



Casen

Observatorio
Social

2017

METODOLOGÍA DE DISEÑO MUESTRAL

(septiembre, 2018)

ÍNDICE

I.	PRESENTACIÓN	8
I.1.	Antecedentes y características generales del diseño Casen 2017.....	9
I.1.1.	Objetivo.....	9
I.1.2.	Población objetivo	10
I.1.3.	Unidad de Información	10
I.1.4.	Marco muestral	10
I.1.5.	Dominios de estudio.....	11
I.1.6.	Objetivos de precisión y tamaño de muestra.....	11
I.1.7.	Estrategia muestral	11
I.1.8.	Método de selección de unidades primarias de muestreo (manzanas y secciones).....	11
I.1.9.	Método de selección de unidades secundarias de muestreo (viviendas)	12
II.	SIMULACIÓN DEL TAMAÑO MUESTRAL	13
II.1.	Objetivo de las simulaciones	13
II.2.	Objetivos específicos de las simulaciones	14
II.3.	Metodología de cálculo del tamaño muestral usando muestras complejas.....	14
II.4.	Resumen del proceso de simulación.....	16
II.4.1.	Primer escenario	19
II.4.2.	Segundo escenario	20
II.4.3.	Tercer escenario.....	21
II.4.4.	Cuarto escenario.....	22
II.4.5.	Quinto escenario	23
II.4.6.	Sexto escenario	24
II.5.	Tamaño muestral propuesto	25
II.5.1.	Distribución de las unidades primarias de muestreo	28
III.	DISEÑO MUESTRAL	33
III.1.	Características del marco muestral	33
III.2.	Cobertura geográfica del marco muestral	38
III.3.	Estratificación del marco muestral	44
III.4.	Estimación del tamaño muestral.....	45
III.5.	Selección de conglomerados de la muestra.....	59
III.5.1.	Selección de manzanas desde el MM2008	59
III.5.2.	Selección de secciones desde el MS2002	61
III.6.	Selección de viviendas de la muestra.....	64
III.6.1.	Verificación de direcciones en los conglomerados de la muestra	64
III.6.2.	Enumeración de los conglomerados de la muestra.....	65
III.6.3.	Selección de las Viviendas	70

IV.	VERIFICACIÓN Y ACTUALIZACIÓN DE UNIDADES DE PRIMERA ETAPA DE LA MUESTRA.....	71
IV.1.	Homologación.....	71
IV.1.1.	Manzanas	72
IV.1.2.	Secciones en Resto de Áreas Urbanas (RAU)	73
IV.1.3.	Secciones Rurales.....	74
IV.2.	Análisis	75
IV.2.1.	Manzanas	76
IV.2.2.	Secciones Resto de Áreas Urbanas (RAU).....	77
IV.2.3.	Secciones Rurales.....	77
IV.3.	Verificación en terreno.....	78
IV.3.1.	Manzanas	78
IV.3.2.	Secciones RAU.....	79
IV.3.3.	Secciones Rurales.....	79
IV.3.4.	Conformación de los listados de viviendas.....	80
IV.4.	Resumen operativo de las actividades de enumeración, verificación y recopilación de información complementaria en terreno.....	80
IV.4.1.	Muestra de manzanas	80
IV.4.2.	Muestra de secciones	82
IV.5.	Conformación definitiva de la muestra	85
V.	DISEÑO FACTORES DE EXPANSIÓN.....	90
V.1.	Introducción.....	90
V.2.	Ponderador de selección de conglomerados	94
V.2.1.	Probabilidad de selección de conglomerados en el Marco de Secciones (MS2002).....	94
V.2.2.	Probabilidad de selección de conglomerados en el Marco de Manzanas (MM2008).....	96
V.2.3.	Ponderador de selección de conglomerados como inverso de la probabilidad de selección.....	98
V.2.4.	Ajuste por omisión de conglomerados	103
V.3.	Ponderador de selección de viviendas	104
V.3.1.	Probabilidad Condicional de selección de viviendas.....	105
V.3.2.	Probabilidad de selección de viviendas como inverso de la probabilidad de selección	106
V.4.	Ponderación por elegibilidad	107
V.4.1.	Ajuste por elegibilidad desconocida.....	108
V.4.2.	Ajuste por no elegibilidad.....	109
V.5.	Ponderación por no respuesta.....	110
V.6.	Ponderación de calibración.....	114
V.6.1.	Ponderador de Calibración Regional.....	116
V.6.2.	Ponderador de Calibración Comunal	119
V.6.3.	Ponderador de Calibración Provincial.....	122

VI. ESTIMACIÓN DE LA VARIANZA COMPLEJA	126
VI.1. Métodos convencionales para estimación de la varianza	126
VI.1.1. Métodos exactos	126
VI.1.2. Método del conglomerado último	127
VI.1.3. Método de linealización	127
VI.1.4. Método de replicación	128
VI.2. Varianza en muestras complejas.....	129
VI.3. Algoritmo de cálculo en muestras complejas.....	131
VI.3.1. Algoritmo de cálculo de los estimadores	132
VI.3.2. Algoritmo de cálculo de las varianzas en muestras complejas.....	133
VI.4. Variables que identifican el diseño muestral complejo en Casen 2017	135
VI.4.1. Creación de pseudo-estratos (varstrat) en Casen 2017.....	136
VI.4.2. Creación de pseudo-conglomerados (varunit) en Casen 2017	137
VI.5. Resultados de la situación de pobreza por ingresos.....	143
ANEXOS.....	148
Anexo N°1. Composición de las nuevas regiones de Biobío y Ñuble	148
Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.....	149
Anexo N°3. Croquis de la manzana (F1).....	166
Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.....	167
Anexo N°5. Ajuste de las estimaciones de población para la encuesta de Caracterización Socioeconómica Nacional Casen 2017	179
Anexo N°6. Comandos y Sintaxis en SPSS y STATA.....	184
BIBLIOGRAFÍA	190

ÍNDICE DE TABLAS

Tabla II.1. Resumen de parámetros utilizados en cada escenario de simulaciones muestrales para Casen 2017.....	19
Tabla II.2. Distribución de los tamaños muestrales de viviendas a nivel nacional y regiones, Casen 2017.....	26
Tabla II.3. Distribución de los tamaños muestrales de viviendas a nivel nacional y regiones, por área urbana y rural, Casen 2017.....	27
Tabla II.4. Errores de estimación para los niveles nacional urbano y nacional rural, Casen 2017.....	27
Tabla II.5. Comparación de los tamaños de muestra objetivo de Casen 2013 y Casen 2017.....	28

Tabla II.6. Distribución de las unidades primarias de muestreo, Casen 2017.....	29
Tabla II.7. Distribución de los tamaños muestrales de viviendas a nivel nacional, regiones y provincias, por área urbana y rural, Casen 2017.....	30
Tabla III.1. Áreas de difícil acceso definidas por el INE.	41
Tabla III.2. Distribución del total de viviendas y secciones excluidas desde el MS2002, según región y área, Casen 2017.....	42
Tabla III.3. Distribución del total de viviendas y manzanas excluidas desde el MM2008, según región, Casen 2017.	43
Tabla III.4. Número de manzanas y viviendas en el MM2008, según grupo de tamaño de manzanas.....	45
Tabla III.5. Tamaños muestrales de viviendas bajo muestreo aleatorio simple y parámetros de cálculo, a nivel nacional y regiones, Casen 2017.....	48
Tabla III.6. Tamaños muestrales de viviendas bajo diseño complejo y parámetros de cálculo, a nivel nacional y regiones, Casen 2017.....	49
Tabla III.7. Distribución de la muestra objetivo de viviendas para las encuestas Casen 2011, 2013, 2015 y 2017.	58
Tabla III.8. Grupos de tamaño de manzanas y promedio de viviendas a encuestar según grupo.60	
Tabla III.9. Número de manzanas seleccionadas del MM2008 y muestra sobredimensionada de viviendas, según región y grupo de tamaño, Casen 2017.....	61
Tabla III.10. Método de selección de unidades primarias secciones, Casen 2017.....	62
Tabla III.11. Número de secciones seleccionadas del MS2002 y muestra sobredimensionada de viviendas, según región y área, Casen 2017.....	63
Tabla IV.1. Distribución de las manzanas gestionadas en terreno, por tipo de trabajo de campo, Casen 2017.....	81
Tabla IV.2. Distribución de las secciones gestionadas, por tipo de trabajo de campo, Casen 2017.	
83	
Tabla IV.3. Distribución de las secciones gestionadas en terreno, por tipo de trabajo de campo, Casen 2017.....	84
Tabla IV.4. Comparación de la distribución inicial y la definitiva del número de manzanas seleccionadas del MM2008, según región y grupo de tamaño, Casen 2017.....	86
Tabla IV.5. Comparación de la distribución inicial y la definitiva de la muestra sobredimensionada de viviendas, según región y grupo de tamaño, Casen 2017.....	87
Tabla IV.6. Comparación de la distribución inicial y la definitiva de secciones seleccionadas del MS2002, según región y área, Casen 2017.....	88
Tabla IV.7. Comparación de la distribución inicial y la definitiva de la muestra sobredimensionada de viviendas, según región y área, Casen 2017.	89

Tabla V.1. Distribución de frecuencias regionales de manzanas y secciones en el Marco Muestral y estimadas en Casen 2017 con ponderador de selección base.....	100
Tabla V.2. Estadísticos de ponderadores de UPM, manzanas y secciones, según MM2008 y MS2002, Casen 2017.....	101
Tabla V.3. Total de viviendas según marco de muestreo, con y sin exclusión de conglomerados.	103
Tabla V.4. Distribución de viviendas seleccionadas por el INE en Casen 2017 por elegibilidad en el trabajo de campo.....	110
Tabla V.5. Ponderador de no respuesta por celda de ajuste.....	113
Tabla V.6. Estimaciones poblacionales basadas en el Censo 2002, al 30 de noviembre de 2017.	115
Tabla V.7. Aplicación de criterio para factibilidad de cálculo de factor de expansión provincial.	122
Tabla VI.1. Total de estratos, conglomerados, varstrat y varunit, según región por área geográfica, para la muestra lograda de Casen 2017.....	138
Tabla VI. 2. Frecuencia de varunit según número de viviendas que lo componen, Casen 2017..	140
Tabla VI. 3. Distribución del número de conglomerados que componen los varunit, Casen 2017.	141
Tabla VI.4. Distribución del número de varunit que componen los varstrat, Casen 2017.....	142
Tabla VI.5. Estimación de porcentaje de la población en situación de pobreza por ingresos, según zona y región, Casen 2017.....	144
Tabla VI.6. Estimación de porcentaje de la hogares en situación de pobreza por ingresos, según zona y región, Casen 2017.....	146

ÍNDICE DE CUADROS

Cuadro II.1. Parámetros comunes utilizados en todos los escenarios de simulaciones muestrales para Casen 2017.....	18
--	----

ÍNDICE DE GRÁFICOS

Gráfico III.1. Distribución de promedios ponderados de las tasas de no respuesta comunales de Casen 2013 y 2015, según región.....	56
Gráfico III.2. Distribución porcentual de la muestra objetivo de viviendas a nivel regional, para las encuestas Casen 2011, 2013 y 2017.....	57

Gráfico V.1. Distribuciones regionales de las probabilidades de selección de las secciones, por Región, Casen 2017.	96
Gráfico V.2. Distribuciones regionales de las probabilidades de selección de las manzanas, por Región, Casen 2017.	98
Gráfico V.3. Distribuciones regionales del ponderador de selección de secciones en Casen 2017. 102	
Gráfico V.4. Distribuciones regionales del ponderador de selección de manzanas en Casen 2017. 102	
Gráfico V.5. Factor de expansión calibración Región–Área (RAU-Rural).....	118
Gráfico V.6. Factor de expansión calibración Región–Área (Urbano).....	119
Gráfico V.7. Distribución regional del Ponderador de Calibración Comunal de secciones (RAU-Rural).....	121
Gráfico V.8. Distribución regional del Ponderador de Calibración Comunal de manzanas (Urbano).....	121
Gráfico V.9. Distribución regional del Ponderador de Calibración Provincial de secciones (RAU-Rural).....	124
Gráfico V.10. Distribución regional del Ponderador de Calibración Provincial de manzanas (Urbano).....	125

ÍNDICE DE FIGURAS

Figura III.1. Estratificación e identificación de unidades primarias de muestreo	38
Figura III.2. Formulario de enumeración en áreas urbanas, Casen 2017 (Formato impreso).....	68
Figura III.3. Formulario de enumeración en áreas rurales, Casen 2017 (Dispositivo móvil).....	69
Figura IV.1. Etapas del proceso de generación de los listados de viviendas particulares ocupadas. 71	
Figura IV.2. Comparativa entre actualización de definiciones marco muestral vigente y Precenso 2016.....	72
Figura IV.3. Resumen actividades de verificación de información de Precenso para zonas urbanas. 73	
Figura IV.4. Resumen actividades de verificación de información de Precenso para zonas rurales. 75	
Figura V.1. Diagrama de procesos de desarrollo de factor de expansión Casen 2017.....	93

I. PRESENTACIÓN¹

Las encuestas de hogares tienen un rol fundamental en la toma de decisiones para la formulación de políticas sociales, de ahí el creciente interés de obtener estimaciones precisas y confiables que puedan representar de manera adecuada a la población de interés. Sin embargo, para lograr esta propiedad deseable en los estimadores, es necesario tener una muestra que responda a las exigencias de los errores muestrales propuestos. En consecuencia, la muestra debe ser construida para los objetivos propios de la encuesta, requerimientos que se deben considerar al momento de determinar un tamaño muestral. Además, el tamaño muestral debe tener un compromiso entre el presupuesto disponible y los errores de muestreo admisibles, logrando conjugar estos tres elementos (de requerimientos, errores muestrales admisibles y presupuesto disponible) se pueden obtener tamaños muestrales óptimos².

Al igual que en las versiones anteriores, la encuesta Casen 2017 fue planificada por el Ministerio de Desarrollo Social (MDS) quien, en diciembre de 2016, mediante un convenio de transferencia, traspasó la elaboración del diseño muestral al Instituto Nacional de Estadísticas (INE).

En ese sentido, mediante el convenio de colaboración interinstitucional entre el MDS y el INE, denominado: Diseño Muestral Encuesta de Caracterización Socioeconómica Nacional Casen 2017, se busca dar continuidad a una nueva versión de la Encuesta Casen, para el año 2017, a través de la cual, se obtenga información relevante y actualizada, para la implementación de políticas públicas, sobre la materia y para uso de la comunidad nacional en general. Este proceso que comienza con diversas simulaciones de tamaño para determinar el número de unidades óptimo a encuestar de acuerdo a los requerimientos y objetivos de la contraparte, para posteriormente seleccionar las unidades muestrales, y elaborar y preparar el material cartográfico que se proporcionará a la institución que se encargará de realizar el levantamiento de la información.

¹ Este documento fue elaborado por el Instituto Nacional de Estadísticas (INE) y editado por el Ministerio de Desarrollo Social. La edición realizada por el Ministerio de Desarrollo Social corresponde a la eliminación de las notas al pie que hacen referencia a: "Manual análisis enumeración Casen 2017", "Manual análisis enumeración rural Casen 2017", "Manual de verificación y enumeración de viviendas en áreas urbanas Casen 2017" y "Manual de verificación y enumeración en áreas rurales y RAU Casen 2017".

² Cabe mencionar que hablar de tamaños muestrales óptimos no denota un proceso matemático de maximización/minimización de una función objetivo, sino el proceso de las múltiples iteraciones para las distintas simulaciones, distintos parámetros y requerimientos. Así como también los acuerdos llegados con la contraparte en las reuniones de trabajo sostenidas.

Además, el INE elabora los factores de expansión, la estimación de los errores de muestreo bajo un diseño de muestreo complejo y este informe metodológico del diseño muestral.

La metodología del diseño muestral de la Encuesta de Caracterización Socioeconómica Nacional Casen 2017, 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.

