

Metodología de Diseño Muestral Encuesta de Caracterización Socioeconómica Nacional 2013

Serie Documentos Metodológicos N°30
24 de Enero de 2015

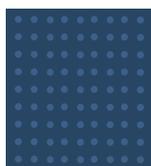
www.desarrollosocial.cl

Observatorio
Social



Ministerio de
Desarrollo
Social

Gobierno de Chile

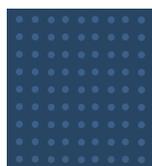


Observatorio
Social

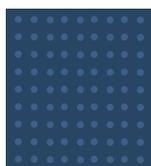


Casen
2013

I.	Objetivo y métodos.....	4
1.	<i>Objetivo</i>	4
2.	<i>Métodos.....</i>	5
II.	Diseño muestral de la encuesta casen 2013	8
1.	<i>Características del marco muestral.....</i>	8
1.1.	Dos Marcos muestrales mutuamente excluyentes.....	8
1.2.	Agrupación de Unidades Primarias de Muestreo en el MM2008	10
1.3.	Cobertura del Marco Muestral	11
1.5.	Estratificación del Marco Muestral por comunas y división censal urbano-rural	17
2.	<i>Estimación del tamaño muestral.....</i>	18
2.1.	Proceso de cálculo	18
2.2.	Distribución inicial de la muestra Casen 2013 en dos submuestras independientes y de mismo diseño.....	25
2.3.	Reseña de los ajustes locales en el tamaño de la muestra de Casen 2013.....	26
2.4.	Ajuste de los tamaños de la muestra objetivo por tasa de no respuesta regional en Casen 2013.....	27
2.5.	Distribución de las viviendas por comunas.....	28
3.	<i>Selección de la muestra</i>	31
3.1.	Selección de conglomerados	31
3.1.1.	Selección de Conglomerados desde el MM2008	31
3.1.2.	Selección de Conglomerados desde el MS2002	32
3.2.	Selección de viviendas	34
3.2.1.	Proceso de enumeración de los Conglomerados en 2012 y reglas de reemplazo en el proceso de enumeración.....	34
3.2.2.	Material de trabajo del enumerador	34
3.2.2.1.	Recorrido y Enumeración Urbana	35
3.2.2.2.	Recorrido y Enumeración Rural.....	35
3.2.2.3.	Razones de reemplazos de conglomerados durante la enumeración.....	35
3.2.2.	Reenumeración del 30% unidades primarias de muestreo ya enumeradas en 2012	36
3.2.2.1	Variables utilizadas.....	37
3.2.2.2	Selección de comunas a reenumerar	38
3.2.2.3	Selección de las Viviendas.....	39
III.	DESARROLLO DE FACTORES DE EXPANSION.....	41
1.	<i>Introducción.....</i>	41
2.	<i>Ponderador de selección de conglomerados</i>	43
2.1.	Probabilidad de selección en el Marco de Secciones (MS2002).....	43
2.2.	Probabilidad de selección en el Marco de Manzanas (MM2008)	46
2.4.	Determinación del Ponderador de Selección de Conglomerados como el inverso de las probabilidades de selección	52
3.	<i>Ajuste por omisión de conglomerados.....</i>	56
4.	<i>Ponderación de selección de viviendas.....</i>	58
4.1.	Probabilidad Condicional de selección de viviendas	58
4.2.	Corrección de probabilidades condicionales por reducción de muestra.....	58
4.3.	Ponderador de selección de viviendas	59



5.	<i>Ponderación por elegibilidad</i>	60
5.1.	<i>Ajuste por elegibilidad desconocida</i>	61
5.2.	<i>Ajuste por no elegibilidad</i>	62
6.	<i>Ponderación por no respuesta</i>	63
7.	<i>Ponderación de calibración</i>	65
7.1.	<i>Ponderador de Calibración Regional</i>	68
7.2.	<i>Ponderador de Calibración Comunal</i>	70
IV.	Estimación de varianzas	74
1.	<i>El estimador</i>	74
2.	<i>Varianza del estimador</i>	76
3.	<i>Variables que identifican el diseño muestral complejo</i>	78
3.1.	<i>Creación de pseudo-estratos (varstrat)</i>	79
3.2.	<i>Creación de pseudo-conglomerados (varunit)</i>	80
4.	<i>Resultados a nivel regional, por áreas geográficas urbana y rural, a nivel nacional</i>	82
5.	<i>Márgenes de error, proyectados y efectivos, a nivel nacional y regional</i>	85
6.	<i>Programas computacionales</i>	87
V.	Anexos	93
	<i>Anexo 1: Referencias</i>	93
	<i>Anexo 2: Constitución de dos marcos muestrales independientes para la selección de muestras para encuesta sociales, INE</i>	94
	<i>Anexo 3: Resto de Áreas Urbanas (RAU) seleccionadas desde el MS2002</i>	95
	<i>Anexo 4: Áreas RAU seleccionadas desde el MM2008</i>	96
	<i>Anexo 5: grupos de tamaño</i>	97
	<i>Anexo 6: total de viviendas según subestrato y área, para cada estrategia</i>	98
	<i>Anexo 8: comunas reenumeradas para casen 2013</i>	100



I. Objetivo y métodos

La metodología de diseño muestral de la Encuesta de Caracterización Socioeconómica Nacional (Casen) 2013 describe el conjunto de opciones metodológicas y de procedimientos estadísticos adoptados a lo largo del proceso de planificación, que fueron considerados los más apropiados para seleccionar una muestra probabilística, estadísticamente representativa de las viviendas particulares de Chile y que permitiera estimar con el nivel de precisión deseado, la tasa de pobreza en cada región del territorio nacional, para áreas geográficas urbana y rural y para el país en su conjunto.

1. Objetivo

La encuesta Casen 2013, al igual que sus predecesoras ha sido pensada como un instrumento de diagnóstico, evaluación y de focalización que permita:

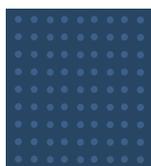
- Conocer periódicamente la situación de los hogares y de la población, especialmente de aquella en situación de pobreza y de aquellos grupos definidos como prioritarios por la política social, con relación a aspectos demográficos, de educación, salud, vivienda, trabajo e ingresos.
- En particular, estimar la magnitud de la pobreza y la distribución del ingreso; identificar carencias y demandas de la población en las áreas señaladas; y evaluar las distintas brechas que separan a los diferentes segmentos sociales y ámbitos territoriales.
- Evaluar el impacto de la política social: estimar la cobertura, la focalización y la distribución del gasto fiscal de los principales programas sociales de alcance nacional entre los hogares, según su nivel de ingreso, para evaluar el impacto de este gasto en el ingreso de los hogares y en la distribución del mismo.

Este documento, de carácter técnico, está orientado a un público de usuarios e investigadores de la encuesta Casen 2013 especializados en temas de encuestas estadísticas y, en general, a todas aquellas personas que deseen entender más específicamente y en profundidad las opciones metodológicas que se tomaron en el diseño de esta encuesta.

En este documento se encuentra la descripción detallada de los aspectos técnicos estadísticos relacionados con el marco muestral del cual se extrajo la muestra, el diseño muestral propiamente tal, con la descripción exhaustiva de sus unidades muestrales, sus etapas y los detalles de su estratificación, los procedimientos empleados en el proceso de selección de las unidades muestrales, las determinación y afijación de la muestra objetivo (o esperada) y de la muestra con sobremuestreo para el programa general de trabajo de campo. Este sobremuestreo tiene como objetivo, entre otros contrarrestar la no respuesta a la unidad.

De manera complementaria, y en espejo con el diseño, en este documento se exponen los distintos ajustes aplicados secuencialmente al inverso de las probabilidades de selección de las unidades muestrales, para generar el factor de expansión regional. Es el uso de este factor el que permite generar las estimaciones a nivel nacional, de áreas geográficas urbana y rural y de regiones, y realizar inferencias a partir de los datos muestrales para la población.

Al igual que el factor de expansión regional, el documento pone a libre disposición de los usuarios de Casen 2013 un "factor comunal". Es importante destacar que el Ministerio no recomienda emplear este factor para la obtención directa de estimaciones



comunales, ya que los datos de la encuesta Casen 2013 no han sido diseñados para ser representativos a ese nivel.

En efecto, el factor comunal es proporcionado en calidad de insumo, entre otros varios, para la generación de estimaciones estadísticas comunales mediante metodologías de estimación para áreas pequeñas (en inglés metodología SAE, cuyas siglas significan "Small Area Estimation"), introducidas por el Ministerio para estimar la tasa de pobreza comunal a partir de Casen 2009.

Otro aspecto importante expuesto en el documento son los procedimientos desarrollados para estimar las varianzas de los estimadores y así caracterizar la precisión de la encuesta Casen 2013. Tratándose de una muestra compleja, para el cálculo de varianzas se requiere de la especificación de las variables que identifican este tipo de diseño: pseudo-estratos (varstrat) y pseudo-conglomerados (varunit) y sus correspondientes rutinas computacionales en SPSS y en Stata.

El conocimiento de estos aspectos del diseño de la encuesta es fundamental para los investigadores que planean analizar los datos de la encuesta Casen 2013.

2. Métodos

La Encuesta de Caracterización Socioeconómica Nacional 2013, diseñada e implementada por el Ministerio de Desarrollo Social, es una encuesta multipropósito que recolecta datos acerca de características sociales y económicas de una muestra representativa de la población residente en hogares particulares en Chile.

Las unidades muestrales de la encuesta Casen 2013 son las viviendas, seleccionadas de manera probabilística, estratificada y multietápica. La muestra es representativa a nivel de país, por áreas geográficas (urbana y rural) y por regiones.

Su cobertura es el territorio nacional y abarca las 15 regiones del país y las 324 comunas incluidas en el marco muestral del Instituto Nacional de Estadísticas (INE), excluyendo aquellos territorios del espacio nacional identificados por tal institución como Áreas de Difícil Acceso (ADAs).

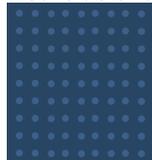
La variable de interés en la determinación del tamaño muestral es la tasa de pobreza, para la cual se fijaron errores absolutos regionales de 1 a 4 puntos porcentuales lo que resulta en un error absoluto de 0.69 puntos porcentuales a nivel nacional¹. Adicionalmente en Casen 2013, el error relativo se acotó a 30%.

La selección de la muestra se realizó en el marco muestral que define el INE para todas las encuestas sociales a hogares, el marco muestral de viviendas, construido con información del Censo de población y viviendas del 2002.

Desde el 2002 a la fecha, el marco se ha desarrollado y ha ido incorporando los cambios territoriales propios de la última década. El marco muestral de viviendas del INE se estructura en la actualidad en dos marcos muestrales independientes: el Marco Muestral de Secciones 2002 (MMS2002 que cubre tanto áreas geográficas rurales como Resto de Áreas Urbanas) y en el Marco Maestro de Manzanas 2008 (MMM2008, que cubre tanto áreas geográficas urbanas como Resto de Áreas Urbanas).

Esquemáticamente, el procedimiento de selección de la muestra se realiza en dos etapas (muestreo bietápico) en el caso del MS2002 y en tres etapas (muestreo trietápico) en el MM2008.

¹ Considerando el tamaño muestral de Casen 2013, n=70.080 viviendas.



Los estratos de muestreo son las áreas geográficas urbanas o rurales de las comunas, no presentándose ambas áreas geográficas sistemáticamente en todas las comunas.

En el MM2008 se seleccionan, en una primera etapa al interior de cada estrato las unidades primarias de muestreo (manzanas) según grupos de manzanas definidos previamente en función del número de viviendas en las manzanas (grupos de tamaño). En una segunda etapa, se seleccionan las manzanas en los grupos de tamaño. En una tercera etapa, habiendo previamente establecido en cada grupo un número fijo de viviendas a seleccionar, se seleccionan al interior de cada manzana las unidades finales de muestreo (las viviendas) con igual probabilidad.

En el MS2002, en una primera etapa, al interior de cada estrato se seleccionan las secciones con probabilidad proporcional al tamaño (definido como el número total de viviendas). En una segunda etapa, al interior de cada sección, se seleccionan las unidades finales de muestreo (las viviendas) con igual probabilidad.

Al interior de la vivienda se identifican todos hogares que allí habitan y las personas miembros de cada hogar.

La entrevista se realiza con un informante por hogar (informante idóneo Casen) que corresponde al jefe de hogar, o a su conyugue o, en su defecto, a un mayor de 18 años, miembro del hogar. A través del informante se recolectan datos de todos los miembros del hogar que son residentes habituales de la vivienda.

Por definición, se consideran miembros de un hogar a todas aquellas personas que, siendo residentes de una misma vivienda, pueden tener (o no) vínculos de parentesco y habitualmente hacen vida en común, es decir, se alojan y se alimentan juntas. Dicho de otra forma: habitan en la misma vivienda y tienen presupuesto de alimentación común.

No se consideran miembros del hogar las personas que pagan pensión (por lo tanto conforman otro hogar), en caso de cumplir con los requisitos de permanencia.

En la encuesta Casen 2013, el programa general de trabajo de campo se diseñó con una muestra que contemplaba un sobremuestreo (87.402 viviendas), para lograr una muestra objetivo o esperada (70.080 viviendas). De la muestra esperada, en la práctica se logró encuestar a 64.842 viviendas (92,5 % de viviendas logradas) lo que se traduce en 66.725 encuestas a hogares logradas.

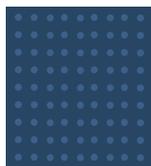
El diseño, la selección y el cálculo de factores de expansión de la muestra, al igual que el desarrollo de estudios complementarios fueron encargados por el Ministerio Desarrollo Social al INE, mediante un convenio de transferencia suscrito entre ambas instituciones.

El levantamiento y procesamiento de los datos de la encuesta, fueron contratados al Centro de Microdatos de la Facultad de Economía y Negocios de la Universidad de Chile. El trabajo de campo se realizó mediante la aplicación de entrevistas cara-a-cara, por encuestadores capacitados para dicha labor, utilizando un cuestionario en papel. La captura de los datos se realizó mediante lectura óptica de los cuestionarios.

Este documento se divide en seis capítulos.

El primer capítulo de introducción presenta sintéticamente las principales características de la muestra, su objetivo y métodos.

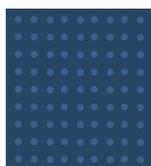
El segundo capítulo describe en detalle el desarrollo, los antecedentes y las principales características del diseño de la muestra de Casen 2013.



El tercer capítulo se refiere al desarrollo de los factores de expansión de la muestra de Casen 2013.

El cuarto capítulo aborda la metodología del cálculo de varianzas para la determinación de la precisión de los estimadores en el contexto de una muestra compleja como la de Casen 2013.

Finalmente el sexto capítulo de anexos reporta documentación relativa al marco muestral al igual que un breve conjunto de referencias metodológicas empleadas en el desarrollo del documento.



II. Diseño muestral de la encuesta casen 2013

La población objetivo de la encuesta Casen 2013, la constituyen todas las personas y hogares que residen en viviendas particulares de las 15 regiones del país, en las 324 comunas incluidas en el marco muestral del INE, excluyendo áreas de difícil acceso identificadas por el INE.

1. Características del marco muestral

“Los marcos muestrales son listas o procedimientos cuyo fin es identificar todos los elementos de una población objetivo. Los marcos pueden ser mapas, de áreas en las cuales se pueden encontrar los elementos, períodos de tiempo durante los cuales los sucesos objetivos pueden ocurrir, o registros de archivos administrativos, entre otros.” (Groves et al. 2009, pág. 70).

El INE mantiene un marco muestral de áreas geográficas que sirven de base para la selección de viviendas requeridas para las muestras de las encuestas de hogares más importantes del país. El marco contiene las unidades geográficas del país organizadas de acuerdo a la división político-administrativa: regiones, provincias y comunas.

Al interior de cada comuna se conforma la división censal que da origen a las áreas geográficas denominadas urbana y rural de acuerdo al Censo 2002², cuya definición se basa en la actividad económica preponderante y da origen a las siguientes entidades:

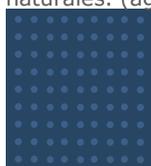
- Ciudad (CD): Es un gran centro urbano conformado por uno o un conjunto de centros urbanos adyacentes con 40.000 o más habitantes.
- Resto de Área Urbana (RAU): Conformado por un conjunto de centros urbanos que totalizan menos de 40.000 y más de 2.000 habitantes. Esta clasificación se da cuando en una 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 denomina simplemente como urbano (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 primarias³.

1.1. Dos Marcos muestrales mutuamente excluyentes

En base a los datos del Censo de Población y Vivienda 2002, el INE realizó una subdivisión de las comunas en **manzanas** censales (en las áreas urbanas) y **secciones** de empadronamiento censal (en las áreas rurales). Estas unidades geográficas son **las unidades primarias de muestreo** (UPM) más comúnmente utilizadas en las muestras de encuestas de hogares diseñadas por el INE.

² La muestra Casen 2013 expande a proyecciones de población residente en viviendas particulares, basadas en el Censo de Población y Vivienda 2002 del INE.

³ Se entiende por Actividad Primaria a toda aquella actividad relacionada con la extracción de recursos naturales. (agricultura, caza, pesca, minería, etc.).



La selección de la muestra de Casen 2013 se realizó a partir del marco muestral que el INE mantenía vigente al año 20120 y 2013:

- En el área rural, el marco de muestreo es aquel generado a partir del Censo de Población y Viviendas del año 2002. Las Unidades Primarias de Muestreo se denominan secciones y corresponden al área que puede recorrer un censista el día del censo. En adelante, denominaremos este marco muestral como **Marco de Secciones (MS2002)**.
- En el área Ciudad y Urbano (U), el Marco de Muestreo es aquel generado a partir de cartografía digital de alta validez métrica, actualizado al segundo semestre del año 2008. La información en el marco se actualiza con información anexa de Mapcity, Dmapas y los registros administrativos asociados a nuevas construcciones, otorgados por los municipios. Las Unidades Primarias de Muestreo se denominan manzanas y corresponden a delimitaciones geográficas fijas. En adelante, denominaremos a este marco como **Marco de Manzanas (MM2008)**.
- En las áreas denominadas Resto de Área Urbana (RAU), sus unidades pueden pertenecer tanto al MS2002 como al MM2008. De un total de 219 áreas de este tipo, 89 son extraídas desde el MS2002⁴ mientras que las restantes 130 se extraen desde el MM2008⁵. Cabe señalar que las 89 áreas urbanas seleccionadas a partir del MS2002 corresponden a áreas de pequeña densidad poblacional, con no más de 40.000 habitantes, y las unidades muestrales, son secciones al igual que en las áreas rurales.

4 Ver Capítulo VI. Anexo 3, para un listado completo de las comunas cuyas áreas RAU fueron seleccionadas desde el MS2002.

5 Ver Capítulo VI. Anexo 4 para un listado completo de las comunas cuyas áreas RAU fueron seleccionadas desde el MM2008.

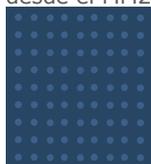
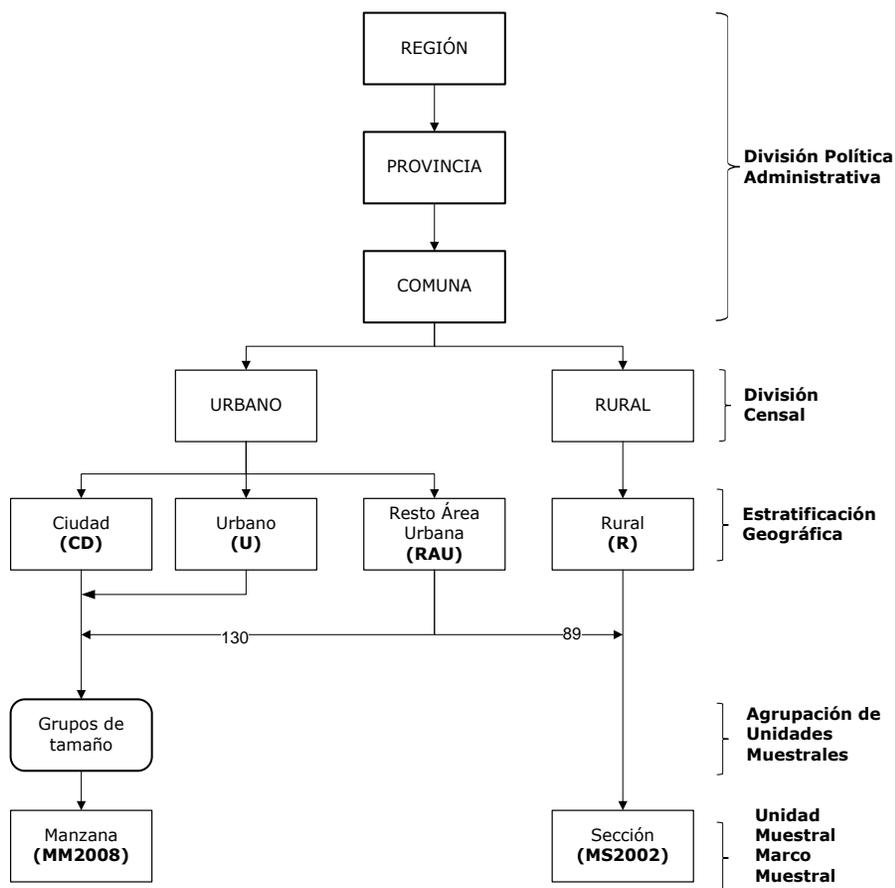


Figura II.1: Estratificación e identificación de Unidades Primarias de Muestreo.



1.2. Agrupación de Unidades Primarias de Muestreo en el MM2008

Debido a la gran variabilidad de tamaño que presentan las manzanas (número de viviendas particulares), el INE las clasifica en 30 grupos homogéneos en función de esta variable.

Esta agrupación por tamaño opera en el MM2008 para la selección de las unidades primarias (las manzanas), y permite reducir el sesgo de selección por tamaño, estableciendo en cada grupo un número fijo de Unidades Secundarias de Muestreo (viviendas) a encuestar (ver Tabla II.1A). Así, se selecciona aleatoriamente uno de los grupos y posteriormente en su interior una manzana. El proceso se repite iterativamente hasta completar el total de manzanas a seleccionar.



Tabla II.1A: Número de manzanas y viviendas en el Marco Muestral de Manzanas, según grupo de tamaño de manzanas.

Grupo de tamaño	Rango de viviendas	Total de Manzanas	% de manzanas	Total de Viviendas	% de viviendas
Grupo 0	1 a 7	13.894	10,4	53.578	1,3
Grupo 1 a 4	8 a 23	63.319	47,5	994.825	24,9
Grupo 5 a 9	24 a 44	39.267	29,4	1.216.7	30,4
Grupo 10 a	45 a 81	10.334	7,7	591.573	14,8
Grupo 20 a 28	82 a 154	3.990	3,0	439.327	11,0
Grupo 29 y	155 y más	2.556	1,9	704.695	17,6
Total		133.3	100,0	4.000.7	100,0

Fuente: Instituto Nacional de Estadísticas.

En la Tabla II.1B se puede apreciar la muestra de manzanas seleccionada para Casen 2013 y sus respectivas viviendas.

La comparación de las Tablas II.1A y II.1B permite apreciar que de las manzanas seleccionadas en la muestra Casen 2013, 66,5% tienen entre 8 y 44 viviendas (subrepresentadas en comparación al MM2008, en donde esta categoría alcanza 76,9%) y 6,8% tienen 155 viviendas y más (sobrerepresentadas en comparación al MM2008, en donde esta categoría alcanza 1,9% de las manzanas del marco muestral).

Tabla II.1B: Número de manzanas y viviendas seleccionadas para la muestra de Casen 2013 en el Marco Muestral de Manzanas del INE, según grupo de tamaño de manzanas.

Grupo de tamaño	Rango de viviendas	Total de Manzanas	% de manzanas	Total de Viviendas	% de viviendas
Grupo 0	1 a 7	0	0,0	0	0,0
Grupo 1 a 4	8 a 23	962	13,0	3.461	4,8
Grupo 5 a 9	24 a 44	3.972	53,5	26.217	36,1
Grupo 10 a 19	45 a 81	1.691	22,8	21.229	29,2
Grupo 20 a 28	82 a 154	286	3,9	6.722	9,3
Grupo 29 y 30	155 a 1.043	507	6,8	15.034	20,7
Total		7.418	100,0	72.663	100,0

Fuente: Instituto Nacional de Estadísticas.

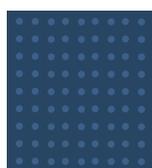
Nótese que el INE descarta del marco muestral las manzanas de tamaño inferior a 7 viviendas previamente a la selección de la muestra para Casen 2013, por lo que esta muestra no cuenta con manzanas de este tamaño, de acuerdo a lo descrito en la siguiente sección.

1.3. Cobertura del Marco Muestral

Las unidades primarias de muestreo (secciones y manzanas) 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 es la falla no planificada por el investigador que excluye del marco muestral, previa selección, ciertos elementos o unidades completas de la población a encuestar, como por ejemplo las fallas en el proceso de conteo e identificación de las viviendas (Kish 1965, pág. 528).

Es importante diferenciar la falta de cobertura (fallas no intencionadas) de aquellas exclusiones que realiza el investigador en forma intencionada.

En el proceso de selección de la muestra Casen 2013 hubo tres tipos de exclusiones intencionadas:

1. las 22 áreas geográficas que han sido catalogadas por INE como áreas de difícil acceso (ADA)⁶;
2. las manzanas y secciones incluidas en otras muestras seleccionadas por el INE para el periodo de recolección de datos de Casen 2013;
3. las manzanas con 7 viviendas o menos, que corresponden al grupo 0 descrito en la Tabla II.1.A, que aparece con frecuencias y porcentajes nulos en la Tabla II.1.B.

El marco muestral de la muestra Casen 2013 (MS2002 y MM2008 en su conjunto) excluye intencionalmente 3,6% de las viviendas del país.

Para algunos efectos, esta cifra puede parecer insignificante, sin embargo es importante que los investigadores que hagan inferencias a partir de los datos Casen la consideren, en particular si los fenómenos de su interés están relacionados con el tamaño de las manzanas (en áreas urbanas) o con la aislación geográfica (en áreas rurales). En ciertos casos puntuales, la falta de cobertura del marco muestral, acotada a los aspectos anteriormente descritos, pudiese incidir en las inferencias hechas a partir de los datos de la encuesta.

1.3.1. Áreas de Dificil Acceso

Las áreas de difícil acceso son zonas geográficas no incluidas en el marco muestral del INE, y no forman parte de las encuestas de hogares cuya muestra es seleccionada por el INE. En total, corresponden al 0,4% de la población de viviendas y al 0,4% de la población de personas del país. La Tabla II.2 muestra las 22 áreas de difícil acceso excluidas de la muestra Casen 2013, en donde se presenta el total de viviendas, de acuerdo al INE en la actualidad, basada en los datos del Censo de Población y Vivienda 2002, y una aproximación del total de personas que allí vivirían, estimadas mediante las proyecciones de población con fecha Noviembre de 2011⁷.

⁶ Ver más antecedentes acerca del marco muestral maestro del INE en: http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/pdf/ene_espanol.pdf.

⁷ También basadas en el Censo de Población y Vivienda del año 2002.

