

# Metodología de Diseño Muestral Encuesta de Caracterización Socioeconómica Nacional, Casen 2015

Serie Documentos Metodológicos Casen N°33

**Casen**  
2015

Encuesta de Caracterización  
Socioeconómica Nacional

---

## CONTENIDO

I.	PRESENTACIÓN.....	5
1.	Antecedentes y características generales del diseño muestral de Casen 2015 ....	6
2.	Innovación de diseño muestral Casen 2015 en comparación a los diseños anteriores.....	8
II.	SIMULACIÓN DE TAMAÑO MUESTRAL .....	10
1.	Metodología de cálculo tamaño muestral usando muestras complejas .....	10
III.	DISEÑO MUESTRAL.....	15
1.	Características del marco muestral.....	15
2.	Cobertura del marco muestral .....	19
3.	Estratificación del marco muestral .....	22
4.	Estimación del tamaño muestral .....	23
4.1.	Determinación del tamaño muestral para las comunas. ....	32
4.2.	Asignación del total de viviendas para la sobredimensión .....	34
4.3.	Distribución de la muestra de viviendas en áreas urbanas y rurales .....	34
5.	Selección de conglomerados de la muestra.....	35
5.1.	Selección de conglomerados desde el MM2008 .....	35
5.1.1.	Selección de conglomerados en Casen 2015.....	35
5.1.2.	Mejora metodológica en el proceso de selección de conglomerados en Casen 2015 .....	37
5.2.	Selección de conglomerados desde el MS2002 .....	37
6.	Selección de viviendas de la muestra.....	39
6.1.	Enumeración de los conglomerados de la muestra.....	39
6.2.	Selección de las Viviendas .....	40
IV.	DESARROLLO DE FACTORES DE EXPANSIÓN.....	42
1.	Introducción .....	42
2.	Ponderador de selección de conglomerados .....	46

2.1.	Probabilidad de selección de conglomerados en el Marco de Secciones (MS2002).....	46
2.2.	Probabilidad de selección de conglomerados en el Marco de Manzanas (MM2008) .....	47
2.3.	Ponderador de selección de conglomerados como inverso de la probabilidad de selección .....	49
2.4.	Ajuste por omisión de conglomerados .....	53
3.	Ponderación de selección de viviendas.....	54
3.1.	Probabilidad Condicional de selección de viviendas .....	54
3.2.	Ponderador de selección de viviendas como inverso de la probabilidad de selección .....	55
4.	Ponderación por elegibilidad.....	56
4.1.	Ajuste por elegibilidad desconocida .....	57
4.2.	Ajuste por no elegibilidad.....	58
5.	Ponderación por no respuesta.....	59
6.	Ponderación de calibración .....	62
6.1.	Ponderador de Calibración Regional .....	64
6.2.	Ponderador de Calibración Comunal .....	66
6.3.	Ponderador de Calibración Provincial.....	68
V.	ESTIMACIÓN DE VARIANZAS .....	71
1.	Métodos convencionales para estimación de la varianza.....	71
2.	Varianza con muestras complejas .....	73
3.	Algoritmo de cálculo en muestras complejas .....	75
3.1.	Algoritmo de cálculo de los estimadores .....	76
3.2.	Algoritmo de cálculo de las varianzas en muestras complejas .....	77
4.	Variables que identifican el diseño muestral complejo .....	79
4.1.	Creación de pseudo-estratos (varstrat).....	80
4.2.	Creación de pseudo-conglomerados (varunit) .....	80
5.	Resultados de la situación de pobreza por ingresos.....	86

VI.	ANEXOS .....	93
1.	Simulaciones .....	93
2.	Actualización del Marco Muestral 2008 .....	100
2.1.	Antecedentes .....	100
2.2.	Fusión y División de manzanas .....	108
3.	Proceso de Enumeración y actualización de la muestra .....	109
3.1.	Antecedentes y Objetivos .....	109
3.2.	Conceptos básicos considerados para la enumeración de viviendas .....	110
3.3.	Procesamiento de la información y aseguramiento de calidad .....	111
3.4.	Cuadro de logros de manzanas y secciones .....	111
4.	Dominios de Estudio .....	117
5.	Comandos y Sintaxis en SPSS y STATA.....	124
5.1.	Comandos utilizados en SPSS para estimar la tasa de pobreza por ingresos y errores muestrales.....	124
5.2.	Comandos utilizados en STATA para estimación de tasa de pobreza por ingresos y errores muestrales.....	126
	BIBLIOGRAFÍA .....	127

## I. PRESENTACIÓN

El Diseño Muestral de la encuesta de Caracterización Socioeconómica Nacional, Casen 2015, se enmarca en un convenio de transferencia entre el Ministerio de Desarrollo Social (Ministerio) y el Instituto Nacional de Estadísticas (INE). En éste el Ministerio encomienda al INE la elaboración de la estrategia de muestreo para la Encuesta Casen 2015, lo cual incluye el diseño muestral, la selección, actualización y enumeración de la muestra, el cálculo de los factores de expansión y la definición de variables de diseño complejo, entre otros. El presente documento metodológico se basa en el informe sobre diseño muestral de Casen 2015 elaborado por el INE en el marco del citado convenio.

El principal desafío del diseño muestral de Casen 2015 fue responder al compromiso del programa de gobierno de la Presidenta Michelle Bachelet de aplicar la Encuesta Casen cada dos años, intercalando mediciones con representatividad nacional/regional con otras representativas a nivel comunal. Dada la complejidad de la tarea, el Ministerio le planteó este requerimiento al INE y se sostuvieron reuniones de trabajo para evaluar escenarios posibles y consensuar el diseño final durante el año 2014 y hasta abril de 2015.

Una vez que se dispuso de una propuesta preliminar de diseño muestral, ésta fue presentada por el Ministerio al Panel de Expertos Casen 2015, cuyos integrantes entregaron comentarios al Ministerio que significaron hacer ajustes al diseño.

El diseño finalmente adoptado consiste en una muestra representativa a nivel nacional, regional y para 139 comunas, compuestas por capitales regionales y provinciales, y otras comunas que en conjunto concentran el 80% o más de la población de viviendas de cada región<sup>1</sup>. Consiste en una muestra objetivo de 82.370 viviendas distribuida en 324 comunas, para lo cual INE seleccionó una muestra con sobremuestreo de 87.420, que permite compensar eventuales pérdidas de muestra por no respuesta (rechazo, no contacto u otras situaciones enfrentadas durante la recolección de datos).

En el marco del convenio, el INE calculó los factores de expansión de la encuesta Casen 2015 para cada nivel de representación de la muestra, lo cual permite realizar estimaciones directas y realizar inferencias utilizando la encuesta a nivel nacional, por zona urbano/rural, por regiones y en 139 comunas. Por otra parte, para fines de investigación, el INE calculó factores de expansión para provincias y comunas, que no son dominios de representación, los que permitirán al Ministerio y a otros usuarios de la encuesta Casen, realizar estimaciones utilizando, por ejemplo, técnicas de estimación para áreas pequeñas (SAE, por sigla en inglés) que combinan datos de la encuesta, con otros provenientes de registros administrativos y censales, permitiendo así

---

<sup>1</sup> Según marcos muestrales del INE.

obtener estimaciones más precisas que la estimación directa especialmente en áreas con menor muestra relativa.

El documento se estructura de la siguiente manera. En una primera sección se presenta la encuesta, sintetizando desafíos, características e innovaciones del diseño muestral de Casen 2015. En una segunda sección, se resume el proceso de desarrollo del diseño de la muestra de Casen 2015 a partir de las orientaciones del programa de gobierno y la viabilidad técnica y presupuestaria. En la tercera sección, se detalla la metodología implementada en la selección del tamaño muestral definitivo. Seguidamente, en una cuarta sección, se presenta el desarrollo de los factores de expansión, considerando los niveles de representación de la muestra y otros niveles de interés. En la sección quinta, y final, se presenta la estimación de varianzas, considerando el diseño complejo de la muestra.

## **1. Antecedentes y características generales del diseño muestral de Casen 2015**

A lo largo de los años, la encuesta Casen ha evolucionado en varios aspectos relacionados con la forma en que es planeada, diseñada y ejecutada. Aquí se revisan las principales características del diseño muestral de Casen 2015 y que se profundizan más adelante en otras secciones.

- a. Población objetivo. La población objetivo de la encuesta es la población que reside en viviendas particulares a lo largo del territorio nacional, excluyendo las áreas de difícil acceso definidas por el INE. Al interior de cada vivienda seleccionada, se entrevista a todos los hogares y a las personas que se declaran residentes habituales. El informante idóneo en Casen 2015 es el jefe de hogar o en su ausencia, algún miembro del hogar de 18 años o más.
- b. Marco Muestral. El INE mantiene vigente al año 2015 un marco muestral conformado por dos marcos independientes, el marco de manzanas para la parte urbana y el marco de secciones para la parte rural. Las unidades que lo componen se denominan conglomerados, que corresponden a conjuntos de viviendas agrupadas y contenidas en áreas geográficas definidas por límites de calles, pasajes, alturas de calles o aglomeraciones de viviendas particulares conformadas a partir de una o más entidades pobladas. A través del tiempo, estos conglomerados han sido denominados segmentos (secciones y manzanas) y corresponden a las unidades primarias de muestreo o UPM, obedeciendo a distintos criterios de agrupación, tanto en límites como en tamaño. A través de estos conglomerados se accede a la vivienda y al hogar, cuyos integrantes son las personas, generalmente el objeto final de estudio. A petición del Ministerio de Desarrollo Social, para la encuesta Casen 2015 el INE realizó una actualización parcial menor y focalizado no superior al 2% de viviendas del marco de manzanas, la que se detalla más adelante en el presente documento.
- c. Dominios de estudio. Para la versión 2015 además del nivel regional y por zona urbano/rural, se incorporaron nuevos dominios que corresponden a las comunas capitales regionales y

provinciales, además de aquellas comunas que en conjunto acumulan el 80% o más de las viviendas contenidas en el marco muestral del INE, resultando en la clasificación de 139 comunas como potenciales dominios de estudio<sup>2</sup>.

- d. Objetivos de precisión y tamaño de muestra. En línea con lo recomendado por el Panel de expertos Casen 2015, se definió el tamaño muestral en función de la optimización de la tasa de pobreza por ingresos a nivel regional, fijando como criterio que el error absoluto no superara 4 puntos porcentuales. Además, se buscó la optimización del tamaño muestral para las 139<sup>3</sup> comunas definidas como dominio de estudio. Los criterios de precisión se diferencian según la tasa de pobreza por ingresos estimada para la versión 2013 de la encuesta:
- A nivel Nacional:
    - (c1) A nivel regional, el error absoluto no debe sobrepasar 4 puntos porcentuales y simultáneamente el error relativo no debe superar 35%;
  - A nivel comunal:
    - (c2) Si la tasa de pobreza por ingresos es mayor o igual a 10%, el error relativo debe ser igual o menor a 36%;
    - (c3) Tasa de pobreza por ingresos menor a 10% y mayor o igual a 5%, el error relativo debe ser igual o menor a 50%;
    - (c4) Tasa de pobreza por ingresos menor a 5% y mayor o igual a 2%, el error relativo debe ser igual o menor a 65%;
    - (c5) Tasa de pobreza por ingresos menor a 2% el error relativo debe ser menor o igual a 80%.
    - (c6) Junto a los criterios de precisión establecidos<sup>4</sup>, se determinó que la fracción de muestreo para las comunas no superara 20%.
- e. Estrategia muestral. En diseño es bi-etápico en el área urbana y rural. Junto a lo anterior, en el área urbana de cada comuna, las manzanas se clasificaron según su tamaño (en

---

<sup>2</sup> Los dominios de estudio definidos en Casen 2011 y 2013 fueron cada una de las 15 regiones del país, y en ellas se fijaron los niveles de precisión en cuanto al error absoluto y relativo. En Casen 2015, se incorporaron además todas las capitales regionales y/o provinciales como potenciales dominios de estudio y algunas comunas que no siendo capitales regionales y/o provinciales, acumularan aproximadamente 80% de las viviendas de la región a la que pertenecen. Para estas comunas, se establecieron criterios de precisión y fracciones de muestreo más flexibles como los expuestos desde el punto (c2) a (c6). Sin embargo, para algunas de ellas fue difícil establecer el tamaño de muestra necesario sin violar alguno de estos criterios de precisión. Tal es el caso de las comunas capitales provinciales: Pozo Almonte, Chañaral, Pichilemu, Natales y Porvenir y las comunas Olívar y Quellón (ver en capítulo VI Anexos, Sección 4 Dominios de estudio, Tabla 4.5). Se recomienda, especialmente, evaluar la precisión de las estimaciones a nivel comunal antes de utilizarlas.

<sup>3</sup> Se sugiere verificar posteriormente al levantamiento y cálculo de la tasa de pobreza, si las comunas establecidas como dominios de estudio cumplen los criterios de precisión pre-establecidos, principalmente las 7 comunas nombradas en el pie de página anterior.

<sup>4</sup> Los criterios antes descritos son concordantes (no pre-establecidos) con que a nivel comunal el error absoluto no supere 9,5 puntos porcentuales, aproximadamente.

número de viviendas), creando 5 estratos de tamaño donde la muestra fue distribuida y seleccionada de forma independiente.

- f. Método de selección de las unidades primarias de muestreo (manzanas y secciones). Para Casen 2015, en el marco de manzanas, se realizó una selección sistemática y con igual probabilidad, al interior de cada uno de los 5 grupos de tamaños definidos en cada una de las comunas pertenecientes a la muestra. Ésta estrategia aseguró que cada uno de los grupos de tamaño de las comunas estén representados en la muestra, y con ello tener una fórmula explícita de sus probabilidades de selección. En el marco de secciones, se realizó una selección sistemática con probabilidad proporcional al tamaño en cuanto al número de viviendas (que registra el MS2002), a nivel comuna-área.
- g. Método de selección de las unidades secundarias de muestreo (viviendas). Tanto en el marco de manzanas como en el de secciones, se realizó una selección sistemática y con igual probabilidad, al interior de cada uno de los conglomerados seleccionados previamente (manzanas o secciones).

## **2. Innovación de diseño muestral Casen 2015 en comparación a los diseños anteriores**

Hasta la encuesta Casen 2009, el Ministerio definió las comunas como dominios de estudio, por lo tanto la muestra consideraba un número mínimo de viviendas seleccionadas en cada comuna, ya que se tenía entre los objetivos de la encuesta la caracterización socioeconómica en comunas. Sin embargo, para algunas comunas en Casen 2009 el margen de error para la estimación de la tasa de pobreza es tan grande<sup>5</sup>, que los resultados se tornan poco informativos (nivel de precisión bajo) para la toma de decisiones.

En este contexto y siguiendo las recomendaciones de expertos<sup>6</sup>, para la encuesta Casen 2011, el Ministerio optó por definir los tamaños muestrales acotando los dominios de estudio, y al mismo tiempo, utilizar la tasa de pobreza, los errores de muestreo, el efecto de diseño, la precisión deseada y la tasa de respuesta observada de sus versiones anteriores. Se definieron entonces, como nuevos dominios de estudio a las regiones, en atención a la creciente demanda de información con que se requiere a este nivel territorial, lo que implicó re-definir el diseño de la muestra en base a las regiones como dominios de estudio, lo que requirió sólo de la presencia de unidades muestrales en las comunas.

---

<sup>5</sup> Para tasas de pobreza cercanas a 5% su margen de error puede ser de 4% o más, y en aquellos casos donde la tasa es superior a 10%, el margen de error puede superar incluso 30%. Existen casos más críticos donde el total de unidades que responde por conglomerado y/o estrato no es suficiente para poder calcular su error estándar.

<sup>6</sup> Ver informe CEPAL (2008, pág. 6) e informe Comisión Técnicos Casen (2010, págs. 6 y 7)

La encuesta Casen 2013, mantiene la representatividad a nivel regional de la encuesta Casen 2011. No obstante, la encuesta Casen 2011 tenía grandes muestras de viviendas en las regiones extremas<sup>7</sup>, en la versión 2013, los tamaños muestrales de estas regiones disminuyen respetando la proporcionalidad o importancia relativa de la región respecto al total nacional.

En el diseño muestral de la encuesta Casen 2015 se realizaron principalmente tres innovaciones respecto a las anteriores:

- Representatividad: Con el objetivo de producir información de mejor precisión para capitales regionales y provinciales, y otras comunas que en conjunto concentran el 80% o más de la población regional de viviendas (conforme información del marco muestral disponible).
- Actualización realizada: Considerando síntomas de desactualización, identificados en experiencias de diferentes levantamientos previos de INE, el Marco de manzanas fue actualizado en 69 urbanos<sup>8</sup> de un total de 241 que contiene. Esta actualización incluyó revisión en gabinete y visitas a terreno.
- Método de selección: En el área urbana y rural el diseño muestral es bi-etápico, y el método de selección de las manzanas es sistemático y con igual probabilidad.

---

<sup>7</sup> Si bien para Casen 2011 se privilegió la precisión en cuanto a la disminución del error de muestreo en las regiones extremas, indujo a que la muestra en términos de fracción muestral quedara sobre-representada respecto a otras regiones, produciendo externalidades negativas en el proceso operativo de levantamiento y la saturación de las unidades primarias de muestreo, sobre todo en las partes rurales y resto de áreas urbanas en que hay pocas unidades contenidas en el Marco del INE, muchas de ellas ocupadas en la encuesta de empleo del INE y que se descuentan previamente para la selección de la muestra Casen.

<sup>8</sup> En el Marco Muestral del INE, según la división política administrativa, las comunas están formadas por entidades urbanas denominadas "urbanos" y entidades rurales (localidades y entidades). Como por ejemplo, la parte urbana de la comuna de los Vilos está formada por los urbanos de Los Vilos, Pichidangui y Quilimarí. De estos urbanos, sólo las UPM de Los Vilos están incorporadas en el Marco de Manzanas o Marco 2008, mientras que las UPM de los otros dos urbanos, están incorporadas en el Marco de Secciones o Marco 2002.

## II. SIMULACIÓN DE TAMAÑO MUESTRAL

Previo al desarrollo de la estrategia de levantamiento de la encuesta Casen 2015, se determinó un tamaño muestral mínimo óptimo para la estimación, con márgenes de precisión razonables para que la encuesta Casen 2015 tuviera representatividad a nivel nacional, nacional urbano-rural, regional y en comunas. Para ello, se simularon distintos escenarios o alternativas de tamaño muestral en base a la unificación de criterios estadísticos solicitados y discutidos por el Ministerio con INE.

A continuación se presenta brevemente la metodología de cálculo de tamaño muestral para muestras complejas, para comprender los parámetros necesarios a definir en los distintos escenarios evaluados por el Ministerio y el INE. Luego se presenta un resumen del proceso de desarrollo del diseño muestral 2015 hasta alcanzar el diseño definitivo que es presentado en la Sección III. Diseño Muestral.

### 1. Metodología de cálculo tamaño muestral usando muestras complejas

La metodología de cálculo del tamaño muestral la utilizada en Casen 2011, 2013 y 2015 es básicamente la misma. Primero, se inicia con la determinación de un tamaño muestral bajo un muestreo aleatorio simple, fijando previamente el error absoluto a alcanzar en base a los valores de encuestas anteriores, con el propósito de reducir el error absoluto y el error relativo esperados a rangos razonables, e incorporando la cuasi-varianza obtenida en base a un muestreo complejo. Luego, se realizan tres ajustes, el primero incorporando el efecto del diseño (obtenido de una encuesta o el promedio de dos o más encuestas), el segundo incorporando un ajuste por finitud, y el tercero incorporando un ajuste con la tasa de no respuesta, que también puede corresponder a un levantamiento o el promedio de más de un levantamiento.

Posteriormente, opera un análisis de refinamiento en los tamaños muestrales para los niveles deseados, donde se busca resguardar coherencia entre los tamaños muestrales y aquéllos de la población, en términos de proporcionalidad<sup>9</sup>.

En el marco de la estimación de un tamaño muestral para Casen 2015 que incorporara el requerimiento de contar con representatividad a nivel comunal, se evaluaron distintos escenarios en función de las definiciones de los parámetros como: variable de interés, error absoluto y error relativo.

---

<sup>9</sup> La proporcionalidad se refiere a que la muestra guarde el orden de tamaño respecto a la población, es decir, a mayor población, mayor muestra.

Al momento de comenzar las estimaciones de tamaños muestrales para Casen 2015, no se encontraba disponible la base de datos de Casen 2013 con las variables de interés para realizar las estimaciones. De modo que se realizó un primer ejercicio de estimación con los datos de Casen 2011, en particular con la submuestra noviembre 2011 - enero 2012 (en adelante, submuestra 2), que es la muestra históricamente comparable.

Inicialmente se consideró determinar un tamaño muestral representativo para las 324 comunas presentes en el marco muestral del INE, usando la variable de interés de Tasa de Pobreza por Ingresos (porcentaje de personas en situación de pobreza por ingresos, estimado bajo metodología tradicional de medición de pobreza por ingresos). El estimador de razón que mide la Tasa de Pobreza por ingresos se consideró confiable y preciso cuando la estimación del error absoluto fuera menor o igual a 5% o 6% y el error relativo no superara 35%.

Se observó que los tamaños muestrales requeridos utilizando este parámetro excedían lo que, dada la experiencia de levantamiento de encuestas Casen anteriores, se consideraba viable de levantar en un período de 3 meses (período de referencia comparable), por razones operativas y de costos, y además se generaban externalidades negativas respecto al número reducido de viviendas disponibles en el marco muestral y distribución no-proporcional, lo que llevó a evaluar la opción de considerar variables de interés alternativas, tales como el ingreso total per cápita de los hogares.

En particular, se consideraron definiciones alternativas del ingreso total del hogar per cápita, basadas en Casen 2011: 1) corregido por no respuesta y ajustado a cuentas nacionales, 2) corregido por no respuesta pero sin ajuste a cuentas nacionales. De esta forma, se determinaron los tamaños muestrales necesarios para los dominios de representación, que permitieran estimar la media del ingreso per cápita, bajo dos definiciones alternativas. Aun cuando se perseguía viabilizar el levantamiento de la muestra objetivo diseñada para Casen 2015, desde el punto de vista operativo, los tamaños muestrales requeridos aún superaban lo que desde la experiencia se estimaba viable de levantar en un período de 3 meses.

Entonces, se analizaron dos alternativas que consideraron como variable de diseño el ingreso total del hogar per cápita corregido (no ajustado a cuentas nacionales, que es la definición que finalmente se aplicó para la medición actualizada de pobreza por ingresos). La primera, buscó contar con representatividad en todas las comunas del marco muestral (324) y la segunda, priorizar representatividad en comunas que son capitales regionales y provinciales. En ambos casos, se contempló recolectar la muestra con representatividad regional en el período comparable con versiones Casen anteriores (noviembre-enero) y disponer de una muestra más amplia (que contenía a la primera) que se recolectaría en un período mayor a 3 meses.

A partir de la evaluación de estos escenarios, se planteó un diseño muestral que perseguía lograr representatividad para la medición de ingreso corregido per cápita en comunas en un período

mayor a 3 meses y la medición de tasa de pobreza por ingresos para todas las regiones del país en el período noviembre-enero. Así, la variable de muestreo a partir de la cual se definen tamaños muestrales a nivel comunal sería la media comunal del ingreso corregido per cápita de los hogares, y a la vez se mantendría como variable analítica la tasa de pobreza por ingresos regional, permitiendo de este modo mantener niveles de precisión a nivel regional similares a los exigidos en versiones previas de la encuesta. Esta propuesta fue calculada utilizando datos actualizados de Casen 2013 y fue presentada al Panel de Expertos Casen 2015, para sus comentarios.

Entre los principales comentarios del Panel de Expertos Casen 2015, cabe destacar la recomendación de evaluar simulaciones de tamaños muestrales que continuasen tomando como variable de interés la tasa de pobreza por ingresos, aun cuando no estuvieran representadas todas las comunas de cada región presentes en la muestra.

Se acotó entonces la propuesta de diseño muestral, considerando los siguientes criterios:

- Se consideran en cada región las capitales regionales y provinciales y algunas comunas más importantes (en cuanto al número de viviendas que contienen en el Marco Muestral) y que acumulen un porcentaje igual o superior a 80%.
- También se considera, para las comunas que queden seleccionadas inicialmente, que el error absoluto para la medición de la tasa de pobreza a nivel regional no supere los 4 puntos porcentuales y simultáneamente el error relativo no exceda el 35%, en forma general. De acuerdo a la prevalencia observada, se aplica un criterio diferenciado según la siguiente tabla.

Tasa de pobreza por ingresos a nivel comunal	Error relativo
< 2%	<= 80%
[ 2%. - 5% )	<= 65%
[ 5%. - 10% )	<= 50 %
>= 10%	<= 36%

- La fracción de muestreo de viviendas respecto al Marco no supere el 20%. Para aquellas comunas con tamaño poblacional de viviendas menor que 10.000, se exige que la fracción de muestreo no supere el 10% (de manera de no saturarlo).
- Se aplica el tamaño muestral de viviendas de la Casen 2013 para las comunas que pertenecen al "Resto de Comunas" en la Región (estas comunas del Resto sólo están incluidas para dar presencia en la región).

Considerando que el tamaño muestral resultante de la aplicación de estos criterios se consideró viable de recolectar en el periodo comparable Casen, se optó por un diseño de una única muestra, incorporándose solo algunos ajustes adicionales conforme otros comentarios recibidos del Panel

de Expertos Casen 2015 (tales como revisar tamaños muestrales en comunas que presentaban errores absolutos superiores a 10 puntos porcentuales).

Como resultado de esta última revisión de los tamaños muestrales, niveles de precisión y fracciones de muestreo de las comunas, se implementan los siguientes criterios:

- i. Aumentar tamaño muestral en aquellas comunas de errores absolutos de 9,5 puntos porcentuales o más, de tal forma de reducirlos bajo este umbral, en la medida que se respeten fracciones de muestreo previamente definidas, según tamaño de población de viviendas de la comuna.
- ii. En aquellos casos que, respetando criterio de fracción de muestreo no sea posible aumentar el tamaño muestral, reemplazar la comuna por la(s) siguiente(s) en participación relativa del total de viviendas de la región, de tal forma de mantener a lo menos la misma proporción acumulada de viviendas regional que la inicial.
- iii. No realizar el reemplazo en caso de capitales regionales o provinciales, si no es posible.

En suma, el diseño muestral de Casen 2015 se resume en:

- utiliza como variable de muestreo la tasa de pobreza a nivel comunal (considerando la nueva metodología de medición de pobreza por ingresos y los datos de Casen 2013),
- sus niveles de representación son: nacional, urbano/rural, regional y 139 comunas capitales regionales y provinciales, y otra(s) que concentran el 80% o más de la población de viviendas de cada región (según marcos muestrales del INE);
- error absoluto esperado de 0,5 puntos porcentuales y error relativo esperado de 3,7% a nivel nacional (muestra objetivo total);
- niveles de precisión esperado a nivel comunal diferenciado según rango de tasa de pobreza por ingresos comunal.

El detalle del diseño muestral finalmente utilizado será desarrollado en la Sección III. Diseño Muestral, mientras los tamaños muestrales se presentan en la tabla II.1.

**Tabla II.1: Distribución tamaño muestral final, considerando tasa de pobreza, Nueva Metodología, Casen 2013.**

Región	Estimador de la razón	Muestra 2013 (viviendas)	Error absoluto efectivo (%)	Error relativo efectivo (%)	Error absoluto propuesto	Error relativo propuesto	Muestra objetivo propuesta	Tamaños corregidos*
<b>Total País</b>	<b>14,4%</b>	<b>64.842</b>	<b>0,6%</b>	<b>4,1%</b>	<b>0,5%</b>	<b>3,7%</b>	<b>82.370</b>	<b>107.932</b>
I de Tarapacá	8,2%	2.657	1,6%	19,7%	2,4%	18,0%	2.428	3.146
II de Antofagasta	4,0%	2.066	1,3%	33,9%	1,4%	34,8%	1.965	2.739
III de Atacama	7,3%	2.191	2,5%	34,7%	1,9%	26,1%	3.861	5.465
IV de Coquimbo	16,2%	2.956	2,4%	15,0%	2,2%	13,3%	3.738	4.587
V de Valparaíso	15,6%	6.337	1,9%	12,4%	1,6%	10,5%	8.847	11.791
VI de O'Higgins	16,0%	4.987	2,0%	12,4%	1,7%	10,4%	7.144	8.468
VII del Maule	22,3%	4.557	2,2%	10,0%	2,0%	9,0%	5.678	6.407
VIII del Biobío	22,3%	9.438	1,9%	8,7%	1,8%	7,9%	11.402	13.666
IX de la Araucanía	27,9%	5.404	2,2%	7,9%	1,9%	7,0%	6.846	7.894
X de Los Lagos	17,6%	4.063	2,1%	12,0%	1,7%	9,8%	6.097	7.108
XI de Aysén	6,8%	1.859	1,6%	24,1%	2,1%	30,8%	1.134	1.450
XII de Magallanes	5,6%	1.904	3,2%	57,9%	3,3%	60,1%	1.770	2.287
XIII Metropolitana	9,2%	10.367	1,0%	10,8%	0,8%	8,4%	17.304	28.132
XIV de Los Ríos	23,1%	3.644	2,5%	10,6%	2,6%	11,1%	3.318	3.775
XV de Arica y Parinacota	14,6%	2.412	2,0%	13,7%	3,4%	23,3%	838	1.017

\* Tasa de no respuesta, promedio ponderado Casen 2011 – 2013

Para detalles a nivel comunal ver Anexos, sección 1, Simulaciones, Tabla 1].

### III. DISEÑO MUESTRAL

La población objetivo de la encuesta Casen la constituyen todas las personas y hogares que residen en viviendas particulares a lo largo del territorio nacional. La cobertura de la encuesta Casen 2015 es todo el territorio nacional, excluyendo aquellos sectores identificados por el INE como áreas de difícil acceso<sup>10</sup>.

El diseño muestral de la encuesta Casen 2015 se puede caracterizar como probabilístico y estratificado, según área geográfica y por tamaño poblacional, tanto en el área urbana como rural. La selección de la muestra se realiza en dos etapas - muestreo bietápico en las áreas rurales y urbanas. Las unidades primarias de muestreo se seleccionan con probabilidad proporcional al tamaño, en cuanto al total de viviendas, en el área rural, mientras que en el área urbana se seleccionan de forma sistemática y con igual probabilidad. Las unidades últimas, las viviendas, se seleccionan de forma sistemática con igual probabilidad.

La unidad última de selección de la muestra es la vivienda. Al interior de la vivienda se identifican todos los hogares y las personas miembros de cada hogar. La entrevista se realiza con un informante por hogar que debe ser el jefe de hogar u otro integrante de 18 años o más. A través del informante se recolectan datos de todos los miembros del hogar que son residentes habituales de la vivienda<sup>11</sup>. Al interior de la vivienda se recolectan datos de todos los residentes habituales y los diferentes hogares que la conforman.

#### 1. Características del marco muestral

Un marco de muestreo se define como la lista o los procedimientos que permiten identificar a todos los elementos de una población objetivo (Groves et al. 2004, pág. 68). Listas de miembros de algunas organizaciones o instituciones son ejemplos de marcos de muestreo sencillos. En Chile, por ejemplo, el Ministerio de Educación tiene el listado de todos los colegios públicos, subvencionados y privados en el país.

---

<sup>10</sup> Incluyen a las comunas de General Lagos, Colchane, Ollagüe, Juan Fernández, Isla de Pascua, Cochamó, Chaitén, Futaleufú, Hualaihué, Palena, Lago Verde, Guaitecas, O'Higgins, Tortel, Laguna Blanca, Río Verde, San Gregorio, Cabo de Hornos (Ex - Navarino), Antártica, Primavera, Timaukel, Torres del Paine. Sin embargo, las comunas de General Lagos, Colchane, Lago verde, Laguna Blanca, Río Verde, San Gregorio, Primavera, Timaukel y Torres del Paine, no se considera la comuna completa como ADA, sino sólo algunos sectores, pues existen unidades muestrales en el marco de dichas comunas (entre 1 y 7 secciones).

<sup>11</sup> Hasta 2009, todas las preguntas en el cuestionario Casen podían ser respondidas por un tercero (el informante). En el cuestionario 2011 y 2013, sin embargo, las preguntas sobre satisfacción con la vida e inclusión financiera sólo pueden ser respondidas por las personas que están físicamente presentes durante la entrevista. Es decir, para estas preguntas no se acepta el reporte de terceros.

La población objetivo de la encuesta Casen 2015 son los hogares que residen en viviendas particulares ocupadas y las personas que residen de forma permanente. El INE mantiene un marco de áreas geográficas que sirve de base para la selección de viviendas, requeridas para las muestras de las encuestas de hogares más importantes del país.

Un marco muestral de área contiene las unidades geográficas de un país organizadas de forma jerárquica. En Chile, esta ordenación se denomina división político-administrativa y las unidades corresponden, en orden descendiente, a región, provincia y comuna. Al interior de cada comuna se conforma la división censal que da origen a las áreas geográficas denominadas urbana y rural.

Estas áreas se encuentran definidas por la actividad económica preponderante y dan origen a las siguientes entidades:

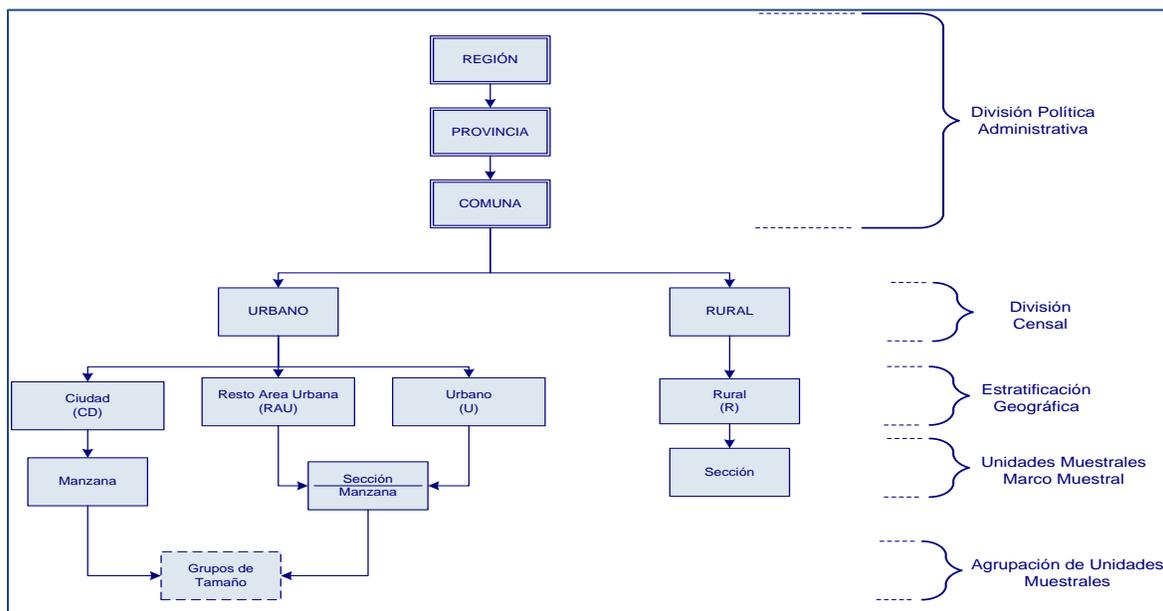
- Ciudad (CD): Es un gran centro urbano conformado por uno o un conjunto de centros urbanos adyacentes con 40.000 o más habitantes.
- Resto de Área Urbana (RAU): Conformado por un conjunto de centros urbanos que totalizan menos de 40.000 y más de 2.000 habitantes. Esta clasificación se da cuando en una comuna existe una ciudad (CD) y entonces todos los centros urbanos restantes, si es que existen, se denominan resto de área urbana (RAU).
- Urbano (U): Es el centro urbano con menos de 40.000 y más de 2.000 habitantes. Esta clasificación se da cuando en la comuna no existe una ciudad (CD), por lo que cada uno de sus centros urbanos se denominan simplemente como urbanos (U).
- Rural (R): Conformado por el conjunto de entidades clasificadas como rurales de acuerdo a un tamaño poblacional menor a 1.000 habitantes o entre 1.001 y 2.000 habitantes con predominio de población económicamente activa (PEA) dedicada a actividades primarias.<sup>12</sup>

Para efectos del Censo 2002, se realizaron sub-divisiones posteriores denominadas manzanas censales (en las áreas urbanas) y secciones de empadronamiento censal (en las áreas rurales). Estas son las unidades geográficas más pequeñas y corresponden a las unidades primarias de muestreo más comúnmente utilizadas en las encuestas de hogares diseñadas por el INE. La Figura III.1 ilustra las unidades geográficas descritas.

---

<sup>12</sup> Se entiende por Actividad Primaria a toda aquella actividad relacionada con la extracción de recursos naturales. (agricultura, caza, pesca, minería, etc.).

Figura III.1: Estratificación e identificación de unidades primarias de muestreo



Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Para cumplir con los objetivos de investigación, el marco de muestreo “ideal” debe ser completo, preciso y actualizado. A mediados del periodo intercensal, investigaciones realizadas por el INE indicaban que el marco muestral en uso (MS2002) presentaba deficiencias en estos aspectos, por lo tanto a partir de 2008 el Instituto desarrolló un nuevo marco de muestreo, definiciones para la conformación de unidades de muestreo y procedimientos para la selección de unidades de muestreo en las áreas urbanas del país<sup>13</sup>.

Bajo el nuevo marco de muestreo urbano, se mantuvo la división político-administrativa, pero se cambió la conformación de las unidades primarias de muestreo. En el nuevo marco, las manzanas censales (predios urbanos delimitados por calles) reemplazaron a las secciones de empadronamiento censal en las áreas urbanas<sup>14</sup>. Esto permitió actualizar más rápidamente las unidades muestrales que en el marco antiguo por medio del plan municipal de edificaciones que se entrega por convenio al Departamento de Gestión del Marco, logrando captar nuevos desarrollos urbanos tanto en áreas urbanas como en áreas previamente catalogadas como agrícolas (por ejemplo, nuevos desarrollos en comunas como Peñalolén, Lampa o Chicureo en la Región Metropolitana).

El INE comenzó a seleccionar las muestras urbanas utilizando el nuevo marco de muestreo a partir del año 2008.

<sup>13</sup> Para mayor información ver documento: “Actualización del diseño muestral Nueva Encuesta Nacional de Empleo”, Julio 2006, Instituto Nacional de Estadísticas. Disponible en la página web: [http://www.ine.cl/canales/chile\\_estadistico/mercado\\_del\\_trabajo/empleo/metodologia/metodologia.php](http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/metodologia.php).

<sup>14</sup> Las secciones están conformadas por una o más manzanas a requerimiento de totalizar en ella una cantidad fija de viviendas de acuerdo al estrato comuna-zona.

En los últimos años se ha observado un importante crecimiento demográfico en las áreas urbanas del país, que ha sido observado a partir de la enumeración muestral de las diversas encuestas en que el INE participa. Con frecuencia, debido a los cambios de estructura y de tamaño de las manzanas, es necesario reemplazar algunas de las unidades seleccionadas. En este contexto, en el año 2014, el Ministerio y el INE acuerdan realizar una actualización del marco muestral para la muestra de la encuesta Casen 2015. En respuesta a tal requerimiento, en el segundo semestre del año 2014, el INE realizó una actualización parcial del marco de manzanas MM2008 y que correspondió a la actualización de 2.351 manzanas 1,96% del total de 19.970 que posee el marco completo que excluye las manzanas de entre 1 y 7 viviendas y 9,7% de un total de 24.182 que poseen las 61 comunas con actualización, correspondientes a 69 urbanos que, de acuerdo a las visitas a terreno de los distintos equipos de levantamiento de encuestas en el INE, poseían altos índices de manzanas no elegibles por cambios de estructura, crecimientos y decrecimientos (ver en Anexos, sección 2, Actualización del Marco Muestral 2008) .

La selección de la muestra 2015 se realizó sobre el marco muestral que el INE mantiene vigente al año 2015, el cual comprende el uso de dos marcos de muestreo mutuamente excluyentes:

- En el área urbana y rural, el marco de muestreo corresponde al generado a partir del Censo de Población y Vivienda del año 2002. Las unidades que componen este marco de muestreo se denominan “secciones” y corresponden, en el área rural, a aglomeraciones de viviendas particulares conformadas a partir de una o más entidades pobladas, enmarcadas generalmente dentro de un distrito censal. En el área urbana, corresponden a aglomeraciones de viviendas particulares conformadas a partir de una o más manzanas según los rangos de viviendas asignados para la sección. Las secciones no sobrepasan los límites del distrito. En adelante, denominaremos este marco muestral como Marco de Secciones (MS2002).
- En el área Ciudad, en el área urbana, que no está incluida en el MS2002, el marco de muestreo es aquel generado a partir de cartografía digital, actualizado al segundo semestre del año 2008. La información en el marco se actualiza con información de los registros administrativos asociados a nuevas construcciones otorgados por los municipios. Las unidades que componen este marco de muestreo se denominan “manzanas” y corresponden a delimitaciones geográficas fijas. En adelante, denominaremos a este marco como Marco de Manzanas (MM2008). Para la encuesta Casen 2015, se utilizó el MM2008 con una actualización parcial de 69<sup>15</sup> entidades urbanas en el año 2014.

Para aquellas áreas denominadas *Resto de Área Urbana (RAU)* y *Urbano (U)* sus unidades pueden pertenecer al MS2002 o al MM2008. De un total de 219 áreas de este tipo, 89 son extraídas desde el MS2002<sup>16</sup>, mientras que las restantes 130 se extraen desde el MM2008. Cabe señalar que las 89

---

<sup>15</sup>Ver Anexos, sección 2, Actualización del Marco Muestral 2008, Tabla 2.1.

<sup>16</sup> Ver Anexos, sección 2, Actualización del Marco Muestral 2008, Tabla 2.4.