I.1. Antecedentes y características generales del diseño Casen 2017

I.1.1. Objetivo

La encuesta Casen 2017, al igual que sus versiones anteriores, ha sido pensada como un instrumento de diagnóstico, evaluación y de focalización que permita:

- 1.** 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.
- 2.** Estimar la magnitud de la pobreza y la distribución del ingreso.
- 3.** Identificar carencias y demandas de la población en las áreas señaladas.
- 4.** Evaluar las distintas brechas que separan a los diferentes segmentos sociales y ámbitos territoriales.
- 5.** Evaluar el impacto de la política social, en cuanto a la cobertura, focalización y distribución del gasto fiscal de los principales programas sociales de alcance nacional entre los hogares, según su nivel de ingreso. Así como también, el impacto de este gasto en el ingreso de los hogares y en la distribución del mismo.

I.1.2. Población objetivo

La población objetivo de la encuesta es la población que reside en viviendas particulares a lo largo del territorio nacional, excluyendo las áreas de difícil acceso definidas por el INE. Al interior de cada vivienda seleccionada, se entrevista a todos los hogares y a las personas que se declaran residentes habituales.

I.1.3. Unidad de Información

El informante idóneo en Casen 2017 es el jefe de hogar o en su ausencia, algún miembro del hogar de 18 años o más.

I.1.4. Marco muestral

El INE mantiene vigente al año 2017 un marco muestral conformado por dos marcos independientes, el marco de manzanas para el área Urbana y el marco de secciones para las áreas Rural y Resto de Área Urbana (RAU). Las unidades que lo componen se denominan conglomerados, que corresponden a conjuntos de viviendas agrupadas y contenidas en áreas geográficas definidas por límites de calles, pasajes, alturas de calles o aglomeraciones de viviendas particulares conformadas a partir de una o más entidades pobladas.

A través del tiempo, estos conglomerados han sido denominados segmentos (secciones y manzanas) y corresponden a las unidades primarias de muestreo o UPM, obedeciendo a distintos criterios de agrupación, tanto en límites como en tamaño. A través de estos conglomerados se accede a la vivienda y al hogar, cuyos integrantes son las personas, generalmente el objeto final de estudio. A petición del Ministerio de Desarrollo Social, para la encuesta Casen 2015 el INE realizó una actualización parcial menor y focalizado no superior a 2% de viviendas del marco de manzanas, la cual se mantiene vigente para la nueva versión de la encuesta Casen 2017.

Para la selección de viviendas o unidades secundarias de muestreo (USM), se considerará la actualización de los conglomerados seleccionados en primera etapa, con base en la información obtenida del Precenso desarrollado durante el año 2016, procedimiento que se describe en el capítulo III, sección 6.

I.1.5. Dominios de estudio

Para la versión 2017, los niveles de estimación o dominios de estudio para los cuales la muestra será representativa serán: nacional, nacional urbano, nacional rural y regional, los cuales coinciden con los dominios de estudio definidos en las versiones de Casen 2011 y 2013. Así mismo, continúa habiendo presencia en las mismas 324 comunas que en las versiones 2011 y 2013 de la encuesta, tanto en áreas urbanas como rurales.

I.1.6. Objetivos de precisión y tamaño de muestra

Los criterios de precisión se mantuvieron iguales a los considerados en Casen 2013, es decir, los errores absolutos nacionales y regionales no deben superar 4 puntos porcentuales y los errores relativos nacionales y regionales no deben superar 30 puntos porcentuales. Además, se estableció un tamaño mínimo de muestra objetivo para el área urbana de 50 viviendas por comuna y para el área rural de 30 viviendas por comuna.

I.1.7. Estrategia muestral

En Casen 2017 el diseño corresponde a la obtención de una muestra probabilística, estratificada y bietápica, siendo los estratos conformados por la dupla Comuna-Área³. Adicionalmente, en el área urbana, proveniente del MM2008, las manzanas se clasificaron según su tamaño (en número de viviendas), creando 5 estratos de tamaño, en los cuales la muestra fue seleccionada de forma independiente, quedando los estratos del área urbana conformados por la terna Comuna-Área-Grupo de tamaño.

I.1.8. Método de selección de unidades primarias de muestreo (manzanas y secciones)

Para Casen 2017, en el marco de manzanas, se realizó una selección sistemática y con igual probabilidad, a nivel de Comuna-Área-Grupo de tamaño, asegurando que cada uno de los grupos de tamaño de las comunas estén representados en la muestra, y con ello tener una fórmula explícita de sus probabilidades de selección. En el marco de secciones, se realizó una selección

³ Las áreas clasificadas como Urbana (Ciudad, Resto de Área Urbana, Urbana) y Rural.

sistemática con probabilidad proporcional al tamaño en cuanto al número de viviendas, a nivel de Comuna-Área.

I.1.9. Método de selección de unidades secundarias de muestreo (viviendas)

Para Casen 2017, ya sea en el marco de manzanas o secciones, se realizó una selección sistemática y con igual probabilidad, al interior de cada uno de los conglomerados seleccionados previamente (manzanas o secciones).

II. SIMULACIÓN DEL TAMAÑO MUESTRAL

Durante el año 2016, el INE tuvo la misión de desarrollar la estrategia de muestreo de la Encuesta Casen 2017, que consistió en elaborar y desarrollar, de acuerdo a los requerimientos planteados por el MDS, simulaciones con términos específicos de tamaño y distribución muestral, optimizando la selección de las unidades muestrales, de acuerdo a criterios de factibilidad y de presupuesto.

Durante esta etapa previa al desarrollo de la estrategia de levantamiento de la Encuesta Casen 2017, se estableció como hito inicial del convenio INE-MDS, encontrar un tamaño muestral mínimo óptimo para la estimación, con márgenes de precisión tolerables con base en los requerimientos, para que la Encuesta Casen 2017 tenga representatividad a nivel nacional, nacional urbano y rural, y regional. Para ello, se simularon distintos escenarios o alternativas de tamaño muestral basados en la unificación de criterios estadísticos solicitados y discutidos con el Ministerio de Desarrollo Social.

Estas simulaciones se realizaron considerando los parámetros de los resultados del trabajo de campo de Casen 2015 y la distribución de la muestra objetivo de Casen 2013, resultando en la formulación de diversos escenarios, los que fueron presentados en reuniones de trabajo conformadas por los equipos de la División del Observatorio Social de la Subsecretaría de Evaluación Social del Ministerio de Desarrollo Social y el Departamento de Investigación y Desarrollo del INE (con apoyo y coordinación de la Subdirección Técnica). Finalmente, se estableció un escenario que se ajusta óptimamente a los requerimientos establecidos por el Ministerio.

En este capítulo se reseñan los aspectos más relevantes de las simulaciones realizadas hasta obtener el tamaño de muestra definitivo. Sin embargo, es importante mencionar que los detalles metodológicos asociados al cálculo del tamaño muestral y su distribución se especifican más adelante en el capítulo III, denominado Diseño Muestral.

II.1. Objetivo de las simulaciones

Encontrar un tamaño muestral que reúna un conjunto de características que puedan representar de manera adecuada a la población chilena, respecto de sus componentes sociodemográficos con

márgenes de error tolerables a nivel nacional, nacional urbano y rural, y regional; considerando distintas estrategias de muestreo, errores de muestreo admisibles, distribución de la muestra objetivo y con sobredimensionamiento.

II.2. Objetivos específicos de las simulaciones

Calcular el tamaño muestral considerando los siguientes elementos:

1. Niveles de estimación:
 - a. Nacional.
 - b. Nacional, urbano y rural.
 - c. Regional.
2. Variable de diseño: Tasa de pobreza (porcentaje de personas en situación de pobreza por ingresos) estimada según resultados obtenidos de Casen 2015.
3. Medida poblacional o parámetro a estimar: Tasa de pobreza (porcentaje de personas en situación de pobreza por ingresos).
4. Errores de muestreo admisibles prefijados por el estudio.
5. Parámetros de los resultados del trabajo de campo de Casen 2015.
6. Distribución de la muestra objetivo de la encuesta Casen 2013.
7. Tasas de no respuesta comunales observadas en la aplicación de las encuestas Casen 2013 y 2015.

II.3. Metodología de cálculo del tamaño muestral usando muestras complejas

La población objetivo, los niveles de estimación deseados y la cobertura geográfica de la encuesta Casen, hacen necesario la confección de un diseño eficiente para la selección de una muestra de viviendas, en términos de ejecución presupuestaria e implementación operativa. El muestreo aleatorio simple de viviendas necesitaría, en principio, de un marco de viviendas con las direcciones de domicilios generado por un censo de población y viviendas, el cual resultaría muy

costoso de mantener actualizado, y por otra parte, las viviendas seleccionadas de un marco de este tipo, se encontrarían tan dispersas geográficamente, que los costos asociados al levantamiento excederían la asignación presupuestaria, por tanto, el muestreo aleatorio simple iría en desmedro de la eficiencia deseada.

Por otra parte, en los diseños de muestra complejos las unidades que finalmente componen la muestra (en este caso viviendas), se determinan después de varias etapas, en consecuencia, no es necesario mantener un marco de viviendas actualizado, lo cual reduce significativamente los costos.

La formación de conglomerados⁴ hace posible agregar nuevas etapas de selección⁵, pasando de un diseño monoetápico a uno polietápico. Adicionalmente, al conformar estratos que permitan incrementar la precisión estadística de los estimadores, se genera la estructura de un diseño de muestra complejo. Para la encuesta Casen, los conglomerados son conjuntos de viviendas que pueden ser manzanas o secciones⁶, lo que permite que la muestra definitiva de viviendas (seleccionada en una segunda etapa) quede lo suficientemente concentrada facilitando el levantamiento y control, dentro de un presupuesto limitado, asegurando niveles razonables de precisión de los estimadores.

Se puede afirmar que cualquier diseño muestral diferente al irrestricto aleatorio o muestreo aleatorio simple, puede ser considerado como un diseño complejo. En este caso, hay varios elementos del diseño que se aplican en la Casen que lo complejizan, específicamente, por el hecho que sea estratificado, polietápico y por conglomerados. Mientras unos elementos tienden a reducir la varianza como la estratificación, otros tienden a aumentarla, como la conglomeración. Las distintas etapas tienden a aumentar la varianza mientras que la disminución del tamaño de los conglomerados tiende a disminuirla al igual que el aumento en la cantidad de ellos. También es necesario considerar así otros efectos tales como la distribución de la prevalencia de la pobreza en las unidades secundarias como la correlación de las unidades secundarias al interior de los conglomerados respecto a la característica de interés, llamada correlación intraclásica, que afecta la varianza y es el principal factor en la medición del efecto del diseño o el efecto provocado por considerar un diseño aleatorio no-simple.

⁴ Conjuntos de unidades muestrales.

⁵ El método de selección en la etapa i es independiente al de la etapa j , pudiendo realizarse con probabilidades iguales o no, perdiendo la condición de una muestra autoponderada en este último caso.

⁶ Ver capítulo III, sección 1, Características del Marco Muestral.

El cociente entre la estimación de la varianza bajo un diseño complejo versus la estimación bajo un diseño aleatorio simple se denomina efecto del diseño, término introducido por Kish (1965), y en términos de tamaño muestral equivale a cuánto se debe aumentar la muestra bajo diseño complejo para alcanzar el mismo nivel de precisión que lo que se obtendría si aplicáramos un muestreo aleatorio simple.

Entonces para determinar el tamaño de la muestra compleja de Casen, se incorpora este factor de ajuste a partir del tamaño de una muestra aleatoria simple, lo que permite aproximarse al número de viviendas necesarias para que el diseño complejo proporcione la misma varianza⁷.

En este contexto, la metodología de cálculo del tamaño muestral utilizada en Casen 2011, 2013, 2015 y en esta nueva versión 2017 es básicamente la misma. Primero, se inicia con la determinación de un tamaño muestral bajo un muestreo aleatorio simple, fijando previamente el error absoluto a alcanzar con base en los valores de encuestas anteriores. Posteriormente, se realizan tres ajustes secuenciales, primero por el efecto del diseño (obtenido de una o como el promedio de dos o más encuestas), el segundo corresponde a un ajuste por finitud, y el tercero se refiere al ajuste por la tasa de no respuesta, que también puede corresponder a un levantamiento o el promedio de más de un levantamiento.

Para la encuesta Casen 2017, se obtuvieron tamaños de muestra regionales, asegurando representatividad a nivel nacional, nacional urbano y rural, y regional. En este proceso, se utilizó el efecto del diseño obtenido del trabajo de campo de Casen 2015 y la tasa de no respuesta se estimó como un promedio ponderado de las tasas de no respuesta comunales de Casen 2013 y 2015.

Posteriormente, operó un procedimiento mediante el cual se distribuyó el tamaño de muestra de cada región, por comuna y área, donde se buscó resguardar coherencia entre los tamaños muestrales y aquéllos de la población, en términos de proporcionalidad.

II.4. Resumen del proceso de simulación

Como se mencionaba anteriormente, la muestra ha sido construida para el logro de los objetivos propios de la encuesta, considerando factores como la variable de interés, el estimador asociado a esa variable, los niveles de estimación, los errores de muestreo tolerables y la disposición de una

⁷ Kish, 1965, 1979, 1987.

fuentes de información para obtener las estimaciones de interés (censo o encuestas anteriores). En ese sentido, se realizó un proceso iterativo donde fueron consideradas distintas estrategias de muestreo, que dieron origen a una variedad de escenarios sobre los cuales se probaron, evaluaron y ajustaron los parámetros hasta obtener un tamaño muestral adecuado a los requerimientos.

En total fueron analizados seis escenarios, el primero de ellos corresponde a la simulación presentada en julio de 2016, la cual se realizó con base en los parámetros obtenidos del trabajo de campo de Casen 2013, y cuyo tamaño muestral sirvió como referencia para la definición de presupuesto y punto de partida para el diseño del segundo escenario, que incorporó la actualización de los cálculos a nivel regional presentados en la primera simulación, con base en los resultados del trabajo de campo de Casen 2015. En este segundo escenario, también se realizó el supuesto de la incorporación de una nueva región de Ñuble, realizando también un ajuste en el tamaño de muestra en la nueva región de Biobío, ya que la región de Ñuble nace de una provincia completa de esta región (ver Anexo N°1 con las comunas que conforman estas nuevas regiones).

El tercer escenario se originó por un ajuste en el efecto del diseño, redondeado a un decimal y dentro de un rango de 1 a 5. Además, se estimó la muestra regional con sobredimensionamiento como una agregación de la sobremuestra comunal.

Dado que los dominios de estudio definidos para Casen 2017 coinciden con los de Casen 2013, se valoró aproximar estos diseños, en consecuencia, el cuarto escenario considera la distribución de la muestra objetivo comunal de Casen 2013 para distribuir la muestra objetivo comunal de Casen 2017, la cual, en el escenario anterior había sido distribuida según la muestra lograda en esa encuesta. Adicionalmente, se obtiene la distribución comunal a nivel urbano y rural, tanto para la muestra objetivo como para la muestra con sobredimensionamiento.

El quinto escenario involucró una modificación de los errores absolutos de las regiones de Atacama y Metropolitana, de tal forma que los tamaños de la muestra objetivo de estas regiones se aproximen a los tamaños de la muestra objetivo de Casen 2013. Además, se consideró modificar los tamaños poblacionales según el Marco 2008 actualizado foco Casen para el área urbana, pasando de un marco de 4.659.944 viviendas a 4.723.693 viviendas.

Finalmente, el sexto escenario es producto del análisis de los tamaños muestrales y errores asociados a nivel provincial, a raíz del cual se incrementó el tamaño de muestra para cinco

provincias, logrando un error absoluto máximo de 6,5% a ese nivel y pasando de una muestra objetivo nacional de 69.261 a 69.816 viviendas.

El siguiente cuadro, presenta los parámetros comunes utilizados en todos los escenarios considerados.

