k) \quad (22.2)$$

Donde,

$P_{hijk}(l)$: Probabilidad incondicional de que la l -ésima persona de la vivienda k de la manzana/localidad j de la comuna/pseudocomuna i del estrato h pertenezca a la muestra.

$P_{hij}^{AJ}(k)$: Probabilidad incondicional, ajustada por no respuesta, de que la k -ésima vivienda de la manzana/sección j de la comuna/pseudocomuna i del estrato h pertenezca a la muestra.

$P_{hijk}(l|k)$: Probabilidad de selección de la persona l condicionada a la selección de la vivienda k de la manzana/localidad j de la comuna i en el estrato h .

w_{hijk}''' Ponderador de viviendas ajustado de la vivienda k de la manzana/localidad j de la comuna i en el estrato h .

^{iv} La edad registrada durante el Empadronamiento, tanto para la persona seleccionada como el resto de los miembros del hogar, es reportada por cualquier adulto de 18 años y más. En consecuencia, es posible que existan ligeras diferencias entre esta edad y la posteriormente reportada por el mismo entrevistado durante la aplicación del formulario F1. En esta etapa del proceso de cálculo, corresponde utilizar la edad reportada durante el empadronamiento ya que esa es la edad que definió las probabilidades de selección del entrevistado. La edad reportada en el formulario F1 se utiliza en la etapa final de calibración a totales de población y corrige, para los efectos de representación poblacional, los errores de reporte de edad en el proceso de Empadronamiento.

La expresión (22.1) indica que, dado que la probabilidad de selección incondicional de la vivienda ($P_{hij}(k)$) fue ajustada por distintas razones durante la etapa de selección de manzanas/localidades ($R_{hij,subm}^{MZ}$, $R_{r,know}^{MZ}$, R_{eleg}^{MZ}) y viviendas ($R_{r,om}^{VIV}$, $R_{r,know}^{VIV}$, R_{eleg}^{VIV} , $R_{g,resp}^{VIV}$), entonces se tiene que la probabilidad incondicional de que la l -ésima persona pertenezca a la muestra de viviendas de la ENS 2016-2017 viene dada por la expresión (21.2), donde el término entre paréntesis cuadrado incluye a la probabilidad de selección incondicional de la vivienda y los distintos ajustes incorporados.

1.4.2. Cálculo del ponderador de selección de personas

Dado que para la ENS 2106-2017 se elegirá a sólo una persona de la vivienda para que responda la encuesta y ésta se elegirá en forma aleatoria con igual probabilidad para las personas de 15-64 años y con doble probabilidad para los de 65 años y más en la vivienda, entonces tenemos que el ponderador de selección de la persona corresponde al producto entre el ponderador de selección de la vivienda ajustado (w''_{hijk}) y el inverso de la probabilidad condicional de elegir una de las p_k personas elegibles de la vivienda. Esto es:

$$w_{hijkl} = \frac{1}{P_{hijk}(l)} \quad (23)$$

Donde,

$P_{hijk}(l)$: Probabilidad incondicional de selección de la persona l de la vivienda k de la manzana/localidad j de la comuna i en el estrato h .

La Tabla 10 presenta estadísticas descriptivas del ponderador de selección de personas.

1.4.3. Ajuste por submuestreo

La ENS 2016-2017 cuenta con una muestra principal y tres submuestras secundarias denominadas CIDI, SMA y SMA2. Las submuestras secundarias fueron seleccionadas aplicando un muestreo sistemático simple sobre la muestra principal ENS. De esta forma, se tiene que las probabilidades de selección (fracciones de muestreo) de cada una de las submuestras son constantes y vienen dadas por las siguientes expresiones:

$$F^{SMA} = \begin{cases} \frac{6.910}{9.894} = 0,6984 & \text{si } \in \text{ a submuestra SMA} \\ 0 & \text{si } \notin \text{ a submuestra SMA} \end{cases} \quad (24.1)$$

$$F^{SMA2} = \begin{cases} \frac{2.425}{9.894} = 0,2451 & \text{si } \in \text{ a submuestra SMA2} \\ 0 & \text{si } \notin \text{ a submuestra SMA2} \end{cases} \quad (24.1)$$

$$F^{CIDI} = \begin{cases} \frac{7.094}{9.894} = 0,7170 & \text{si } \in \text{ a submuestra CIDI} \\ 0 & \text{si } \notin \text{ a submuestra CIDI} \end{cases} \quad (24.1)$$

Donde,

F^{SMA} : Fracción de submuestreo SMA aplicada sobre la muestra principal ENS seleccionada^v.

F^{SMA2} : Fracción de submuestreo SMA2 aplicada sobre la muestra principal ENS seleccionada^{vi}.

F^{CIDI} : Fracción de submuestreo CIDI aplicada sobre la muestra principal ENS seleccionada^{vii}.

Los ajustes por submuestreo se aplican solo para el cálculo de los factores relacionados a los casos pertenecientes a las submuestras respectivas. El ponderador de personas ajustado por submuestreo viene dado por:

$$w'_{ijkl} = \begin{cases} w_{ijkl} \cdot (F^{CIDI})^{-1} & \text{si } \in \text{ a submuestra CIDI} \\ w_{ijkl} \cdot (F^{SMA})^{-1} & \text{si } \in \text{ a submuestra SMA} \\ w_{ijkl} \cdot (F^{SMA2})^{-1} & \text{si } \in \text{ a submuestra SMA2} \end{cases} \quad (25)$$

Donde,

w_{ijkl} : Ponderador de selección de la persona l de la vivienda k de la manzana/localidad j de la comuna i en el estrato h .

Los ponderadores relacionados a casos que no pertenecen a las submuestras no incluyen este ajuste.

^v La fracción de muestreo efectiva de la submuestra SMA fue de 62% (= 3.846/ 6.233, exámenes SMA realizados sobre el total de entrevistas F1).

^{vi} La fracción de muestreo efectiva de la submuestra SMA2 fue de 22% (= 1.376/ 6.233, exámenes SMA2 realizados sobre el total de entrevistas F1).

^{vii} La fracción de muestreo efectiva de la submuestra CIDI fue de 55% (= 3.430 /6.233, entrevistas realizados CIDI sobre el total de entrevistas F1).