áreas urbanas seleccionadas a partir del MS2002<sup>17</sup>, corresponden a áreas de pequeña densidad poblacional, con no más de 40.000 habitantes, y las unidades muestrales, al igual que en las áreas rurales, son secciones.

Para fines de selección muestral, de distintos instrumentos de recolección de datos en que INE ha participado, se han conglomerado las unidades primarias de muestreo que conforman el MM2008; el objetivo, es poder generar grupos homogéneos de manzanas, respecto de una cierta característica y, de este modo, establecer en cada grupo un número fijo de unidades secundarias de muestreo para encuestar. Para la consecución de tal propósito, se realizó un análisis de *cluster*<sup>18</sup>, que utilizó el número de viviendas particulares, al interior de cada manzana, como variable de conglomeración. Como resultado de esta segmentación se pudo asociar las manzanas del MM2008 dentro de 5 grupos de tamaño. Desde esta última clasificación surge la variable de conglomeración que se añade al marco muestral, previo a la selección de cualquier muestra, y que permite escoger las manzanas, pero considerando grupos de manzanas homogéneas en cuanto a tamaño.

## 2. Cobertura del marco muestral

Las unidades de muestreo tienen cuatro características fundamentales para el diseño muestral: (1) cubren, usualmente, la totalidad del territorio del país; (2) tienen sus límites bien definidos; (3) existen estimaciones poblacionales para las unidades; y (4) existen mapas para las unidades (Turner, 2003).

La cobertura es una propiedad estadística asociada al marco muestral que se utiliza para la selección de la muestra. La falta de cobertura denota la falla al incluir ciertos elementos (o unidades completas) de la población a encuestar a partir del marco muestral que se ha definido (Kish 1965, pág. 528). Estas fallas no son planeadas por el investigador (ej. fallas en el proceso de conteo e identificación de las viviendas previo a la selección).

Es importante distinguir la falta de cobertura (fallas no intencionadas), de las exclusiones que realiza el investigador en forma intencionada. En el proceso de selección de la encuesta Casen 2015, al igual que en las versiones anteriores, se pueden identificar tres tipos de exclusiones de este último tipo:

- las 22 áreas geográficas que han sido catalogadas por INE como áreas de difícil acceso (ADA)<sup>19</sup>;
- las manzanas y secciones incluidas en otras muestras seleccionadas por el INE para el periodo de recolección de datos de la encuesta Casen 2015;

---

<sup>17</sup> Ver Anexos, sección 2, Actualización del Marco Muestral 2008, Tabla 2.5.

<sup>18</sup> Técnica estadística que utiliza distintas medidas de distancia para conformar grupos homogéneos en cuanto a características multivariadas.

<sup>19</sup> Ver tabla III.1 más adelante para más detalle.

- las manzanas con menos de 8 viviendas.

Las áreas de difícil acceso corresponden a zonas geográficas no incluidas en el marco muestral del INE. Estas áreas no están presentes en las muestras de ninguna de las encuestas de hogares seleccionadas por el INE. En total, corresponden a 0,4% de la población de viviendas y a 0,4% de la población de personas del país<sup>20</sup>. La Tabla III.1 muestra las 22 áreas de difícil acceso excluidas de la muestra de la encuesta Casen 2015, donde se presenta el total de viviendas según la información del Censo de Población y Vivienda del año 2002, y una aproximación del total de personas, estimadas mediante las proyecciones de población con fecha noviembre de 2015.

Por otra parte, la Encuesta Nacional del Empleo (ENE) se levanta todos los meses, lo que tiene como consecuencia que durante los meses de octubre, noviembre y diciembre, coincide temporalmente con el levantamiento de la encuesta Casen. Para evitar complicaciones en el trabajo de campo de ambas encuestas, el 2015 el INE excluyó del marco de muestreo, previo a la selección de la muestra Casen 2015, todas aquellas manzanas y secciones que forman parte de la muestra ENE. Así, el total de unidades excluidas del marco de muestreo porque se encuentran en la ENE corresponden a 1,7% de manzanas (1,8% de viviendas aproximadamente) en el MM2008, y 6,6% de secciones (6,7% de viviendas aproximadamente)<sup>21</sup>. Esto equivale, aproximadamente, a 2.314 manzanas y 497 secciones según trimestre móvil.

En las zonas urbanas, el INE excluye del MM2008 las manzanas con 7 o menos viviendas debido a potenciales problemas operativos y de costos. En total, estas manzanas representan 10,3% del total (1,3% de viviendas) del MM2008.

Por la exclusión de los conglomerados que estaban levantándose en la Encuesta Nacional de Empleo del INE en el mismo período de Casen 2015 y de aquellos de tamaño menor a 8 viviendas, el INE aplica una corrección denominada "ajuste por omisión de conglomerados", propio del cálculo de las probabilidades de selección de éstos. Sin embargo, para las áreas de difícil acceso que quedan excluidas del marco y en las cuales por defecto no existen conglomerados definidos como tales, el ajuste se hace indirectamente en el proceso de calibración o ajuste por las proyecciones de población, las cuales incluyen a nivel agregado como por ejemplo región, la población contenida en estas áreas.

En total, el marco muestral a partir del cual se seleccionó la muestra, MS2002 y MM2008 (actualizado con foco Casen 2015) en su conjunto, excluye aproximadamente 2,2% de las viviendas del país. Para algunos efectos, esta cifra puede parecer insignificante, sin embargo, es

---

<sup>20</sup> Estimaciones propias del Instituto Nacional de Estadísticas, obtenidas a partir del Censo de Población y Vivienda del año 2002.

<sup>21</sup> Los valores son referenciales, ya que existe una permanente rotación de unidades primarias muestrales y viviendas lo que puede generar que el total de viviendas que contienen las manzanas y/o secciones sean distintos de un trimestre a otro.

importante que los investigadores que hagan inferencias a partir de los datos Casen tengan estas exclusiones en consideración. Si los fenómenos de interés no están relacionados con el tamaño de las manzanas (en áreas urbanas) o no están relacionados con la aislación geográfica (en áreas rurales), entonces es probable que la falta de cobertura del marco muestral no afecte las inferencias hechas a partir de los datos de la encuesta. Si lo contrario es cierto, entonces los investigadores deberán ser cautelosos al momento de hacer inferencia a la población bajo estudio.

**Tabla III.1: Áreas de Difícil Acceso definidas por el INE**

Región	Provincia	Comuna	Total Viviendas Particulares ocupadas Censo 2002	Total de personas Proyecciones de Población*
Arica y Parinacota	Parinacota	General Lagos	260	1.207
Tarapacá	Tamarugal	Colchane	461	1.588
Antofagasta	El Loa	Ollagüe	66	232
Valparaíso	Valparaíso	Juan Fernández	206	965
	Isla de Pascua	Isla de Pascua	1.136	5.637
	Llanquihue	Cochamó	1.345	4.223
Los Lagos	Palena	Chaitén	1.830	6.870
		Futaleufú	606	1.797
		Hualaihué	2.249	8.316
		Palena	558	1.610
		Lago Verde	338	886
Aysén del General Carlos Ibáñez del Campo	Aysén	Guaitecas	383	1.943
	Capitán Prat	O'Higgins	154	779
		Tortel	145	686
		Laguna Blanca	116	636
Magallanes y La Antártica Chilena	Magallanes	Río Verde	86	360
		San Gregorio	212	604
		Cabo de Hornos	520	3.002
	Antártica Chilena	Antártica	10	74
	Tierra el Fuego	Primavera	228	473
		Timaukel	82	1.047
		Torres del Paine	114	1.325

\*30 Noviembre de 2015

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

### 3. Estratificación del marco muestral

En el diseño de una muestra, la estratificación corresponde al proceso de agrupar a los elementos de una población en grupos homogéneos previo a la selección de la muestra. Su propósito es mejorar la precisión estadística de los estimadores agrupando las unidades del marco en clases homogéneas en su interior y que difieran de las características del resto. Los estratos deben ser mutuamente excluyentes, es decir, cada elemento en la población debe ser asignado a un solo estrato. Además, los estratos deben ser exhaustivos colectivamente, es decir, ningún elemento de la población puede quedar excluido.

La encuesta Casen ha definido tradicionalmente los estratos de selección de la muestra a partir del criterio de proximidad geográfica. Hasta 1996 los estratos se conforman a partir de comunas y grupos de comunas, según área urbano/rural<sup>22</sup>. A partir de 1998 los estratos se constituyen a partir de las comunas, según área urbano/rural<sup>23</sup>. Para la encuesta Casen 2015 se conformaron  $h = 588$  estratos, a partir de la interposición de la división político-administrativa (a nivel de comunas) y la división censal (urbano-rural). De éstos, 312 estratos se encuentran en zonas urbanas y 276 en zonas rurales. Además, en la parte urbana, se incorporó una estratificación adicional que, a diferencia de las anteriores de carácter principalmente geográfico, corresponden a la clasificación de todas las manzanas del marco en grupos de tamaño (ver Tabla III.2) de acuerdo a su tamaño en viviendas.

En el MM2008, al interior de cada comuna, las manzanas fueron clasificadas según su tamaño en 5 grupos, con el objetivo de poder distribuir la muestra de viviendas, en cada uno de estos grupos de manzanas y de esta forma, tener representadas en la muestra los diversos grupos de tamaño<sup>24</sup>. La distribución de los grupos de tamaño en el marco y en la muestra seleccionada se presenta a continuación en las Tablas III.2 y III.3.

---

<sup>22</sup> En 1994 se conforman 146 estratos y en 1996, 246 estratos.

<sup>23</sup> En 1998 se conformaron 358 estratos. En 2000, 2003, 2006 y 2009 se conformaron 529, 553, 605 y 602 estratos respectivamente.

<sup>24</sup> Es posible que en algunas comunas no existan los 5 grupos de tamaño porque no existen manzanas en el rango del grupo faltante.

**Tabla III.2: Número de manzanas y viviendas en el MM2008, según grupo de tamaño de manzanas**

Grupo de tamaño	Rango de viviendas	Total de Manzanas	% de manzanas	Total de Viviendas	% de viviendas
<b>Total</b>		<b>133.759</b>	<b>100,0</b>	<b>4.064.511</b>	<b>100,0</b>
Grupo 0	1 a 7	13.789	10,3	53.425	1,3
Grupo 1 a 4	8 a 23	63.355	47,4	995.329	24,5
Grupo 5 a 9	24 a 44	39.454	29,5	1.222.921	30,1
Grupo 10 a 19	45 a 81	10.382	7,8	594.380	14,6
Grupo 20 a 28	82 a 154	4.099	3,1	451.936	11,1
Grupo 29 y 30	155 y más	2.680	2,0	746.520	18,4

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

**Tabla III.3: Número de manzanas y viviendas seleccionadas en el MM2008, según grupo de tamaño**

Grupo de tamaño	Rango de viviendas	Total de Manzanas	% de manzanas	Total de Viviendas	% de viviendas
<b>Total</b>		<b>10.201</b>	<b>100</b>	<b>82.174</b>	<b>100,0</b>
Grupo 0	1 a 7	0	0,0	0	0,0
Grupo 1 a 4	8 a 23	5.529	54,2	22.883	27,8
Grupo 5 a 9	24 a 44	3.107	30,5	25.966	31,6
Grupo 10 a 19	45 a 81	817	8,0	12.220	14,9
Grupo 20 a 28	82 a 154	345	3,4	8.949	10,9
Grupo 29 y 30	155 a 1.043	403	4,0	12.156	14,8

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

#### 4. Estimación del tamaño muestral

Uno de los objetivos del diseño muestral de Casen 2015, fue alcanzar mejores niveles de precisión para las estimaciones comunales<sup>25</sup>. Para ello se realizaron, en primera instancia, un conjunto de simulaciones donde el parámetro asociado o variable de interés fue: (1) La tasa de pobreza por ingresos estimada bajo metodología tradicional de medición de pobreza e ingresos, aplicada a datos de Casen 2011, (2) el ingreso corregido total del hogar per cápita, según metodología tradicional, considerando datos de Casen 2011<sup>26</sup>, y (3) cuando estuvo disponible la base de datos, el ingreso corregido total del hogar per cápita estimado con nueva metodología de medición, y datos de Casen 2013. Los niveles de precisión, específicamente el error relativo, establecido en estas simulaciones para la variable de interés a nivel comunal oscilaron entre 10% y 35%. Así, para estimar la tasa de pobreza por ingresos a nivel comunal, bajo estos criterios, se obtuvo que se

<sup>25</sup> Con excepción de las 22 comunas clasificadas como áreas de difícil acceso.

<sup>26</sup> Submuestra recolectada entre noviembre 2011 y enero 2012.

requeriría una muestra objetivo superior a las 190.000 viviendas, mientras que para el ingreso corregido total del hogar per cápita, la muestra superaría las 120.000 viviendas.

Sin embargo, al presentar los diferentes escenarios<sup>27</sup> al Panel de expertos Casen 2015, éste recomendó que el tamaño muestral fuera determinado considerando como principal variable de interés la tasa de pobreza por ingresos obtenida en Casen 2013. Junto con lo anterior, sugirió priorizar la precisión de aquellas comunas de mayor importancia como capitales regionales y provinciales del país, entre otras que el Ministerio considerara prioritarias respecto a la necesidad de información.

En este contexto, para la encuesta Casen 2015 se determinaron 139 comunas de interés para obtener estimaciones de la tasa de pobreza con márgenes de error tolerables. Estas comunas fueron determinadas a partir de los siguientes criterios: (1) capitales regionales, (2) capitales provinciales; y (3) comunas más importantes de la región en cuanto al total de viviendas que contiene el marco muestral, y que ellas en conjunto con las capitales regionales y provinciales acumularan al menos 80% del total de viviendas de la región.

El tamaño muestral fue determinado en forma independiente para cada una de las regiones del país, y posteriormente en las comunas definidas como dominios de estudio y las restantes. La determinación del tamaño muestral se realizó en etapas, considerando los parámetros estimados a partir de los datos de la encuesta Casen de 2013 (tasa de pobreza por ingresos, efecto diseño, tasa de respuesta) y versión 2011 (tasa de respuesta) (Ver tablas III.4 y III.5).

Así, el tamaño muestral de la encuesta Casen 2015 fue determinado de acuerdo al siguiente diseño:

1. Se definió un tamaño muestral para la muestra objetivo para cada una de las 15 regiones del país, a partir de la tasa de pobreza por ingresos de la encuesta Casen 2013, utilizando la nueva metodología de cálculo de éste indicador, de manera que el error absoluto de estas tasas no superara 4 puntos porcentuales, y en lo posible el error relativo no superara 35%.
2. Se distribuyó el total de viviendas de la región de forma proporcional en las comunas en cuanto al total de viviendas al interior de cada región.
3. Se modificó el tamaño muestral de aquellas comunas definidas como dominio de estudio, con el objetivo de que éstas tuvieran los márgenes de error predeterminados, y que la fracción de muestreo no superara 20%.
4. Para determinar el tamaño muestral de las restantes comunas, se estableció que estas comunas debían tener, aproximadamente, el mismo tamaño que tuvieron en Casen 2013.

---

<sup>27</sup> Mayor detalle de todas las simulaciones de tamaño muestral ver documento: "Informe de Resultados de las Simulaciones de Tamaños Muestrales, para que la Encuesta Casen 2015 tenga representatividad a nivel Nacional, Nacional Urbano-Rural, Regional y en Comunas".

5. Luego, para todas las comunas se calculó el tamaño muestral sobredimensionado, y con ello se determinó el tamaño sobredimensionado de la región a partir de la agregación de las comunas.
6. Finalmente, se distribuyó la muestra de acuerdo a la desagregación geográfica de cada comuna, urbana y/o rural, en aquellos casos que fuera posible.

El detalle de los niveles de precisión esperados (error absoluto y error relativo) y los tamaños muestrales en cada etapa del proceso se encuentran en los siguientes párrafos.

**Tabla III.4: Tamaño muestral bajo muestreo aleatorio simple Casen 2015 y parámetros, por región**

Región	Población Vivienda Casen 2013 $M_o$	Estimador de la Razón Casen 2013 $P_{i(2013)}$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Tamaño propuesto bajo (M.A.S) Casen 2015 $m_{or}$	Efecto Diseño Casen 2013 $d_{eff}$
<b>Nivel País</b>	<b>4.441.666</b>	<b>14,4%</b>	<b>0,5%</b>	<b>3,7%</b>	<b>40.119</b>	<b>3,7</b>
I de Tarapacá	73.430	8,2%	2,4	18,0	3.418	0,73
II de Antofagasta	134.676	4,0%	1,4%	34,8%	1.375	1,45
III de Atacama	71.377	7,3%	1,9%	26,1%	2.320	1,76
IV de Coquimbo	186.606	16,2%	2,2%	13,3%	1.708	2,23
V de Valparaíso	531.339	15,6%	1,6%	10,5%	2.259	3,98
VI de O'Higgins	191.955	16,0%	1,7%	10,4%	3.622	2,05
VII del Maule	257.056	22,3%	2,0%	9,0%	2.290	2,54
VIII del Biobío	526.671	22,3%	1,8%	7,9%	3.415	3,41
IX de La Araucanía	241.585	27,9%	1,9%	7,0%	3.787	1,86
X de Los Lagos	190.036	17,6%	1,7%	9,8%	2.983	2,11
XI de Aysén	26.510	6,8%	2,1%	30,8%	3.419	0,35
XII de Magallanes	45.198	5,6%	3,3%	60,1%	874	2,11
XIII Metropolitana	1.814.485	9,2%	0,8%	8,4%	2.906	6,01
XIV de Los Ríos	94.621	23,1%	2,6%	11,1%	3.398	1,01
XV de Arica y Parinacota	56.121	14,6%	3,4%	23,3%	2.345	0,36

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

**Tabla III.5: Tamaño muestral bajo diseño complejo Casen 2015 y parámetros, por región**

Región	Tamaño M.A.S ajuste diseño complejo Casen 2015 $m_{1r}$	Fracción de muestreo $f$	Tamaño corregido por finitud Casen 2015 $m_{2r}$	Tasa de no respuesta $tnr^*$	Tamaño muestral corregido por tasa de no respuesta
<b>Nivel País</b>	<b>84.411</b>	<b>1,9%</b>	<b>82.370</b>	<b>23,7%</b>	<b>107.932</b>
I de Tarapacá	2.511	3,4%	2.428	22,8%	3.146
II de Antofagasta	1.994	1,5%	1.965	28,3%	2.739
III de Atacama	4.082	5,7%	3.861	29,4%	5.465
IV de Coquimbo	3.814	2,0%	3.738	18,5%	4.587
V de Valparaíso	8.997	1,7%	8.847	25,0%	11.791
VI de O'Higgins	7.420	3,9%	7.144	15,6%	8.468
VII del Maule	5.806	2,3%	5.678	11,4%	6.407
VIII del Biobío	11.654	2,2%	11.402	16,6%	13.666
IX de La Araucanía	7.046	2,9%	6.846	13,3%	7.894
X de Los Lagos	6.299	3,3%	6.097	14,2%	7.108
XI de Aysén	1.185	4,5%	1.134	21,8%	1.450
XII de Magallanes	1.842	4,1%	1.770	22,6%	2.287
XIII Metropolitana	17.471	1,0%	17.304	38,5%	28.132
XIV de Los Ríos	3.439	3,6%	3.318	12,1%	3.775
XV de Arica y Parinacota	851	1,5%	838	17,6%	1.017

Fuente: Elaborado por el Instituto Nacional de Estadísticas

\* Para la tasa de no respuesta a nivel comunal, se utilizó el promedio ponderado entre las tasas de no respuesta de Casen 2011 submuestra2 y la de Casen 2013. La tasa de no respuesta a nivel regional, se obtiene por despeje entre la muestra objetivo esperada y la muestra con sobremuestreo a nivel regional (ver fórmula 5 del paso 6, más adelante).

La nomenclatura para describir las fórmulas que siguen, se presenta a continuación, asumiendo que el índice  $r$  corresponde a región;  $h$  identifica el estrato de muestreo Casen;  $i$  corresponde a la unidad primaria de muestreo (es decir manzanas o secciones);  $j$  es la unidad secundaria de muestreo (es decir. viviendas).

Considérese,

$p$  : Prevalencia de la variable cualitativa de interés en el estudio, corresponde a la tasa de pobreza en la región  $r$ .

$n$  : Número de conglomerados o unidades primarias de muestreo resultantes en la región  $r$ .

$\bar{m}$  : Número promedio de viviendas a encuestar por unidad primaria de muestreo resultantes en la región  $r$ .

$n * \bar{m} = m_r$  : Número de viviendas a encuestar en la región  $r$ .

$M_r$  : Número de viviendas en la región  $r$ .

- $Deff(p)_r$  : Efecto del diseño asociado a la tasa de pobreza ( $p$ ) en la región  $r$ , que se puede interpretar como el aumento o disminución en la varianza, debido a considerar un muestreo complejo (ejemplo. estratificado, bietápico, por conglomerados) en vez de un muestreo aleatorio simple de viviendas.
- $SE(p)$  : Error estándar de la estimación de la pobreza ( $p$ ) en la región  $r$ , igual a la raíz cuadrada de la varianza de la estimación.
- $S(p)_r^2$  : Cuasivarianza poblacional de la tasa de pobreza ( $p$ ) en la región  $r$ . Para variables cualitativas y en un muestreo aleatorio simple, corresponde aproximadamente a  $(M_r/(M_r - 1)) \cdot (P \cdot Q)$ , siendo  $P$  la Tasa de pobreza por ingresos regional, con  $Q = (1 - P)$  y  $M_r$  el total de viviendas en la región  $r$ .
- $Z_{1-\alpha/2}$  : Percentil de nivel  $(1 - \alpha/2)$  de la distribución Normal, correspondiente a una estimación intervalar de  $(1 - \alpha)$  de confianza.

A continuación se describe en detalle cada uno de los seis pasos realizados para definir la muestra con el total de viviendas a encuestar en cada región  $r$ .

### **Paso 1**

Se obtienen estimaciones, a nivel regional, de la tasa de pobreza por ingresos, el error estándar asociado a la tasa de pobreza por ingresos y el efecto del diseño asociado a este indicador, usando Casen 2013. La cuasi-varianza poblacional  $S(p)_r^2$  se deriva a partir del error estándar de la tasa de pobreza 2013 utilizando la siguiente fórmula:

$$S(p)_r^2 = \frac{n \cdot \bar{m}}{Deff(p)_r} \cdot SE(p)_r^2 \quad (1)$$

### **Paso 2**

Se establecen parámetros para los errores absolutos "objetivo" por región ( $e_{0r}$ ) que varían entre 1 y 4 puntos porcentuales y el error relativo asociado según la estimación de la tasa de pobreza por ingresos, en lo posible, que no sobrepase 35%, salvo en la región de Magallanes en la cual el error relativo alcanza 60,1%. También se establece el nivel de confianza de 95% para obtener el percentil de la distribución normal  $Z_{1-\alpha/2}$ . Con estos y la cuasi-varianza de la tasa de pobreza derivada en (1), se obtiene un tamaño muestral considerando un muestreo aleatorio simple mono-etápico  $m_{0r}$ .

$$m_{0r} = \frac{Z_{1-\alpha/2}^2 \cdot S(p)_r^2}{e_{0r}^2} \quad (2)$$

### **Paso 3**

Se estima el tamaño muestral considerando el efecto del diseño al aplicar un muestreo en dos etapas. El efecto del diseño representa el ajuste del tamaño anterior debido al cambio que experimentará la varianza al pasar de un muestreo mono-etápico a uno bi-etápico, y con el

objetivo de mantener el error absoluto inicial. Para esto se utilizan el efecto del diseño por región asociado a la tasa de pobreza por ingresos  $Def(p)_r$  (ver columna 7 en Tabla III.6). El  $Def$  más bajo se observa en las regiones de Arica y Parinacota y Aysén, mientras que el más alto en la Región Metropolitana.

La fórmula utilizada para el cálculo del tamaño muestral  $m_{1r}$ , incorporando el efecto del diseño se presenta a continuación:

$$m_{1r} = m_{0r} \cdot Def(p)_r \quad (3)$$

Donde  $m_{0r}$  representa el total de viviendas estimadas a encuestar en la región  $r$  mediante un muestreo aleatorio simple y  $m_{1r}$  representa el número de viviendas a encuestar en la región  $r$ , ajustado por efecto de diseño. En la Tabla III.6, columna 6 se presentan los valores de  $m_{0r}$ , mientras que la Tabla III.7 columna 2 los valores de  $m_{1r}$ , según región.

#### **Paso 4**

El tamaño anterior a su vez se corrige por un ajuste para muestreo en poblaciones finitas o corrección por finitud  $(1 - m_{1r}/M_r)$  llegando al cálculo de  $m_{2r}$ . La fórmula utilizada se presenta a continuación:

$$m_{2r} = \frac{Z_{1-\alpha/2}^2 \cdot S(p)_r^2 \cdot Def(p)_r}{e_{0r}^2 + Z_{1-\alpha/2}^2 \cdot Def(p)_r \cdot S(p)_r^2 / M_r} = \frac{m_{1r}}{1 + m_{1r}/M_r} \quad (4)$$

Donde  $m_{2r}$  representa el total de viviendas estimadas a encuestar en la región  $r$  mediante un diseño muestral bietápico que incorpora el efecto del diseño y el ajuste por finitud o ajuste para poblaciones finitas, y  $M_r$  corresponde al total de viviendas de la Población en la región  $r$ .

#### **Paso 5**

Posteriormente se determinó el tamaño muestral de las comunas  $m_{2c}$ , para ello se distribuyó el total de viviendas regional de forma proporcional en las comunas en cuanto al total de viviendas al interior de cada región, y se modificó el tamaño muestral de aquellas comunas definidas como dominio de estudio con el objetivo de que éstas tengan los márgenes de error predeterminados, y que la fracción de muestreo no supere 20%. Luego, para determinar el tamaño muestral de las restantes comunas, se estableció que estas comunas debían tener, aproximadamente, el mismo tamaño que tuvieron en la versión 2013.

#### **Paso 6**

Una vez definido el total de unidades muestrales a encuestar para estimar la tasa de pobreza por ingresos, se debe tener en consideración que es posible no lograr el total de unidades por diversas razones, tales como rechazos, moradores ausentes, etc. Con la finalidad de salvaguardar la

precisión de la estimación de la tasa de pobreza, se aplica un factor de corrección a la fórmula (4) en base a una estimación de la tasa de no respuesta.

En años anteriores a 2011, se usó un factor de ajuste por no respuesta parejo para todas las áreas geográficas de aproximadamente 25%. El año 2011, se utilizó la tasa de respuesta regional, estimada a partir de los resultados del trabajo de campo de la Encuesta Casen 2009.

En 2013 se utilizaron las tasas de no-respuesta obtenidas de Casen 2011, submuestra 2. Como se esperaba de la submuestra 2 de Casen 2011, el aumento desproporcionado de la muestra en las regiones extremas significó mayores costos operativos y mayores dificultades de logro de las encuestas a levantar, junto con altas tasas de no respuesta.

Para la encuesta Casen 2015 se utilizó el promedio ponderado de la tasa de no respuesta de las encuestas Casen 2011 (submuestra2) y Casen 2013, a nivel comunal. Así, se obtuvo el tamaño muestral con sobre-muestreo para cada comuna del país, y luego por agregación se determinó el tamaño a nivel regional. En la figura 2 se presenta la distribución de los promedios de tasas comunales de no respuesta, según región.

La fórmula que define el tamaño muestral final de la encuesta Casen 2015, que incorpora finalmente los ajustes conjuntos de finitud, efecto de diseño y tasa de no-respuesta implementada para cada comuna del país, está dada por:

$$m_{3c} = \frac{m_{2c}}{1 - tnr_c} \quad (5)$$

Donde  $m_{3c}$  representa el número de viviendas final a encuestar en la comuna  $c$ .

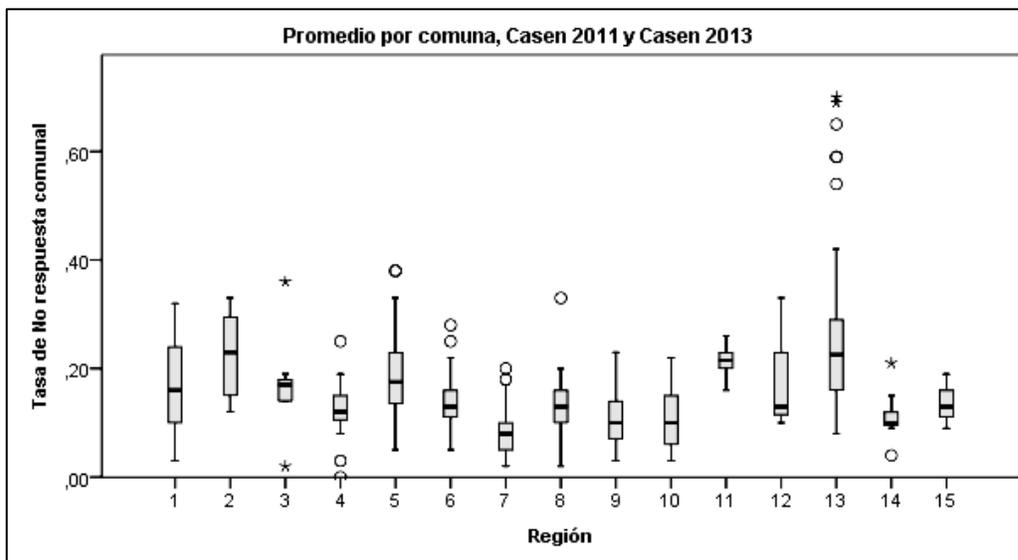
### **Paso 7**

Cabe señalar que el tamaño muestral regional con sobre-muestreo  $m_{3r}$ , fue calculado como la suma o agregación comunal del total de viviendas determinado en cada una de las comunas de la región que están presentes en la muestra de Casen 2015. Luego, de la expresión (5) se puede despejar la tasa de no respuesta y estimar la tasa regional a partir de la siguiente fórmula:

$$tnr_r = \frac{m_{3r} - m_{2r}}{m_{3r}} \quad (6)$$

La Figura III.2 informa la distribución de tasas comunales de no respuesta en cada región del país, considerando el valor promedio simple entre Casen 2011 y Casen 2013.

Figura III.2: Distribución de tasas de no-respuesta comunal, según región



Fuente: Elaborado por el Instituto Nacional de Estadísticas.

En la Tabla III.6 se presenta el total de viviendas y su distribución según el marco muestral INE utilizado en diseño muestral de Casen 2015, el cual fue desarrollado en base al Censo de Población y Viviendas de 2002. Las columnas que siguen presentan el total de viviendas y la distribución de la muestra seleccionada para la encuesta Casen 2009, 2011, 2013<sup>28</sup> y 2015.

Al comparar las muestras, se observa que la distribución de la muestra 2009 se aproxima más a la distribución de la población de viviendas. Esto es consistente con el diseño muestral empleado hasta 2009, que asignaba más muestra a las regiones con más comunas (las cuales tienen mayor población de viviendas).

En la muestra 2013 se disminuyó el tamaño en las regiones extremas, dados los costos operativos que llevó a levantar una muestra relativamente grande para esas regiones (Casen 2011) acercándola a su vez, a la distribución porcentual a la de Casen 2009 como a la del marco muestral de viviendas de 2002.

Finalmente, si se compara el tamaño muestral de las tres últimas versiones de la encuesta, sin considerar la región Metropolitana, la muestra de la encuesta Casen 2015 se asemeja más a la Casen 2009, que también persiguió representatividad comunal, que a las otras dos consideradas.

<sup>28</sup> En primera instancia, la muestra objetivo de Casen 2013 comprendía un total de 75.080 viviendas. Luego, a solicitud del Ministerio, ésta fue reducida a 70.080 viviendas.

**Tabla III.6: Distribución de la muestra objetivo para las encuestas Casen 2009, 2011, 2013 y 2015**

Región	Censo 2002		Casen 2009		Casen 2011		Casen 2013		Casen 2015	
	Población de viviendas	%	Muestra	%	Muestra	%	Muestra	%	Muestra	%
<b>País</b>	<b>4.594.668</b>	<b>100</b>	<b>74.339</b>	<b>100</b>	<b>90.122</b>	<b>100</b>	<b>70.080</b>	<b>100</b>	<b>82.370</b>	<b>100</b>
I de Tarapacá	71.684	1,6	1.418	1,9	6.241	6,9	2.977	4,2	2.428	2,9
II de Antofagasta	123.340	2,7	1.995	2,7	6.108	6,8	2.513	3,6	1.965	2,4
III de Atacama	68.332	1,5	1.953	2,6	4.382	4,9	2.071	3,0	3.861	4,7
IV de Coquimbo	198.469	4,3	3.255	4,4	3.780	4,2	3.018	4,3	3.738	4,5
V de Valparaíso	491.287	10,7	8.045	10,8	8.064	8,9	6.737	9,6	8.847	10,7
VI de O'Higgins	244.622	5,3	6.874	9,2	5.469	6,1	5.137	7,3	7.144	8,7
VII del Maule	288.879	6,3	6.430	8,6	7.464	8,3	4.872	7,0	5.678	6,9
VIII del Biobío	553.948	12,1	11.850	15,9	8.652	9,6	9.826	14,0	11.402	13,8
IX de La Araucanía	267.546	5,8	6.760	9,1	5.782	6,4	5.387	7,7	6.846	8,3
X de Los Lagos	228.440	5,0	5.825	7,8	6.391	7,1	4.192	6,0	6.097	7,4
XI de Aysén	28.211	0,6	1.300	1,7	3.941	4,4	2.051	2,9	1.134	1,4
XII de Magallanes	45.118	1,0	1.210	1,6	2.315	2,6	2.085	3,0	1.770	2,1
XIII Metropolitana	1.833.728	39,9	14.125	19,0	11.461	12,7	12.865	18,4	17.304	21,0
XIV de Los Ríos	107.635	2,3	2.535	3,4	6.576	7,3	3.768	5,4	3.318	4,0
XV de Arica y Parinacota	43.429	0,9	764	1,0	3.495	3,9	2.581	3,7	838	1,0

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

#### **4.1. Determinación del tamaño muestral para las comunas.**

Una vez obtenido el tamaño muestral regional se determinó, al interior de cada región, el total de viviendas a seleccionar en cada una de las comunas.

En primera instancia, se determinó el tamaño muestral de las comunas que se desea sean dominios de estudio, según los mismos criterios establecidos para las regiones, es decir, que el error absoluto asociado a la tasa de pobreza no supere 4 puntos porcentuales y, al mismo tiempo, el error relativo no supere 35%. Con estas restricciones se obtuvo que algunas comunas incrementaban tanto su tamaño muestral que el tamaño regional perdía el orden de proporcionalidad establecido originalmente. Esto debido a que algunas comunas poseen alta variabilidad, o simplemente porque las estimaciones de la tasa son tan pequeñas que aun cuando poseen un error absoluto pequeño (menor a 4%) el error relativo asociado puede ser mayor a 35%.

Para enmendar este problema, el INE en conjunto con el Ministerio de Desarrollo Social, establecieron ciertos criterios relacionados con la precisión deseada y fracción de muestreo, a fin de obtener, en lo posible, un tamaño muestral nacional, regional y comunal que respetara, en la medida de lo posible, el orden de proporcionalidad nacional, regional y comunal de la población.

#### **Criterios para considerar las comunas como dominios de estudio**

Dentro de los criterios que se establecieron, los dos primeros fueron de tipo político administrativo<sup>29</sup>, mientras que los dos últimos son de precisión o errores de las estimaciones de la tasa de pobreza por ingresos:

1. Las comunas deben ser capitales regionales y/o provinciales.
2. Deben estar entre las más grandes que acumulen aproximadamente 80% de las viviendas a nivel regional.
3. La fracción de muestreo no debe superar 20%, es decir, el total de viviendas a encuestar en cada comuna no debe ser más de 20% de las viviendas de la población, según lo registrado en el marco muestral. En el caso de las comunas con tamaño poblacional menor de 10.000 viviendas, la fracción de muestreo no debe superar 10%.
4. De acuerdo a la tasa de pobreza por ingresos observada de la encuesta Casen 2013, se aplica un criterio diferenciado en la precisión tal como se señala en la Tabla III.7.

---

<sup>29</sup> Con esto, puede ocurrir que algunas comunas no cumplan con los criterios de errores absolutos y/o relativos establecidos en los puntos 3 y 4 y aún así ser dominio de estudio. Ver listado de comunas según los criterios en Anexos, sección 4, Dominios de Estudio.

**Tabla III.7: Criterio de precisión comunas auto-representadas**

Tasa de Pobreza	Error relativo
<2%	<= 80%
[2% - 5%)	<=65%
[5% - 10%)	<=50%
>=10%	<=36%

Fuente: Elaborado por el Instituto Nacional de Estadísticas

Luego, al aplicar estos criterios, el tamaño muestral de ciertas comunas fue modificado (puede haber aumentado o disminuido) en virtud de conservar el tamaño original de la región y de cumplir las restricciones impuestas. Así, de las 139 comunas definidas inicialmente como dominios de estudio,

- 29 poseen error absoluto menor o igual a 4% y error relativo igual o inferior a 35%;
- 73 comunas tienen tasa de pobreza estimada mayor o igual a 10%, fracción de muestreo menor a 20% y error relativo menor a 36%;
- 19 tienen tasa de pobreza estimada mayor o igual a 5% y menor a 10%, fracción de muestreo menor a 20% y error relativo menor o igual a 50%; 9 tienen tasa de pobreza estimada mayor o igual a 2% y menor a 5%, fracción de muestreo menor a 20% y error relativo menor o igual a 65%;
- finalmente solo 2 poseen tasa menor a 2% y error relativo igual o menor a 80%.
- Las restantes 7 comunas no cumplen los criterios de error 3 y 4 pre-establecidos, sin embargo, fueron igualmente consideradas como dominios de estudio por ser capitales provinciales o porque con ellas se lograba el objetivo de tener representadas por lo menos 80% de las viviendas en las comunas más importantes de la región en cuanto a tamaño.

Si bien se flexibilizaron los criterios de precisión en aquellos casos donde el error relativo es superior a 35% se sugiere que los usuarios revisen errores efectivos<sup>30</sup>, para evaluar si las estimaciones que se realice con la encuesta son suficientemente precisas para sus fines de investigación.

En las tablas de Anexos, Dominios de Estudio, se presenta el listado de comunas consideradas como dominios de estudio, y se expone la tasa de pobreza por ingresos estimada directamente de la encuesta Casen 2013 (aun cuando ésta no era representativa a nivel comunal) el error absoluto y relativo asociado, la fracción de muestreo y la muestra objetivo.

Para aquellas comunas no consideradas como dominio de estudio, se asignó como muestra objetivo el total de viviendas entrevistado en Casen 2013, o aproximadamente, el mismo valor, tal que la agregación de todas las comunas de la región no sobrepasara el total de viviendas establecido.

<sup>30</sup> De las 139 comunas definidas como dominios de estudio, 7 de ellas entraron a esta categoría por ser capitales provinciales y/o regionales, o bien porque con ellas se llegaba al 80% de la población de viviendas de la región, presentando errores relativos que no cumplen los criterios 3 y/o 4.

#### 4.2. Asignación del total de viviendas para la sobredimensión

La finalidad de sobredimensionar la muestra objetivo ( $n=82.370$ ) es contrarrestar las pérdidas que se originan producto de distintas formas de no-respuesta (ej. unidades no elegibles, falta de contacto, rechazo a la entrevista, etc.). El diseño 2015, toma un ajuste diferenciado por región y comuna, según la no-respuesta observada en el proceso de encuestaje Casen 2011, específicamente de la submuestra 2, y de Casen 2013. Las tasas de respuesta se pueden interpretar como una estimación de la probabilidad de responder de los hogares en las comunas seleccionadas<sup>31</sup>.

Luego, en cada comuna se aplicó la fórmula (5) y se obtuvo el tamaño muestral sobredimensionado de la comuna, y con ello la muestra sobredimensionada de la región.

Finalmente, se debe distribuir la muestra en el área urbana y rural, lo que se presenta en el siguiente apartado.

#### 4.3. Distribución de la muestra de viviendas en áreas urbanas y rurales

Una vez establecido el total de viviendas en cada región y comuna se realiza la distribución de las viviendas por área geográfica de las 107.932 viviendas con sobremuestreo  $m_{3r}$ . A continuación, se describe el protocolo de distribución de la muestra Casen 2015.

Como cada comuna puede tener área urbana y/o rural, se hace necesaria la distribución al interior de cada comuna según área geográfica. Para esto se obtiene el peso o el aporte relativo, en cuanto al total de viviendas, de cada área geográfica en la comuna respectiva según el Censo 2002. La distribución de viviendas al interior de cada comuna se realiza en función de estos pesos.

Para las versiones anteriores de la encuesta, se determinaron algunas condiciones adicionales dentro del protocolo de distribución de la muestra:

- Se estableció un número mínimo de viviendas por estrato: 50 en estratos urbanos y 30 en estratos rurales<sup>32</sup>.
- Se estableció un mínimo de 2 manzanas (o secciones) por estrato, tanto para estratos urbanos como rurales.

La primera condición, que implica fijar un mínimo de viviendas por estrato, busca estabilizar las estimaciones de varianza. Dicho de otro modo, reducir la variabilidad de las estimaciones de la varianza. La segunda condición, que implica fijar un mínimo de 2 manzanas (o secciones) por estrato, busca hacer

---

<sup>31</sup> Ver más detalles del marco conceptual de la no respuesta en encuestas de hogares en "Nonresponse in Household Surveys" (Groves and Couper, 1998).

<sup>32</sup> Con 50 viviendas en el área urbana, se asegura un mínimo de 6 manzanas (tomando en cuenta que en promedio se encuestan 8 viviendas por manzana). En el área rural, se asegura un mínimo de dos secciones, tomando también en cuenta que en esta área se encuestan en promedio 15 viviendas por sección.

factible la estimación de varianza bajo muestreo complejo al evitar, por diseño, el caso de un solo conglomerado por estrato.