Cuadro II.1. Parámetros comunes utilizados en todos los escenarios de simulaciones muestrales para Casen 2017.

Parámetro	Descripción
Variable de diseño	Tasa de pobreza por ingreso (porcentaje de personas en situación de pobreza por ingresos) estimada según resultados obtenidos de Casen 2015.
Estimador asociado	Estimador de razón: $r = \frac{\text{Número de personas pobres}}{\text{Número total de personas}}$
Parámetro asociado	Tasa de pobreza por ingresos
Niveles de estimación	Nacional Nacional, urbano y rural Regional
Errores de muestreo	Los errores absolutos nacionales y regionales no deben superar 4% y los errores relativos nacionales y regionales no deben superar 30%.
Muestra utilizada para el cálculo del efecto del diseño y Cuasi-varianza	Casen 2015 (Excepto el escenario N°1 que se utilizó Casen 2013)

Elaborado por el Instituto Nacional de Estadísticas.

La Tabla II.1, presenta un resumen de todos los parámetros utilizados en cada una de las simulaciones, enumeradas del 1 al 6, por orden cronológico.

Tabla II.1. Resumen de parámetros utilizados en cada escenario de simulaciones muestrales para Casen 2017.

Nro.	Regiones	Viviendas en el Marco	Error Absoluto Estimado (Nacional)	Error Relativo Estimado (Nacional)	Efecto Diseño	Tasa de no Respuesta ⁸	Tamaño Muestral Objetivo (País)	Tamaño Muestral con Sobremuestreo (País)
1	15	4.659.944	0,6%	4,0%	Casen 2013	25,4%	70.359	94.314
2	15	4.659.944	0,5%	4,0%	Casen 2015	23,7%	68.886	90.276
	16	4.659.944	0,5%	4,0%	Casen 2015	26,0%	68.885	90.256
3	15	4.659.944	0,5%	3,9%	Casen 2015 Ajustado	23,9%	69.890	91.781
	16	4.659.944	0,5%	3,9%	Casen 2015 Ajustado	23,9%	69.868	91.752
4	16	4.659.944	0,5%	3,9%	Casen 2015 Ajustado	26,0%	69.868	94.444
5	16	4.723.693	0,5%	4,0%	Casen 2015 Ajustado	26,7%	69.261	94.478
6	16	4.723.693	0,5%	4,0%	Casen 2015 Ajustado	26,6%	69.816	95.159

Elaborado por el Instituto Nacional de Estadísticas.

II.4.1. Primer escenario

Como ya se ha mencionado antes, los niveles de estimación o dominios de estudio para los cuales la muestra Casen 2017 es representativa son: nacional, nacional-urbano, nacional-rural y regional, lo que prevé que el diseño muestral de esta nueva versión de la encuesta se aproxime al diseño planteado para Casen 2013.

En consecuencia, se utilizaron los tamaños de la muestra objetivo regional de Casen 2013 como punto de partida para definir los tamaños de la muestra objetivo regional de Casen 2017, para los cuales se estimaron los errores con base en el trabajo de campo de Casen 2013.

Una vez obtenidos los errores a nivel regional se evaluó que estuvieran dentro de los rangos definidos, es decir, los errores absolutos menores a 4% y los relativos menores a 30%, encontrando

⁸ Dependiendo del escenario, el cálculo de la sobremuestra regional se obtuvo de tres formas: La primera, como agregación de la sobremuestra comunal (escenarios N°1 y 3); la segunda, a través de la estimación de tasas de no respuesta regionales (escenario N°2); y la tercera, como agregación de la sobremuestra comunal según área (escenarios N°4, 5 y 6); lo que implicó el cambio en la estimación de las tasas de no respuesta regionales, y en consecuencia, la nacional.

que los errores relativos de las regiones de Atacama y Magallanes sobrepasaban considerablemente la cota de 30%.

En el caso de Atacama se aumentó el tamaño de muestra, logrando disminuir el error relativo a 33,5%, y, por otra parte, en la región de Magallanes se hizo la excepción, ya que siendo la tasa de pobreza 5,6% (obtenido de Casen 2013) implicaría un aumento significativo del tamaño de muestra.

Para determinar el tamaño de muestra con sobredimensionamiento, inicialmente se obtuvo la sobremuestra a nivel comunal, la cual se estimó en función del promedio simple de las tasas de no respuesta de Casen 2013 y Casen 2015. Posteriormente, se calculó la sobremuestra a nivel regional y nacional por agregación.

II.4.2. Segundo escenario

El segundo escenario surge de actualizar los cálculos anteriores en función de los resultados obtenidos del trabajo de campo de Casen 2015, para ello se tomaron los tamaños muestrales definidos a nivel regional en el primer escenario y se estimaron los errores absolutos y relativos con base en los parámetros de Casen 2015.

Los errores estimados fueron utilizados para obtener un primer tamaño muestral a nivel regional (15 y 16 regiones), asumiendo muestreo aleatorio simple, el cual fue corregido en forma secuencial por los tres factores mencionados anteriormente, es decir, el efecto del diseño obtenido de Casen 2015, la finitud de la población, y la estimación de la tasa de no respuesta.

Si bien se estableció que los errores relativos regionales no deben superar 30%, el caso de la región de Magallanes se aceptó como un caso de excepción a esta regla, siendo la tasa de pobreza de esta región 4,4%, se mantuvo un error relativo de 32,6%, ya que una disminución de éste implica un aumento considerable en el tamaño de muestra.

En cuanto al ajuste dado por el efecto del diseño, destacan los tres siguientes casos:

- 1.** En la Región de Atacama el efecto del diseño es 0,48, lo cual implica disminuir la muestra obtenida bajo muestreo aleatorio simple a poco menos de la mitad.
- 2.** En la Región de Magallanes, el efecto del diseño es 0,86, lo cual también implica disminuir la muestra, en este caso un 14%.

3. Y, en la Región Metropolitana la muestra es corregida por un efecto del diseño de 5,76, aumentando la muestra bajo M.A.S. en más de 5 veces.

El análisis de los casos anteriores hizo notar la necesidad de establecer cotas mínima y máxima para el efecto del diseño. En consecuencia, se estableció un mínimo de 1,0, basados en el supuesto de que, para un mismo tamaño de muestra, una muestra polietápica tiene un mayor error de muestreo que bajo muestreo aleatorio simple. Este supuesto atribuye que los efectos del diseño menores que 1,0, obtenidos para Atacama y Magallanes (0,48 y 0,86 respectivamente) responden a factores particulares⁹ del diseño y la muestra recolectada en Casen 2015.

Por otra parte, se estableció un máximo de 5,0, ya que la evidencia empírica en estudios de este tipo apunta a que la ganancia en la precisión, producto de corregir el tamaño de muestra por un efecto del diseño mayor a 5, es marginal respecto al incremento de los costos de levantamiento del nuevo tamaño de muestra.

II.4.3. Tercer escenario

El tercer escenario fue producto del ajuste en el efecto del diseño, el cual se redondeó a un decimal y se acotó entre un mínimo de 1,0 y un máximo de 5,0, obteniendo así, nuevos tamaños de muestra a nivel regional.

Una vez obtenidos los tamaños de muestra regionales, se realizó la primera distribución de la muestra objetivo por comuna, la cual se basó en la distribución comunal de la muestra lograda de Casen 2013. Posteriormente, se calculó la muestra con sobredimensionamiento a nivel comunal y por agregación se obtuvo a nivel regional y nacional.

Para el cálculo de la sobremuestra a nivel comunal, se estimó una tasa de no respuesta de dos maneras, la primera, utilizando el promedio ponderado (por el número de viviendas elegibles) de las tasas de no respuesta comunales 2013 y 2015, y la segunda, utilizando su promedio simple. Finalmente, se estableció calcular la sobremuestra a través de la tasa de no respuesta ponderada por el número de viviendas elegibles, obteniendo la sobremuestra regional como agregación de los tamaños comunales, modificando así, la estimación de las tasas de no respuesta regionales, y en consecuencia, la nacional.

⁹ Dado que la inferencia se realiza sobre una muestra particular de un gran número de muestras posibles.

II.4.4. Cuarto escenario

El cuarto escenario se originó por una modificación en la distribución del tamaño de muestra comunal, el cual se ajustó de forma proporcional a la muestra objetivo (en vez de la muestra lograda) comunal Casen 2013, en consecuencia, se modificó la sobremuestra comunal, regional y nacional.

Cabe señalar que también fue considerado realizar la distribución comunal proporcional al número de viviendas en el marco, sin embargo, la alta concentración de viviendas en algunas comunas respecto a otras, origina tamaños de muestra muy pequeños. Esta situación fue analizada y corregida en el diseño de Casen 2013, razón por la cual fue utilizada dicha distribución para Casen 2017.

Una vez obtenidos los tamaños de muestra comunales, se distribuyó la muestra objetivo urbana y rural en función de la proporción de éstas en Casen 2013 en cada comuna. Además, se estableció un tamaño mínimo de muestra objetivo para el área urbana de 50 viviendas por comuna y para el área rural de 30 viviendas por comuna.

Luego, se obtuvo el tamaño de muestra sobredimensionado tanto para el área urbana como para el área rural comunal, aplicando en cada área la tasa de no respuesta estimada por el promedio ponderado (por el número de viviendas elegibles) de las tasas de no respuesta comunales 2013 y 2015. Lo cual originó un cambio tanto en los tamaños de sobremuestra regionales como en la estimación de las tasas de no respuesta a ese nivel.

Para la distribución resultante se realizó una comparación con el diseño muestral de Casen 2013, ya que, como se ha mencionado antes, se esperaba que ambos diseños se aproximaran, al contar con los mismos dominios de representatividad. Finalmente, esto resultó en tamaños de muestra objetivo para las regiones de Atacama y Metropolitana considerablemente distantes de los tamaños definidos para Casen 2013.

Las diferencias presentes en estas regiones se debieron a que los efectos de diseño obtenidos en Casen 2015, fueron ajustados al rango entre 1 y 5. Esto afectó principalmente a la región de Atacama, donde el efecto del diseño fue ajustado de 0,48 a 1, provocando un aumento a más del doble de la muestra inicial, pasando de 2.350 en Casen 2013 a 4.657 viviendas. Mientras que, en la Región Metropolitana, el efecto del diseño fue ajustado de 5,76 a 5, lo que provocó una disminución del tamaño muestral inicial de 12.865 en Casen 2013 a 11.100 viviendas.

II.4.5. Quinto escenario

En el quinto escenario, se buscó una mejor aproximación entre la distribución obtenida hasta el escenario anterior y el diseño de Casen 2013. Si bien los ajustes en el efecto del diseño han afectado la similitud entre las distribuciones de las muestras regionales 2013 y 2017, se consideró modificar los errores absolutos de Atacama y Región Metropolitana.

En este sentido, para la Región de Atacama el error absoluto propuesto inicialmente, aumentaría de 1,5% a 2,1%, obteniendo un tamaño de muestra objetivo de 2.331 viviendas. Por su parte, para la Región Metropolitana el error absoluto inicialmente propuesto, disminuiría de 0,90% a 0,83%, obteniendo un tamaño de muestra objetivo de 12.766 viviendas.

En este escenario, también se incluye la actualización de los tamaños poblacionales según Marco 2008 actualizado foco Casen (marco Casen 2015) para el área urbana, pasando de un marco de 4.659.944 viviendas a 4.723.693 viviendas. Lo cual modificó los tamaños de muestra ajustados por finitud para cada región.

Adicionalmente, en este escenario se analizaron en detalle los tamaños muestrales a nivel provincial, en el cual fueron consideradas aquellas provincias respecto de las cuales se espera contar con estimaciones, cuyos errores absolutos no superen 6,5 puntos porcentuales. Estas provincias son aquellas que cumplen con los criterios¹⁰ que sustentan la posibilidad de disponer de factores de expansión provinciales, los cuales fueron definidos e implementados para Casen 2015.

Encontrándose que, en 11 de las 53 provincias es superado el límite de 5% de error absoluto, siendo mayor en Parinacota (12,4%)¹¹, seguido por Cauquenes y Chacabuco con 8,3% y 8,1% respectivamente.

¹⁰ Los factores de expansión provinciales serán calculados para todas las provincias que no sean Áreas de Difícil Acceso (ADA) y que tengan todas las comunas presentes en el marco muestral del INE o, para aquellas provincias cuyas comunas faltantes no representen más del 10% de la población de personas total de la provincia, según estimaciones poblacionales utilizadas para el mismo cálculo del factor.

¹¹ Si bien el error absoluto de la provincia de Parinacota alcanzará previsiblemente 12,4 puntos porcentuales, no se considera en el análisis, dado que esta provincia no cumple con los criterios que sustentan la posibilidad de disponer de factores de expansión provinciales.

Por otra parte, en 17 de las 53 provincias es superado el límite de 35% de error relativo, siendo mayor en Tierra del Fuego (256,4%)¹², lo cual se justifica por la baja tasa de pobreza de 1,1%. También destaca Chacabuco con un error relativo de 70,4%. Un caso particular es el observado en la región de Aysén, donde sus 4 provincias superan este límite, siendo a nivel regional de 28,8%.

Por último, cabe señalar que en 5 provincias se superan ambos umbrales, siendo éstas: Tamarugal, Tocopilla, San Antonio, Chacabuco y Parinacota, pertenecientes a las regiones I, II, V, XIII y XV, respectivamente.

II.4.6. Sexto escenario

El sexto escenario, considerado el definitivo, es producto del análisis de los tamaños muestrales y errores asociados a nivel provincial, a partir del cual se ajustó la distribución muestral de las provincias Tamarugal, Tocopilla, Cauquenes, Chacabuco y Punilla, pertenecientes a las regiones I, II, VII, XIII y nueva región XVI, respectivamente.

El ajuste consistió en aumentar el tamaño de muestra para cada una de estas provincias, de tal manera que el error absoluto alcanzara un máximo de 6,5% a este nivel de estimación, con la excepción de la provincia de Parinacota, cuyo error máximo se mantuvo en 12,4%.

Dentro de cada una de estas cinco provincias, se distribuyó el nuevo tamaño de muestra comunal conforme a la distribución anterior, es decir, la muestra objetivo comunal de Casen 2013.

A nivel nacional, el tamaño de muestra objetivo se incrementó de 69.261 a 69.816 viviendas y con sobredimensionamiento, pasó de 94.478 a 95.159 viviendas.

¹² En el caso de Tierra del Fuego se estima un posible error relativo de 256,4%, sin embargo, esta provincia no cumple los criterios mencionados en la nota 13, por esta razón no se dispondrá de factores de expansión para esta provincia.

II.5. Tamaño muestral propuesto

Como se ha mencionado antes, las simulaciones se realizaron asegurando representatividad a nivel nacional, nacional urbano y rural, y regional, procurando, que los errores absolutos a nivel provincial se mantuvieran dentro de un rango razonable.

Se estimó un tamaño de muestra tanto para las 15 regiones que conforman la división política administrativa actual del territorio nacional, así como, bajo el supuesto de la nueva región XVI Ñuble. En ese sentido, el tamaño de muestra obtenido en el sexto escenario, asegura dicho nivel de representatividad tanto para 15 como para 16 regiones. A continuación, se presentan los parámetros de cálculo, errores estimados y distribución de la muestra propuesta para Casen 2017.

Tabla II.2. Distribución de los tamaños muestrales de viviendas a nivel nacional y regiones, Casen 2017.