1.4.4. Ajuste por participación

La encuesta ENS 2016-2017 requiere que sus entrevistados accedan a participar en varias instancias de medición. Para empezar, todos los seleccionados son elegibles para participar de la entrevista F1, las mediciones F2 y los Exámenes EX1. La participación en entrevista F3 está condicionada a ser parte de la submuestra CIDI. La obtención de exámenes EX2 y EX3 está condicionada a ser parte de las submuestras SMA y SMA2 respectivamente. Cada una de estas instancias puede tener como resultado la participación o no participación del entrevistado, lo cual genera una multiplicidad de alternativas de combinación para efectos del cálculo de los ponderadores.

Para la ENS 2016-2017 se optó por desarrollar ponderadores para un subconjunto de las posibles combinaciones de mayor interés analítico. De esta forma, se generaron indicadores de participación en 13 combinaciones de interés:

$$I^{F1} = \begin{cases} 1 & \text{si tiene datos en F1} \\ 0 & \text{resto} \end{cases} \quad (26.1)$$

$$I^{F2} = \begin{cases} 1 & \text{si tiene datos en F2} \\ 0 & \text{resto} \end{cases} \quad (26.2)$$

$$I^{F3} = \begin{cases} 1 & \text{si tiene datos en F3} \\ 0 & \text{resto} \end{cases} \quad (26.3)$$

$$I^{EX1} = \begin{cases} 1 & \text{si tiene datos en EX1} \\ 0 & \text{resto} \end{cases} \quad (26.4)$$

$$I^{EX2} = \begin{cases} 1 & \text{si tiene datos en EX2} \\ 0 & \text{resto} \end{cases} \quad (26.5)$$

$$I^{EX3} = \begin{cases} 1 & \text{si tiene datos en EX3} \\ 0 & \text{resto} \end{cases} \quad (26.6)$$

$$I^{F1F2} = \begin{cases} 1 & \text{si tiene datos en F1 y F2} \\ 0 & \text{resto} \end{cases} \quad (26.7)$$

$$I^{F1F2EX1} = \begin{cases} 1 & \text{si tiene datos en F1 y F2 y EX1} \\ 0 & \text{resto} \end{cases} \quad (26.8)$$

$$I^{F1F2EX2} = \begin{cases} 1 & \text{si tiene datos en F1 y F2 y EX2} \\ 0 & \text{resto} \end{cases} \quad (26.9)$$

$$I^{F1F2EX3} = \begin{cases} 1 & \text{si tiene datos en F1 y F2 y EX3} \\ 0 & \text{resto} \end{cases} \quad (26.10)$$

$$I^{F1F3} = \begin{cases} 1 & \text{si tiene datos en F1 y F3} \\ 0 & \text{resto} \end{cases} \quad (26.11)$$

$$I^{F1F2F3} = \begin{cases} 1 & \text{si tiene datos en F1 y F2 y F3} \\ 0 & \text{resto} \end{cases} \quad (26.12)$$

$$I^{F1F2F3EX1} = \begin{cases} 1 & \text{si tiene datos en F1 y F2 y F3 y EX1} \\ 0 & \text{resto} \end{cases} \quad (26.13)$$

Cabe mencionar que todos los casos en la base de datos ENS 2016-2017 participaron al menos en la entrevista inicial F1, por lo tanto, el ajuste por defecto en esta instancia es "1". La Tabla 11 muestra el recuento de casos para cada una de las variables dummy creadas.

El ponderador de personas ajustado por submuestreo y participación viene dado por:

$$w_{hijkl}^{F1} = w_{hijkl} \cdot 1 \cdot I^{F1} \quad (27.1)$$

$$w_{hijkl}^{F2} = w_{hijkl} \cdot 1 \cdot I^{F2} \quad (27.2)$$

$$w_{hijkl}^{F3} = w_{hijkl} \cdot (F^{CIDI})^{-1} \cdot I^{F3} \quad (27.3)$$

$$w_{hijkl}^{EX1} = w_{hijkl} \cdot 1 \cdot I^{EX1} \quad (27.4)$$

$$w_{hijkl}^{EX2} = w_{hijkl} \cdot (F^{SMA})^{-1} \cdot I^{EX2} \quad (27.5)$$

$$w_{hijkl}^{EX3} = w_{hijkl} \cdot (F^{SMA2})^{-1} \cdot I^{EX3} \quad (27.6)$$

$$w_{hijkl}^{F1F2} = w_{hijkl} \cdot 1 \cdot I^{F1F2} \quad (27.7)$$

$$w_{hijkl}^{F1F2EX1} = w_{hijkl} \cdot 1 \cdot I^{F1F2EX1} \quad (27.8)$$

$$w_{hijkl}^{F1F2EX2} = w_{hijkl} \cdot (F^{SMA})^{-1} \cdot I^{F1F2EX2} \quad (27.9)$$

$$w_{hijkl}^{F1F2EX3} = w_{hijkl} \cdot (F^{SMA2})^{-1} \cdot I^{F1F2EX3} \quad (27.10)$$

$$w_{hijkl}^{F1F3} = w_{hijkl} \cdot (F^{CIDI})^{-1} \cdot I^{F1F3} \quad (27.11)$$

$$w_{hijkl}^{F1F2F3} = w_{hijkl} \cdot (F^{CIDI})^{-1} \cdot I^{F1F2F3} \quad (27.12)$$

$$w_{hijkl}^{F1F2F3EX1} = w_{hijkl} \cdot (F^{CIDI})^{-1} \cdot I^{F1F2F3EX1} \quad (27.13)$$

La Tabla 12 presenta las estadísticas descriptivas del ponderador de personas ajustado por submuestreo y participación de las submuestras respectivas.