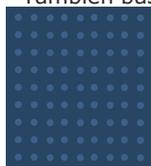


Tabla II.2: Áreas de Difícil Acceso definidas por el INE

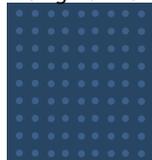
Región	Provincia	Comuna	Total Viviendas Particulares ocupadas Censo 2002	Total de personas Proyecciones de Población*
Arica y Parinacota	Parinacota	General Lagos	260	1.227
Tarapacá	Tamarugal	Colchane	461	1.623
Antofagasta	El Loa	Ollagüe	66	245
Valparaíso	Valparaíso	Juan Fernández	206	869
	Isla de Pascua	Isla de Pascua	1.136	5.090
Los Lagos	Llanquihue	Cochamó	1.345	4.295
	Palena	Chaitén	1.830	7.037
		Futaleufú	606	1.819
		Hualaihué	2.249	8.394
		Palena	558	1.642
Aysén del General Carlos Ibáñez del Campo	Coihaique	Lago Verde	338	935
	Aysén	Guaitecas	383	1.855
	Capitán Prat	O'Higgins	154	688
		Tortel	145	641
Magallanes y La Antártica Chilena	Magallanes	Laguna Blanca	116	631
		Río Verde	86	367
		San Gregorio	212	714
	Antártica Chilena	Cabo de Hornos (Ex - Navarino)	520	2.816
		Antártica	10	90
	Tierra el Fuego	Primavera	228	572
		Timaukel	82	840
	Última Esperanza	Torres del Paine	114	1.137

*30 Noviembre de 2011.

Fuente: Instituto Nacional de Estadísticas.

1.3.2. Manzanas y Secciones incluidas en otras muestras seleccionadas por el INE para el periodo de recolección de datos de Casen 2013⁸

Dado el levantamiento mensual de la Encuesta Nacional del Empleo (ENE), y su coincidencia temporal con el levantamiento de la Encuesta Casen durante los meses de noviembre, diciembre y enero, para evitar complicaciones en el trabajo de campo de ambas encuestas, el 2012 el INE excluyó del marco de muestreo previo a la selección

⁸ Originalmente seleccionada en 2012.

de la muestra Casen 2012 (que fue utilizada en Casen 2013), todas aquellas manzanas y secciones que formaban parte de la muestra ENE.

Al igual se excluyeron todas las manzanas y secciones de la muestra de la encuesta del Servicio Nacional para la Prevención y Rehabilitación del Consumo de Drogas y Alcohol (Senda) 2012.

Así, el total de unidades excluidas del marco de muestreo por ese motivo correspondieron a 2,7% de manzanas (2,8% de viviendas aproximadamente) en el MM2008, y 6,6% de secciones (6,5% de viviendas aproximadamente) en el MS2002.

Esto equivale, aproximadamente, a 3.661 manzanas y 503 secciones según trimestre móvil⁹. En las tablas II.3A y II.3B, se presentan las distribuciones absolutas y relativas de viviendas y *secciones* excluidas por región y área geográfica (urbana o rural) en el MS2002. En las tablas II.4A y II.4B, se presentan las distribuciones absolutas y relativas de viviendas y manzanas excluidas por región y área geográfica (urbana o rural) en el MM2008.

⁹ 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.

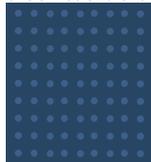


Tabla II.3A: Distribuciones de frecuencias de viviendas y secciones excluidas desde el MS2002, según región y área.

Región	Número de excluidas			
	Rural		Urbano	
	Viviendas	Secciones	Viviendas	Secciones
Total País	35.473	457	2.953	40
I de Tarapacá	207	4	491	8
II de Antofagasta	50	1	0	0
III de Atacama	592	11	0	0
IV de Coquimbo	5.097	64	220	3
V de Valparaíso	2.074	28	156	2
VI de O'Higgins	4.131	58	388	5
VII del Maule	4.561	56	697	8
VIII del Biobío	2.848	37	164	2
IX de La Araucanía	3.703	45	0	0
X de Los Lagos	6.068	74	143	2
XI de Aysén	353	7	324	5
XII de Magallanes y Antártica	92	2	67	1
XIII Metropolitana	2.579	25	0	0
XIV de Los Ríos	2.056	25	303	4
XV de Arica y Parinacota	1.062	20	0	0

Fuente: Instituto Nacional de Estadísticas.

Tabla II.3B: Distribuciones relativas de viviendas y secciones excluidas desde el MS2002, según región y área.

Región	Porcentaje excluidas					
	Rural		Urbano		Total	
	Viviendas	Secciones	Viviendas	Secciones	Viviendas	Secciones
Total País	6,9%	7,0%	3,8%	4,1%	6,5%	6,6%
I de Tarapacá	8,4%	8,0%	62,1%	61,5%	21,4%	19,0%
II de Antofagasta	2,6%	2,6%	0,0%	0,0%	2,1%	2,2%
III de Atacama	11,4%	10,6%	0,0%	0,0%	11,4%	10,6%
IV de Coquimbo	13,7%	13,4%	11,8%	13,0%	13,6%	13,4%
V de Valparaíso	7,0%	6,9%	2,9%	2,8%	6,5%	6,3%
VI de O'Higgins	7,0%	7,4%	4,6%	4,5%	6,7%	7,1%
VII del Maule	5,6%	5,6%	4,5%	4,2%	5,4%	5,4%
VIII del Biobío	3,3%	3,3%	1,1%	1,1%	3,0%	3,0%
IX de La Araucanía	4,9%	4,8%	0,0%	0,0%	4,2%	4,1%
X de Los Lagos	10,9%	10,8%	1,6%	1,9%	9,6%	9,6%
XI de Aysén	7,6%	7,6%	14,1%	13,5%	9,8%	9,3%
XII de Magallanes y	9,8%	10,7%	5,1%	4,5%	7,6%	8,0%
XIII Metropolitana	5,8%	5,8%	0,0%	0,0%	5,6%	5,7%
XIV de Los Ríos	7,1%	6,9%	11,5%	12,1%	7,4%	7,3%
XV de Arica y Parinacota	36,5%	35,6%	0,0%	0,0%	34,0%	33,3%

Fuente: Instituto Nacional de Estadísticas.

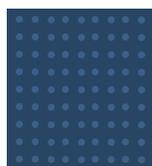


Tabla II.4A. Distribuciones de frecuencias absolutas y relativas de viviendas y manzanas excluidas desde el MM2008, según región.

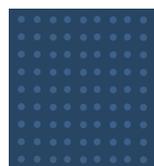
Región	Manzanas con 7 o menos viviendas				Excluido ENE y Senda 2012			
	Frecuencias Absolutas (n)		Frecuencias relativas (%)		Frecuencias Absolutas (n)		Frecuencias relativas (%)	
	Viviendas	Manzanas	Viviendas	Manzanas	Viviendas	Manzanas	Viviendas	Manzanas
Total País	53.5	13.89	1,3%	10,4%	111.0	3.661	2,8%	2,7%
I de Tarapacá	652	194	0,9%	8,8%	3.532	104	5,0%	4,7%
II de Antofagasta	2.524	710	1,9%	12,6%	3.956	150	3,0%	2,7%
III de Atacama	2.869	763	3,8%	19,1%	2.869	103	3,8%	2,6%
IV de Coquimbo	3.605	899	2,4%	13,2%	4.348	165	2,8%	2,4%
V de Valparaíso	12.58	3.004	2,5%	15,6%	18.024	589	3,5%	3,1%
VI del O'Higgins	3.071	809	1,9%	13,8%	4.088	132	2,6%	2,2%
VII del Maule	2.572	614	1,4%	9,2%	4.624	179	2,5%	2,7%
VIII del Biobío	7.481	1.893	1,7%	11,5%	16.548	544	3,7%	3,3%
IX de La Araucanía	3.261	849	1,9%	11,9%	4.756	201	2,8%	2,8%
X de Los Lagos	2.317	604	1,7%	11,6%	4.476	167	3,3%	3,2%
XI de Aysén	487	132	2,5%	13,3%	2.656	110	13,4%	11,1%
XII de Magallanes	899	187	2,1%	9,2%	1.438	75	3,4%	3,7%
XIII Metropolitana	9.116	2.658	0,5%	5,8%	33.546	880	1,9%	1,9%
XIV de Los Ríos	1.044	286	1,6%	11,2%	2.275	70	3,4%	2,7%
XV de Arica y	1.092	292	2,1%	12,1%	3.864	192	7,3%	8,0%

Fuente: Instituto Nacional de Estadísticas.

Tabla II.4B: Distribuciones de frecuencias absolutas y relativas de viviendas y manzanas excluidas desde el MM2008, según región.

Región	Total			
	Frecuencias absolutas (n)		Frecuencias relativas (%)	
	Viviendas	Manzanas	Viviendas	Manzanas
Total País	164.578	17.555	4,1%	13,2%
I de Tarapacá	4.184	298	6,0%	13,6%
II de Antofagasta	6.480	860	4,9%	15,3%
III de Atacama	5.738	866	7,7%	21,7%
IV de Coquimbo	7.953	1064	5,2%	15,6%
V de Valparaíso	30.612	3.593	6,0%	18,7%
VI del O'Higgins	7.159	941	4,5%	16,0%
VII del Maule	7.196	793	3,9%	11,8%
VIII del Biobío	24.029	2.437	5,4%	14,8%
IX de La Araucanía	8.017	1050	4,7%	14,7%
X de Los Lagos	6.793	771	5,0%	14,8%
XI de Aysén	3.143	242	15,9%	24,4%
XII de Magallanes y	2.337	262	5,5%	12,9%
XIII Metropolitana	42.662	3.538	2,4%	7,7%
XIV de Los Ríos	3.319	356	5,0%	14,0%
XV de Arica y Parinacota	4.956	484	9,4%	20,0%

Fuente: Instituto Nacional de Estadísticas.



1.3.3. Manzanas con 7 viviendas o menos

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 el 10,4% del total de manzanas, correspondientes a 1,3% de viviendas del marco muestral MM2008.

1.4. Presencia de comunas en la muestra

Al igual que en 2011, el nivel de estimación más desagregado en la encuesta Casen 2013 es el nivel regional, lo que implica la **presencia** de las comunas en la muestra. La presencia de las comunas no les atribuye representatividad estadística en la muestra, al no contar para todas las comunas con tamaños muestrales suficientes para generar estimaciones confiables y con niveles de precisión adecuados.

Cabe mencionar sin embargo, que para subsanar la carencia de estimaciones directas, válidas a nivel comunal, el Ministerio de Desarrollo Social pone a disposición documentos técnicos para generar estimaciones estadísticamente válidas a nivel comunal, basadas en la metodología de estimación para Áreas Pequeñas¹⁰.

Al igual, al interior de cada comuna se conforma la muestra unidades seleccionadas de manzanas y secciones lo que implica presencia de áreas geográficas urbanas y rurales al interior de las comunas y de las regiones. Sin embargo los tamaños muestrales urbanos y rurales así definidos al interior de comunas no permite calcular estimaciones con niveles adecuados de precisión, y por tanto es importante también recordar que la muestra es diseñada con representatividad de áreas geográficas (urbana y rural) a nivel nacional.

1.5. Estratificación del Marco Muestral por comunas y división censal urbano-rural

La estratificación es el proceso de clasificación de las unidades primarias de muestreo en grupos (o estratos), previo a la selección de la muestra. Se adopta esta estrategia con el fin de mejorar la precisión estadística de los estimadores.

Los estratos deben cumplir con las siguientes condiciones:

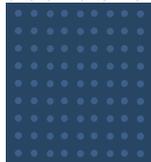
- Los estratos deben ser lo más homogéneos posibles y deben diferenciarse lo mejor posible.
- Los estratos han de ser mutuamente excluyentes, es decir que cada unidad poblacional debe ser asignada a un solo estrato.
- Los estratos deben ser exhaustivos colectivamente, es decir, ningún elemento de la población puede quedar excluido de la estratificación.

Para Casen 2013, se conformaron 585 estratos, a partir del cruce de la división político-administrativa (a nivel de comunas) y la división censal (urbano-rural).

Se entiende entonces por estrato el área geográfica urbana o rural de una comuna.

Como no todas las comunas presentan ambas áreas geográficas urbana y rural, 312 estratos se encuentran en áreas geográficas urbanas y 273 en áreas geográficas rurales.

¹⁰ Ver documentos metodológicos "Small Area Estimation", SAE de Casen 2011, disponibles en: <http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen-documentos.php?c=82&a=2011> (1 diciembre 2014).



2. Estimación del tamaño muestral

En Casen 2013, el tamaño de la muestra objetivo o esperada en el período noviembre 2013-enero 2014 era de 70.080 viviendas.

El tamaño muestral fue determinado en forma independiente para cada una de las regiones del país, siendo la variable de interés la tasa de pobreza. Se determinó que el error absoluto asociado a la tasa de pobreza regional debía fluctuar entre 1 y 4 puntos porcentuales y el error relativo no debía superar 30%.

La repartición del tamaño muestral entre las regiones se realizó respetando el peso relativo de las regiones dentro del país. La repartición del tamaño al interior de cada región operó en forma proporcional al tamaño de la población de viviendas entre las comunas, como también la distribución en las partes urbana o rural de éstas.

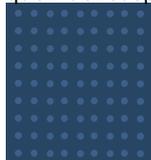
2.1. Proceso de cálculo

En la muestra Casen 2013, la determinación del tamaño muestral es un proceso que consta de cinco etapas. Las fórmulas empleadas utilizan las notaciones a continuación:

p_r	<i>La tasa de pobreza en la región r.</i>
n_r	<i>Número de unidades primarias de muestreo (manzanas o secciones) en la región r.</i>
\bar{m}_r	<i>Número promedio de viviendas a encuestar por unidad primaria de muestreo (manzana o sección) en la región r.</i>
$m_r = n_r * \bar{m}_r$	<i>Número de viviendas a encuestar en la región r.</i>
M_r	<i>Número de viviendas del marco muestral en la región r.</i>
$Def(p_r)$	<i>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, al considerar un muestreo complejo (estratificado, bietápico, por conglomerados) en vez de un muestreo aleatorio simple de viviendas.</i>
$SE(p_r)$	<i>Error estándar de la estimación de la pobreza p en la región r, igual a la raíz cuadrada de la varianza de la estimación.</i>
$S^2(p_r)$	<i>Coasivarianza poblacional de la tasa de pobreza en la región r. Para variables cualitativas y en un muestreo aleatorio simple, corresponde aproximadamente a $\left(\frac{M_r}{M_r-1}\right)(p_r)(1-p_r)$, en donde p_r es la tasa de pobreza en la región r y M_r el número de viviendas del marco muestral en la región r.</i>
$Z_{1-\frac{\alpha}{2}}$	<i>Percentil de nivel $\left(1-\frac{\alpha}{2}\right)$ de la distribución Normal, correspondiente a una estimación por intervalo de $(1-\alpha)$ de confianza.</i>

A continuación se describen los cinco pasos realizados para definir los tamaños muestrales regionales la muestra¹¹.

¹¹ En primera instancia, la muestra total de Casen 2013 (diseñada originalmente para Casen 2012, encuesta que fue suspendida previa al trabajo de campo) se distribuyó por región y luego, se dividió en dos



Paso 1: Cálculo de las cuasi varianzas de las tasas de pobreza regionales.

Usando datos de Casen 2009, se calcularon las estimaciones regionales de la tasa de pobreza, los errores estándar asociados a las tasas de pobreza regionales y los efectos de diseño regionales asociados a las tasas de pobreza regionales usando Casen 2009¹².

La cuasi-varianza de la tasa de pobreza regional¹³ $S^2(p_r)$ se calculó a partir del error estándar de la tasa de pobreza regional 2009, $SE(p_r)$, utilizando la siguiente fórmula:

$$S^2(p_r) = \frac{n_r * \bar{m}_r}{Def(p_r)} \cdot SE^2(p_r) = \frac{m_r}{Def(p_r)} \cdot SE^2(p_r) \quad (1)$$

Paso 2: Cálculo de los tamaños muestrales de las regiones considerando un muestreo aleatorio simple monoetápico.

Se establecieron parámetros para los errores absolutos "objetivo" por región (e_{0r}), los que deberán variar entre 1 y 4 puntos porcentuales, al igual que para el error relativo regional para la estimación de la tasa de pobreza, que deberá ser inferior a 30%.

También se estableció el nivel de confianza al 95% para obtener el percentil de la distribución normal $Z_{1-\alpha/2}$.

Con estos parámetros y las cuasi-varianzas regionales de las tasas de pobreza calculadas en (1), se calcularon los tamaños muestrales regionales considerando un muestreo aleatorio simple monoetápico m_{0r} .

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

Paso 3: Ajuste por el efecto de diseño de los tamaños muestrales regionales.

Se ajustaron los tamaños muestrales regionales obtenidos en el paso anterior considerando el efecto de diseño al aplicar un muestreo en dos etapas.

El efecto del diseño regional es el ajuste que se le aplica al tamaño muestral regional considerando un muestreo aleatorio simple monoetápico, tomando en cuenta el cambio que experimentaría la varianza al pasar de un muestreo monoetápico a uno multietápico¹⁴, manteniendo el error absoluto inicial.

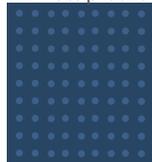
Por lo tanto, el tamaño muestral m_{1r} que incorpora el efecto del diseño se calcula como:

submuestras, correspondiendo aproximadamente a 33% y 67% para las submuestras 1 y 2 respectivamente. Sin embargo, por recomendación del Panel de Expertos Casen, comité asesor ministerial que ha acompañado y supervisado el proceso Casen 2013, finalmente las dos submuestras se levantaron durante un solo período de recolección de datos, como se describe más adelante en la sección "2.2. Distribución inicial de la muestra Casen 2013 en dos submuestras independientes y de mismo diseño."

¹² Estas etapas de la determinación del tamaño muestral se realizaron originalmente para la muestra de Casen 2012. Entonces no se contaba con los datos de Casen 2011, por lo que se emplearon los de Casen 2009.

¹³ La cuasi-varianza de la tasa de pobreza regional es la varianza de la estimación tasa de pobreza regional corregida por el número de viviendas de la región r , calibrada por el efecto de diseño asociado a la tasa de pobreza de la región.

¹⁴ Bietápico en el MS2002 y trietápico en el MM2008.



$$m_{1r} = m_{0r} \cdot Deff(p_r) \quad (3)$$

Donde

m_{0r} *Número de viviendas estimadas a encuestar en la región r bajo muestreo aleatorio simple.*

m_{1r} *Número de viviendas a encuestar en la región r ajustado por efecto de diseño (muestra compleja).*

$Deff(p_r)$ *Efecto del diseño asociado a la tasa de pobreza p en la región r.*

Se estimaron los efectos de diseño regionales de las tasas de pobreza regionales en Casen 2013, como el promedio de las encuestas Casen 2006 y 2009¹⁵, truncado a un máximo de 5. Sólo dos regiones que sobrepasaban este valor promedio, la región Metropolitana (5,2) y la de Los Ríos (5,8). Los efectos de diseño regional más bajos se observaron en Magallanes (1,9) y Aysén (2,1).

Tabla II.5: Efectos de diseño de Casen 2006, 2009, promedio y de Casen 2013.

Región	Casen 2006	Casen 2009	Promedio 2006 - 2009	Casen 2013 (*)
XV de Arica y Parinacota	3,3	2,6	3,0	3,0
I de Tarapacá	2,8	4,8	3,8	3,7
II de Antofagasta	3,7	4,7	4,2	4,2
III de Atacama	2,8	3,1	2,9	3,0
IV de Coquimbo	3,6	3,2	3,4	3,4
V de Valparaíso	3,8	4,4	4,1	4,1
VI de O'Higgins	2,4	2,0	2,2	2,2
VII del Maule	3,2	6,4	4,8	4,8
VIII del Biobío	2,9	3,4	3,2	3,2
IX de La Araucanía	1,9	3,0	2,5	2,5
X de Los Lagos	2,6	4,8	3,7	3,6
XIV de Los Ríos	2,8	8,8	5,8	5,0
XI de Aysén	2,7	1,4	2,1	2,1
XII de Magallanes	2,2	1,6	1,9	2,0
XIII Metropolitana	5,9	4,5	5,2	5,0
País	3,96	4,29	4,12	4,12

(*)Efectos de diseño regionales estimados.

Fuente: Instituto Nacional de Estadísticas.

Paso 4: Ajuste por finitud de los tamaños muestrales regionales.

En seguida se calibraron los tamaños muestrales regionales con el efecto de diseño al incorporarlos en los cálculos mediante la fórmula (3). A demás, se realizó un ajuste por muestreo en poblaciones finitas o corrección por finitud $(1 + m_{1r}/M_r)$ llegando al cálculo de m_{2r} :

¹⁵ Tal como se explicó anteriormente, al momento de determinar el tamaño muestral de Casen 2013 aún no se disponía de la información de la Casen 2011.



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

Donde

m_{2r} Número de viviendas estimadas a encuestar en la región r , mediante un diseño muestral multietápico¹⁶ que incorpora el efecto del diseño y el ajuste por finitud o ajuste para poblaciones finitas

M_r Número de viviendas de la población en la región r .

Paso 5: Ajuste por no respuesta a la unidad de los tamaños muestrales regionales.

Una vez definido el total regional de unidades muestrales a encuestar para estimar la tasa de pobreza, se consideró además que es posible no lograr ese total de unidades por diversas razones tales como rechazos, moradores ausentes, etc.

En este sentido, y con el objetivo de mantener la precisión de la estimación de la tasa de pobreza, se aplica un factor de ajuste basado en la tasa de no respuesta tnr_r al tamaño calculado en la fórmula (4).

Así para Casen 2013, se obtiene m_{3r} el número de viviendas final a encuestar en la región r con los 3 ajustes que se aplicaron secuencialmente (de diseño, por finitud y por no respuesta a la unidad); cuya fórmula está dada por:

$$m_{3r} = \frac{Z_{1-\alpha/2}^2 \cdot S^2(p_r) \cdot Deff(p_r)}{e_{0r}^2 + Z_{1-\alpha/2}^2 \cdot Deff(p_r) \cdot S^2(p_r)/M_r} \cdot \frac{1}{(1-tnr_r)} = \frac{m_{2r}}{1-tnr_r} \quad (5)$$

Este tamaño fue el objeto de un ajuste inferior al 2% a nivel nacional, debido a la imposición de un mínimo de unidades secundarias (50 viviendas en estrato urbano, correspondiente a 2 manzanas y de 30 viviendas en el caso de estrato rural, correspondiente a 2 secciones), para fines de cálculo de la varianza del estimador.

A continuación, en la Tabla II.6, se muestran los pasos involucrados en el proceso secuencial implementado para calcular el tamaño muestral de Casen 2013 y su correspondiente distribución regional.

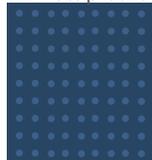
Paso 1: Cálculo de las cuasi varianzas de las tasas de pobreza regionales.

- La columna 2 presenta p_r los valores estimados de la tasa de pobreza regional en Casen 2009.

Paso 2: Cálculo de los tamaños muestrales de las regiones considerando un muestreo aleatorio simple monoetápico.

- La columna 3 presenta los errores absolutos regionales calculados d_i a partir de las estimaciones de la tasa de pobreza regional, distintos de los errores absolutos "objetivo" por región (e_{0r}), cuyos rangos son 1 a 4 puntos porcentuales y que determinan los tamaños muestrales iniciales (sin ajuste) m_{0r} .

¹⁶ Bietápico en el MS2002 y trietápico en MM2008.



- La columna 4 presenta los errores relativos calculados e_i , a partir de las estimaciones de la tasa de pobreza regional.
- La columna 5 presenta los valores de m_{0r} , los tamaños muestrales regionales iniciales sin ajustes.

Paso 3: Ajuste por el efecto del diseño de los tamaños muestrales regionales.

- La columna 6 muestra los efectos de diseño regionales $Def(p_r)$ en las estimaciones de las tasas de pobreza regionales.
- La columna 7 presenta m_{1r} los tamaños muestrales regionales iniciales m_{0r} ajustados por el efecto de diseño $Def(p_r)$.

Paso 4: Ajuste por finitud de los tamaños muestrales regionales.

- La columna 9 presenta m_{2r} los tamaños muestrales regionales ajustados por el efecto de diseño m_{1r} ajustados por finitud.

Paso 5: Ajuste por no respuesta a la unidad de los tamaños muestrales regionales.

- La columna 11 presenta m_{3r} , los tamaños muestrales ajustados por el efecto de diseño regional, por finitud y por tasas de no respuesta regionales tnr_r .
- La columna 12 presenta m'_{3r} , los tamaños muestrales objetivos regionales, que han sido ajustados en el momento de la selección para garantizar la selección de al menos dos conglomerados en cada estrato, además de 50 viviendas en estratos urbanos y 30 en estratos rurales.

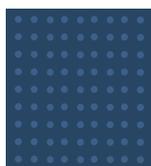


Tabla II.6. Tamaño muestral inicial y parámetros utilizados, por región, Casen 2013.

	(1) Población Viviendas año 2009 ¹⁷	(2) Tasa de Pobreza Casen 2009	(3) Error Absoluto Propuesto (en p. p.)	(4) Error Relativo Propuesto (%)	(5) Tamaños Propuestos Muestreo Aleatorio Simple (M.A.S)	(6) Efecto Diseño	(7) Tamaños m_{0r} ajustados por efecto de diseño	(8) Fracción de muestreo	(9) Tamaños m_{1r} ajustados por finitud	(10) Tasa de No Respuesta (tnr_r) Casen 2009	(11) Tamaños m_{2r} ajustados por tasa de no respuesta	(12) Tamaños Ajustados en Selección ¹⁸
Región		p_r	d_i	e_i	m_{0r}	$Def f(p_r)$	m_{1r}	f_r	m_{2r}	tnr_r	m_{3r}	m'_{3r}
XV de Arica y Parinacota	43.429	0,128	2,40	18,7	1.002	3,0	3.007	6,9%	2.812	15%	3.308	3.332
I de Tarapacá	71.684	0,158	3,00	18,9	918	3,7	3.397	4,7%	3.244	20%	4.055	4.084
II de Antofagasta	123.340	0,080	2,30	28,7	665	4,2	2.792	2,3%	2.730	20%	3.413	3.449
III de Atacama	68.332	0,175	3,80	21,7	772	3,0	2.317	3,4%	2.241	20%	2.801	2.843
IV de Coquimbo	198.469	0,166	2,80	16,9	970	3,4	3.297	1,7%	3.243	20%	4.054	4.139
V de Valparaíso	491.287	0,151	1,95	13,0	1.786	4,1	7.322	1,5%	7.215	25%	9.620	9.866
VI de O´Higgins	244.622	0,128	2,00	15,6	2.522	2,2	5.549	2,3%	5.426	15%	6.384	6.633
VII del Maule	288.879	0,207	3,35	16,1	1.100	4,8	5.281	1,8%	5.186	10%	5.762	5.945
VIII del Biobío	553.948	0,210	1,85	8,8	3.349	3,2	10.718	1,9%	10.515	15%	12.371	12.693
IX de La Araucanía	267.546	0,271	2,55	9,4	2.353	2,5	5.882	2,2%	5.755	15%	6.771	6.957
X de Los Lagos	228.440	0,142	2,80	19,7	1.262	3,6	4.542	2,0%	4.453	15%	5.239	5.411
XIV de Los Ríos	107.635	0,204	4,00	19,6	799	5,0	3.996	3,7%	3.853	15%	4.533	4.601
XI de Aysén	28.211	0,151	4,00	26,5	1.079	2,1	2.265	8,0%	2.097	20%	2.621	2.674
XII de Magallanes	45.118	0,091	2,00	22,1	1.199	2,0	2.399	5,3%	2.278	20%	2.848	2.861
XIII Metropolitana	1.833.728	0,115	0,95	8,2	2.828	5,0	14.141	0,8%	14.032	25%	18.709	18.833
País	4.594.668	0,151	0,67	4,42	22.605	4,12	76.906	1,7%	75.080	20,3%	92.489	94.321

Fuente: Instituto Nacional de Estadísticas.