Para esta nueva versión de la encuesta, dado que en la parte urbana se introdujo una estratificación adicional por grupo de tamaño, estas condiciones no fueron aplicadas a este nivel de grupo de tamaño, pero sí a nivel de estrato geográfico comuna-área, como en las versiones anteriores. No obstante, esto será subsanado, al igual que lo realizado en los años anteriores, mediante la construcción de pseudo-estratos y pseudo-conglomerados para la estimación de la varianza.

## **5. Selección de conglomerados de la muestra**

Como se mencionó en la sección 2 apartado 1, el INE mantiene un marco de áreas que contiene un listado de unidades (que denominaremos conglomerados) que sirven de base para la identificación y posterior selección de las viviendas, que son las unidades últimas de selección de la muestra.

En esta sección, se describen los procedimientos utilizados en la selección de conglomerados. La selección se realiza de forma distinta dependiendo del marco de muestreo en que se encuentra registrada la parte urbana o rural de una comuna. A continuación, se describen los protocolos utilizados para la selección de manzanas (a partir del MM2008) y de secciones (a partir del MS2002).

### **5.1. Selección de conglomerados desde el MM2008**

A diferencia de Casen 2011 y 2013, el INE, a través de procesos continuos de mejora, ha realizado un cambio en la forma de seleccionar los conglomerados desde el MM2008, actualizado con foco Casen 2015. Este cambio obedece principalmente a la optimización de las estimaciones y minimización de las varianzas de éstas. Para conocer la metodología utilizada en 2011 y 2013 refiérase a los Manuales de Diseño Muestral de dichos años.

#### **5.1.1. Selección de conglomerados en Casen 2015**

Con el objeto de optimizar las estimaciones y minimizar las varianzas. El MM2008 está particionado en manzanas y estratificado en 5 grupos de tamaño<sup>33</sup>, que reflejan el tamaño de las manzanas en el marco. La selección de las manzanas se realiza en cada comuna de la muestra y estrato de tamaño de forma independiente, sistemática y con igual probabilidad.

---

<sup>33</sup> Esta partición en 5 grupos de tamaño fue la partición original con la que se formaron los grupos o estratos de tamaño de las encuestas del INE. Posteriormente, se subdividieron en 30 subgrupos de tamaño para facilitar la tarea del encuestador en el sentido de cuántas viviendas debía levantar en cada manzana, considerando el 25% del total de viviendas de la manzana, siempre con un tope de 30.

La información de las manzanas agrupadas en 5 grupos de tamaño a nivel del marco, se cruza con las áreas urbanas de cada comuna, de modo que en cada comuna-área pueden existir a lo más 5 grupos ( $i = 1, 2, \dots, z$  ; donde  $z \leq 5$ ).

Posteriormente, se cuenta el total de viviendas que contiene cada grupo de tamaño, según información del marco y se determina la proporción de viviendas del total que contiene cada uno de estos grupos. De acuerdo a esta proporción se asigna el total de viviendas a seleccionar en cada grupo de tamaño. Como cada grupo de tamaño tiene un número fijo<sup>34</sup> de viviendas a seleccionar por conglomerado, se calcula el total de manzanas a seleccionar, dividiendo la cantidad de viviendas asignada de acuerdo a la proporción, por este número fijo de viviendas a encuestar por conglomerado<sup>35</sup>.

Una vez asignado el total de viviendas y manzanas por grupo de tamaño, la manzana se elige con igual probabilidad y de forma sistemática. De acuerdo al número total de viviendas que tiene la manzana en el marco, cada una de éstas tiene predeterminado el total de viviendas que pueden ser seleccionadas para una muestra, tal como se muestra en la Tabla III.8.

Determinado el número de manzanas a seleccionar en cada grupo de tamaño, la selección se hace en un paso:

1. Se selecciona la cantidad de manzanas determinadas previamente, con igual probabilidad dentro de cada grupo.

**Tabla III.8: Grupos de tamaño de manzanas y total de viviendas a encuestar según grupo**

Grupo Tamaño	Manzanas con Número de viviendas entre		Viviendas a Encuestar (aprox. 25% con tope de 30)		Tamaño fijo de viviendas a encuestar por manzana dentro del grupo de tamaño
	Mínimo	Máximo	Mínimo	Máximo	
1	8	23	2	7	4
2	24	44	5	13	8
3	45	81	9	24	14
4	82	154	19	31	26
5	157	931	28	31	30

Fuente: Elaborado por el Instituto Nacional de Estadísticas

<sup>34</sup> Este número fijo corresponde al promedio ponderado de viviendas a encuestar por conglomerado dentro del grupo y se calcula primero sumando el número de viviendas a encuestar por conglomerado (aproximadamente igual al 25% de las viviendas disponibles de la manzana, con un tope de treinta) y segundo, se divide esta suma ponderada por el total de manzanas que contiene el grupo.

<sup>35</sup> Por ejemplo, si en el grupo de tamaño 2 de la Tabla III.12 corresponde levantar 90 viviendas con sobremuestreo, de acuerdo a la distribución según proporción de viviendas que contiene este grupo de tamaño, el número de manzanas a seleccionar en este grupo, será aproximadamente de  $90/8 = 11,25 \approx 11$  manzanas.

### 5.1.2. Mejora metodológica en el proceso de selección de conglomerados en Casen 2015

El cambio en el método de selección de los conglomerados en el marco de manzanas, se debe principalmente a una optimización de las estimaciones por un lado y a la minimización de la varianza por otro, que se enmarca en el proceso de mejoras continuas que realiza el INE al conjunto de sus productos.

En cuanto a la optimización de las estimaciones, en Casen 2015, agregar una estratificación adicional como los grupos de tamaño y distribuir la muestra en ellos, asegura que estén mejor representados todos los conglomerados de distinto tamaño. A diferencia del método utilizado en Casen 2011 y 2013, en que, en promedio, se seleccionaban conglomerados de mayor tamaño, restando cierta dispersión geográfica<sup>36</sup>. Adicionalmente, en Casen 2015, con la implementación de este método, las probabilidades de selección se pueden obtener directamente mediante expresión directa o explícita, en contraste con Casen 2011 y 2013, en las que no se contaba con probabilidades explícitas, sino que debieron ser generadas mediante un método indirecto de remuestreo, demandando mayores recursos en tiempo para su estimación. Además, con este método de remuestreo se obtienen aproximaciones a las probabilidades de selección, mientras que con el nuevo método se obtienen las probabilidades reales con mayor precisión.

En cuanto a la minimización de la varianza de las estimaciones, en Casen 2015 se logra desde dos aristas. La primera es que la estratificación por sí sola y la selección con igual probabilidad al interior de estos estratos de tamaño, generan probabilidades de selección de los conglomerados más homogéneas que en Casen 2011 y 2013, y por tanto los factores de expansión, como inversos de estas probabilidades, también son más homogéneos, es decir, tienen menor variabilidad. La segunda arista tiene que ver con que los grupos de tamaño en Casen 2015 son cinco y operan como estratos, a diferencia de Casen 2011 y 2013, que ascienden a 30 y operan como una etapa más de muestreo<sup>37</sup>. En consecuencia, la varianza de los estimadores en promedio es mayor.

## 5.2. Selección de conglomerados desde el MS2002

El marco muestral de secciones está particionado en 6.650 secciones en la parte Rural (R) y 1.135<sup>38</sup> secciones en la parte RAU. La selección preliminar de las secciones se realiza al interior de cada comuna en forma proporcional al total de viviendas ocupadas que éstas contienen, según información del Censo

---

<sup>36</sup> Efectivamente y poniéndose en situaciones extremas para entenderlo, un tamaño de viviendas a seleccionar fijo en un estrato geográfico queda distribuido en pocas manzanas si la mayoría son grandes; y queda distribuido en muchas manzanas si la mayoría son pequeñas. En este contexto, si hay pocas manzanas, es más difícil que cubran o queden dispersas en más territorio distinto que si hubiera muchas y, por tanto, pocas manzanas (de mayor tamaño) presentan menor dispersión geográfica.

<sup>37</sup> Cada etapa de muestreo adicional aumenta la varianza de los estimadores porque aumenta el efecto del diseño. Casen 2011 y 2013 tienen una etapa más de muestreo, que tiende a aumentar la varianza de las estimaciones; mientras que Casen 2015 tiene una estratificación adicional, que tiende a disminuir la varianza.

<sup>38</sup> Previo a la selección se descartan las secciones de la Encuesta Nacional de Empleo, por ello el total de secciones difiere entre las versiones de la encuesta Casen 2011, 2013 y 2015.

2002. Esta selección se realiza en forma sistemática proporcional al tamaño<sup>39</sup>, mediante un algoritmo de acumulación de viviendas en el estrato o comuna, en el que se elige un período y arranque aleatorio, como sigue:

### **Paso 1**

Se construyen Q intervalos (equivalente al número de secciones del Estrato o comuna) de la manera como se muestra en la Tabla III.g:

**Tabla III.g. Método de selección de unidades primarias secciones.**

Sección	Nº Viviendas	Acumulado del número de viviendas	Límite inferior	Límite superior
<i>i</i>	$M_i$	$\Sigma M_i$	$L_{inf}$	$L_{sup}$
1	$M_1$	$\Sigma_1 = M_1$	1	$\Sigma_1$
2	$M_2$	$\Sigma_2 = M_1 + M_2$	$\Sigma_1 + 1$	$\Sigma_2$
3	$M_3$	$\Sigma_3 = M_1 + M_2 + M_3$	$\Sigma_2 + 1$	$\Sigma_3$
⋮	⋮	⋮	⋮	⋮
Q-1	$M_{Q-1}$	$\Sigma_{Q-1} = M_1 + M_2 + \dots + M_{Q-1}$	$\Sigma_{Q-2} + 1$	$\Sigma_{Q-1}$
Q	$M_Q$	$\Sigma_Q = M_1 + M_2 + \dots + M_Q$	$\Sigma_{Q-1} + 1$	$\Sigma_Q$

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

### **Paso 2**

Se calcula el período dado por:  $K = \Sigma Q / n$  con  $\Sigma Q$  siendo el Total de viviendas de la comuna o estrato y n siendo el número de secciones a seleccionar (se supone  $n \leq Q$ , es decir, que el número de secciones a seleccionar es menor o igual al número de secciones de la comuna o estrato).

### **Paso 3**

Se genera un número aleatorio<sup>40</sup> entero A entre 1 y el período K.

### **Paso 4**

Enseguida se va sumando sucesivamente el período K para obtener distintos valores que pertenecerán a los distintos intervalos que indican qué sección es la seleccionada. De acuerdo a este algoritmo, las secciones seleccionadas son aquellas cuyo intervalo asociado contiene a los valores dados por:

$$A, A + K, A + 2K, A + 3K, \dots, A + (n - 1)K.$$

El procedimiento anterior, para secciones homogéneas en cuanto al tamaño (de una cantidad de viviendas similar), no permite repeticiones. Por el contrario, cuando las secciones son de tamaño muy distinto, llegando a ser algunas el doble y más del doble de las demás, el procedimiento o algoritmo

<sup>39</sup> En la práctica, mediante el módulo de muestras complejas de SPSS u otro software estadístico, la selección aleatoria en forma sistemática es muy fácil de implementar, y ciertamente el software aplica el mismo algoritmo de selección.

<sup>40</sup> Para la generación de este número aleatorio, se puede introducir una semilla para conservar la selección de los conglomerados y no cambien en caso que se replique el procedimiento.

selecciona menos de las requeridas, porque esencialmente puede seleccionar dos o más veces una misma sección, si ésta es demasiado grande respecto a las demás. Se puede demostrar que para secciones homogéneas en cuanto al tamaño, la probabilidad de inclusión de una sección con  $M_i$  viviendas está dada por  $n \cdot M_i / \Sigma Q$ .

## 6. Selección de viviendas de la muestra

Como se desprende de la argumentación anterior, la selección de conglomerados (manzanas y secciones) se realiza en una etapa en el marco de manzanas y secciones. Previo a la selección de viviendas, sin embargo, se debe actualizar el registro del total de viviendas y la cartografía asociada a los conglomerados en la muestra (ver Anexos, sección 3, Proceso de Enumeración y Actualización de la Muestra)

A continuación, se detalla cual es el procedimiento de enumeración de los conglomerados y el método de selección de las viviendas.

### 6.1. Enumeración de los conglomerados de la muestra

La selección inicial de conglomerados, ya sea a partir del marco de manzanas o secciones, es de carácter preliminar, debido a que el número final de conglomerados se ajusta en función de los resultados del proceso de actualización que se realiza en terreno. Este proceso es conocido como enumeración o empadronamiento.

La enumeración es una operación de registro de viviendas en un área geográfica específica, que tiene como finalidad identificar el universo en el cual se realizará la selección de viviendas. Para realizar esta tarea, cada enumerador debe cumplir con las etapas y objetivos del referido estudio. La enumeración realizada para la muestra de la encuesta Casen 2015, contempló áreas urbanas y rurales.

Durante el proceso de enumeración, algunos conglomerados todavía pueden estar sujetos a cambios (reemplazos) por diversos motivos:

- el conglomerado puede estar temporalmente destruido porque se van a levantar nuevas edificaciones, ya sean casas nuevas o edificios;
- la mayoría de las viviendas cambiaron de uso o destino a otro que no es de vivienda;
- es imposible ingresar a las viviendas del conglomerado porque están dentro de un edificio que tiene un conserje que les impide el ingreso;
- el conglomerado no tiene el número de viviendas mínimo requerido para la selección de las viviendas;
- la estructura en terreno del conglomerado difiere de la enviada en el croquis, debido a que ésta pudo haberse fusionado con otra o dividido debido al cambio de límites por cierre o construcción de nuevas calles. Este último es el caso más frecuente;

- el conglomerado se encuentra afectado por algún tipo de catástrofe natural como aluviones, terremotos u otros.

En total, los motivos inciden aproximadamente en el remplazo de 3% de las manzanas/secciones originalmente seleccionadas. Para remplazo se busca, dentro del entorno geográfico, un conglomerado con características similares en cuanto al tamaño (número de viviendas) y en cuanto a la ubicación (la misma comuna, distrito y zona, o la misma comuna y distrito, o la misma comuna, dependiendo si al nivel más desagregado se encuentre una de tamaño similar). El remplazo de manzanas/secciones en esta etapa del proceso es práctica habitual en muestras para encuestas de hogares desarrolladas por el INE.

## 6.2. Selección de las Viviendas

Una vez determinadas las manzanas y secciones definitivas de la muestra, se procede a la selección de las viviendas al interior de los respectivos conglomerados.

En el marco de manzanas y secciones, la selección de viviendas se realiza al interior de cada una de las seleccionadas previamente, en forma sistemática, con igual probabilidad. Específicamente, al algoritmo de selección sistemático de viviendas, que es aplicado también en otras encuestas, se introduce una semilla de aleatorización al igual que para la selección de manzanas, para que, independientemente de quien aplique el mismo algoritmo, bajo las mismas condiciones de ordenamiento previo<sup>41</sup>, seleccione las mismas viviendas.

Sean  $M_i$  el número de viviendas que posee la manzana o sección y  $m_i$  el número de viviendas a seleccionar. Para la selección de  $m_i$  viviendas se siguen los siguientes pasos:

### **Paso 1**

Se calcula el período  $K = M_i / m_i$

### **Paso 2**

Se genera un número aleatorio entero  $A$  entre 1 y el período  $K$ .

### **Paso 3**

Enseguida se va sumando sucesivamente el período  $K$  al arranque  $A$  para obtener distintos valores que al redondearlos van generando las sucesivas selecciones:

$A, A + K, A + 2K, A + 3K, \dots, A + (m_i - 1)K.$

---

<sup>41</sup> Si se trata de manzanas, el ordenamiento es comuna-distrito-zona-manzana-orden vivienda. Si se trata de secciones, el ordenamiento es comuna-estrato-sección-manzana-orden vivienda. (se incorpora la variable manzana sólo en áreas urbanas cuyas unidades muestrales sean secciones)

La primera selección  $A$  ya es un número redondeado a entero, la segunda es el redondeo de  $A + K$ , la tercera es el redondeo de  $A + 2K$ , y así sucesivamente hasta la  $m_i$  selección dada por el redondeo de  $A + (m_i - 1)K$ .

## IV. DESARROLLO DE FACTORES DE EXPANSIÓN

### 1. Introducción

Los análisis y estudios basados en encuestas a hogares con diseño muestral probabilístico, para que tengan validez sobre la población objetivo, utilizan un ponderador en la estimación de las variables de interés del estudio, que tiene que ver con las probabilidades de selección de las distintas unidades de muestreo y que da cuenta del número de personas de la población que representa un individuo que participa en dicha encuesta. Este ponderador se conoce como factor de expansión.

Históricamente, para cada encuesta Casen, se han desarrollado dos factores de expansión para cada persona entrevistada: un factor que expande a la proyección de población regional (*expr*) y otro que expande a la población comunal (*expc*)<sup>42</sup>. En esta ocasión, en Casen 2015, dado que se definieron varias comunas que son capitales provinciales como dominios de estudio, se ha incorporado además un factor de expansión a nivel provincial (*expp*).

Al igual que el factor regional, el factor provincial lleva incorporado el “peso” de aquellas comunas que pertenecen al marco como de aquellas que no tienen presencia en él, por ejemplo, las comunas catalogadas como áreas de difícil acceso o ADA.

Los factores de expansión provinciales han sido calculados, para fines de investigación, para todas las provincias que no sean ADA y que tengan todas las comunas presentes en el marco muestral del INE o, para aquellas provincias cuyas comunas faltantes no representen más del 10% de la población de personas total de la provincia, según proyecciones de población utilizadas para el mismo cálculo del factor.

De las 54 Provincias del País, Casen no cuenta con 3 de ellas por estar excluidas del marco del INE debido a que fueron provincias catalogadas como ADA. De las restantes 51 provincias, existen 6 provincias que cuentan con sólo una comuna en la muestra. De éstas, de acuerdo al criterio establecido, se generaron factores de expansión provinciales sólo para Coyhaique, Magallanes y Última Esperanza (ver cuadro IV.1), quedando Capitán Prat, Tierra del Fuego y Paríacota como las únicas provincias del marco a las cuales no se les calcularon factores de expansión.

En cuanto a los factores de expansión comunales, han sido publicados como parte de la base de datos Casen 2015, aquellos correspondientes a las 139 comunas que fueron definidas como dominios de estudio.

Los factores regionales, han sido calculados para todas las regiones.

---

<sup>42</sup> La suma de los factores regionales es levemente mayor a la suma de los factores comunales ( $\sum expr > \sum expc$ ), ya que por construcción el factor regional incluye la población en las comunas no incluidas en la muestra.

**Tabla IV.1. Aplicación de criterio para factibilidad de cálculo de factor de expansión provincial**

Región	Provincia	Tipo	N° Comunas en Provincia		Comuna Única	Proyecciones		% de Comuna en Provincia	Provincia con Factor
			Total	Muestra		Provincial	Comunal		
5	52 Isla de Pascua	ADA	1	0		5.637			
10	104 Palena	ADA	4	0		18.593			
11	111 Coyhaique		2	1	Coyhaique	62.126	61.240	98,6	Si
	113 Capitán Prat		3	1	Cochrane	4.163	2.698	64,8	No
12	121 Magallanes		4	1	Punta Arenas	127.882	126.282	98,7	Si
	122 Antártica Chilena	ADA	2	0		3.076			
	123 Tierra del Fuego		3	1	Porvenir	7.222	5.702	79,0	No
	124 Última Esperanza		2	1	Natales	23.158	21.833	94,3	Si
15	152 Parinacota		2	1	Putre	2.265	1.058	46,7	No

Fuente: Elaboración conjunta entre INE y MDS

La metodología de cálculo de los factores de expansión consiste en la aplicación secuencial de cuatro ponderadores:

1. El Ponderador de Selección, refleja cuantas viviendas en la población representa la vivienda seleccionada en la muestra. En términos generales se calcula como el inverso de la probabilidad de selección de la unidad j (vivienda), condicional a la selección del conglomerado i (sección o manzana) del estrato h (área geográfica urbana o rural de una comuna) en la que se localiza.
2. El Ponderador Corregido por Elegibilidad busca corregir las fallas que llevan a la inclusión en la muestra de viviendas no elegibles y aquellas que terminan clasificadas como de elegibilidad desconocida.
3. El Ponderador Corregido por No Respuesta se aplica con el objetivo de reducir el sesgo asociado a la falla en conseguir la cooperación de los hogares residentes en las viviendas seleccionadas en la muestra. La técnica implementada en Casen 2015 fue la misma utilizada en Casen 2013 y Casen 2011, con la cual se busca desarrollar grupos homogéneos (en relación a responder la encuesta) al interior de los cuales se calcula un ajuste de razón que permite corregir los ponderadores de selección.

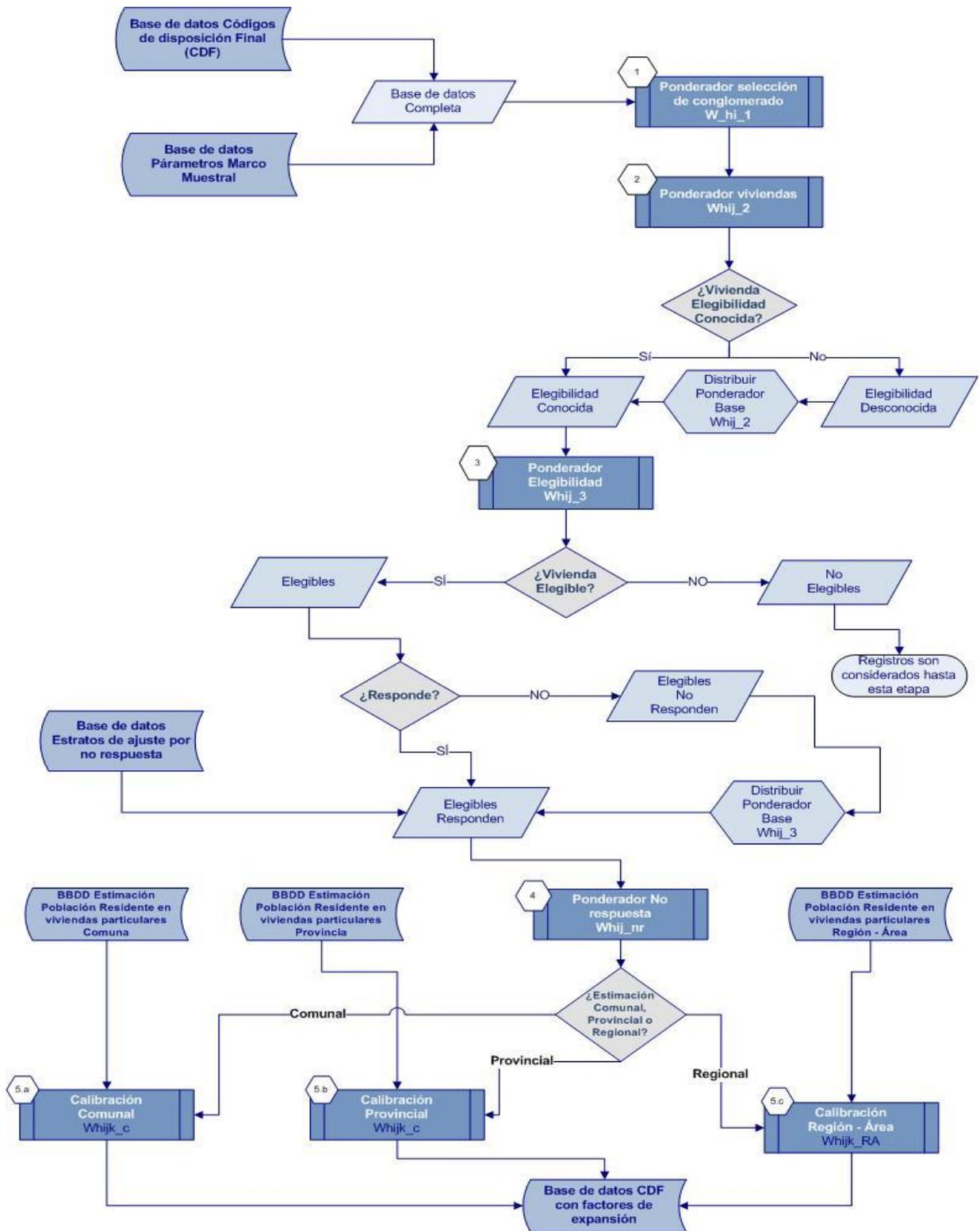
4. Los Ponderadores de Calibración Regional, Provincial y Comunal buscan cuadrar las estimaciones poblacionales obtenidas a partir de la encuesta con las estimaciones poblacionales vigentes publicadas por el INE<sup>43</sup>, mediante la corrección de fallas en la cobertura de subpoblaciones de interés. Este ajuste se conoce también como “corrección de población total”. El ajuste “comunal”, se obtiene como la razón entre el total de personas en la comuna según las proyecciones demográficas del INE y el total de personas residentes en viviendas particulares estimadas para cada comuna a partir de la encuesta. Asimismo, el ajuste “provincial”, se obtiene como la razón entre el total de personas residentes en viviendas particulares estimado para cada provincia a partir de la encuesta y el total de personas en este grupo poblacional según las proyecciones demográficas del INE. Por último, el ajuste “regional”, se realiza separadamente en el área urbana y rural, a través de la razón entre el total de personas estimado a partir de la encuesta y el total de la población residentes en viviendas particulares según las proyecciones de población del INE por región-área.

La Figura IV.1 resume el proceso del cálculo de los factores de expansión para Casen 2015. El capítulo repasa en detalle el proceso de construcción de los factores de expansión de la muestra. Cada apartado hace referencia a las rutinas correspondientes a cada ponderador.

---

<sup>43</sup> Estas proyecciones son las que ocupa el INE en base al Censo 2002 y utilizadas habitualmente en las encuestas propias como la ENE, la ESI, ENUSC, etc. No son las que fueron publicadas en base al Censo 2012 y que nunca se han utilizado en encuesta alguna.

Figura IV.1 Diagrama de procesos de desarrollo de factor de expansión Casen 2015.



Fuente: Instituto Nacional de Estadísticas.

## 2. Ponderador de selección de conglomerados

Casen 2015, posee un diseño probabilístico, estratificado según área geográfica y tamaño poblacional. La muestra de viviendas fue seleccionada a partir de dos marcos muestrales, el Marco de Secciones (MS2002) y el Marco de Manzanas (MM2008), mutuamente excluyentes y de cobertura nacional.

En ambos marcos muestrales la selección de conglomerados, manzanas o secciones, se realizó independientemente en cada estrato geográfico definido por el área (urbano o rural) de una comuna. Para la encuesta Casen 2015 se seleccionaron unidades muestrales en 588<sup>44</sup> estratos de los 610 que posee el INE en sus Marcos de selección.

Dado que el diseño muestral fue realizado en varias etapas y que la probabilidad de seleccionar una unidad de muestreo (vivienda particular) está condicionada por la unidad primaria de muestreo (sección o manzana) que la contiene, se exponen a continuación las probabilidades de selección según el marco muestral de procedencia.

### 2.1. Probabilidad de selección de conglomerados en el Marco de Secciones (MS2002)

Un elemento importante de recordar, es que para la selección de los conglomerados de la muestra Casen 2015 fueron excluidas las secciones ya seleccionadas en la muestra de la Encuesta Nacional de Empleo, cuyo periodo de levantamiento coincidía con Casen 2015. El número de secciones excluidas<sup>45</sup> corresponde a 515 y representan 6,6% del MM2002 (6,5% en cuanto a viviendas del MM2002).

Tras acotar el Marco Muestral de secciones, las unidades muestrales (viviendas particulares) provenientes del MS2002 fueron seleccionadas en dos etapas (se seleccionan las secciones en primera etapa y posteriormente, en segunda etapa, las viviendas al interior de cada sección seleccionada).

En una primera etapa se seleccionaron los conglomerados o “secciones” del MS2002 (Unidades Primarias de Muestreo) con probabilidad proporcional al tamaño (es decir proporcional al número de viviendas particulares que la sección registraba en el Censo de Población y Vivienda 2002). Las secciones fueron seleccionadas todas a la vez de forma sistemática y con probabilidades definidas previamente.

En ese contexto,  $P_h(i)$  la probabilidad de selección de la sección  $i$  del estrato  $h$  es calculada como sigue a continuación,

$$P_h(i) = \frac{n_h M_{hi}}{M_h} \quad (1)$$

---

<sup>44</sup> Se incorporaron los estratos de Putre rural, Atacama rural, La estrella rural, Porvenir rural.

<sup>45</sup> Ver Tabla III.2 de la sección 3 del capítulo III.

Donde:

$i$  Es el índice de la unidad primaria de muestreo o sección.

$h$  Es el índice del estrato de muestreo, o área (urbana o rural) de una comuna.

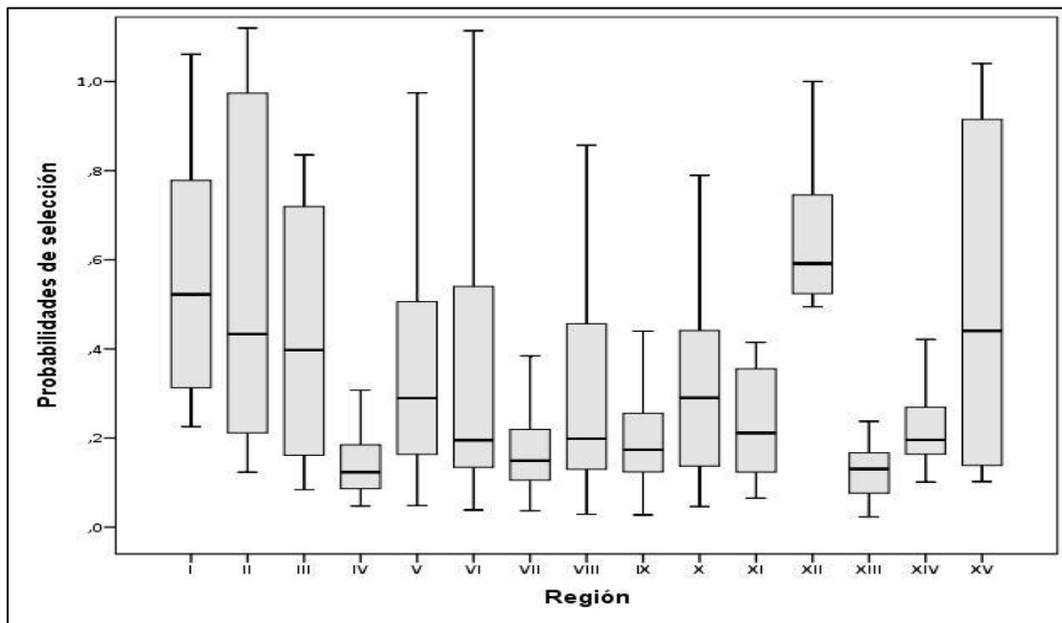
$n_h$  Es el número de secciones seleccionadas del estrato  $h$ .

$M_{hi}$  Es el número de viviendas en la sección  $i$  del estrato  $h$  según el MS2002.

$M_h$  Es el número de viviendas en el estrato  $h$  según el MS2002 previamente acotado para la selección de Casen 2015<sup>46</sup>.

La Figura IV.2 presenta la distribución regional de probabilidades de selección de secciones. La mediana regional es baja (línea negra al medio de cada caja) para las regiones en el centro (en torno a 20%) y relativamente más alta para las regiones en los extremos con menor población.

**Figura IV.2 Distribuciones regionales de las probabilidades de selección de las Secciones, por Región, Casen 2015.**



Fuente: Instituto Nacional de Estadísticas.

## 2.2. Probabilidad de selección de conglomerados en el Marco de Manzanas (MM2008)

Al igual que en el marco de secciones, en el marco de manzanas, para la selección de los conglomerados de la muestra Casen 2015, fueron excluidas las manzanas ya seleccionadas en la muestra de la Encuesta Nacional de Empleo, cuyo periodo de levantamiento coincidía con Casen 2015. El número de manzanas excluidas corresponde a 2.175 y representan el 1,7% del MM2008, actualizado con foco Casen (1,6% en cuanto a viviendas).

<sup>46</sup> Este total de viviendas excluye todas aquellas viviendas pertenecientes a secciones previamente seleccionados para la Encuesta Nacional de Empleo, cuyo periodo de levantamiento coincidía con el de Casen 2015.

A diferencia de la selección en el marco de secciones, en el marco de manzanas, cada estrato geográfico  $h$  dado por la comuna, fue particionado en 5 estratos o grupos de tamaño y la selección se realizó al interior de estos grupos en cada comuna.

En una primera etapa se seleccionaron los conglomerados o “manzanas” del MS2008, actualizado con foco Casen (Unidades Primarias de Muestreo) con igual probabilidad. Las manzanas fueron seleccionadas todas a la vez de forma sistemática dentro de cada grupo de tamaño y con probabilidades definidas previamente.

En ese contexto,  $P_h(i)$  la probabilidad de selección de la manzana  $i$  del grupo de tamaño  $g$ , en el estrato  $h$ , es calculada como sigue a continuación,

$$P_h(i) = \frac{n_{hg}}{N_{hg}} \quad (2)$$

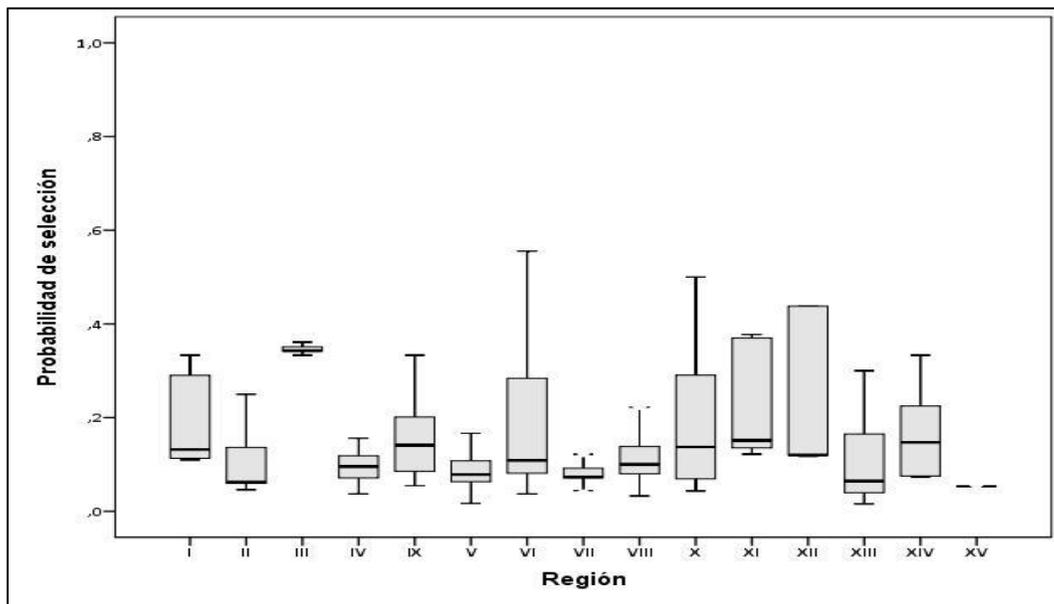
Donde,

$N_{hg}$  Es el número de manzanas en el Marco que posee el grupo de tamaño  $g$ , en la comuna  $h$ .

$n_{hg}$  Es el número de manzanas en la muestra que fueron seleccionadas en el grupo de tamaño  $g$  de la comuna  $h$  ( $g$  puede variar de 1 a 5, ya que el marco está particionado en 5 grupos de tamaño).

La Figura IV.3 presenta la distribución regional de probabilidades de selección de manzanas.

**Figura IV.3 Distribuciones regionales de las probabilidades de selección de las manzanas, por Región, Casen 2015.**



Fuente: Instituto Nacional de Estadísticas.

### 2.3. Ponderador de selección de conglomerados como inverso de la probabilidad de selección

Posterior al cálculo de las probabilidades de selección de los conglomerados (secciones y manzanas), se calculó el Ponderador de Selección de los Conglomerados como el inverso de las probabilidades de selección de secciones o manzanas (según corresponda), tal como se indica a continuación:

$$w_{hi} = \frac{1}{P_h(i)} = \begin{cases} \frac{M_h}{n_h \cdot M_{hi}} & , \text{ Si unidad } i \text{ es una sección} \\ \frac{N_{hg}}{n_{hg}} & , \text{ Si unidad } i \text{ es una manzana} \end{cases} \quad (3)$$

Donde,

$w_{hi}$  Es el ponderador de selección base de la unidad primaria de muestreo o conglomerado (sección o manzana).

$i$  Es el índice de la unidad primaria de muestreo o conglomerado (manzana o sección).

$h$  Es el índice del estrato de muestreo, o área (urbana o rural) de una comuna.

$n_h$  Es el número de secciones seleccionadas en el estrato  $h$ .

$M_{hi}$  Es el número de viviendas en la sección  $i$  del estrato  $h$  según el MS2002.

$M_h$  Es el número de viviendas en el estrato  $h$  según el MS2002 previamente acotado para la selección de Casen 2015.

$N_{hg}$  Es el número de manzanas según el MM2008, en el grupo de tamaño  $g$  del estrato  $h$ .

$n_{hg}$  Es el número de manzanas en la muestra, seleccionadas en el grupo de tamaño  $g$  del estrato  $h$ .

El ponderador de selección de conglomerados puede ser interpretado como “el total de unidades de su tipo que en la población representa”. Por ejemplo, cada manzana presente en la muestra representa a  $w_{hi}$  manzanas de la población. Del mismo modo, la suma de los ponderadores de selección de todas las manzanas en la muestra puede ser interpretada como una estimación del total de manzanas existentes en la población.

Para realizar dichas estimaciones se debe utilizar la base de datos a nivel de manzanas y secciones. Para obtener la estimación del total de manzanas y secciones, basta sumar el ponderador de selección de todas las unidades en la muestra. Mientras que para estimar el total de viviendas de la población, se debe ponderar previamente el total de viviendas por conglomerado según información del Marco de muestreo ( $M_{hi}$ ) por el Ponderador de selección de conglomerados ( $w_{hi}$ ).

Al realizar este ejercicio se obtiene una estimación de 124.009 conglomerados en el total del país (secciones y manzanas).

La Tabla IV.2 muestra la distribución regional de manzanas y secciones en cada Marco Muestral y las correspondientes estimaciones desarrolladas a partir del ponderador de manzanas y secciones. Se observa una subestimación del total de manzanas y secciones en todas las regiones del país, a excepción de Antofagasta en el marco de secciones. Esto se explica porque previa a la selección de la muestra Casen 2015, se acotó el Marco Muestral, excluyendo previamente las secciones y manzanas ya seleccionadas en la muestra de la Encuesta Nacional de Empleo, cuyo periodo de levantamiento coincidía con Casen 2015 y en el marco urbano se excluyeron las manzanas con 7 viviendas y menos, al igual que la Áreas de Difícil Acceso definidas por el INE.

Para detalle de algunos estadísticos de estos ponderadores de selección, ver la tabla IV.3 más adelante.

**Tabla IV.2 Distribución de frecuencias regionales de manzanas y secciones en el Marco Muestral y estimadas en Casen 2015 con ponderador de selección base de selección.**

Región	Marco Muestral		Estimadas en Casen 2015	
	Nº de Manzanas	Nº de Secciones	Nº de Manzanas	Nº de Secciones
<b>Total País</b>	<b>133.759</b>	<b>7.785</b>	<b>116.673</b>	<b>7.336</b>
I Tarapacá	2.295	63	2.042	54
II de Antofagasta	5.698	53	4.874	47
III de Atacama	3.966	106	3.142	96
IV de Coquimbo	6.824	545	5.779	467
V de Valparaíso	19.346	578	15.765	527
VI de O'Higgins	5.789	897	4.874	853
VII del Maule	6.703	1.216	5.916	1.165
VIII del Biobío	16.492	1.336	14.130	1.282
IX de La Araucanía	7.202	1.096	6.187	1.086
X de Los Lagos	5.332	791	4.586	742
XI de Aisén	1.029	129	833	121
XII de Magallanes y La Antártica	2.073	50	1.833	48
XIII Metropolitana	46.017	458	42.451	420
XIV de Los Ríos	2.569	404	2.224	378
XV de Arica y Parinacota	2.424	63	2.037	50

Fuente: Instituto Nacional de Estadísticas.

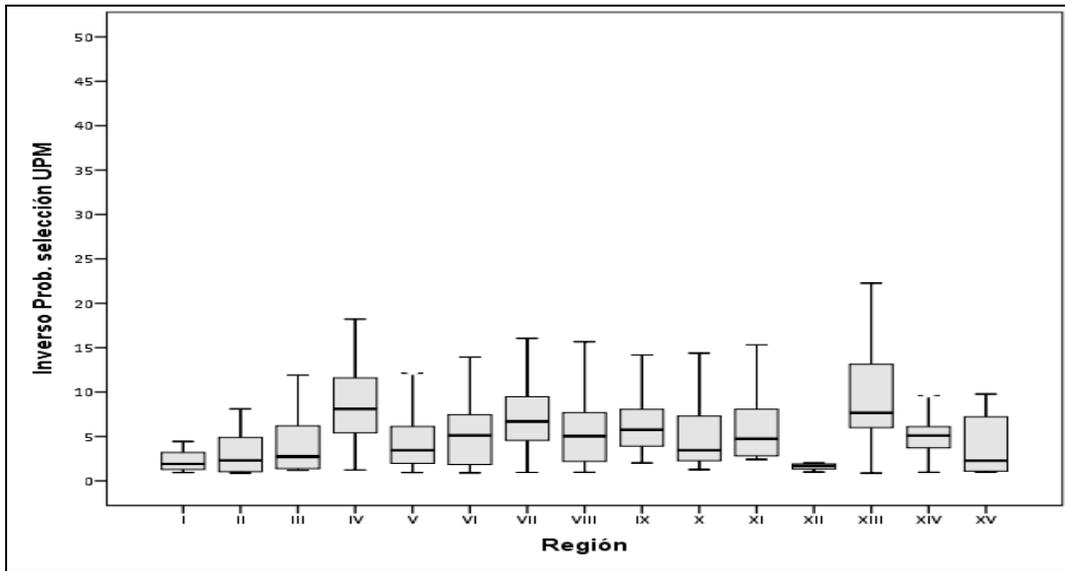
La tabla IV.2 y las Figuras IV.4 y IV.5 presentan las distribuciones regionales de los ponderadores de selección de secciones y manzanas respectivamente, las que presentan diferencias notables: la mediana de los ponderadores de selección de las manzanas es superior al de las secciones entre 1,1 y 8,2 veces, mientras que la media es superior entre 2,1 y 4,8 veces.