Región	Población Viviendas en el Marco	Muestra Viviendas Efectivas Casen 2015	Tasa de Pobreza Casen 2015	Error Absoluto Inicial	Error Relativo Inicial	Tamaño Propuesto bajo M.A.S. Casen 2017	Efecto Diseño Casen 2015 Ajustado	Tamaños m_{0r} ajustados por efecto de diseño	Error Absoluto Estimado	Error Relativo Estimado	Fracción de Muestreo	Tamaños m_{1r} ajustados por finitud	Tasa de No Respuesta	Tamaños m_{2r} ajustados por tasa de no respuesta
	M_r	m_r	p_r	d_{0r}	e_{0r}	m_{0r}	$Deff(p_r)_{ajus}$	m_{1r}	d_{1r}	e_{1r}	f_r	m_{2r}	tnr_r	m_{3r}
Nivel País	4.723.693	82.120	11,7%	0,5%	4,0%	38.450	2,9	71.278	0,5%	4,00%	1,5%	69.816	26,6%	95.159
Arica y Parinacota	61.313	811	9,7%	1,4%	14,8%	2.506	1,0	2.506	1,5%	15,00%	4,1%	2.408	21,5%	3.066
Tarapacá	81.824	2.426	7,1%	1,5%	20,3%	2.805	1,1	3.086	1,4%	20,10%	3,8%	2.974	29,3%	4.206
Antofagasta	144.072	1.908	5,4%	1,6%	29,6%	881	2,9	2.555	1,6%	29,50%	1,8%	2.511	33,0%	3.746
Atacama	81.301	3.839	6,9%	2,1%	30,6%	2.399	1,0	2.399	1,4%	21,10%	3,0%	2.331	23,5%	3.048
Coquimbo	199.142	3.676	13,8%	1,6%	11,4%	2.562	1,2	3.074	1,6%	11,30%	1,5%	3.028	21,2%	3.841
Valparaíso	571.793	8.884	12,0%	1,3%	11,1%	3.089	2,2	6.797	1,3%	11,10%	1,2%	6.717	27,8%	9.298
Metropolitana	1.842.153	16.993	7,1%	0,8%	11,6%	2.609	5,0	13.045	0,9%	12,50%	0,7%	12.954	43,8%	23.051
O'Higgins	241.593	7.060	13,7%	2,0%	14,5%	2.367	2,2	5.208	2,0%	14,40%	2,2%	5.099	17,7%	6.194
Maule	285.612	5.633	18,7%	1,8%	9,4%	2.682	1,9	5.095	1,8%	9,40%	1,8%	5.007	14,8%	5.878
Biobío (Antigua)	555.327	11.320	17,5%	1,3%	7,6%	4.504	2,2	9.908	1,3%	7,60%	1,8%	9.735	18,5%	11.942
Biobío	423.448	8.967	16,6%	1,6%	9,5%	3.189	2,2	7.015	1,6%	9,50%	1,7%	6.901	19,7%	8.589
Ñuble	131.879	2.353	20,8%	2,5%	12,1%	1.259	2,3	2.895	2,5%	12,00%	2,2%	2.834	15,5%	3.353
La Araucanía	269.653	7.012	23,6%	2,1%	8,8%	3.079	1,7	5.235	2,1%	9,00%	1,9%	5.136	13,4%	5.933
Los Ríos	103.351	3.364	16,7%	2,5%	15,0%	2.086	1,8	3.755	2,5%	15,00%	3,6%	3.624	18,0%	4.419
Los Lagos	208.523	6.166	16,0%	2,0%	12,5%	2.478	1,7	4.212	2,0%	12,40%	2,0%	4.129	16,4%	4.938
Aysén	29.541	1.151	6,5%	1,8%	28,4%	1.986	1,0	1.986	1,9%	28,80%	6,7%	1.862	25,7%	2.506
Magallanes	48.495	1.877	4,4%	1,4%	32,6%	2.415	1,0	2.415	1,3%	30,20%	5,0%	2.301	25,6%	3.093

Elaborado por el Instituto Nacional de Estadísticas.

Tabla II.3. Distribución de los tamaños muestrales de viviendas a nivel nacional y regiones, por área urbana y rural, Casen 2017.

Región	Muestra objetivo			Muestra con sobredimensionamiento		
	Total m_{2r}	Urbana $m_{U 2r}$	Rural $m_{R 2r}$	Total m_{3r}	Urbana $m_{U 3r}$	Rural $m_{R 3r}$
Nivel País	69.816	57.087	12.729	95.159	79.992	15.167
Arica y Parinacota	2.408	2.155	253	3.066	2.756	310
Tarapacá	2.974	2.733	241	4.206	3.926	280
Antofagasta	2.511	2.348	163	3.746	3.533	213
Atacama	2.331	2.054	277	3.048	2.722	326
Coquimbo	3.028	2.390	638	3.841	3.075	766
Valparaíso	6.717	5.621	1.096	9.298	7.899	1.399
Metropolitana	12.954	12.227	727	23.051	22.079	972
O'Higgins	5.099	3.585	1.514	6.194	4.382	1.812
Maule	5.007	3.610	1.397	5.878	4.293	1.585
Biobío (Antigua)	9.735	7.676	2.059	11.942	9.507	2.435
Biobío	6.901	5.761	1.140	8.589	7.233	1.356
Ñuble	2.834	1.915	919	3.353	2.274	1.079
La Araucanía	5.136	3.622	1.514	5.933	4.242	1.691
Los Ríos	3.624	2.487	1.137	4.419	3.110	1.309
Los Lagos	4.129	2.927	1.202	4.938	3.554	1.384
Aysén	1.862	1.484	378	2.506	1.989	517
Magallanes	2.301	2.168	133	3.093	2.925	168

Elaborado por el Instituto Nacional de Estadísticas.

La siguiente tabla muestra los errores de muestreo tanto absolutos como relativos para los niveles de estimación nacional rural y nacional urbano.

Tabla II.4. Errores de estimación para los niveles nacional urbano y nacional rural, Casen 2017.

Nivel de Estimación	Muestra Viviendas Efectivas Casen 2015 m_{2015}	Tasa de Pobreza Casen 2015 p_r	Error Absoluto Efectivo Casen 2015 d_{2015r}	Error Relativo Efectivo Casen 2015 e_{2015r}	Muestra objetivo propuesta Casen 2017 m_{2r}	Error Absoluto Estimado Casen 2017 d_{1r}	Error Relativo Estimado Casen 2017 e_{1r}
	Nacional	82.120	11,7%	0,43%	3,68%	69.816	0,46%
Nacional urbano	63.382	10,1%	0,46%	4,52%	57.087	0,48%	4,76%
Nacional rural	18.738	22,1%	1,14%	5,18%	12.729	1,39%	6,28%

Elaborado por el Instituto Nacional de Estadísticas.

En la tabla siguiente se observa la comparación de la distribución final de la muestra Casen 2017 con Casen 2013, donde puede observarse que se ajusta en gran medida a esta última.

Tabla II.5. Comparación de los tamaños de muestra objetivo de Casen 2013 y Casen 2017.

Región	Muestra Viviendas Objetivo Casen 2013		Muestra Viviendas Objetivo Casen 2017	
	m ₂₀₁₃	%m ₂₀₁₃	m ₂₀₁₇	%m ₂₀₁₇
Nivel País	70.080	100%	69.816	100%
Arica y Parinacota	2.581	3,7%	2.408	3,4%
Tarapacá	2.977	4,2%	2.974	4,3%
Antofagasta	2.513	3,6%	2.511	3,6%
Atacama	2.071	3,0%	2.331	3,3%
Coquimbo	3.018	4,3%	3.028	4,3%
Valparaíso	6.737	9,6%	6.717	9,6%
Metropolitana	12.865	18,4%	12.954	18,6%
O'Higgins	5.137	7,3%	5.099	7,3%
Maule	4.872	7,0%	5.007	7,2%
Biobío (Antigua)	9.826	14,0%	9.735	13,9%
Biobío	7.195	10,3%	6.901	9,9%
Ñuble	2.631	3,8%	2.834	4,1%
La Araucanía	5.387	7,7%	5.136	7,4%
Los Ríos	3.768	5,4%	3.624	5,2%
Los Lagos	4.192	6,0%	4.129	5,9%
Aysén	2.051	2,9%	1.862	2,7%
Magallanes	2.085	3,0%	2.301	3,3%

Elaborado por el Instituto Nacional de Estadísticas.

II.5.1. Distribución de las unidades primarias de muestreo

Por último, se estimó el tamaño de muestra de las unidades primaria, la que alcanzó un total de 10.229 UPM, el cual se prevé sea modificado al momento de llevar a cabo la selección, ya que es en esta etapa del proceso donde se determina la existencia y disponibilidad tanto de manzanas como de secciones. Así mismo, se estima que las modificaciones en el tamaño de muestra de unidades primarias no superen 3%.

Tabla II.6. Distribución de las unidades primarias de muestreo, Casen 2017.

Región	Muestra Urbana				Muestra Rural	
	Manzanas MM2008 ¹³	Muestra de Manzanas	Secciones MM2002 RAU ¹⁴	Muestra de Secciones RAU	Secciones MM2002 ¹⁵	Muestra de Secciones
	N _U	n _U	N _{RAU}	n _{RAU}	N _R	n _R
Nivel País	119.970	9.179	1.630	274	6.650	776
Arica y Parinacota	2.144	384	4	3	59	16
Tarapacá	2.103	406	13	3	50	14
Antofagasta	4.975	466	7	3	46	11
Atacama	3.230	446	10	0	106	17
Coquimbo	5.921	430	90	9	478	40
Valparaíso	16.362	972	181	16	474	71
Metropolitana	43.364	2.000	77	5	432	49
O'Higgins	5.008	469	213	31	781	90
Maule	6.089	458	240	49	1.004	81
Biobío (Antigua)	14.599	1.092	288	61	1.121	127
Biobío	11.880	876	164	24	611	72
Ñuble	2.719	216	124	37	510	55
La Araucanía	6.374	535	239	41	934	89
Los Ríos	2.283	391	79	9	362	66
Los Lagos	4.721	410	125	30	683	71
Aysén	911	277	42	11	92	25
Magallanes	1.886	443	22	3	28	9

Elaborado por el Instituto Nacional de Estadísticas.

Los tamaños de la muestra objetivo y total con sobredimensionamiento, por región, así como también la distribución de la muestra diseñada para las áreas urbanas y rurales por provincias, se presenta en la Tabla II.7.

La distribución de la muestra a nivel de comunas se muestra en el Anexo N°2 Distribución de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

¹³ Excluye manzanas con menos de 8 viviendas.

¹⁴ Excluye áreas de difícil acceso (ADA) (Ver Tabla III.1).

¹⁵ Ídem nota anterior.

Tabla II.7. Distribución de los tamaños muestrales de viviendas a nivel nacional, regiones y provincias, por área urbana y rural, Casen 2017.

Región / Provincia	Población Viviendas en el Marco	Muestra objetivo Casen 2017	Muestra objetivo (Urbana) Casen 2017	Muestra objetivo (Rural) Casen 2017	Error Absoluto Estimado ¹⁶	Error Relativo Estimado	Tasa de No Resp.	Sobre- muestra Casen 2017	Sobre- muestra (Urbana) Casen 2017	Sobre- muestra (Rural) Casen 2017
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	d_{1r}	e_{1r}	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$
Nivel País	4.723.693	69.816	57.087	12.729	0,5%	4,0%	26,6%	95.159	79.992	15.167
Arica y Parinacota	61.313	2.408	2.155	253	1,5%	15,0%	21,5%	3.066	2.756	310
Arica	60.747	2.325	2.105	220	1,3%	14,1%	21,9%	2.976	2.702	274
Parinacota	566	83	50	33	12,4%	39,3%	7,8%	90	54	36
Tarapacá	81.824	2.974	2.733	241	1,4%	20,1%	29,3%	4.206	3.926	280
Iquique	76.724	2.664	2.626	38	1,3%	19,8%	31,0%	3.861	3.803	58
Tamarugal	5.100	310	107	203	6,5%	44,0%	10,1%	345	123	222
Antofagasta	144.072	2.511	2.348	163	1,6%	29,5%	33,0%	3.746	3.533	213
Antofagasta	91.754	1.672	1.609	63	1,6%	32,0%	33,7%	2.523	2.448	75
El Loa	41.513	644	576	68	3,3%	66,8%	34,1%	977	879	98
Tocopilla	10.805	195	163	32	6,5%	57,8%	20,7%	246	206	40
Atacama	81.301	2.331	2.054	277	1,4%	21,1%	23,5%	3.048	2.722	326
Copiapó	50.253	1.485	1.404	81	2,0%	38,9%	27,9%	2.060	1.956	104
Chañaral	10.680	238	208	30	2,3%	46,1%	17,4%	288	252	36
Huasco	20.368	608	442	166	2,5%	22,5%	13,1%	700	514	186
Coquimbo	199.142	3.028	2.390	638	1,6%	11,3%	21,2%	3.841	3.075	766
Elqui	124.450	1.863	1.625	238	1,7%	17,3%	23,0%	2.418	2.139	279
Choapa	26.038	461	317	144	5,3%	29,7%	16,9%	555	384	171
Limarí	48.654	704	448	256	4,1%	16,3%	18,9%	868	552	316
Valparaíso	571.793	6.717	5.621	1.096	1,3%	11,1%	27,8%	9.298	7.899	1.399
Valparaíso	247.414	2.467	2.347	120	2,1%	16,5%	34,4%	3.762	3.613	149
Los Andes	31.952	535	412	123	3,8%	40,4%	28,4%	747	580	167
Petorca	20.543	408	283	125	5,7%	31,1%	20,2%	511	352	159
Quillota	55.728	680	528	152	2,9%	22,1%	19,4%	844	662	182
San Antonio	74.340	644	523	121	5,7%	68,7%	22,1%	827	672	155
S. Felipe de Aconcagua	43.302	623	437	186	4,2%	25,4%	30,1%	891	632	259

¹⁶ Los tamaños de muestra provinciales son producto de la agregación de la muestra objetivo comunal, por lo que, para este nivel de estimación, los errores absolutos y relativos se calculan sin considerar la finitud de la población.

Tabla II.7. Distribución de los tamaños muestrales de viviendas a nivel nacional, regiones y provincias, por área urbana y rural, Casen 2017.

Región / Provincia	Población Viviendas en el Marco	Muestra objetivo Casen 2017	Muestra objetivo (Urbana) Casen 2017	Muestra objetivo (Rural) Casen 2017	Error Absoluto Estimado ¹⁶	Error Relativo Estimado	Tasa de No Resp.	Sobre- muestra Casen 2017	Sobre- muestra (Urbana) Casen 2017	Sobre- muestra (Rural) Casen 2017
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	d_{1r}	e_{1r}	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$
Nivel País	4.723.693	69.816	57.087	12.729	0,5%	4,0%	26,6%	95.159	79.992	15.167
Marga Marga	98.514	1.360	1.091	269	2,8%	27,9%	20,7%	1.716	1.388	328
Metropolitana	1.842.153	12.954	12.227	727	0,9%	12,5%	43,8%	23.051	22.079	972
Santiago	1.427.399	9.872	9.872	0	1,0%	15,5%	47,9%	18.950	18.950	0
Cordillera	159.980	891	827	64	3,2%	40,5%	22,8%	1.154	1.059	95
Chacabuco	42.767	534	324	210	6,5%	56,7%	26,0%	722	442	280
Maipo	109.087	725	601	124	4,1%	48,5%	29,3%	1.025	848	177
Melipilla	41.747	442	267	175	3,5%	30,4%	21,9%	566	346	220
Talagante	61.173	490	336	154	3,3%	34,1%	22,7%	634	434	200
O'Higgins	241.593	5.099	3.585	1.514	2,0%	14,4%	17,7%	6.194	4.382	1.812
Cachapoal	166.587	3.290	2.496	794	2,8%	21,5%	16,6%	3.944	3.025	919
Cardenal Caro	16.077	510	301	209	1,9%	12,5%	23,4%	666	393	273
Colchagua	58.929	1.299	788	511	3,0%	18,6%	18,0%	1.584	964	620
Maule	285.612	5.007	3.610	1.397	1,8%	9,4%	14,8%	5.878	4.293	1.585
Talca	109.952	1.598	1.181	417	2,8%	16,8%	15,5%	1.891	1.427	464
Cauquenes	18.499	505	313	192	6,5%	26,6%	14,4%	590	365	225
Curicó	77.506	1.859	1.471	388	2,9%	15,5%	17,0%	2.241	1.784	457
Linares	79.655	1.045	645	400	4,2%	21,4%	9,6%	1.156	717	439
Biobío (Antigua)	555.327	9.735	7.676	2.059	1,3%	7,6%	18,5%	11.942	9.507	2.435
Concepción	277.304	4.142	3.825	317	1,8%	13,1%	23,1%	5.385	4.981	404
Arauco	41.116	883	644	239	4,9%	19,2%	17,4%	1.069	780	289
Biobío	105.028	1.876	1.292	584	3,2%	15,1%	12,1%	2.135	1.472	663
Ñuble	131.879	2.834	1.915	919	2,5%	12,2%	15,5%	3.353	2.274	1.079
Diguillín	89.451	1.527	1.118	409	3,1%	16,4%	16,3%	1.825	1.342	483
Itata	21.140	707	425	282	5,8%	22,6%	14,5%	827	496	331
Punilla	21.288	600	372	228	6,5%	23,6%	14,4%	701	436	265
La Araucanía	269.653	5.136	3.622	1.514	2,1%	9,0%	13,4%	5.933	4.242	1.691
Cautín	207.676	3.852	2.722	1.130	2,2%	10,0%	14,5%	4.506	3.241	1.265
Malleco	61.977	1.284	900	384	5,4%	18,1%	10,0%	1.427	1.001	426

Tabla II.7. Distribución de los tamaños muestrales de viviendas a nivel nacional, regiones y provincias, por área urbana y rural, Casen 2017.