¹⁷ Para estas estimaciones se excluyeron las áreas censadas en Casen 2009. En efecto, puntualmente, en Futaleufú, Palena y Hualaihue (que son Áreas de Dificil Acceso para el INE, todas de la Xª Región, que se integraron en la muestra Casen 2006, se mantuvieron en la muestra de Casen 2009, pero que se descartaron del marco muestral previa selección de la muestra de Casen 2011), el tamaño muestral resultó igual al tamaño total de viviendas, y en consecuencia, todas las unidades resultaron encuestadas (la información levantada allí fue censal y no muestral).

¹⁸ Los tamaños finales fueron modificados, con el objetivo de garantizar la selección de al menos dos conglomerados en cada estrato, además de 50 viviendas en estratos urbanos y 30 en estratos rurales.

Metodolog

Para determinar los tamaños muestrales objetivo y con sobremuestreo independientemente por regiones, en Casen 2013, se consideraron las tasas de no respuesta de Casen 2011 (submuestra noviembre 2011-enero 2012).

En Casen 2011 (muestra noviembre 2011-enero 2012), las tasas de no respuesta regionales más altas se observaron en las regiones de Copiapó (38%), Magallanes (36%), Antofagasta (32%), Tarapacá (32%), Valparaíso (30%) y Metropolitana (29%); mientras que la tasa de no-respuesta más baja se observa en región del Maule (11%).

La muestra Casen 2011 fue diseñada fijando estimaciones para niveles de precisión de tasas de pobreza a nivel regional comprendidos entre 1 y 4 puntos porcentuales, y por esta razón, se requirió fijar tamaños muestrales relativamente similares para lograr similares niveles de precisión en cada región.

El levantamiento de Casen 2011 generó externalidades negativas en las operaciones de levantamiento de la Casen 2011, en las regiones extremas, debido a un aumento sobresaliente de muestra en las regiones extremas respecto a la proporcionalidad general dada por el tamaño relativo de las regiones de Acuerdo al Censo de Población y Vivienda 2002.

Tomando en cuenta lo anterior es que en Casen 2013 se decidió reducir los tamaños muestrales en las regiones extremas, aumentando los tamaños muestrales de las regiones de Valparaíso, del Biobío y Metropolitana, y respetando así el orden de los "pesos" regionales de manera similar a la distribución de la Casen 2009, acercándose entonces a la distribución relativa regional del Censo de población y vivienda 2002, manteniendo claro está el rango de variación de los errores absolutos entre 1 y 4 puntos porcentuales para la tasas de pobreza y fijando como error relativo máximo 30%.

La Tabla II.4 permite comparar las distribuciones regionales de la muestra objetivo en las tres últimas muestras de la encuesta Casen con la distribución regional de viviendas vigentes (Censo de Población y Viviendas de 2002), para así poder apreciar la similitud de ambos perfiles.

Las dos primeras columnas de la Tabla II.4 presentan el total de viviendas y su distribución regional según el marco muestral desarrollado en base al Censo de Población y Viviendas de 2002. Las columnas siguientes presentan el total de viviendas y la distribución regional de las muestras de las encuestas Casen 2009 y 2011 (muestra completa). Al comparar las dos primeras muestras, se observa que la distribución regional de la muestra 2009 se asemeja más a la distribución de la población de viviendas 2002, lo que es consistente con el diseño muestral empleado hasta 2009, que asignaba mayor tamaño muestral a las regiones con más comunas (las que contaban con un mayor número de viviendas).

Finalmente, en Casen 2013, el aumento del tamaño muestral de las regiones de Valparaíso, del Biobío y Metropolitana tuvo como efecto la disminución de los tamaños relativos de las regiones extremas dentro de la muestra total, estando así más cercana del perfil regional de la muestra de Casen 2009 y a la del Marco Muestral de viviendas de 2002 que de la distribución relativa de la muestra por regiones en Casen 2011.

Tabla II.7: Distribución de la muestra objetivo inicial, encuestas Casen 2009, 2011 y 2013.

Región	Marco Muestral de Viviendas 2002		Casen 2009		Casen 2011		Casen 2013	
	Población	%	Muestra	%	Muestra	%	Muestra	%
XV de Arica y Parinacota	43.429	0,9	764	1	3.495	3,9	2.812	3,7
I de Tarapacá	71.684	1,6	1.418	1,9	6.241	6,9	3.244	4,3
II de Antofagasta	123.340	2,7	1.995	2,7	6.108	6,8	2.730	3,6
III de Atacama	68.332	1,5	1.953	2,6	4.382	4,9	2.241	3,0
IV de Coquimbo	198.469	4,3	3.255	4,4	3.780	4,2	3.243	4,3
V de Valparaíso	491.287	10,7	8.045	10,8	8.064	8,9	7.215	9,6
VI de O'Higgins	244.622	5,3	6.874	9,2	5.469	6,1	5.426	7,2
VII del Maule	288.879	6,3	6.430	8,6	7.464	8,3	5.186	6,9
VIII del Biobío	553.948	12,1	11.850	15,9	8.652	9,6	10.515	14,0
IX de La Araucanía	267.546	5,8	6.760	9,1	5.782	6,4	5.755	7,7
X de Los Lagos	228.440	5,0	5.825	7,8	6.391	7,1	4.453	5,9
XIV de Los Ríos	107.635	2,3	2.535	3,4	6.576	7,3	3.853	5,1
XI de Aysén	28.211	0,6	1.300	1,7	3.941	4,4	2.097	2,8
XII de Magallanes	45.118	1,0	1.210	1,6	2.315	2,6	2.278	3,0
XIII Metropolitana	1.833.728	39,9	14.125	19	11.461	12,7	14.032	18,7
País	4.594.668	100	74.339	100	90.122	100	75.080	100

Fuente: Instituto Nacional de Estadísticas.

2.2. Distribución inicial de la muestra Casen 2013 en dos submuestras independientes y de mismo diseño

La muestra Casen 2013, fue diseñada para ser aplicada inicialmente en 2012 basándose en dos submuestras con mismo diseño, que correspondían aproximadamente al 33% y a 67% del tamaño total de la muestra. La primera submuestra tenía que ser levantada entre octubre y noviembre 2012 y la segunda entre noviembre 2012 y enero 2013 (coincidente con el período histórico de las encuestas Casen). Ante la suspensión de la encuesta Casen 2012, previo al levantamiento, la muestra fue utilizada en Casen 2013, considerando como distribución inicial la indicada en la Tabla II.8.



Tabla II.8: Distribución de la muestra objetivo inicial, encuesta Casen 2013.

Región	Muestra teórica de viviendas	SM1		SM2	
		Viviendas	%	Viviendas	%
XV de Arica y Parinacota	2.812	956	34,0	1.856	66,0
I de Tarapacá	3.244	1.102	34,0	2.142	66,0
II de Antofagasta	2.730	927	34,0	1.803	66,0
III de Atacama	2.241	758	33,8	1.483	66,2
IV de Coquimbo	3.243	1.070	33,0	2.173	67,0
V de Valparaíso	7.215	2.396	33,2	4.819	66,8
VI de O'Higgins	5.426	1.734	32,0	3.692	68,0
VII del Maule	5.186	1.696	32,7	3.490	67,3
VIII del Biobío	10.515	3.476	33,1	7.039	66,9
IX de La Araucanía	5.755	1.886	32,8	3.869	67,2
X de Los Lagos	4.453	1.434	32,2	3.019	67,8
XIV de Los Ríos	3.768	1.284	33,3	2.569	66,7
XI de Aysén	2.097	700	33,4	1.397	66,6
XII de Magallanes	2.278	784	34,4	1.494	65,6
XIII Metropolitana	14.032	4.797	34,2	9.235	65,8
País	75.080	25.000	33,3	50.080	66,7

Fuente: Instituto Nacional de Estadísticas.

2.3. Reseña de los ajustes locales en el tamaño de la muestra de Casen 2013

El año 2013 el levantamiento de la encuesta Casen 2013 fue licitado públicamente, pero la licitación debió declararse desierta por falta de oferentes. A objeto de viabilizar la realización de la encuesta Casen 2013, el Ministerio de Desarrollo Social, solicitó al INE estudiar la factibilidad de reducir el tamaño de la muestra teórica de Casen 2013, disminuyendo de 25.000 a 20.000 las entrevistas a lograr en la primera submuestra (diseñada para ser recolectada entre octubre y noviembre), y manteniendo sin modificación la segunda submuestra (la que se levantaría en período históricamente comparable).

La operación resultó delicada operativamente debido a que se intentaba mantener al menos dos unidades primarias de muestreo por estrato, teniendo menos tamaño muestral a repartir proporcionalmente entre las regiones. La Tabla II.9 traduce el resultado de esta reducción de tamaño muestral en 5.000 unidades, en el perfil regional de la muestra teórica de viviendas y las correspondientes distribuciones en cada submuestra.

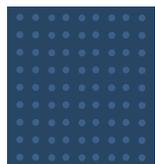


Tabla II.9: Distribución de la muestra objetivo inicial con disminución de 5.000 viviendas de la muestra (octubre-noviembre), Encuesta Casen 2013.

Región	Muestra teórica de viviendas	Muestra octubre-noviembre		Muestra noviembre-enero	
		Viviendas	%	Viviendas	%
XV de Arica y Parinacota	2.581	725	28,1	1.856	71,9
I de Tarapacá	2.977	835	28,0	2.142	72,0
II de Antofagasta	2.513	710	28,3	1.803	71,7
III de Atacama	2.071	588	28,4	1.483	71,6
IV de Coquimbo	3.018	845	28,0	2.173	72,0
V de Valparaíso	6.737	1.918	28,5	4.819	71,5
VI de O'Higgins	5.137	1.445	28,1	3.692	71,9
VII del Maule	4.872	1.382	28,4	3.490	71,6
VIII del Biobío	9.826	2.787	28,4	7.039	71,6
IX de La Araucanía	5.387	1.518	28,2	3.869	71,8
X de Los Lagos	4.192	1.173	28,0	3.019	72,0
XIV de Los Ríos	3.768	1.199	31,8	2.569	68,2
XI de Aysén	2.051	654	31,9	1.397	68,1
XII de Magallanes	2.085	591	28,3	1.494	71,7
XIII Metropolitana	12.865	3.630	28,2	9.235	71,8
País	70.080	20.000	28,5	50.080	71,5

Fuente: Instituto Nacional de Estadísticas.

Se estimaron los errores de muestreo para la muestra Casen 2013.

A nivel regional, en la muestra objetivo (70.080 viviendas), los errores absolutos esperados para la estimación de la tasa de pobreza regional se mantuvieron entre 1 y 4 puntos porcentuales y los errores relativos permanecieron bajo el 30%.

Es importante mencionar que hasta esta etapa, por separado, las submuestras de Casen 2013, no fueron diseñadas para cumplir separadamente los requisitos de precisión (error absoluto y relativo de la estimación de tasa de pobreza regional) establecidos para la muestra completa de Casen 2013.

En Octubre de 2013, el Panel de Expertos Casen recomendó implementar el levantamiento de la encuesta de Casen 2013 como una muestra única de viviendas, levantando ambas submuestras en el mismo período de recolección, de tal forma que la encuesta completa aplicada en el período históricamente comparable satisficiera los objetivos de precisión señalados aquel contemplado originalmente para la muestra (noviembre 2013 a enero 2014).

2.4. Ajuste de los tamaños de la muestra objetivo por tasa de no respuesta regional en Casen 2013.

A los tamaños muestrales regionales objetivo de cada muestra se les aplicó un doble ajuste: un ajuste basado en las tasas de no respuesta regionales y un ajuste que asegurase un número mínimo de unidades primarias por estrato.

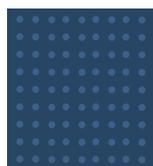


Tabla II.10: Distribución regional de la muestra objetivo inicial con sobremuestreo de viviendas, encuesta Casen 2013.

Región	Muestra objetivo		Muestra con sobremuestreo	
	Viviendas	%	Viviendas	%
XV de Arica y Parinacota	2.581	3,7	3.077	3,5
I de Tarapacá	2.977	4,2	3.770	4,3
II de Antofagasta	2.513	3,6	3.183	3,6
III de Atacama	2.071	3,0	2.622	3,0
IV de Coquimbo	3.018	4,3	3.821	4,4
V de Valparaíso	6.737	9,6	9.098	10,4
VI de O'Higgins	5.137	7,3	6.122	7,0
VII del Maule	4.872	7,0	5.484	6,3
VIII del Biobío	9.826	14,0	11.709	13,4
IX de La Araucanía	5.387	7,7	6.420	7,3
X de Los Lagos	4.192	6,0	4.997	5,7
XIV de Los Ríos	3.768	5,4	4.488	5,1
XI de Aysén	2.051	2,9	2.595	3,0
XII Magallanes	2.085	3,0	2.640	3,0
XIII Metropolitana	12.865	18,4	17.376	19,9
País	70.080	100,0	87.402	100,0

Fuente: Elaboración propia en base a datos del Instituto Nacional de Estadísticas, 2014.

2.5. Distribución de las viviendas por comunas.

A continuación se describe paso a paso el protocolo de distribución de la muestra Casen 2013, una vez establecido el tamaño muestral en cada región, en número de viviendas.

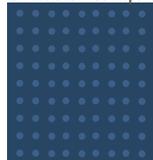
Paso 1

Al interior de cada región, la muestra de viviendas se distribuyó entre las comunas, en forma proporcional a los tamaños comunales dentro de sus respectivas regiones, dados por el número de viviendas que cada una posee en el marco muestral. En seguida, el tamaño comunal asignado de viviendas se distribuyó entre áreas urbanas y rurales de la comuna, también en forma proporcional al tamaño de viviendas que presenta en el marco muestral.

Como cada comuna podía tener áreas urbanas y/o rurales, se hizo necesaria una segunda distribución al interior de cada comuna según área geográfica. La distribución de viviendas al interior de cada comuna se realizó en función del peso o el aporte relativo de cada área geográfica en el total de viviendas de la comuna según el Censo 2002.

Cabe mencionar que se determinaron dos restricciones adicionales dentro del protocolo de distribución de la muestra 2013, al igual que en Casen 2011: se estableció un número mínimo de viviendas por estrato: 50 en estratos urbanos y 30 en estratos rurales¹⁹, para lograr la estimación de varianza bajo muestreo complejo al evitar, por diseño, la eventualidad de un solo conglomerado (manzanas o secciones) por estrato.

¹⁹ 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).



Se obtiene así una distribución de la muestra objetivo por regiones, con presencia en comunas y áreas geográficas urbana y rural.

Paso 2

El siguiente paso consistió en sobredimensionar la muestra objetivo (n=70.080) para contrarrestar las pérdidas que se originen producto de distintas formas de no-respuesta (ej. unidades no elegibles, falta de contacto, rechazo a la entrevista, etc.). Inicialmente se aplicaron tasas de no respuesta regionales y se ajustó por los tamaños mínimos de viviendas (50 en urbano y 30 en rural). El diseño 2013, realiza un ajuste diferenciado por región y comuna, tomando como referencia la tasa de no-respuesta observada de la submuestra noviembre-enero de Casen 2011.

En 2013 se utilizaron las tasas regionales de no-respuesta observadas en Casen 2011 (muestra noviembre 2011-enero 2012) "amortiguadas" y reducidas, basándose en el análisis las tasas de no respuestas regionales observadas en Casen anteriores y aquellas obtenidas en otras encuestas realizadas por el INE. Como resultado de este análisis se consideró que era posible eventualmente reducir las tasas de no respuesta en terreno al tener un mayor control y supervisión, y se redujeron al considerar que en Casen 2011 estaban sobrestimadas.

Paralelamente, también a nivel comunal se calcularon tasas de no respuesta como el promedio de las tasas observadas en Casen 2009 y Casen 2011 (muestra noviembre 2011-enero 2012), las que fueron suavizadas en 6 categorías. Las tasas de no-respuesta se pueden interpretar como una estimación de la probabilidad de no responder de los hogares en las comunas seleccionadas²⁰. La distribución de las tasas de no-respuesta a nivel comunal presentaba alta variabilidad en cada región. En atención a esto, se optó por definir cinco grandes categorías para las tasas de no-respuesta y utilizar esta aproximación fijando valores de tasas de no-respuesta "suavizadas" para estos intervalos en vez de las estimaciones directas de las tasas de respuesta comunales correspondientes.

Estas tasas suavizadas se aplicaron a nivel comunal para la determinación de tamaños muestrales comunales con sobremuestreo, sometiéndolos a la restricción de que la suma de los tamaños comunales con sobre muestreo derivados de éstas, no sobrepasaran el total regional pre-establecido.

La Tabla II.11 muestra que las tasas de no-respuesta suavizadas fueron fijadas con valores cercanos a los extremos superiores de cada intervalo (y no por ejemplo en su centro), con el objeto de aplicar un criterio conservador y asegurar el tamaño de muestra suficiente al corregir por este factor.

²⁰ Ver más detalles del marco conceptual de la no respuesta en encuestas de hogares en "Nonresponse in Household Surveys" (Groves and Couper, 1998).

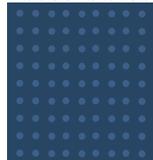
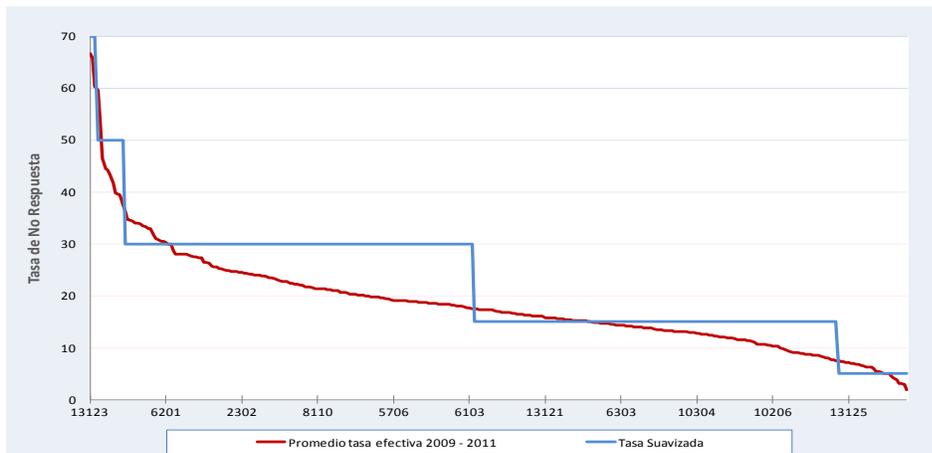


Tabla II.11: Reglas de paso de tasas de no respuesta observadas a tasas de no respuesta suavizadas.

Tasas de no respuesta	
Observadas	"suavizadas"
[0; 7,5[5
[7,5; 17,5[15
[17,5; 37,5[30
[37,5; 60[50
[60; 100]	70

Fuente: Elaboración propia en base a datos del Instituto Nacional de Estadísticas, 2014.

Para las tasas de no-respuesta suavizadas por comuna se calcularon los promedios de las tasas de no-respuesta de Casen 2009 y de la segunda submuestra de la encuesta Casen 2011, para enseguida aplicar el suavizamiento.

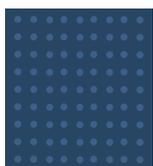
Figura II.3: Tasas de No-Respuesta suavizadas utilizadas para Sobredimensionar la muestra Casen 2013 en cada comuna.

Fuente: Instituto Nacional de Estadísticas

Paso 3

Finalmente, el sobredimensionamiento de la muestra generada por la no-respuesta a nivel comunal es ajustada para coincidir con la no-respuesta a nivel regional fijada inicialmente, empleando un simple ajuste de razón que re-distribuye la sobredimensión en forma proporcional al peso (en número de viviendas del marco muestral) de cada comuna en la región.

Una vez determinados los tamaños de muestra requeridos por comuna y región se procede a la selección de las unidades de muestreo. En la próxima sección se describen los procedimientos implementados con el marco de manzanas y el marco de secciones.



3. Selección de la muestra

3.1. Selección de conglomerados

Como se ha mencionado, el INE no dispone de un directorio de viviendas particulares del país. En cambio cuenta con un marco de áreas que contiene un listado de Unidades Primarias de Muestreo (UPM) o conglomerados de viviendas (secciones o manzanas) que sirven de base para selección inicial de unidades para la enumeración, la identificación y posterior selección de 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 (urbano o rural). A continuación se describen los protocolos utilizados para la selección de manzanas (a partir del MM2008) y de secciones (a partir del MS2002).

3.1.1. Selección de Conglomerados desde el MM2008

El marco muestral de manzanas está particionado en 119.466²¹ unidades y estratificado en 30 estratos, que reflejan el tamaño de las manzanas en el marco²². La selección de las manzanas se realiza en cada estrato de la muestra de forma independiente, aleatoria y proporcional al tamaño del total de viviendas y manzanas por cada grupo de tamaño²³.

Las manzanas pueden estar ubicadas en Ciudades (CD), áreas urbanas (U) o Resto de Áreas Urbanas (RAU). La selección se implementa en base a un algoritmo de selección que considera dos pasos, que se describen a continuación. El procedimiento descrito permite seleccionar en la muestra manzanas grandes y pequeñas.

Paso 1 Selección de un grupo de tamaño

Las manzanas se agrupan en i grupos de tamaño en el marco de manzanas (ver Tabla II.12). Esta información se cruza con las comunas, de modo que en cada comuna existan a lo más 30 grupos ($i = 1, 2, \dots, z$ donde $z \leq 30$).

El procedimiento aplicado es un proceso iterativo²⁴ en que se escoge en forma aleatoria un grupo de tamaño i proporcionalmente a su tamaño medido como W_i igual al producto entre el número total de manzanas y el número total de viviendas de cada grupo i , ($W_i = N_i * M_i$).

²¹ Sin considerar al grupo cero (manzanas con menos de 8 viviendas), se llega a 119.466.

²² Ver Tabla II.1A.

²³ Los grupos de tamaño son conglomerados de manzanas censales. Estos grupos de tamaño son las unidades primarias de muestreo que se utilizan en el procedimiento de selección implementado en el marco de manzanas de INE.

²⁴ En el programa R se aplica la función SAMPLE con probabilidad P_i .

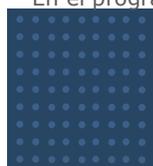


Tabla II.12: Método de selección del grupo de tamaño y manzanas.

Grupo de Tamaño	Número de Manzanas	Número de Viviendas	Peso del Grupo de Tamaño	Ponderación
G_i	N_i	M_i	$N_i * M_i$	$P_i = 1/W_i$
1	N_1	M_1	$N_1 * M_1$	P_1
2	N_2	M_2	$N_2 * M_2$	P_2
3	N_3	M_3	$N_3 * M_3$	P_3
⋮	⋮	⋮		
⋮	⋮	⋮		
⋮	⋮	⋮		
Z	N_Z	M_Z	$N_Z * M_Z$	P_Z
Total	N_h	M_h	$\sum W_i$	1

Fuente: Instituto Nacional de Estadísticas.

Donde,

$$N_h = \sum_{i=1}^z N_i \quad ; \quad M_h = \sum_{i=1}^z M_i \quad ; \quad W_i = N_i * M_i \quad ; \quad P_i = \frac{W_i}{\sum_{i=1}^z W_i} \quad ; \quad \sum_{i=1}^z P_i = 1$$

Paso 2: Selección de una manzana.

Una vez seleccionado el grupo de tamaño, la manzana se elige con igual²⁵ probabilidad.

Enseguida se descuenta la manzana (disminuye en uno el número de manzanas en el grupo y en el total de la comuna y/o estrato) y también se descuenta el número de viviendas²⁶ (disminuye el número de viviendas en el grupo seleccionado y en el total de la comuna y/o estrato), para finalmente actualizar tanto el número de manzanas y viviendas en cada grupo de tamaño y repetir el procedimiento hasta seleccionar todas las manzanas que completan el número de viviendas requeridas.

3.1.2. Selección de Conglomerados desde el MS2002

El marco muestral de secciones está particionado en 6.679 secciones en la parte Rural (R) y 968 secciones en la parte Resto de Áreas Urbanas (RAU). La selección preliminar de las secciones se realiza al interior de cada comuna en forma proporcional al total de viviendas ocupadas en ellas según información del Censo 2002. Esta selección se realiza en forma sistemática, proporcional al tamaño, 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) como se muestra en la Tabla II.13:

²⁵ Otra vez se aplica en R la función SAMPLE, pero ahora con Probabilidad = 1/N_i, donde N_i es el número de manzanas que contiene el grupo i.

²⁶ De acuerdo al número total de viviendas que se obtiene de la enumeración, cada manzana tiene predeterminado el total de viviendas que pueden ser seleccionadas para una muestra. Este total es consistente con una fracción de muestreo de aproximadamente 25% (ver tabla C en los Anexos). Para efectos de selección, sin embargo, se descuenta el número de viviendas enumeradas con el fin de recalcular las probabilidades de selección P_i.

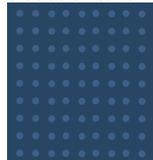


Tabla II.13: Método de selección de unidades primarias secciones.

Sección	N° Viviendas	Acumulado del número de viviendas	Límite inferior	Límite superior
i	M _i	ΣM _i	L _{inf}	L _{sup}
1	M ₁	Σ ₁ = M ₁	1	Σ ₁
2	M ₂	Σ ₂ = M ₁ + M ₂	Σ ₁ + 1	Σ ₂
3	M ₃	Σ ₃ = M ₁ + M ₂ + M ₃	Σ ₂ + 1	Σ ₃
⋮	⋮	⋮	⋮	⋮
Q-1	M _{Q-1}	Σ _{Q-1} = M ₁ + M ₂ + ... + M _{Q-1}	Σ _{Q-2} + 1	Σ _{Q-1}
Q	M _Q	Σ _Q = M ₁ + M ₂ + ... + M _Q	Σ _{Q-1} + 1	Σ _Q

Fuente: Instituto Nacional de Estadísticas.

Paso 2

Se calcula el período dado por: $K = \Sigma Q / n$ con ΣQ igual al total de viviendas de la comuna o estrato y n 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

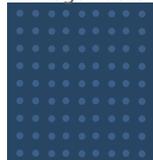
Se genera un número aleatorio entero A entre 1 y el período K .

Paso 4

Se suma 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 distinto²⁷, el procedimiento o algoritmo selecciona menos de las requeridas, porque 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 $n \cdot M_i / \Sigma Q$.

²⁷ Llegando a ser algunas el doble y más del doble de las demás.



3.2. Selección de viviendas

Como se desprende de la argumentación anterior, la selección de conglomerados (manzanas y secciones) se realiza en tres etapas en el marco de manzanas²⁸ y en dos etapas en el marco de secciones.

Previo a la selección de viviendas, se debe actualizar el registro del total de viviendas y la cartografía asociada a los conglomerados en la muestra. El Ministerio de Desarrollo Social ha encomendado al INE realizar también la enumeración de viviendas y la actualización del directorio y cartografía correspondientes. A diferencia de Casen 2011, en Casen 2013, este proceso se realizó en dos tiempos distintos:

- En el año 2012 donde se enumeró el 100% de la muestra (proceso Casen que se suspendió previo al levantamiento).
- En el año 2013, donde se re-enumeró un 30% de la muestra.