Tabla 10. Estadísticas descriptivas del Ponderador de selección de persona (w_{hijkl}), según tipo de USM. ENS 2016-2017

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|------------------|------------|--------------|----------------------|------------|------------|
| Manz 2008 | 4.977 | 1.537,342 | 2.260,865 | 1,896 | 32.126,300 |
| Manz 2002 | 265 | 1.779,576 | 7.341,988 | 87,724 | 83.871,470 |
| Loc 2001 | 991 | 1.054,814 | 989,122 | 43,832 | 9.929,522 |

Tabla 11. Número de casos para cada ponderador en la base de datos. ENS 2016-2017.

| Tipo de caso | Muestra | Obs |
|--------------------------|-----------------|------------|
| Tiene F1 | Principal ENS | 6.233 |
| Tiene F2 | Principal ENS | 5.520 |
| Tiene F3 | Submuestra CIDI | 3.430 |
| Tiene Ex1 | Principal ENS | 5.466 |
| Tiene Ex2 | Submuestra SMA | 3.861 |
| Tiene Ex3 | Submuestra SMA2 | 1.380 |
| Tiene F1 y F2 | Principal ENS | 5.520 |
| Tiene F1 y F2 y Ex1 | Principal ENS | 5.428 |
| Tiene F1 y F2 y Ex2 | Submuestra SMA | 3.847 |
| Tiene F1 y F2 y Ex3 | Submuestra SMA2 | 1.376 |
| Tiene F1 y F3 | Submuestra CIDI | 3.430 |
| Tiene F1 y F2 y F3 | Submuestra CIDI | 3.215 |
| Tiene F1 y F2 y F3 y Ex1 | Submuestra CIDI | 3.172 |

Tabla 12. Estadísticas descriptivas de los Ponderadores de selección de persona ajustados por submuestreo y participación. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|-------------|-------|----------|---------------|------|------------|
| w_f1 | 6.233 | 1.470,92 | 2.560,35 | 1,90 | 83.871,47 |
| w_f2 | 5.520 | 1.494,07 | 2.647,80 | 1,90 | 83.871,47 |
| w_f3 | 3.430 | 1.914,80 | 2.873,37 | 2,64 | 44.806,55 |
| w_ex1 | 5.445 | 1.493,63 | 2.649,45 | 1,90 | 83.871,47 |
| w_ex2 | 3.861 | 2.107,12 | 3.467,45 | 2,72 | 96.072,70 |
| w_ex3 | 1.380 | 6.106,50 | 11.192,69 | 7,74 | 273.754,30 |
| w_f1f2 | 5.520 | 1.494,07 | 2.647,80 | 1,90 | 83.871,47 |
| w_f1f2ex1 | 5.428 | 1.494,74 | 2.652,57 | 1,90 | 83.871,47 |
| w_f1f2ex2 | 3.847 | 2.110,03 | 3.471,97 | 2,72 | 96.072,70 |
| w_f1f2ex3 | 1.376 | 6.111,80 | 11.205,94 | 7,74 | 273.754,30 |
| w_f1f3 | 3.430 | 1.914,80 | 2.873,37 | 2,64 | 44.806,55 |
| w_f1f2f3 | 3.215 | 1.938,12 | 2.922,55 | 2,64 | 44.806,55 |
| w_f1f2f3ex1 | 3.172 | 1.934,49 | 2.911,45 | 2,64 | 44.806,55 |

1.4.5. Cálculo del ponderador de personas calibrado

La muestra principal ENS 2016-2017 corresponde a una muestra de personas 15 o más años y, hasta este punto, los ponderadores calculados sirven para llevar a las unidades muestrales a sus respectivos stocks poblacionales. Sin embargo, debido a la obsolescencia y falta de cobertura de los marcos de muestreo utilizados para la selección de la muestra, partes de la población objetivo pueden quedar excluidas (sin cobertura) por el estudio. Por esta razón, es importante corregir las discrepancias entre las estimaciones de la encuesta y las “mejores estimaciones externas disponibles” para la población objetivo de la encuesta.

1.4.5.1. Calibración utilizando el método de raking

El método de calibración que se propone utilizar para ENS 2016-2017, conocido como *raking*, permite ajustar en forma simultánea estas múltiples subpoblaciones de interés con un nivel de precisión adecuado. Esta es una de las ventajas del método de *raking* por

sobre métodos tradicionales como el ajuste de razón (o postestratificación) tradicionalmente utilizado para calibrar factores de expansión en encuestas en Chile^{viii}.

La ENS 2016-2017 utilizará el comando `ipkraking` (Kolenikov, 2014) en Stata 14 para el desarrollo de los ponderadores calibrados utilizando el método de *raking*. A nivel básico, el algoritmo consiste en un ciclo externo que chequea criterios de convergencia y un ciclo interno que itera sobre las variables de control. La notación multi-índice de los ponderadores intermedios, $w_f^{k,v}$, indica el ponderador de la unidad f calculado en el ciclo externo k luego de postestratificar con respecto a la variable v -ésima. De esta forma k va desde 1 a un número predeterminado de iteraciones K ; v va desde “0” (que indica el ponderador de inicio para una determinada iteración) pasando por “1” (que indica ajuste con respecto a la primera variable de control) y llegando a “p” (que indica ajuste con respecto a la última variable de control).

El algoritmo de *raking* básico implementado en `ipfraking` consiste en 7 pasos:

- 1) Iniciar el contador de iteraciones $k \leftarrow 0$ y los ponderadores como $w_f^{0,p}$. (Es decir, use los ponderadores bases para iniciar el raking; el superíndice “0,p” se usa solo por consistencia con la notación de los siguientes pasos).
- 2) Incremente el contador de la iteración $k \leftarrow k + 1$, y actualice los ponderadores $w_f^{k,0} \leftarrow w_f^{k-1,p}$. (Es decir, use el resultado de la iteración del ciclo externo previo para inicializar los ponderadores para el ciclo externo de iteración actual).
- 3) Ciclo interno: vaya desde la variable de control $v = 1, \dots, p$, y actualice los ponderadores

$$w_f^{k,v} = \begin{cases} w_f^{k,v-1} \frac{T(X_v)}{\sum_{l \in S} w_l^{k,v-1} x_{vl}} & , x_{vf} \neq 0 \\ w_f^{k,v-1} & x_{vf} = 0 \end{cases} \quad (28)$$

(Es decir, postestratificar con respecto a la v -ésima variable de control.)

- 4) Si las discrepancias entre los totales ponderados $\sum_{f \in S} w_f^{k,p} x_v$ y los totales de control $T(X_v)$ están dentro de los márgenes de tolerancia especificados para todas las variables $v = 1, \dots, p$, declare la convergencia y vaya al paso 7.

^{viii} Para una discusión acerca de las ventajas y desventajas de distintos métodos de calibración ver Kalton y Flores Cervantes (2003).

- 5) Si el número de iteraciones k alcanza el límite pre-especificado K , declare no convergencia y vaya al paso 7.
- 6) De otra forma, regrese al paso 2. (Es decir, la precisión alcanzada de los totales objetivo es insuficiente, y se requiere más trabajo.)
- 7) Entregue los ponderadores $w_f^{k,p}$ de la etapa final como los ponderadores calibrados.