**Tabla IV.3 Estadísticos de ponderadores de UPM, manzanas y secciones, según MM2008 y MS2002, Casen 2015.**

Región	Mínimo		Máximo		Media		Mediana		Desviación estándar	
	MM2008	MS2002	MM2008	MS2002	MM2008	MS2002	MM2008	MS2002	MM2008	MS2002
<b>Total País</b>	<b>1,0</b>	<b>0,9</b>	<b>63,0</b>	<b>43,0</b>	<b>11,5</b>	<b>6,0</b>	<b>9,6</b>	<b>5,2</b>	<b>8,2</b>	<b>4,6</b>
I Tarapacá	1,0	0,9	9,1	4,4	6,5	2,2	7,6	1,9	2,9	1,1
II de Antofagasta	2,0	0,9	21,7	8,1	14,0	3,1	16,1	2,3	5,3	2,5
III de Atacama	1,9	1,2	12,0	11,9	4,1	4,4	2,9	2,7	2,6	3,7
IV de Coquimbo	4,0	1,2	27,0	21,0	10,7	8,5	10,5	8,1	2,9	4,6
V de Valparaíso	2,0	0,9	59,5	20,5	13,2	5,0	12,8	3,5	7,5	4,1
VI de O'Higgins	1,0	0,9	27,0	25,8	8,4	5,2	9,2	5,1	4,5	3,7
VII del Maule	3,0	0,9	27,3	27,2	12,8	7,8	13,7	6,7	4,6	4,8
VIII del Biobío	2,5	1,0	30,5	34,6	10,8	6,2	10,0	5,0	5,4	5,3
IX de La Araucanía	3,0	2,0	18,3	36,2	8,6	6,4	7,1	5,8	4,7	3,7
X de Los Lagos	1,2	1,3	23,0	21,7	8,7	5,2	7,3	3,4	5,4	4,0
XI de Aisén	2,7	2,4	8,2	15,3	5,1	6,1	6,7	4,7	2,4	3,7
XII de Magallanes y La Antártica	2,3	1,0	8,5	9,2	5,8	2,3	8,3	1,7	3,0	2,2
XIII Metropolitana	1,9	0,9	63,0	43,0	16,6	10,5	15,5	7,7	11,3	8,5
XIV de Los Ríos	2,0	1,0	13,5	9,9	7,9	5,0	6,8	5,1	3,8	1,8
XV de Arica y Parinacota	6,5	1,0	33,0	9,8	18,4	3,9	18,7	2,3	2,7	3,2

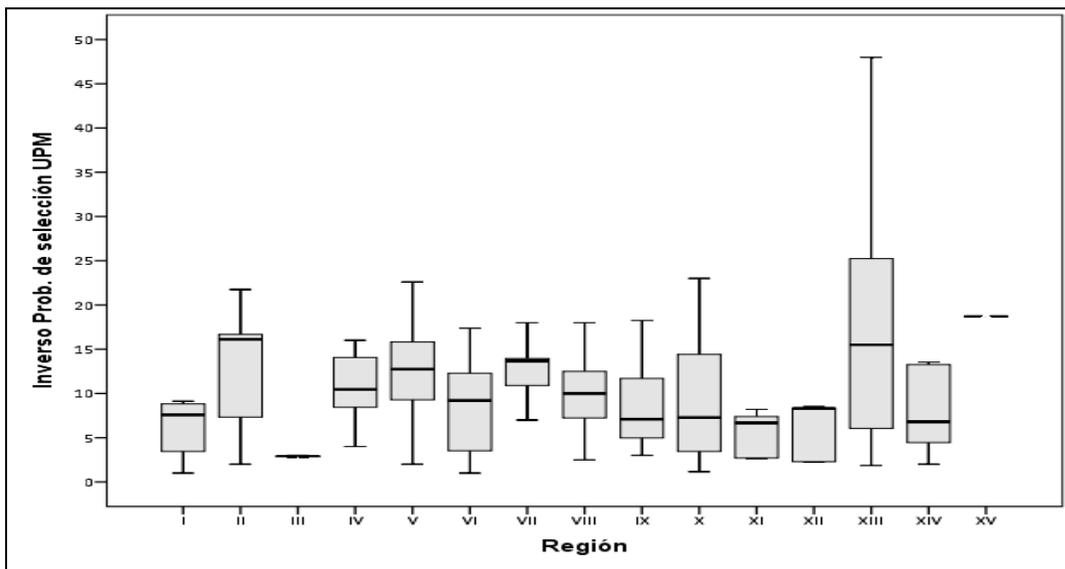
Fuente: Instituto Nacional de Estadísticas.

Figura IV.4 Distribuciones regionales del Ponderador de Selección de Secciones en Casen 2015.



Fuente: Instituto Nacional de Estadísticas.

Figura IV.5 Distribuciones regionales del Ponderador de Selección de Manzanas en Casen 2015.



Fuente: Instituto Nacional de Estadísticas.

## 2.4. Ajuste por omisión de conglomerados

El ajuste por omisión de conglomerados, tiene como finalidad corregir los ponderadores de selección de los conglomerados (manzanas y secciones) según la cantidad de viviendas que se excluyeron por estrato de muestreo en Casen 2015 por motivos operativos.

Como anteriormente, previa selección de la muestra de Casen 2015, se descartaron *a priori* del marco muestral de Manzanas (MM2008) (los conglomerados de tamaño 7 viviendas o menos, y aquellos ya seleccionados para las muestras de otras encuestas).

Análogamente, se excluyeron del marco muestral de secciones (MS2002) las secciones previamente seleccionadas para la muestra de la Encuesta Nacional de Empleo.

El número de manzanas y secciones descartadas *a priori* de ambos marcos muestrales equivale a un total de 160.000 viviendas aproximadamente, lo que corresponde a 3,5% de la población de viviendas pertenecientes a los estratos Casen<sup>47</sup>.

**Tabla IV.4 Total de viviendas según marco de muestreo, con y sin exclusión de conglomerados.**

Total de viviendas MM	Marco de Muestreo		Total Viviendas
	MM 2008	MS 2002	
Marco con exclusión de unidades*	3.944.627	619.018	4.563.645
Marco sin Exclusión de unidades*	4.064.511	659.182	4.723.693

\*unidades del marco con exclusiones

Fuente: Instituto Nacional de Estadísticas.

En este contexto, utilizar el ponderador de selección de los conglomerados sin ajustar por omisión de conglomerado para la estimación de totales de conglomerados y viviendas, resultaría en la subestimación del número total de viviendas por conglomerado. Para evitarlo, se realizó un ajuste, que consistió en aumentar los ponderadores de selección de los conglomerados proporcionalmente a la cantidad de viviendas excluidas por estrato de muestreo. De este modo, la estimación del número de viviendas resultó consistente (4.661.833 a nivel nacional) con el número de viviendas de la población

Para determinar el ajuste  $\hat{R}_h$  a realizar por estrato  $h$  (según el marco muestral que corresponda), se calculó la razón entre  $M_h$ , el número total de viviendas del estrato  $h$  y el número total de viviendas estimadas con el ponderador de selección del estrato  $h$ ,

$$\hat{R}_h = \frac{M_h}{\sum_{i \in \Omega_h} w_{hi} \cdot M_{hi}} \quad (4)$$

Donde,

$M_h$  Es el número de viviendas en el estrato  $h$ .

$\Omega_h$  Es el conjunto de conglomerados seleccionados en el estrato  $h$ .

<sup>47</sup> El total de viviendas en el marco muestral de Casen 2015 es de 4.594.668 viviendas.

- $M_{hi}$  Es el número de viviendas del conglomerado  $i$  del estrato  $h$ .  
 $w_{hi}$  Es el ponderador de selección del conglomerado  $i$  del estrato  $h$ .

Luego, se calcula el Ponderador de Selección de Conglomerados Corregido por Omisión  $w'_{hi}$ , en adelante, "Ponderador de Selección de Conglomerado Corregido" como,

$$w'_{hi} = \hat{R}_h \cdot w_{hi} \quad (5)$$

### 3. Ponderación de selección de viviendas

Los pasos descritos anteriormente están vinculados con la probabilidad de selección de las manzanas y secciones. Sin embargo, la última unidad de selección de la muestra es la vivienda al interior de los conglomerados previamente seleccionados.

En los apartados siguientes se detallan las probabilidades de selección de las viviendas, así como los ajustes y ponderadores asociados.

#### 3.1. Probabilidad Condicional de selección de viviendas

Al interior de cada conglomerado, independientemente del marco de procedencia, las viviendas fueron seleccionadas bajo un algoritmo de selección sistemático, de modo que todas las viviendas particulares ocupadas en el conglomerado, registradas en el empadronamiento, tuvieron la misma probabilidad de ser seleccionadas.

Luego, la probabilidad condicional  $P_{hi}(j|i)$  de incluir la  $j$ -ésima vivienda, dado que (condicional a que) el  $i$ -ésimo conglomerado (manzana o sección) fue seleccionado en la muestra del estrato  $h$ , está dada por

$$P_{hi}(j|i) = \frac{m_{hi}}{M'_{hi}} \quad (6)$$

Donde,

$j$  Es el índice de la vivienda.

$M'_{hi}$  Es el número de viviendas ocupadas en la manzana o sección  $i$ , del estrato  $h$ , según enumeración previa a la recolección de datos.

$m_{hi}$  Es el número de viviendas seleccionadas en la manzana o sección  $i$ , del estrato  $h$ .

Al aplicar la definición de la probabilidad condicional, se puede determinar la probabilidad  $P_{hi}(j)$  de selección de la vivienda  $j$  ubicada en el conglomerado (manzana o sección)  $i$  del estrato  $h$  en la muestra Casen 2015:

$$P_{hi}(j) = P_h(i) \cdot P_{hi}(j|i) \quad (7)$$

Donde,

$P_{hi}(j|i)$  Es la probabilidad condicional de que la vivienda  $j$  pertenezca al conglomerado  $i$  del estrato  $h$ .

$P_h(i)$  Es la probabilidad de que el conglomerado  $i$  (manzana o sección) del estrato  $h$  pertenezca a la muestra.

### 3.2. Ponderador de selección de viviendas como inverso de la probabilidad de selección

El ponderador de selección de viviendas es calculado como el inverso de la probabilidad  $P_{hi}(j)$  de selección de la vivienda  $j$  ubicada en el conglomerado  $i$  (manzana o sección) del estrato  $h$ . Esto es:

$$w_{hij} = \frac{1}{P_h(i) \cdot P_{hi}(j|i)} = \frac{1}{P_h(i)} \cdot \frac{1}{P_{hi}(j|i)} = w_{hi} \cdot \frac{1}{P_{hi}(j|i)} \quad (8)$$

Pero el ponderador de selección de conglomerado  $w_{hi}$  tuvo un ajuste por omisión dado en el punto 2.4 de este capítulo y expresado como el ponderador  $w'_{hi}$ , que desde otro punto de vista, se ajustaron simultáneamente las probabilidades de selección de los conglomerados por la omisión de algunos que fueron descartados a priori por estar en otras muestras o conglomerados urbanos con menos de 7 viviendas. Reescribiendo el ponderador de selección de la vivienda, con la incorporación del ajuste por omisión de conglomerado, la expresión queda:

$$w_{hij} = w'_{hi} \cdot \frac{1}{P_{hi}(j|i)} = w_{hi} \cdot \hat{R}_h \cdot \frac{1}{P_{hi}(j|i)} = w_{hi} \cdot \hat{R}_h \cdot \frac{M'_{hi}}{m_{hi}} \quad (9)$$

Donde,

$h, i, j$  Es el índice del estrato, conglomerado (sección o manzana) y vivienda, respectivamente.

$w_{hij}$  Es el ponderador de selección de la vivienda  $j$  ubicada en el conglomerado  $i$  del estrato  $h$ .

$w'_{hi}$  Es el ponderador de selección del conglomerado  $i$  del estrato  $h$ , corregido por omisión de conglomerados (descartados a priori de los respectivos marcos muestrales).

$P_{hi}(j|i)$  de selección de la vivienda  $j$  ubicada en el conglomerado (manzana o sección)  $i$  del estrato  $h$  en la muestra (definitiva) de Casen 2015 (formula (14)).

$w_{hi}$  Es el ponderador de selección base de la unidad primaria de muestreo o conglomerado (sección o manzana).

$\hat{R}_h$  Es el factor de ajuste por omisión de conglomerados (secciones o manzanas).

$M'_{hi}$  Es el número de viviendas ocupadas en la manzana o sección  $i$ , del estrato  $h$ , según enumeración previa a la recolección de datos.

$m_{hi}$  Es el número de viviendas seleccionadas en la manzana o sección  $i$ , del estrato  $h$ .

Esta expresión, al diferenciarla por el Marco desde el cual fue seleccionada la vivienda, se puede escribir como:

$$w_{hij} = \begin{cases} \frac{M_h}{n_h \cdot M_{hi}} \cdot \hat{R}_h \cdot \frac{M'_{hi}}{m_{hi}} & , \text{ Si unidad } i \text{ es una sección} \\ \frac{N_{hg}}{n_{hg}} \cdot \hat{R}_h \cdot \frac{M'_{hi}}{m_{hi}} & , \text{ Si unidad } i \text{ es una manzana} \end{cases} \quad (10)$$

Donde se agrega la nomenclatura:

- $M_h$  Es el número de viviendas en el estrato  $h$  según el MS2002 previamente acotado para la selección de Casen 2015.
- $n_h$  Es el número de secciones seleccionadas en el estrato  $h$ .
- $M_{hi}$  Es el número de viviendas en la sección  $i$  del estrato  $h$  según el MS2002.
- $N_{hg}$  Es el número de manzanas según el MM2008, en el grupo de tamaño  $g$  del estrato  $h$ .
- $n_{hg}$  Es el número de manzanas en la muestra, seleccionadas en el grupo de tamaño  $g$  del estrato  $h$ .

Este ponderador puede ser interpretado como el número de viviendas en la población que representan las viviendas seleccionadas en la muestra Casen 2015 en la población. Al interior de cada conglomerado, la ponderación es la misma para todas las viviendas.

#### 4. Ponderación por elegibilidad

La población objetivo de la encuesta Casen (y por lo tanto a la que apunta el proceso de inferencia) son las personas que residen en forma habitual en *viviendas particulares ocupadas*. Esta definición es consistente con la definición de los marcos muestrales urbano y rural elaborados por el INE.

Sin embargo, a pesar de todos los esfuerzos que se realicen por mantener marcos muestrales válidos para esta población siempre existe la posibilidad de errores de inclusión (o de exclusión) en el marco.

Por ejemplo, fallas en el proceso de enumeración pueden derivar en el registro de inmuebles “no elegibles” (ej. casas de veraneo, negocios, etc.) en el marco<sup>48</sup>.

Cambios en el uso de la vivienda también pueden ocurrir entre el momento de enumeración y el momento de la entrevista.

Por esta razón, es importante ajustar las probabilidades de selección de las viviendas seleccionadas para incorporar el hecho de que una proporción de las viviendas seleccionadas en la muestra no es elegible y para otras viviendas simplemente se desconoce su elegibilidad.

<sup>48</sup> Ver Proceso de Enumeración y Actualización de la Muestra, capítulo de Anexos, sección 3.

Al término del trabajo de campo, todas las viviendas seleccionadas inicialmente por el INE terminan siendo clasificadas en tres grandes grupos:

1. Elegibles: las edificaciones en que el encuestador pudo determinar que se trataban de viviendas particulares ocupadas (incluye tanto entrevistas como no entrevistas).
2. No elegibles: las edificaciones identificadas como negocios, viviendas colectivas, viviendas deshabitadas, viviendas de veraneo, viviendas destruidas, etc.
3. De elegibilidad desconocida: las edificaciones en que no se pudo determinar su estado. Este es el caso, por ejemplo, de unidades que nunca fueron enviadas a terreno, viviendas a las cuales no se pudo llegar o encontrar, y otros casos similares.

El ponderador de selección de viviendas tiene valores válidos para las viviendas elegibles, no elegibles y de elegibilidad desconocida.

En lo que sigue del proceso sólo se dejarán valores válidos para las viviendas elegibles. A continuación, se describen los ajustes aplicados.

#### 4.1. Ajuste por elegibilidad desconocida

Inicialmente, todas las viviendas tienen un estatus conocido de elegibilidad: son o no viviendas particulares ocupadas. La muestra de viviendas seleccionadas incluye, en principio solo viviendas elegibles (según información disponible con posterioridad de procesos de enumeración y re-enumeración en el caso de Casen 2015).

Durante el trabajo de campo, sin embargo, este estado no se puede comprobar en algunas viviendas, las que terminan clasificadas como de “Elegibilidad Desconocida”.

Las edificaciones consideradas, en principio, como viviendas con Elegibilidad Conocida (elegibles y no elegibles), inicialmente seleccionadas, pero que finalmente terminan siendo clasificadas como edificaciones de “Elegibilidad Desconocida”, son redistribuidas al interior de cada estrato, junto con las viviendas de elegibilidad Conocida (elegibles y no elegibles).

La proporción de los pesos a redistribuir  $\hat{R}_{h,known}$  se calcula como la razón entre dos estimaciones realizadas en base al ponderador de selección de viviendas. El numerador corresponde a la estimación del total de viviendas en la población y el denominador corresponde a la estimación del total de viviendas con elegibilidad conocida (viviendas elegibles y no elegibles). Ambas estimaciones realizadas con el ponderador de selección de viviendas, es decir,

$$\hat{R}_{h,known} = \frac{\sum_{i \in \Omega_h} \sum_{j \in \Theta_i} w_{hij}}{\sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,known}} w_{hij}} \quad (11)$$

Donde,

$\Theta_i$  Es el conjunto de viviendas  $j$  seleccionadas en la manzana o sección  $i$ .

$\Theta_{i,known}$  Es el conjunto de viviendas  $j$  seleccionadas en la en la manzana o sección  $i$  y clasificadas como elegibles o no elegibles.

$\Omega_h$  Es el conjunto de estratos  $h$  (áreas geográficas urbana o rural de las comunas).

$w_{hij}$  Es el ponderador de selección de la vivienda  $j$  de selección de conglomerado  $i$  del estrato  $h$ ,

La razón antes descrita es la misma para todas las unidades muestrales perteneciente a un mismo estrato.

Luego, el ponderador de selección de vivienda corregido por elegibilidad desconocida  $w'_{hij}$  está dado por:

$$w'_{hij} = \hat{R}_{h,known} \cdot w_{hij} \quad (12)$$

Donde  $w_{hij}$  es el ponderador de selección de la vivienda  $j$  desde el conglomerado  $i$  seleccionado del estrato  $h$ .

#### 4.2. Ajuste por no elegibilidad

Originalmente, la muestra definitiva Casen 2015 contemplaba 107.932 viviendas con sobremuestreo. De ese total 1.030 resultaron de "elegibilidad desconocida" una vez finalizado el trabajo de campo. Entre las 106.902 viviendas con elegibilidad conocida, hubo 104.237 que correspondían a viviendas "elegibles" y 2.665 que correspondieron a viviendas "no elegibles" (ver Tabla IV.5).

El objetivo analítico de la Encuesta Casen es producir inferencias hacia la población que reside en viviendas particulares (elegibles). Por lo tanto, a partir de esta etapa no se consideran para fines analíticos aquellas viviendas que no conforman la población objetivo de la encuesta (viviendas no elegibles, tales como oficinas de empresas, viviendas abandonadas, viviendas de veraneo, viviendas demolidas, etc.).

Esto se implementa asignando un valor blanco ("missing"), en el ponderador de selección de viviendas corregido por elegibilidad, a las viviendas con clasificación "no elegible".

Entre las 104.237 viviendas elegibles, hubo 82.120 viviendas para las cuales se logró concretar una o más entrevistas<sup>49</sup> y 22.117 para las cuales no se logró concretar ninguna entrevista.

El siguiente capítulo discute los problemas que esta situación conlleva y el tratamiento correspondiente implementado en la encuesta Casen 2015.

---

<sup>49</sup> Recordar que, bajo el protocolo de trabajo de campo Casen, al interior de una vivienda seleccionada se completan tantas entrevistas como hogares sean identificados en la vivienda y por lo tanto el encuestador debe intentar realizar una entrevista para cada hogar identificado.

**Tabla IV.5 Distribución de viviendas seleccionadas por el INE en Casen 2015 por elegibilidad en el trabajo de campo**

Vivienda Elegible	Frecuencia	Porcentaje
No elegible	2.665	2,5
Elegible	104.237	96,6
Elegibilidad desconocida	1.030	1,0
<b>Total</b>	<b>107.932</b>	<b>100,0</b>

Fuente: Instituto Nacional de Estadísticas.

## 5. Ponderación por no respuesta

En esta etapa sólo son consideradas las viviendas elegibles, pues son éstas las que cumplen con las características necesarias para participar en la encuesta (ser viviendas particulares ocupadas). Sin embargo, aun cuando dichas viviendas cumplen con los atributos necesarios, es posible que los residentes de algunas de ellas no deseen participar.

En Casen 2015 de las 104.237 viviendas elegibles seleccionadas en la muestra, se lograron concretar 82.120 entrevistas. Es decir, la tasa de respuesta de la encuesta fue de 78,8%<sup>50</sup>.

Entonces, ¿qué sucede con la información de esas viviendas elegibles seleccionadas en la muestra y para las cuales no se logró concretar entrevistas? ¿Qué se puede decir de sus residentes? ¿Qué implicancias podría tener la no respuesta sobre las estimaciones?

Utilizar un ponderador de selección sin tomar en cuenta el fenómeno de la no respuesta a la unidad, significa obtener estimaciones representativas solo de aquellas unidades que participan en la encuesta, y por lo tanto una parte de la población no sería cubierta.

Si no hay relación entre el “patrón de no respuesta” y la variable de interés, entonces es posible que la no respuesta sólo afecte la precisión (varianza) de las estimaciones de interés.

Sin embargo, si existe alguna relación entre la no respuesta y la variable de interés, existe riesgo de sesgo en las estimaciones a partir de la muestra.

Por ejemplo, si existe evidencia de que la no respuesta es mayor en una parte de la población con características especiales, por ejemplo, nivel socioeconómico, no ajustar por la no respuesta, podría generar una subestimación (o sobreestimación) de algunos estadísticos de interés (ya que en la encuesta estaría sub-representada o sobre-representada una parte de la población).

<sup>50</sup> Para fines de cálculo de factores de expansión la tasa de respuesta se obtiene como el cociente entre el total de viviendas que responde la encuesta completa, sobre el total de viviendas elegibles, ya que, en el ajuste de los factores antes realizado, el peso de las viviendas con elegibilidad desconocida es distribuido entre las viviendas elegibles y no elegibles. En el caso de obtener la tasa de respuesta general, ésta se obtiene como el cociente entre las viviendas que responde y aquellas viviendas elegibles y de elegibilidad desconocida 78%.

Para minimizar los problemas ocasionados por la ausencia de respuesta a la unidad en la Encuesta Casen 2015, se realiza un ajuste para compensar la pérdida de las unidades, denominado "propensity score".

En general, este método de ajuste consiste en estimar la probabilidad de las viviendas con hogares que responden y aquellas que no responden. Como no se cuenta con información de los residentes de las viviendas donde no se realiza la entrevista, ni tampoco de las características de éstas, sino sólo la información de ubicación (manzana o sección, comuna, provincia, área, etc.), resulta imposible realizar el ajuste a nivel de viviendas, entonces el ajuste se realiza a través de la tasa de respuesta por unidad de muestreo primaria (sección o manzana).

El ajuste consistió, al igual que en Casen 2013, en implementar un modelo de regresión lineal para predecir la participación en la encuesta (tasa de respuesta) de las unidades de muestreo (secciones y manzanas) utilizando un conjunto de variables auxiliares con información de las unidades mismas (total de viviendas según el Marco de muestreo, tipo de unidad - manzana o sección - entre otras), así como también información de la comuna, proveniente de otras encuestas o incluso de registros administrativos, con el objeto de predecir la tasa de respuesta al interior de las unidades, para posteriormente ordenarlas en orden creciente.

Finalmente, todas las unidades de muestreo (secciones y manzanas) fueron agrupadas en veintiles (20 grupos de igual tamaño), siendo las viviendas pertenecientes a la misma unidad asignadas al mismo veintil.

En resumen, la determinación de las celdas de ajuste, también denominados "estratos" de ajuste, se realiza en tres pasos:

1. Selección del modelo de predicción de la tasa de respuesta de manzanas y secciones.
2. Predicción de la tasa de respuesta para cada manzana y sección.
3. Ordenamiento de las tasas predichas, de mayor a menor, y creación de veintiles.

Si el modelo de predicción es correcto, al interior de las 20 celdas de ajuste las viviendas son similares en sus características de interés y, por consiguiente, los entrevistados pueden representar a los no entrevistados.

Bajo este supuesto, los ponderadores asociados a las viviendas elegibles que no respondieron deben ser redistribuidos entre las viviendas elegibles que si respondieron. Para estos efectos, se debe calcular una razón de ajuste.

La determinación de la razón de ajuste viene dada por la ecuación (18) y se realiza al interior de cada veintil utilizando el ponderador de selección de viviendas corregido por elegibilidad.

En el numerador se incluye la estimación del total de viviendas elegibles, y en el denominador se incluye la estimación del total de viviendas elegibles que responde,

$$\hat{R}_{g,R} = \frac{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i,eleg}} w'_{hij}}{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i,eleg,R}} w'_{hij}} \quad (13)$$

Donde,

$g$  Es el índice del grupo para la corrección de no respuesta.

$\Phi_g$  Es el conjunto de manzanas o secciones asignadas al veintil  $g$ .

$\Theta_{g,eleg}$  Es el conjunto de viviendas pertenecientes a la unidad  $i$  y catalogadas como elegibles.

$\Theta_{g,eleg,R}$  Es el conjunto de viviendas pertenecientes a la unidad  $i$ , catalogadas como elegibles y que responden la encuesta.

$w'_{hij}$  Es el ponderador de selección de viviendas corregido por elegibilidad.

La Tabla IV.6 presenta la composición del ponderador de no respuesta (factor de ajuste) estimado para cada veintil. Aquí se aprecia que el grupo que tiene una mayor ponderación es el primer veintil incrementando los ponderadores de elegibilidad desconocida a más del doble, mientras que el ponderador de las viviendas pertenecientes al grupo 20 son los que tienen menor impacto por la corrección de no respuesta, cercano a 1.

Tabla IV.6 Ponderador de no respuesta por celda de ajuste.

Celda de ajuste de no respuesta o Veintiles	Componentes del ponderador de corrección por no respuesta			
	ponderador de selección de viviendas	Viviendas Seleccionadas o elegibles	Viviendas Responden	$\hat{R}_{g,R}$
Total	22.117	104.237	82.120	
1	5.124	8.009	2.885	2,4
2	3.023	6.014	2.991	1,9
3	1.782	4.267	2.485	1,7
4	1.551	5.033	3.482	1,4
5	1.411	5.321	3.910	1,4
6	1.308	5.782	4.474	1,3
7	817	4.080	3.263	1,2
8	1024	5.900	4.876	1,2
9	820	4.494	3.674	1,2
10	723	4.281	3.558	1,2
11	621	4.023	3.402	1,2
12	605	4.634	4.029	1,1
13	547	4.020	3.473	1,2
14	514	4.187	3.673	1,1
15	474	4.393	3.919	1,1
16	389	3.814	3.425	1,1
17	379	4.708	4.329	1,1
18	336	5.222	4.886	1,1
19	286	6.038	5.752	1,1
20	383	10.017	9.634	1,0

Fuente: Ventiles desarrollado por MDS.

Elaboración: INE

Una vez estimados los factores de ajuste, el ponderador de selección corregido por no respuesta  $w_{hij}^{NR}$  puede ser expresado como,

$$w_{hij}^{NR} = \hat{R}_{g,R} \cdot w'_{hij} \quad (14)$$

Este ponderador se asigna a las viviendas elegibles entrevistadas. A las viviendas elegibles no entrevistadas se asigna un valor blanco ("missing").

## 6. Ponderación de calibración

La muestra Casen es una muestra de viviendas y, hasta este punto, los factores de expansión calculados sirven para llevar estas unidades muestrales a sus respectivos stocks poblacionales (totales de viviendas). Sin embargo, como las principales unidades de interés analítico de la encuesta son las

personas que residen en las viviendas seleccionadas se requiere incorporar información acerca de este stock poblacional a los factores de expansión.

Con este objetivo, el INE preparó proyecciones poblacionales, estimadas en base al Censo 2002<sup>51</sup>, al 30 de noviembre 2015 para todas las comunas de la muestra y además para las 30 divisiones dadas por el cruce de región y área (urbano, rural) (ver Tabla IV.7). Estas proyecciones fueron desarrolladas a partir de modelos demográficos de toda la población, sin diferenciar entre aquellos que residen en viviendas particulares y aquellos que se encuentran en viviendas colectivas.

**Tabla IV.7 Proyecciones de población en base al Censo 2002, proyectadas al 30 de noviembre de 2015**

Región	Proyecciones de población Total			Proyecciones de población residentes en viviendas particulares		
	Urbano	Rural	Total	Urbano	Rural	Total
<b>Total País</b>	<b>15.609.007</b>	<b>2.313.183</b>	<b>17.922.190</b>	<b>15.322.580</b>	<b>2.232.286</b>	<b>17.554.866</b>
I Tarapacá	326.878	26.536	353.414	311.110	15.949	327.059
II de Antofagasta	593.716	16.060	609.776	565.050	9.323	574.373
III de Atacama	263.360	28.120	291.480	258.510	22.778	281.288
IV de Coquimbo	630.329	143.499	773.828	618.429	139.357	757.786
V de Valparaíso	1.699.708	157.995	1.857.703	1.670.686	154.042	1.824.728
VI de O'Higgins	660.894	267.706	928.600	650.934	263.846	914.780
VII del Maule	709.597	340.859	1.050.456	696.282	336.638	1.032.920
VIII del Biobío	1.770.456	333.306	2.103.762	1.740.411	327.084	2.067.495
IX de La Araucanía	688.945	324.438	1.013.383	666.270	317.671	983.941
X de Los Lagos	634.317	257.837	892.154	613.029	250.563	863.592
XI de Aisén	94.811	15.508	110.319	90.004	13.261	103.265
XII de Magallanes y La Antártica	150.327	11.011	161.338	144.382	6.260	150.642
XIII Metropolitana	6.966.422	249.549	7.215.971	6.892.417	243.132	7.135.549
XIV de Los Ríos	264.204	120.662	384.866	253.295	116.304	369.599
XV de Arica y Parinacota	155.043	20.097	175.140	151.771	16.078	167.849

Fuente: Instituto Nacional de Estadísticas.

Como la población de inferencia de la Encuesta Casen está restringida a las personas que residen en viviendas particulares ocupadas, las proyecciones de población elaboradas por INE se ajustan por un factor que representa la proporción de personas que reside en viviendas particulares sobre el total de la población. Este factor se estima, para cada estrato, en base a los resultados del Censo 2002.

Por ejemplo, la proyección de población nacional al 30 de noviembre de 2015 era de 17.922.190 personas<sup>52</sup>. Según el Censo 2002, la proporción de personas residentes en viviendas particulares era de

<sup>51</sup> Estas proyecciones, tal como se indicó en una nota anterior, son las utilizadas actualmente por el INE en todas las encuestas de hogares y están desarrolladas en base al Censo 2002 e incluye información de estadísticas vitales y migratorias también del año 2002. No incluye información más reciente o actualizada.

<sup>52</sup> Con fines analíticos, el departamento de Demografía del INE genera proyecciones poblacionales regionales por sexo y por edad y comunales por zona al 30 de noviembre del año correspondiente a la encuesta Casen, que envía al Ministerio.

97% aproximadamente, por lo cual se estima que el total de la población que reside en viviendas particulares es 17.554.866 personas<sup>53</sup>.

A continuación, se describen los procedimientos para la última etapa de desarrollo de los factores de expansión de Casen 2015. Primero se describe el desarrollo del Ponderador de Calibración Regional, Comunal y Provincial.

### 6.1. Ponderador de Calibración Regional

El Ponderador de Calibración Regional, también conocido como “Factor de Expansión Regional” (*expr*), se asocia a cada persona en la muestra. Esto, a diferencia de los ponderadores anteriores, que están asociados a las viviendas en la muestra. El procedimiento de cálculo de este ponderador puede ser resumido en cuatro pasos:

1. Ajustar las proyecciones de población y estimar el total de personas en viviendas particulares. Se solicitó al INE las proyecciones de población para los 30 estratos formados por el cruce de región y área (urbano, rural). Las proyecciones fueron ajustadas, según la proporción de personas residentes en viviendas particulares reportadas por el Censo de Población y Vivienda del año 2002. Este ajuste fue realizado por comuna-área y luego se sumó el total de población en la región y área, utilizando la información disponible en todas las comunas de la región (incluyendo comunas con y sin presencia en la muestra Casen 2015). De este modo se estimó el total de personas residentes en viviendas particulares ( $N_{RA}$ ) en los 30 estratos formados por el cruce de región-área.
2. Estimar el total de personas en hogares particulares a partir de Casen 2015. En cada vivienda entrevistada se multiplicó el ponderador corregido por no respuesta por el número de personas en la vivienda para obtener una estimación del total de personas residentes en viviendas particulares en cada región-área.
3. Construir el ajuste a población total. Se construyó la razón  $\hat{R}_{RA}$ , a partir del cociente entre la estimación del total de personas residentes en viviendas particulares desarrollado a partir de las proyecciones de población ( $N_{RA}$ , en paso 1) y la estimación de la Encuesta Casen 2015 (paso 2).

$$\hat{R}_{RA} = \frac{N_{RA}}{\sum_{h \in RA} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,elg,R}} N_{hij} \cdot w_{hij}^{NR}} \quad R = 1, \dots, 15; \quad A = \text{Urbano, Rural} \quad (15)$$

Donde,

$N_{RA}$  Es el número de personas residentes en viviendas particulares, en la región  $R$  y área  $A$ .

$N_{hij}$  Es el número de personas residentes en la vivienda  $j$ , del conglomerado  $i$ , en el estrato  $h$ .

<sup>53</sup> Las proyecciones de población de residentes en viviendas particulares, se obtienen realizando el ajuste por un factor que representa la proporción de personas que residen en viviendas particulares sobre el total de la población del Censo 2002.

4. Construir el Ponderador de Calibración Regional. Finalmente, se multiplica el Ponderador corregido por No Respuesta con el ajuste de población total (paso 3).

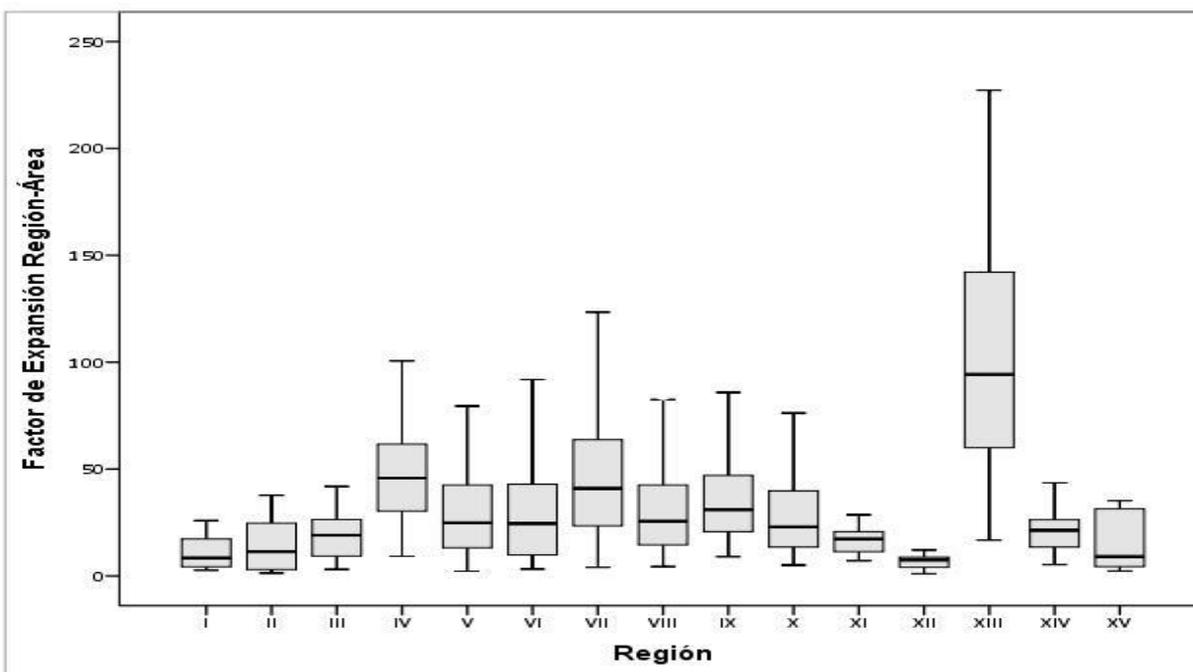
El factor de expansión regional tiene la siguiente expresión:

$$w_{hijk}^{RA} = \hat{R}_{RA} \cdot w_{hij}^{NR} \quad (16)$$

Las Figuras IV.6a y IV.6b muestran la distribución regional de los ponderadores de calibración para el marco rural y urbano respectivamente.

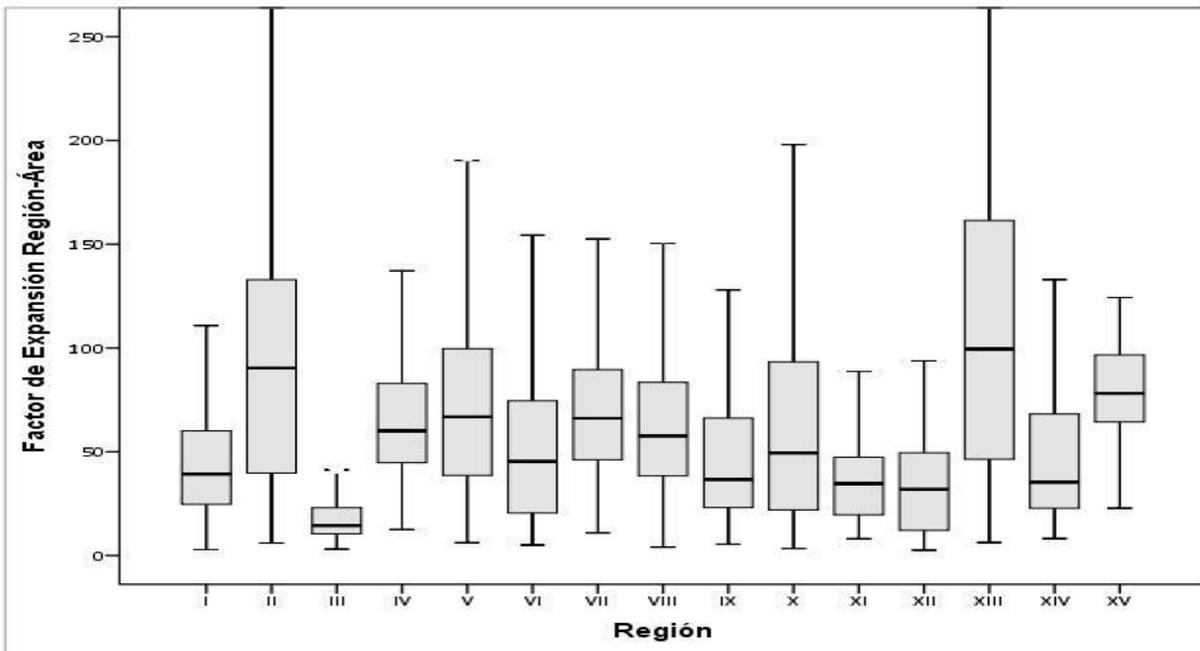
Se puede apreciar que los patrones observados en ponderadores preliminares se mantienen y que los mayores ponderadores se concentran en la región Metropolitana (XIII) (tanto para el marco urbano como rural). Además, se observa que en general, las unidades provenientes del MM2008 son las que poseen mayor variabilidad, lo que se explica, en parte, porque en la parte urbana se concentra gran parte de la población total, así como también en ellas existe gran diversidad en estructura de sus unidades (manzanas grandes y chicas).

**Figura IV.6a Factor de expansión calibración Región-área MS2002**



Fuente: Instituto Nacional de Estadísticas.

Figura IV.6b Factor de expansión calibración Región-área MM2008



Fuente: Instituto Nacional de Estadísticas.

## 6.2. Ponderador de Calibración Comunal

El Ponderador de Calibración Comunal, también conocido como "Factor de Expansión Comunal" ( $exp_c$ ), también se asocia a cada persona en la muestra. El procedimiento de cálculo de este ponderador puede ser resumido en cuatro pasos:

1. Ajustar las proyecciones de población y estimar el total de personas en hogares particulares. Se solicitó al INE las proyecciones de población para los 588 estratos formados por la intersección comuna-área en la muestra Casen 2015. Las proyecciones fueron ajustadas, según la proporción de personas residentes en viviendas particulares reportadas por el Censo de Población y Vivienda del año 2002. Este ajuste fue realizado por comuna-área y luego se sumó el total de población en cada comuna. De este modo se estimó el total de personas residentes en viviendas particulares ( $N_c$ ) en las 324 comunas con presencia en la muestra 2015.
2. Estimar el total de personas en hogares particulares a partir de Casen 2015. En cada vivienda entrevistada se multiplicó el Ponderador corregido por no respuesta por el número de personas en la vivienda para obtener una estimación del total de personas residentes en viviendas particulares en cada comuna en la muestra 2015.
3. Construir el ajuste a población total. Se construyó la razón  $\hat{R}_c$ , a partir del cociente entre la estimación del total de personas residentes en viviendas particulares desarrollado a partir de las

proyecciones de población ( $N_{C_i}$  en paso 1) y el número de personas encuestadas en la comuna por la encuesta Casen 2015 (paso 2),

$$\hat{R}_C = \frac{N_C}{\sum_{h \in C} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,elg,R}} N_{hij} \cdot w_{hij}^{NR}} \quad (17)$$

Donde,

$N_C$  Es el número de personas residentes en viviendas particulares, en la comuna  $C$ .