A continuación se detallará primero el proceso de enumeración de conglomerados y luego se explicitarán los métodos y criterios para definir aquellos conglomerados que forman parte de la enumeración realizada el año 2012 y el 30% actualizado durante el 2013.

3.2.1. Proceso de enumeración de los Conglomerados en 2012 y reglas de reemplazo en el proceso de enumeración

La selección inicial de conglomerados, ya sea a partir del marco de manzanas o de 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 y obedece a una pauta sistemática, la que se describe detalladamente a continuación.

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 2013, contempló áreas urbanas y rurales.

3.2.2. Material de trabajo del enumerador

A cada enumerador(a) se le entrega el siguiente set de materiales:

- Plano de ubicación comunal: Plano de la comuna donde se encuentran identificadas todas las manzanas y secciones seleccionadas en la muestra.
- Croquis de la manzana o sección: Plano a menor escala que identifica sólo la manzana o sección seleccionada que debe visitar el enumerador.
- Formulario de enumeración urbano o rural: en él se registran todas las construcciones encontradas en la manzana o sección según el recorrido u orden en que fueron encontradas en terreno²⁹.

²⁸ En el MM2008, la primera unidad de selección corresponde a un *grupo de tamaño* y la segunda unidad de selección corresponde a una *manzana* dentro del *grupo de tamaño* seleccionado.

²⁹ Ver Formulario de Enumeración en Áreas Urbanas en Figura IV.4 (los formularios de áreas rurales son similares).



Al recorrer la localidad o entidad de la sección o la manzana, el (la) enumerador(a) debe anotar en el formulario el uso o destino de cada una de las edificaciones, a medida que se vaya llegando a ellas.

Si es una vivienda, procede a anotar el número de orden de dicha vivienda y luego marca un punto en el plano de ubicación de la vivienda, y junto a él anota el mismo número de orden que registró en el Formulario de Enumeración.

3.2.2.1. Recorrido y Enumeración Urbana

Una vez que llega a una dirección, el (la) enumerador(a) debe confirmar que se encuentra en la manzana seleccionada, verificándolo en el plano de ubicación y croquis respectivo.

En seguida, debe situarse en la esquina norponiente de la manzana, identificada con un punto en el croquis, y comenzar el recorrido siguiendo el sentido de los punteros del reloj. Como método de control, la línea de edificación de la manzana debe estar siempre a la derecha del enumerador.

Durante el recorrido se realiza la enumeración de las viviendas, o de las unidades destinadas a otros usos, registrando los datos asociados a cada una de ellas en los campos correspondientes del Formulario de Enumeración.

3.2.2.2. Recorrido y Enumeración Rural.

El (la) enumerador(a) debe realizar la enumeración de las viviendas en forma correlativa dentro de cada localidad o entidad. Es decir, que en cada localidad, debe asignar un número correlativo de 1 al N (número total de viviendas en la localidad visitada) a cada vivienda en el área rural definida en el o los plano(s) que se le encargan de enumerar.

Siempre que sea posible, se debe iniciar el trabajo de un área rural desde los accesos señalados en sus límites hacia el interior de ella. El nombre de la localidad o entidad trazada se debe obtener de los propios habitantes de las viviendas enumeradas en el lugar. Ocasionalmente ellos la conocen por distintos nombres, en estos casos, predomina el que se menciona con mayor frecuencia.

3.2.2.3. Razones de reemplazos de conglomerados durante la enumeración

Durante el proceso de enumeración, algunos conglomerados todavía pueden ser objeto de reemplazo por diversos motivos:

- el conglomerado puede estar temporalmente destruido porque se van a levantar nuevas edificaciones (casas nuevas o edificios);
- la mayoría de las viviendas del conglomerado pueden haber cambiado de uso o destino a otro que no es residencial;
- 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, es el caso de fusiones o cuando un conglomerado se segmenta entre dos conglomerados debido al cambio de límites por cierre o construcción de nuevas calles (éste es el caso más frecuente).



Los motivos anteriores inciden conjuntamente en el remplazo de aproximadamente un 3% de las manzanas/secciones originalmente seleccionadas.

Para el remplazo de las manzanas y secciones se busca dentro del entorno geográfico un conglomerado con características similares en cuanto a

- Tamaño (número de viviendas)
- Ubicación (la misma comuna, distrito y zona, o la misma comuna y distrito, o la misma comuna, dependiendo si al nivel más desagregado se encuentra 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.

Figura II.4: Formulario de Enumeración en Áreas Urbanas, Casen 2013.

INSTITUTO NACIONAL DE ESTADÍSTICAS
 Subdirección de Operaciones
 Departamento de Estadísticas de Hogares
 Subdepartamento de Proyectos de Hogares
 Paseo Bulnes N° 209 Oficina 31, Santiago - Chile
 Teléfono: 62-27962400 / 62-27962417 / 62-27962448
 www.ine.cl

ENUMERACIÓN DE VIVIENDAS EN ÁREAS GEOGRÁFICAS URBANAS
 LOS DATOS QUE EL INE SOLICITA SON CONFIDENCIALES. EL ARTÍCULO 29 LEY 17.374
 ESTABLECE EL SECRETO ESTADÍSTICO

IDENTIFICACIÓN GEOGRÁFICA Hoja N° ___ de ___

Uso Exclusivo de Secciones: Estrato [] Sección []

Comuna [] Distrito [] Zona [] Manzana [] Nombre de Comuna []

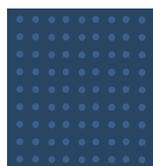
N° de Vivienda	N° de Orden en la Edificación	Dirección						Código del Registro	N° de Orden de la Vivienda	N° de Orden del Hogar (Incluir manzanas de 15 años)	Anotaciones complementarias		
		Tipo de Dirección	Nombre de Dirección	Número de Domicilio	Bloque o Torre	N° de Piso	N° o letra del Depto. o Píez				Añade el número del código en la columna correspondiente y detalla en la columna Descripción y Observación		
1		1 - Avenida 2 - Calle 3 - Paseo 4 - Centro 5 - Finca 6 - Otro	(Nombre de Dirección) (Otro: Indique nombre si corresponde)					1 - Morador 2 - Asesores 3 - Desocupado 4 - Temporal 5 - Sin Vivienda 6 - Otro/No			1 - Vivienda sin numeración (puede descripción detallada para identificar y diferenciar la vivienda) 2 - Cualquier anotación importante para identificar la vivienda y/o su ubicación al interior del catastro 3 - Fachazo 4 - Describir tipo de vivienda colectiva 5 - Observaciones sin vivienda 6 - Describir actividad de otro uso de la edificación	Código	Descripción y Observación
2													

Fuente: Instituto Nacional de Estadísticas.

3.2.2. Reenumeración del 30% unidades primarias de muestreo ya enumeradas en 2012

El año 2012, el Ministerio de Desarrollo Social encomendó al Instituto Nacional de Estadísticas la elaboración del diseño muestral de la encuesta Casen, lo que abarcaba el diseño y selección de la muestra, la elaboración del material cartográfico, la construcción y el cálculo de los factores de expansión, y la estimación de los errores de muestreo o precisión de la tasa de pobreza. Este trabajo fue llevado a cabo hasta la etapa de elaboración del material cartográfico. Sin embargo, previo al inicio del trabajo de campo, el gobierno decidió suspender la encuesta.

El año 2013, ante la decisión ministerial de implementar un nuevo levantamiento de la encuesta Casen, el Ministerio de Desarrollo Social solicitó nuevamente apoyo al INE para el diseño, selección y reenumeración de la muestra, elaboración de directorios y mapas de la nueva muestra, el cálculo de factores de expansión y de errores muestrales, la custodia de la llave Dirección-Folio, la recepción del Material Cartográfico entregado por el tercero designado por el Ministerio, y el desarrollo de estudios complementarios que fueran acordados entre ambas instituciones en materias vinculadas con la encuesta. Este acuerdo de colaboración quedó plasmado en el convenio de transferencia suscrito entre el Ministerio y el INE el 1 de marzo de 2013.



En resguardo de la inversión pública realizada el 2012 y de los recursos fiscales presupuestados en el marco del convenio de 2013, el Ministerio de Desarrollo Social evaluó junto con INE la forma de utilizar como muestra Casen 2013 las unidades primarias y secundarias de muestreo seleccionadas y enumeradas por el INE el año 2012.

Como resultado de esta evaluación, se diseñó una estrategia de actualización de la muestra Casen 2012, que diese mayor probabilidad de selección a aquellas unidades que mostraron un mayor dinamismo en la edificación de viviendas, de acuerdo a información proveniente de registros administrativos, y que diera alguna probabilidad de selección a aquellas con menor o cero crecimiento observado.

Así, con el objetivo de reducir posibles problemas de cobertura³⁰ debido al crecimiento de la población o al cambio de estado de las viviendas, se acordó actualizar el 30% de las unidades primarias muestrales ya enumeradas en el año 2012 (2.337 manzanas y 471 secciones).

Antes de proceder a la re-enumeración, el Ministerio y el INE evaluaron distintas formas para determinar cuántas y cuales unidades primarias debían ser sometidas a este procedimiento.

El objetivo principal de la reenumeración 2013, fue captar el crecimiento de la población de personas y de viviendas en el año transcurrido en aquellas áreas geográficas donde, de acuerdo a diversos antecedentes, se sabía que había existido cierto dinamismo de sus unidades (nuevas construcciones, demoliciones, erradicación de viviendas, etc.). Mientras que un objetivo secundario correspondía a actualizar el estado de aquellas Unidades Secundarias de Muestreo (USM) en áreas con mayor probabilidad de cambio de estado y/o uso de las viviendas (Viviendas de Temporada, Colectivas, desocupadas, etc.).

A continuación, se detallan los distintos criterios evaluados para determinar el total de Unidades Primarias de Muestreo (UPM) a enumerar y las comunas donde éstas se concentran.

3.2.2.1 Variables utilizadas

El marco muestral que posee el INE, es un marco de áreas geográficas, donde sus unidades muestrales son distintas de acuerdo al área de procedencia. Así, en el área urbana las unidades muestrales corresponden a delimitaciones geográficas bien definidas denominadas manzanas, mientras que en el área rural las unidades se denominan secciones y corresponden a unidades creadas a partir de la unión de un set de viviendas particulares según información proveniente del Censo 2002³¹.

Las manzanas del marco muestral urbano han sido actualizadas, en cuanto al total de viviendas que ellas contienen, con información proveniente del plan municipal de edificaciones, logrando captar de esta forma los nuevos desarrollos urbanos. Así las últimas actualizaciones del marco muestral fueron realizadas el año 2007 y posteriormente el año 2008.

³⁰Omitir el crecimiento de la población al interior de las unidades primarias seleccionadas incide directamente en la cobertura y las probabilidades de selección de las viviendas.

³¹Mayor antecedentes ver "Metodología del Diseño Muestral y Factores de Expansión Encuesta de Caracterización Socioeconómica Nacional 2011", disponible en la página web del Ministerio de Desarrollo Social.

http://observatorio.ministeriodesarrollosocial.gob.cl/layout/doc/casen/Informe%20Diseno%20Muestral_Revision_1_3sep12.pdf (a Diciembre 2014).



De acuerdo a lo anterior y siguiendo el objetivo principal se trabajó, en primera instancia, con un set de variables disponibles en el marco de muestreo, que dan cuenta del crecimiento respecto al total de viviendas pertenecientes a las UPM que se encuentran en el MM2008 (marco de muestreo actualizado el año 2008) y en el MM2007 (marco de muestreo actualizado el año 2007).

Respecto al crecimiento de las comunas, se analizaron dos variables:

- una que da cuenta de la variación del total de las USM, al interior de cada comuna, entre el año 2007 y 2008 (crecimiento o decrecimiento al interior de las UPM e incorporación de USM en nuevas UPM),
- y otra de la variación de USM que existió en cada manzana presente en ambos marcos muestrales (crecimiento o decrecimiento al interior de cada UPM).

Adicionalmente, se analizaron otras variables provenientes de fuentes externas al Marco Muestral:

- Información proveniente de la Superintendencia de Servicios Sanitarios (SISS), que consiste en el total de inmuebles y total de inmuebles conectados a la red de agua potable y servicios de alcantarillado y tratamiento de aguas servidas, por comuna para los años 2011 y 2012 (diciembre).
- Como información auxiliar y en atención al segundo objetivo, se utilizó la proporción de viviendas no elegibles según comuna, observado en el trabajo de campo de Casen 2011.

Previo a cualquier análisis se determinó revisar la variación del crecimiento de las comunas y junto con ello construir un indicador que diera cuenta del cambio o crecimiento de éstas.

3.2.2.2 Selección de comunas a reenumerar

Una vez creados los estratos se debieron determinar las comunas y/o las unidades primarias de muestreo que deberían ser reenumeradas, de acuerdo a los objetivos definidos.

Para ello se determinaron las siguientes medidas considerando los estratos y subestratos.

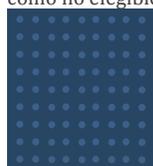
1. En el Estrato 1³² se estableció reenumerar todas las comunas clasificadas en él, ya que en Casen 2011 la principal causa de pérdida de unidades en algunas regiones se debió a las viviendas no elegibles. Tal fue el caso de la V Región, donde la tasa de no elegibilidad alcanzó 11,2% aproximadamente.
2. En el estrato 2³³, se asignó un total de comunas a seleccionar diferenciado según el subestrato de pertenencia. Así, acorde a la probabilidad de cambio se asignó una proporción de comunas a seleccionar, seleccionando más de aquellas comunas con mayor probabilidad de cambio y menos comunas donde se registró menor dinamismo entre los años 2011 y 2012, según información de la SISS.

A continuación, se detalla la fracción de muestreo al interior de cada subestrato:

- Subestrato 1: 70% de las comunas (30 de 43 comunas).
- Subestrato 2: 40% de las comunas (29 de 72 comunas).
- Subestrato 3: 15% de las comunas (26 de 171 comunas).

³² Es el estrato compuesto de aquellas comunas que según Casen 2011, poseen 10% o más viviendas de la muestra clasificadas como no elegibles.

³³ Es el estrato compuesto de aquellas comunas que según Casen 2011 poseen menos de un 10% de viviendas de la muestra clasificadas como no elegibles.



Al interior de cada estrato, las comunas fueron seleccionadas aleatoriamente y con igual probabilidad. Posteriormente, se estableció que todas las unidades primarias de las comunas seleccionadas deberían ser reenumeradas en las áreas urbanas y rurales.

Finalmente, en la Tabla II.21 se exponen el total de comunas seleccionadas³⁴ en cada estrato y subestrato, y el total de conglomerados y viviendas que ellas reúnen.

Tabla II.21: Distribución regional del total de comunas y unidades a reenumerar.

Región	Nº Comunas	Nº Manzanas	Nº Secciones	Nº Viviendas
XV de Arica y Parinacota	1	0	6	101
I Tarapacá	4	125	17	1.405
II de Antofagasta	3	227	3	2.225
III de Atacama	4	96	7	591
IV de Coquimbo	5	170	22	1.619
V de Valparaíso	19	274	63	4.056
VI de O'Higgins	8	178	37	2.433
VII del Maule	14	111	67	2.141
VIII del Biobío	23	486	102	6.415
IX de La Araucanía	9	80	48	1.400
X de Los Lagos	9	174	40	2.340
XIV de Los Ríos	4	28	25	639
XI de Aisén	0	0	0	0
XII de Magallanes y La Antártica	0	0	0	0
XIII Metropolitana	20	388	34	6.657
Total País	123	2.337	471	32.022

Nota: el número total de viviendas a reenumerar (32.022) corresponde al 30% la muestra sobredimensionada estimada en alrededor de 96.000 viviendas para Casen 2012, cuando se tenía como muestra objetivo un tamaño muestral de 75.080 viviendas. Posteriormente, se decidió reducir la muestra Casen 2013 a una muestra objetivo de 70.080 viviendas.

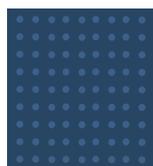
Fuente: Instituto Nacional de Estadísticas.

3.2.2.3 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 la fracción de muestreo que se aplicó al interior de cada manzana fue de 25% aproximadamente, con un mínimo de 2 viviendas a encuestar. Para no concentrar la muestra en manzanas con gran número de viviendas, se establece una cota superior de 31 viviendas a encuestar. Esto permite que manzanas de todos los tamaños estén representadas en la muestra en forma proporcional, a través de un número fijo de viviendas, según el grupo al que pertenezca. Este criterio de selección se desarrolló originalmente para la Encuesta Nacional de Empleo y el INE la aplica para la selección de

³⁴ En el Anexo E se encuentra el listado de las comunas que fueron re-enumeradas.



muestras urbanas de todas las encuestas de hogares. La Tabla V.1 en el Anexo presenta el total de viviendas a seleccionar en cada manzana según su grupo de tamaño.

En el marco de secciones la selección de viviendas se realizó al interior de cada sección, en forma sistemática, con igual probabilidad. Específicamente, el algoritmo de selección sistemático de viviendas, que es aplicado también en otras encuestas, se estandariza para que independientemente de quien aplique el mismo algoritmo, bajo las mismas condiciones de ordenamiento previo³⁵, seleccione las mismas viviendas. Este procedimiento es análogo a introducir una semilla de aleatorización para selecciones aleatorias, las cuales arrojan también las mismas selecciones.

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: cálculo del período.

Se calcula el período $K = M_i / m_i$

Paso 2: cálculo de la primera selección.

Se calcula el arranque A o primera selección como la mitad entre el número uno y el período K. Al sumarle 0,5 al período K, se evita que el arranque empiece con cero:
 $A = \text{Redondea}([K+0,5] / 2 ; 0)$

Paso 3: generación de las selecciones.

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, ... , A+($m_i - 1$)K.

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.

Podría pensarse que se resta aleatoriedad al partir la selección siempre con la mitad entre uno y el período, pero el arranque es tan impredecible debido a que se pre-seleccionan sólo las elegibles, generando que la primera vivienda elegible no corresponda con la primera vivienda de acuerdo al orden de la vivienda según recorrido de la enumeración.

³⁵ 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)



III. DESARROLLO DE FACTORES DE EXPANSION

1. INTRODUCCIÓN

Para que los análisis y estudios basados en encuestas a hogares con diseño muestral probabilístico tengan validez sobre toda la población objetivo, se debe utilizar un ponderador en la estimación de cualquier variable de interés, 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*)³⁶.

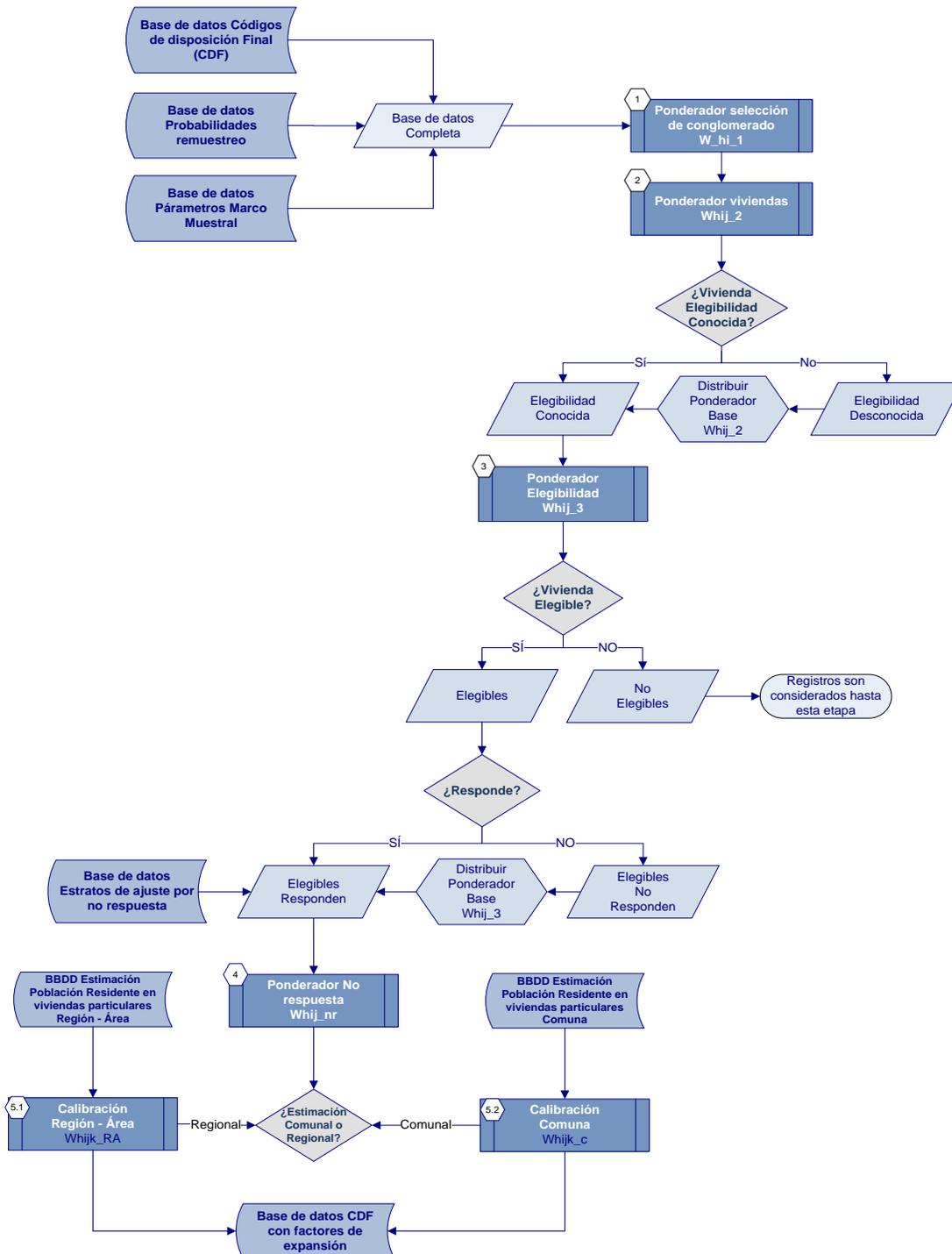
La metodología de cálculo de los factores de expansión consiste en la aplicación secuencial de cuatro ponderadores:

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 2013 fue la misma utilizada en 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 y Comunal buscan cuadrar las estimaciones poblacionales obtenidas a partir de la encuesta con las estimaciones poblacionales vigentes publicadas por el INE, 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 residentes en viviendas particulares estimado para cada comuna a partir de la encuesta y el total de personas en este grupo poblacional según las proyecciones demográficas del INE. Mientras que el ajuste "regional", se realiza 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 en cada región-área. La Figura III.1 resume el proceso del cálculo de los factores de expansión para Casen 2013. 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.

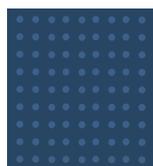
³⁶ La suma de los factores regionales es levemente mayor a la suma de los factores comunales ($\sum expr > \sum expc$), ya que por construcción el factor regional incluye la población en las comunas no incluidas en la muestra.



Figura III.1. Diagrama de procesos de desarrollo de factor de expansión Casen 2013



Fuente: Instituto Nacional de Estadísticas.



2. Ponderador de selección de conglomerados

Casen 2013, 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 definido por el área (urbano o rural) de una comuna y también de forma independiente en cada submuestra³⁷. Para la encuesta Casen 2013 se seleccionaron unidades muestrales en 589³⁸ 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 en el Marco de Secciones (MS2002)

Un elemento importante de recordar en este contexto, es que los conglomerados de la muestra Casen 2013 fueron seleccionados para la muestra de Casen 2012, y en este proceso, se excluyeron las secciones ya seleccionadas en la muestra de la Encuesta Nacional de Empleo, cuyo periodo de levantamiento coincidía con Casen 2012.

La Tabla III.1³⁹ presenta el total de secciones y viviendas que posee el marco de secciones y el total de unidades, secciones y viviendas excluidas del marco previo a la selección.

³⁷ En esa etapa de diseño aún se planeaba el levantamiento secuencial de dos submuestras independientes, con diseño idéntico y tamaño diferente.

³⁸ Se incorporaron los estratos de Putre rural, Atacama rural, La estrella rural, Porvenir rural.

³⁹ Existen diferencias entre el total de unidades en el MM2002 de las versiones de los años 2011 y 2013, pues en Casen 2013 se incorporaron nuevos estratos (las comunas no cambian, sin embargo, se incluyeron áreas de algunas comunas que no estaban presentes en Casen 2011, tales como Putre rural, Atacama rural, La estrella rural, Porvenir rural)



Tabla III.1: Número de unidades del MS2002 (secciones y viviendas) consideradas y excluidas de la selección de la muestra Casen 2013 por regiones.

Región	Unidades MM2002		Unidades MM2002 excluidas	
	Número de Secciones	Número de Viviendas	Número de Secciones	Número de Viviendas
XV de Arica y Parinacota	63	3.256	21	1.107
I Tarapacá	63	3.268	12	698
II de Antofagasta	46	2.347	1	50
III de Atacama	104	5.233	11	597
IV de Coquimbo	501	39.412	67	5.358
V de Valparaíso	538	38.768	34	2.508
VI de O'Higgins	893	67.856	63	4.517
VII del Maule	1.192	96.332	64	5.248
VIII del Biobío	1.293	100.123	39	3.015
IX de La Araucanía	1.096	89.333	45	3.725
X de Los Lagos	791	64.581	76	6.212
XIV de Los Ríos	395	31.685	29	2.356
XI de Aisén	129	6.951	12	678
XII de Magallanes y La Antártica	50	2.784	4	212
XIII Metropolitana	441	45.434	25	2.549
TOTAL	7.595	597.363	503	38.830

Fuente: Instituto Nacional de Estadísticas.

Tras acotar el Marco Muestral de secciones, las unidades muestrales (viviendas particulares) provenientes del MS2002 fueron seleccionadas en dos etapas.

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 una segunda etapa, se seleccionaron con igual probabilidad en cada sección, las viviendas particulares clasificadas como elegibles al momento de la enumeración de Casen .

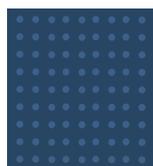
Es importante señalar que al momento de selección, se pensaba aún en un diseño muestral con dos submuestras, con mismo diseño, diferente tamaño y levantamiento en períodos secuenciales. En ese contexto, $P_h(i)$ la probabilidad de selección de la sección i del estrato h dependía de la submuestra para la cual había sido seleccionada, como se indica a continuación,

$$P_h(i) = \begin{cases} \frac{n_{1h}M_{hi}}{M_h} & \text{si } i \in \text{Submuestra 1} \\ \frac{n_{2h}M_{hi}}{M_h} & \text{si } i \in \text{Submuestra 2} \end{cases} \quad (1)$$

Donde:

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

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



n_{1h}	Es el número de secciones seleccionadas en el estrato h , para la submuestra 1.
n_{2h}	Es el número de secciones seleccionadas en el estrato h , para la submuestra 2.
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 2013 ⁴⁰ .

Si bien, las secciones fueron seleccionadas de forma independiente en cada una de las dos submuestras, por recomendación del Panel de Expertos de Casen, la muestra en su totalidad fue levantada en un mismo período⁴¹, y tratada como si se hubiera seleccionado toda a la vez. Por lo tanto, para fines de análisis y comparación se debe considerar solo una muestra o muestra completa. Lo anterior debe estar plasmado en las probabilidades de selección de las secciones.

Por lo tanto, esto implica que la probabilidad de que una sección pertenezca a la muestra completa está condicionada por haber sido seleccionada en la submuestra 1 o la submuestra 2.