Región / Provincia	Población Viviendas en el Marco M_r	Muestra objetivo Casen 2017 m_{2r}	Muestra objetivo (Urbana) Casen 2017 $m_{U\ 2r}$	Muestra objetivo (Rural) Casen 2017 $m_{R\ 2r}$	Error Absoluto Estimado ¹⁶ $d_{1\ r}$	Error Relativo Estimado $e_{1\ r}$	Tasa de No Resp. tnr_r	Sobre-muestra Casen 2017 m_{3r}	Sobre-muestra (Urbana) Casen 2017 $m_{U\ 3r}$	Sobre-muestra (Rural) Casen 2017 $m_{R\ 3r}$
Nivel País	4.723.693	69.816	57.087	12.729	0,5%	4,0%	26,6%	95.159	79.992	15.167
Los Ríos	103.351	3.624	2.487	1.137	2,5%	15,0%	18,0%	4.419	3.110	1.309
Valdivia	75.935	2.814	2.014	800	2,9%	18,1%	20,2%	3.528	2.589	939
Ranco	27.416	810	473	337	4,8%	22,9%	9,1%	891	521	370
Los Lagos	208.523	4.129	2.927	1.202	2,0%	12,4%	16,4%	4.938	3.554	1.384
Llanquihue	97.312	1.678	1.164	514	2,8%	19,1%	17,3%	2.029	1.433	596
Chiloé	43.834	1.068	653	415	4,8%	27,5%	13,2%	1.230	758	472
Osorno	67.377	1.383	1.110	273	3,3%	18,7%	17,6%	1.679	1.363	316
Aysén	29.541	1.862	1.484	378	1,9%	28,8%	25,7%	2.506	1.989	517
Coyhaique	18.990	1.118	909	209	2,5%	38,3%	25,9%	1.508	1.226	282
Aysén	7.211	531	469	62	2,3%	36,8%	23,3%	692	610	82
Capitán Prat	990	81	50	31	4,0%	45,3%	32,5%	120	74	46
General Carrera	2.350	132	56	76	3,6%	53,8%	29,0%	186	79	107
Magallanes	48.495	2.301	2.168	133	1,3%	30,2%	25,6%	3.093	2.925	168
Magallanes	40.437	1.981	1.908	73	1,2%	26,7%	27,2%	2.722	2.622	100
Tierra del Fuego	1.458	84	54	30	2,9%	256,4%	9,7%	93	60	33
Última Esperanza	6.600	236	206	30	3,3%	64,5%	15,1%	278	243	35

Elaborado por el Instituto Nacional de Estadísticas.

III. DISEÑO MUESTRAL

La población objetivo de la encuesta Casen la constituyen todas las personas y hogares que residen en viviendas particulares a lo largo del territorio nacional. La cobertura de la encuesta Casen 2017 es todo el territorio nacional, excluyendo aquellos sectores identificados por el INE como áreas de difícil acceso (ADA)¹⁷.

El diseño muestral de la encuesta Casen 2017 se puede caracterizar como probabilístico y estratificado, según área geográfica y por tamaño poblacional, tanto en el área urbana como rural. La selección de la muestra se realiza en dos etapas - muestreo bietápico en las áreas rurales y urbanas. Las unidades primarias de muestreo, provenientes del MS2002, se seleccionan con probabilidad proporcional al tamaño, en cuanto al total de viviendas, mientras que aquellas unidades provenientes del MM2008 se seleccionan de forma sistemática y con igual probabilidad. Las unidades últimas, las viviendas, se seleccionan de forma sistemática con igual probabilidad.

La unidad última de selección de la muestra es la vivienda. Al interior de la vivienda se identifican todos los hogares y las personas miembros de cada hogar. La entrevista se realiza con un informante por hogar que debe ser el jefe de hogar u otro integrante de 18 años o más. A través del informante se recolectan datos de todos los miembros del hogar que son residentes habituales de la vivienda¹⁸. Al interior de la vivienda se recolectan datos de todos los residentes habituales y los diferentes hogares que la conforman.

III.1. Características del marco muestral

Un marco de muestreo se define como la lista o los procedimientos que permiten identificar a todos los elementos de una población objetivo (Groves et al. 2004, pág. 68). Listas de miembros de algunas organizaciones o instituciones son ejemplos de marcos de muestreo sencillos. En Chile,

¹⁷ Incluyen a las comunas de General Lagos, Colchane, Ollagüe, Juan Fernández, Isla de Pascua, Cochamó, Chaitén, Futaleufú, Hualaihué, Palena, Lago Verde, Guaitecas, O'Higgins, Tortel, Laguna Blanca, Río Verde, San Gregorio, Cabo de Hornos (Ex - Navarino), Antártica, Primavera, Timaukel, Torres del Paine.

¹⁸ Hasta 2009, todas las preguntas en el cuestionario Casen podían ser respondidas por un tercero (el informante). En el caso específico de Casen 2015, sin embargo, las preguntas sobre orientación sexual e identidad de género (Módulo Residentes, preguntas r21 y r22, para personas de 18 años y más), solo pueden responder personas que físicamente están presentes durante la entrevista. Es decir, no se acepta el reporte de terceros.

por ejemplo, el Ministerio de Educación tiene el listado de todos los colegios públicos, subvencionados y privados en el país.

La población objetivo de la encuesta Casen 2017 son los hogares que residen en viviendas particulares ocupadas y las personas que residen de forma permanente. El INE mantiene un marco de áreas geográficas que sirve de base para la selección de viviendas, requeridas para las muestras de las encuestas de hogares más importantes del país.

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

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

- 1. Ciudad (CD):** Es un gran centro urbano conformado por uno o un conjunto de centros urbanos adyacentes con 40.000 o más habitantes.
- 2. 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).
- 3. Urbano (U):** Es el centro urbano con menos de 40.000 y más de 2.000 habitantes. Esta clasificación se da cuando en la comuna no existe una ciudad (CD), por lo que cada uno de sus centros urbanos se denominan simplemente como urbanos (U).
- 4. 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 primarias¹⁹.

Para cumplir con los objetivos de investigación, el marco de muestreo “ideal” debe ser completo, preciso y actualizado. En consecuencia, a mediados del periodo intercensal, investigaciones realizadas por el INE indicaban que el marco muestral en uso (MS2002) presentaba deficiencias

¹⁹ Se entiende por Actividad Primaria a toda aquella actividad relacionada con la extracción de recursos naturales. (agricultura, caza, pesca, minería, etc.).

en estos aspectos, por lo tanto, a partir de 2008 el Instituto desarrolló un nuevo marco de muestreo, junto a definiciones para la conformación de unidades de muestreo y procedimientos para la selección de unidades de muestreo en las áreas urbanas del país²⁰.

Bajo el nuevo marco de muestreo urbano, se mantuvo la división político-administrativa, pero se cambió la conformación de las unidades primarias de muestreo. En el nuevo marco, las manzanas censales (predios urbanos delimitados por calles) reemplazaron a las secciones de empadronamiento censal en las áreas urbanas²¹. Esto permitió actualizar el número de viviendas por conglomerado por medio del plan municipal de edificaciones que se entrega por convenio al Departamento de Geografía, logrando captar nuevos desarrollos urbanos tanto en áreas urbanas como en áreas previamente catalogadas como agrícolas (por ejemplo, nuevos desarrollos en comunas como Peñalolén, Lampa o Chicureo en la Región Metropolitana).

En ese sentido, las sub-divisiones denominadas manzanas censales (en las áreas urbanas) y secciones de empadronamiento censal (en las áreas rurales), definen las unidades geográficas más pequeñas y corresponden a las unidades primarias de muestreo (UPM) más comúnmente utilizadas en las encuestas de hogares diseñadas por el INE.

El INE comenzó a seleccionar las muestras urbanas utilizando el nuevo marco de muestreo a partir del año 2008 (MM2008). Las primeras encuestas a nivel nacional, cuyas muestras urbanas fueron seleccionadas a partir del nuevo marco, fueron la Nueva Encuesta del Empleo (NENE) en 2009 y la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) en 2008. En las zonas rurales, sin embargo, el INE ha seguido utilizando el marco de secciones (MS2002) para la selección de muestras.

En los últimos años se ha observado un importante crecimiento demográfico en las áreas urbanas del país, que ha sido observado a partir de la enumeración muestral de las diversas encuestas en que el INE participa. Con frecuencia, debido a los cambios de estructura y de tamaño de las manzanas, es necesario reemplazar algunas de las unidades seleccionadas. En este contexto, en el año 2014, el Ministerio y el INE acuerdan realizar una actualización del marco muestral para la

²⁰ 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://historico.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/pdf/actualizaciondeldisenomuestral.pdf

²¹ 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.

muestra de la encuesta Casen 2015. En respuesta a tal requerimiento, en el segundo semestre del año 2014, el INE realizó una actualización²² parcial del marco de manzanas MM2008 y que correspondió a la actualización de 2.351 manzanas 1,96% del total de 119.970 que posee el marco completo que excluye las manzanas de entre 1 y 7 viviendas y 9,7% de un total de 24.182 que poseen las 61 comunas con actualización, correspondientes a 69 urbanos que, de acuerdo a las visitas a terreno de los distintos equipos de levantamiento de encuestas en el INE, poseían altos índices de manzanas no elegibles por cambios de estructura, crecimientos y decrecimientos. De esta actualización resultó el marco muestral de manzanas actualizado foco Casen (MM2008) el cual se encuentra vigente para la selección en áreas urbanas de Casen 2017.

Finalmente, la selección de la muestra 2017 se realizó sobre el marco muestral que el INE mantiene vigente al año 2017, el cual comprende el uso de dos marcos de muestreo mutuamente excluyentes:

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

²² Ver documentos metodológicos "Metodología de Diseño Muestral Encuesta de Caracterización Socioeconómica Nacional, Casen 2015", anexos, sección 2, donde se describe en detalle el proceso de actualización, disponible en <http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/metodologia.php>

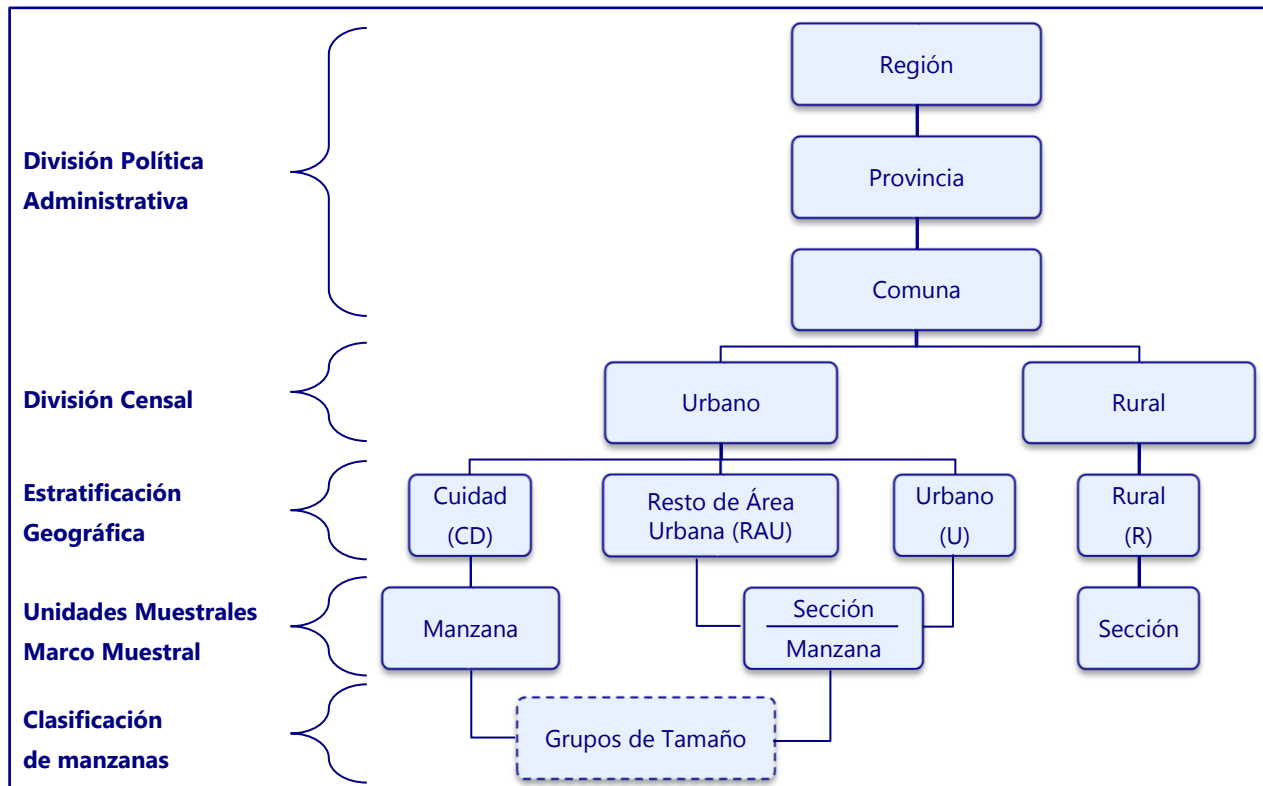
Para aquellas áreas denominadas Resto de Área Urbana (RAU) y Urbano (U) sus unidades pueden pertenecer al marco de secciones MS2002 o al marco de manzanas MM2008. De un total de 400 urbanos, correspondientes a 246 comunas, 252 pertenecen al MS2002; y 148, al marco MM2008. Cabe señalar que las 252 áreas urbanas presentes en el MS2002, corresponden a áreas de pequeña densidad poblacional, con no más de 40.000 habitantes, y las unidades muestrales, al igual que en las áreas rurales, son secciones.

Para fines de selección muestral, de distintos instrumentos de recolección de datos en que INE ha participado, se han clasificado las unidades primarias de muestreo que conforman el MM2008; el objetivo, es poder generar grupos homogéneos de manzanas, respecto de una cierta característica y, de este modo, establecer en cada grupo un número promedio de unidades secundarias de muestreo para encuestar. Para la consecución de tal propósito, se realizó un análisis de clúster²³, que utilizó el número de viviendas particulares, al interior de cada manzana, como variable de clasificación. Como resultado de este procedimiento se pudo clasificar las manzanas del MM2008 según 5 grupos de tamaño. Desde esta última clasificación surge la variable “grupo de tamaño” que se añade al marco muestral, previo a la selección de cualquier muestra, y permite que ésta quede conformada por manzanas de diferentes tamaños.