El algoritmo descrito en los pasos 1 a 7 corresponde a un ajuste de postestratificación que trata a cada variable de control como la variable de postestratificación y comienza el ciclo sobre cada una de estas variables en cada iteración. En términos de optimización multivariada, este algoritmo procede a través de la optimización de cada uno de los márgenes en forma secuencial.

El algoritmo básico de *raking* implementado en *ipfraking* puede dar origen a ponderadores calibrados de alta variabilidad o con algunos valores extremos (*outlayers*). Valores extremos en los ponderadores pueden generar problemas al momento de realizar los análisis, especialmente en subpoblaciones, por esta razón usualmente se truncan los ponderadores extremos.

Ipfraking tienen también disponible un algoritmo de *raking* modificado que permite implementar distintas opciones para truncar los ponderadores de manera simultánea a la implementación del procedimiento de calibración. Para ENS 2016-2017, todos los ponderadores se truncan en el percentil 0,5% más alto. Se utilizó la opción de truncado de valor *absoluto* y se aplicó el procedimiento al término de una ronda completa de iteraciones (opción *sometimes* en *Ipfraking*). Para detalles acerca del algoritmo modificado ver Kolenikov (2014, pág. 29-31).

1.4.5.2. Total de personas utilizado para calibración

Para los análisis a nivel de personas, se consideró que las mejores referencias externas (a la encuesta) del stock de población de personas de 15 años y más corresponden a la Actualización y Proyecciones Cortas de Población, elaborada por el Instituto Nacional de Estadísticas (INE 2014a, 2014b), para el 30 de junio de 2016. Estas proyecciones de población publicadas por el INE en 2014 permiten disponer de la población actualizada para el periodo 2002-2012 y proyectada hasta 2020, por sexo y edad, para el país y regiones, reemplazando así a las proyecciones publicadas por el mismo organismo en

2005 con información disponible hasta el Censo de 2002. Los datos y la descripción de los métodos utilizados en la elaboración de las proyecciones de población están disponibles en la página web del INE⁶.

Para las combinaciones de casos asociados exclusivamente a la muestra principal ENS 2016-2017 (*i.e.* F1, F1-F2, F1-F2-EX1) se utiliza de referencia un margen “C1” de 120 totales de población, extraídos a partir del cruce de tres variables disponibles en las proyecciones de población del INE para el año 2016, en base a las siguientes variables:

- totales de población definidos por 2 categorías de sexo (hombre, mujer).
- totales de población dados por y 4 categorías de edad (= 15-24 años, 25-44 años, 45-64 años, 65 y más años).
- totales de población definidos por las 15 subpoblaciones dadas por la agrupación de 15 regiones del país

Para las combinaciones de casos asociados a las submuestras SMA (*i.e.* F1-F2-EX2) y SMA2 (*i.e.* F1-F2-EX3) se utiliza de referencia un margen “C2” de 14 totales de población, extraídos a partir del cruce de dos variables disponibles en las proyecciones de población del INE para el año 2016, en base a las siguientes variables:

- totales de población definidos por 2 categorías de sexo (hombre, mujer).
- totales de población dados por 7 categorías de edad (= 15-24 años, 25-34 años, 35-44 años, 45-54 años, 55-64 años, 65-74 años, y 75 y más años).

Para las combinaciones de casos asociados a la submuestra CIDI (*i.e.* F1-F3, F1-F2-F3, F1-F2-F3-EX1) se utiliza de referencia un margen “C3” de 14 totales de población, extraídos a partir del cruce de dos variables disponibles en las proyecciones de población del INE para el año 2016, en base a las siguientes variables:

- totales de población definidos por 2 categorías de sexo (hombre, mujer).
- totales de población dados por 7 categorías de edad^{ix} (= 18-24 años, 25-34 años, 35-44 años, 45-54 años, 55-64 años, 65-74 años, y 75 y más años).

Cabe mencionar que el raking en un margen de 120 celdas, como está definido aquí, es equivalente al tradicionalmente conocido “ajuste de post-estratificación”. Se optó por esta

^{ix} La diferencia entre los márgenes C2 y C3 es solo respecto a la categoría de menor tramo etario que en C2 incluye personas de 15 a 24 años y en C3 corresponde a personas de 18 a 24 años.

alternativa para fomentar una mayor comparabilidad con los ajustes realizados en versiones anteriores de la ENS. Los totales de población utilizados para el proceso de calibración se presenta en las Tabla 13 (muestra principal ENS), Tabla 14 (submuestras SMA y SMA2), y Tabla 15 (submuestra CIDI).

1.4.5.3. Cálculo del ponderador de personas calibrado

La expresión (29) ilustra la forma del ponderador final – el ponderador de personas calibrado. El primer término corresponde al ponderador de selección de personas ajustado por submuestreo (si corresponde) y ajustado por participación (si corresponde). El segundo término corresponde al ajuste de calibración que se obtiene del proceso de *raking* implementado para la población adulta de interés.

$$w_{hijkl}^{F1,cal} = w_{hijkl}^{F1} \cdot R_{c1}^{CAL} \quad (29.1)$$

$$w_{hijkl}^{F2,cal} = w_{hijkl}^{F2} \cdot R_{c1}^{CAL} \quad (29.2)$$

$$w_{hijkl}^{F3,cal} = w_{hijkl}^{F3} \cdot R_{c3}^{CAL} \quad (29.3)$$

$$w_{hijkl}^{EX1,cal} = w_{hijkl}^{EX1} \cdot R_{c1}^{CAL} \quad (29.4)$$

$$w_{hijkl}^{EX2,cal} = w_{hijkl}^{EX2} \cdot R_{c2}^{CAL} \quad (29.5)$$

$$w_{hijkl}^{EX3,cal} = w_{hijkl}^{EX3} \cdot R_{c2}^{CAL} \quad (29.6)$$

$$w_{hijkl}^{F1F2,cal} = w_{hijkl}^{F1F2} \cdot R_{c1}^{CAL} \quad (29.7)$$

$$w_{hijkl}^{F1F2EX1,cal} = w_{hijkl}^{F1F2EX1} \cdot R_{c1}^{CAL} \quad (29.8)$$

$$w_{hijkl}^{F1F2EX2,cal} = w_{hijkl}^{F1F2EX2} \cdot R_{c2}^{CAL} \quad (29.9)$$

$$w_{hijkl}^{F1F2EX3,cal} = w_{hijkl}^{F1F2EX3} \cdot R_{c2}^{CAL} \quad (29.10)$$

$$w_{hijkl}^{F1F3,cal} = w_{hijkl}^{F1F3} \cdot R_{c3}^{CAL} \quad (29.11)$$

$$w_{hijkl}^{F1F2F3,cal} = w_{hijkl}^{F1F2F3} \cdot R_{c3}^{CAL} \quad (29.12)$$

$$w_{hijkl}^{F1F2F3EX1,cal} = w_{hijkl}^{F1F2F3EX1} \cdot R_{c3}^{CAL} \quad (29.13)$$

Donde,

R_{c1}^{CAL} : Ajuste de calibración en base a los 120 controles de población C_1 , disponibles para la muestra principal ENS.