La probabilidad $P_h(i)$ de seleccionar una sección i en una muestra de tamaño S_h en el estrato h para la muestra completa ha de ser calibrada de acuerdo a:

$$P_h(i) = \frac{n_h \cdot M_{hi}}{M_h} = \frac{n_{1h}M_{hi}}{M_h} + \frac{n_{2h}M_{hi}}{M_h} \quad (2)$$

Donde,

i	Es el índice de la unidad de muestreo o sección.
h	Es el índice del estrato de muestreo, o área (urbana o rural) de una comuna.
n_{1h}	Es el número de secciones seleccionadas en el estrato h , para la submuestra 1.
n_{2h}	Es el número de secciones seleccionadas en el estrato h , para la submuestra 2.
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 2013 ⁴² .

La Figura III.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 al 20%) y relativamente más alta para las regiones en los extremos con menor población.

La región con menor variabilidad es la región de Coquimbo, donde el 50% de sus secciones tienen una probabilidad superior o igual al 0,23 de ser seleccionadas. En la Región del Maule se concentran las secciones con menor probabilidad de selección. En cambio, en aquellas áreas de menor concentración de la población (principalmente rural), como las regiones I, II, XI, XII y XV, aproximadamente el 50% de sus secciones poseen probabilidad de selección superior a 0,4.

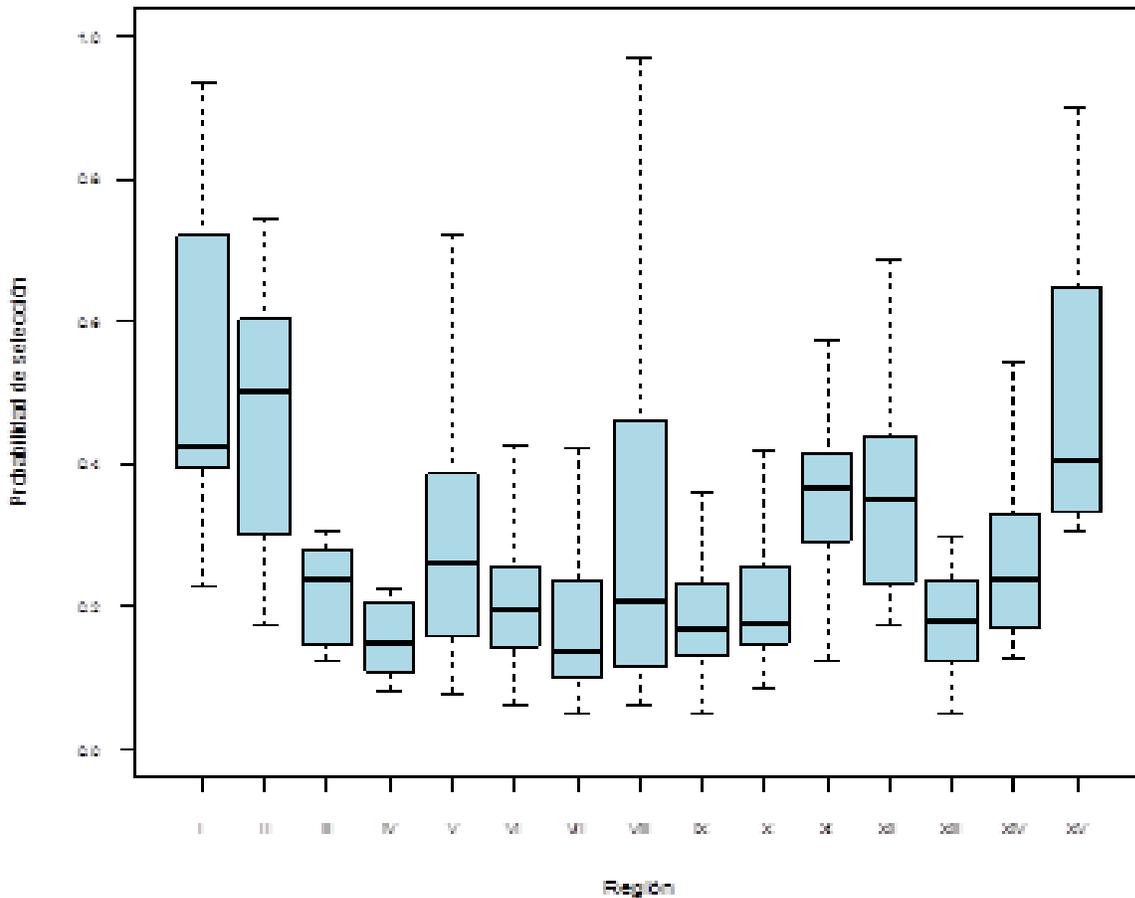
⁴⁰ Este total de viviendas excluye todas aquellas viviendas pertenecientes a secciones previamente seleccionados para la Encuesta Nacional de Empleo, cuyo periodo de levantamiento coincidía con el de Casen 2013.

⁴¹ Período comparable con las versiones anteriores de Casen y con la submuestra 2 de Casen 2011 (recolectada entre noviembre 2011 y diciembre y enero 2013).

⁴² 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 2013.



Figura III.2: Distribuciones regionales de las probabilidades de selección de las Secciones, por Región, Casen 2013.



Fuente: Instituto Nacional de Estadísticas.

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

La Tabla III.2 presenta, para cada región del país, el total de manzanas y viviendas en el marco de manzanas y el total de manzanas y viviendas excluidas del MM2008 para la selección de la muestra Casen 2013.



Tabla III.2: Número de unidades del MM2008 (manzanas y viviendas) consideradas y excluidas de la selección de la muestra Casen 2013 por regiones.

Región	Unidades MM2008		Unidades MM2008 excluidas	
	Número de Manzanas	Número de Viviendas	Número de Manzanas	Número de Viviendas
XV de Arica y Parinacota	2.415	52.865	484	4.956
I Tarapacá	2.198	70.162	298	4.184
II de Antofagasta	5.614	131.946	860	6.480
III de Atacama	4.000	74.686	866	5.738
IV de Coquimbo	6.817	152.956	1.064	7.953
V de Valparaíso	19.259	507.778	3.593	30.612
VI de O'Higgins	5.881	159.390	941	7.159
VII del Maule	6.703	185.055	793	7.196
VIII del Biobío	16.492	445.810	2.437	24.029
IX de La Araucanía	7.158	170.383	1.050	8.017
X de Los Lagos	5.221	136.763	771	6.793
XIV de Los Ríos	2.547	66.790	356	3.319
XI de Aisén	990	19.782	242	3.143
XII de Magallanes y La Antártica	2.038	42.374	262	2.337
XIII Metropolitana	46.027	1.784.022	3.538	42.662
TOTAL	133.360	4.000.762	17.555	164.578

Fuente: Instituto Nacional de Estadísticas.

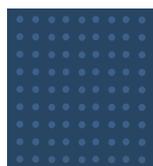
Las unidades muestrales (viviendas particulares) provenientes del MM2008 fueron seleccionadas en tres etapas.

Una vez establecido el marco de selección, se determinaron 30 "grupos de tamaño" (definidos en función del número de viviendas) en los cuales se ordenaron todas las manzanas, para enseguida proceder a la selección de la muestra de la encuesta Casen 2013.

En una primera etapa se seleccionaron grupos de tamaño, en los cuales en una segunda etapa se seleccionaron las manzanas, mediante un método iterativo y sin remplazo, con probabilidad proporcional a una medida de tamaño, definida según el total de manzanas y viviendas que contenía el grupo de tamaño seleccionado en primera etapa.

El método iterativo consistió en recalcular después de la selección de una manzana en un grupo de tamaño, las probabilidades de selección de los 30 grupos de tamaño n , descontando el total de viviendas de la manzana seleccionada, para volver a realizar la selección de un "grupo de tamaño" y así reiterar la selección de una manzana en su interior.

Cada vez que se seleccionó una manzana operó la tercera etapa de selección, en la cual las viviendas particulares fueron seleccionadas con igual probabilidad al interior de cada manzana. El proceso se realizó repetidamente hasta alcanzar el total predeterminado de viviendas en una comuna. Al tratarse de un método iterativo, la probabilidad de selección de los conglomerados no tiene asociada una fórmula lineal explícita, siendo estimada a través de un método numérico.



Paso 1: Selección del grupo de tamaño g

$$P(g) = \frac{N_{hg} \cdot M_{hg}}{\sum_{k=1}^{30} N_{hg} \cdot M_{hk}} \quad (3)$$

Donde,

M_{hg} Es el número de viviendas ubicadas en las manzanas contenidas en el grupo de tamaño g del área (urbana o rural) de la comuna h .

N_{hg} Es el número de manzanas que posee el grupo de tamaño g , en el área (urbana o rural) de la comuna h .

M_{hk} Es el número de viviendas ubicadas en las manzanas contenidas en el grupo de tamaño k del área (urbana o rural) de la comuna h . (k puede variar de 1 a 30, ya que hay 30 grupos de tamaño).

Luego, se recalculan las probabilidades de selección del grupo de tamaño, descontando el número de viviendas en la manzana seleccionada, así como también reduciendo el número de manzanas que contiene el grupo del cual fue extraída.

Probabilidad de selección de la primera manzana en el grupo de tamaño g :

La probabilidad $P_h(i \cap g)$ de seleccionar una primera manzana i desde un grupo de tamaño g seleccionado, en el área (urbana o rural) de la comuna h , está dada por:

$$P_h(i \cap g) = P(i | g) \cdot P(g) \quad (4)$$

Donde,

$P(i | g)$ Probabilidad de seleccionar la manzana i en el área (urbana o rural) de la comuna h , dado que fue seleccionado el grupo de tamaño g .

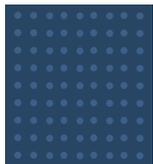
$P(g)$ Probabilidad de seleccionar el grupo de tamaño g (paso 1).

Entonces, la probabilidad de seleccionar la manzana i del área (urbana o rural) de la comuna h en el grupo de tamaño g , dado que resultó seleccionado el grupo de tamaño g , se obtiene como:

$$P(i | g) = \frac{1}{N_{hg}} \quad (5)$$

Donde,

N_{hg} Es el número de manzanas del grupo de tamaño g , en el área (urbana o rural) de la comuna h .



Paso 2: Selección de la segunda manzana.

Al seleccionar la segunda manzana se repite el paso anterior, actualizando N_{hg} y M_{hg} , y descontando en el grupo de tamaño g la selección de la primera manzana.

Paso 3 y siguientes.

El proceso se itera hasta acumular el número de viviendas establecidas al diseñar y determinar el tamaño muestral.

2.3.Desarrollo de un estudio de probabilidades de selección empíricas.

La probabilidad de selección $P(i \in S_n)$ de una manzana i en la muestra de manzanas seleccionadas S_n es igual a la suma de todas las probabilidades de selección determinadas por el proceso iterativo descrito por los pasos anteriores, como se refleja en la fórmula:

$$P(i \in S_n) = \sum_{k=1}^n P(i \text{ seleccionada en } k^{\circ} \text{ extracción}) \quad (6)$$

Formular la probabilidad de selección de manzanas resulta complejo mediante una fórmula lineal, ya que depende de la probabilidad de seleccionar un grupo de tamaño g y una manzana en cada paso, actualizando iterativamente el conjunto de unidades (grupos de tamaño g y manzanas) desde los cuales se realiza la selección.

Ante esto, para la encuesta Casen 2011, el INE desarrolló un estudio de simulaciones que sentó el precedente de la pertinencia de aproximar la estimación de las probabilidades de selección de las manzanas utilizando las fórmulas del muestreo trietápico proporcional al tamaño, mediante un estudio de simulaciones.

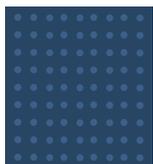
Para la muestra de la encuesta Casen 2013, aprovechando las bases instaladas en la ronda anterior para estos efectos, se realizaron 2.000 simulaciones de selección de manzanas, en cada una de las 223 comunas pertenecientes al MM2008, generando así para cada comuna una distribución de probabilidad de selección empírica.

En los 82 estratos o áreas geográficas (urbana o rural) de comunas, en las cuales las 2.000 simulaciones no mostraban resultados concluyentes para producir estimaciones confiables acerca de la distribución de probabilidad de selección de las manzanas, se extendieron los procesos hasta alcanzar 6.000 simulaciones de selección de manzanas, generando también, para estos estratos, distribuciones probabilidad de selección empírica.

El proceso de simulación consideró el remuestreo de manzanas asociadas a una muestra de 32.616 viviendas.

El INE estimó la probabilidad de selección de las manzanas mediante dos métodos distintos:

- Método basado en una aproximación lineal fundamentada en las probabilidades empíricas obtenidas de las simulaciones;
- Método basado en la "probabilidad de remuestreo" o proporción de muestras en las que fue seleccionada una manzana (Valliant *et al.*, 2008; Valliant & Rust, 2010); Sitter, 1992).



Siguiendo las recomendaciones del Profesor R. Valliant, experto consultado en la encuesta Casen 2011, se guardaron las probabilidades de selección condicionales de la muestra y, por otra parte, se conservó el estudio de simulaciones realizado en el contexto de Casen 2011.

Al momento de seleccionar la muestra de manzanas de Casen 2013, se procedió a estimar las probabilidades de selección según el método basado en la "probabilidad de remuestreo", al igual que para Casen 2011.

Este método consistió en calcular, en cada estrato (área geográfica urbana o rural de una comuna) perteneciente al MM2008, el número de muestras en las que cada manzana fue seleccionada, respecto a la totalidad de las muestras simuladas.

Así, la probabilidad de selección de la manzana i del estrato h , estimada mediante técnicas de remuestreo, es igual a la número de veces que resultó seleccionada en el total de simulaciones de muestreo procesadas:

$$\pi_{hi} = \begin{cases} \frac{n_{hi}}{2000} & , \text{ Si manzana } i \in \text{ comuna } h, \text{ para la que se realizaron 2000 simulaciones} \\ \frac{n_{hi}}{6000} & , \text{ Si manzana } i \in \text{ comuna } h, \text{ para la que se realizaron 6000 simulaciones} \end{cases}$$

Donde,

n_{hi} Es el número de muestras en las simulaciones en que la manzana i del estrato h fue seleccionada.

π_{hi} Es la probabilidad de de selección estimada mediante técnicas de remuestreo, de la manzana i del estrato h , en el marco de simulaciones que consideraron un tamaño muestral nacional de 32.616 viviendas.

En Casen 2013, las simulaciones se generaron considerando dos submuestras cuyos tamaños respectivos fueron diferentes al tamaño con el cual se generaron las simulaciones ($n=32.616$). Por ello, las probabilidades de selección estimadas a partir del método de remuestreo (π_{hi}) deben ser calibradas para expandir en forma consistente con los tamaños muestrales de cada submuestra.

De esta forma, la probabilidad de selección $P_h(i)$ de que la manzana i del estrato h pertenezca a la muestra de Casen 2013 ha de ser estimada como:

$$P_h(i) = (f_1 + f_2) \cdot \pi_{hi} \quad (8)$$

Donde,

f_1 Es la razón entre el total de viviendas asignadas a la submuestra 1 en el área urbana (24.230) y el tamaño de la muestra con el cual se realizaron las simulaciones (32.616).

f_2 Es la razón entre el total de viviendas asignadas a la submuestra 2 en el área urbana (48.433) y el tamaño de la muestra con el cual se realizaron las simulaciones (32.616).

El 75% de las manzanas tienen una probabilidad estimada de selección inferior o igual al 12%. La Figura III.3A, muestra la asimetría de esta distribución.

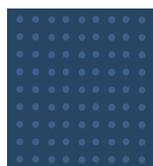
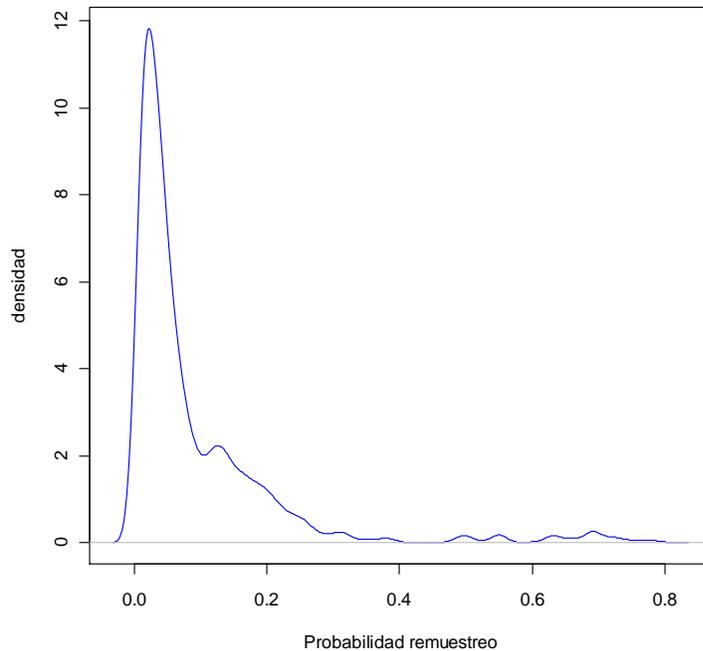


Figura III. 3A: Densidad de la probabilidad de selección de las manzanas mediante método de remuestreo - Casen 2013.



Fuente: Instituto Nacional de Estadísticas.

La Figura III.3B compara las distribuciones regionales de probabilidad de selección de manzanas. Globalmente, la mediana de las probabilidades de selección es baja para las regiones en el centro del país (en torno a 0.05) y relativamente más alta para las regiones extremas del norte y del sur, las que tienen menos población urbana. Se puede apreciar que las probabilidades estimadas son relativamente homogéneas al interior de cada región (cajas pequeñas).

En la Región de Aysén, sin embargo, se observan mayor heterogeneidad y un mayor nivel mediano de probabilidad de selección (igual a 0.5). Esto último se puede explicar por la gran cantidad de manzanas seleccionadas en la muestra en comparación al total de unidades existentes en los estratos de la región.

En la Región Metropolitana, en cambio, se observa una gran homogeneidad y una probabilidad de selección mediana reducida, lo cual puede ser explicado por la alta concentración de manzanas existentes en los estratos de muestreo de esta región.

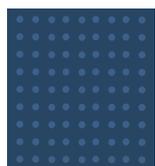
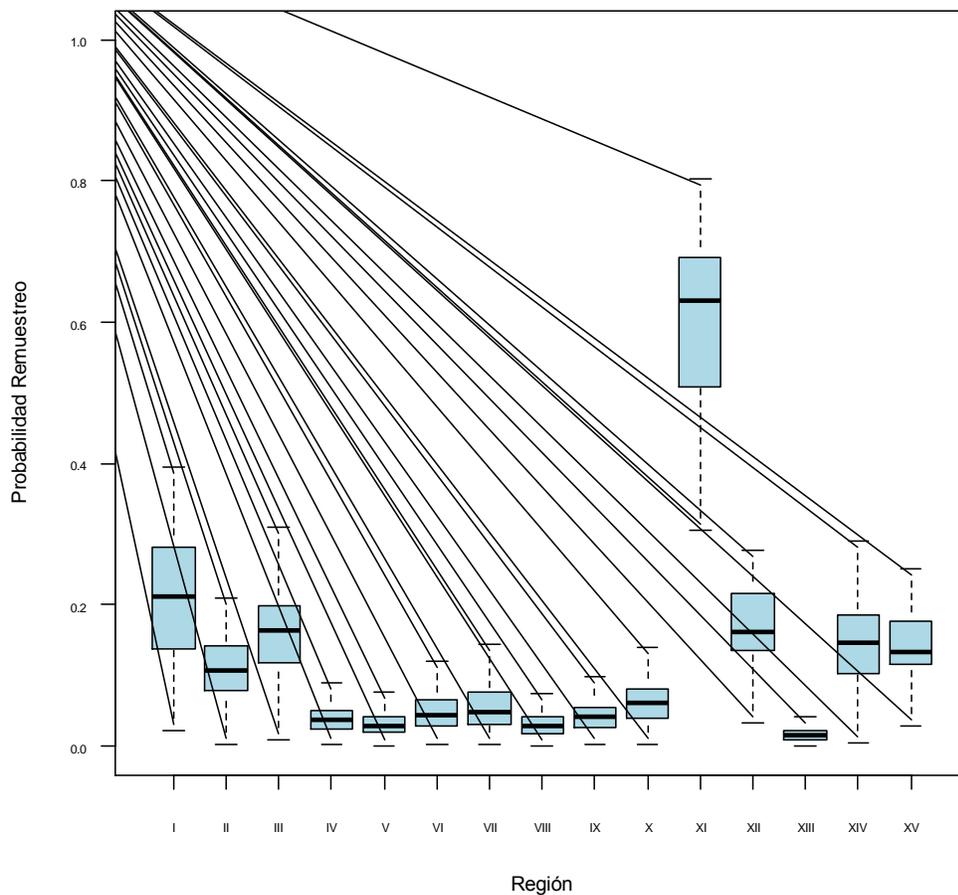


Figura III. 3B: Distribuciones regionales de las probabilidades de selección de las manzanas mediante el método de remuestreo - Casen 2013.



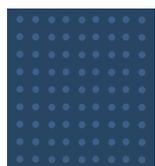
Fuente: Instituto Nacional de Estadísticas.

2.4. Determinación del Ponderador de Selección de Conglomerados como el inverso de las probabilidades 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:

$$w_{hi} = \frac{1}{P_h(i)} = \begin{cases} \frac{M_h}{S_h \cdot M_{hi}} & , \text{ Si unidad } i \text{ es una sección} \\ \frac{1}{(f_1 + f_2) \cdot \pi_{hi}} & , \text{ Si unidad } i \text{ es una manzana} \end{cases} \quad (9)$$

Donde,



h	Es el índice del estrato de muestreo, o área (urbana o rural) de una comuna.
w_{hi}	Es el ponderador de selección del conglomerado i del estrato h .
$P_h(i)$	Es la probabilidad de selección del conglomerado i del estrato h .
$\frac{S_h \cdot M_{hi}}{M_h}$	Es la probabilidad de selección de la sección i del estrato h (de tamaño S_h).
$(f_1 + f_2) \cdot \pi_{hi}$	Es la probabilidad de selección de la manzana i 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 125.818 conglomerados en el total del país (secciones y manzanas).

La Tabla III.5 muestra la distribución regional de manzanas y secciones en cada Marco Muestral y las correspondientes estimaciones muestrales 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. Esto se explica porque previa a la selección de la muestra Casen 2013, 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 2012 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.

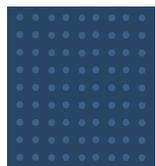


Tabla III.5: Distribución de frecuencias regionales de manzanas y secciones en el Marco Muestral y estimadas en Casen 2013.

Región	Marco Muestral		Estimadas en Casen 2013	
	Nº de Manzanas	Nº de Secciones	Nº de Manzanas	Nº de Secciones
Arica y Parinacota	2.415	63	1.289	44
Tarapacá	2.198	63	797	51
Antofagasta	5.614	46	1.714	45
Atacama	4.000	104	1.217	96
Coquimbo	6.817	501	5.769	435
Valparaíso	19.259	538	15.349	501
Libertador General Bernardo O'Higgins	5.881	893	4.929	836
Maule	6.703	1.192	4.258	1.149
Biobío	16.492	1.293	18.270	1.258
Araucanía	7.158	1.096	6.687	1.053
Los Lagos	5.221	791	2.968	710
Los Ríos	2.547	395	1.182	366
Aysén del General Carlos Ibáñez del Campo	990	129	203	116
Magallanes y La Antártica Chilena	2.038	50	1.059	45
Metropolitana de Santiago	46.027	441	53.007	417
Total País	133.360	7.595	118.697	7.121

Fuente: Instituto Nacional de Estadísticas

Las Figuras III.4a y III.4b 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 secciones es inferior a la de las manzanas y la variabilidad de los ponderadores de selección de las manzanas es mayor, lo que se explica por la variabilidad de los tamaños de las manzanas (en número de viviendas que poseen oscila entre 8 y más de 1.000 unidades).

Para las secciones, la distribución regional de los ponderadores de selección es más homogénea debido principalmente a que los tamaños de secciones son más homogéneos al interior de cada región. Sin embargo, en las regiones donde la población rural es pequeña, la mediana del ponderador de selección de los conglomerados es inferior a 8 viviendas que en comparación a las manzanas, cuyas medianas regionales toman valores comprendidos entre 0 y 25 viviendas. La región de La Araucanía es la región que presenta el valor máximo observado del ponderador, en donde una sección representa a 20 viviendas.

Para las manzanas, la Región Metropolitana presenta el ponderador de selección con mayor variabilidad, pues sus manzanas tienen tamaños que varían entre 1 y 772 viviendas, con una mediana de 25. Por otra parte, las manzanas de la Región de Aysén (XI región), Magallanes (XII), Arica y Parinacota (XV), Tarapacá (I), Antofagasta (II) y Atacama (III) presentan valores medianos bajo 5 viviendas y la vez las variabilidad de tamaño de las manzanas al interior de ésta regiones son las más homogéneas del país.

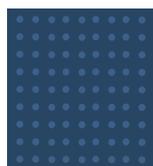
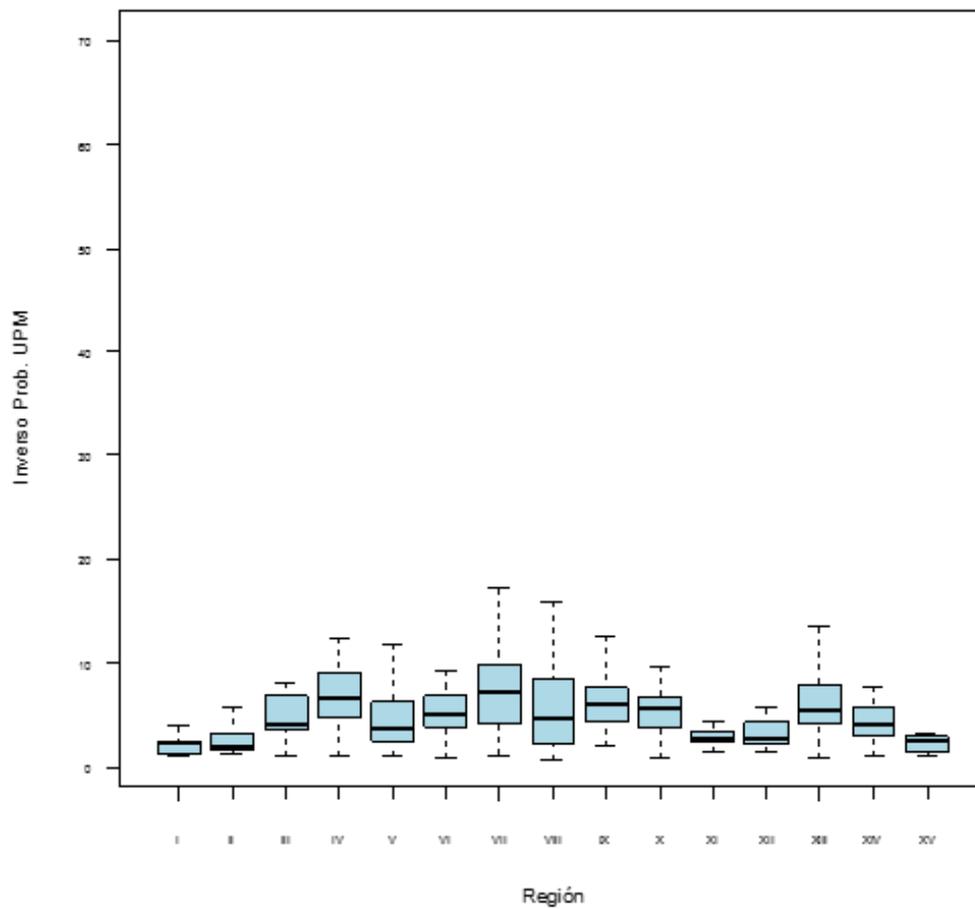


Figura III.4: Distribuciones regionales del Ponderador de Selección de Secciones en Casen 2013.