La Figura III.1 resume las unidades geográficas descritas en esta sección.

²³ Técnica estadística que utiliza distintas medidas de distancia para conformar grupos homogéneos en cuanto a características multivariadas.

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



Elaborado por el Instituto Nacional de Estadísticas.

III.2. Cobertura geográfica del marco muestral

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

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

Es importante distinguir la falta de cobertura (fallas no intencionadas), de las exclusiones que realiza el investigador en forma intencionada. En el proceso de selección de la encuesta Casen

2017, al igual que en las versiones anteriores, se pueden identificar tres tipos de exclusiones, según se detalla a continuación:

1. Las 22 áreas geográficas que han sido clasificadas por el INE como áreas de difícil acceso (ADA);
2. Las manzanas y secciones incluidas en otras muestras seleccionadas por el INE para el periodo de recolección de datos de la encuesta Casen 2017;
3. Las manzanas con 7 o menos viviendas.

Las áreas de difícil acceso corresponden a zonas geográficas que, por razones de orden climático, topográfico o por ausencias de vías y medios de comunicación expeditos permanecen aisladas parcial o totalmente en el año, lo cual ha impedido su correcta incorporación al marco muestral del INE. En este sentido, el alto costo y las dificultades operativas que implican ampliar la cobertura del marco a estas áreas, así como asegurar la oportunidad y calidad en el levantamiento de información de éstas, no justifica la pequeña ganancia en la precisión de los estimadores, por tanto, estas áreas no están presentes en ninguna de las muestras de las encuestas de hogares del INE. En total, corresponden a 0,28% de la población de viviendas según datos del Censo 2002 y en términos de personas representa el 0,24% de la población del país²⁴.

La Tabla III.1 muestra las 22 áreas de difícil acceso excluidas de la muestra de la encuesta Casen 2017, donde se presenta el total de viviendas según la información del Censo de Población y Vivienda del año 2002, y una aproximación del total de personas, estimadas mediante las estimaciones poblacionales.

Por otra parte, la Encuesta Nacional del Empleo (ENE) se levanta todos los meses, lo que tiene como consecuencia que, durante los meses de octubre, noviembre y diciembre, coincide temporalmente con el levantamiento de la encuesta Casen. Para evitar seleccionar viviendas que se encuentren presentes en la muestra de la Encuesta Nacional de Empleo, previendo su rechazo a participar en una o ambas encuestas, se excluyen²⁵ del marco de muestreo todas aquellas manzanas y secciones que forman parte de esa muestra. Así, el total de unidades excluidas del marco de muestreo porque se encuentran en la Encuesta Nacional de Empleo corresponden a 2,3% de manzanas (2,3% de viviendas aproximadamente) en el MM2008, y 6,5% de secciones

²⁴ calculado en base a estimaciones poblacionales desarrolladas por INE al 30 de noviembre de 2017.

²⁵ Se excluyen previo a la selección de la muestra Casen 2017.

(6,5% de viviendas aproximadamente)²⁶. Esto equivale, aproximadamente, a 3.073 manzanas y 543 secciones según trimestre móvil. En las Tablas III.2 y III.3, se presenta el detalle según marco de procedencia y región.

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

Por la exclusión de los conglomerados seleccionados para el levantamiento de la Encuesta Nacional de Empleo del INE en el mismo período de Casen 2017 y de aquellos de tamaño menor a 8 viviendas, el INE aplica una corrección denominada “ajuste por omisión de conglomerados”, propio del cálculo de las probabilidades de selección de éstos. Sin embargo, para las áreas de difícil acceso que quedan excluidas del marco y en las cuales por defecto no existen conglomerados definidos como tales, el ajuste se hace indirectamente en el proceso de calibración o ajuste por las estimaciones poblacionales, las cuales incluyen a nivel agregado como por ejemplo región, la población contenida en estas áreas.

En total, el marco muestral a partir del cual se seleccionó la muestra, MS2002 y MM2008 (actualizado con foco Casen 2015) en su conjunto, excluye aproximadamente 4,0% 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 tengan estas exclusiones en consideración. Si los fenómenos de interés no están relacionados con el tamaño de las manzanas (en áreas urbanas) o no están relacionados con la aislación geográfica (en áreas rurales), entonces es probable que la falta de cobertura del marco muestral no afecte las inferencias hechas a partir de los datos de la encuesta. Si lo contrario es cierto, entonces los investigadores deberán ser cautelosos al momento de hacer inferencia a la población bajo estudio.

Como se puede evidenciar en las Tablas III.2 y III.3 la exclusión de unidades muestrales no sigue un patrón uniforme en las regiones del país, observándose que en el área rural de ciertas regiones la exclusión puede superar 10%, como es el caso de las regiones de Arica y Parinacota (35,2%), Coquimbo (13,6%), Atacama (11,1%), Los Lagos (10,7%) y Aysén (10,1%). Mientras que, en el resto de áreas urbanas, la exclusión más significativa ocurre en la Región de Tarapacá (25,0%), seguido por la Región Metropolitana (12,5%), Región de Aysén (11,7%) y Región de Magallanes (10,7%).

²⁶ 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.

Las unidades excluidas en el MS2002 son aquellas que pertenecen a la muestra de la Encuesta Nacional de Empleo. Con su exclusión se busca reducir la posibilidad de tener rechazos en las futuras entrevistas de las viviendas rurales de la ENE (cada una de ellas entrevistada en 12 oportunidades) y además rechazos en Casen por causa de cansancio del informante.

Tabla III.1. Áreas de difícil acceso definidas por el INE.

Región	Provincia	Comuna	Total viviendas particulares Ocupadas Censo 2002	Total de personas Estimaciones poblacionales*
Totales			11.202	44.513
Arica y Parinacota	Parinacota	General Lagos	263	1.183
Tarapacá	Tamarugal	Colchane	464	1.569
Antofagasta	El Loa	Ollagüe	66	213
Valparaíso	Valparaíso	Juan Fernández	209	1.005
	Isla de Pascua	Isla de Pascua	1.154	5.910
Los Lagos	Llanquihue	Cochamó	1.357	4.187
		Chaitén	1.845	6.763
	Palena	Futaleufú	613	1.781
		Hualaihué	2.252	8.250
		Palena	559	1.602
Aysén	Coyhaique	Lago Verde	338	862
	Aysén	Guaitecas	383	1.983
	Capitán Prat	O'Higgins	154	816
		Tortel	145	699
Magallanes	Magallanes	Laguna Blanca	118	637
		Río Verde	86	352
		San Gregorio	216	548
	Antártica Chilena	Cabo de Hornos	523	3.099
		Antártica	10	62
	Tierra del Fuego	Primavera	232	436
		Timaukel	85	1.153
Última Esperanza	Torres del Paine	130	1.403	

*Estimaciones poblacionales al 30 de noviembre de 2017.

Elaborado por el Instituto Nacional de Estadísticas, 2017.

Tabla III.2. Distribución del total de viviendas y secciones excluidas desde el MS2002, según región y área, Casen 2017.

Región	Total en el marco		Total excluido ENE						% Excluido ENE					
			Total		RAU		Rural		Total		RAU		Rural	
País	Secciones	Viviendas	Secciones	Viviendas	Secciones	Viviendas	Secciones	Viviendas	Secciones	Viviendas	Secciones	Viviendas	Secciones	Viviendas
	8.323	661.702	543	42.985	79	7.174	464	35.811	6,5%	6,5%	4,8%	5,2%	6,9%	6,8%
Arica y Parinacota	64	3.301	21	1.082	0	0	21	1.082	32,8%	32,8%	0,0%	0,0%	35,0%	35,2%
Tarapacá	71	3.685	7	405	3	198	4	207	9,9%	11,0%	23,1%	25,0%	6,9%	7,2%
Antofagasta	53	2.730	1	43	0	0	1	43	1,9%	1,6%	0,0%	0,0%	2,2%	1,9%
Atacama	116	6.012	11	590	0	0	11	590	9,5%	9,8%	0,0%	0,0%	10,4%	11,1%
Coquimbo	568	44.602	68	5.389	4	297	64	5.092	12,0%	12,1%	4,4%	4,2%	13,4%	13,6%
Valparaíso	655	47.844	38	2.830	6	474	32	2.356	5,8%	5,9%	3,3%	3,4%	6,8%	6,9%
Metropolitana	509	58.080	34	4.300	9	1.722	25	2.578	6,7%	7,4%	11,7%	12,5%	5,8%	5,8%
O'Higgins	994	76.214	69	5.007	11	880	58	4.127	6,9%	6,6%	5,2%	5,2%	7,4%	7,0%
Maule	1.244	100.557	70	5.731	14	1.149	56	4.582	5,6%	5,7%	5,8%	5,9%	5,6%	5,7%
Biobío (Antigua)	1.409	109.517	49	3.797	12	988	37	2.809	3,5%	3,5%	4,2%	4,2%	3,3%	3,3%
Biobío	775	58.268	37	2.819	11	922	26	1.897	4,8%	4,8%	6,7%	6,8%	4,3%	4,2%
Ñuble	634	51.249	12	978	1	66	11	912	1,9%	1,9%	0,8%	0,7%	2,2%	2,2%
La Araucanía	1.173	95.603	48	3.928	3	255	45	3.673	4,1%	4,1%	1,3%	1,3%	4,8%	4,8%
Los Ríos	441	35.438	32	2.603	7	549	25	2.054	7,3%	7,3%	8,9%	8,6%	6,9%	7,1%
Los Lagos	821	67.021	75	6.184	3	216	72	5.968	9,1%	9,2%	2,2%	1,9%	10,5%	10,7%
Aysén	141	7.604	15	809	5	307	10	502	10,6%	10,6%	11,9%	11,7%	10,1%	10,1%
Magallanes	64	3.494	5	287	2	139	3	148	7,8%	8,2%	9,1%	10,7%	7,1%	6,8%

Elaborado por el Instituto Nacional de Estadísticas.

Tabla III.3. Distribución del total de viviendas y manzanas excluidas desde el MM2008, según región, Casen 2017.

Región	Total en el marco		Total excluido						% Total Excluido					
	Manzanas	Viviendas	Total		Manzanas con 7 o menos viviendas		Excluido ENE		% Total		% Manzanas con 7 o menos viviendas		% Excluido ENE	
País	133.759	4.064.511	Manzanas	Viviendas	Manzanas	Viviendas	Manzanas	Viviendas	Manzanas	Viviendas	Manzanas	Viviendas	Manzanas	Viviendas
Arica y Parinacota	2.424	58.057	434	4.164	280	1.050	154	3.114	17,9%	7,2%	11,6%	1,8%	6,4%	5,4%
Tarapacá	2.295	78.556	278	3.742	192	659	86	3.083	12,1%	4,8%	8,4%	0,8%	3,7%	3,9%
Antofagasta	5.698	141.342	851	5.918	723	2.587	128	3.331	14,9%	4,2%	12,7%	1,8%	2,2%	2,4%
Atacama	3.966	75.289	825	5.088	736	2.785	89	2.303	20,8%	6,8%	18,6%	3,7%	2,2%	3,1%
Coquimbo	6.824	154.540	1.036	7.025	903	3.631	133	3.394	15,2%	4,5%	13,2%	2,3%	1,9%	2,2%
Valparaíso	19.346	523.949	3.481	27.701	2.984	12.569	497	15.132	18,0%	5,3%	15,4%	2,4%	2,6%	2,9%
Metropolitana	46.017	1.784.073	3.394	38.484	2.653	9.099	741	29.385	7,4%	2,2%	5,8%	0,5%	1,6%	1,6%
O'Higgins	5.789	165.379	897	6.683	781	3.019	116	3.664	15,5%	4,0%	13,5%	1,8%	2,0%	2,2%
Maule	6.703	185.055	760	6.530	614	2.572	146	3.958	11,3%	3,5%	9,2%	1,4%	2,2%	2,1%
Biobío (Antigua)	16.492	445.810	2.357	21.650	1.893	7.481	464	14.169	14,3%	4,9%	11,5%	1,7%	2,8%	3,2%
Biobío	13.460	365.180	1.987	18.776	1.580	6.262	407	12.514	14,8%	5,1%	11,7%	1,7%	3,0%	3,4%
Ñuble	3.032	80.630	370	2.874	313	1.219	57	1.655	12,2%	3,6%	10,3%	1,5%	1,9%	2,1%
La Araucanía	7.202	174.050	989	7.179	828	3.212	161	3.967	13,7%	4,1%	11,5%	1,8%	2,2%	2,3%
Los Ríos	2.569	67.913	345	3.012	286	1.052	59	1.960	13,4%	4,4%	11,1%	1,5%	2,3%	2,9%
Los Lagos	5.332	142.534	746	6.132	611	2.373	135	3.759	14,0%	4,3%	11,5%	1,7%	2,5%	2,6%
Aysén	1.029	22.253	214	2.665	118	454	96	2.211	20,8%	12,0%	11,5%	2,0%	9,3%	9,9%
Magallanes	2.073	45.711	255	2.187	187	882	68	1.305	12,3%	4,8%	9,0%	1,9%	3,3%	2,9%

Elaborado por el Instituto Nacional de Estadísticas.

III.3. Estratificación del marco muestral

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

La encuesta Casen ha definido tradicionalmente los estratos de selección de la muestra a partir del criterio de proximidad geográfica. Hasta 1996 los estratos se conforman a partir de comunas y grupos de comunas, según área urbano/rural²⁷. A partir de 1998 los estratos se constituyen a partir de las comunas, según área urbano/rural²⁸. Para la encuesta Casen 2017 se conformaron h=1.129 estratos, a partir de la interposición de la división político-administrativa (a nivel de comunas) y la división censal (urbano-rural). De éstos, 853 estratos se encuentran en zonas urbanas y 276 en zonas rurales. Para la parte urbana, se incorporó una estratificación que, a diferencia de las anteriores de carácter principalmente geográfico, corresponden a la clasificación de todas las manzanas del marco en grupos de tamaño (ver Tabla III.4) de acuerdo a su tamaño en viviendas.

En el MM2008, al interior de cada comuna, las manzanas fueron clasificadas según su tamaño en 5 grupos, con el objetivo de poder distribuir la muestra de viviendas, en cada uno de estos grupos de manzanas y de esta forma, tener representadas en la muestra los diversos grupos de tamaño²⁹. La distribución de los grupos de tamaño en el marco se presenta a continuación en la Tabla III.4.

²⁷ En 1994 se conforman 146 estratos y en 1996, 246 estratos.

²⁸ En 1998 se conformaron 358 estratos. En 2000, 2003, 2006 y 2009 se conformaron 529, 553, 605 y 602 estratos respectivamente.

²⁹ Es posible que en algunas comunas no existan los 5 grupos de tamaño porque no existen manzanas en el rango del grupo faltante.

Tabla III.4. Número de manzanas y viviendas en el MM2008, según grupo de tamaño de manzanas.

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

Elaborado por el Instituto Nacional de Estadísticas.

III.4. Estimación del tamaño muestral

El tamaño y distribución de la muestra de la encuesta Casen 2013 se determinó en forma similar a la encuesta Casen 2011 que a su vez presenta características que la diferencian de sus predecesoras (2003 a 2009): la muestra Casen 2011, submuestra 2, comparable³⁰ con las versiones anteriores, es más pequeña (57.361 viviendas logradas, 59.084 hogares) en comparación con Casen 2009 (70.241 viviendas logradas, 71.460 hogares), mientras que para Casen 2013 se lograron 64.842 viviendas (66.725 hogares).

En la encuesta Casen 2011, dado que las zonas extremas del norte y sur del país presentaban bajas tasas de pobreza y desde el Ministerio se solicitó contar con un buen nivel de precisión en estimación regional, la muestra fue distribuida en mayor proporción en las regiones del extremo norte y sur del país. En la encuesta Casen 2013, atendidas las tasas de pobreza, los valores actualizados de parámetros y las dificultades en levantamiento, no hacía sentido mantener una muestra tan grande en regiones extremas para encontrarse en rango de errores muestrales razonables.