R_{c2}^{CAL} : Ajuste de calibración en base a los 14 controles de población C_2 , disponibles para las submuestras SMA y SMA2.

R_{c3}^{CAL} : Ajuste de calibración en base a los 14 controles de población C_3 , disponibles para la submuestra CIDI.

La Tabla 16 presenta las estadísticas descriptivas del ponderador calibrado truncado al 0,5% superior para los 13 factores de expansión desarrollados. Solo con el 0,5% truncado los *outliers* se reducen bastante, pasando para F1 por ejemplo, de un máximo de 67.737 a 24.000. Para terminar, la Figura 6 presenta la distribución por región y área del ponderador calibrado y truncado al 0,5% superior para el F1.

Tabla 13. Totales de población utilizados para calibrar en $C_1=120$ celdas de ajusta conformadas por el cruce completo de región (15), sexo (2) y tramo etario (4). En base a proyección de población de 15 años y más a junio de 2016. ENS 2016-2017.

| Región | Hombre | | | | Mujer | | | | Total |
|--------------|------------------|------------------|------------------|----------------|------------------|------------------|------------------|------------------|-------------------|
| | 15-24 | 25-44 | 45-64 | 65+ | 15-24 | 25-44 | 45-64 | 65+ | |
| 1 | 27.233 | 57.119 | 39.570 | 11.730 | 25.008 | 51.367 | 36.026 | 13.461 | 261.514 |
| 2 | 49.467 | 107.729 | 79.663 | 22.072 | 45.752 | 91.637 | 67.891 | 26.169 | 490.380 |
| 3 | 24.799 | 48.916 | 39.546 | 13.941 | 23.049 | 43.652 | 36.004 | 15.025 | 244.932 |
| 4 | 62.224 | 112.100 | 92.258 | 38.221 | 60.065 | 111.661 | 94.980 | 47.838 | 619.347 |
| 5 | 142.018 | 268.503 | 217.634 | 96.112 | 133.888 | 262.373 | 235.630 | 132.644 | 1.488.802 |
| 6 | 72.041 | 129.087 | 120.039 | 47.438 | 69.172 | 127.221 | 118.390 | 55.350 | 738.738 |
| 7 | 82.726 | 145.019 | 132.593 | 54.171 | 80.226 | 148.939 | 134.841 | 63.497 | 842.012 |
| 8 | 169.223 | 299.879 | 260.111 | 104.850 | 164.199 | 309.041 | 271.703 | 131.763 | 1.710.769 |
| 9 | 81.850 | 138.821 | 117.663 | 49.327 | 79.540 | 144.469 | 119.720 | 62.790 | 794.180 |
| 10 | 66.477 | 123.102 | 105.766 | 38.746 | 63.456 | 119.689 | 101.385 | 48.361 | 666.982 |
| 11 | 8.665 | 16.925 | 14.244 | 4.476 | 8.009 | 15.695 | 12.195 | 4.616 | 84.825 |
| 12 | 11.981 | 26.527 | 22.452 | 7.712 | 11.052 | 23.137 | 20.415 | 9.415 | 132.691 |
| 13 | 545.657 | 1.149.782 | 859.075 | 315.350 | 527.761 | 1.152.088 | 932.957 | 444.056 | 5.926.726 |
| 14 | 33.354 | 58.996 | 50.649 | 20.228 | 31.850 | 57.620 | 49.496 | 25.340 | 327.533 |
| 15 | 19.140 | 38.236 | 27.032 | 11.061 | 18.049 | 35.360 | 27.552 | 13.108 | 189.538 |
| Total | 1.396.855 | 2.720.741 | 2.178.295 | 835.435 | 1.341.076 | 2.693.949 | 2.259.185 | 1.093.433 | 14.518.969 |

Tabla 14. Totales de población utilizados para calibrar las submuestras SMA y SMA2 en $C_2=14$ celdas de ajuste conformadas por el cruce completo de sexo (2) y tramo etario (7). En base a Proyección de población de 15 años y más a junio de 2016. ENS 2016-2017.

| Edad | Hombre | Mujer | Total |
|----------------------|------------------|------------------|-------------------|
| 15-24 años | 1.396.855 | 1.341.076 | 2.737.931 |
| 25-34 años | 1.471.093 | 1.443.241 | 2.914.334 |
| 35-44 años | 1.249.648 | 1.250.708 | 2.500.356 |
| 45-54 años | 1.230.743 | 1.259.019 | 2.489.762 |
| 55-64 años | 947.552 | 1.000.166 | 1.947.718 |
| 65-74 años | 543.411 | 629.670 | 1.173.081 |
| 75 y más años | 292.024 | 463.763 | 755.787 |
| Total | 7.131.326 | 7.387.643 | 14.518.969 |