$N_{hij}$  Es el número de personas encuestadas en la vivienda  $j$ , del conglomerado  $i$ , en el estrato  $h$ .

4. Construir el Ponderador de Calibración comunal. Finalmente se multiplica el Ponderador corregido por no respuesta con el ajuste de población total (paso3).

El factor de expansión comunal tiene la siguiente expresión:

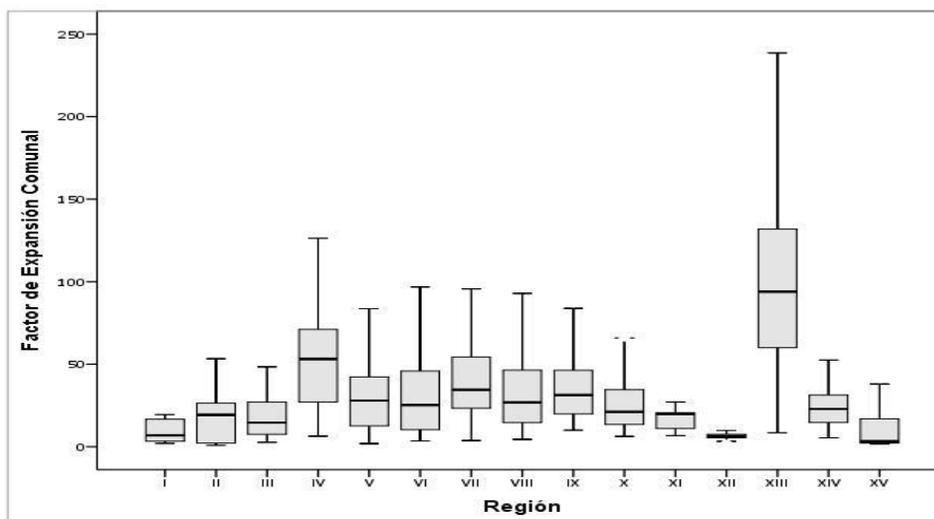
$$w_{hijk}^C = \hat{R}_C \cdot w_{hij}^{NR} \quad (18)$$

Las Figuras IV.7a y IV.7b muestran la distribución regional de los ponderadores de calibración comunal para las unidades en los marcos rural y urbano, respectivamente.

De la comparación entre las Figuras IV.7a y IV.7b se desprende que las unidades provenientes del marco de manzanas, en general, poseen mayor variabilidad, destacándose nuevamente la Región Metropolitana (XIII) con la mayor variabilidad, seguida de la Región de Atacama (III).

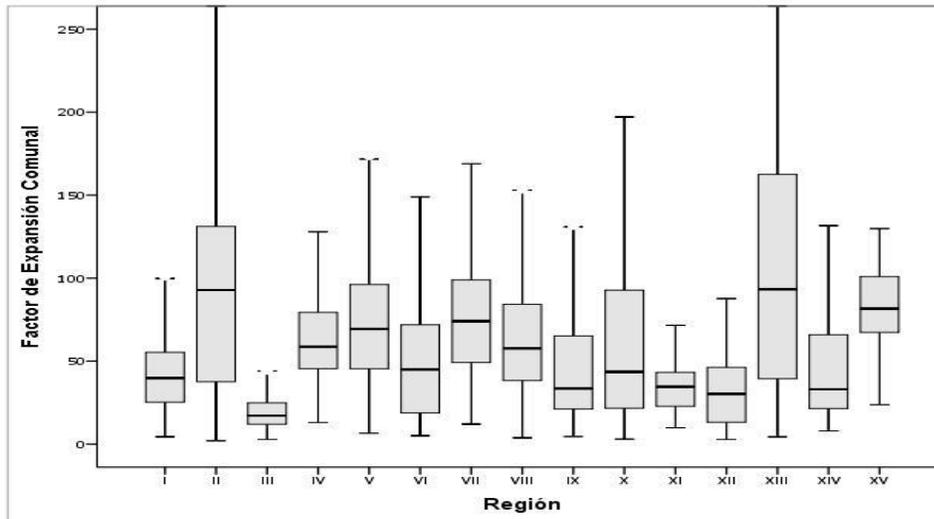
La región de Magallanes (XII) en el Marco de secciones posee los ponderadores más pequeños y con menor variabilidad, mientras que en el Marco de manzanas, la Región de Antofagasta (II) posee los factores más pequeños y de menor variabilidad.

**Figura IV.7a Distribución regional del Ponderador de Calibración Comunal de secciones (MS2002).**



Fuente: Instituto Nacional de Estadísticas.

Figura IV.7b Distribución regional del Ponderador de Calibración Comunal de manzanas (MM2008).



Fuente: Instituto Nacional de Estadísticas.

### 6.3. Ponderador de Calibración Provincial<sup>54</sup>

El Ponderador de Calibración Provincial, denominado “Factor de Expansión Provincial” (*expp*), también se asocia a cada persona en la muestra. El procedimiento de cálculo de este ponderador es similar al de las regiones, pero se obtiene en base a la agregación de todas las comunas que conforman la provincia en vez de la región. El procedimiento puede ser resumido también en cuatro pasos:

1. Ajustar las proyecciones de población y estimar el total de personas en hogares particulares. Las proyecciones fueron ajustadas, según la proporción de personas residentes en viviendas particulares reportadas por el Censo de Población y Vivienda del año 2002, por comuna-área y luego se obtuvo el total de población en cada provincia, a través de la suma sobre todas las comunas que conforman la provincia. De este modo se estimó el total de personas residentes en viviendas particulares ( $N_p$ ) en cada provincia con presencia en la muestra 2015.
2. Estimar el total de personas en hogares particulares a partir de Casen 2015. En cada vivienda entrevistada se multiplicó el Ponderador corregido por no respuesta por el número de personas en la vivienda para obtener una estimación del total de personas residentes en viviendas particulares en cada comuna en la muestra 2015. Enseguida se agregó a nivel de cada provincia
3. Construir el ajuste a población total. Se construyó la razón  $\hat{R}_p$ , a partir del cociente entre la estimación del total de personas residentes en viviendas particulares desarrollado a partir de las

<sup>54</sup> El criterio para desarrollar el factor provincial, ya descrito en la introducción de este capítulo es: “Los factores de expansión provinciales serán calculados para todas las provincias que no sean ADA y que tengan todas las comunas presentes en el marco muestral del INE o, para aquellas provincias cuyas comunas faltantes no representen más del 10% de la población de personas total de la provincia, según proyecciones de población utilizadas para el mismo cálculo del factor”.

proyecciones de población ( $N_P$ , en paso 1) y el número de personas encuestadas en la Provincia por la encuesta Casen 2015 (paso 2),

$$\hat{R}_P = \frac{N_P}{\sum_{h \in P} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,elg,R}} N_{hij} \cdot w_{hij}^{NR}} \quad (19)$$

Donde,

$N_P$  Es el número de personas residentes en viviendas particulares, en la provincia  $P$ .

$N_{hij}$  Es el número de personas encuestadas en la vivienda  $j$ , del conglomerado  $i$ , en el estrato  $h$ .

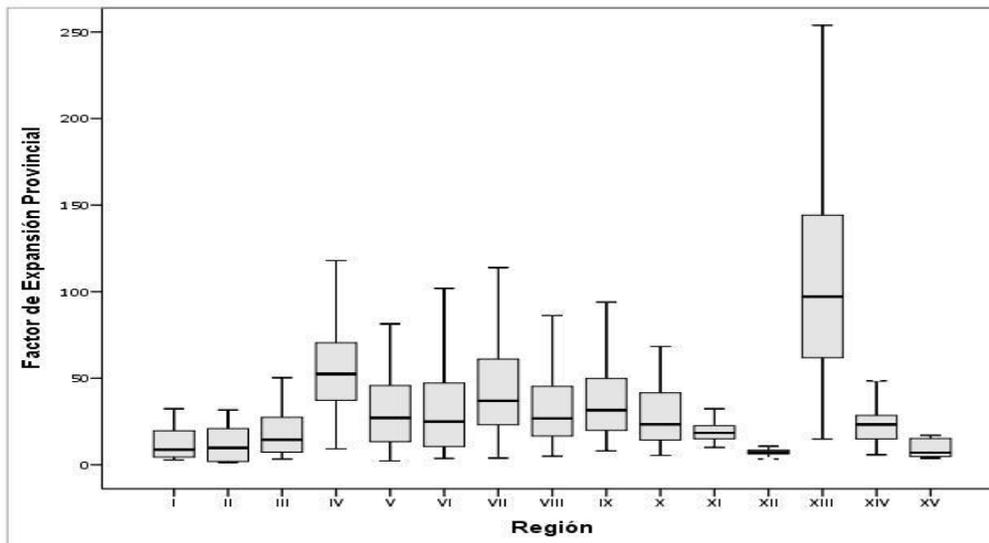
4. Construir el Ponderador de Calibración provincial. Finalmente se multiplica el Ponderador corregido por no respuesta con el ajuste de población total (paso3).

El factor de expansión provincial tiene la siguiente expresión:

$$w_{hijk}^P = \hat{R}_P \cdot w_{hij}^{NR} \quad (20)$$

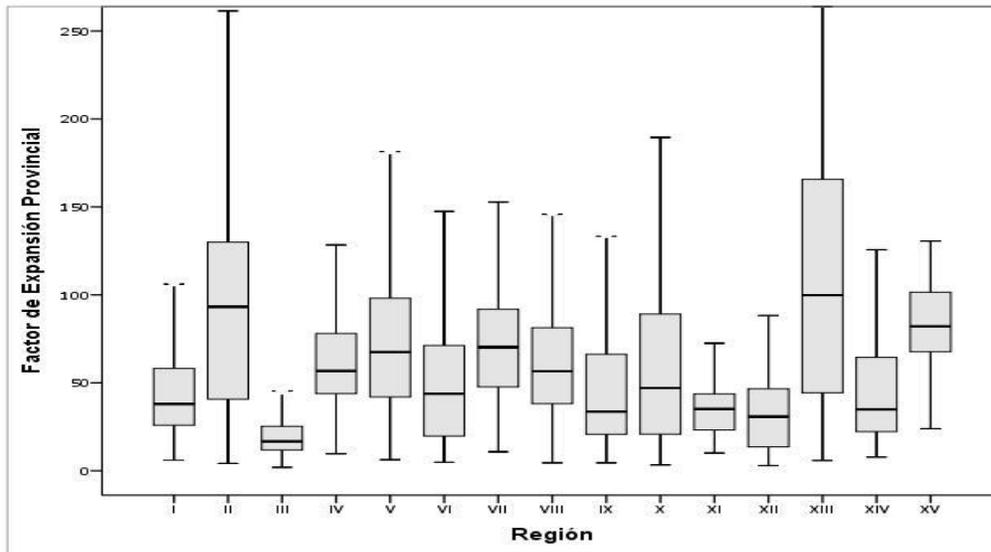
Las Figuras IV.8a y IV.8b muestran la distribución regional de los ponderadores de calibración provincial para las unidades en los marcos rural y urbano, respectivamente.

**Figura IV.8a Distribución regional del Ponderador de Calibración Provincial de secciones (MS2002).**



Fuente: Instituto Nacional de Estadísticas.

Figura IV.8b Distribución regional del Ponderador de Calibración Provincial de manzanas (MM2008).



Fuente: Instituto Nacional de Estadísticas.

De la comparación entre las Figuras IV.8a y IV.8b se desprende que las unidades provenientes del marco de manzanas, en general, poseen mayor variabilidad y las distribuciones de los factores tanto en el Marco de Manzanas como en el de Secciones es muy similar, observándose valores de factores de expansión entre 1 y 250 unidades, aproximadamente (gráficamente se excluyen los valores extremos, 5% de los valores más altos del factor de expansión).

## V. ESTIMACIÓN DE VARIANZAS

Los métodos exactos de estimación de la varianza en diseños de muestras estándares, cuando son aplicables, son la mejor forma de estimar la varianza. Sin embargo, los diseños muestrales empleados en la mayoría de las encuestas de hogares como la Casen son más complejos que el muestreo aleatorio simple o estratificado, debido a que las estimaciones de interés puede que no sean funciones lineales simples de los valores observados, por lo que la varianza de muestreo no siempre puede expresarse por una fórmula de forma cerrada como la media de la muestra en el muestreo aleatorio simple o el muestreo estratificado.

Es cada vez más intensivo el uso de software estadísticos, como SAS, SPSS, STATA y otros como R, que incorporan en los módulos los paquetes estadísticos del cálculo de estimaciones, varianzas y coeficientes de variación, incorporando los típicos efectos de diseño cuando se trata de muestreo complejo, por lo general polietápico.

### 1. Métodos convencionales para estimación de la varianza

Existen principalmente cuatro métodos convencionales para estimar las varianzas o errores muestrales para estimaciones basadas en datos de una encuesta<sup>55</sup>:

1. Métodos exactos;
2. Métodos del conglomerado último;
3. Aproximaciones por linealización;
4. Técnicas de replicación;

En general se utiliza una combinación de métodos para el cálculo tanto de estimaciones como de la varianza de ellas. A continuación, pasamos a describir brevemente en qué consisten estos métodos de estimación de la varianza.

#### Métodos exactos

Los métodos exactos de estimación de la varianza en diseños de muestras estándares, cuando son aplicables, son la mejor forma de estimar la varianza. Sin embargo, los diseños muestrales empleados en la mayoría de las encuestas de hogares son más complejos que el muestreo aleatorio simple o estratificado. En segundo lugar, las estimaciones de interés puede que no sean funciones lineales simples de los valores observados, por lo que la varianza de muestreo no

---

<sup>55</sup> Para más información ver Kish y Frankel (1974), Wolter (1985) y Lehtonen y Pahkinen (1995).

siempre puede expresarse por una fórmula de forma cerrada como la media de la muestra en el muestreo aleatorio simple o el muestreo estratificado. Además, los métodos exactos dependen del diseño de la muestra en cada caso, de la estimación de interés y de los procedimientos de ponderación empleados.

### **Método del conglomerado último o final**

El método del conglomerado último para estimar la varianza (véase Hansen, Hurwitz y Madow, 1953, págs. 257-259) se emplea para estimar las varianzas de las estimaciones basadas en una muestra obtenida a partir de un diseño muestral complejo, generalmente polietápico o polifásico. Según este método, el conglomerado último consiste en considerar a la unidad primaria de muestreo UPM como la unidad de información última que conglopera la información de todas las unidades de muestreo de etapas posteriores dentro de ella, como por ejemplo, si la unidad primaria de muestreo corresponde a un conglomerado que contiene unidades secundarias de muestreo como las viviendas, y a su vez estas unidades secundarias de muestreo contienen unidades terciarias como las personas, y se está interesado en alguna variable de interés como la ocupación, entonces cada UPM tendrá asociado el total de ocupados, como la sumatoria sobre todas las personas de todas las viviendas contenidas en la UPM. Las estimaciones de la varianza se calculan utilizando sólo totales entre UPM, sin tener que calcular los componentes de la varianza en cada etapa de selección.

### **Método de linealización**

La linealización es ampliamente utilizada en la práctica porque puede aplicarse a casi todos los diseños muestrales y a cualquier estadística que pueda linealizarse o, lo que es lo mismo, expresarse como una función lineal de las estadísticas comunes como medias o totales, cuyos coeficientes se extraen de derivadas parciales necesarias para la expansión en serie de Taylor.

Se aplica la linealización generalmente para la estimación de razones o cocientes de dos variables las que se desarrollan en una serie de Taylor infinita centrada en el valor esperado (estimado) del numerador y el valor esperado (estimado) del denominador. Seguidamente, el estimador no lineal se somete a aproximación algebraica conservando sólo los primeros términos de la serie infinita de Taylor, lo que produce una expresión algebraica que ya no es una función no lineal, sino una función lineal de los datos muestrales; esto es, el estimador no lineal cociente se ha "linealizado". Ahora es posible obtener directamente la varianza estimada de la función linealizada (incluidos los términos de covarianza pertinentes). Una vez linealizada, la varianza de la estimación no lineal puede aproximarse mediante los métodos exactos descritos anteriormente

(véanse Cochran, 1997, y Lohr, 1999, para consultar información técnica detallada sobre el proceso de linealización ilustrada con ejemplos).

### **Método de replicación**

Este método consiste en tomar submuestras o réplicas de la muestra total. Se estima con la muestra Total la variable de interés  $Y$ , de acuerdo al diseño probabilístico, sea cual fuere. Enseguida se generan réplicas o se seleccionan submuestras de la muestra total, de manera que cada réplica refleje el plan de muestreo y los ajustes y procedimientos de ponderación de la muestra completa, de modo que con cada una de ellas se llegue a la misma estimación que con la muestra completa. Existen varias técnicas para hacerlo y en algunos software ya vienen incluidos los módulos de selección de réplicas, entre los cuales las más utilizadas son los grupos aleatorios; la replicación repetida equilibrada (BRR); la replicación jackknife (JK1, JK2, y JKn); el bootstrap. Por ejemplo, supongamos que se crean  $K$  réplicas a partir de una muestra, cada una con estimaciones  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K$  de un parámetro  $\hat{\theta}$ , y supongamos que la estimación basada en la muestra completa es  $\hat{\theta}_0$ . La estimación de la varianza basada en la replicación viene dada por:  $V(\hat{\theta}) = \frac{1}{c} \sum_{r=1}^K (\hat{\theta}_r - \hat{\theta}_0)^2$ , donde  $c$  es una constante que depende del método de estimación.

Estos métodos de estimación de varianzas son antiguos y pocos Software los incorporaban o desarrollaban. Sin embargo, con el avance tecnológico y de la computación en particular, hoy en día, la mayoría de los softwares estadísticos traen incorporados módulos especiales para el tratamiento de los datos enfocados a muestras complejas, entendiendo por muestra compleja, a aquella que incorpora variables de ponderación, de estratificación y de conglomeración para las estimaciones de los principales estadísticos asociados a alguna variable de interés y también la estimación de sus varianzas.

## **2. Varianza con muestras complejas**

Una aproximación común (WR: *With Replacement*) para la descripción de muchos planes de muestreo complejo, normalmente es utilizada en las encuestas complejas de muestreo polietápico por conglomerados como la Casen. Además, es habitual la estratificación de las UPM de la población antes del muestreo en la primera etapa. Por otra parte, la estratificación de las unidades de la segunda y posteriores etapas (en una UPM seleccionada) puede tener lugar antes del muestreo en estas etapas. Sin embargo, con los métodos aproximados de estimación de varianzas comúnmente utilizados para estos diseños complejos, no es necesario tomar en consideración todas las etapas de muestreo y estratificación. El muestreo complejo en las etapas posteriores queda automáticamente cubierto de manera adecuada con la aproximación "con

reemplazo” para la primera etapa de muestreo que hemos visto más arriba. De hecho, pocos paquetes de software tienen la capacidad de incluir por separado todas las etapas de muestreo en la estimación de varianzas, en los casos en los que no se practica la aproximación con reemplazo en la primera etapa.

La estimación de varianzas por conglomerados finales (EVCF) puede efectuarse indistintamente por linealización mediante el desarrollo en serie de Taylor o con una técnica de replicación. Con este enfoque, las UPM se tratan como si se hubieran seleccionado con reemplazo entre los estratos de la primera etapa. En tal caso, cada elemento o individuo sólo tiene que identificarse por el estrato de la primera etapa y por la UPM (dentro del estrato) de la que se haya seleccionado. A los efectos de la estimación de varianzas no se necesita información sobre las etapas de muestreo por debajo del nivel de las UPM. De este modo la descripción del plan de muestreo se simplifica de forma que se asemeja a un muestreo por conglomerados estratificados en una etapa, esto es, una muestra estratificada de conglomerados finales completamente enumerados. Este enfoque por conglomerados finales produce una buena aproximación para la estimación de la varianza, siempre que sea razonable partir de la hipótesis de una primera etapa con reemplazo.

Esta aproximación común (EVCF) se denomina a veces WR (por las siglas inglesas de la expresión “con reemplazo”) en la bibliografía sobre encuestas, Así, cuando el plan de muestreo se describe como WR sólo se necesitan tres variables del diseño de la encuesta para la estimación de varianzas:

- a. La variable de ponderación de la muestra o factor de expansión (necesaria también para las estimaciones) en Casen 2015 designado expr o expc o expp, según el nivel territorial de expansión: regional, comunal o provincial, respectivamente..
- b. La variable de estratificación (o variable de pseudo-estratificación) en Casen 2015 designado VarStrat.
- c. La variable UPM (o pseudo-conglomerado) en Casen 2015 designada como VarUnit

El empleo de WR para la aproximación al plan de muestreo complejo puede dar lugar a una ligera sobreestimación de las varianzas. Sin embargo, los analistas de datos de encuestas no tienen, por lo general, problemas en aceptar un cierto grado de sobreestimación por la relativa simplicidad de la aproximación mediante WR. No obstante, hay que tener en cuenta que la sobrestimación puede ser apreciable si hay varios estratos en los que el muestreo en la primera etapa se hace sin reemplazo y con grandes fracciones de muestreo. En este caso, puede ser preferible optar por un

software que cuente con la opción de incorporar los factores de corrección por población finita *cpf* o corrección por finitud para la primera etapa.

### 3. Algoritmo de cálculo en muestras complejas

En el cálculo de varianzas con muestras complejas, tanto en STATA como SPSS se utiliza el método de conglomerado último, en que la contribución de la segunda etapa se asume con reemplazo, lo que simplifica las fórmulas de cálculo.

Para el desarrollo de la estimación de las varianzas, es necesario contar con las ponderaciones de cada unidad de la muestra de modo que reflejen la divergencia de las probabilidades de selección (probabilidades desiguales). La ponderación o factor de expansión de una unidad es el recíproco de su probabilidad de selección para incluirla en la muestra.

Si una vivienda se incluye en la muestra con una probabilidad  $p_{hij}$ , su ponderación básica, viene dada por  $w_{hij} = 1/p_{hij}$ , es decir, el inverso de la probabilidad de selección y se denomina factor de expansión teórico, de acuerdo al diseño probabilístico. Este ponderador es igual para todas las viviendas dentro del conglomerado (manzana) y, por lo tanto, para todas las personas dentro de ellas. En la Casen, posteriormente, este factor se ajustó por omisión de conglomerados, elegibilidad desconocida y también por no-respuesta de las viviendas y además se calibró para alcanzar los stocks poblacionales de personas según las proyecciones de población.

Sea  $y_{hij}$  la variable observada en la vivienda  $j$  del conglomerado  $i$  del estrato  $h$ , con ponderador base  $w_{hij}$ . Se utiliza la siguiente nomenclatura para las estimaciones y cálculo de las varianzas:

$z_{hij} = w_{hij} \cdot y_{hij}$  es el valor de la variable observada en la vivienda  $j$  expandida a la población.

$z_{hi} = \sum_{j=1}^{m_{hi}} z_{hij}$  es el valor de la variable observada expandida y agregada en el conglomerado  $i$  del estrato  $h$ .

$\bar{z}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} z_{hi}$  es el promedio de la variable observada expandida en el estrato  $h$ .

$n_h$  es el número de UPM (manzanas o secciones) seleccionadas en el estrato  $h$ , en la muestra Casen 2015.

$N_h$  es el número de UPM (manzanas o secciones) en el estrato  $h$ , contenidas en el marco muestral utilizado en Casen 2015.

$m_{hi}$  es el número de viviendas seleccionadas y que responden dentro de la UPM  $i$  del estrato  $h$ .

$S_h^2 = \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} (z_{hi} - \bar{z}_h)^2$  es la cuasivarianza en el estrato  $h$ , que corresponde a la variabilidad entre los conglomerados de la variable observada expandida en ellos. Asumiendo muestreo con reemplazo, se le agrega el factor de corrección por finitud  $(1 - f_h) = (1 - n_h/N_h)$ . Para muestreo sin reemplazo, se utiliza una expresión más complicada, que se puede estudiar en Yates and Grundy (1953) and Sen (1953).

### 3.1. Algoritmo de cálculo de los estimadores

Para la estimación de totales, medias y razones, ya sea para cada uno de los estratos de muestreo  $h$ , como para los niveles de estimación o dominios de estudio, se utilizan los siguientes algoritmos:

$$\hat{Y}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} z_{hij} = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \quad (1)$$

Es el Total estimado en el estrato  $h$  de la variable de interés  $Y$

$$\hat{Y} = \sum_{h=1}^H \hat{Y}_h \quad (2)$$

Es la estimación total a nivel agregado de la variable  $Y$ , es decir, la estimación agregada sobre todos los estratos y que forman algún nivel de estimación.

$$\hat{N}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \quad (3)$$

Es la estimación del número de unidades en la población en el estrato  $h$ .

$$\hat{N} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \quad (4)$$

Es la estimación del número de unidades en la población en el nivel de estimación requerido, como agregación sobre todos los estratos de muestreo.

$$\hat{\bar{Y}}_h = \frac{1}{\hat{N}_h} \hat{Y}_h \quad (5)$$

Es la media estimada de la variable de interés  $Y$ , en el estrato  $h$ .

$$\hat{\bar{Y}} = \frac{1}{\hat{N}} \hat{Y} \quad (6)$$

Es la media estimada de la variable de interés  $Y$ , en el nivel de estimación requerido.

$$\hat{R}_h = \frac{\hat{Y}_h}{\hat{X}_h} = \frac{\sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij}}{\sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij}} = \frac{\hat{Y}_h}{\hat{X}_h} \quad (7)$$

Es el estimador de razón entre las variables  $Y$  y  $X$ , en el estrato  $h$ , es decir,  $R_h = \frac{Y_h}{X_h} = \frac{\bar{Y}_h}{\bar{X}_h}$

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij}}{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij}} = \frac{\hat{Y}}{\hat{X}} \quad (8)$$

Es el estimador de razón entre las variables  $Y$  y  $X$ , en el nivel de estimación requerido del estudio.

Existen también los estimadores del Total  $Y$  y la media  $\bar{Y}$ , pero a través de los estimadores de razón, esto es, si despejamos ya sea el Total  $Y$  o la media  $\bar{Y}$  de la razón  $R$ , obtenemos los estimadores:

$$\hat{Y}_{R_h} = \hat{R}_h \cdot X_h \quad \text{y también} \quad \hat{Y}_{R_h} = \hat{R}_h \cdot \hat{X}_h \quad (9)$$

Son las estimaciones del Total y la Media de la variable  $Y$  en el estrato  $h$ , obtenidas mediante el estimador de razón

$$\hat{Y}_R = \hat{R} \cdot X \quad \text{y también} \quad \hat{Y}_R = \hat{R} \cdot \hat{X} \quad (10)$$

Son las estimaciones del Total y la Media de la variable  $Y$  a nivel agregado, obtenidas mediante el estimador de razón.

Para las fórmulas (7), (8), (9) y (10), se asume que el Total  $X$  o la media  $\bar{X}$  son conocidas o dadas como fijas<sup>56</sup>. Pero, generalmente se desconoce el valor poblacional y entonces son estimados por  $\hat{X}$  y  $\hat{\bar{X}}$  en la expresión (7), o son estimadas por  $\hat{X}_h$  y  $\hat{\bar{X}}_h$  en la expresión (8).

### 3.2. Algoritmo de cálculo de las varianzas en muestras complejas

Como se ha planteado, el método de conglomerado último asume que la varianza total es debida a la variabilidad entre los conglomerados para la variable de interés (inter-varianza) y el aporte de la etapa secundaria y posterior es despreciable.

Definida la estratificación en los estratos  $h$  en la encuesta Casen, la varianza del estimador del total, se puede expresar como:

$$\hat{v}(\hat{Y}) = \sum_{h=1}^H \hat{v}(\hat{Y}_h) = \sum_{h=1}^H S_h^2 = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (z_{hi} - \bar{z}_h)^2 \quad (11)$$

Desglosando  $z_{hi}$  y  $\bar{z}_h$  por las expresiones definidas previamente, queda:

$$\hat{v}(\hat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( \sum_{j=1}^{m_{hi}} z_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} z_{hi} \right)^2 \quad (12)$$

<sup>56</sup> Por ejemplo, en la calibración a las proyecciones de población de personas, los stocks poblacionales son estimaciones obtenidas mediante modelos demográficos que son consideradas fijas, como los verdaderos valores en la población. Además si la variable es dicotómica,  $X$  coincide con  $N$  como asimismo  $\hat{X}$  coincide con  $\hat{N}$ .

Reemplazando  $z_{hij} = w_{hij} * y_{hij}$  y  $z_{hi} = \sum_{j=1}^{m_{hi}} z_{hij}$ , estas expresiones pueden finalmente ser escritas como:

$$\hat{V}(\hat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right)^2 \quad (13)$$

Finalmente, se muestran las estimaciones de las varianzas de los estimadores de Totales, Medias y Razones.

$$\hat{V}(\hat{Y}_h) = \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right)^2 \quad (14)$$

Es la varianza o cuasivarianza estimada de la variable de interés expandida en el estrato h.

$$V(\hat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right)^2 \quad (15)$$

Es la estimación de la varianza del Total de la variable de interés Y, en el nivel de estimación requerido.

$$V(\hat{Y}_h) = \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \left( \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right) / \hat{N}_h \right]^2 \quad (16)$$

Es la estimación de la varianza en el estrato h, de la Media de la variable de interés Y.

$$V(\hat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \left( \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right) / \hat{N} \right]^2 \quad (17)$$

Es la estimación de la varianza de la Media de la variable de interés Y, en el nivel de estimación requerido.

$$\hat{V}(\hat{R}_h) = \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \left( \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \hat{R}_h \cdot \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij} \right) / \hat{X}_h \right]^2 \quad (18)$$

Es la varianza del estimador de razón entre las variables Y y X, en el estrato h, con:

$$\hat{X}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij}.$$

$$\hat{V}(\hat{R}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \left( \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \hat{R} \cdot \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij} \right) / \hat{X} \right]^2 \quad (19)$$

Es la varianza del estimador de razón entre las variables Y y X, en el nivel de estimación requerido.

La estimación de la varianza para el Total y la Media respectivamente de la variable Y en base a la razón entre Y y X o la razón entre  $\bar{Y}$  y  $\bar{X}$  (ver expresiones (7), (8), (9) y (10) es:

$$\hat{V}(\hat{Y}_{R_h}) = \hat{V}(\hat{R}_h \cdot X_h) \doteq \hat{V}(\hat{R}_h) \cdot \hat{X}_h^2 \quad ; \quad \hat{V}(\hat{Y}_{\hat{R}_h}) = \hat{V}(\hat{R}_h \cdot \bar{X}_h) \doteq \hat{V}(\hat{R}_h) \cdot \hat{X}_h^2 \quad (20)$$

Es la estimación de la varianza del Total y la Media de la variable Y, en el estrato h, mediante el estimador de razón

$$\hat{V}(\hat{Y}_R) = \hat{V}(\hat{R} \cdot X) \doteq \hat{V}(\hat{R}) \cdot \hat{X}^2 \quad ; \quad \hat{V}(\hat{Y}_R) = \hat{V}(\hat{R} \cdot \bar{X}) \doteq \hat{V}(\hat{R}) \cdot \hat{X}^2 \quad (21)$$

Es la estimación de la varianza del Total y la Media de la variable Y a nivel agregado, mediante el estimador de razón,

#### 4. Variables que identifican el diseño muestral complejo

Un diseño de muestreo complejo se origina cuando interviene más de una etapa de selección. Esta selección en más de una etapa suele ser generalmente además estratificada, lo que añade otro grado de complejidad. Bajo la estratificación, en un muestreo bietápico, las unidades muestrales se suelen llamar conglomerados o unidades de conglomeración, que son grupos de unidades de selección de las últimas unidades de muestreo sobre las cuales se obtienen las estimaciones de las variables de interés.

La encuesta Casen 2015, como también sus predecesoras, presenta un diseño muestral que se puede caracterizar como probabilístico, estratificado y polietápico. En el muestreo estratificado la estimación de la varianza total se obtiene mediante la suma de las varianzas estimadas en cada estrato y la varianza en cada estrato se estima a partir de la variabilidad de la característica de interés entre los conglomerados.

En la base de datos de la Encuesta Casen, la variable que identifica los estratos "originales" utilizados en la selección de la muestra se denomina "Estrato" y la variable que identifica los conglomerados "originalmente" seleccionados en la muestra se denomina "Segmento". En total considerando la muestra completa 2015 consideró 588 estratos y 11.274 segmentos (10.049 manzanas y 1.225 secciones).

Para hacer factible la estimación de la varianza en un muestreo estratificado, se deben seleccionar como mínimo 2 conglomerados por estrato. Para evitar estimaciones de varianza volátiles, se recomienda seleccionar un número mínimo de unidades muestrales (ej. viviendas) por conglomerado. En consecuencia, diseños muestrales adecuados consideran al menos dos conglomerados por estrato al momento de selección de la muestra y un número mínimo de unidades por conglomerado. El diseño Casen 2015 estableció, para la muestra completa, un mínimo de 2 conglomerados por estrato y un mínimo de 30 y 50 viviendas por conglomerado para los segmentos rurales (secciones) y urbanos (manzanas) respectivamente.

Ya sea por diseño o por falta de respuesta, el problema de terminar con estratos con un solo conglomerado, es recurrente en encuestas de hogares. Hay tres opciones para lidiar con el problema de estratos unitarios: (1) no hacer nada, lo que implica que no es posible estimar varianzas; (2) asumir que los conglomerados unitarios fueron seleccionados "con certeza" en cuyo caso estos estratos no contribuyen a la estimación de la varianza (se le asigna varianza igual a 0) generando una subestimación de la varianza; y (3) crear unidades que agrupen estratos unitarios junto a otros estratos (pseudo-estratos) para efectos de cálculo de varianza.

Siguiendo las mejores prácticas en la materia, se optó por esta última opción, con la finalidad de poner a disposición de los investigadores variables que permitan identificar las características del diseño complejo de Casen 2015 y producir estimaciones de varianza más estables.

Las estimaciones oficiales de la varianza compleja de Casen 2015 han sido desarrolladas utilizando pseudo-estrato, pseudo-conglomerado y corrección por poblaciones finitas<sup>57</sup>. A continuación, se describen los procedimientos seguidos para la construcción de estas variables.

#### **4.1. Creación de pseudo-estratos (varstrat)**

La creación de pseudo-estratos tiene como objetivo corregir el problema de estimación de varianza que generan los estratos unitarios (1 conglomerado por estrato), el cual se traduce en una subestimación de la varianza de cualquier variable de interés. Los pseudo-estratos, en adelante varstrat, son una re-agrupación de los 588 estratos<sup>58</sup> de muestreo originales (agrupaciones de comuna-área) en que cada uno contiene al menos dos conglomerados. Cada varstrat fue creado de acuerdo a reglas de ordenamiento jerárquico en cuanto a la división político administrativa, marco de procedencia y al total de conglomerados que estos poseen. El procedimiento es como sigue:

1. Si el estrato unitario se encuentra en una provincia y área que posee dos o más estratos, se combina con otro estrato al interior de la provincia y área.
2. Si el estrato unitario se encuentra en una provincia y área que posee sólo un estrato, se une a un estrato en otra provincia al interior de la misma región, siempre perteneciente a la misma área (urbana o rural).

De un total de 588 estratos muestrales se crearon 313 varstrat, los que poseen al menos dos conglomerados. La Tabla V.1 presenta el total de estratos y varstrat por región.

#### **4.2. Creación de pseudo-conglomerados (varunit)**

La creación de pseudo-conglomerados tiene como objetivo estabilizar las estimaciones de varianza, la cual tiende a incrementarse cuando las estimaciones se basan en conglomerados con muy pocas unidades muestrales. Es importante recordar que tanto por diseño como por falta de respuesta existen conglomerados pequeños en la muestra. Por una parte, en manzanas con 8 viviendas sólo 2 se seleccionan en la muestra, por otra parte, por efecto de no respuesta un conglomerado con 15 viviendas seleccionadas puede terminar con solo 8 viviendas entrevistadas. Los pseudo-conglomerados, en adelante varunit, son una re-agrupación de los 11.274 conglomerados originales, en donde cada uno contiene un número mínimo de viviendas.

---

<sup>57</sup> Se pueden encontrar en principales estadísticas obtenidas con datos Casen 2015. Ver [http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen\\_obj.php](http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen_obj.php)

<sup>58</sup> Si se considera la estratificación por grupo de tamaño al interior de la parte urbana del marco, el número total de estratos asciende a 1.193 estratos, antes de ser generados los pseudo-estratos o VarStrat.

A partir de un total de 11.274 conglomerados seleccionados en la muestra Casen 2015, se crearon 2.304 varunit, los cuales agrupan entre 16 y 59 viviendas entrevistadas (cifras menor y mayor a los límites mínimos y máximos establecidos inicialmente de 30 y 50 viviendas por varstrat). La Tabla V.1 presenta el total de conglomerados (Manzanas/Secciones) y varunit por región.

**Tabla V.1 Total de estratos, conglomerados, varstrat y varunit, según región por área geográfica.**

Región	Tipo	Nº Estratos Casen 2015*	Nº Manzanas o Secciones	Nº de Viviendas**	Nº VarStrat	Nº VarUnit
<b>Total País</b>		<b>1.193</b>	<b>11.274</b>	<b>82.120</b>	<b>313</b>	<b>2.304</b>
Tarapacá	Urbano	14	308	1.889	4	55
	Rural	5	21	537	4	15
Antofagasta	Urbano	23	340	1.666	6	50
	Rural	5	12	242	2	7
Atacama	Urbano	22	755	3.525	6	107
	Rural	7	22	314	2	8
Coquimbo	Urbano	35	544	2.826	7	83
	Rural	15	48	850	9	22
Valparaíso	Urbano	116	1.197	7.237	19	207
	Rural	31	91	1.647	10	47
O'Higgins	Urbano	96	612	4.631	19	126
	Rural	33	130	2.429	15	64
Maule	Urbano	69	503	3.647	15	103
	Rural	30	107	1.986	14	53
Biobío	Urbano	134	1.347	8.641	29	244
	Rural	50	151	2.679	11	68
Araucanía	Urbano	78	746	4.406	19	125
	Rural	32	136	2.606	15	67
Los Lagos	Urbano	63	542	3.596	15	104
	Rural	25	117	2.570	13	67
Aysén	Urbano	9	171	905	4	29
	Rural	6	11	246	2	6
Magallanes	Urbano	8	324	1.603	3	48
	Rural	3	9	274	3	7
Metropolitana	Urbano	208	2.513	16.439	43	469
	Rural	19	36	554	4	15
Los Ríos	Urbano	36	290	1.737	8	49
	Rural	12	67	1.627	10	35
Arica y Parinacota	Urbano	6	115	634	1	19
	Rural	3	9	177	1	5

\* Estratos de diseño según Casen 2015, desagregación Comuna-área (Urbano, rural) y Grupo de tamaño (En el caso del área urbana)

\*\* Número de viviendas que respondió Casen 2015

Fuente: Instituto Nacional de Estadísticas y Ministerio de Desarrollo Social.

Los varunit fueron contruidos de acuerdo a criterios de orden jerárquico al interior de cada varstrat, según el total de viviendas entrevistadas en la encuesta. El procedimiento consistió en

unir cada manzana o sección con menos de 30 viviendas entrevistadas. El procedimiento utilizado inicialmente es el siguiente:

1. Se ordenan los conglomerados al interior de cada varstrat según el total de viviendas entrevistadas.
2. Al interior de un varstrat, cada uno de los conglomerados con menos de 30 viviendas se une al conglomerado siguiente hasta acumular un total de 30 o más viviendas, con un tope máximo de 50 viviendas.
3. Si al combinar los conglomerados de un mismo varstrat la cantidad de viviendas sigue siendo inferior a 30, o la cantidad de varunit es igual a 1, el varstrat en el cual se realiza la agrupación de conglomerados se une a otro varstrat según los criterios establecidos para la creación de los varstrat.

El procedimiento fue repetido hasta que cada varunit tuviera entre 30 y 50 viviendas, y hasta que cada varstrat tuviera dos o más varunit. En la implementación del algoritmo se observó que 86 conglomerados quedaban por debajo del criterio establecido (menos de 30 viviendas), lo que se debió a que el total de viviendas que respondieron en dichos conglomerados se encontraba por debajo de 30 (superior a 15 e inferior a 30). Ver Tabla V.2 siguiente.

En las tablas V.3 y V.4 se presentan las distribuciones totales del número de conglomerados (manzanas o secciones) que componen los varunit y el número de varunit que componen los varstrat.