En el diseño de Casen 2015, se planteó como objetivo alcanzar mejores niveles de precisión para las estimaciones comunales³¹, dando prioridad a aquellas comunas de mayor importancia como capitales regionales y provinciales del país, entre otras que el Ministerio consideró prioritarias

³⁰ La submuestra 2 de Casen 2011 se levantó aproximadamente en el mismo período de Casen 2009 y anteriores.

³¹ Con excepción de las 22 comunas clasificadas como áreas de difícil acceso.

respecto a la necesidad de información. En este sentido, en la aplicación de la encuesta Casen 2015 se logró una muestra de 82.120 viviendas y 83.887 hogares.

En Casen 2017, al igual que en 2011 y 2013, el nivel de estimación más desagregado en la encuesta es el nivel regional, lo que implica la presencia de las comunas³² 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 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 validas a nivel comunal, basadas en la metodología de estimación para Áreas Pequeñas³³.

Al igual, al interior de cada comuna se conforma la muestra de 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.

En este contexto, el tamaño de muestra objetivo o esperado de Casen 2017 es de 69.816 viviendas, el cual fue determinado en forma independiente para cada una de las regiones del país, siendo la variable de interés la tasa de pobreza por ingreso.

En general, el tamaño muestral de la encuesta Casen 2017 fue determinado de acuerdo al siguiente diseño:

- 1.** A partir los tamaños de muestra objetivo de Casen 2013, se definió el tamaño de muestra objetivo de Casen 2017 para cada una de las regiones del país. Para los cuales se estimaron los errores con base en la tasa de pobreza por ingresos de la encuesta Casen 2015, cuidando de que el error absoluto de estas tasas no superara 4 puntos porcentuales, y en

³² La encuesta Casen 2017 tendrá presencia en las mismas 324 comunas que en versiones anteriores.

³³ Ver documentos metodológicos "Metodología actualizada de estimación para áreas pequeñas (SAE): Tasa de pobreza por ingresos a nivel comunal (2011-2013)" disponible en:

http://observatorio.ministeriodesarrollosocial.gob.cl/documentos/Metodologia_SAE_Tasa_Pobreza_por_Ingresos_2011_2013.pdf

lo posible el error relativo no superara 30 puntos porcentuales. Así mismo, se procuró que a nivel provincial el error absoluto no superara, en lo posible, 6,5 puntos porcentuales.

2. Se distribuyó el total de viviendas de la región de forma proporcional en las comunas en cuanto a la distribución de la muestra objetivo comunal de Casen 2013, considerando también, la distribución de la muestra según área urbana y rural.
3. Se aseguró que el tamaño mínimo de muestra objetivo para el área urbana fuera de 50 viviendas por comuna y para el área rural de 30 viviendas por comuna.
4. Finalmente, para todas las comunas se calculó el tamaño muestral sobredimensionado según área urbana y rural, y con ello se determinó el tamaño sobredimensionado de la región a partir de la agregación de las comunas.

El detalle de los niveles de precisión esperados (error absoluto y error relativo) y los tamaños muestrales en cada etapa del proceso se encuentra a continuación.

Tabla III.5. Tamaños muestrales de viviendas bajo muestreo aleatorio simple y parámetros de cálculo, a nivel nacional y regiones, Casen 2017.

Región	Población Viviendas en el Marco	Muestra Viviendas Efectivas Casen 2015	Tasa de Pobreza Casen 2015	Efecto Diseño Casen 2015	Cuasi-Varianza Casen 2015	Error Absoluto Efectivo Casen 2015	Error Relativo Efectivo Casen 2015	Error Absoluto Inicial	Error Relativo Inicial	Tamaño Propuesto bajo M.A.S. Casen 2017
	M_r	$m_{2015 r}$	p_r	$Deff(p_r)$	$S(p)^2_r$	$d_{2015 r}$	$e_{2015 r}$	$d_{0 r}$	$e_{0 r}$	$m_{0 r}$
Nivel País	4.723.693	82.120	11,7%	2,9	0,13	0,4%	3,7%	0,5%	4,0%	38.450
Arica y Parinacota	61.313	811	9,7%	1,0	0,14	2,6%	26,4%	1,4%	14,8%	2.506
Tarapacá	81.824	2.426	7,1%	1,1	0,15	1,6%	22,6%	1,5%	20,3%	2.805
Antofagasta	144.072	1.908	5,4%	2,9	0,06	1,8%	34,2%	1,6%	29,6%	881
Atacama	81.301	3.839	6,9%	0,5	0,28	1,1%	16,7%	2,1%	30,6%	2.399
Coquimbo	199.142	3.676	13,8%	1,2	0,17	1,4%	10,4%	1,6%	11,4%	2.562
Valparaíso	571.793	8.884	12,0%	2,2	0,14	1,2%	9,7%	1,3%	11,1%	3.089
Metropolitana	1.842.153	16.993	7,1%	5,8	0,05	0,8%	10,9%	0,8%	11,6%	2.609
O'Higgins	241.593	7.060	13,7%	2,2	0,24	1,7%	12,4%	2,0%	14,5%	2.367
Maule	285.612	5.633	18,7%	1,9	0,22	1,7%	8,9%	1,8%	9,4%	2.682
Biobío (Antigua)	555.327	11.320	17,5%	2,2	0,21	1,2%	7,1%	1,3%	7,6%	4.504
Biobío	423.448	8.967	16,6%	2,2	0,2	1,4%	8,4%	1,6%	9,5%	3.189
Ñuble	131.879	2.353	20,8%	2,3	0,21	2,8%	13,3%	2,5%	12,1%	1.259
La Araucanía	269.653	7.012	23,6%	1,7	0,35	1,8%	7,7%	2,1%	8,8%	3.079
Los Ríos	103.351	3.364	16,7%	1,8	0,34	2,7%	15,8%	2,5%	15,0%	2.086
Los Lagos	208.523	6.166	16,0%	1,7	0,26	1,6%	10,3%	2,0%	12,5%	2.478
Aysén	29.541	1.151	6,5%	1,0	0,17	2,4%	37,9%	1,8%	28,4%	1.986
Magallanes	48.495	1.877	4,4%	0,9	0,13	1,5%	34,3%	1,4%	32,6%	2.415

Elaborado por el Instituto Nacional de Estadísticas.

Tabla III.6. Tamaños muestrales de viviendas bajo diseño complejo y parámetros de cálculo, a nivel nacional y regiones, Casen 2017.

Región	Efecto Diseño Ajustado Casen 2015	Tamaños m_{0r} ajustados por efecto de diseño	Error Absoluto Estimado	Error Relativo Estimado	Fracción de Muestreo	Tamaños m_{1r} ajustados por finitud	Tasa de No Respuesta	Tamaños m_{2r} ajustados por tasa de no respuesta
	$Deff(p_r)_{ajus}$	m_{1r}	d_{1r}	e_{1r}	f_r	m_{2r}	tnr_r	m_{3r}
Nivel País	2,9	71.278	0,5%	4,0%	1,5%	69.816	26,6%	95.159
Arica y Parinacota	1,0	2.506	1,5%	15,0%	4,1%	2.408	21,5%	3.066
Tarapacá	1,1	3.086	1,4%	20,1%	3,8%	2.974	29,3%	4.206
Antofagasta	2,9	2.555	1,6%	29,5%	1,8%	2.511	33,0%	3.746
Atacama	1,0	2.399	1,4%	21,1%	3,0%	2.331	23,5%	3.048
Coquimbo	1,2	3.074	1,6%	11,3%	1,5%	3.028	21,2%	3.841
Valparaíso	2,2	6.797	1,3%	11,1%	1,2%	6.717	27,8%	9.298
Metropolitana	5,0	13.045	0,9%	12,5%	0,7%	12.954	43,8%	23.051
O'Higgins	2,2	5.208	2,0%	14,4%	2,2%	5.099	17,7%	6.194
Maule	1,9	5.095	1,8%	9,4%	1,8%	5.007	14,8%	5.878
Biobío (Antigua)	2,2	9.908	1,3%	7,6%	1,8%	9.735	18,5%	11.942
Biobío	2,2	7.015	1,6%	9,5%	1,7%	6.901	19,7%	8.589
Ñuble	2,3	2.895	2,5%	12,0%	2,2%	2.834	15,5%	3.353
La Araucanía	1,7	5.235	2,1%	9,0%	1,9%	5.136	13,4%	5.933
Los Ríos	1,8	3.755	2,5%	15,0%	3,6%	3.624	18,0%	4.419
Los Lagos	1,7	4.212	2,0%	12,4%	2,0%	4.129	16,4%	4.938
Aysén	1,0	1.986	1,9%	28,8%	6,7%	1.862	25,7%	2.506
Magallanes	1,0	2.415	1,3%	30,2%	5,0%	2.301	25,6%	3.093

Elaborado por el Instituto Nacional de Estadísticas.

Como se ha mencionado en el capítulo II, sección 3, el cálculo del tamaño muestral de Casen 2017 prevé una serie de procedimientos que corresponden, en principio, al paso de un diseño bajo muestreo aleatorio simple a un diseño complejo, seguido de otros dos ajustes, determinados por finitud de la población y la tasa de no respuesta, obteniendo así el tamaño de muestra sobredimensionado. En los párrafos siguientes, se describe en detalle cada uno de los ocho pasos realizados para definir la muestra con el total de viviendas a encuestar en cada región.

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

Considérese,

- p : Prevalencia de la variable cualitativa de interés en el estudio, corresponde a la tasa de pobreza por ingreso en la región r .
- n_r : Número de conglomerados o unidades primarias de muestreo (manzanas o secciones) resultantes en la región r .
- \bar{m}_r : Número promedio de viviendas a encuestar por unidad primaria de muestreo resultantes en la región r .
- $n \cdot \bar{m} = m_r$: Número de viviendas a encuestar en la región r .
- M_r : Número de viviendas en el marco muestral para la región r .
- $Def(p)_r$: Efecto del diseño³⁴ asociado a la tasa de pobreza (p) en la región r . Por definición el efecto del diseño es $Def(p) = \sigma^2(p)_{MC} / \sigma^2(p)_{MAS}$, donde $\sigma^2(p)_{MC}$ es la varianza bajo un diseño complejo y $\sigma^2(p)_{MAS}$ es la varianza bajo un muestreo aleatorio simple.
- $SE(p)_r$: Error estándar de la estimación de la pobreza (p) en la región r , igual a la raíz cuadrada de la varianza de la estimación.
- $S(p)_r^2$: Cuasivarianza poblacional de la tasa de pobreza (p) en la región r . Para variables cualitativas y en un muestreo aleatorio simple, corresponde aproximadamente a $(M_r / (M_r - 1)) \cdot (P \cdot Q)$, siendo P la Tasa de pobreza por ingresos regional, con $Q = (1 - P)$ y M_r el total de viviendas en la región r .
- $Z_{1-\alpha/2}$: Percentil de nivel $(1 - \alpha/2)$ de la distribución Normal, correspondiente a una estimación intervalar de $(1 - \alpha)$ de confianza.

Paso 1

Se obtienen estimaciones, a nivel regional, de la tasa de pobreza por ingresos, el error estándar asociado a la tasa de pobreza por ingresos y el efecto del diseño asociado a este indicador, usando

³⁴ Ver capítulo II, sección 3, Metodología de cálculo del tamaño muestral usando muestras complejas.

los resultados del trabajo de campo de Casen 2015. La cuasi-varianza poblacional $S(p)_r^2$ se deriva a partir del error estándar de la tasa de pobreza por ingreso 2015 utilizando la siguiente fórmula:

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

Paso 2

Se establecen parámetros para los errores absolutos “objetivo” por región (d_{0r}), cuyo valor máximo es 4 puntos porcentuales y el error relativo (e_{0r}) asociado según la estimación de la tasa de pobreza por ingresos, en lo posible, que no sobrepase 30%. También se establece el nivel de confianza de 95% para obtener el percentil de la distribución normal $Z_{1-\alpha/2}$. Con éstos y la cuasi-varianza de la tasa de pobreza derivada en (1), se obtiene un tamaño muestral inicial considerando un muestreo aleatorio simple monoetápico m_{0r} .

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

Paso 3

Se estima el tamaño muestral considerando el efecto del diseño al aplicar un muestreo en dos etapas. El efecto del diseño representa el ajuste del tamaño anterior debido al cambio que experimentará la varianza al pasar de un muestreo monoetápico a uno bietápico, y con el objetivo de mantener el error absoluto inicial. Para esto se utilizan el efecto del diseño por región asociado a la tasa de pobreza por ingresos $Def(p)_r$.

El Def fue ajustado a un mínimo de 1,0 y un máximo 5,0, redondeado a un decimal (ver columna $Def(pr)_{ajus}$ en Tabla III.6). En consecuencia, el Def de las regiones de Atacama y Magallanes fueron aumentados, pasando de 0,5 a 1,0 y de 0,9 a 1,0, respectivamente. Por su parte, para la Región Metropolitana se disminuyó de 5,8 a 5,0.

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

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

Donde m_{0r} representa el total de viviendas estimadas a encuestar en la región r mediante un muestreo aleatorio simple y m_{1r} representa el número de viviendas a encuestar en la región r , ajustado por efecto de diseño. En las Tablas III.5 (columna m_{0r}) y III.6 (columna m_{1r}) se presentan estos valores, según región.

Paso 4

Dado que el efecto del diseño utilizado en el cálculo fue ajustado³⁵, obteniendo así, un tamaño de muestra bajo muestreo complejo (m_{1r}) cuyos errores absoluto (d_{1r}) y relativo (e_{1r}) difieren de los definidos inicialmente en el paso 2³⁶, es necesario estimar nuevamente estos errores. Para lo cual, se emplea la siguiente fórmula:

$$d_{1r} = d_{2015r} \cdot \sqrt{\frac{m_{2015r}}{m_{1r}}} \quad (4)$$

Donde:

d_{1r} : Error absoluto para el tamaño de muestra m_{1r} .

d_{2015r} : Error absoluto efectivo para la región r en Casen 2015.

m_{1r} : Número de viviendas a encuestar en la región r , ajustado por efecto de diseño.

m_{2015r} : Tamaño de muestra logrado en la región r , Casen 2015.

De forma análoga se estima el error relativo.

Cabe señalar que el error relativo de la Región de Magallanes alcanza 30,2%, lo cual se consideró un caso de excepción, justificado por la baja tasa de pobreza observada en 2015 (4,4%), lo que implicaría un aumento considerable del tamaño de muestra para lograr su disminución.

Por otra parte, se previó que los errores absolutos a nivel provincial no superaran 6,5%, dándose la excepción en la provincia de Parícuta, donde éste valor alcanzó 12,4%

³⁵ El efecto del diseño fue ajustado al rango de 1 a 5, redondeado a un decimal.

³⁶ Si m_{0r} es corregido por el efecto del diseño teórico (obtenido de Casen 2015 sin ajuste), el tamaño de muestra m_{1r} conserva los mismos errores absoluto y relativo definidos originalmente. Sin embargo, en caso de corregir m_{0r} por un efecto del diseño distinto al teórico, los errores definidos originalmente se modifican, por tanto, deben ser estimados nuevamente.

Paso 5

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

$$m_{2r} = \frac{Z_{1-\alpha/2}^2 \cdot S(p)_r^2 \cdot Def(p)_r}{d_{0r}^2 + Z_{1-\alpha/2}^2 \cdot Def(p)_r \cdot S(p)_r^2 / M_r} = \frac{m_{1r}}{1 + m_{1r}/M_r} \quad (5)$$

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

Paso 6

Posteriormente, se determinó el tamaño muestral objetivo de las comunas m_{2c} , para ello se distribuyó el total de viviendas regional de forma proporcional en las comunas de acuerdo a la distribución de la muestra objetivo comunal de Casen 2013, considerando también, la distribución de la muestra según área urbana y rural.