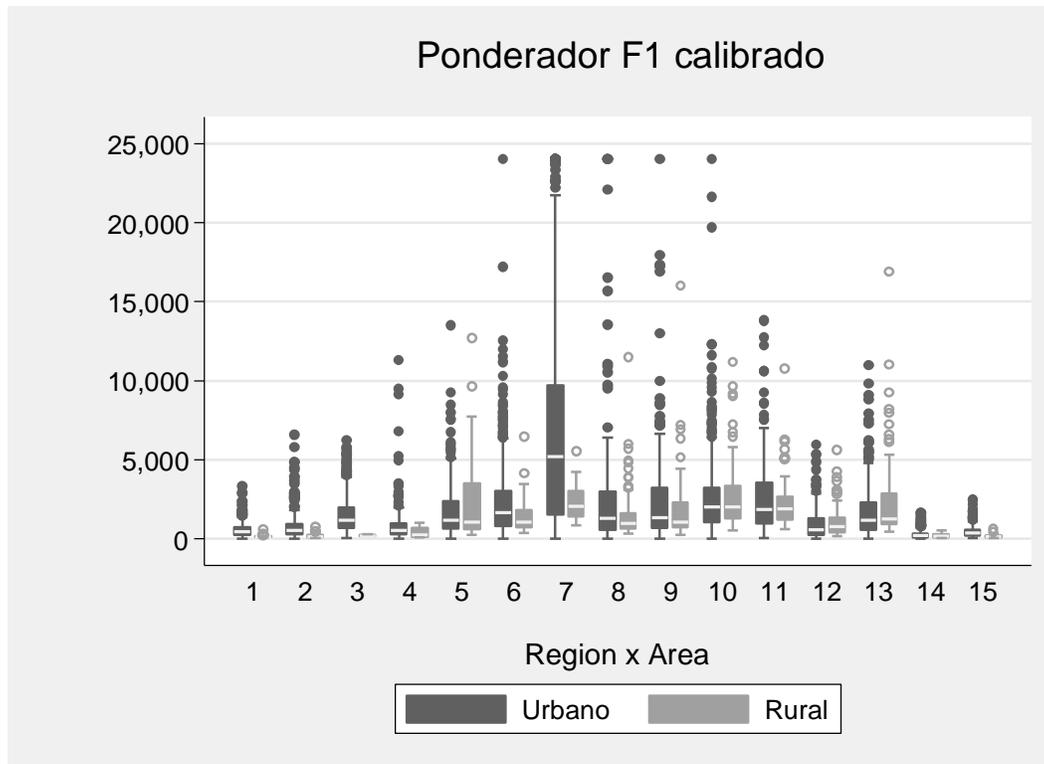
Tabla 15. Totales de población utilizados para calibrar la submuestra CIDI en $C_2=14$ celdas de ajuste conformadas por el cruce completo de sexo (2) y tramo etario (7). En base a Proyección de población de 18 años y más a junio de 2016. ENS 2016-2017.

| Edad | Hombre | Mujer | Total |
|----------------------|------------------|------------------|-------------------|
| 18-24 años | 1.006.254 | 966.636 | 1.972.890 |
| 25-34 años | 1.471.093 | 1.443.241 | 2.914.334 |
| 35-44 años | 1.249.648 | 1.250.708 | 2.500.356 |
| 45-54 años | 1.230.743 | 1.259.019 | 2.489.762 |
| 55-64 años | 947.552 | 1.000.166 | 1.947.718 |
| 65-74 años | 543.411 | 629.670 | 1.173.081 |
| 75 y más años | 292.024 | 463.763 | 755.787 |
| Total | 6.740.725 | 7.013.203 | 13.753.928 |

Tabla 16. Estadísticas descriptivas de los Ponderadores calibrados y truncados al 0,5% superior. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|----------------------|------------|--------------|----------------------|------------|------------|
| Fexp_F1p_Corr | 6.233 | 2.329,37 | 3.526,21 | 2,59 | 24.000 |
| Fexp_F2p_Corr | 5.520 | 2.630,25 | 3.888,86 | 3,13 | 24.000 |
| Fexp_F3p_Corr | 3.430 | 4.009,89 | 5.802,74 | 4,48 | 40.000 |
| Fexp_EX1p_Corr | 5.445 | 2.666,48 | 3.928,00 | 3,16 | 24.000 |
| Fexp_EX2p_Corr | 3.847 | 3.765,63 | 5.411,16 | 4,32 | 34.000 |
| Fexp_EX3p_Corr | 1.376 | 10.532,52 | 14.973,73 | 10,27 | 98.000 |
| Fexp_F1F2p_Corr | 5.520 | 2.630,25 | 3.888,86 | 3,13 | 24.000 |
| Fexp_F1F2EX1p_Corr | 5.428 | 2.674,83 | 3.933,49 | 3,16 | 24.000 |
| Fexp_F1F2EX2p_Corr | 3.847 | 3.774,10 | 5.419,26 | 4,34 | 34.000 |
| Fexp_F1F2EX3p_Corr | 1.376 | 10.551,58 | 14.988,75 | 10,41 | 98.000 |
| Fexp_F1F3p_Corr | 3.430 | 4.009,89 | 5.802,74 | 4,48 | 40.000 |
| Fexp_F1F2F3p_Corr | 3.215 | 4.278,05 | 6.166,79 | 4,76 | 40.000 |
| Fexp_F1F2F3EX1p_Corr | 3.172 | 4.336,04 | 6.201,41 | 4,78 | 40.000 |

Figura 6. Distribución del Ponderador F1 calibrado ($F_{exp_F1p_Corr}$), según región y área. ENS 2016-2017.



1.5. Cálculo del ponderador de niños

Para la implementación de la Encuesta de Desarrollo Infantil (EDI) se identificaron las viviendas con niños entre 7-59 meses de edad entre las viviendas seleccionadas para la muestra ENS 2016-2017 y se aplicó la entrevista EDI en aquellas viviendas con niños donde se haya completado la entrevista F1. Se trató de una aplicación censal a todos los niños en edad elegible del hogar de la persona seleccionada para la ENS 2016-2017. Como consecuencia, la construcción del ponderador de selección de niños en la muestra EDI requiere como insumo los ponderadores que dan cuenta de la probabilidad de ser entrevistado en las viviendas seleccionadas en la muestra ENS 2016-2017. En esta sección se describe la metodología de cálculo de los ponderadores asociados a la EDI.

1.5.1. Cálculo del ponderador base de niños

Las viviendas seleccionadas para la muestra EDI son una submuestra de las viviendas seleccionadas para la muestra ENS 2016-2017 – aquellas viviendas que tienen al menos un niño entre 7 y 59 meses de edad. Para la construcción de ponderadores de la EDI se requiere utilizar como punto de partida el ponderador de la ENS 2016-2017 que permite hacer inferencia al total de la población de viviendas del país. El ponderador de viviendas ajustado por no respuesta de la ENS ($w''_{hijk} = w''_{hijk} \cdot R_{g,resp}^{VIV}$; en la expresión (19) sirve bien este propósito ya que expande a la población de viviendas estimadas al momento del trabajo de campo de la ENS 2016-2017. La suma de los w''_{hijk} corresponde a una estimación del stock de viviendas particulares del país, ajustado por el crecimiento en la población de viviendas entre la fecha del Censo 2002 y la fecha del levantamiento de la ENS 2016-2017.