**Tabla V.2 Frecuencia de varunit según número de viviendas que lo componen**

Nº de viviendas que responde	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
16	1	0,04	0,04	0,04
17	1	0,04	0,04	0,09
18	4	0,17	0,17	0,26
20	8	0,35	0,35	0,61
21	7	0,30	0,30	0,91
22	9	0,39	0,39	1,30
23	8	0,35	0,35	1,65
24	2	0,09	0,09	1,74
25	9	0,39	0,39	2,13
26	8	0,35	0,35	2,47
27	10	0,43	0,43	2,91
28	10	0,43	0,43	3,34
29	9	0,39	0,39	3,73
30	331	14,37	14,37	18,10
31	167	7,25	7,25	25,35
32	415	18,01	18,01	43,36
33	154	6,68	6,68	50,04
34	139	6,03	6,03	56,08
35	157	6,81	6,81	62,89
36	115	4,99	4,99	67,88
37	72	3,13	3,13	71,01
38	68	2,95	2,95	73,96
39	70	3,04	3,04	77,00
40	64	2,78	2,78	79,77
41	49	2,13	2,13	81,90
42	65	2,82	2,82	84,72
43	46	2,00	2,00	86,72
44	38	1,65	1,65	88,37
45	27	1,17	1,17	89,54
46	35	1,52	1,52	91,06
47	25	1,09	1,09	92,14
48	30	1,30	1,30	93,45
49	32	1,39	1,39	94,84
50	30	1,30	1,30	96,14
51	16	0,69	0,69	96,83
52	19	0,82	0,82	97,66
53	15	0,65	0,65	98,31
54	15	0,65	0,65	98,96
55	6	0,26	0,26	99,22
56	7	0,30	0,30	99,52
57	2	0,09	0,09	99,61
58	7	0,30	0,30	99,91
59	2	0,09	0,09	100,00
<b>Total</b>	<b>2.304</b>	<b>100,00</b>	<b>100,00</b>	

Fuente: Instituto Nacional de Estadísticas.

**Tabla V.3 distribución del número de conglomerados que componen los varunit**

<b>N° de Manzanas o Secciones</b>	<b>Frecuencia</b>	<b>Porcentaje</b>	<b>Porcentaje válido</b>	<b>Porcentaje acumulado</b>
1	122	5,30	5,30	5,30
2	643	27,91	27,91	33,20
3	294	12,76	12,76	45,96
4	270	11,72	11,72	57,68
5	266	11,55	11,55	69,23
6	92	3,99	3,99	73,22
7	83	3,60	3,60	76,82
8	209	9,07	9,07	85,89
9	49	2,13	2,13	88,02
10	133	5,77	5,77	93,79
11	33	1,43	1,43	95,23
12	30	1,30	1,30	96,53
13	18	0,78	0,78	97,31
14	12	0,52	0,52	97,83
15	13	0,56	0,56	98,39
16	11	0,48	0,48	98,87
17	10	0,43	0,43	99,31
18	3	0,13	0,13	99,44
19	6	0,26	0,26	99,70
21	1	0,04	0,04	99,74
22	3	0,13	0,13	99,87
23	1	0,04	0,04	99,91
24	1	0,04	0,04	99,96
30	1	0,04	0,04	100,00
<b>Total</b>	<b>2.304</b>	<b>100,00</b>	<b>100,00</b>	

Fuente: Instituto Nacional de Estadísticas.

**Tabla V.4 Distribución del número de VarUnit que componen los VarStrat**

Nº de VarUnit	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
2	44	14,1	14,1	14,1
3	75	24,0	24,0	38,0
4	47	15,0	15,0	53,0
5	32	10,2	10,2	63,3
6	14	4,5	4,5	67,7
7	10	3,2	3,2	70,9
8	11	3,5	3,5	74,4
9	12	3,8	3,8	78,3
10	6	1,9	1,9	80,2
11	10	3,2	3,2	83,4
12	3	1,0	1,0	84,3
13	9	2,9	2,9	87,2
14	4	1,3	1,3	88,5
16	5	1,6	1,6	90,1
17	3	1,0	1,0	91,1
18	2	,6	,6	91,7
19	2	,6	,6	92,3
20	3	1,0	1,0	93,3
21	1	,3	,3	93,6
22	1	,3	,3	93,9
23	1	,3	,3	94,2
24	1	,3	,3	94,6
25	2	,6	,6	95,2
26	2	,6	,6	95,8
27	3	1,0	1,0	96,8
28	2	,6	,6	97,4
32	2	,6	,6	98,1
33	2	,6	,6	98,7
36	1	,3	,3	99,0
41	2	,6	,6	99,7
72	1	,3	,3	100,0
<b>Total</b>	<b>313</b>	<b>100,0</b>	<b>100,0</b>	

Fuente: Instituto Nacional de Estadísticas.

## 5. Resultados de la situación de pobreza por ingresos

En Casen 2015, al igual que en Casen 2013, las estimaciones y márgenes de error de la situación de pobreza por ingresos, fueron determinadas utilizando los parámetros varstrat, varunit y expr (factor de expansión regional), considerando la utilización de pseudo-estratos y pseudo-conglomerados, siendo éstos implementados mediante cálculos realizados en paralelo utilizando los programas SPSS y STATA.

En el año 2011, los márgenes de error fueron estimados utilizando los parámetros varstrat, varunit, expr\_r2 y fmr2. Por otro lado, para los años 1996 a 2009, los márgenes de error fueron estimados utilizando los parámetros estrato, segmento, expr, ya que no se disponía para la serie histórica de pseudo-estratos (varstrat) y pseudo-conglomerados (varunit).

En la Tabla V.5 se presentan las tasas de pobreza por ingresos para las personas y los hogares para la encuesta Casen 2015, a nivel nacional, por área geográfica urbana y rural y a nivel regional. Al igual, se calcularon los márgenes de error al 95% de confianza, asociados a estas estimaciones.

**Tabla V.5 Estimación de porcentaje de la población en situación de pobreza por ingresos, según zona y región, Casen 2015.**

Nivel	Estadísticos		Situación de pobreza por ingresos			
			Pobreza extrema	Pobreza no extrema	Total Pobreza	No pobre
Nacional	Estimación		3,5%	8,1%	11,7%	88,3%
	Error estándar		0,1%	0,2%	0,2%	0,2%
	95% de intervalo de confianza	Inferior	3,3%	7,8%	11,3%	87,9%
		Superior	3,8%	8,5%	12,1%	88,7%
	Coeficiente de variación		3,2%	2,0%	1,9%	0,2%
Urbano	Estimación		3,0%	7,1%	10,2%	89,8%
	Error estándar		0,1%	0,2%	0,2%	0,2%
	95% de intervalo de confianza	Inferior	2,8%	6,8%	9,7%	89,4%
		Superior	3,3%	7,5%	10,6%	90,3%
	Coeficiente de variación		3,9%	2,5%	2,3%	0,3%
Rural	Estimación		7,0%	15,1%	22,1%	77,9%
	Error estándar		0,3%	0,4%	0,6%	0,6%
	95% de intervalo de confianza	Inferior	6,4%	14,3%	21,0%	76,8%
		Superior	7,7%	15,9%	23,2%	79,0%
	Coeficiente de variación		4,6%	2,8%	2,6%	0,7%
1 Región de Tarapacá	Estimación		2,3%	4,8%	7,1%	92,9%
	Error estándar		0,5%	0,6%	0,8%	0,8%
	95% de intervalo de confianza	Inferior	1,5%	3,7%	5,7%	91,1%
		Superior	3,5%	6,3%	8,9%	94,3%
	Coeficiente de variación		21,5%	13,1%	11,6%	0,9%
2 Región de Antofagasta	Estimación		1,6%	3,8%	5,4%	94,6%
	Error estándar		0,4%	0,7%	0,9%	0,9%
	95% de intervalo de confianza	Inferior	1,0%	2,6%	3,8%	92,4%
		Superior	2,6%	5,6%	7,6%	96,2%
	Coeficiente de variación		23,9%	19,5%	17,4%	1,0%
3 Región de Atacama	Estimación		1,7%	5,1%	6,9%	93,1%
	Error estándar		0,4%	0,4%	0,6%	0,6%
	95% de intervalo de confianza	Inferior	1,1%	4,4%	5,8%	91,9%
		Superior	2,9%	6,0%	8,1%	94,2%
	Coeficiente de variación		25,4%	8,0%	8,5%	0,6%
4 Región de Coquimbo	Estimación		4,0%	9,8%	13,8%	86,2%
	Error estándar		0,4%	0,7%	0,7%	0,7%
	95% de intervalo de confianza	Inferior	3,2%	8,6%	12,4%	84,7%
		Superior	4,9%	11,3%	15,3%	87,6%
	Coeficiente de variación		10,4%	7,0%	5,3%	0,8%

**Tabla V5 Estimación de porcentaje de la población en situación de pobreza por ingresos, según zona y región, Casen 2015.**

Nivel	Estadísticos		Situación de pobreza por ingresos			
			Pobreza extrema	Pobreza no extrema	Total Pobreza	No pobre
5 Región de Valparaíso	Estimación		3,2%	8,8%	12,0%	88,0%
	Error estándar		0,3%	0,5%	0,6%	0,6%
	95% de intervalo de confianza	Inferior	2,6%	7,8%	10,9%	86,8%
		Superior	3,9%	9,9%	13,2%	89,1%
	Coeficiente de variación		10,3%	5,8%	4,9%	0,7%
6 Región del Libertador Gral. Bernardo O'Higgins	Estimación		3,8%	10,0%	13,7%	86,3%
	Error estándar		0,4%	0,7%	0,9%	0,9%
	95% de intervalo de confianza	Inferior	3,0%	8,7%	12,1%	84,5%
		Superior	4,7%	11,3%	15,5%	87,9%
	Coeficiente de variación		10,9%	6,6%	6,3%	1,0%
7 Región del Maule	Estimación		5,1%	13,6%	18,7%	81,3%
	Error estándar		0,4%	0,7%	0,8%	0,8%
	95% de intervalo de confianza	Inferior	4,4%	12,2%	17,1%	79,6%
		Superior	6,0%	15,0%	20,4%	82,9%
	Coeficiente de variación		8,2%	5,3%	4,5%	1,0%
8 Región del Biobío	Estimación		5,8%	11,8%	17,6%	82,4%
	Error estándar		0,3%	0,5%	0,6%	0,6%
	95% de intervalo de confianza	Inferior	5,2%	10,8%	16,3%	81,2%
		Superior	6,5%	12,7%	18,8%	83,7%
	Coeficiente de variación		5,7%	4,1%	3,6%	0,8%
9 Región de La Araucanía	Estimación		8,4%	15,2%	23,6%	76,4%
	Error estándar		0,5%	0,6%	0,9%	0,9%
	95% de intervalo de confianza	Inferior	7,4%	14,0%	21,8%	74,5%
		Superior	9,5%	16,5%	25,5%	78,2%
	Coeficiente de variación		6,5%	4,1%	4,0%	1,2%
10 Región de Los Lagos	Estimación		5,0%	11,1%	16,1%	83,9%
	Error estándar		0,4%	0,7%	0,8%	0,8%
	95% de intervalo de confianza	Inferior	4,2%	9,8%	14,5%	82,2%
		Superior	5,9%	12,5%	17,8%	85,5%
	Coeficiente de variación		8,5%	6,2%	5,2%	1,0%
11 Región de Aysén del Gral. Carlos Ibáñez del Campo	Estimación		1,6%	4,9%	6,5%	93,5%
	Error estándar		0,5%	1,0%	1,2%	1,2%
	95% de intervalo de confianza	Inferior	0,9%	3,2%	4,4%	90,6%
		Superior	2,9%	7,3%	9,4%	95,6%
	Coeficiente de variación		30,1%	20,7%	19,3%	1,3%

**Tabla V5 Estimación de porcentaje de la población en situación de pobreza por ingresos, según zona y región, Casen 2015.**

Nivel	Estadísticos		Situación de pobreza por ingresos			
			Pobreza extrema	Pobreza no extrema	Total Pobreza	No pobre
12 Región de Magallanes y de la Antártica Chilena	Estimación		1,6%	2,9%	4,4%	95,6%
	Error estándar		0,4%	0,6%	0,8%	0,8%
	95% de intervalo de confianza	Inferior	1,0%	1,9%	3,1%	93,8%
		Superior	2,5%	4,4%	6,2%	96,9%
	Coeficiente de variación		24,0%	22,0%	17,5%	0,8%
13 Región Metropolitana de Santiago	Estimación		2,1%	5,0%	7,1%	92,9%
	Error estándar		0,2%	0,3%	0,4%	0,4%
	95% de intervalo de confianza	Inferior	1,8%	4,5%	6,4%	92,1%
		Superior	2,6%	5,6%	7,9%	93,6%
	Coeficiente de variación		9,5%	5,7%	5,6%	0,4%
14 Región de Los Ríos	Estimación		4,8%	11,9%	16,8%	83,2%
	Error estándar		0,7%	1,0%	1,4%	1,4%
	95% de intervalo de confianza	Inferior	3,7%	10,1%	14,3%	80,4%
		Superior	6,3%	14,0%	19,6%	85,7%
	Coeficiente de variación		13,6%	8,3%	8,1%	1,6%
15 Región de Arica y Parinacota	Estimación		2,0%	7,7%	9,7%	90,3%
	Error estándar		0,6%	1,1%	1,3%	1,3%
	95% de intervalo de confianza	Inferior	1,1%	5,9%	7,4%	87,4%
		Superior	3,7%	10,1%	12,6%	92,6%
	Coeficiente de variación		31,6%	13,8%	13,5%	1,5%

**Tabla V.6 Estimación de porcentaje de hogares en situación de pobreza por ingresos, según zona y región, Casen 2015.**

Nivel	Estadísticos		Situación de pobreza por ingresos			
			Pobreza extrema	Pobreza no extrema	Total Pobreza	No pobre
Nacional	Estimación		3,2%	7,2%	10,4%	89,6%
	Error estándar		0,1%	0,1%	0,2%	0,2%
	95% de intervalo de confianza	Inferior	3,0%	6,9%	10,0%	89,2%
		Superior	3,4%	7,5%	10,8%	90,0%
	Coeficiente de variación		3,2%	2,1%	1,9%	0,2%
Urbano	Estimación		2,7%	6,2%	8,9%	91,1%
	Error estándar		0,1%	0,2%	0,2%	0,2%
	95% de intervalo de confianza	Inferior	2,5%	5,9%	8,5%	90,6%
		Superior	3,0%	6,5%	9,4%	91,5%
	Coeficiente de variación		4,0%	2,6%	2,4%	0,2%
Rural	Estimación		6,1%	14,1%	20,2%	79,8%
	Error estándar		0,3%	0,4%	0,5%	0,5%
	95% de intervalo de confianza	Inferior	5,6%	13,4%	19,2%	78,8%
		Superior	6,7%	14,8%	21,2%	80,8%
	Coeficiente de variación		4,3%	2,6%	2,5%	0,6%
1 Región de Tarapacá	Estimación		2,1%	4,0%	6,1%	93,9%
	Error estándar		0,4%	0,5%	0,7%	0,7%
	95% de intervalo de confianza	Inferior	1,4%	3,1%	4,8%	92,2%
		Superior	3,1%	5,2%	7,8%	95,2%
	Coeficiente de variación		20,3%	13,2%	12,1%	0,8%
2 Región de Antofagasta	Estimación		1,4%	3,1%	4,6%	95,4%
	Error estándar		0,3%	0,5%	0,7%	0,7%
	95% de intervalo de confianza	Inferior	1,0%	2,2%	3,4%	93,9%
		Superior	2,2%	4,4%	6,1%	96,6%
	Coeficiente de variación		20,9%	17,3%	15,3%	0,7%
3 Región de Atacama	Estimación		1,4%	4,4%	5,8%	94,2%
	Error estándar		0,3%	0,4%	0,5%	0,5%
	95% de intervalo de confianza	Inferior	1,0%	3,7%	4,9%	93,1%
		Superior	2,1%	5,2%	6,9%	95,1%
	Coeficiente de variación		18,7%	9,0%	8,5%	0,5%
4 Región de Coquimbo	Estimación		3,5%	8,6%	12,2%	87,8%
	Error estándar		0,3%	0,6%	0,7%	0,7%
	95% de intervalo de confianza	Inferior	2,9%	7,6%	10,9%	86,5%
		Superior	4,3%	9,8%	13,5%	89,1%
	Coeficiente de variación		9,8%	6,6%	5,4%	0,7%

Nivel	Estadísticos	Situación de pobreza por ingresos				
		Pobreza extrema	Pobreza no extrema	Total Pobreza	No pobre	
5 Región de Valparaíso	Estimación		2,9%	7,2%	10,1%	89,9%
	Error estándar		0,3%	0,4%	0,5%	0,5%
	95% de intervalo de confianza	Inferior	2,3%	6,4%	9,1%	88,8%
		Superior	3,5%	8,2%	11,2%	90,9%
	Coefficiente de variación		10,3%	6,2%	5,2%	0,6%
6 Región del Libertador Gral. Bernardo O Higgins	Estimación		3,5%	9,2%	12,7%	87,3%
	Error estándar		0,4%	0,6%	0,8%	0,8%
	95% de intervalo de confianza	Inferior	2,8%	8,2%	11,3%	85,7%
		Superior	4,3%	10,4%	14,3%	88,7%
	Coefficiente de variación		11,2%	6,2%	6,1%	0,9%
7 Región del Maule	Estimación		4,6%	12,2%	16,8%	83,2%
	Error estándar		0,4%	0,6%	0,7%	0,7%
	95% de intervalo de confianza	Inferior	4,0%	11,1%	15,4%	81,7%
		Superior	5,4%	13,4%	18,3%	84,6%
	Coefficiente de variación		7,8%	4,9%	4,3%	0,9%
8 Región del Biobío	Estimación		5,4%	10,8%	16,2%	83,8%
	Error estándar		0,3%	0,4%	0,6%	0,6%
	95% de intervalo de confianza	Inferior	4,8%	10,0%	15,1%	82,7%
		Superior	5,9%	11,7%	17,3%	84,9%
	Coefficiente de variación		5,2%	4,1%	3,5%	0,7%
9 Región de La Araucanía	Estimación		7,0%	13,7%	20,7%	79,3%
	Error estándar		0,4%	0,6%	0,8%	0,8%
	95% de intervalo de confianza	Inferior	6,3%	12,6%	19,2%	77,7%
		Superior	7,9%	14,8%	22,3%	80,8%
	Coefficiente de variación		5,8%	4,0%	3,8%	1,0%
10 Región de Los Lagos	Estimación		4,5%	9,8%	14,3%	85,7%
	Error estándar		0,4%	0,5%	0,7%	0,7%
	95% de intervalo de confianza	Inferior	3,8%	8,8%	13,0%	84,4%
		Superior	5,3%	10,8%	15,6%	87,0%
	Coefficiente de variación		8,3%	5,2%	4,8%	0,8%
11 Región de Aysén del Gral. Carlos Ibáñez del Campo	Estimación		1,5%	4,2%	5,7%	94,3%
	Error estándar		0,4%	0,7%	0,8%	0,8%
	95% de intervalo de confianza	Inferior	0,9%	3,0%	4,3%	92,5%
		Superior	2,5%	5,9%	7,5%	95,7%
	Coefficiente de variación		27,1%	17,3%	14,5%	0,9%

Nivel	Estadísticos		Situación de pobreza por ingresos			
			Pobreza extrema	Pobreza no extrema	Total Pobreza	No pobre
12 Región de Magallanes y de la Antártica Chilena	Estimación		1,3%	2,5%	3,8%	96,2%
	Error estándar		0,3%	0,6%	0,7%	0,7%
	95% de intervalo de confianza	Inferior	0,9%	1,5%	2,6%	94,6%
		Superior	2,1%	4,0%	5,4%	97,4%
	Coeficiente de variación		22,6%	25,0%	18,5%	0,7%
13 Región Metropolitana de Santiago	Estimación		1,9%	4,3%	6,2%	93,8%
	Error estándar		0,2%	0,3%	0,4%	0,4%
	95% de intervalo de confianza	Inferior	1,6%	3,8%	5,5%	93,0%
		Superior	2,3%	4,8%	7,0%	94,5%
	Coeficiente de variación		10,1%	6,2%	6,0%	0,4%
14 Región de Los Ríos	Estimación		4,1%	10,5%	14,5%	85,5%
	Error estándar		0,5%	0,8%	1,1%	1,1%
	95% de intervalo de confianza	Inferior	3,2%	9,0%	12,5%	83,2%
		Superior	5,2%	12,2%	16,8%	87,5%
	Coeficiente de variación		12,7%	7,7%	7,5%	1,3%
15 Región de Arica y Parinacota	Estimación		1,7%	6,1%	7,8%	92,2%
	Error estándar		0,6%	0,7%	1,1%	1,1%
	95% de intervalo de confianza	Inferior	0,9%	4,8%	5,9%	89,8%
		Superior	3,3%	7,7%	10,2%	94,1%
	Coeficiente de variación		33,0%	12,1%	13,8%	1,2%

## VI. ANEXOS

### 1. Simulaciones

**Tabla 1: Distribución tamaño muestral final Casen 2015, con variable de interés tasa de pobreza por ingresos**

Región	Comuna	Estimador de la razón Casen 2013	Muestra Casen 2013 (viviendas)	Error absoluto efectivo Casen 2013	Error relativo efectivo Casen 2013 (%)	Error absoluto propuesto 2015	Error relativo propuesto 2015	Muestra objetivo Casen 2015	Muestra con sobremuestreo Casen 2015
<b>Total País (Tasa de pobreza)</b>		<b>14,4%</b>	<b>64.842</b>	<b>0,6%</b>	<b>4,1%</b>	<b>0,5%</b>	<b>3,7%</b>	<b>82.370</b>	<b>107.932</b>
I	Iquique **	6,2%	1.516	1,9%	30,6%	2,5%	39,8%	900	1.319
	Alto Hospicio **	10,3%	844	3,1%	30,3%	3,4%	33,2%	700	872
	Pozo Almonte **	14,0%	86	13,7%	97,9%	5,2%	37,1%	600	685
	Camíña	31,1%	39	20,6%	66,2%	20,6%	66,2%	39	40
	Huara	17,3%	89	14,8%	85,8%	14,8%	85,8%	89	99
	Pica	25,5%	83	9,9%	38,7%	9,0%	35,3%	100	131
<b>Total Regional</b>		<b>8,2%</b>	<b>2.657</b>	<b>1,6%</b>	<b>19,7%</b>	<b>2,4%</b>	<b>18,0%</b>	<b>2.428</b>	<b>3.146</b>
II	Antofagasta **	3,7%	1.255	1,8%	49,7%	2,3%	62,3%	800	1.192
	Mejillones	8,7%	46	18,5%	212,4%	12,6%	144,0%	100	129
	Sierra Gorda	19,6%	37	4,2%	21,3%	4,3%	21,9%	35	41
	Taltal	6,4%	77	6,1%	95,8%	4,4%	68,7%	150	177
	Calama **	3,4%	447	2,1%	60,0%	2,2%	63,4%	400	581
	San Pedro de Atacama	17,0%	76	8,6%	50,8%	8,4%	49,5%	80	111
	Tocopilla **	2,2%	93	2,4%	108,2%	1,3%	60,2%	300	394
	María Elena	13,0%	35	9,9%	76,3%	5,9%	45,1%	100	114
<b>Total Regional</b>		<b>4,0%</b>	<b>2.066</b>	<b>1,3%</b>	<b>33,9%</b>	<b>1,4%</b>	<b>34,8%</b>	<b>1.965</b>	<b>2.739</b>
III	Copiapó **	6,3%	1.070	4,6%	73,6%	3,1%	49,6%	2.350	3.663
	Caldera	7,0%	151	2,1%	30,4%	2,1%	30,4%	151	182
	Tierra Amarilla	7,3%	110	1,6%	22,4%	1,6%	22,4%	110	134
	Chañaral **	10,0%	96	12,0%	119,0%	5,0%	49,7%	550	648
	Diego de Almagro	5,9%	163	6,2%	106,3%	6,2%	106,3%	163	199
	Vallenar **	7,8%	364	3,0%	38,0%	3,3%	41,9%	300	371
	Alto del Carmen	23,3%	49	10,4%	44,8%	10,4%	44,8%	49	50
	Freirina	7,5%	92	4,7%	63,4%	4,7%	63,4%	92	107
Huasco	10,2%	96	7,5%	73,5%	7,5%	73,5%	96	111	
<b>Total Regional</b>		<b>7,3%</b>	<b>2.191</b>	<b>2,5%</b>	<b>34,7%</b>	<b>1,9%</b>	<b>26,1%</b>	<b>3.861</b>	<b>5.465</b>
IV	La Serena **	13,7%	796	6,0%	43,9%	4,9%	35,7%	1.200	1.603
	Coquimbo **	13,2%	677	3,1%	23,7%	2,9%	21,8%	800	990
	Andacollo	15,5%	82	5,4%	34,9%	5,4%	34,9%	82	90
	La Higuera	24,5%	91	12,3%	50,0%	12,3%	50,0%	91	94
	Paiguano	18,3%	42	7,5%	41,1%	7,5%	41,1%	42	42
	Vicuña **	26,2%	114	8,5%	32,3%	6,4%	24,4%	200	223
	Illapel **	16,3%	139	5,8%	35,6%	4,8%	29,7%	200	231
	Canela	44,8%	72	24,6%	54,9%	24,6%	54,9%	72	81
	Los Vilos	27,5%	163	12,1%	44,0%	12,1%	44,0%	163	190
	Salamanca	20,4%	90	14,9%	73,0%	14,9%	73,0%	90	103
	Ovalle **	15,6%	392	6,2%	39,9%	5,5%	35,3%	500	597
	Combarbalá	37,1%	67	23,7%	63,7%	23,7%	63,7%	67	80
	Monte Patria	31,1%	126	3,6%	11,7%	3,6%	11,7%	126	142
	Punitaqui	27,3%	79	7,2%	26,2%	7,2%	26,2%	79	92
Río Hurtado	48,4%	26	12,7%	26,3%	12,7%	26,3%	26	29	
<b>Total Regional</b>		<b>16,2%</b>	<b>2.956</b>	<b>2,4%</b>	<b>15,0%</b>	<b>2,2%</b>	<b>13,3%</b>	<b>3.738</b>	<b>4.587</b>

\*Tamaños corregidos por tasa de no respuesta promedio ponderado Casen 2011 – 2013

\*\* Comunas definidas como dominios de estudio.

Continúa ►

Región	Comuna	Estimador de la razón Casen 2013	Muestra Casen 2013 (viviendas)	Error absoluto	Error relativo	Error absoluto	Error relativo	Muestra objetivo	Tamaños corregidos*
				Casen 2013	Casen 2013	Casen 2013	Casen 2013	Casen 2015	Casen 2015
V	Valparaíso **	17,3%	953	4,9%	28,5%	4,8%	27,8%	1.000	1.437
	Casablanca	25,3%	90	20,2%	80,0%	20,2%	80,0%	90	110
	Concón **	7,4%	72	5,6%	76,3%	2,6%	34,6%	350	565
	Puchuncaví	14,2%	75	6,2%	43,9%	6,2%	43,9%	75	83
	Quintero	22,6%	86	20,0%	88,4%	20,0%	88,4%	86	97
	Viña del Mar **	13,3%	911	4,1%	30,9%	3,7%	28,1%	1.100	1.783
	Los Andes **	12,0%	259	4,6%	38,4%	4,0%	33,0%	350	474
	Calle Larga	14,1%	63	7,0%	49,9%	7,0%	49,9%	63	80
	Rinconada	20,6%	72	5,1%	24,6%	5,1%	24,6%	72	85
	San Esteban	12,5%	70	7,9%	62,9%	7,9%	62,9%	70	84
	La Ligua **	18,6%	127	11,9%	64,2%	6,3%	34,1%	450	534
	Cabildo	18,8%	80	6,3%	33,4%	6,3%	33,4%	80	90
	Papudo	5,7%	56	10,3%	179,2%	10,3%	179,2%	56	59
	Petorca	29,0%	86	16,1%	55,5%	16,1%	55,5%	86	96
	Zapallar	24,8%	78	14,4%	57,8%	14,4%	57,8%	78	96
	Quillota **	14,0%	270	4,2%	30,3%	4,2%	30,3%	270	352
	Calera **	25,3%	172	9,6%	37,9%	8,9%	35,1%	200	233
	Hijuelas	19,2%	79	8,3%	43,3%	8,3%	43,3%	79	90
	La Cruz	28,1%	87	13,6%	48,4%	13,6%	48,4%	87	105
	Nogales	17,0%	66	20,1%	118,2%	20,1%	118,2%	66	75
	San Antonio **	18,9%	308	8,3%	43,9%	6,5%	34,5%	500	593
	Algarrobo	6,4%	72	6,9%	107,4%	6,9%	107,4%	72	88
	Cartagena	26,1%	69	7,2%	27,7%	7,2%	27,7%	69	80
	El Quisco **	8,7%	34	12,9%	148,6%	4,3%	50,0%	300	387
	El Tabo	2,2%	39	4,7%	213,9%	4,7%	213,9%	39	48
	Santo Domingo	8,5%	99	2,5%	29,9%	2,5%	29,9%	99	119
	San Felipe **	11,0%	162	6,7%	61,1%	3,5%	31,8%	600	894
	Catemu	20,5%	70	13,2%	64,6%	13,2%	64,6%	70	82
	Llailay	31,2%	89	7,4%	23,7%	7,4%	23,7%	89	116
	Panquehue	10,5%	69	7,2%	69,0%	7,2%	69,0%	69	89
Putaendo	13,9%	56	9,8%	70,0%	9,8%	70,0%	56	80	
Santa María	30,2%	69	12,7%	42,1%	12,7%	42,1%	69	91	
Quilpué **	8,3%	595	4,8%	58,1%	4,1%	49,5%	820	1.063	
Limache **	26,8%	330	14,2%	53,1%	9,4%	35,2%	750	862	
Olmué	24,0%	87	4,3%	17,9%	4,3%	17,9%	87	100	
Villa Alemana **	12,8%	437	3,7%	28,9%	3,6%	28,5%	450	571	
<b>Total Regional</b>	<b>15,6%</b>	<b>6.337</b>	<b>1,9%</b>	<b>12,4%</b>	<b>1,6%</b>	<b>10,5%</b>	<b>8.847</b>	<b>11.791</b>	
VI	Rancagua **	14,2%	1.209	4,0%	28,4%	4,2%	29,7%	1.100	1.323
	Codegua **	20,9%	84	6,8%	32,5%	6,2%	29,8%	100	117
	Coinco	20,6%	80	7,1%	34,6%	7,1%	34,6%	80	96
	Coltauco **	10,9%	75	10,6%	97,2%	3,8%	34,4%	600	692
	Doñihue	20,0%	124	4,6%	23,0%	4,6%	23,0%	124	134
	Graneros **	13,9%	247	4,3%	30,7%	4,7%	34,1%	200	226
	Las Cabras	33,1%	112	18,5%	55,8%	13,8%	41,7%	200	234
	Machalí **	5,8%	170	5,8%	99,4%	2,9%	49,0%	700	977
	Malloa	24,6%	73	7,0%	28,4%	7,0%	28,4%	73	78
	Mostazal	6,9%	143	5,8%	83,7%	5,8%	83,7%	143	162
	Olivar **	16,2%	84	7,8%	48,1%	7,1%	44,1%	100	108
	Peumo **	8,4%	103	6,2%	73,6%	4,0%	47,3%	250	285
	Pichidegua **	26,7%	77	5,6%	21,0%	3,5%	13,1%	200	211
	Quinta de Tilcoco	20,0%	81	14,5%	72,4%	14,5%	72,4%	81	89
	Rengo **	21,7%	291	5,8%	27,0%	5,8%	26,5%	300	346
	Requínoa **	6,4%	107	7,8%	123,2%	3,2%	50,0%	650	735
	San Vicente **	17,0%	267	5,7%	33,6%	5,9%	34,7%	250	272
	Pichilemu **	13,7%	78	21,0%	152,9%	13,1%	95,5%	200	251
	La Estrella	13,0%	78	11,5%	88,4%	11,5%	88,4%	78	100
	Litueche	21,7%	68	15,9%	73,2%	15,9%	73,2%	68	81

\*Tamaños corregidos por tasa de no respuesta promedio ponderado Casen 2011 – 2013

\*\* Comunas definidas como dominios de estudio.

Continúa ►

Región	Comuna	Estimador de la razón Casen 2013	Muestra Casen 2013 (viviendas)	Error absoluto	Error relativo	Error absoluto	Error relativo	Muestra objetivo	Tamaños corregidos*
				Casen 2013	Casen 2013	Casen 2013 (%)	propuesto Casen 2015	propuesto Casen 2015	propuesta Casen 2015
VI	Marchihue	17,5%	75	13,2%	75,2%	13,2%	75,2%	75	90
	Navidad	22,4%	55	17,2%	76,7%	17,2%	76,7%	55	73
	Paredones	32,9%	66	2,0%	6,1%	2,0%	6,1%	66	83
	San Fernando **	17,6%	385	6,7%	38,1%	6,2%	35,3%	450	552
	Chépica	25,2%	73	9,2%	36,5%	9,2%	36,5%	73	87
	Chimbarongo **	15,9%	149	4,3%	26,8%	3,7%	23,2%	200	225
	Lolol	33,5%	78	13,0%	38,7%	13,0%	38,7%	78	92
	Nancagua	18,7%	81	13,3%	71,4%	13,3%	71,4%	81	94
	Palmilla	24,3%	78	7,6%	31,4%	7,6%	31,4%	78	86
	Peralillo	22,3%	84	10,5%	47,0%	10,5%	47,0%	84	92
	Placilla	20,1%	73	15,7%	77,9%	15,7%	77,9%	73	84
	Pumanque	25,4%	34	18,6%	73,0%	18,6%	73,0%	34	39
	Santa Cruz **	15,2%	205	6,1%	40,4%	5,1%	33,4%	300	354
<b>Total Regional</b>	<b>16,0%</b>	<b>4.987</b>	<b>2,0%</b>	<b>12,4%</b>	<b>1,7%</b>	<b>10,4%</b>	<b>7.144</b>	<b>8.468</b>	
VII	Talca **	14,7%	624	3,8%	25,9%	3,0%	20,5%	1.000	1.211
	Constitución **	19,9%	167	8,4%	42,2%	6,9%	34,5%	250	285
	Curepto	43,3%	87	10,7%	24,7%	10,7%	24,7%	87	92
	Empedrado	12,2%	70	8,7%	70,9%	8,7%	70,9%	70	76
	Maule	11,0%	80	10,6%	96,2%	10,6%	96,2%	80	84
	Pelarco	26,1%	77	8,5%	32,7%	8,5%	32,7%	77	80
	Pencahue	25,4%	86	15,9%	62,4%	15,9%	62,4%	86	92
	Río Claro	19,7%	82	5,2%	26,4%	5,2%	26,4%	82	86
	San Clemente **	23,5%	144	6,7%	28,6%	5,7%	24,2%	200	213
	San Rafael	17,5%	78	7,6%	43,3%	7,6%	43,3%	78	83
	Cauquenes **	33,0%	148	13,0%	39,2%	9,3%	28,0%	290	322
	Chanco	23,7%	75	15,1%	63,4%	15,1%	63,4%	75	84
	Pelluhue	16,0%	58	8,0%	50,0%	8,0%	50,0%	58	61
	Curicó **	18,7%	1.073	4,1%	21,8%	5,5%	29,2%	600	731
	Hualañé	34,9%	63	16,3%	46,6%	16,3%	46,6%	63	76
	Licantén	25,3%	82	14,2%	56,2%	14,2%	56,2%	82	90
	Molina	25,6%	132	3,4%	13,4%	3,4%	13,4%	132	146
	Rauco	26,6%	79	12,2%	45,9%	12,2%	45,9%	79	86
	Romerol	23,1%	86	2,9%	12,7%	2,9%	12,7%	86	93
	Sagrada Familia	26,6%	72	4,7%	17,8%	4,7%	17,8%	72	80
	Teno **	28,9%	114	3,7%	12,9%	2,8%	9,7%	200	229
	Vichuquén	38,0%	65	18,0%	47,5%	18,0%	47,5%	65	81
	Linares **	20,3%	316	6,4%	31,3%	6,3%	31,1%	320	356
Colbún **	35,9%	73	11,7%	32,5%	7,0%	19,6%	200	206	
Longaví	19,8%	115	24,0%	121,1%	14,9%	75,0%	300	319	
Parral **	31,2%	119	16,9%	54,1%	9,4%	30,3%	380	448	
Retiro	18,4%	86	12,6%	68,8%	12,6%	68,8%	86	91	
San Javier **	34,4%	153	9,6%	28,0%	8,9%	25,8%	180	186	
Villa Alegre **	30,6%	68	7,8%	25,4%	4,5%	14,8%	200	216	
Yerbas Buenas **	32,1%	85	13,5%	42,2%	8,8%	27,5%	200	204	
<b>Total Regional</b>	<b>22,3%</b>	<b>4.557</b>	<b>2,2%</b>	<b>10,0%</b>	<b>2,0%</b>	<b>9,0%</b>	<b>5.678</b>	<b>6.407</b>	
VIII	Concepción **	9,1%	816	4,2%	46,3%	3,8%	41,9%	1.000	1.495
	Coronel **	20,7%	482	4,3%	20,7%	4,2%	20,3%	500	594
	Chiguayante **	14,0%	467	3,9%	27,7%	3,7%	26,7%	500	626
	Florida	33,0%	72	14,9%	45,1%	14,9%	45,1%	72	85
	Hualqui	32,5%	93	1,7%	5,3%	1,7%	5,3%	93	114
	Lota **	19,4%	244	7,9%	40,6%	6,9%	35,5%	320	364
	Penco **	22,4%	207	6,4%	28,5%	6,5%	29,0%	200	238
	San Pedro de la Paz **	25,5%	297	11,1%	43,5%	7,8%	30,6%	600	749
	Santa Juana	20,6%	79	12,5%	60,6%	12,5%	60,6%	79	98
	Talcahuano **	15,3%	650	4,9%	32,1%	5,3%	34,9%	550	688
	Tomé **	24,5%	228	8,2%	33,4%	8,8%	35,7%	200	232
	Hualpén **	13,1%	381	2,5%	19,1%	3,4%	26,3%	200	242

\*Tamaños corregidos por tasa de no respuesta promedio ponderado Casen 2011 – 2013

\*\* Comunas definidas como dominios de estudio.

Continúa ►

Región	Comuna	Estimador de la razón Casen 2013	Muestra Casen 2013 (viviendas)	Error absoluto	Error relativo	Error absoluto	Error relativo	Muestra objetiva	Tamaños corregidos*
				Casen 2013	Casen 2013 (%)	Casen 2015	Casen 2015	Casen 2015	Casen 2015
VIII	Lebu **	31,2%	113	6,4%	20,4%	4,8%	15,3%	200	250
	Arauco **	19,3%	169	4,7%	24,1%	4,3%	22,1%	200	241
	Cañete **	51,6%	141	20,3%	39,3%	9,4%	18,3%	650	754
	Contulmo	23,5%	84	12,2%	51,9%	12,2%	51,9%	84	99
	Curanilahue **	16,4%	215	7,8%	47,4%	5,7%	34,8%	400	493
	Los Álamos	21,9%	94	3,8%	17,5%	3,8%	17,5%	94	113
	Tirúa	32,4%	88	15,8%	48,7%	15,8%	48,7%	88	103
	Los Ángeles **	22,3%	741	11,1%	49,9%	7,8%	35,0%	1.500	1.759
	Antuco	37,8%	97	9,5%	25,2%	9,5%	25,2%	97	102
	Cabrero	35,1%	145	9,6%	27,3%	9,6%	27,3%	145	162
	Laja **	22,6%	78	6,7%	29,6%	4,2%	18,5%	200	228
	Mulchén **	25,4%	127	8,4%	33,2%	6,7%	26,4%	200	214
	Nacimiento **	32,2%	115	11,6%	36,0%	8,8%	27,3%	200	231
	Negrete	38,8%	75	7,6%	19,5%	7,6%	19,5%	75	78
	Quilaco	23,6%	63	4,3%	18,3%	4,3%	18,3%	63	72
	Quilleco	38,4%	78	14,9%	38,8%	14,9%	38,8%	78	80
	San Rosendo	25,3%	74	15,8%	62,5%	15,8%	62,5%	74	89
	Santa Bárbara	27,5%	73	11,6%	42,3%	11,6%	42,3%	73	80
	Tucapel	33,4%	85	17,5%	52,5%	17,5%	52,5%	85	90
	Yumbel	18,3%	65	10,0%	54,9%	10,0%	54,9%	65	77
	Alto Biobío	69,2%	77	22,1%	32,0%	22,1%	32,0%	77	79
	Chillán **	23,2%	743	4,8%	20,6%	5,8%	25,1%	500	577
	Bulnes	37,8%	90	17,8%	47,0%	17,8%	47,0%	90	97
	Cobquecura	37,6%	73	29,6%	78,8%	29,6%	78,8%	73	80
	Coelemu	21,0%	102	18,6%	88,6%	18,6%	88,6%	102	118
	Coihueco	44,2%	113	10,7%	24,3%	10,7%	24,3%	113	126
	Chillán Viejo **	22,1%	142	5,0%	22,6%	4,2%	19,0%	200	223
	El Carmen	41,1%	76	22,1%	53,8%	22,1%	53,8%	76	80
	Ninhue	41,4%	82	11,2%	27,1%	11,2%	27,1%	82	91
	Ñiquén	31,4%	78	12,2%	39,0%	12,2%	39,0%	78	88
	Pemuco	29,3%	86	8,3%	28,2%	8,3%	28,2%	86	95
	Pinto	35,6%	95	8,3%	23,4%	8,3%	23,4%	95	104
	Portezuelo	25,5%	83	10,1%	39,6%	10,1%	39,6%	83	95
Quillón	24,0%	96	7,9%	32,9%	7,9%	32,9%	96	104	
Quirihue	28,5%	77	12,9%	45,2%	12,9%	45,2%	77	89	
Ránquil	31,6%	77	12,2%	38,6%	12,2%	38,6%	77	92	
San Carlos	34,9%	214	10,8%	30,9%	10,8%	30,9%	214	247	
San Fabián	35,0%	83	12,7%	36,4%	12,7%	36,4%	83	93	
San Ignacio	43,8%	76	13,4%	30,6%	13,4%	30,6%	76	85	
San Nicolás	28,8%	79	9,9%	34,2%	9,9%	34,2%	79	88	
Treguaco	46,5%	76	14,9%	32,1%	14,9%	32,1%	76	85	
Yungay	35,9%	84	8,9%	24,7%	8,9%	24,7%	84	90	
<b>Total Regional</b>	<b>22,3%</b>	<b>9.438</b>	<b>1,9%</b>	<b>8,7%</b>	<b>1,8%</b>	<b>7,9%</b>	<b>11.402</b>	<b>13.666</b>	
IX	Temuco **	18,3%	1.366	3,2%	17,6%	4,0%	21,7%	900	1.170
	Carahue **	44,6%	116	11,5%	25,9%	8,8%	19,7%	200	211
	Cunco **	36,7%	99	11,0%	29,8%	7,2%	19,6%	230	268
	Curarrehue	29,4%	83	18,3%	62,3%	18,3%	62,3%	83	87
	Freire **	34,4%	143	5,3%	15,4%	4,5%	13,1%	200	223
	Galvarino	31,4%	80	10,5%	33,3%	10,5%	33,3%	80	89
	Gorbea	40,9%	90	5,1%	12,5%	5,1%	12,5%	90	100
	Lautaro **	34,7%	142	6,2%	17,8%	4,7%	13,4%	250	282
	Loncoche	24,5%	137	15,2%	62,1%	15,2%	62,1%	137	167
	Melipeuco	45,3%	82	10,4%	22,9%	10,4%	22,9%	82	89
	Nueva Imperial **	52,6%	148	9,1%	17,4%	7,9%	15,0%	200	216
	Padre Las Casas **	29,3%	382	6,1%	20,8%	6,4%	21,7%	350	393
	Perquenco	55,8%	70	12,0%	21,5%	12,0%	21,5%	70	75
Pitrufquén **	27,8%	158	15,1%	54,3%	9,4%	33,7%	410	449	

\*Tamaños corregidos por tasa de no respuesta promedio ponderado Casen 2011 – 2013

\*\* Comunas definidas como dominios de estudio.