Fuente: Instituto Nacional de Estadísticas.

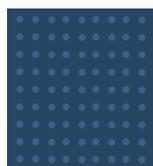
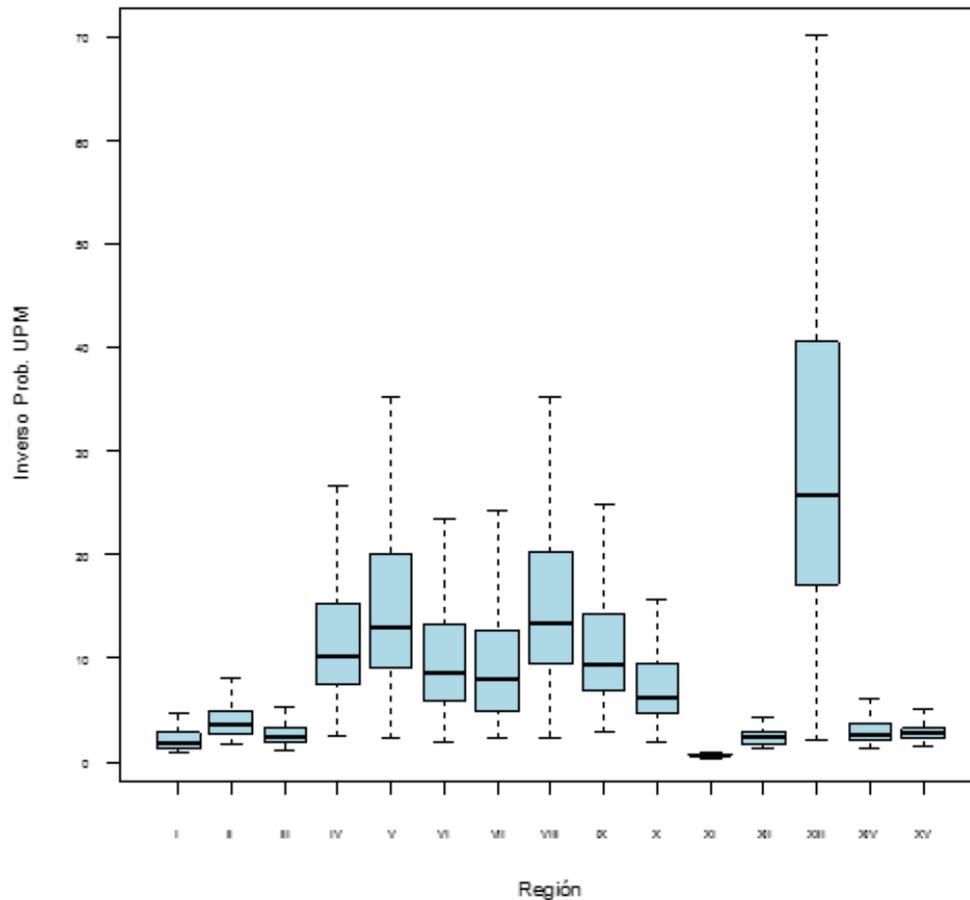


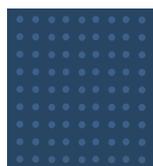
Figura III.5: Distribuciones regionales del Ponderador de Selección de Manzanas en Casen 2013.



Fuente: Instituto Nacional de Estadísticas.

3. 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 2013 por motivos operativos.



Como anteriormente, previa selección de la muestra de Casen 2013, 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 encuestas⁴³).

Análogamente, se excluyeron del marco muestral de secciones (MS2002) las secciones previamente seleccionadas para la muestra de la Encuesta Nacional de Empleo⁴⁴.

El número de manzanas y secciones descartadas *a priori* de ambos marcos muestrales equivale a un total de 203.000 viviendas aproximadamente, lo que corresponde a un 4,4% de la población de viviendas pertenecientes a los estratos Casen⁴⁵.

Tabla III.6: 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*	3.836.184	558.533	4.394.717
Marco sin Exclusión de unidades*	4.000.762	597.363	4.598.125

*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 estimadas resultó consistente con el número de viviendas de la población (4.598.125 a nivel nacional).

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 el número total de viviendas estimadas con el ponderador de selección del estrato h ,

$$\hat{R}_h = \frac{M_h}{\sum_{i \in \Omega_h} w_{hi} \cdot M_{hi}} \quad (10)$$

Donde,

M_h Es el número de viviendas en el estrato h .

Ω_h Es el conjunto de conglomerados seleccionados en el estrato h .

M_{hi} Es el número de viviendas del conglomerado i del estrato h .

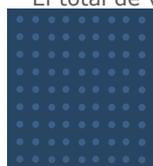
w_{hi} Es el ponderador de selección del conglomerado i del estrato h .

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

⁴³ Note que si bien el trabajo de campo de Casen 2013 se realizó entre Noviembre – Diciembre de 2013 y Enero de 2014, la omisión de unidades hace referencia a muestras seleccionadas para encuestas levantadas el 2012, ya que la selección se realizó para Casen 2012, versión que finalmente no se realizó y por lo tanto estas unidades fueron utilizadas para Casen 2013.

⁴⁴ Idem nota anterior.

⁴⁵ El total de viviendas en el marco muestral de Casen 2013 es de 4.594.668 viviendas.



$$w'_{hi} = \hat{R}_h \cdot w_{hi} \quad (11)$$

4. 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.

4.1. 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, tuvieran 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

$$P_{hi}(j|i) = \frac{m_{hi}}{M'_{hi}} \quad (12)$$

Donde,

j Es el índice de la vivienda.

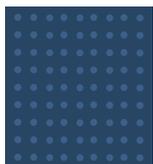
M'_{hi} Es el número de viviendas ocupadas en la manzana o sección i , del estrato h , según enumeración previa a la recolección de datos.

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

4.2. Corrección de probabilidades condicionales por reducción de muestra

La probabilidad condicional de selección de viviendas antes descrita está asociada a la muestra total de viviendas seleccionadas (94.321), sin embargo, para Casen 2013 se realizó una reducción del tamaño muestral.

Por ello, la probabilidad de selección de una vivienda depende de la submuestra en la cual fue seleccionada, al no haber sido descartada en una segunda instancia cuando se redujo la muestra de 94.321 a 87.402 viviendas, como consecuencia de la disminución del tamaño muestral objetivo en 5.000 viviendas por motivos operativos.



El siguiente apartado presenta la probabilidad de seleccionar una vivienda en Casen 2013 y que no haya sido descartada de la muestra.

Esta nueva selección consistió en elegir de entre las 94.321 viviendas, aquellas unidades que no serían enviadas al trabajo de campo a causa de la reducción de 6.919 viviendas de la muestra con sobremuestreo. La selección de dichas unidades se realizó a través de un procedimiento de selección sistemática y con igual probabilidad, al interior de cada estrato.

La probabilidad condicional $P'_{hi}(j|i)$ de que una vivienda j ubicada en conglomerado i del estrato h , conglomerado del que no haya sido descartada a causa de la reducción de la muestra puede ser calculada como,

$$P'_{hi}(j|i) = \frac{m_{hi}}{M'_{hi}} \cdot (1 - P_h(j \in \phi_h)) = \frac{m_{hi}}{M'_{hi}} \cdot \left(1 - \frac{m_{h,drop}}{m_h}\right) \quad (13)$$

Donde,

- j Es el índice de la vivienda.
- ϕ_h Es el conjunto de viviendas en el estrato h que se descartan de la muestra final.
- $P_h(j \in \phi_h)$ Es la probabilidad de que la vivienda seleccionada originalmente en el estrato h sea descartada de la muestra final.
- m_{hi} Es el número de viviendas seleccionadas en la manzana o sección i , del estrato h .
- 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.
- $m_{h,drop}$ Es el número de viviendas seleccionadas en el estrato h y que se descartan de la muestra final.
- m_h Es el número de viviendas seleccionadas en el estrato h (selección inicial)

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 (final) de Casen 2013:

$$P_{hi}(j) = P'_{hi}(j|i) \cdot P_h(i) \quad (14)$$

Donde,

- $P'_{hi}(j|i)$ Es la probabilidad condicional de que la vivienda j pertenezca a la muestra, producto de la reducción de la muestra objetivo en 5.000 unidades.
- $P_h(i)$ Es la probabilidad de que el conglomerado i (manzana o sección) del estrato h pertenezca a la muestra.

4.3. Ponderador de selección de viviendas

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 .



Luego de reordenar la expresión (13), el ponderador de selección de las viviendas puede ser expresado como,

$$w_{hij} = w'_{hi} \cdot \frac{1}{P_{hi}(j)} = w'_{hi} \cdot \frac{M'_{hi}}{m_{hi}} \cdot \frac{m_h}{(m_h - m_{h,drop})} \quad (15)$$

Donde,

- j Es el índice de la vivienda.
- 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 de conglomerado i del estrato h , corregido por omisión de conglomerados (descartados a priori del 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 2013 (formula (14)).
- 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 .
- m_h Es el número de viviendas seleccionadas en el estrato h .
- $m_{h,drop}$ Es el número de viviendas seleccionadas en el estrato h y que se descartan de la muestra final.

Este ponderador puede ser interpretado como el número de viviendas en la población que representan las viviendas seleccionadas en la muestra Casen 2013 en la población. Al interior de cada conglomerado, la ponderación es la misma para todas las viviendas.

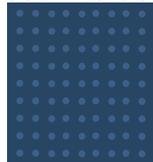
5. Ponderación por elegibilidad

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 marco⁴⁶.

46 El proceso de enumeración consiste en el registro de direcciones asociadas a las manzanas y secciones seleccionadas en la muestra. Este proceso lo realizó el INE entre los meses de Mayo y Agosto de 2012 y luego para un 30% de la muestra se reenumeró durante Junio y Julio de 2013. El trabajo se realizó en terreno por personal del INE entre 3 y 6 meses antes del trabajo de campo.



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.

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.

5.1. 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 reenumeración en el caso de Casen 2013).

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

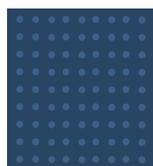
Las edificaciones consideradas, en principio, como viviendas con Elegibilidad Conocida (elegibles y no elegibles), inicialmente seleccionadas, pero que finalmente terminan siendo clasificadas como edificaciones de "Elegibilidad Desconocida", son redistribuidas al interior de cada estrato, junto con las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

La proporción de los pesos a redistribuir $\hat{R}_{h, know}$ 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,

$$\hat{R}_{h, know} = \frac{\sum_{i \in \Omega_h} \sum_{j \in \Theta_i} w_{hij}}{\sum_{i \in \Omega_h} \sum_{j \in \Theta_{i, know}} w_{hij}} \quad (16)$$

Donde,

Θ_i Es el conjunto de viviendas j seleccionadas en la manzana o sección i .



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

Ω_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

$$w'_{hij} = \hat{R}_{h, know} \cdot w_{hij} \quad (17)$$

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

5.2. Ajuste por no elegibilidad

Originalmente, la muestra definitiva Casen 2013 contemplaba 87.402 viviendas con sobremuestreo. De ese total 1.155 resultaron de "elegibilidad desconocida" una vez finalizado el trabajo de campo. Entre las 86.247 viviendas con elegibilidad conocida, hubo 83.659 que correspondían a viviendas "elegibles" y 2.588 que correspondieron a viviendas "no elegibles" (ver Tabla III.8).

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".

A partir de este ajuste, la suma del ponderador corresponde a la estimación del total de viviendas elegibles (viviendas particulares ocupadas), la que expandida es de 4.739.872 unidades.

Entre las 83.659 viviendas elegibles, hubo 64.842 viviendas para las cuales se logró concretar una o más entrevistas⁴⁷ y 18.817 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 2013.

⁴⁷ 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. Así en Casen 2013 se realizaron 66.725 entrevistas a hogares en estas 64.842 viviendas.

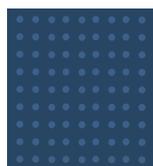


Tabla III.8: Distribución de viviendas seleccionadas por el INE en Casen 2013 por elegibilidad en el trabajo de campo

Vivienda Elegible	Frecuencia	Porcentaje
No elegible	2.588	3,0
Elegible	83.659	95,7
Elegibilidad desconocida	1.155	1,3
Total	87.402	100,0

Fuente: Instituto Nacional de Estadísticas.

6. 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 2013 de las 83.659 viviendas elegibles seleccionadas en la muestra, se lograron concretar 64.842 entrevistas. Es decir, la tasa de respuesta de la encuesta fue de 77,5%⁴⁸.

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.

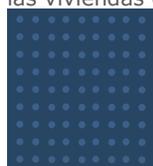
Si no hay relación entre el "patrón de no respuesta" y la variable de interés, entonces es posible que la no respuesta sólo afecte la precisión (varianza) de las estimaciones de interés.

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).

Para minimizar los problemas ocasionados por la ausencia de respuesta a la unidad en la Encuesta Casen 2013, se realiza un ajuste para compensar la pérdida de las unidades, denominado "propensity score".

⁴⁸ 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 76,5%



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 (sección o manzana).

El ajuste consistió, al igual que en Casen 2011, 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,

$$\hat{R}_{g,R} = \frac{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i,eleg}} w'_{hij}}{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i,eleg,R}} w'_{hij}} \quad (18)$$

Donde,

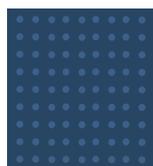
g Es el índice del grupo para la corrección de no respuesta.

Φ_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 III.9 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 III.9 Ponderador de no respuesta por celda de ajuste.

Celda de ajuste de no respuesta O Veintiles	Componentes del ponderador de corrección por no respuesta				$\hat{R}_{g,R}$
	ponderador de selección de viviendas	Viviendas Seleccionadas o elegibles	Viviendas Responden		
1	4.119	6.136	2.017		2,73
2	2.110	4.346	2.236		1,81
3	1.770	4.537	2.767		1,60
4	1.536	4.425	2.889		1,46
5	1.356	4.600	3.244		1,39
6	1.487	5.130	3.643		1,31
7	1.294	4.902	3.608		1,25
8	999	4.130	3.131		1,34
9	1.199	5.004	3.805		1,28
10	871	4.118	3.247		1,25
11	960	4.560	3.600		1,19
12	621	3.725	3.104		1,16
13	551	3.329	2.778		1,19
14	639	4.228	3.589		1,14
15	559	3.487	2.928		1,13
16	566	4.048	3.482		1,12
17	503	3.674	3.171		1,09
18	526	4.323	3.797		1,07
19	445	4.322	3.877		1,08
20	449	4.378	3.929		1,06

Fuente: Instituto Nacional de Estadísticas.

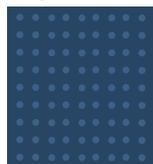
Una vez estimados los factores de ajuste, el ponderador de selección corregido por no respuesta w_{hij}^{NR} puede ser expresado como,

$$\widehat{w_{hij}^{NR}} = \hat{R}_{g,R} \cdot w'_{hij} \quad (19)$$

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

7. 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 al 30 de noviembre 2013 para todas las comunas de la muestra y además para las 30 divisiones dadas por el cruce de región y área (urbano, rural). 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.

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 2013 era de 17.633.833 personas⁴⁹. Según el Censo 2002, la proporción de personas residentes en viviendas particulares era de 97% aproximadamente, por lo cual se estima que el total de la población que reside en viviendas particulares es 17.272.759 personas⁵⁰.

⁴⁹ 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.

⁵⁰ 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.



Tabla III.10: Proyecciones de población Total

Región	Proyecciones de población Total			Proyecciones de población residentes en viviendas particulares		
	Urbano	Rural	Total	Urbano	Rural	Total
XV de Arica y Parinacota	159.477	19.254	178.731	156.105	15.403	171.508
I Tarapacá	315.379	24.320	339.699	300.903	14.851	315.754
II de Antofagasta	582.285	15.489	597.774	554.351	8.985	563.336
III de Atacama	260.376	27.282	287.658	255.562	22.154	277.716
IV de Coquimbo	612.015	142.462	754.477	600.454	138.210	738.664
V de Valparaíso	1.668.394	154.832	1.823.226	1.639.664	151.076	1.790.740
VI de O'Higgins	650.872	261.853	912.725	641.041	258.018	899.059
VII del Maule	698.071	337.515	1.035.586	684.968	333.355	1.018.323
VIII del Biobío	1.745.892	334.471	2.080.363	1.716.148	328.311	2.044.459
IX de La Araucanía	678.356	320.006	998.362	656.008	313.250	969.258
X de Los Lagos	620.117	252.359	872.476	599.358	245.216	844.574
XIV de Los Ríos	263.065	120.163	383.228	252.288	115.873	368.161
XI de Aisén	92.586	15.844	108.430	87.893	13.542	101.435
XII de Magallanes y La Antártica	149.346	11.079	160.425	143.446	6.350	149.796
XIII Metropolitana	6.859.306	241.367	7.100.673	6.784.821	235.155	7.019.976
Total País	15.355.537	2.278.296	17.633.833	15.073.010	2.199.749	17.272.759

Fuente: Instituto Nacional de Estadísticas.

A continuación se describen los procedimientos para la última etapa de desarrollo de los factores de expansión de Casen 2013. Primero se describe el desarrollo del Ponderador de Calibración Regional y luego el Ponderador de Calibración Comunal.

7.1. 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 2013). 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 2013. 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 2013 (paso 2).

$$\hat{R}_{RA} = \frac{N_{RA}}{\sum_{h \in RA} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,elg,R}} N_{hij} \cdot w_{hij}^{NR}} \quad R = 1, \dots, 15; \quad A = \text{Urbano, Rural} \quad (20)$$

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 .

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:

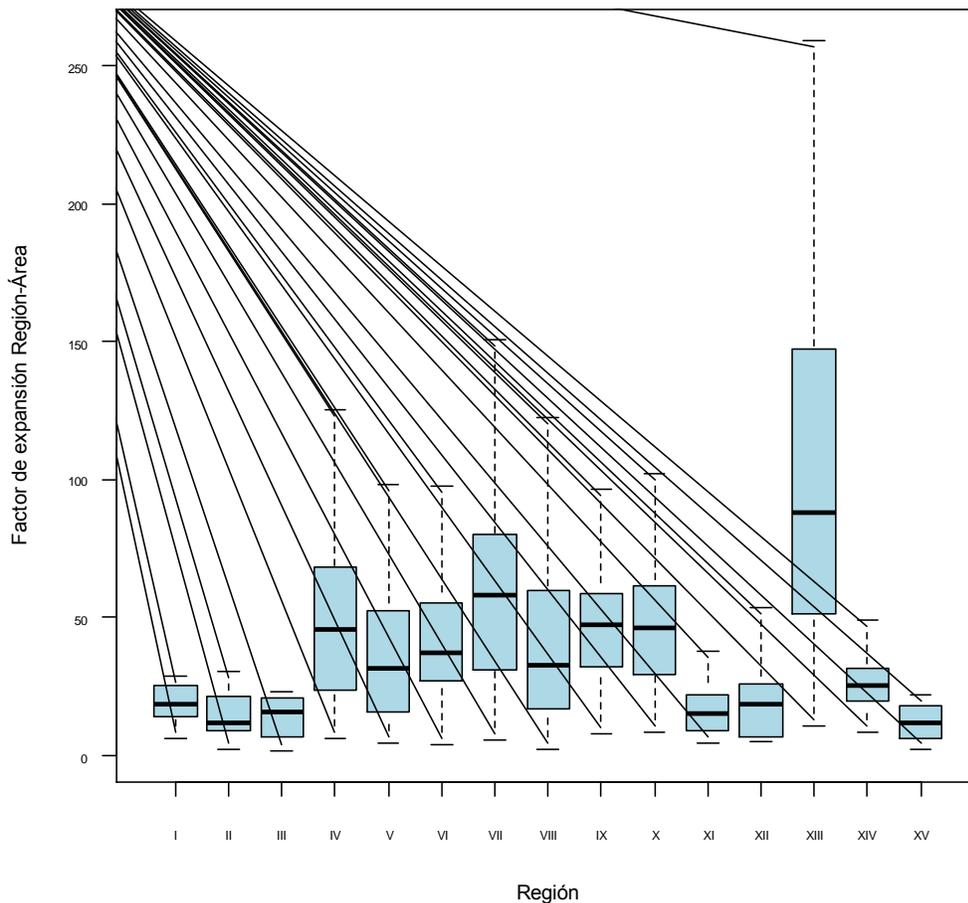
$$w_{hijk}^{RA} = \hat{R}_{RA} \cdot w_{hij}^{NR} \quad (21)$$

Las Figuras III.14 y III.15 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 (tanto para el marco urbano como rural).

Además, se observa que las unidades provenientes del MM2008 pertenecientes a las regiones Metropolitana, V y VIII son las que poseen mayor variabilidad, lo que se explica, en parte, porque en ellas 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 III.14: Factor de expansión calibración Región-área MS2002



Fuente: Instituto Nacional de Estadísticas.

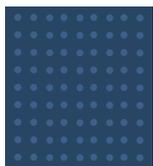
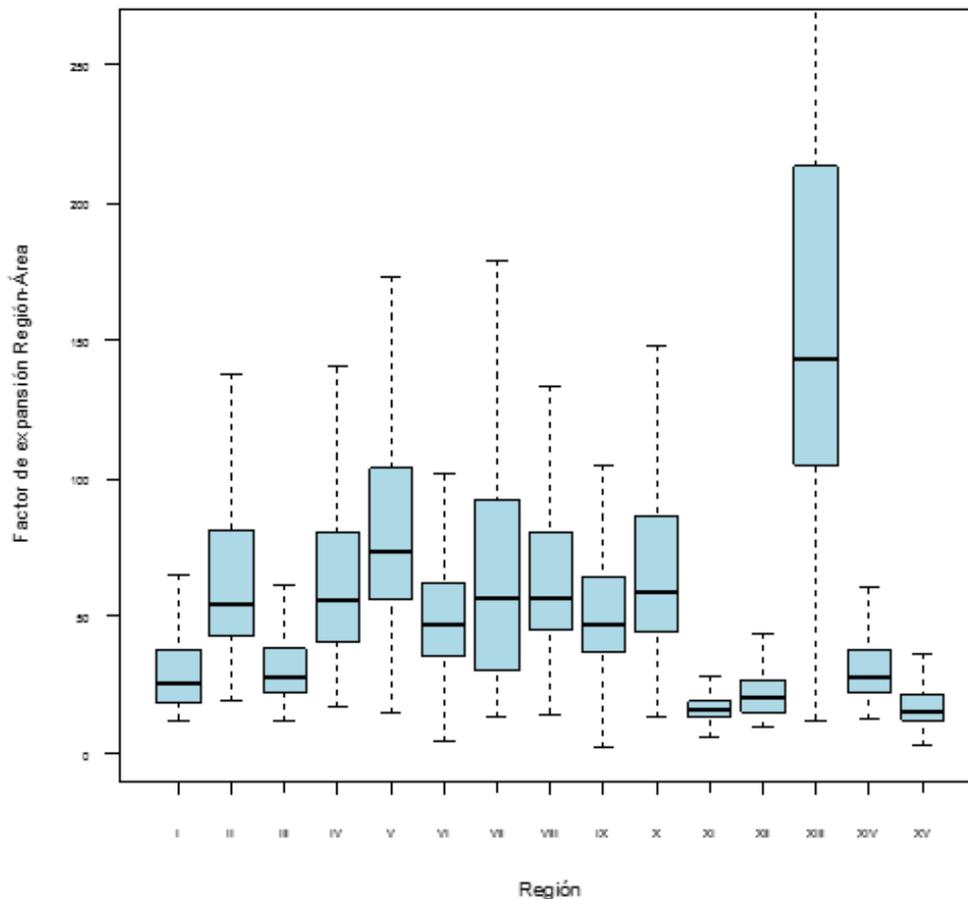


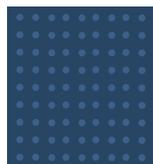
Figura III.15: Factor de expansión calibración Región-área MM2008

Fuente: Instituto Nacional de Estadísticas.

7.2. Ponderador de Calibración Comunal

El Ponderador de Calibración Comunal, también conocido como "Factor de Expansión Comunal" (*expc*), también se asocia a cada persona en la muestra. El procedimiento 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 589 estratos formados por la intersección comuna área en la muestra Casen 2013. 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 2013.



2. Estimar el total de personas en hogares particulares a partir de Casen 2013. 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 2013.
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 , en paso 1) y el número de personas encuestadas en la comuna por la encuesta Casen 2013 (paso 2),

$$\hat{R}_C = \frac{N_C}{\sum_{h \in C} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,el,g,R}} N_{hij} \cdot w_{hij}^{NR}} \quad (22)$$

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:

$$w_{hijk}^C = \hat{R}_C \cdot w_{hij}^{NR} \quad (23)$$

ADVERTENCIA

La encuesta Casen 2013 es representativa para el país a nivel de áreas geográficas urbana y rural, y a nivel regional.

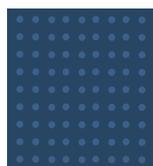
La encuesta Casen 2013 no es representativa a nivel comunal. Sin embargo por apego a la política de transparencia de la información generada en Casen 2013, y para fines de investigación es que se pone a disposición de los usuarios el factor de expansión comunal.

Por construcción del factor de expansión comunal, todos los individuos residentes en un mismo conglomerado (manzana o sección), comparten el mismo ponderador, que se interpreta como el número de personas en la comuna de residencia que representa un individuo de la muestra.

Si bien el ponderador se puede generar numéricamente, esto no significa que tenga sentido estadísticamente en todas las comunas.

Para poder ser utilizado a nivel de una comuna es necesario, antes de realizar cualquier análisis, evaluar el error estándar asociado a cada estimación de interés, ya que en algunas comunas el tamaño muestral no es suficiente para obtener estimaciones con suficiente precisión.

Al usar este factor de expansión comunal, es importante recordar que se expande sólo al total de la población residente en las comunas con presencia en la muestra Casen 2013. Esto, a diferencia del ponderador regional que expande a la población nacional.