El diseño de la Encuesta de Desarrollo Infantil establece que todas las viviendas de la muestra principal ENS 2016-2017 con al menos un niño/a entre 7 y 59 meses de edad son elegibles para la muestra EDI. La expresión (31) corresponde a un indicador que toma el valor “1” cuando la vivienda seleccionada en la muestra ENS tiene al menos un niño de edad elegible y “0” si no.

$$I_{hijk}^N = \begin{cases} 1 & \text{si hay algún niño de 7 – 59 meses} \\ 0 & \text{si no hay niños de 7 – 59 meses} \end{cases} \quad (31)$$

Utilizando la expresión (31) y el ponderador de viviendas ajustado por no respuesta (w''_{hijk}), se construye el ponderador base de la encuesta EDI:

$$w_{hijk}^N = w''_{hijk} \cdot I_{hijk}^N \quad (32)$$

La muestra lograda EDI tiene un total de 984 casos, de los cuales 878 corresponden a hogares “únicos” y 106 corresponden a “duplicados” (es decir, hogares donde se recolectó información acerca de más de un niño). La suma de los 878 w_{hijk}^N asociados a los “hogares únicos” produce una estimación del stock de viviendas particulares del país con al menos un niño de 7 a 59 meses de edad. La suma de los 984 w_{hijk}^N asociados a todos los casos de la muestra EDI (hogares únicos+duplicados) produce una estimación del total de niños de 7 a 59 meses de edad. La Tabla 46 presenta estadísticas descriptivas del ponderador base de los niños.

1.5.2. Ajuste por submuestreo

La Encuesta de Desarrollo Infantil se planteó como un censo de todos los hogares con al menos un niño de 7 a 59 meses de edad que pertenecieran a la muestra ENS 2016-2017. Para reflejar el hecho de que se realizó un censo, y no un submuestreo de casos, se deja explícita una fracción de muestreo unitaria para la muestra EDI:

$$F^{EDI} = 1 \quad (33)$$

1.5.3. Ajuste por participación

La participación en la entrevista F4 está condicionada a la participación del adulto seleccionado en la entrevista F1 y al cumplimiento de los criterios de inclusión/exclusión de los niños en la muestra¹⁰. Para la EDI se tiene el siguiente indicador de participación en la entrevista F4:

$$I^{F4} = \begin{cases} 1 & \text{si cumple criterios F4} \\ 0 & \text{resto} \end{cases} \quad (34)$$

El ponderador de niños ajustado por submuestreo y participación viene dado por:

$$w_{hijkl}^{F4} = w_{hijkl}^N \cdot (F^{EDI})^{-1} \cdot I^{F4} \quad (35)$$

La Tabla 12 presenta las estadísticas descriptivas del ponderador de niños ajustado por submuestreo y participación.

Tabla 17. Estadísticas descriptivas del Ponderador base de niños (w_{hijkl}^N), y del Ponderadores de niños ajustados por submuestreo y participación. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|----------|-----|--------|---------------|-------|---------|
| w3_hijk | 984 | 623,52 | 676,83 | 1,896 | 5354,38 |
| w_f4 | 984 | 623,52 | 676,83 | 1,896 | 5354,38 |

¹⁰ Si por ejemplo en una vivienda se completó F4, pero luego se verifica que el niño seleccionado no tenía la edad correspondiente al momento de la entrevista, entonces se debe asignar un "0" al indicador de participación "I_F4". Lo mismo se debe hacer si, en forma posterior al levantamiento de F4, se verifica que no se cuenta con datos de la entrevista F1 del adulto seleccionado.

1.5.4. Cálculo del ponderador de niños calibrado

La suma del ponderador en (35), para los 984 niños entrevistados, corresponde a una estimación del total de niños entre 7-59 meses de edad en la población chilena. Sin embargo, dado que se producen fallas en la cobertura de la encuesta de la población objetivo, y fallas en conseguir la cooperación de los hogares seleccionados, es importante corregir las discrepancias entre las estimaciones desarrolladas con el ponderador (35) y las proyecciones externas disponibles para la población bajo estudio. El ajuste de calibración permite corregir las discrepancias entre las estimaciones y las proyecciones de población externas. A continuación se describe el procedimiento de cálculo del ajuste poblacional que da origen al ponderador de calibración, los totales poblacionales utilizados de referencia, y la fórmula de cálculo del ponderador de calibración de niños.

1.5.4.1. Calibración utilizando el método de raking

El método de calibración que se propone utilizar para EDI, conocido como *raking*, permite ajustar en forma simultánea estas múltiples subpoblaciones de interés con un nivel de precisión adecuado. Esta es una de las ventajas del método de *raking* por sobre métodos tradicionales como el ajuste de razón (o postestratificación) tradicionalmente utilizado para calibrar factores de expansión en encuestas en Chile¹¹. Ver la descripción de la metodología del método de raking en el capítulo 4.4.5.1.

1.5.4.2. Total de personas utilizado para calibración

La población a la cual pretende hacer inferencia la EDI corresponde a todas las personas comprendidas en el grupo de edades de 7 a 59 meses, cumplidos al momento de aplicación de la encuesta, que residen en viviendas particulares a lo largo del territorio nacional. En consecuencia, se requiere contar con proyecciones de población para este grupo poblacional. Para los análisis a nivel de personas, se consideró que las mejores referencias externas del stock de población de personas de 7 a 59 meses corresponden a la Actualización y Proyecciones Cortas de Población, elaborada por el Instituto Nacional de Estadísticas (INE 2014a, 2014b), para el 30 de Junio de 2015. Estas proyecciones de población publicadas por el INE en 2014 permiten disponer de la población actualizada para el periodo 2002-2012 y proyectada hasta 2020, por sexo y edad, para el país y

¹¹ Para una discusión acerca de las ventajas y desventajas de distintos métodos de calibración ver Kalton y Flores Cervantes (2003).

regiones, reemplazando así a las proyecciones publicadas por el mismo organismo en 2005 con información disponible hasta el Censo de 2002. Los datos y la descripción de los métodos utilizados en la elaboración de las proyecciones de población están disponibles en la página web del INE⁷.

Para la encuesta EDI se utilizan 2 márgenes de referencia, “C4” y “C5” de 10 y 2 totales de población, extraídos a partir de las proyecciones de población del INE para el año 2016 de las siguientes variables:

- los totales de población definidos por 10 subpoblaciones que vienen dados por el cruce completo de 2 categorías de sexo y 5 categorías de edad (= 7-11 meses, 12-23 meses, 24-35 meses, 36-47 meses, 48-59 meses).
- los totales de población definidos por 2 áreas, urbana y rural, a nivel nacional.