Continúa ►

Región	Comuna	Estimador de la razón Casen 2013	Muestra Casen 2013 (viviendas)	Error absoluto	Error relativo	Error absoluto	Error relativo	Muestra objetivo	Tamaños corregidos*
				Casen 2013	Casen 2013 (%)	Casen 2015	Casen 2015	Casen 2015	Casen 2015
IX	Pucón **	22,4%	136	6,3%	28,0%	4,6%	20,6%	250	306
	Saavedra	46,5%	73	7,2%	15,5%	7,2%	15,5%	73	75
	Teodoro Schmidt	52,2%	92	7,9%	15,1%	7,9%	15,1%	92	98
	Toltén	52,1%	76	8,4%	16,2%	8,4%	16,2%	76	78
	Vilcún	33,4%	184	5,5%	16,5%	5,5%	16,5%	184	206
	Villarrica **	14,2%	284	6,9%	48,5%	4,9%	34,8%	550	616
	Cholchol	34,3%	71	25,5%	74,4%	25,5%	74,4%	71	75
	Angol **	11,3%	331	4,3%	37,6%	3,9%	34,2%	400	461
	Collipulli **	35,1%	101	3,1%	8,8%	1,7%	4,7%	350	394
	Curacautín **	26,9%	103	13,9%	51,6%	8,9%	33,1%	250	272
	Ercilla	41,0%	81	8,2%	20,0%	8,2%	20,0%	81	101
	Lonquimay	61,4%	94	14,2%	23,2%	14,2%	23,2%	94	99
	Los Sauces	43,1%	87	22,1%	51,4%	22,1%	51,4%	87	103
	Lumaco	51,9%	82	21,5%	41,4%	21,5%	41,4%	82	89
	Purén	37,4%	88	12,1%	32,3%	12,1%	32,3%	88	102
	Renaico	27,5%	86	10,7%	38,9%	10,7%	38,9%	86	97
	Traiguén **	29,0%	146	3,3%	11,2%	2,8%	9,6%	200	235
Victoria **	21,2%	193	12,7%	60,1%	7,5%	35,6%	550	668	
<b>Total Regional</b>	<b>27,9%</b>	<b>5.404</b>	<b>2,2%</b>	<b>7,9%</b>	<b>1,9%</b>	<b>7,0%</b>	<b>6.846</b>	<b>7.894</b>	
X	Puerto Montt **	14,6%	808	3,3%	22,5%	3,1%	21,3%	900	1.105
	Calbuco **	30,2%	194	6,7%	22,2%	6,6%	21,9%	200	206
	Fresia	42,3%	83	14,2%	33,6%	14,2%	33,6%	83	89
	Frutillar **	25,9%	82	16,4%	63,4%	9,1%	34,9%	270	321
	Los Muermos **	20,6%	88	16,7%	81,0%	7,4%	35,8%	450	470
	Llanquihue **	13,6%	96	6,8%	49,7%	3,8%	28,1%	300	321
	Mauñín	17,4%	117	6,2%	35,6%	6,2%	35,6%	117	123
	Puerto Varas	6,2%	173	6,2%	100,8%	6,2%	100,8%	173	219
	Castro **	11,7%	206	7,9%	67,6%	4,1%	35,4%	750	890
	Ancud **	15,1%	208	7,6%	50,2%	5,4%	35,3%	420	482
	Chonchi	16,0%	89	10,1%	63,3%	10,1%	63,3%	89	100
	Curaco de Vélez	39,2%	31	15,7%	40,0%	15,7%	40,0%	31	32
	Dalcahue	13,4%	94	7,7%	57,8%	7,7%	57,8%	94	100
	Puqueldón	8,5%	42	5,2%	61,1%	5,2%	61,1%	42	44
	Queilén	31,3%	87	16,6%	53,0%	16,6%	53,0%	87	93
	Quellón **	13,2%	156	12,7%	96,2%	6,6%	49,9%	580	699
	Quemchi	16,5%	73	2,4%	14,3%	2,4%	14,3%	73	78
Quinchao	17,6%	79	8,4%	47,7%	8,4%	47,7%	79	93	
Osorno **	16,6%	809	3,1%	18,6%	3,6%	21,6%	600	774	
Puerto Octay	16,5%	71	8,1%	49,1%	8,1%	49,1%	71	76	
Purranque	22,0%	127	21,6%	98,3%	21,6%	98,3%	127	149	
Puyehue	20,9%	97	12,2%	58,4%	12,2%	58,4%	97	108	
Río Negro **	22,3%	89	14,5%	64,7%	7,9%	35,3%	300	353	
San Juan de la Costa	48,9%	83	15,7%	32,1%	15,7%	32,1%	83	91	
San Pablo	38,3%	81	12,0%	31,3%	12,0%	31,3%	81	92	
<b>Total Regional</b>	<b>17,6%</b>	<b>4.063</b>	<b>2,1%</b>	<b>12,0%</b>	<b>1,7%</b>	<b>9,8%</b>	<b>6.097</b>	<b>7.108</b>	
XI	Coyhaique **	6,9%	1.121	2,2%	32,4%	3,4%	49,5%	480	621
	Aysén **	7,9%	489	3,3%	41,8%	3,9%	49,4%	350	439
	Cisnes	4,1%	64	3,4%	81,5%	3,4%	81,5%	64	76
	Cochrane **	8,4%	66	2,9%	34,8%	2,4%	28,2%	100	136
	Chile Chico **	4,6%	79	2,0%	42,5%	1,7%	37,8%	100	126
Río Ibáñez	6,0%	40	5,7%	94,3%	5,7%	94,3%	40	52	
<b>Total Regional</b>	<b>6,8%</b>	<b>1.859</b>	<b>1,6%</b>	<b>24,1%</b>	<b>2,1%</b>	<b>30,8%</b>	<b>1.134</b>	<b>1.450</b>	

\*Tamaños corregidos por tasa de no respuesta promedio ponderado Casen 2011 – 2013

\*\* Comunas definidas como dominios de estudio.

Continúa ►

Región	Comuna	Estimador de la razón Casen 2013	Muestra Casen 2013 (viviendas)	Error absoluto	Error relativo	Error absoluto	Error relativo	Muestra objetivo	Tamaños corregidos*
				Casen 2013	Casen 2013	Casen 2013 (%)	propuesto Casen 2015	propuesto Casen 2015	Casen 2015
XII	Punta Arenas **	3,9%	1.593	1,3%	33,8%	1,8%	47,7%	800	1.187
	Porvenir **	0,8%	86	1,8%	215,9%	1,0%	115,6%	300	334
	Natales **	17,2%	225	14,8%	85,9%	8,6%	49,8%	670	766
	<b>Total Regional</b>	<b>5,6%</b>	<b>1.904</b>	<b>3,2%</b>	<b>57,9%</b>	<b>3,3%</b>	<b>60,1%</b>	<b>1.770</b>	<b>2.287</b>
XIII	Santiago **	6,5%	398	4,0%	62,0%	3,2%	48,5%	650	1.413
	Cerrillos	6,3%	125	9,2%	146,4%	9,2%	146,4%	125	150
	Cerro Navia **	15,8%	220	6,7%	42,6%	5,6%	35,3%	320	364
	Conchalí **	17,9%	208	16,8%	93,9%	6,3%	35,0%	1.500	1.799
	El Bosque **	13,7%	226	10,2%	74,1%	4,8%	35,2%	1.000	1.287
	Estación Central **	9,9%	231	4,2%	42,0%	4,5%	45,1%	200	257
	Huechuraba	4,4%	114	1,9%	44,2%	1,9%	44,2%	114	136
	Independencia	17,3%	124	7,6%	43,8%	7,6%	43,8%	124	167
	La Cisterna	7,5%	127	4,2%	55,2%	4,2%	55,2%	127	176
	La Florida **	7,8%	502	2,5%	32,4%	2,5%	32,5%	500	782
	La Granja **	21,8%	235	6,5%	29,8%	6,3%	28,9%	250	290
	La Pintana **	17,4%	269	3,6%	20,7%	3,4%	19,6%	300	374
	La Reina	0,9%	110	1,8%	203,7%	1,8%	203,7%	110	188
	Las Condes **	1,5%	406	1,9%	132,1%	1,2%	80,2%	1.100	2.655
	Lo Barnechea	4,5%	117	5,1%	113,7%	5,1%	113,7%	117	332
	Lo Espejo	5,4%	137	4,0%	73,8%	4,0%	73,8%	137	162
	Lo Prado	6,2%	179	4,5%	71,9%	4,5%	71,9%	179	201
	Macul **	6,0%	186	2,4%	39,9%	2,3%	38,5%	200	274
	Maipú **	3,6%	829	1,9%	53,4%	2,3%	62,7%	600	839
	Nuñoa **	0,4%	283	0,6%	160,9%	0,3%	79,8%	1.150	2.825
	Pedro Aguirre Cerda **	7,3%	168	5,0%	68,8%	3,5%	47,7%	350	408
	Peñalolén **	9,3%	364	2,6%	28,1%	2,6%	27,5%	380	509
	Providencia **	3,2%	355	3,7%	114,2%	2,1%	64,9%	1.100	3.703
	Pudahuel **	8,1%	320	4,9%	61,2%	4,0%	48,9%	500	588
	Quilicura **	9,9%	260	6,8%	69,0%	4,9%	49,8%	500	575
	Quinta Normal **	9,9%	167	6,9%	69,5%	4,7%	48,0%	350	419
	Recoleta **	11,0%	232	5,0%	45,1%	3,9%	35,2%	380	497
	Renca **	8,6%	220	5,0%	58,3%	4,3%	49,9%	300	342
	San Joaquín	4,6%	122	4,6%	99,8%	4,6%	99,8%	122	187
	San Miguel **	4,7%	104	6,7%	142,6%	3,1%	65,0%	500	838
	San Ramón	8,4%	139	7,0%	83,2%	7,0%	83,2%	139	175
	Vitacura	0,2%	150	0,5%	221,7%	0,5%	221,7%	150	487
	Puente Alto **	13,6%	776	6,7%	49,2%	4,8%	35,4%	1.500	1.836
	Pirque	1,3%	71	1,6%	121,6%	1,6%	121,6%	71	94
	San José de Maipo	11,9%	56	17,5%	147,9%	17,5%	147,9%	56	97
	Colina **	4,8%	113	4,5%	93,7%	3,0%	63,0%	250	332
	Lampa	8,0%	121	4,5%	56,1%	4,5%	56,1%	121	160
	Tiltil	9,3%	78	4,6%	48,9%	4,6%	48,9%	78	92
	San Bernardo **	19,0%	389	5,6%	29,6%	5,5%	29,2%	400	540
	Buín	11,3%	93	6,4%	56,9%	6,4%	56,9%	93	113
Calera de Tango	16,9%	56	4,1%	24,4%	4,1%	24,4%	56	86	
Paine	17,2%	104	2,2%	13,0%	2,2%	13,0%	104	137	
Melipilla **	21,4%	168	4,2%	19,8%	3,7%	17,3%	220	277	
Alhué	17,5%	86	2,0%	11,4%	2,0%	11,4%	86	93	
Curacaví	4,8%	65	3,7%	76,5%	3,7%	76,5%	65	84	
María Pinto	10,5%	56	10,3%	98,3%	10,3%	98,3%	56	72	
San Pedro	34,2%	36	19,6%	57,4%	19,6%	57,4%	36	39	
Talagante **	9,8%	134	3,4%	35,1%	2,8%	28,7%	200	237	
El Monte	7,5%	71	7,6%	101,1%	7,6%	101,1%	71	84	
Isla de Maipo	16,4%	68	9,2%	56,3%	9,2%	56,3%	68	97	
Padre Hurtado	12,3%	71	13,9%	113,3%	13,9%	113,3%	71	98	
Peñaflor	8,6%	128	5,9%	68,4%	5,9%	68,4%	128	165	
<b>Total Regional</b>	<b>9,2%</b>	<b>10.367</b>	<b>1,0%</b>	<b>10,8%</b>	<b>0,8%</b>	<b>8,4%</b>	<b>17.304</b>	<b>28.132</b>	

\*Tamaños corregidos por tasa de no respuesta promedio ponderado Casen 2011 – 2013

\*\* Comunas definidas como dominios de estudio.

Continúa ►

Región	Comuna	Estimador de la razón Casen 2013	Muestra Casen 2013 (viviendas)	Error absoluto efectivo Casen 2013	Error relativo efectivo Casen 2013 (%)	Error absoluto propuesto Casen 2015	Error relativo propuesto Casen 2015	Muestra objetiva propuesta Casen 2015	Tamaños corregidos* Casen 2015
XIV	Valdivia **	14,2%	1.551	3,0%	20,9%	4,4%	31,1%	700	887
	Corral	22,9%	85	2,6%	11,3%	2,6%	11,3%	85	96
	Lanco	37,4%	136	5,5%	14,7%	5,5%	14,7%	136	160
	Los Lagos **	27,6%	174	9,3%	33,7%	5,8%	21,0%	450	499
	Máfil	30,1%	94	13,3%	44,1%	13,3%	44,1%	94	104
	Mariquina	40,3%	248	16,2%	40,3%	16,2%	40,3%	248	273
	Paillaco **	31,9%	179	11,7%	36,8%	9,4%	29,4%	280	310
	Panguipulli **	34,5%	282	10,7%	31,2%	9,4%	27,2%	370	405
	La Unión **	23,0%	318	5,2%	22,8%	4,8%	21,1%	370	411
	Futrono	22,7%	129	13,2%	58,2%	13,2%	58,2%	129	148
	Lago Ranco	41,8%	86	22,1%	52,8%	22,1%	52,8%	86	95
Río Bueno **	24,1%	362	6,2%	25,8%	6,2%	25,6%	370	387	
<b>Total Regional</b>		<b>23,1%</b>	<b>3.644</b>	<b>2,5%</b>	<b>10,6%</b>	<b>2,6%</b>	<b>11,1%</b>	<b>3.318</b>	<b>3.775</b>
XV	Arica **	14,1%	2.297	2,0%	14,5%	3,7%	26,3%	700	860
	Camarones	15,6%	38	2,5%	15,8%	2,5%	15,8%	38	42
	Putre **	29,2%	77	9,9%	34,0%	8,7%	29,9%	100	115
	<b>Total Regional</b>		<b>14,6%</b>	<b>2.412</b>	<b>2,0%</b>	<b>13,7%</b>	<b>3,4%</b>	<b>23,3%</b>	<b>838</b>

\*Tamaños corregidos por tasa de no respuesta promedio ponderado Casen 2011 – 2013

\*\* Comunas definidas como dominios de estudio.

Continúa ►

## 2. Actualización del Marco Muestral 2008

### 2.1. Antecedentes

En julio del año 2014, el Ministerio de Desarrollo Social y el Instituto Nacional de Estadísticas suscriben un convenio bajo el cual se encuentra circunscrita la actividad “Enumeración unidades muestrales en comunas específicas”, que consiste en el mejoramiento del marco muestral 2008, a partir de criterios acordados con el Ministerio de Desarrollo Social, que permite la incorporación de nuevas manzanas y actualizar el registro del total de viviendas particulares que estas contienen.

Cabe señalar que esta actualización era recomendable por la obsolescencia del marco en cuanto a que las manzanas en mayor proporción, así como las secciones, han sufrido cambios importantes debido a crecimientos y cambios de estructura como fusiones y divisiones.

Es importante también señalar que esta actualización no configura un cambio de marco propiamente tal (en que todas las unidades primarias y secundarias se actualizan de algún modo) sino que obedece a solo una actualización menor y parcial, focalizada en algunos urbanos (dentro de las comunas) que poseían, de acuerdo a información de levantamiento de varias encuestas levantadas por el INE, altos índices de solicitudes de reemplazo de manzanas por cambio de estructura y crecimiento. Esta actualización focalizada representa menos del 1,76% de las viviendas totales del marco de manzanas (2.351 manzanas de un total de 133.970 que posee el MM2008) y en cuanto a viviendas disponibles para selección, se incorporaron 63.749 de un total de 4.000.762 que poseía el marco, es decir apenas el 1,59%.

En principio, la actualización de dicho marco consistió en la enumeración de aquellas manzanas catalogadas como *no elegibles*, ya que cartográficamente tenían cambio de estructura (divisiones o fusiones de manzanas)<sup>59</sup> o cambios importantes en el total de viviendas, según el proceso de enumeración realizado para las encuestas Casen 2011, Casen 2013 y otras encuestas realizadas por el INE. Sin embargo, debido a la gran dispersión geográfica y el alto costo que implica el proceso de enumeración, se seleccionaron, en primera instancia, 59 entidades urbanas con las tasas más altas de “no elegibilidad”, correspondiente a 909 manzanas, pertenecientes a las regiones de Atacama, Coquimbo, Valparaíso, O’Higgins, La Araucanía, Los Ríos, Los Lagos y Metropolitana. Posteriormente, se incorporaron aquellas entidades urbanas de mayor envergadura en cuanto al total poblacional y número de manzanas, incorporando así 453 nuevas manzanas en las regiones de Arica y Parinacota, Tarapacá, Antofagasta, Aysén y Magallanes, alcanzando un total de 1.362 manzanas a revisar.

El MM2008 fue actualizado mediante dos métodos. Primero, se definió la posibilidad de recuperar la manzana en “gabinete”, a partir de la confrontación de la información de la manzana según el MM2008 y la información obtenida en el precenso 2011; y en segundo lugar, se recuperaron manzanas en “terreno”, lo que se realizó a partir de la enumeración de manzanas para verificar la cantidad de viviendas que estas contenían, las cuales fueron determinadas previamente según ciertos análisis cartográficos.

En el proceso de gabinete se revisaron las 1.362 manzanas catalogadas como no elegibles y se comparó, paralelamente, su estructura según la información de precenso 2011 y la imagen satelital más actualizada disponible. Si el motivo de no elegibilidad era cambio de estructura (fusión o división de manzana) y estaba registrado en precenso, se revisó la estructura y el total de viviendas con la imagen satelital, si ambos coincidían entonces la manzana no fue enviada al proceso de enumeración para su actualización y se registró como una manzana actualizada en gabinete. Si la manzana en estructura, total de viviendas y/o imagen no es coincidente con el precenso, entonces la manzana fue enviada a terreno para verificar su condición.

---

<sup>59</sup> Ver punto 2.2 de este anexo, “Fusión y División de manzanas”.

En este contexto, se determinó que 386 manzanas debían ser enviadas a terreno. Sin embargo, al revisar los casos de las manzanas con problemas, se visualizó que en varias ocasiones las manzanas colindantes presentaban las mismas características, por lo que era necesario realizar un análisis más exhaustivo y general. Con un barrido de imagen satelital se determinó incorporar una vecindad, de aquellas manzanas no elegibles y que no fueron actualizadas en gabinete, para su actualización en “terreno”, incorporando así un total de 989 nuevas manzanas a revisión.

En resumen, de un total de 24.182 manzanas contenidas en los 69 urbanos<sup>60</sup> actualizados, se revisaron 2.351 manzanas correspondientes a 9,7%. De éstas, 1.375 (58%) fueron enviadas a terreno para el proceso de enumeración, mientras que 976 fueron recuperadas en gabinete.

En la Tabla 1 se puede observar el cambio en cuanto al total de manzanas y viviendas en cada uno de los urbanos actualizados. En total se incrementó en 1,6% las manzanas y en 11% del total de viviendas de los urbanos actualizados (1,6% del total del MM2008 sin actualizar). Sin embargo, al realizar un análisis más detallado respecto al cambio en cada urbano, se observa que existen aumentos y disminuciones en el total de manzanas, mientras que en el total de viviendas se observan incrementos a excepción de los urbanos de El Salvador y Los Lagos que presentaron disminución en el total de manzanas y viviendas. El mayor incremento en términos absolutos se observó en Alto Hospicio con un total de 108 manzanas, y equivalentes a 5.827 viviendas. Cambios no menos importantes se observan en la zona norte en Arica, Iquique, Antofagasta y Calama se produce un incremento en más de 1.000 viviendas, aun cuando existe una disminución en el total de manzanas. En la zona sur, se observa que Loncoche, Pucón, Calbuco, Frutillar, Puerto Varas, Quellón, Coyhaique y Punta Arenas tienen un crecimiento de más de 1.000 viviendas, sin embargo, Loncoche, Frutillar, Coyhaique y Punta Arenas presentan los mayores incrementos de manzanas respectivamente.

---

<sup>60</sup> Se repite explicación de contexto para urbano, en el ámbito de la actualización y/o enumeración de conglomerados. “En el Marco Muestral del INE, según la división política administrativa, las comunas están formadas por entidades urbanas denominadas “urbanos” y entidades rurales (localidades y entidades). Como por ejemplo, la parte urbana de la comuna de los Vilos está formada por los urbanos de Los Vilos, Pichidangui y Quilimarí. De estos urbanos, sólo las UPM de Los Vilos están incorporadas en el Marco de Manzanas o MM2008, mientras que las UPM de los otros dos urbanos, están incorporadas en el Marco de Secciones o MS2002.”.

**Tabla 2.1. Total de viviendas y manzanas vigentes en MM2o8 y MM2oo8 actualizado con foco Casen, en urbanos actualizados**

Región	Comuna	Urbano	MM2oo8		MM2oo8 foco Casen		Variación	
			Nº de manzanas	Nº de viviendas	Nº de manzanas	Nº de viviendas	Nº de manzanas	Nº de viviendas
<b>Total 69 urbanos actualizados</b>			<b>24.182</b>	<b>580.367</b>	<b>24.581</b>	<b>644.116</b>	<b>1,6%</b>	<b>11,0%</b>
15	Arica	Arica	2.415	52.865	2.424	58.057	0,4%	9,8%
1	Iquique	Iquique	1.344	47.049	1.333	49.616	-0,8%	5,5%
1	Alto Hospicio	Alto Hospicio	722	20.822	830	26.649	15,0%	28,0%
2	Antofagasta	Antofagasta	3.213	80.425	3.199	85.395	-0,4%	6,2%
2	Calama	Calama	1.632	36.318	1.712	39.776	4,9%	9,5%
2	Tocopilla	Tocopilla	325	7.098	343	8.066	5,5%	13,6%
3	Tierra Amarilla	Tierra Amarilla	145	2.363	142	2.941	-2,1%	24,5%
3	Diego de Almagro	Diego de Almagro	210	2.899	191	2.906	-9,0%	0,2%
3	Diego de Almagro	El Salvador	198	2.945	186	2.873	-6,1%	-2,4%
3	Freirina	Freirina	122	1.056	122	1.146	0,0%	8,5%
4	Vicuña	Vicuña	183	4.715	198	5.506	8,2%	16,8%
4	Salamanca	Salamanca	167	3.952	149	4.341	-10,8%	9,8%
4	Combarbalá	Combarbalá	124	2.131	124	2.352	0,0%	10,4%
4	Monte Patria	El Palqui	102	1.788	99	1.803	-2,9%	0,8%
4	Monte Patria	Monte Patria	119	1.831	132	1.999	10,9%	9,2%
5	Valparaíso	Placilla de Peñuelas	307	7.933	341	10.896	11,1%	37,4%
5	Valparaíso	Valparaíso	2.974	81.982	2.982	84.572	0,3%	3,2%
5	Casablanca	Casablanca	156	4.990	160	5.966	2,6%	19,6%
5	Puchuncaví	Las Ventanas	202	5.694	204	7.291	1,0%	28,0%
5	Puchuncaví	Puchuncaví	70	1.365	74	1.710	5,7%	25,3%
5	Calle Larga	Calle Larga	55	1.773	59	2.223	7,3%	25,4%
5	Rinconada	Rinconada	61	1.801	93	2.510	52,5%	39,4%
5	San Esteban	San Esteban	85	2.870	111	3.978	30,6%	38,6%
5	Cabildo	Cabildo	185	3.606	178	3.629	-3,8%	0,6%
5	Hijuelas	Hijuelas	83	2.378	73	2.684	-12,0%	12,9%
5	Nogales	El Melón	155	2.730	140	2.881	-9,7%	5,5%
5	Nogales	Nogales	125	2.606	121	2.820	-3,2%	8,2%
5	Algarrobo	Algarrobo	352	8.727	387	10.400	9,9%	19,2%
5	Cartagena	Cartagena	623	11.727	589	12.318	-5,5%	5,0%
5	El Quisco	El Quisco	731	12.569	724	13.012	-1,0%	3,5%
5	Las Cruces	Las Cruces	261	5.845	259	5.945	-0,8%	1,7%
5	Catemu	Catemu	105	1.989	97	2.201	-7,6%	10,7%
5	Llailay	Llailay	215	5.230	230	5.989	7,0%	14,5%
5	Putendo	Putendo	109	2.288	109	2.478	0,0%	8,3%
5	Santa María	Santa María	54	2.093	61	2.578	13,0%	23,2%
5	Olmué	Olmué	193	3.968	196	4.254	1,6%	7,2%

Continúa

**Tabla 2.1. Total de viviendas y manzanas vigentes en MM2o8 y MM2oo8 foco Casen, en urbanos actualizados**

Región	Comuna	Urbano	MM2oo8		MM2oo8 Foco Casen		Variación	
			Nº de manzanas	Nº de viviendas	Nº de manzanas	Nº de viviendas	Nº de manzanas	Nº de viviendas
6	Codegua	Codegua	57	1.376	56	1.588	-1,8%	15,4%
6	Doñihue	Doñihue	112	2.177	98	2.391	-12,5%	9,8%
6	Doñihue	Lo Miranda	90	2.229	100	2.828	11,1%	26,9%
6	Graneros	Graneros	205	6.889	216	7.621	5,4%	10,6%
6	Las Cabras	Las Cabras	79	1.969	67	2.335	-15,2%	18,6%
6	San Francisco de Mostazal	San Francisco de Mostazal	121	3.403	110	3.572	-9,1%	5,0%
6	Olivar Alto	Gultro	51	1.416	48	1.517	-5,9%	7,1%
6	Olivar Alto	Olivar Alto	22	686	14	823	-36,4%	20,0%
6	Peumo	Peumo	114	2.096	81	2.619	-28,9%	25,0%
6	Pichidegua	Pichidegua	66	1.529	52	1.800	-21,2%	17,7%
6	Quinta de Tilcoco	Quinta de Tilcoco	74	1.635	68	1.942	-8,1%	18,8%
6	Pichilemu	Pichilemu	430	5.678	446	6.830	3,7%	20,3%
6	Litueche	Litueche	53	928	52	1.119	-1,9%	20,6%
6	Chépica	Chépica	85	1.556	73	1.989	-14,1%	27,8%
6	Chimbarongo	Chimbarongo	186	4.572	172	5.154	-7,5%	12,7%
9	Carahue	Carahue	121	3.418	118	3.918	-2,5%	14,6%
9	Cunco	Cunco	85	2.282	87	2.357	2,4%	3,3%
9	Loncoche	Loncoche	208	4.389	267	5.867	28,4%	33,7%
9	Pucón	Pucón	245	6.041	222	7.429	-9,4%	23,0%
9	Purén	Purén	119	2.771	127	2.843	6,7%	2,6%
9	Renaico	Renaico	83	1.835	84	1.989	1,2%	8,4%
14	Los Lagos	Los Lagos	141	2.902	140	2.888	-0,7%	-0,5%
14	Panguipulli	Panguipulli	124	3.114	142	3.655	14,5%	17,4%
14	Futrono	Futrono	53	1.868	59	2.373	11,3%	27,0%
14	Futrono	Llifén	9	288	8	379	-11,1%	31,6%
10	Calbuco	Calbuco	95	2.983	114	4.270	20,0%	43,1%
10	Fresia	Fresia	82	2.135	98	2.592	19,5%	21,4%
10	Frutillar	Frutillar	107	3.331	147	4.569	37,4%	37,2%
10	Puerto Varas	Puerto Varas	290	6.658	300	8.438	3,4%	26,7%
10	Quellón	Quellón	108	3.695	134	4.704	24,1%	27,3%
11	Coyhaique	Coyhaique	684	14.584	723	17.055	5,7%	16,9%
12	Punta Arenas	Punta Arenas	1.703	36.163	1.738	39.500	2,1%	9,2%
13	Pirque	Pirque	58	1.350	48	1.401	-17,2%	3,8%

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

El MM2oo8 clasifica las manzanas en 5 grupos de tamaño de acuerdo al número total de viviendas en la manzana. Más de 70% de las manzanas en el marco tienen entre 8 y 44 viviendas, mientras que el grupo de tamaño más grande corresponde a manzanas con 155 o más viviendas, las cuales representan 2% del total de las manzanas en el marco (ver Tabla III.4 del capítulo III, sección 4, Estratificación del marco muestral. En la Tabla 2.2 que le sigue se puede apreciar la muestra seleccionada de manzanas y sus respectivas viviendas).

**Tabla 2.2. Total de viviendas y manzanas vigentes en MM2o8 y MM2oo8 actualizado con foco Casen, según región**

Región	Viviendas		Manzanas		Variación	
	MM2oo8	MM2oo8 foco Casen	MM2oo8	MM2oo8 foco Casen	Viviendas	Manzanas
<b>Total País</b>	<b>4.000.762</b>	<b>4.064.511</b>	<b>133.360</b>	<b>133.759</b>	<b>1,6%</b>	<b>0,3%</b>
I de Tarapacá	70.162	78.556	2.198	2.295	12,0%	4,4%
II de Antofagasta	131.946	141.342	5.614	5.698	7,1%	1,5%
III de Atacama	74.686	75.289	4.000	3.966	0,8%	-0,9%
IV de Coquimbo	152.956	154.540	6.817	6.824	1,0%	0,1%
V de Valparaíso	507.778	523.949	19.259	19.346	3,2%	0,5%
VI de O'Higgins	159.390	165.379	5.881	5.789	3,8%	-1,6%
VII del Maule	185.055	185.055	6.703	6.703	0,0%	0,0%
VIII del Biobío	445.810	445.810	16.492	16.492	0,0%	0,0%
IX de La Araucanía	170.383	174.050	7.158	7.202	2,2%	0,6%
X de Los Lagos	136.763	142.534	5.221	5.332	4,2%	2,1%
XI de Aysén	19.782	22.253	990	1.029	12,5%	3,9%
XII de Magallanes y	42.374	45.711	2.038	2.073	7,9%	1,7%
XIII Metropolitana	1.784.022	1.784.073	46.027	46.017	0,0%	0,0%
XIV de Los Ríos	66.790	67.913	2.547	2.569	1,7%	0,9%
XV de Arica y Parinacota	52.865	58.057	2.415	2.424	9,8%	0,4%

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

**Tabla 2.3. Listado de centros urbanos actualizados en MM2008**

Región	Comuna	Urbano	Región	Comuna	Urbano
15	Arica	Arica	6	Codegua	Codegua
1	Iquique	Iquique	6	Doñihue	Doñihue
1	Alto Hospicio	Alto Hospicio	6	Doñihue	Lo Miranda
2	Antofagasta	Antofagasta	6	Graneros	Graneros
2	Calama	Calama	6	Las Cabras	Las Cabras
2	Tocopilla	Tocopilla	6	San Francisco de Mostazal	San Francisco de Mostazal
3	Tierra Amarilla	Tierra Amarilla	6	Olivar Alto	Gultro
3	Diego de Almagro	Diego de Almagro	6	Olivar Alto	Olivar Alto
3	Diego de Almagro	El Salvador	6	Peumo	Peumo
3	Freirina	Freirina	6	Pichidegua	Pichidegua
4	Vicuña	Vicuña	6	Quinta de Tilcoco	Quinta de Tilcoco
4	Salamanca	Salamanca	6	Pichilemu	Pichilemu
4	Combarbalá	Combarbalá	6	Litueche	Litueche
4	Monte Patria	El Palqui	6	Chépica	Chépica
4	Monte Patria	Monte Patria	6	Chimbarongo	Chimbarongo
5	Valparaíso	Placilla de Peñuelas	9	Carahue	Carahue
5	Valparaíso	Valparaíso	9	Cunco	Cunco
5	Casablanca	Casablanca	9	Loncoche	Loncoche
5	Puchuncaví	Las Ventanas	9	Pucón	Pucón
5	Puchuncaví	Puchuncaví	9	Purén	Purén
5	Calle Larga	Calle Larga	9	Renaico	Renaico
5	Rinconada	Rinconada	14	Los Lagos	Los Lagos
5	San Esteban	San Esteban	14	Panguipulli	Panguipulli
5	Cabildo	Cabildo	14	Futrono	Futrono
5	Hijuelas	Hijuelas	14	Futrono	Llifen
5	Nogales	El Melón	10	Calbuco	Calbuco
5	Nogales	Nogales	10	Fresia	Fresia
5	Algarrobo	Algarrobo	10	Frutillar	Frutillar
5	Cartagena	Cartagena	10	Puerto Varas	Puerto Varas
5	El Quisco	El Quisco	10	Quellón	Quellón
5	Las Cruces	Las Cruces	11	Coyhaique	Coyhaique
5	Catemu	Catemu	12	Punta Arenas	Punta Arenas
5	Llailay	Llailay	13	Pirque	Pirque
5	Putendo	Putendo			
5	Santa María	Santa María			
5	Olmué	Olmué			

**Tabla 2.4. Listado de comunas cuyas áreas RAU fueron seleccionadas desde el MS2002**

N°	Nombre Comuna	N°	Nombre Comuna
1	Alhué	45	Pelarco
2	Alto Biobío	46	Pelluhue
3	Antuco	47	Pemuco
4	Canela	48	Pencahue
5	Chile Chico	49	Peralillo
6	Cholchol	50	Perquenco
7	Chonchi	51	Petorca
8	Cisnes	52	Pica
9	Cobquecura	53	Pinto
10	Cochrane	54	Placilla
11	Coínco	55	Portezuelo
12	Colbún	56	Porvenir
13	Coltauco	57	Puerto Octay
14	Contulmo	58	Punitaqui
15	Corral	59	Putre
16	Curarrehue	60	Puyehue
17	Curepto	61	Queilén
18	Dalcahue	62	Quemchi
19	El Carmen	63	Quilaco
20	Empedrado	64	Quilleco
21	Ercilla	65	Quinchao
22	Galvarino	66	Ránquil
23	La Estrella	67	Rauco
24	La Higuera	68	Retiro
25	Lago Ranco	69	Río Claro
26	Licantén	70	Romeral
27	Lolol	71	Saavedra
28	Lonquimay	72	Sagrada Familia
29	Los Sauces	73	San Fabián
30	Lumaco	74	San Ignacio
31	Máfil	75	San Juan de La Costa
32	Malloa	76	San Nicolás
33	María Pinto	77	San Pablo
34	Maule	78	San Pedro de Atacama
35	Mauñín	79	San Rafael
36	Melipeuco	80	San Rosendo
37	Navidad	81	Santo Domingo
38	Negrete	82	Teodoro Schmidt
39	Ninhue	83	Tirúa
40	Ñiquén	84	Toltén
41	Palmilla	85	Treguaco
42	Panquehue	86	Vichuquén
43	Papudo	87	Vilcún
44	Paredones	88	Yerbas Buenas
		89	Zapallar

**Tabla 2.5. Listado de comunas cuyas áreas RAU fueron seleccionadas desde el MM2008.**

N°	Nombre Comuna	N°	Nombre Comuna	N°	Nombre Comuna
1	Algarrobo	45	Hualqui	89	Pozo Almonte
2	Ancud	46	Huasco	90	Puchuncaví
3	Andacollo	47	Illapel	91	Pucón
4	Arauco	48	Isla de Maipo	92	Puerto Varas
5	Bulnes	49	La Ligua	93	Purén
6	Cabildo	50	La Unión	94	Purranque
7	Cabrero	51	Laja	95	Putendo
8	Calbuco	52	Lampa	96	Quellón
9	Caldera	53	Lanco	97	Quillón
10	Calera de Tango	54	Las Cabras	98	Quinta de Tilcoco
11	Calle Larga	55	Lautaro	99	Quintero
12	Cañete	56	Lebu	100	Quirihue
13	Carahue	57	Limache	101	Renaico
14	Casablanca	58	Litueche	102	Rengo
15	Castro	59	Llailay	103	Requínoa
16	Catemu	60	Llanquihue	104	Rinconada
17	Cauquenes	61	Loncoche	105	Río Bueno
18	Chanco	62	Longaví	106	Río Negro
19	Chañaral	63	Los Alamos	107	Salamanca
20	Chépica	64	Los Lagos	108	San Carlos
21	Chimbarongo	65	Los Muermos	109	San Clemente
22	Codegua	66	Los Vilos	110	San Esteban
23	Coelemu	67	Marchihue	111	San Javier
24	Coihueco	68	María Elena	112	San José de Maipo
25	Collipulli	69	Mariquina	113	San Vicente
26	Combarbalá	70	Mejillones	114	Santa Bárbara
27	Constitución	71	Molina	115	Santa Cruz
28	Cunco	72	Monte Patria	116	Santa Juana
29	Curacautín	73	Mostazal	117	Santa María
30	Curacaví	74	Mulchén	118	Taltal
31	Curanilahue	75	Nacimiento	119	Teno
32	Diego De Almagro	76	Nancagua	120	Tierra Amarilla
33	Doñihue	77	Natales	121	Tiltil
34	El Monte	78	Nogales	122	Tocopilla
35	El Quisco	79	Nueva Imperial	123	Traiguén
36	Florida	80	Olmué	124	Tucapel
37	Freire	81	Paillaco	125	Victoria
38	Freirina	82	Paine	126	Vicuña
39	Fresia	83	Panguipulli	127	Villa Alegre
40	Frutillar	84	Parral	128	Villarrica
41	Futrono	85	Peumo	129	Yumbel
42	Gorbea	86	Pichidegua	130	Yungay
43	Graneros	87	Pichilemu		
44	Hualañé	88	Pitrufquen		

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

## 2.2. Fusión y División de manzanas

A través del tiempo, las manzanas sufren cambios tanto en la forma como en la composición al interior, lo que origina problemas en terreno desde el momento de la identificación, que no coincide con la original del marco. Estos cambios se pueden clasificar como estructurales (fusión o división de ellas) y tienen que ver con el cambio de los límites o polígonos cartográficos de las manzanas con respecto a los que tenían en el marco originalmente.

También los cambios en la composición interna pueden ser estructurales aún cuando no existan cambios en los polígonos cartográficos, como por ejemplo, cuando se construye un edificio en una manzana o cuando por medio de una catástrofe natural como un terremoto se destruyen viviendas quedando inhabitables.

Generalmente, el aumento o disminución de viviendas particulares al interior de una manzana, por cambio de uso o destino, son considerados no estructurales, aunque se consideraran para el cálculo de las probabilidades de selección, que requiere del número de viviendas particulares disponibles o seleccionables.

Para mayor claridad, a continuación, se da un ejemplo de fusión y uno de división de las manzanas.

### Fusión de manzanas

Manzanas vecinas o contiguas que no se encuentran definidas por elementos permanentes y/o reconocibles en terreno las cuales se estructuran como una sola manzana. En el ejemplo, según el MM2008 existían las manzanas 14, 19, 20, 30, 52. Sin embargo, a partir de la actualización del marco, en el MM2008 foco Casen, todas ellas conforman la nueva manzana 65, y por lo tanto salen del marco mejorado quedando solo la manzana 65.

Figura 2.1. Ejemplo de manzana fusionada en MM2008 foco Casen



Fuente: Informe sobre actualización de manzanas para mejora del MM2008, Instituto Nacional de Estadísticas.

### División de manzanas

Manzana subdividida principalmente por crecimiento urbano y/o modificaciones viales. En el ejemplo, en el MM2008 se encontraba la manzana 31, sin embargo, al actualizarlo se observó que la manzana fue subdividida conformando 4 nuevas manzanas: 48, 49, 50, 51. En el MM2008 foco Casen ya no se encontrará la manzana 31, en su lugar se encuentran las 4 nuevas manzanas.

Figura 2.2: Ejemplo de manzana dividida en MM2008 foco Casen



Fuente: Informe sobre actualización de manzanas para mejora del MM2008, Instituto Nacional de Estadísticas.

### 3. Proceso de Enumeración y actualización de la muestra

#### 3.1. Antecedentes y Objetivos

En el marco del convenio entre el Instituto Nacional de Estadísticas (INE) y el Ministerio de Desarrollo Social, se realizó durante el año 2015, la enumeración de viviendas en manzanas y secciones rurales seleccionadas en la muestra Casen 2015.