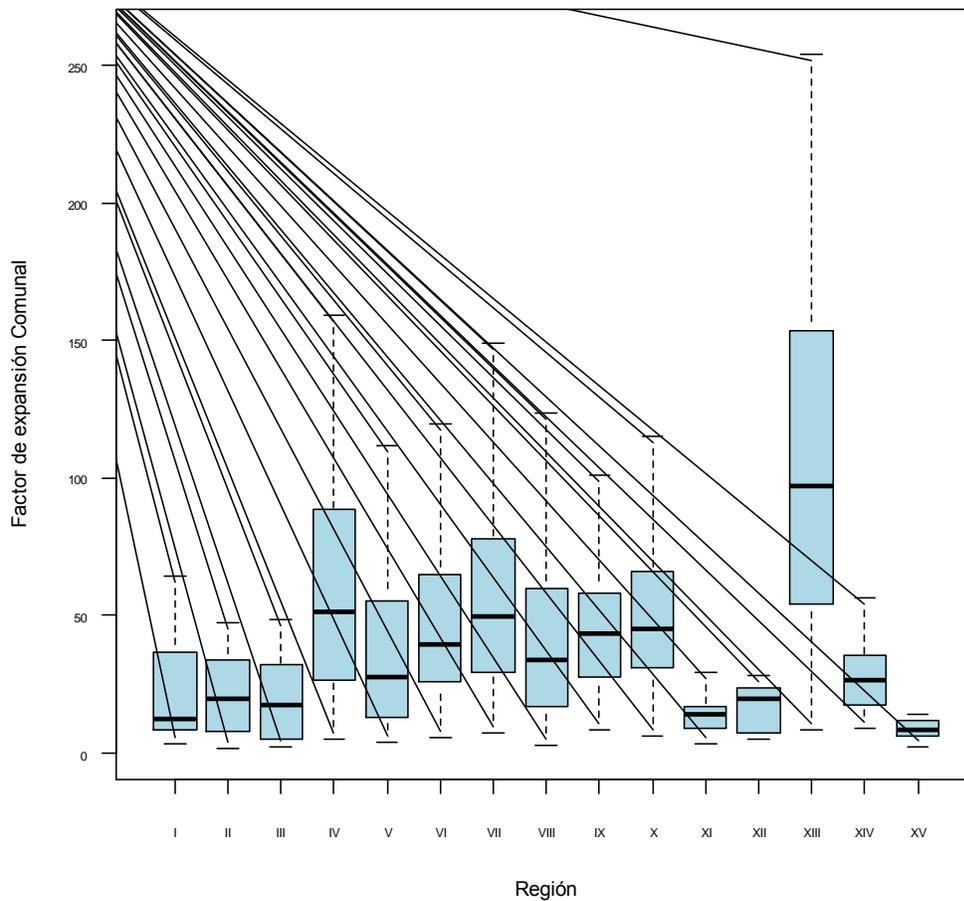


Las Figuras III. 16a y III.16b muestran la distribución regional de los ponderadores de calibración comunal para las unidades en los marcos rural y urbano, respectivamente.

De la comparación entre las Figuras III.16a y III.16b se desprende que las unidades provenientes del marco de manzanas, en general, poseen mayor variabilidad.

La región de Aysén posee los ponderadores más pequeños y con menor variabilidad, para las manzanas y secciones. Mientras que en la región Metropolitana se observa la mayor variabilidad y los ponderadores más grandes.

Figura III.16a: Distribución regional del Ponderador de Calibración Comunal de secciones (MS2002).



Fuente: Instituto Nacional de Estadísticas.

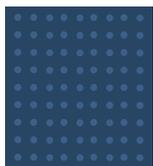
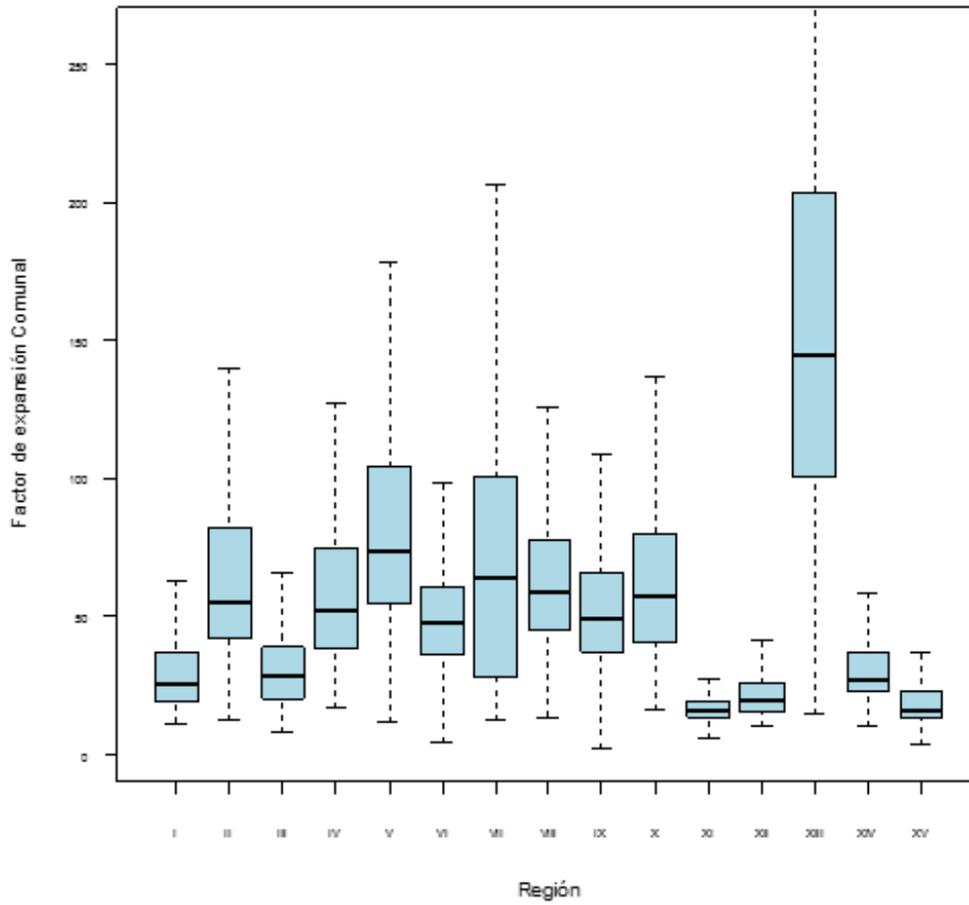
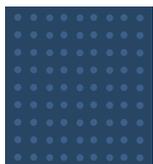


Figura III.16b: Distribución regional del Ponderador de Calibración Comunal de manzanas (MM2008).



Fuente: Instituto Nacional de Estadísticas.



IV. Estimación de varianzas

La configuración del Marco Muestral (manzanas/secciones) presenta una gran variedad de tamaños respecto al número de viviendas disponibles, esto motiva que la Casen considere probabilidades de selección desigual, lo que se debe considerar en los análisis e incorporar en los factores de expansión. Si no se usan factores de expansión, las viviendas con mayor probabilidad de selección pesan más en la muestra y distorsionan los resultados. Pero si se usan los factores con probabilidad desigual, cada persona en la muestra pesa lo que corresponde para representar a su región (o área geográfica).

Cuando se utiliza un diseño muestral Aleatorio Simple, las varianzas de los estimadores lineales como proporciones, medias y totales, tienen expresiones analíticas simples. Sin embargo, en diseños complejos como multietápicos y por conglomerados, las varianzas suelen tener expresiones analíticas más complejas de determinar e implementar, por lo que se recurre habitualmente a dos métodos de aproximación (Rust, 1985), el de linealización mediante el desarrollo de una serie de Taylor tanto del estimador como de la estimación de la varianza, y el de replicación a través de los métodos de Replicación Repetida de Jackknife (JRR), el de Replicación Repetida Equilibrada (BRR) y de Bootstrap.

En esta sección se presenta en forma concisa los puntos a considerar en el análisis de una encuesta de diseño complejo como Casen. Se sugiere a los investigadores revisar bibliografía especializada en estos temas (Wolter, 2007; Valliant et al., 2013; Westat, 2007; Heeringa et al., 2010).

1. El estimador

Casen es una encuesta en que según el diseño, las unidades de selección no tienen la misma probabilidad de ser elegidas, o en otras palabras, tienen probabilidades desiguales de ser seleccionadas. Debido a esto, las variables de interés, asociadas a las unidades de selección, deben ser ponderadas por el inverso de estas probabilidades, que comúnmente se denominan factores de expansión, y que indican a cuántos elementos en la población representa un elemento en la muestra seleccionada.

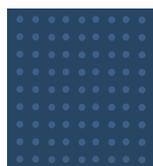
Supóngase que de una población U particionada en N unidades, $U = \{1, 2, 3, \dots, N\}$ se toma una muestra probabilística S de tamaño de tamaño n .

La probabilidad de inclusión de la unidad k está dada por $\pi_k = P(k \in S) > 0$ (diseño aleatorio), donde el término $w_k = 1/\pi_k$ es el inverso de la probabilidad de inclusión o ponderador del diseño asociado al individuo k .

La probabilidad que los elementos k y q pertenezcan a la muestra, está dada por $\pi_{kq} = P(k \in S \wedge q \in S) > 0$, siendo $w_{kq} = 1/\pi_{kq}$ el ponderador asociado a la inclusión conjunta de los elementos k y q en la muestra.

Algunos de los parámetros que se desea estimar son:

$t = \sum_{k \in U} y_k$ el total poblacional de la variable de interés Y ,
 $\bar{y}_U = \sum_{k \in U} y_k / N$, la media poblacional de la variable Y .



En un diseño aleatorio simple, el estimador Horvitz-Thompson permite realizar estimaciones insesgadas del total poblacional y de la media:

$$\hat{t}_{HT} = \sum_{i \in S} \frac{y_i}{\pi_i} = \sum_{i \in S} w_i y_i \quad (1)$$

$$\hat{y}_{HT} = \frac{\hat{t}_{HT}}{\hat{N}} = \frac{1}{\hat{N}} \sum_{i \in S} \frac{y_i}{\pi_i} = \frac{1}{\hat{N}} \sum_{i \in S} w_i y_i \quad (2)$$

Donde:

w_i , es el peso o ponderador del i -ésimo elemento según el diseño e indica a cuántos elementos representa en la población y

$\hat{N} = \sum w_i$, es el tamaño de la población estimada.

En **muestreo aleatorio simple** para poblaciones finitas, donde se extrae una muestra aleatoria de n elementos de una población que contiene N de ellos, el ponderador muestral es $w_k = N/n$, y el estimador de Horvitz-Thompson para el tamaño total es:

$$\hat{t}_{HT} = \sum_{i \in S} \frac{N}{n} y_i = N \sum_{i \in S} \frac{y_i}{n} = N \hat{y} \quad (3)$$

Donde $\hat{y} = \sum_{i \in S} \frac{y_i}{n}$, el estimador para la media poblacional.

En **muestreo aleatorio estratificado** en cada estrato independiente, el estimador del total dado por el estimador de Horvitz-Thompson es:

$$\hat{t}_{HT} = \sum_{k=1}^H \hat{t}_{HT_h} = \sum_{k=1}^H N_k \hat{y}_h \quad (4)$$

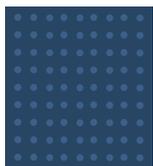
Donde:

$\hat{y}_h = \sum_{k \in H} \frac{y_h}{n_h}$, la media muestral en el estrato h , y

n_h es el número de elementos en la muestra dentro del estrato h .

Los **estimadores calibrados** son usados frecuentemente para obtener estimaciones más precisas, siendo necesario contar con información auxiliar correlacionada con las variables de interés y en que los totales son conocidos o son fijados a nivel poblacional.

La Encuesta Casen utiliza la calibración para ajustar las estimaciones de personas a los stocks poblacionales estimados por medio de proyecciones que son fijadas como las poblacionales. Adicionalmente, utiliza otros ajustes como la corrección por no respuesta, de modo que el factor de expansión contiene varios componentes que son necesarios para la obtención de las estimaciones.



Los estimadores calibrados al ser no lineales, son ligeramente sesgados, pero el sesgo tiende a cero rápidamente en muestras lo suficientemente grandes como la Encuesta Casen.

Para la estimación del total $t = \sum_{k \in U} y_k$ el estimador calibrado viene dado por:

$$\hat{t}_{cal,h} = \sum_{k \in S} w_k y_k \quad (5)$$

Para la estimación de una razón $R_{y/z} = \frac{\sum_{k \in U} y_k}{\sum_{k \in U} z_k}$ el estimador calibrado viene dado por:

$$\hat{R}_{cal,y/z} = \frac{\sum_{k \in S} w_k y_k}{\sum_{k \in S} w_k z_k} \quad (6)$$

Con

$$\hat{t}_{r_h} = \sum_{k \in S} w_k r_k,$$

Donde

$$r_k = y_k - \hat{R}_{cal,y/z} \cdot z_k$$

Estimaciones de totales desarrolladas a partir de la encuesta Casen incluyen, entre otras, el total de personas bajo la línea de pobreza y el total de personas de 60 años y más. Estimaciones de razón incluyen ejemplos tales como las tasas de cobertura de programas escolares y la tasa de desocupación.

2. Varianza del estimador

Las fórmulas anteriores para los estimadores Horvitz-Thompson (HT) de la población total son muy similares a las fórmulas que se aplican bajo muestreo con probabilidades iguales. Sin embargo, las fórmulas para el cálculo de la varianza del estimador HT para el total de la población, son más complejas:

$$\widehat{var}(\hat{t}_{HT}) = \sum_{k \in U} \sum_{q \in U} \left(\frac{w_k w_q}{w_{kq}} - 1 \right) y_k y_q \quad (7)$$

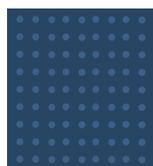
Un estimador insesgado de esta varianza del Total \hat{t}_{HT} es:

$$\widehat{var}(\hat{t}_{HT}) = \sum_{k \in S} \sum_{q \in S} (w_k w_q - w_{kq}) y_k y_q \quad (8)$$

Bajo el mismo diseño aleatorio simple para poblaciones finitas, los ponderadores son

$$w_k = \frac{N}{n} \quad y \quad w_{kq} = \frac{N}{n} \cdot \frac{(N-1)}{(n-1)}$$

El estimador de H-T para la varianza es, reemplazando en la expresión anterior (7):



$$\text{Var}(\hat{t}_{HT}) = N^2(1-f) \frac{S_U^2}{n} \quad (9)$$

Donde

$$S_U^2 = \sum_{k=1}^N \frac{(y_k - \bar{y})^2}{N-1} \quad \text{la varianza poblacional de la variable } Y.$$

El estimador de esta varianza de \hat{t}_{HT} es:

$$\widehat{\text{Var}}(\hat{t}_{HT}) = N^2(1-f) \frac{S_S^2}{n} \quad (10)$$

Donde

$$S_S^2 = \sum_{k=1}^n \frac{(y_k - \bar{y})^2}{n-1} \quad \text{la varianza muestral de la variable } Y.$$

Tanto en la expresión (9) como en la (10), $f = n/N$ es la fracción de muestreo y $(1-f) = fpc$ es la corrección por finitud o factor de ajuste para poblaciones finitas.

Bajo el diseño aleatorio estratificado, la varianza de este estimador HT viene dada por:

$$\widehat{\text{Var}}(\hat{t}_{HT}) = \sum_{h=1}^H N_h^2(1-f_h) \frac{S_{S_h}^2}{n_h} \quad (11)$$

Donde

$$S_{S_h}^2 = \sum_{k \in S_h} \frac{(y_k - \bar{y}_k)^2}{n_k - 1} \quad \text{Es la varianza muestral en el estrato } h \text{ de la variable } Y.$$

Donde $f_h = n_h/N_h$ es la fracción de muestreo en el estrato h , y $(1-f_h) = fpc_h$ es la corrección por finitud o factor de ajuste para poblaciones finitas en el estrato h .

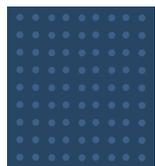
Para la estimación de varianzas calibradas y bajo el supuesto de muestreo aleatorio estratificado con remplazo, los software STATA (svyset), SPSS (csplan, en muestras complejas) y R (Survey, svydesign) utilizan las siguientes fórmulas con ponderadores w_k , lo que facilita los cálculos:

$$\widehat{\text{Var}}(\hat{t}_{cal}) = \sum_{k=1}^H \sum_{k \in S_h} \frac{(y_k w_k n_k - \hat{t}_k)^2}{n_k(n_k - 1)} \quad (12)$$

$$\widehat{\text{Var}}(\hat{R}_{cal,y/z}) = \sum_{k=1}^H \sum_{k \in S_h} \frac{(r_k w_k n_k - \hat{t}_k)^2}{n_k(n_k - 1)} \quad (13)$$

Es importante mencionar dos supuestos que son utilizados frecuentemente por los paquetes estadísticos en la estimación de la varianza:

1. Se asume que las unidades primarias de muestreo, en los diseños multi-etápicas, fueron seleccionadas con remplazo a partir de los estratos de muestreo. Se ignora cualquier



corrección por población finita. Como resultado, las estimaciones de varianza estarán levemente sobre-estimadas.

2. Se ignoran las etapas de selección al interior de la unidad primaria de muestreo. Es decir, se asume que las unidades fueron seleccionadas en una etapa al interior del conglomerado, por lo tanto la fórmula de estimación de la varianza sólo requiere el identificador de las unidades primarias de muestreo y los estratos primarios. Toda la variabilidad al interior de la unidad primaria se captura en el estimador compuesto de varianza.

3. 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.

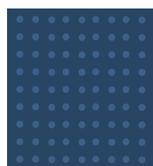
La encuesta Casen 2013 presenta un diseño muestral que se puede caracterizar como probabilístico, estratificado y multietápico. En el muestreo estratificado la estimación de la varianza total se obtiene mediante la suma de las varianzas estimadas en cada estrato y la varianza en cada estrato se estima 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 2013 consideró 589 estratos y 8.703 segmentos (7.723 manzanas y 980 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 2013 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 2013 y producir estimaciones de varianza más estables.



Las estimaciones oficiales de la varianza compleja de Casen 2013 han sido desarrolladas utilizando pseudo-estrato, pseudo-conglomerado y corrección por poblaciones finitas⁵¹. A continuación se describen los procedimientos seguidos para la construcción de estas variables.

3.1. 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 589 estratos 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 589 estratos muestrales se crearon 266 varstrat, los que poseen al menos dos conglomerados. La Tabla IV.1 presenta el total de estratos y varstrat por región.

⁵¹ Se pueden encontrar en principales estadísticas obtenidas con datos Casen 2013. Ver http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen_obj.php

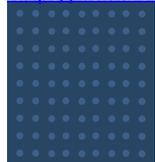


Tabla IV.1: Total de estratos y varstrat según región por área geográfica.

Región	Estratos	Varstrat	Urbano		Rural	
			Número de VarStrat	%	Número de VarStrat	%
País	589	266	180	67,4	87	32,6
XV de Arica y Parinacota	5	2	1	0,4	1	0,4
I de Tarapacá	9	4	2	0,7	2	0,7
II de Antofagasta	12	5	4	1,5	1	0,4
III de Atacama	16	9	7	2,6	2	0,7
IV de Coquimbo	28	11	6	2,2	5	1,9
V de Valparaíso	67	24	18	6,7	6	2,2
VI de O´Higgins	65	26	16	6,0	10	3,7
VII del Maule	60	23	13	4,9	10	3,7
VIII del Biobío	104	40	30	11,2	10	3,7
IX de La Araucanía	64	29	17	6,4	12	4,5
X de Los Lagos	48	19	9	3,7	10	3,7
XIV de Los Ríos	24	20	10	3,7	10	3,7
XI de Aysén	11	5	2	0,7	3	1,1
XII de Magallanes	6	3	2	0,7	1	0,4
XIII Metropolitana	70	46	42	15,7	4	1,5

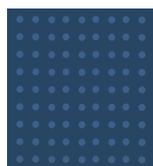
Fuente: Instituto Nacional de Estadísticas y Ministerio de Desarrollo Social.

3.2. 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 8.703 conglomerados originales, en donde cada uno contiene un número mínimo de viviendas.

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 171 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 muy cercano a 30 (superior a 24 e inferior a 30).

A partir de un total de 8.703 conglomerados seleccionados en la muestra Casen 2013, se crearon 1.827 varunit, los cuales agrupan entre 25 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 IV.2 presenta el total de conglomerados (Manzanas/Secciones) y varunit por región.

Tabla IV.2: Total de conglomerados y varunit (pseudo-conglomerados) según región.

Región	Manzanas /Secciones	Varunit
I de Tarapacá	361	75
II de Antofagasta	370	61
III de Atacama	386	65
IV de Coquimbo	469	84
V de Valparaíso	878	179
VI de O´Higgins	558	136
VII del Maule	515	127
VIII del Biobío	1.152	262
IX de La Araucanía	677	151
X de Los Lagos	464	115
XI de Aysén	328	56
XII de Magallanes	365	58
XIII Metropolitana	1389	288
XIV de Los Ríos	404	99
XV de Arica y Parinacota	387	71
País	8.703	1.827

Fuente: Instituto Nacional de Estadísticas.

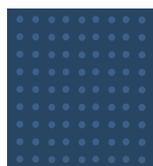


Tabla IV.3: Total de conglomerados y varunit (pseudo-conglomerados) según región por número de viviendas.

Región	Varunit / número de viviendas por Varstrat				
	25-29	30-39	40-49	50-59	Total
País	66	1.376	292	93	1.827
XV Arica y Parinacota	1	61	8	1	71
I Tarapacá	0	62	8	5	75
II Antofagasta	1	53	5	2	61
III Atacama	3	56	4	2	65
IV Coquimbo	1	66	14	3	84
V Valparaíso	4	141	27	7	179
VI Bdo. O'Higgins	7	90	30	9	136
VII Maule	5	89	27	6	127
VIII BíoBío	11	186	51	14	262
IX La Araucanía	6	107	29	9	151
X Los Lagos	4	89	16	6	115
XIV Los Ríos	7	61	25	6	99
XI Aysén	1	52	1	2	56
XII Magallanes	1	52	4	1	58
XIII Metropolitana	14	211	43	20	288

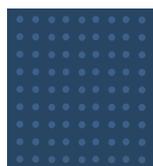
Fuente: Instituto Nacional de Estadísticas, Ministerio de Desarrollo Social.

4. Resultados a nivel regional, por áreas geográficas urbana y rural, a nivel nacional.

Las Tablas IV.4a y IV.4b presentan las estimaciones de la tasa de pobreza y pobreza extrema a nivel nacional y sus márgenes de error (al 95% de confianza) para las muestras Casen 1996 a 2013. La tabla IV.4a informa las estimaciones para la metodología tradicional de medición de la pobreza y la Tabla IV.4b lo hace para la metodología nueva que actualiza la línea de pobreza y la forma de medir ingresos.

En Casen 2013 los márgenes de error de estas estimaciones (para ambas metodologías de medición de ingresos y pobreza) fueron determinados utilizando los parámetros varstrat, varunit y expr (factor de expansión), considerando la utilización de pseudo-estratos y pseudo-conglomerados, siendo éstos implementados mediante cálculos realizados en paralelo utilizando los programas SPSS v.20 y STATA v.12.

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 de fueron estimados utilizando los parámetros estrato, segmento, expr_r2, ya que no se



dispone para la serie histórica de pseudo-estratos (varstrat) y pseudo-conglomerados (varunit).

Tabla VI.4a: Incidencia de la Pobreza y de la Pobreza Extrema, 1996-2013 (Estimación (%) y Margen de Error) según metodología tradicional

Año	Pobreza		Pobreza Extrema	
	Estimación (%)	Margen de error	Estimación (%)	Margen de error
1996	23,21	0,94	5,75	0,43
1998	21,66	0,75	5,62	0,39
2000	20,22	0,76	5,58	0,38
2003	18,66	0,62	4,67	0,3
2006	13,71	0,58	3,21	0,26
2009	11,44	0,55	3,59	0,28
2011	10,89	0,73	3,14	0,30
2013	7,80	0,44	2,50	0,21

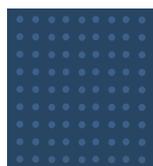
Nota: Estimaciones 2009 y 2011 realizadas por Ministerio de Desarrollo Social.
Fuente: Instituto Nacional de Estadística, Ministerio de Desarrollo Social.

Tabla VI.4a: Incidencia de la Pobreza y de la Pobreza Extrema, 2006-2013 (Estimación (%) y Margen de Error) según nueva metodología

Año	Pobreza		Pobreza Extrema	
	Estimación (%)	Margen de error	Estimación (%)	Margen de error
2006	29,11	0,95	12,58	0,56
2009	25,27	0,89	9,88	0,51
2011	22,19	0,95	8,10	0,52
2013	14,38	0,59	4,51	0,29

Nota: Estimaciones 2006, 2009 y 2011 realizadas por Ministerio de Desarrollo Social.
Fuente: Instituto Nacional de Estadística, Ministerio de Desarrollo Social.

En la tabla IV.5 se presentan las tasas de pobreza y pobreza extrema para las personas y los hogares para los tres niveles de representatividad de la encuesta Casen 2013 (Regional, por área geográfica urbana y rural y nacional). Al igual, se calcularon los márgenes de error al 95% de confianza, asociados a estas estimaciones, los que figuran en la tabla IV.6. El margen de error fue estimado utilizando los parámetros varstrat, varunit y expr, considerando la utilización de pseudo-estratos y pseudo-conglomerados. Los tamaños



muestrales asociados a las entrevistas efectivas de cada año se presentan en la última fila de la Tabla VI.7.

La estimación de las tasas de pobreza y de pobreza extrema, al igual que los errores muestrales en Casen 2013, se calcularon utilizando la metodología nueva y tradicional.

Tabla VI.5: Estimaciones de tasas de pobreza y de pobreza extrema de personas y hogares en Casen 2013, según nivel de representatividad, por metodología de medición de ingresos y pobreza.

Niveles de representatividad de Casen 2013		Personas				Hogares			
		Pobres		Pobres Extremos		Pobres		Pobres Extremos	
		MT	MN	MT	MN	MT	MN	MT	MN
Regiones	1 Tarapacá	4,6	8,2	1,6	2,2	4,1	7,5	1,7	2,0
	2 Antofagasta	4,4	4,0	1,4	1,1	3,7	3,2	1,3	1,0
	3 Atacama	6,8	7,3	3,1	2,0	5,6	6,1	2,4	1,7
	4 Coquimbo	9,4	16,2	3,4	5,3	7,5	14,6	2,9	4,4
	5 Valparaíso	9,1	15,6	2,8	4,5	7,6	13,6	2,8	3,9
	6 O'Higgins	6,7	16,0	2,3	4,7	5,5	14,2	2,1	4,2
	7 Maule	8,0	22,3	2,6	5,9	6,9	20,7	2,4	5,4
	8 Bío Bío	12,3	22,3	4,1	8,0	10,5	20,4	3,7	7,1
	9 La Araucanía	12,0	27,9	3,4	10,6	9,8	25,1	2,9	9,0
	10 Los Lagos	8,1	17,6	2,9	5,7	6,6	15,8	2,6	4,9
	11 Aysén	4,7	6,8	0,9	1,6	3,9	6,2	1,0	1,6
	12 Magallanes	3,2	5,6	1,0	3,2	2,9	4,3	1,1	2,3
	13 Metropolitana	5,9	9,2	1,8	2,6	4,9	7,8	1,7	2,2
	14 Los Ríos	10,4	23,1	3,2	8,0	8,6	19,9	2,8	6,7
	15 Arica y Parinacota	9,9	14,6	3,7	4,7	7,8	11,7	3,1	3,8
Áreas geográficas	Urbano	8,0	12,4	2,5	3,8	6,7	10,9	2,3	3,3
	Rural	6,7	27,9	2,8	9,6	5,5	25,4	2,5	8,3
Nacional		7,8	14,4	2,5	4,5	6,5	12,8	2,3	3,9

Nota: MT: estimación según metodología tradicional;

MN: estimación según metodología nueva.

Fuente: Instituto Nacional de Estadística, Ministerio de Desarrollo Social.

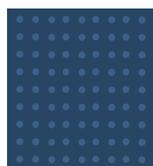


Tabla VI.6: Márgenes de error para tasas de pobreza y pobreza extrema de personas y hogares en Casen 2013, por niveles de representatividad, según metodología de medición de ingresos y pobreza.