1.5.4.3. Cálculo del ponderador de niños calibrado

La expresión (36) ilustra la forma del ponderador final – el ponderador de niños calibrado. El primer término corresponde al ponderador de selección de niños y el segundo término corresponde al ajuste de calibración que se obtiene del proceso de raking implementado para la población infantil de interés (ver metodología raking en subcapítulo 4.4.5.1).

Para la encuesta EDI se desarrolló un solo set de ponderadores en atención a la combinación de formularios respondidos por los entrevistados. El grupo de niños para quienes se completó el formulario F4 tiene asociado el ponderador $Fexp_{F4p_Corr}$ en la base de datos, el cual corresponde a la expresión:

$$w_{hijkl}^{F4,cal} = w_{hijkl}^{F4} \cdot R_{l,c4c5}^{CAL} \quad (36)$$

Donde,

R_{c4c5}^{CAL} : Ajuste de calibración en base a los 10 controles de población C_4 y los 2 controles C_5 disponibles para la muestra EDI.

Tabla 18. Totales de población utilizados para calibrar a los niños entrevistados en la encuesta de desarrollo infantil en $C_4=10$ celdas de ajuste, conformadas por el cruce completo de sexo (2) y tramo etario (5), y $C_5=2$ celdas de área urbano/rural. En base a Proyección de población de 7 a 59 meses a Junio de 2016.

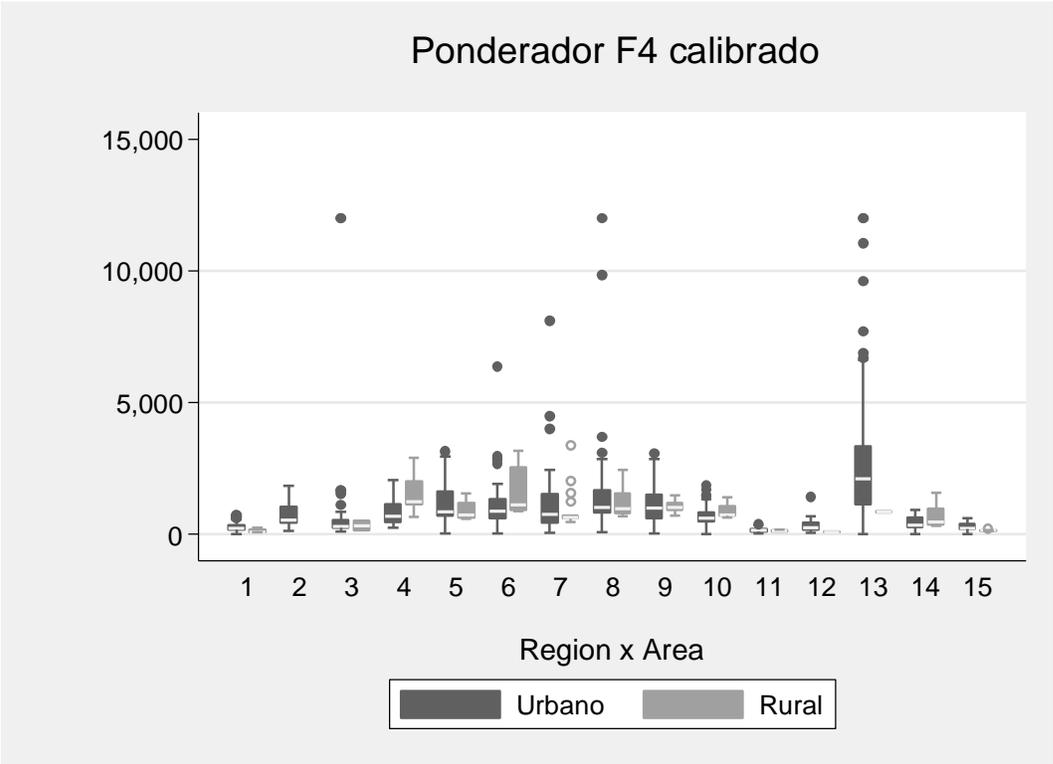
| Tramo etario | Hombre | Mujer | Total |
|---------------------|----------------|----------------|------------------|
| 0 | 52.923 | 51.057 | 103.980 |
| 1 | 126.836 | 122.385 | 249.221 |
| 2 | 125.910 | 121.501 | 247.411 |
| 3 | 124.822 | 120.474 | 245.296 |
| 4 | 125.705 | 120.775 | 246.480 |
| Total | 556.196 | 536.192 | 1.092.388 |

| Área | Total |
|--------------|------------------|
| Urbano | 973.303 |
| Rural | 119.085 |
| Total | 1.092.388 |

Tabla 19. Estadísticas descriptivas de los Ponderadores de Niños calibrados y truncados al 0,5% superior. ENS 2016-2017.

| Variable | Obs | Media | Desv.Estandar | Min | Max |
|-----------------|------------|--------------|----------------------|------------|------------|
| Fexp_F4p_Corr | 984 | 1110.15 | 1451.8 | 4.79 | 12000 |

Figura 7. Distribución del Ponderador F4 calibrado, según región y área. ENS 2016-2017.



5. REFERENCIAS BIBLIOGRÁFICAS

¹ Ministerio de Desarrollo Social (2012a). “Diseño Muestral y Cálculo de Factores de Expansión Encuesta de Caracterización Socioeconómica Nacional (Casen) 2011”. Santiago, Chile.

² Ministerio de Desarrollo Social (2012b). “Diseño Muestral y Cálculo de Factores de Expansión Encuesta de Actividades de Niños, Niñas y Adolescentes (EANNA) 2012”. Santiago, Chile.

³ Valliant, R. Dever, J.A., Kreuter, F. (2013). “Practical Tools for Designing and Weighting Survey Samples”. New York: Springer.

⁴ Rosenbaum, P. R., Rubin, D. B. (1983). „The Central Role of the Propensity Score in Observational Studies for Causal Effects”. *Biometrika*, Vol. 70, No. 1. (Apr., 1983), pp. 41-55.

⁵ Kolenikov, S. (2014). “Calibrating survey data using iterative proportional fitting (raking)”. *The Stata Journal*, Number 1, pp. 22–59.

⁶ Ver en http://www.ine.cl/canales/chile_estadistico/familias/demograficas_vitales.php

⁷ Ver en http://www.ine.cl/canales/chile_estadistico/familias/demograficas_vitales.php