Entre las actividades desarrolladas en el levantamiento, se consideraron la enumeración de la totalidad de viviendas y otros usos de cada unidad elegida, de una muestra de manzanas y secciones rurales y urbanas, pertenecientes a las comunas seleccionadas, por lo tanto, se debieron aplicar todos los conceptos asociados al proceso INE de enumeración de viviendas y actualización de la cartografía.

La enumeración y actualización de viviendas, corresponde a la etapa previa a la selección de la muestra de viviendas que se desea levantar, que se desarrolla en áreas geográficas determinadas, en la instancia de selección de primera etapa de las unidades de información por parte del departamento de Investigación y Desarrollo (I&D) del INE.

Es importante tener presente que las áreas geográficas sufren constantemente cambios, tanto en su conformación como en sus características. Mediante esta etapa es posible obtener información actualizada, detallada y completa en cuanto al número de viviendas y hogares, a objeto que el departamento I&D, realice las correcciones pertinentes de manera oportuna, antes de la aplicación de la muestra efectiva de los respectivos estudios.

#### Objetivos específicos de la enumeración de viviendas

- Disponer de información actualizada a nivel nacional con criterio único y técnico del área geográfica seleccionada.
- Disponer de información actualizada de viviendas, hogares y edificaciones de otros usos, tanto en áreas urbanas y rurales.
- Disponer de información de viviendas particulares ocupadas tanto en áreas urbanas y rurales.

- Determinar la existencia de hogares en las viviendas ocupadas particulares.
- Caracterizar aquellas viviendas que no cuentan con identidad domiciliaria, a objeto que los encuestadores las identifiquen con facilidad al momento de la entrevista.

### 3.2. Conceptos básicos considerados para la enumeración de viviendas

#### Enumeración

Es una operación de registro sistemático de viviendas en un área geográfica específica con el fin de identificar y caracterizar el segmento del universo en el cual se realizará la selección de viviendas. Para realizar esta tarea, cada enumerador o supervisor debe conocer los conceptos básicos, a objeto de obtener y registrar información en un formulario y lograr la caracterización fiel del área representada en un croquis, para que cualquier supervisor y/o encuestador pueda ubicarlas posteriormente cuando deba volver a ella a realizar la supervisión o las encuestas.

#### Ámbito urbano: Manzana

Es un área geográfica delimitada o todo conjunto aislado de edificaciones contiguas o separadas, que queda circunscrito por calles de uso público o eventualmente en alguna de sus partes por caminos; línea férrea; canal o accidentes geográficos, tales como: cerro, estero, playa, río, etc. Es la mínima unidad geográfica en el ámbito urbano, considerada como unidad de observación para estudios estadísticos.

#### Ámbito urbano: Sección urbana

Conglomerado de manzanas que contiene aproximadamente 150 a 200 viviendas o eventualmente más, con límites previamente predeterminados. Es la mínima unidad geográfica en el ámbito urbano, considerada como unidad de observación para estudios estadísticos.

#### Ámbito rural: Sección

Área que contiene aproximadamente de 100 a 150 viviendas o eventualmente más, que abarcan una o más localidades pobladas o parte de una localidad poblada.

Es la mínima unidad geográfica en el ámbito rural, considerada como unidad de observación para estudios estadísticos.

#### Ámbito rural: Localidad poblada

Es todo lugar geográfico, que se identifica con un nombre propio de carácter genérico. En una Localidad Poblada pueden existir una o varias **Entidades de Población**.

#### Entidad de población

Es todo lugar poblado, inserto dentro de una Localidad y que, se diferencia de ésta última por su ámbito de referencia más específico, vale decir, más restringido. Las Entidades de Población se han clasificado en categorías factibles de distinguir dentro de una Localidad.

#### Vivienda

Es toda edificación construida, convertida o dispuesta para el alojamiento permanente o temporal de personas, así como cualquier clase de albergue, fijo o móvil, ocupado como lugar de residencia.

#### Vivienda particular

La vivienda particular para ser considerada como tal debe, necesariamente, disponer de ACCESO INDEPENDIENTE para sus moradores. Se considera que la vivienda tiene acceso independiente cuando sus ocupantes NO necesitan transitar por recintos que pertenezcan a otra vivienda particular (living, comedor, cocina, dormitorios, etc.) para entrar a las dependencias que ocupan. El acceso debe ser directamente desde la calle o transitando por un patio, corredor, pasillo, galería, etc., de manera de no interrumpir la privacidad de los moradores de otra vivienda particular.

Se debe tener presente que en una edificación, se puede encontrar más de una vivienda, siempre y cuando sus moradores cuenten con acceso independiente en la condición descrita anteriormente.

#### **Hogar particular**

El hogar está constituido por una o más personas, que unidas o no por vínculos de parentesco, ocupan completamente una vivienda o parte de ella y se benefician de un mismo presupuesto, principalmente para alimentación.

#### **Vivienda particular con más de un hogar particular**

Del estudio de las definiciones anteriores se establece que en algunas viviendas particulares se pueden encontrar dos o más hogares particulares. Para que se cumpla esta situación se debe tener presente que ambos hogares tengan un acceso común y distinto presupuesto principalmente para alimentación, es decir, NO tienen dependencia económica entre ellos. Estas condiciones se deben cumplir en forma conjunta, para ser considerados como hogares separados dentro de una misma vivienda. Recuerde que una vivienda particular necesariamente debe tener acceso independiente.

### **3.3. Procesamiento de la información y aseguramiento de calidad**

Concluida la etapa de terreno, toda la información recogida, (registros en los formularios conteniendo los antecedentes de las viviendas particulares ocupadas, desocupadas, de temporada, colectiva y edificaciones de otros usos (Industrias, Comercio y servicios), tanto para el ámbito urbano como rural se somete a una exhaustiva revisión y análisis de consistencia, como también a la completitud del llenado de cada campo del formulario.

Este proceso de revisión se realiza en todas las sedes regionales y cuyas tareas son controladas por INE central mediante un sistema informático de gestión, el que entrega los antecedentes de avance de cada etapa de los procesos y los contenidos de las bases en rutinas diarias.

Esta revisión, la cual también es aprobada en la Sede Central del INE, consiste en comprobar, que toda la información contenida en los respectivos formularios recepcionados desde las regiones, esté contenida en los planos y tengan correspondencia con las Localidades y Entidades de población de dicha cartografía.

Se verifica además que las observaciones consignadas dispongan de información suficiente con descripción y/o ubicación de las viviendas de tal manera sean identificadas de manera inequívoca por el personal de campo de Casen 2015.

Concluida esta etapa, finalmente la unidad de I&D del INE, contará con información de la totalidad de las viviendas particulares ocupadas, de cada unidad de primera etapa seleccionada para la muestra, de tal forma concluir este producto con la entrega de la selección de las unidades de segunda etapa para la versión Casen 2015.

### **3.4. Cuadro de logros de manzanas y secciones**

El siguiente cuadro muestra el total de manzanas visitadas para realizar la respectiva enumeración y actualización de viviendas, donde en teoría, según MM2008 existían 379.199 viviendas y se registraron al cierre del trabajo de campo un total de 426.166 viviendas, lo que significó un incremento de 46.961 viviendas nuevas que fueron incorporadas al MM2008.

Para el caso de secciones urbanas, las viviendas se incrementaron en 13.062 respecto a las teóricas y fueron incorporadas al MS2002.

Para el caso de secciones rurales, las viviendas se incrementaron en 58.327 respecto a las teóricas y fueron incorporadas al MS2002.

El detalle de los aumentos en cantidad y porcentaje, se muestran en las tablas siguientes.

**Tabla 3.1. Distribución regional de Total de manzanas y viviendas asociadas a la enumeración de la muestra Casen 2015, en MM2008.**

Región	Total manzanas Rurales en muestra Casen 2015	Viviendas teóricas en manzanas de la muestra Según MM2008	Viviendas seleccionadas en manzanas de muestra de Casen 2015	Viviendas Procesadas después de enumeración de manzanas	Aumento de viviendas procesadas e incorporadas al MM2008 respecto a las teóricas	Porcentaje de viviendas procesadas respecto a las teóricas según MM2008	Registros Procesados (incluye todo tipo de vivienda)
<b>Total País</b>	<b>10.201</b>	<b>379.199</b>	<b>82.174</b>	<b>426.166</b>	<b>46.967</b>	<b>12,4%</b>	<b>477.286</b>
I de Tarapacá	312	13.151	2.508	16.735	3.584	27,3%	18.361
II de Antofagasta	347	9.578	2.370	11.084	1.506	15,7%	12.422
III de Atacama	765	18.003	5.070	21.199	3.196	17,8%	24.087
IV de Coquimbo	539	14.280	3.523	15.500	1.220	8,5%	17.334
V de Valparaíso	1.203	42.031	9.578	48.101	6.070	14,4%	53.763
VI de O'Higgins	580	21.895	4.835	22.895	1.000	4,6%	26.059
VII del Maule	461	14.516	3.439	14.997	481	3,3%	17.238
VIII del Biobío	1.308	42.053	9.630	45.628	3.575	8,5%	51.842
IX de la Araucanía	717	18.208	4.486	19.887	1.679	9,2%	22.928
X de Los Lagos	524	15.838	3.955	17.244	1.406	8,9%	20.406
XI de Aysén	164	3.610	941	4.115	505	14,0%	4.912
XII de Magallanes	315	7.050	1.784	7.858	808	11,5%	8.784
XIII Metropolitana	2.572	147.732	27.325	167.904	20.172	13,7%	184.239
XIV de Los Ríos	283	7.787	1.954	9.246	1.459	18,7%	10.734
XV de Arica y Parinacota	111	3.467	776	3.773	306	8,8%	4.177

Fuente: Elaborado por el Instituto Nacional de Estadísticas

**Tabla 3.2. Distribución regional de Total de secciones y viviendas asociadas a la enumeración de la muestra Casen 2015, en MS2002, área urbana.**

Región	Total Secciones Rurales en muestra Casen 2015	Viviendas en secciones de la muestra Según MS2002	Viviendas seleccionadas en secciones de muestra de Casen 2015	Viviendas Procesadas después de enumeración de secciones	Aumento de viviendas procesadas e incorporadas al MS2002 respecto a las teóricas	Porcentaje de viviendas procesadas respecto a las teóricas según MS2002	Registros Procesados (incluye todo tipo de vivienda)
<b>Total País</b>	<b>258</b>	<b>20.725</b>	<b>5.284</b>	<b>33.787</b>	<b>13.062</b>	<b>63,0%</b>	<b>43.662</b>
I de Tarapacá	4	268	82	372	104	38,8%	471
II de Antofagasta	3	191	61	585	394	206,3%	1.102
III de Atacama	-	-	-	-	-	-	-
IV de Coquimbo	7	560	146	1.167	607	108,4%	1.481
V de Valparaíso	15	1.133	308	1.985	852	75,2%	2.600
VI de O'Higgins	35	2.665	796	4.266	1.601	60,1%	5.336
VII del Maule	43	3.701	888	6.527	2.826	76,4%	8.188
VIII del Biobío	54	4.374	1.032	6.911	2.537	58,0%	8.736
IX de la Araucanía	33	2.758	690	3.994	1.236	44,8%	5.379
X de Los Lagos	27	2.271	507	3.490	1.219	53,7%	4.484
XI de Aysén	9	570	226	759	189	33,2%	1.006
XII de Magallanes	12	702	228	1.140	438	62,4%	1.474
XIII Metropolitana	4	615	92	1.133	518	84,2%	1.437
XIV de Los Ríos	8	694	168	1.034	340	49,0%	1.341
XV de Arica y Parinacota	4	223	60	424	201	90,1%	627

Fuente: Elaborado por el Instituto Nacional de Estadísticas

**Tabla 3.3. Distribución regional de Total de secciones y viviendas asociadas a la enumeración de la muestra Casen 2015, en MS2002, área rural.**

Región	Total Secciones Rurales en muestra Casen 2015	Viviendas en secciones de la muestra Según MS2002	Viviendas seleccionadas en secciones de muestra de Casen 2015	Viviendas Procesadas después de enumeración de secciones	Aumento de viviendas procesadas e incorporadas al MS2002 respecto a las teóricas	Porcentaje de viviendas procesadas respecto a las teóricas según MS2002	Registros Procesados (incluye todo tipo de vivienda)
<b>Total País</b>	<b>968</b>	<b>73.978</b>	<b>20.474</b>	<b>132.305</b>	<b>58.327</b>	<b>78,8%</b>	<b>144.566</b>
I de Tarapacá	21	1.047	556	2.916	1.869	178,5%	3.189
II de Antofagasta	12	568	308	2.216	1.648	290,1%	2.726
III de Atacama	22	1.095	395	2.512	1.417	129,4%	2.751
IV de Coquimbo	48	3.743	918	6.962	3.219	86,0%	7.564
V de Valparaíso	91	6.471	1.905	15.359	8.888	137,4%	16.687
VI de O'Higgins	130	9.980	2.837	17.467	7.487	75,0%	19.020
VII del Maule	107	8.605	2.080	11.544	2.939	34,2%	12.817
VIII del Biobío	152	11.316	3.004	18.916	7.600	67,2%	20.304
IX de la Araucanía	136	11.020	2.718	18.771	7.751	70,3%	20.873
X de Los Lagos	117	9.484	2.646	15.255	5.771	60,8%	16.632
XI de Aysén	11	569	283	1.022	453	79,6%	1.164
XII de Magallanes	9	498	275	1.802	1.304	261,8%	1.882
XIII Metropolitana	36	3.734	715	6.854	3.120	83,6%	7.608
XIV de Los Ríos	67	5.388	1.653	9.458	4.070	75,5%	9.922
XV de Arica y Parinacota	9	460	181	1.251	791	172,0%	1.427

Fuente: Elaborado por el Instituto Nacional de Estadísticas

A nivel general, tanto para el ámbito urbano como rural, las viviendas se incrementaron en 118.356 unidades lo que representa una variación de 25% en relación al total nacional de viviendas contenidas en los 11.247 conglomerados contenidos en la muestra Casen 2015. Con respecto al marco completo, MM2008 y MS2002 juntos, este incremento es de 2,5% (118.336 / 4.6793.623).

**Tabla 3.4: Total de viviendas encuesta Casen 2015**

Región	Total conglomerados Rurales en muestra Casen 2015	Viviendas en conglomerados de la muestra Según MS2002	Viviendas seleccionadas en conglomerados de muestra de Casen 2015	Viviendas Procesadas después de enumeración de conglomerados	Aumento de viviendas procesadas e incorporadas al Marco respecto a las teóricas	Porcentaje de viviendas procesadas respecto a las teóricas según Marco	Registros Procesados (incluye todo tipo de vivienda)
<b>Total País</b>	<b>11.427</b>	<b>473.902</b>	<b>107.932</b>	<b>592.258</b>	<b>118.356</b>	<b>25,0%</b>	<b>665.514</b>
I de Tarapacá	337	14.466	3.146	20.023	5.557	38,4%	22.021
II de Antofagasta	362	10.337	2.739	13.885	3.548	34,3%	16.250
III de Atacama		19.098	5.465	23.711	4.613	24,2%	26.838
IV de Coquimbo	594	18.583	4.587	23.629	5.046	27,2%	26.379
V de Valparaíso	1309	49.635	11.791	65.445	15.810	31,9%	73.050
VI de O'Higgins	745	34.540	8.468	44.628	10.088	29,2%	50.415
VII del Maule	611	26.822	6.407	33.068	6.246	23,3%	38.243
VIII del Biobío	1514	57.743	13.666	71.455	13.712	23,7%	80.882
IX de la Araucanía	886	31.986	7.894	42.652	10.666	33,3%	49.180
X de Los Lagos	668	27.593	7.108	35.989	8.396	30,4%	41.522
XI de Aysén	184	4.749	1.450	5.896	1.147	24,2%	7.082
XII de Magallanes	336	8.250	2.287	10.800	2.550	30,9%	12.140
XIII Metropolitana	2612	152.081	28.132	175.891	23.810	15,7%	193.284
XIV de Los Ríos	358	13.869	3.775	19.738	5.869	42,3%	21.997
XV de Arica y Parinacota	124	4.150	1.017	5.448	1.298	31,3%	6.231

Fuente: Elaborado por el Instituto Nacional de Estadísticas

#### 4. Dominios de Estudio

Tabla 4.1. Comunas definidas como dominio de estudio con tasa de pobreza por ingresos (nueva metodología) mayor a 10% y error relativo menor o igual a 36%.

Región	Comuna		Estimador de la Razón $P_{i(2013)}$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Muestra Objetiva Propuesta $m_2$	Fracción de Muestreo $f$	Tasa de No Respuesta Promedio Ponderado 2011-2013	Muestra con Sobremuestreo
I de Tarapacá	Alto Hospicio		10,3%	3,4%	33,2%	700	3,4%	0,20	872
IV de Coquimbo	Coquimbo	Capital Provincial	13,2%	2,9%	21,8%	800	1,5%	0,19	990
	La Serena	Capital Regional y Provincial	13,7%	4,9%	35,7%	1.200	2,2%	0,25	1.603
	Ovalle	Capital Provincial	15,6%	5,5%	35,3%	500	1,8%	0,16	597
	Illapel	Capital Provincial	16,3%	4,8%	29,7%	200	2,2%	0,13	231
	Vicuña		26,2%	6,4%	24,4%	200	2,6%	0,10	223
V de Valparaíso	Viña del Mar		13,3%	3,7%	28,1%	1.100	1,0%	0,38	1.783
	Valparaíso	Capital Regional y Provincial	17,3%	4,8%	27,8%	1.000	1,2%	0,30	1.437
	Villa Alemana		12,8%	3,6%	28,5%	450	1,3%	0,21	571
	San Antonio	Capital Provincial	18,9%	6,5%	34,5%	500	1,8%	0,16	593
	Quillota	Capital Provincial	14,0%	4,2%	30,3%	270	1,1%	0,23	352
	San Felipe	Capital Provincial	11,0%	3,5%	31,8%	600	2,9%	0,33	894
	Los Andes	Capital Provincial	12,0%	4,0%	33,0%	350	1,8%	0,26	474
	Calera		25,3%	8,9%	35,1%	200	1,3%	0,14	233
	Limache		26,8%	9,4%	35,2%	750	5,8%	0,13	862
La Ligua	Capital Provincial	18,6%	6,3%	34,1%	450	5,8%	0,16	534	
VI de O'Higgins	Rancagua	Capital Regional y Provincial	14,2%	4,2%	29,7%	1.100	1,6%	0,17	1.323
	San Fernando	Capital Provincial	17,6%	6,2%	35,3%	450	2,1%	0,18	552
	Santa Cruz		15,2%	5,1%	33,4%	300	3,1%	0,15	354
	Chimbarongo		15,9%	3,7%	23,2%	200	2,3%	0,11	225
	Graneros		13,9%	4,7%	34,1%	200	2,6%	0,11	226
	San Vicente		17,0%	5,9%	34,7%	250	4,6%	0,08	272
	Pichidegua		26,7%	3,5%	13,1%	200	5,1%	0,05	211
	Rengo		21,7%	5,8%	26,5%	300	7,7%	0,13	346
	Coltauco		10,9%	3,8%	34,4%	600	17,5%	0,13	692
Codegua		20,9%	6,2%	29,8%	100	3,7%	0,15	117	

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Continúa ►

Región	Comuna	Estimador de la Razón $P_{i(2013)}$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Muestra Objetiva Propuesta $m_2$	Fracción de Muestreo $f$	Tasa de No Respuesta Promedio Ponderado 2011-2013	Muestra con Sobremuestreo	
VII del Maule	Talca	Capital Regional y Provincial	14,7%	3,0%	20,5%	1.000	1,5%	0,17	1.211
	Curicó	Capital Provincial	18,7%	5,5%	29,2%	600	1,6%	0,18	731
	Linares	Capital Provincial	20,3%	6,3%	31,1%	320	1,2%	0,10	356
	Cauquenes	Capital Provincial	33,0%	9,3%	28,0%	290	2,1%	0,10	322
	Constitución		19,9%	6,9%	34,5%	250	2,0%	0,12	285
	Parral		31,2%	9,4%	30,3%	380	3,1%	0,15	448
	San Clemente		23,5%	5,7%	24,2%	200	1,8%	0,06	213
	Teno		28,9%	2,8%	9,7%	200	2,9%	0,13	229
	Villa Alegre		30,6%	4,5%	14,8%	200	4,0%	0,08	216
	San Javier		34,4%	8,9%	25,8%	180	3,8%	0,03	186
	Colbún		35,9%	7,0%	19,6%	200	4,6%	0,03	206
	Yerbas Buenas		32,1%	8,8%	27,5%	200	4,8%	0,02	204
VIII del Biobío	Chillán	Capital Provincial	23,2%	5,8%	25,1%	500	1,0%	0,13	577
	Los Ángeles	Capital Provincial	22,3%	7,8%	35,0%	1.500	3,1%	0,15	1.759
	Talcahuano		15,3%	5,3%	34,9%	550	1,2%	0,20	688
	San Pedro de la Paz		25,5%	7,8%	30,6%	600	1,8%	0,20	749
	Coronel		20,7%	4,2%	20,3%	500	1,7%	0,16	594
	Hualpén		13,1%	3,4%	26,3%	200	0,8%	0,17	242
	Chiguayante		14,0%	3,7%	26,7%	500	2,1%	0,20	626
	Tomé		24,5%	8,8%	35,7%	200	1,4%	0,14	232
	Lota		19,4%	6,9%	35,5%	320	2,3%	0,12	364
	Penco		22,4%	6,5%	29,0%	200	1,6%	0,16	238
	Cañete		51,6%	9,4%	18,3%	650	7,3%	0,14	754
	Mulchén		25,4%	6,7%	26,4%	200	2,3%	0,07	214
	Curanilahue		16,4%	5,7%	34,8%	400	4,7%	0,19	493
	Nacimiento		32,2%	8,8%	27,3%	200	2,5%	0,13	231
	Chillán Viejo		22,1%	4,2%	19,0%	200	2,6%	0,10	223
	Laja		22,6%	4,2%	18,5%	200	2,7%	0,12	228
	Arauco		19,3%	4,3%	22,1%	200	2,8%	0,17	241
Lebu	Capital Provincial	31,2%	4,8%	15,3%	200	3,2%	0,20	250	

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Continúa ►

Región	Comuna	Estimador de la Razón $P_{1(2013)}$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Muestra Objetiva Propuesta $m_2$	Fración de Muestreo $f$	Tasa de No Respuesta Promedio Ponderado 2011-2013	Muestra con Sobremuestreo	
IX de La Araucanía	Temuco	Capital Regional y Provincial	18,3%	4,0%	21,7%	900	1,2%	0,23	1.170
	Padre Las Casas		29,3%	6,4%	21,7%	350	2,0%	0,11	393
	Angol	Capital Provincial	11,3%	3,9%	34,2%	400	2,7%	0,13	461
	Victoria		21,2%	7,5%	35,6%	550	5,2%	0,18	668
	Nueva Imperial		52,6%	7,9%	15,0%	200	2,1%	0,07	216
	Lautaro		34,7%	4,7%	13,4%	250	2,7%	0,11	282
	Pucón		22,4%	4,6%	20,6%	250	3,1%	0,18	306
	Pitrufquén		27,8%	9,4%	33,7%	410	5,9%	0,09	449
	Carahue		44,6%	8,8%	19,7%	200	2,9%	0,05	211
	Collipulli		35,1%	1,7%	4,7%	350	5,2%	0,11	394
	Freire		34,4%	4,5%	13,1%	200	3,0%	0,10	223
	Traiguén		29,0%	2,8%	9,6%	200	3,3%	0,15	235
	Curacautín		26,9%	8,9%	33,1%	250	4,3%	0,08	272
Curco		36,7%	7,2%	19,6%	230	4,5%	0,14	268	
Villarrica		14,2%	4,9%	34,8%	550	11,6%	0,11	616	
X de Los Lagos	Puerto Montt	Capital Regional y Provincial	14,6%	3,1%	21,3%	900	1,7%	0,19	1.105
	Osorno	Capital Provincial	16,6%	3,6%	21,6%	600	1,3%	0,22	774
	Ancud		15,1%	5,4%	35,3%	420	3,4%	0,13	482
	Castro	Capital Provincial	11,7%	4,1%	35,4%	750	6,5%	0,16	890
	Calbuco		30,2%	6,6%	21,9%	200	2,7%	0,03	206
	Frutillar		25,9%	9,1%	34,9%	270	5,3%	0,16	321
	Río Negro		22,3%	7,9%	35,3%	300	6,7%	0,15	353
	Llanquihue		13,6%	3,8%	28,1%	300	6,9%	0,07	321
	Los Muermos		20,6%	7,4%	35,8%	450	10,9%	0,04	470
XIII Metropolitana	Puente Alto	Capital Provincial	13,6%	4,8%	35,4%	1.500	1,0%	0,18	1.836
	San Bernardo	Capital Provincial	19,0%	5,5%	29,2%	400	0,6%	0,26	540
	La Pintana		17,4%	3,4%	19,6%	300	0,7%	0,20	374
	El Bosque		13,7%	4,8%	35,2%	1.000	2,3%	0,22	1.287
	Recoleta		11,0%	3,9%	35,2%	380	1,0%	0,24	497
	Cerro Navia		15,8%	5,6%	35,3%	320	0,9%	0,12	364
	Conchalí		17,9%	6,3%	35,0%	1.500	4,6%	0,17	1.799
	La Granja		21,8%	6,3%	28,9%	250	0,8%	0,14	290
	Melipilla	Capital Provincial	21,4%	3,7%	17,3%	220	0,8%	0,21	277

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Continúa ►

Región	Comuna		Estimador de la Razón $P_{i(2013)}$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Muestra Objetiva Propuesta $m_2$	Fracción de Muestreo $f$	Tasa de No Respuesta Promedio Ponderado 2011-2013	Muestra con Sobremuestreo
XIV de Los Ríos	Valdivia	Capital Regional y Provincial	14,2%	4,4%	31,1%	700	1,7%	0,21	887
	La Unión	Capital Provincial	23,0%	4,8%	21,1%	370	3,3%	0,10	411
	Río Bueno		24,1%	6,2%	25,6%	370	3,7%	0,04	387
	Panguipulli		34,5%	9,4%	27,2%	370	6,4%	0,09	405
	Los Lagos		27,6%	5,8%	21,0%	450	7,9%	0,10	499
	Paillaco		31,9%	9,4%	29,4%	280	5,0%	0,10	310
XV de Arica y Parinacota	Arica	Capital Regional y Provincial	14,1%	3,7%	26,3%	700	1,3%	0,19	860
	Putre	Capital Provincial	29,2%	8,7%	29,9%	100	17,7%	0,13	115

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

**Tabla 4.2. Comunas definidas como dominio de estudio con tasa de pobreza por ingresos (nueva metodología) mayor o igual a 5% y menor a 10% y error relativo menor o igual a 50%**

Región	Comuna		Estimador de la Razón $P_{i(2013)}$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Muestra Objetivo Propuesta $m_2$	Fracción de Muestreo	Tasa de No Respuesta Promedio Ponderado 2011-2013	Muestra con Sobremuestreo
I de Tarapacá	Iquique	Capital Regional y Provincial	6,2%	2,5%	39,8%	900	1,9%	0,32	1.319
III de Atacama	Copiapó	Capital Regional y Provincial	6,3%	3,1%	49,6%	2.350	5,8%	0,36	3.663
	Vallenar	Capital Provincial	7,8%	3,3%	41,9%	300	2,1%	0,19	371
V de Valparaíso	Quilpué	Capital Provincial	8,3%	4,1%	49,5%	820	1,8%	0,23	1.063
	El Quisco		8,7%	4,3%	50,0%	300	2,4%	0,22	387
	Concón		7,4%	2,6%	34,6%	350	2,8%	0,38	565
VI de O'Higgins	Machalí		5,8%	2,9%	49,0%	700	6,8%	0,28	977
	Peumo		8,4%	4,0%	47,3%	250	7,0%	0,12	285
	Requínoa		6,4%	3,2%	50,0%	650	19,0%	0,12	735
VIII del Biobío	Concepción	Capital Regional y Provincial	9,1%	3,8%	41,9%	1.000	1,5%	0,33	1.495
XI de Aysén	Coyhaique		6,9%	3,4%	49,5%	480	2,9%	0,23	621
	Aysén		7,9%	3,9%	49,4%	350	6,5%	0,20	439
	Cochrane	Capital Provincial	8,4%	2,4%	28,2%	100	10,1%	0,26	136
XIII Metropolitana	Santiago	Capital Regional y Provincial	6,5%	3,2%	48,5%	650	0,6%	0,54	1.413
	La Florida		7,8%	2,5%	32,5%	500	0,5%	0,36	782
	Peñalolén		9,3%	2,6%	27,5%	380	0,7%	0,25	509
	Pudahuel		8,1%	4,0%	48,9%	500	0,9%	0,15	588
	Quilicura		9,9%	4,9%	49,8%	500	1,0%	0,13	575
	Renca		8,6%	4,3%	49,9%	300	0,8%	0,12	342
	Estación Central		9,9%	4,5%	45,1%	200	0,6%	0,22	257
	Macul		6,0%	2,3%	38,5%	200	0,6%	0,27	274
	Quinta Normal		9,9%	4,7%	48,0%	350	1,2%	0,17	419
	Pedro Aguirre Cerda		7,3%	3,5%	47,7%	350	1,3%	0,14	408
	Talagante	Capital Provincial	9,8%	2,8%	28,7%	200	1,3%	0,16	237

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

**Tabla 4.3. Comunas definidas como dominio de estudio con tasa de pobreza por ingresos (nueva metodología) mayor o igual a 2% y menor a 5% y error relativo menor o igual a 65%**

Región	Comuna		Estimador de la Razón $P_i (2013)$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Muestra Objetivo Propuesta $m_2$	Fracción de Muestreo	Tasa de No Respuesta Promedio Ponderado 2011-2013	Muestra con Sobremuestreo
II de Antofagasta	Antofagasta	Capital Regional y Provincial	3,7%	2,3%	62,3%	800	1,0%	0,33	1.192
	Calama	Capital Provincial	3,4%	2,2%	63,4%	400	1,1%	0,31	581
	Tocopilla	Capital Provincial	2,2%	1,3%	60,2%	300	4,2%	0,24	394
XI de Aysén	Chile Chico	Capital Provincial	4,6%	1,7%	37,8%	100	6,8%	0,20	126
XII de Magallanes	Punta Arenas	Capital Regional y Provincial	3,9%	1,8%	47,7%	800	2,2%	0,33	1.187
XIII Metropolitana	Maipú		3,6%	2,3%	62,7%	600	0,4%	0,28	839
	Providencia		3,2%	2,1%	64,9%	1.100	1,7%	0,70	3.703
	San Miguel		4,7%	3,1%	65,0%	500	1,7%	0,40	838
	Colina	Capital Provincial	4,8%	3,0%	63,0%	250	1,1%	0,25	332

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

**Tabla 4.4. Comunas definidas como dominio de estudio con tasa de pobreza por ingresos (nueva metodología) menor a 2% y error relativo menor o igual a 80%**

Región	Comuna		Estimador de la Razón $P_i (2013)$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Muestra Objetivo Propuesta $m_2$	Fracción de Muestreo	Tasa de No Respuesta Promedio Ponderado 2011-2013	Muestra con Sobremuestreo
XIII Metropolitana	Las Condes		1,5%	1,2%	80,2%	1.100	1,1%	0,59	2.655
	Ñuñoa		0,4%	0,3%	79,8%	1.150	1,7%	0,59	2.825

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

**Tabla 4.5. Comunas definidas como dominio de estudio que no cumplen con los criterios de error pre-establecidos, pero son consideradas por ser capitales provinciales y/o acumulan aproximadamente el 80% de las viviendas de la región\*.**

Región	Comuna		Estimador de la Razón $P_{i(2013)}$	Error Absoluto Propuesto $e_i$	Error Relativo Propuesto $e_i$	Muestra Objetiva Propuesta (m 2)	Fracción de Muestreo	Tasa de No Respuesta Promedio Ponderado 2011-2013	Muestra con Sobremuestreo
I de Tarapacá	Pozo Almonte	Capital Provincial	14,0%	5,2%	37,1%	600	20,6%	0,12	685
III de Atacama	Chañaral	Capital Provincial	10,0%	5,0%	49,7%	550	13,2%	0,15	648
VI de O'Higgins	Olívar		16,2%	7,1%	44,1%	100	3,2%	0,07	108
	Pichilemu	Capital Provincial	13,7%	13,1%	95,5%	200	22,0%	0,20	251
X de Los Lagos	Quellón		13,2%	6,6%	49,9%	580	10,0%	0,17	699
XII de Magallanes	Natales	Capital Provincial	17,2%	8,6%	49,8%	670	10,2%	0,13	766
	Porvenir	Capital Provincial	0,8%	1,0%	115,6%	300	20,6%	0,10	334

\*Datos a ser evaluados considerando Casen 2015, en base los errores efectivos.

## 5. Comandos y Sintaxis en SPSS y STATA

### 5.1. Comandos utilizados en SPSS para estimar la tasa de pobreza por ingresos y errores muestrales

#### Plan de Muestra Estimación Tasa de Pobreza y Tasa de Pobreza Extrema por Ingresos, CASEN 2015

```
CSPLAN ANALYSIS  
/PLAN FILE='C:\BASES CASEN 2015\plan_casen2015.csaplan'  
/PLANVARS ANALYSISWEIGHT=expr  
/SRSESTIMATOR TYPE=WOR  
/PRINT PLAN  
/DESIGN STRATA=varstrat CLUSTER=varunit  
/ESTIMATOR TYPE=WR.
```

#### Transformación de Variables para Identificar Personas y Hogares en situación de pobreza por ingresos.

```
DATASET ACTIVATE Conjunto_de_datos1.  
RECODE pobreza (1=1) (2=1) (3=0) INTO pobre.  
VARIABLE LABELS pobre 'Pobreza Extrema y No Extrema (Pobre)'.  
VALUE LABELS pobre 1 "pobres" 0 "no pobres".  
EXECUTE.
```

#### **Cálculo para personas**

##### 1. Porcentaje de personas en situación de pobreza extrema por ingresos, por región, año 2015

```
CSTABULATE  
/PLAN FILE='C:\Base de datos Casen2015\Plan_Casen2015.csaplan'  
/TABLES VARIABLES=region BY pobreza  
/CELLS ROWPCT  
/STATISTICS SE CV CIN(95) COUNT  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

##### 2. Porcentaje de personas en situación de pobreza por ingresos, por región, año 2015

```
CSTABULATE  
/PLAN FILE='C:\Base de datos Casen2015\Plan_Region_Casen2015.csaplan'  
/TABLES VARIABLES=region BY pobre  
/CELLS ROWPCT  
/STATISTICS SE CV CIN(95) COUNT  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

##### 3. Porcentaje de personas en situación de pobreza extrema por ingresos, por zona, año 2015

```
CSTABULATE  
/PLAN FILE='C:\Base de datos Casen2015\Plan_Region_Casen2015.csaplan'  
/TABLES VARIABLES=zona BY pobreza  
/CELLS ROWPCT  
/STATISTICS SE CV CIN(95)  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

##### 4. Porcentaje de personas en situación de pobreza por ingresos, por zona, año 2015

```
CSTABULATE  
/PLAN FILE='C:\Base de datos Casen2015\Plan_Region_Casen2015.csaplan'  
/TABLES VARIABLES=zona BY pobre  
/CELLS ROWPCT
```

```
/STATISTICS SE CV CIN(95)
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

### **Cálculo para hogares**

\* Se filtra la base de datos a nivel de jefe de hogar

```
USE ALL.
COMPUTE filter_$=(pco1 = 1).
FILTER BY filter_$.
EXECUTE.
```

#### **1. Porcentaje de hogares en situación de pobreza extrema por ingresos, por región, año 2015**

```
CSTABULATE
/PLAN FILE='C:\Base de datos Casen2015\Plan_Casen2015.csaplan'
/TABLES VARIABLES=region BY pobreza
/CELLS ROWPCT
/STATISTICS SE CV CIN(95) COUNT
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

#### **2. Porcentaje de hogares en situación de pobreza por ingresos, por región, año 2015**

```
CSTABULATE
/PLAN FILE='C:\Base de datos Casen2015\Plan_Region_Casen2015.csaplan'
/TABLES VARIABLES=region BY pobre
/CELLS ROWPCT
/STATISTICS SE CV CIN(95) COUNT
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

#### **3. Porcentaje de hogares en situación de pobreza extrema por ingresos, por zona, año 2015**

```
CSTABULATE
/PLAN FILE='C:\Base de datos Casen2015\Plan_Region_Casen2015.csaplan'
/TABLES VARIABLES=zona BY pobreza
/CELLS ROWPCT
/STATISTICS SE CV CIN(95)
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

#### **4. Porcentaje de hogares en situación de pobreza por ingresos, por zona, año 2015**

```
CSTABULATE
/PLAN FILE='C:\Base de datos Casen2015\Plan_Region_Casen2015.csaplan'
/TABLES VARIABLES=zona BY pobre
/CELLS ROWPCT
/STATISTICS SE CV CIN(95)
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

## 5.2. Comandos utilizados en STATA para estimación de tasa de pobreza por ingresos y errores muestrales

### Plan de Muestro Cálculo Tasa de Pobreza y Tasa de Pobreza Extrema por ingresos CASEN 2015

```
svyset varunit [w=expr], strata(varstrat) vce(linearized) singleunit(certainty)
```

### Transformación de Variables para Identificar Personas y Hogares en situación de pobreza por ingresos.

```
gen pobre =1 if (pobreza ==1 | pobreza ==2)
```

```
replace pobre=0 if (pobreza ==3)
```

```
label var pobre "pobreza_ingreso"
```

```
label define pobre 1 "pobres" o "no pobres"
```

```
label values pobre pobre
```

### Cálculo para personas

#### 1. Porcentaje de personas en situación de pobreza extrema por ingresos, por región, año 2015

```
svy: prop pobreza, level(95)
```

```
svy: prop pobreza, level(95) over(region)
```

#### 2. Porcentaje de personas en situación de pobreza por ingresos, por región, año 2015

```
svy: prop pobre, level(95)
```

```
svy: prop pobre, level(95) over(region)
```

#### 3. Porcentaje de personas en situación de pobreza extrema por ingresos, por zona, año 2015

```
svy: prop pobreza, level(95)
```

```
svy: prop pobreza, level(95) over(zona)
```

#### 4. Porcentaje de personas en situación de pobreza por ingresos, por zona, año 2015

```
svy: prop pobre, level(95)
```

```
svy: prop pobre, level(95) over(zona)
```

### Cálculo para hogares

#### 1. Porcentaje de hogares en situación de pobreza extrema por ingresos, por región, año 2015

```
svy: prop pobreza if pco1==1, level(95)
```

```
svy: prop pobreza if pco1==1, level(95) over(region)
```

#### 2. Porcentaje de hogares en situación de pobreza por ingresos, por región, año 2015

```
svy: prop pobre if pco1==1, level(95)
```

```
svy: prop pobre if pco1==1, level(95) over(region)
```

#### 3. Porcentaje de hogares en situación de pobreza extrema por ingresos, por zona, año 2015

```
svy: prop pobreza if pco1==1, level(95)
```

```
svy: prop pobreza if pco1==1, level(95) over(zona)
```

#### 4. Porcentaje de hogares en situación de pobreza por ingresos, por zona, año 2015

```
svy: prop pobre if pco1==1, level(95)
```

```
svy: prop pobre if pco1==1, level(95) over(zona)
```

## BIBLIOGRAFÍA

- INE Chile. (2006). Encuesta Nacional de Empleo. Diseño muestral.
- Hansen, M; Hurwitz, W; Madow, W.(1953). Sample Survey Method and Theory. Volume II-Theory. John Wiley & Sons, Inc. New York.
- Cochran, W. (1977). Sampling Techniques. Third Edition. John Wiley & Sons, Inc. New York.
- Kish, L. (1965). Survey Sampling. John Wiley & Sons, Inc. New York.
- Groves, R. (1989, 2004). Survey Errors and Survey Cost. John Wiley & Sons, Inc. Hoboken. New Jersey.
- Levy, P; Lemeshow, S. (1999). Sampling of Populations: Methods and Applications. Third Edition. Wiley & Sons, Inc. Canada.
- Lehtonen, R; Pahkinen, E. (2004). Practical Methods for Design and Analysis of Complex Surveys. Second Edition. John Wiley & Sons, Ltd. England.
- Rao, P. (2000). Sampling Methodologies with Applications. Chapman & Hall/CRC.US.
- Naciones Unidas. (2009). Encuesta de hogares en los países de desarrollo y transición. Departamento de Asuntos Económicos. División Estadística. New York.
- Czajka, J.L., Hirabayashi, S.M., Little, R.J.A. and Rubin, D.B. (1992). "Projecting from Advance Data Using Propensity Modeling: An Application to Income and Tax Statistics." *Journal of Business and Economic Statistics*, 10(2), 117-132.
- Ehling, M., and T. Körner (2007), *Handbook on Data Quality Assessment Methods and Tools*, Wiesbaden: European Commission, Eurostat.
- Sitter RR (1992) Comparing Three Bootstrap Methods for Survey Data. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 20:135-154
- Valliant R, Brick JM, Dever JA (2008) Weight Adjustments for the Grouped Jackknife Variance Estimator. *Journal of Social Statistics* 24(3):469-488.
- Valliant, R. Dever, J.A., Kreuter,F. (2013, chapter 14).
- Valliant, R. Dever, J.A., Kreuter,F. (2013, section 6.5.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- Valliant R, Rust KF (2010) Degrees of Freedom Approximations and Rules-of-Thumb. *Journal of Social Statistics* 26:585{602.
- Wolter KM (2007) *Introduction to Variance Estimation*, 2nd edn. Springer, New York Valliant, R. Dever, J.A., Kreuter,F. (2013, chapter 15) Westat (2007) *WesVar 4.3 Users Guide*. Westat, Rockville MD, URL [www.westat.com](http://www.westat.com)