Niveles de representatividad de Casen 2013		Personas				Hogares			
		Pobres		Pobres Extremos		Pobres		Pobres Extremos	
		MT	MN	MT	MN	MT	MN	MT	MN
Regiones	1 Tarapacá	1,3	1,6	0,6	0,7	1,1	1,3	0,6	0,6
	2 Antofagasta	1,3	1,3	0,7	0,6	0,9	0,8	0,6	0,4
	3 Atacama	2,9	2,5	2,5	0,8	2,0	1,7	1,8	0,6
	4 Coquimbo	1,9	2,4	1,1	1,6	1,5	2,1	0,9	1,2
	5 Valparaíso	1,4	1,9	0,7	0,8	1,2	1,7	0,8	0,6
	6 O'Higgins	1,2	2,0	0,6	1,0	0,9	1,6	0,5	0,7
	7 Maule	1,4	2,2	0,7	1,3	1,2	2,1	0,7	1,1
	8 Bío Bío	1,3	1,9	0,7	1,0	1,0	1,6	0,5	0,8
	9 La Araucanía	1,7	2,2	0,7	1,5	1,2	1,8	0,5	1,2
	10 Los Lagos	1,3	2,1	0,7	1,1	1,0	1,8	0,6	0,9
	11 Aysén	1,5	1,6	0,4	0,8	1,1	1,2	0,4	0,7
	12 Magallanes	1,0	3,2	0,5	3,3	0,8	1,7	0,5	1,8
	13 Metropolitana	0,8	1,0	0,4	0,5	0,6	0,9	0,3	0,4
	14 Los Ríos	1,8	2,5	0,7	1,3	1,4	2,2	0,7	1,1
	15 Arica y Parinacota	1,5	2,0	1,1	1,3	1,2	1,4	0,9	1,0
Áreas geográficas	Urbano	0,5	0,6	0,2	0,3	0,4	0,5	0,2	0,2
	Rural	0,8	1,4	0,5	1,0	0,6	1,3	0,4	0,8
Nacional		0,4	0,6	0,2	0,3	0,4	0,5	0,2	0,2

Nota: Márgenes de error (MOE(p)) al 95% de Confianza de la Tasa de Pobreza (p) en la muestra Casen 2013.

MT: estimación según metodología tradicional;

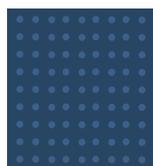
MN: estimación según metodología nueva.

Fuente: Instituto Nacional de Estadística, Ministerio de Desarrollo Social.

5. Márgenes de error, proyectados y efectivos, a nivel nacional y regional

La muestra de la encuesta Casen 2013 fue optimizada a nivel regional esto quiere decir que su diseño buscaba lograr márgenes de error entre 1 y 4 puntos porcentuales en cada región del país.

La Tabla VI.7, muestra para cada región los márgenes de error de la tasa de pobreza (variable de interés) calculados en base al diseño teórico de la muestra (entrevistas proyectadas) y los márgenes de error calculados en base a la muestra lograda (entrevistas efectivas). La columna (2) muestra los márgenes de error proyectados para cada región en base a la muestra completa (n=70.080 viviendas).

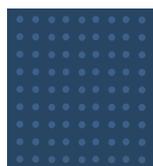


Los errores efectivos en Casen 2013 (n=66.725 hogares) se presenta en la columna (4), que incluye la apertura MT (Metodología Tradicional) y MN (Metodología Nueva). Los márgenes de error a nivel regional se mantuvieron dentro de los límites esperados de 1 a 4 puntos porcentuales.

Tabla VI.7: Márgenes de error proyectados y efectivos a nivel regional y nacional, Casen 2013.

Muestra 2013					
Región	Entrevistas Proyectadas	MOE Proyectado	Entrevistas Efectivas	MOE Efecivo	
				MT	MN
1 Tarapacá	2.977	3,10	2.786	1,33	1,62
2 Antofagasta	2.513	2,40	2.230	1,29	1,35
3 Atacama	2.071	4,00	2.282	2,91	2,52
4 Coquimbo	3.018	2,90	3.038	1,88	2,43
5 Valparaíso	6.737	2,00	6.462	1,36	1,94
6 O´Higgins	5.137	2,10	5.097	1,17	1,99
7 Maule	4.872	3,50	4.658	1,36	2,24
8 Bío Bío	9.826	1,90	9.599	1,31	1,94
9 La Araucanía	5.387	2,60	5.436	1,66	2,19
10 Los Lagos	4.192	2,90	4.077	1,31	2,12
11 Aysén	2.051	4,00	1.865	1,53	1,63
12 Magallanes	2.085	2,10	1.916	1,01	3,23
13 Metropolitana	12.865	1,00	10.981	0,82	0,99
14 Los Ríos	3.768	4,00	3.660	1,75	2,46
15 Arica y Parinacota	2.581	2,50	2.638	1,47	2,00
Nacional	70.080	0,69	66.725	0,4	0,59

Fuente: Instituto Nacional de Estadística, Ministerio de Desarrollo Social.



6. Programas computacionales

Figura IV.1: Sintaxis de Estimación en SPSS 22.

```

*****Plan de Muestro Cálculo Tasa de Pobreza y Tasa de Pobreza Extrema CASEN 2013*****

*Diseño Plan de Muestras Complejas.
CSPLAN ANALYSIS
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/PLANVARS ANALYSISWEIGHT=expr
/SRSESTIMATOR TYPE=WOR
/PRINT PLAN
/DESIGN STRATA=VarStrat CLUSTER=VarUnit
/ESTIMATOR TYPE=WR.

*Transformación de Variables para Identificar Pobres y Pobres Extremos.
DATASET ACTIVATE Conjunto_de_datos1.
RECODE pobreza_ingreso (1=1) (2=1) (3=0) INTO pobreza_MN1.
VARIABLE LABELS pobreza_MT1 'Pobreza Extrema y No Extrema (Pobres) Metodo Nuevo'.
EXECUTE.

DATASET ACTIVATE Conjunto_de_datos1.
RECODE pobreza_ingreso (1=1) (2=0) (3=0) INTO pobreza_MN2.
VARIABLE LABELS pobreza_MT2 'Pobreza Extrema Metodo Nuevo'.
EXECUTE.

DATASET ACTIVATE Conjunto_de_datos1.
RECODE pobreza_MT (1=1) (2=1) (3=0) INTO pobreza_MT1.
VARIABLE LABELS pobreza_MN1 'Pobreza Extrema y No Extrema (Pobres) Metodo Tradicional'.
EXECUTE.

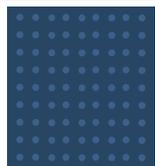
DATASET ACTIVATE Conjunto_de_datos1.
RECODE pobreza_MT (1=1) (2=0) (3=0) INTO pobreza_MT2.
VARIABLE LABELS pobreza_MN2 'Pobreza Extrema Metodo Tradicional'.
EXECUTE.

*****ESTIMACIÓN DE PERSONAS*****.

*Estimaciones de Pobreza y Pobreza Extrema a Nivel Nacional, usando el Método Tradicional y Nuevo.
(Personas).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT1 pobreza_MT2 pobreza_MN1 pobreza_MN2
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Pobreza a Nivel Regional, usando el Método Nuevo.(Personas).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MN1
/SUBPOP TABLE=region DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

```



```
*Estimaciones de Extrema Pobreza a Nivel Regional, usando el Método Nuevo.(Personas).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MN2
/SUBPOP TABLE=region DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```

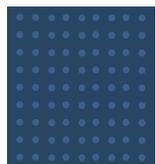
```
*Estimaciones de Pobreza a Nivel Regional, usando el Método Tradicional.(Personas).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT1
/SUBPOP TABLE=region DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```

```
*Estimaciones de Extrema Pobreza a Nivel Regional, usando el Método Tradicional.(Personas).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT2
/SUBPOP TABLE=region DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```

```
*Estimaciones de Pobreza a Nivel Urbano-Rural, usando el Método Nuevo.(Personas).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MN1
/SUBPOP TABLE=zona DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```

```
*Estimaciones de Pobreza Extrema a Nivel Urbano-Rural, usando el Método Nuevo.(Personas).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MN2
/SUBPOP TABLE=zona DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```

```
*Estimaciones de Pobreza a Nivel Urbano-Rural, usando el Método Tradicional.(Personas).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT1
/SUBPOP TABLE=zona DISPLAY=LAYERED
```



```

/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Pobreza Extrema a Nivel Urbano-Rural, usando el Método Tradicional.(Personas).
CSD DESCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT2
/SUBPOP TABLE=zona DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*****ESTIMACIÓN DE HOGARES*****

*Transformación de Base de Datos de Personas a Hogar, filtrando por Jefe de Hogar.
USE ALL.
COMPUTE filter_$=(pco1 = 1).
VARIABLE LABELS filter_$ 'pco1 = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

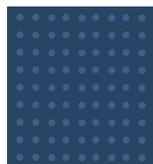
*Estimaciones de Pobreza y Pobreza Extrema a Nivel Nacional, usando el Método Tradicional y Nuevo.
(Hogares).
CSD DESCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT1 pobreza_MT2 pobreza_MN1 pobreza_MN2
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Pobreza a Nivel Regional, usando el Método Nuevo.(Hogares).
CSD DESCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MN1
/SUBPOP TABLE=region DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Extrema Pobreza a Nivel Regional, usando el Método Nuevo.(Hogares).
CSD DESCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MN2
/SUBPOP TABLE=region DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Pobreza a Nivel Regional, usando el Método Tradicional.(Hogares).
CSD DESCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN

```



```
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT1
/SUBPOP TABLE=region DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

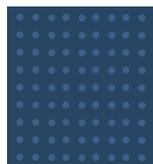
*Estimaciones de Extrema Pobreza a Nivel Regional, usando el Método Tradicional.(Hogares).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT2
/SUBPOP TABLE=region DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Pobreza a Nivel Urbano-Rural, usando el Método Nuevo.(Hogares).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MN1
/SUBPOP TABLE=zona DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Pobreza Extrema a Nivel Urbano-Rural, usando el Método Nuevo.(Hogares).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MN2
/SUBPOP TABLE=zona DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Pobreza a Nivel Urbano-Rural, usando el Método Tradicional.(Hogares).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT1
/SUBPOP TABLE=zona DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

*Estimaciones de Pobreza Extrema a Nivel Urbano-Rural, usando el Método Tradicional.(Hogares).
CSDSCRIPTIVES
/PLAN FILE='C:\Users\mjeria\Desktop\ESTIMACION DE VARIANZAS CASEN 2013 INE\BASES CASEN
'+
'2013\plan_casen2013.csaplan'
/SUMMARY VARIABLES=pobreza_MT2
/SUBPOP TABLE=zona DISPLAY=LAYERED
/MEAN
/STATISTICS SE CV COUNT POPSIZE DEFF
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```



*Transformación de Base de Datos de Hogar a Personas.
Filter Off.
EXECUTE.

Figura IV.2: Sintaxis de Estimación en SPSS Stata 12

Metodología Nueva

En primer lugar se crea la variable corte_p_MN: Pobreza_ingreso 2013 MN

```
gen cortep_mn = 1 if (pobreza_ingreso==1 | pobreza_ingreso==2)
replace cortep_mn=0 if (pobreza_ingreso==3)
label var cortep_mn "corte pobreza_ingreso"
label define cortep_mn 1 "pobres" 0 "no pobres"
label values cortep_mn cortep_mn
```

Calculo para PERSONAS:

1. Porcentaje de personas en pobreza, por región, año 2013 MN

```
svyset varunit [w=expr], strata(varstrat)
svy: prop cortep_mn, level(95)
svy: prop cortep_mn, level(95) over(region)
```

2. Porcentaje de personas en pobreza extrema y pobres no extremo, por región, año 2013 MN

```
svyset varunit [pw=expr], strata(varstrat)
svy: prop pobreza_ingreso, level(95)
svy: prop pobreza_ingreso, level(95) over(region)
```

3. Porcentaje de personas en pobreza, por zona, año 2013 MN

```
svyset varunit [w=expr], strata(varstrat)
svy: prop cortep_mn, level(95)
svy: prop cortep_mn, level(95) over(zona)
```

4. Porcentaje de personas en pobreza extrema y pobres no extremo, por zona, año 2013 MN

```
svyset varunit [pw=expr], strata(varstrat)
svy: prop pobreza_ingreso, level(95)
svy: prop pobreza_ingreso, level(95) over(zona)
```

Calculo para HOGARES

5. Porcentaje de Hogares en pobreza, por región, año 2013 MN

```
svyset varunit [w=expr], strata(varstrat)
svy: prop cortep_mn if pco1==1, level(95)
svy: prop cortep_mn if pco1==1, level(95) over(region)
```

6. Porcentaje de Hogares en pobreza extrema y pobres no extremo, por región, año 2013 MN

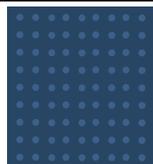
```
svyset varunit [pw=expr], strata(varstrat)
svy: prop pobreza_ingreso if pco1==1, level(95)
svy: prop pobreza_ingreso if pco1==1, level(95) over(region)
```

7. Porcentaje de Hogares en pobreza, por zona, año 2013 MN

```
svyset varunit [w=expr], strata(varstrat)
svy: prop cortep_mn if pco1==1, level(95)
svy: prop cortep_mn if pco1==1, level(95) over(zona)
```

8. Porcentaje de Hogares en pobreza extrema y pobres no extremo, por zona, año 2013 MN

```
svyset varunit [pw=expr], strata(varstrat)
svy: prop pobreza_ingreso if pco1==1, level(95)
svy: prop pobreza_ingreso if pco1==1, level(95) over(zona)
```



Metodología Tradicional

En primer lugar se crea la variable cortept_MT: Pobreza_mt 2013 MT

```
gen cortept_mt =1 if (pobreza_mt==1 | pobreza_mt==2)
replace cortept_mt=0 if (pobreza_ingreso==3)
label var cortept_mt "corte pobreza_mt"
label define cortept_mt 1 "pobres" 0 "no pobres"
label values cortept_mt cortept_mt
```

Calculo para PERSONAS:

9. Porcentaje de personas en pobreza, por región, año 2013 MT

```
svyset varunit [w=expr], strata(varstrat)
svy: prop cortept_mt, level(95)
svy: prop cortept_mt, level(95) over(region)
```

10. Porcentaje de personas en pobreza extrema y pobres no extremo, por región, año 2013 MT

```
svyset varunit [pw=expr], strata(varstrat)
svy: prop pobreza_mt, level(95)
svy: prop pobreza_mt, level(95) over(region)
```

11. Porcentaje de personas en pobreza, por zona, año 2013 MT

```
svyset varunit [w=expr], strata(varstrat)
svy: prop cortept_mt, level(95)
svy: prop cortept_mt, level(95) over(zona)
```

12. Porcentaje de personas en pobreza extrema y pobres no extremo, por zona, año 2013 MT

```
svyset varunit [pw=expr], strata(varstrat)
svy: prop pobreza_mt, level(95)
svy: prop pobreza_mt, level(95) over(zona)
```

Calculo para HOGARES

13. Porcentaje de Hogares en pobreza, por región, año 2013 MT

```
svyset varunit [w=expr], strata(varstrat)
svy: prop cortept_mt if pco1==1, level(95)
svy: prop cortept_mt if pco1==1, level(95) over(region)
```

14. Porcentaje de Hogares en pobreza extrema y pobres no extremo, por región, año 2013 MT

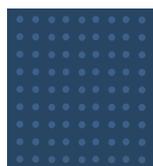
```
svyset varunit [pw=expr], strata(varstrat)
svy: prop pobreza_mt if pco1==1, level(95)
svy: prop pobreza_mt if pco1==1, level(95) over(region)
```

15. Porcentaje de Hogares en pobreza, por zona, año 2013 MT

```
svyset varunit [w=expr], strata(varstrat)
svy: prop cortept_mt if pco1==1, level(95)
svy: prop cortept_mt if pco1==1, level(95) over(zona)
```

16. Porcentaje de Hogares en pobreza extrema y pobres no extremo, por zona, año 2013 MT

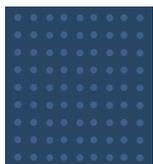
```
svyset varunit [pw=expr], strata(varstrat)
svy: prop pobreza_mt if pco1==1, level(95)
svy: prop pobreza_mt if pco1==1, level(95) over(zon)
```



V. Anexos

Anexo 1: Referencias

- 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.
- Groves et Al. (2009) *Survey Methodology*. New Jersey: John Wiley & Sons.
- Kish (1965), *Survey Sampling*, J. Wiley.
- Sitter RR (1992) Comparing Three Bootstrap Methods for Survey Data. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 20:135-154
- Turner A. (2003) Sampling frames and master samples. Expert Group Meeting to Review the Draft Handbook on Designing of Household Sample Surveys 3-5 December 2003. UNITED NATIONS SECRETARIAT ESA/STAT/AC.93/. Statistics Division, 03 November 2003.
- Valliant R, Brick JM, Dever JA (2008) Weight Adjustments for the Grouped Jackknife Variance Estimator. *Journal of Oficial Statistics* 24(3):469-488.
- 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 Ocial 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



Anexo 2: Constitución de dos marcos muestrales independientes para la selección de muestras para encuesta sociales, INE

La figura de dos marcos muestrales mutuamente excluyentes en las encuestas sociales, en las que el INE selecciona la muestra, obedece a opciones metodológicas y operativas, cuyos principales elementos se exponen en Capítulo 6, Anexo 2.

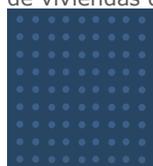
A mediados del periodo intercensal, investigaciones realizadas por el INE indicaban que el marco muestral en uso (MS2002), en el cual las unidades geográficas más pequeñas eran las secciones de empadronamiento censal, presentaba deficiencias, por lo que a partir de 2008 el INE desarrolló un nuevo marco de muestreo, definiciones para la conformación de unidades de muestreo y procedimientos para la selección de unidades de muestreo en las áreas urbanas del país⁵².

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 marco muestral 2008, las manzanas censales (predios urbanos delimitados por calles) reemplazan a las secciones de empadronamiento censal en las áreas urbanas⁵³. Esto se realizó con la idea de actualizar más rápidamente las unidades muestrales que el marco antiguo y así 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), por medio del plan municipal de edificaciones entregado por convenio al Departamento de Gestión del Marco del INE.

El INE comenzó a seleccionar las muestras urbanas utilizando el nuevo marco de muestreo a partir del año 2008. Las primeras encuestas cuyas muestras urbanas representativas a nivel nacional que fueron seleccionadas a partir del nuevo marco fueron la Nueva Encuesta del Empleo (NENE) en 2009, la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) en 2008, y la Encuesta de Uso de Drogas en Población General en 2008. En las zonas rurales y RAU, el INE sigue seleccionando sus muestras a partir del marco de secciones (MS2002).

⁵² 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.

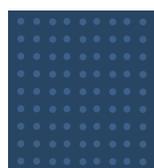
⁵³ Las secciones están conformadas por una o más manzanas a requerimiento de totalizar en ella una cantidad fija de viviendas de acuerdo al estrato comuna-zona.



Anexo 3: Resto de Áreas Urbanas (RAU) seleccionadas desde el MS2002.**Tabla V.1: Listado de comunas cuyas áreas RAU fueron seleccionadas desde el MS2002.**

Nº	Nombre Comuna	Nº	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	Laque 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	Trequaco
42	Panquehue	86	Vichuquén
43	Papudo	87	Vilcún
44	Paredones	88	Yerbas Buenas
		89	Zapallar

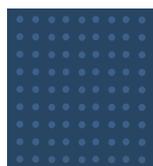
Fuente: Instituto Nacional de Estadísticas.



Anexo 4: Áreas RAU seleccionadas desde el MM2008.**Tabla V.2: 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	Pitrufquén		

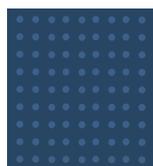
Fuente: Instituto Nacional de Estadísticas.



Anexo 5: grupos de tamaño**Tabla V.3: Grupos de Tamaño de marco muestral INE en Áreas Urbanas.**

Grupo	Límites (viviendas)	Número de viviendas a encuestar	Fracción de muestreo mínima	Fracción de muestreo máxima
0	[1 - 7]			
1	[8 - 9]	2	22%	25%
2	[10 - 13]	3	23%	30%
3	[14 - 17]	4	24%	29%
4	[18 - 23]	5	22%	28%
5	[24 - 25]	6	24%	25%
6	[26 - 29]	7	24%	27%
7	[30 - 33]	8	24%	27%
8	[34 - 37]	9	24%	26%
9	[38 - 44]	10	23%	26%
10	[45 - 45]	11	24%	24%
11	[46 - 49]	12	24%	26%
12	[50 - 53]	13	25%	26%
13	[54 - 57]	14	25%	26%
14	[58 - 61]	15	25%	26%
15	[62 - 65]	16	25%	26%
16	[66 - 69]	17	25%	26%
17	[70 - 73]	18	25%	26%
18	[74 - 77]	19	25%	26%
19	[78 - 81]	20	25%	26%
20	[82 - 85]	21	25%	26%
21	[86 - 89]	22	25%	26%
22	[90 - 93]	23	25%	26%
23	[94 - 97]	24	25%	26%
24	[98 - 102]	25	25%	26%
25	[103 - 105]	26	25%	25%
26	[106 - 109]	27	25%	25%
27	[110 - 113]	28	25%	25%
28	[114 - 154]	29	19%	25%
29	[155 - 303]	30	10%	19%
30	[304 - 1043]	31	3%	10%

Fuente: Instituto Nacional de Estadísticas.



Anexo 6: total de viviendas según subestrato y área, para cada estrategia**Tabla V.4: Total de viviendas seleccionadas contenidas en comunas de acuerdo a cada subestrato, de acuerdo a la estrategia 1 y según área.**

Estrategia 1	Ciudad(CD)⁵⁴	RAU	Rural	Total
1	8.595	602	2.465	11.662
2	21.305	1.281	3.628	26.214
3	39.576	2.280	8.921	50.777
IF	3.187	1.040	1.441	5.668
Total	72.663	5.203	16.455	94.321

Fuente: Instituto Nacional de Estadísticas.

Tabla V.5: Total de viviendas seleccionadas contenidas en comunas de acuerdo a cada subestrato, de acuerdo a la estrategia 2 y según área.

Estrategia 2	Ciudad	RAU	Rural	Total
1	18.755	107	2.181	21.043
2	11.588	1.959	3.277	16.824
3	39.133	2.097	9.556	50.786
IF	3.187	1.040	1.441	5.668
Total	72.663	5.203	16.455	94.321

Fuente: Instituto Nacional de Estadísticas.

Tabla V.6: Total de viviendas seleccionadas contenidas en comunas de acuerdo a cada subestrato, de acuerdo a la estrategia 3 y según área.

Estrategia 3	Ciudad	RAU	Rural	Total
1	7.892	219	2.752	10.863
2	21.429	1.254	3.393	26.076
3	40.155	2.690	8.869	51.714
IF	3.187	1.040	1.441	5.668
Total	72.663	5.203	16.455	94.321

Fuente: Instituto Nacional de Estadísticas.

Tabla V.7: Total de viviendas seleccionadas contenidas en comunas de acuerdo a cada subestrato, de acuerdo a la estrategia 4 y según área.

Estrategia 4	Ciudad	RAU	Rural	Total
1	7.014	107	1.842	8.963
2	17.692	1.016	2.985	21.693
3	44.770	3.040	10.187	57.997
IF	3.187	1.040	1.441	5.668
Total	72.663	5.203	16.455	94.321

Fuente: Instituto Nacional de Estadísticas.

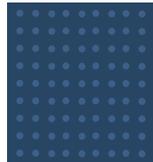
⁵⁴ CD=ciudad

Tabla V.8: Total de viviendas seleccionadas contenidas en comunas de acuerdo a cada subestrato, de acuerdo a la estrategia 5 y según área.

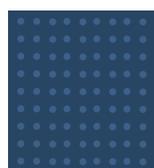
Estrategia 5	Ciudad	RAU	Rural	Total
1	12.747	107	1.247	14.101
2	15.188	1.277	2.891	19.356
3	41.541	2.779	10.876	55.196
IF	3.187	1.040	1.441	5.668
Total	72.663	5.203	16.455	94.321

Fuente: Instituto Nacional de Estadísticas.

Tabla V.9: Total de viviendas seleccionadas contenidas en comunas de acuerdo a cada subestrato, de acuerdo a la estrategia 6 y según área.

Estrategia 6	Ciudad	RAU	Rural	Total
1	5.222	219	1.896	7.337
2	13.842	774	2.749	17.365
3	50.412	3.170	10.369	63.951
IF	3.187	1.040	1.441	5.668
Total	72.663	5.203	16.455	94.321

Fuente: Instituto Nacional de Estadísticas.



Anexo 8: comunas reenumeradas para casen 2013**Tabla V.10: Comunas cuyas Unidades Primarias de Muestreo fueron seleccionadas para Casen 2012 y reenumeradas para Casen 2013.**

RPC⁵⁵	Nombre Comuna	RPC	Nombre Comuna	RPC	Nombre Comuna
1107	Alto Hospicio	6309	Pumanque	9107	Gorbea
1402	Camiña	6310	Santa Cruz	9108	Lautaro
1404	Huara	7102	Constitución	9111	Nueva Imperial
1405	Pica	7103	Curepto	9115	Pucón
2101	Antofagasta	7109	San Clemente	9119	Vilcún
2103	Sierra Gorda	7203	Pelluhue	9202	Collipulli
2302	María Elena	7302	Hualañé	9203	Curacautín
3102	Caldera	7304	Molina	10105	Frutillar
3202	Diego de Almagro	7307	Sagrada Familia	10109	Puerto Varas
3303	Freirina	7308	Teno	10202	Ancud
3304	Huasco	7309	Vichuquén	10206	Puqueldón
4102	Coquimbo	7401	Linares	10207	Queilén
4106	Vicuña	7404	Parral	10208	Quellón
4201	Illapel	7405	Retiro	10301	Osorno
4203	Los Vilos	7406	San Javier	10302	Puerto Octay
4304	Punitaqui	7407	Villa Alegre	10305	Río Negro
5102	Casablanca	8101	Concepción	13101	Santiago
5103	Concón	8102	Coronel	13103	Cerro Navia
5105	Puchuncaví	8106	Lota	13107	Huechuraba
5107	Quintero	8110	Talcahuano	13109	La Cisterna
5109	Viña del Mar	8111	Tomé	13111	La Granja
5302	Calle Larga	8207	Tirúa	13120	Ñuñoa
5401	La Ligua	8301	Los Ángeles	13123	Providencia
5403	Papudo	8303	Cabrero	13125	Quilicura
5405	Zapallar	8306	Nacimiento	13126	Quinta Normal
5504	La Cruz	8308	Quilaco	13127	Recoleta
5602	Algarrobo	8311	Santa Bárbara	13130	San Miguel
5603	Cartagena	8312	Tucapel	13202	Pirque
5604	El Quisco	8402	Bulnes	13302	Lampa
5605	El Tabo	8403	Cobquecura	13501	Melipilla
5606	Santo Domingo	8406	Chillán Viejo	13502	Alhué
5703	Llaillay	8407	El Carmen	13505	San Pedro
5704	Panquehue	8408	Ninhue	13601	Talagante
5802	Limache	8412	Portezuelo	13602	El Monte
5803	Olmué	8413	Quillón	13603	Isla de Maipo
6101	Rancagua	8414	Quirihue	13604	Padre Hurtado
6109	Malloa	8415	Ránquil	14102	Corral
6114	Quinta de Tilcoco	8418	San Ignacio	14105	Máfil
6201	Pichilemu	8420	Treguaco	14106	Mariquina
6202	La Estrella	9104	Curarrehue	14202	Futrono
6205	Navidad	9105	Freire	15201	Putre

⁵⁵ Identificador único de la comuna compuesto de código de Región, código de Provincia y Código de comuna. Conocido también como código único territorial.