Se asignó un tamaño mínimo de muestra objetivo para el área urbana de 50 viviendas por comuna y para el área rural de 30 viviendas por comuna. En consecuencia, fue necesario aumentar el tamaño de muestra para algunas áreas en determinadas comunas, y disminuirlo en otras, siempre y cuando perteneciera a la misma región, de tal forma que se respetaran los tamaños de muestra regionales definidos en el paso 5 (m_{2r}). En ese sentido, se busca estabilizar las estimaciones de varianza o, dicho de otro modo, reducir la variabilidad de las estimaciones de la varianza.

En versiones anteriores de la encuesta, también se establecía un tamaño mínimo muestral de dos manzanas o secciones para áreas urbanas o rurales respectivamente, con lo cual se busca hacer factible la estimación de varianza bajo muestreo complejo, al evitar, por diseño, el caso de un solo conglomerado por estrato. Sin embargo, para esta nueva versión de la encuesta, dado que en la parte urbana existe una estratificación adicional por grupo de tamaño, esta condición no fue aplicada a este nivel de grupo de tamaño, pero sí a nivel de estrato geográfico comuna-área, como en las versiones anteriores. No obstante, esto será subsanado, al igual que lo realizado en los años anteriores, mediante la construcción de pseudo-estratos y pseudo-conglomerados para la estimación de la varianza.

Finalmente, con 50 viviendas en el área urbana, se asegura un mínimo de 5 manzanas (tomando en cuenta que en promedio se encuestan 8 viviendas por manzana). Y en el área rural, se asegura un mínimo de dos secciones, considerando que en esta área se encuestan en promedio 19 viviendas por sección

Paso 7

Una vez definido el total de unidades muestrales a encuestar para estimar la tasa de pobreza por ingresos, se debe tener en consideración que es posible no lograr el total de unidades por diversas razones, tales como rechazos, moradores ausentes, etc. Con la finalidad de salvaguardar la precisión de la estimación de la tasa de pobreza, se aplica un factor de corrección a la fórmula (5) con base en una estimación de la tasa de no respuesta, ya que ésta puede ser interpretada como una estimación de la probabilidad de responder de los hogares en las comunas seleccionadas³⁷.

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

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

En Casen 2015 se utilizó el promedio ponderado de la tasa de no respuesta de las encuestas Casen 2011 (submuestra2) y Casen 2013, a nivel comunal. Así, se obtuvo el tamaño muestral con sobremuestreo para cada comuna del país, y luego por agregación se determinó el tamaño a nivel regional.

Para Casen 2017 se utilizó un promedio ponderado³⁸ de la tasa de no respuesta de las encuestas Casen 2013 y 2015, a nivel comunal. Obteniendo un tamaño muestral sobredimensionado tanto para el área urbana como rural de cada comuna, y el de esta última, por agregación.

³⁷ Ver más detalles del marco conceptual de la no respuesta en encuestas de hogares en "Nonresponse in Household Surveys" (Groves and Couper, 1998).

³⁸ En este caso se ponderó por el número de viviendas elegibles dentro de la comuna.

Finalmente, al igual que en 2015, se determinó el tamaño de muestra regional por agregación. En el Gráfico III.1 se presenta la distribución de los promedios de tasas comunales de no respuesta, según región.

La fórmula que define el tamaño muestral final de la encuesta Casen 2017, que incorpora los ajustes conjuntos de efecto del diseño, finitud y tasa de no respuesta, implementada para cada comuna, según área urbana y rural del país, está dada por:

$$m_{3c} = \frac{m_{2c}}{1 - tnr_c} \quad (6)$$

Donde m_{3c} representa el número de viviendas final a encuestar en la comuna c , según área urbana y rural.

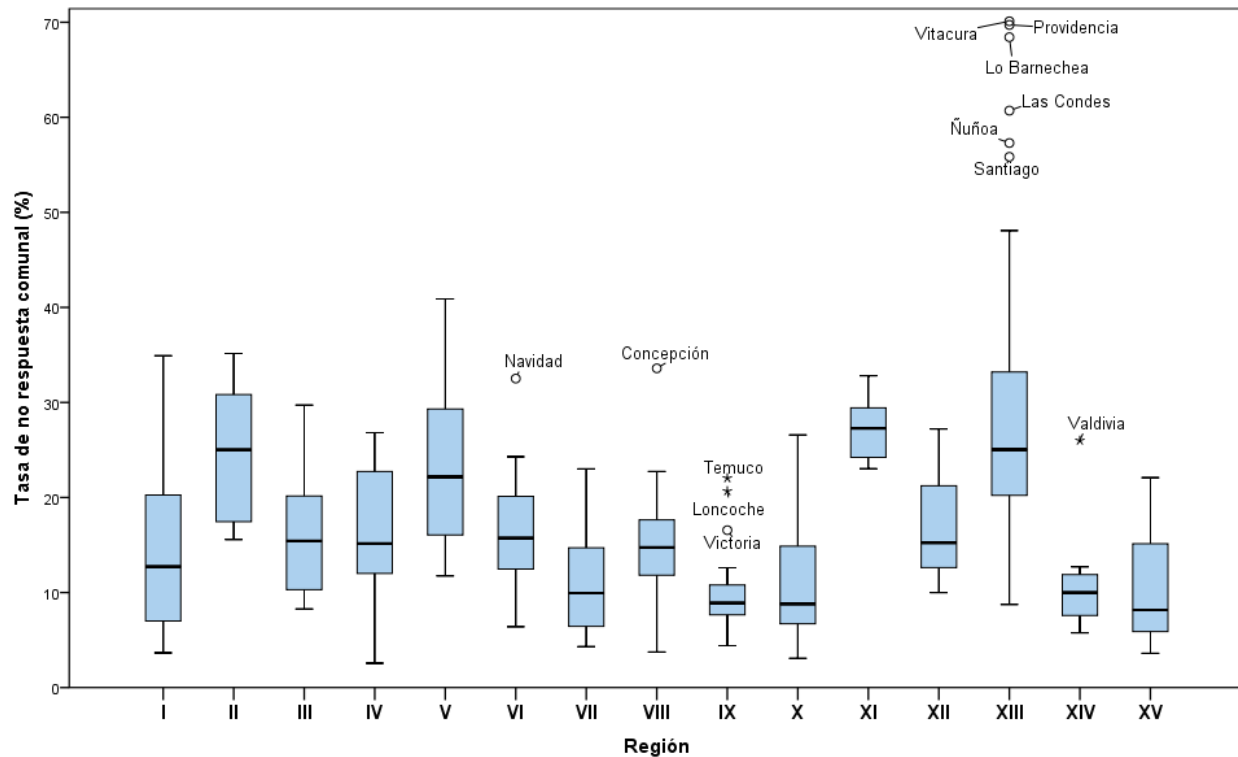
Paso 8

Cabe señalar que el tamaño muestral regional con sobre-muestreo m_{3r} , fue calculado como la suma o agregación comunal, según área urbana y rural, del total de viviendas en cada una de las comunas de la región que están presentes en la muestra de Casen 2017. Luego, de la expresión (6) se puede despejar la tasa de no respuesta y estimar la tasa regional a partir de la siguiente fórmula:

$$tnr_r = \frac{m_{3r} - m_{2r}}{m_{3r}} \quad (7)$$

En las columnas tnr_r y m_{3r} de la Tabla III.6 se presentan la tasa de no respuesta regional estimada y el tamaño muestral regional ajustado por tasa de no respuesta, correspondiente al sobredimensionamiento de la muestra.

Gráfico III.1. Distribución de promedios ponderados de las tasas de no respuesta comunales³⁹ de Casen 2013 y 2015, según región.



Elaborado por el Instituto Nacional de Estadísticas.

Las dos primeras columnas de la Tabla III.7 presentan el total de viviendas y su distribución según el marco muestral desarrollado en base al Censo de Población y Viviendas de 2002. Las columnas que siguen presentan el total de viviendas y la distribución de la muestra seleccionada para las encuestas Casen 2011, 2013⁴⁰, 2015 y 2017.

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

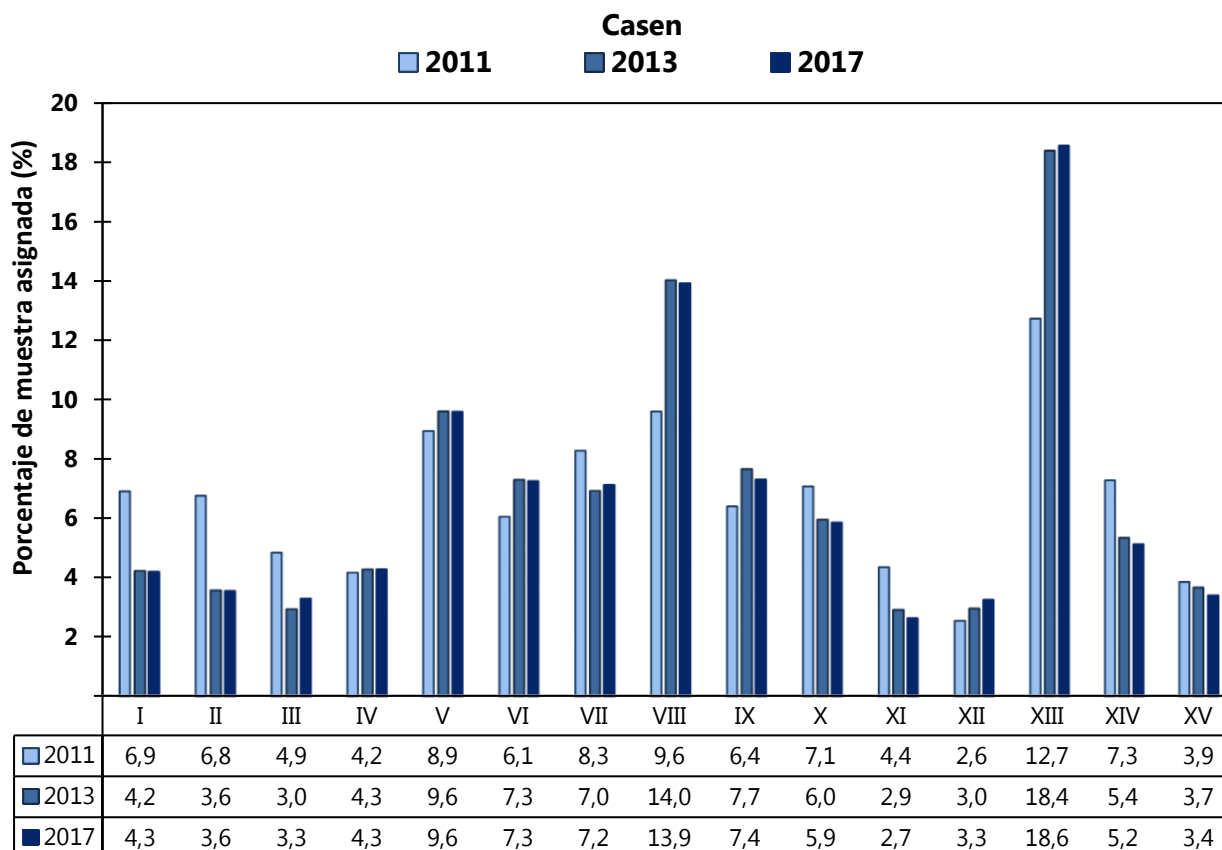
En la muestra 2015, dado que se establecieron determinadas provincias y comunas como dominios de estudios, originó que se incrementara la proporción muestral de regiones como Atacama, Valparaíso y Metropolitana, disminuyendo así, la de otras regiones.

³⁹ Para la tasa de no respuesta a nivel comunal, se utilizó el promedio ponderado entre las tasas de no respuesta comunales de Casen 2013 y 2015. En la figura, se muestra la distribución de estas tasas de no-respuesta comunales según región.

⁴⁰ En primera instancia, la muestra objetivo de Casen 2013 comprendía un total de 75.080 viviendas. Luego, a solicitud del Ministerio, ésta fue reducida a 70.080 viviendas.

Finalmente, el diseño de Casen 2017 busca recuperar el diseño de la versión 2013, por lo que, si se compara el tamaño muestral de las últimas versiones de la encuesta, es claro que la muestra de la encuesta Casen 2017 se asemeja más a la Casen 2013 (ver Gráfico III.2 o Tabla III.7).

Gráfico III.2. Distribución porcentual de la muestra objetivo de viviendas a nivel regional, para las encuestas Casen 2011, 2013 y 2017.



Elaborado por el Instituto Nacional de Estadísticas.

Tabla III.7. Distribución de la muestra objetivo de viviendas para las encuestas Casen 2011, 2013, 2015 y 2017.

Región	Censo 2002		Casen 2011		Casen 2013		Casen 2015 ⁴¹		Casen 2017	
	Viviendas en el Marco Muestral	%	Muestra	%	Muestra	%	Muestra	%	Muestra	%
País	4.594.668	100	90.122	100	70.080	100	82.370	100	69.816	100
Arica y Parinacota	43.429	0,9	3.495	3,9	2.581	3,7	838	1,0	2408	3,4
Tarapacá	71.684	1,6	6.241	6,9	2.977	4,2	2.428	2,9	2.974	4,3
Antofagasta	123.340	2,7	6.108	6,8	2.513	3,6	1.965	2,4	2.511	3,6
Atacama	68.332	1,5	4.382	4,9	2.071	3,0	3.861	4,7	2.331	3,3
Coquimbo	198.469	4,3	3.780	4,2	3.018	4,3	3.738	4,5	3.028	4,3
Valparaíso	491.287	10,7	8.064	8,9	6.737	9,6	8.847	10,7	6.717	9,6
Metropolitana	1.833.728	39,9	11.461	12,7	12.865	18,4	17.304	21,0	12.954	18,6
O'Higgins	244.622	5,3	5.469	6,1	5.137	7,3	7.144	8,7	5.099	7,3
Maule	288.879	6,3	7.464	8,3	4.872	7,0	5.678	6,9	5.007	7,2
Biobío	553.948	12,1	8.652	9,6	9.826	14,0	11.402	13,8	9.735	13,9
La Araucanía	267.546	5,8	5.782	6,4	5.387	7,7	6.846	8,3	5.136	7,4
Los Ríos	107.635	2,3	6.576	7,3	3.768	5,4	3.318	4,0	3.624	5,2
Los Lagos	228.440	5,0	6.391	7,1	4.192	6,0	6.097	7,4	4.129	5,9
Aysén	28.211	0,6	3.941	4,4	2.051	2,9	1.134	1,4	1.862	2,7
Magallanes	45.118	1,0	2.315	2,6	2.085	3,0	1.770	2,1	2.301	3,3

Elaborado por el Instituto Nacional de Estadísticas.

⁴¹ El diseño de Casen 2015 consideró 139 comunas auto representadas.

III.5. Selección de conglomerados de la muestra

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

En esta sección, se resumen los procedimientos utilizados en la selección de conglomerados, la cual se realiza de forma distinta dependiendo del marco de muestreo en que se encuentra registrada la parte urbana o rural de una comuna.

A continuación, se describen los protocolos utilizados para la selección de manzanas (a partir del MM2008) y de secciones (a partir del MS2002).

III.5.1. Selección de manzanas desde el MM2008

El MM2008 contiene información de las manzanas correspondientes al área urbana de cada comuna, las cuales se han estratificado en 5 grupos de tamaño, según la cantidad de viviendas que éstas integran.

Previo a la selección de manzanas dentro del área urbana de cada comuna, se determina la muestra de viviendas a seleccionar en cada grupo de tamaño. Para lo cual, se toma el tamaño de muestra urbana de la comuna y se distribuye entre los cinco⁴² grupos de tamaño, proporcionalmente a la cantidad de viviendas que tiene cada grupo, de esa comuna, dentro del marco.

Además, como a cada grupo de tamaño le ha sido asociado un número promedio⁴³ de viviendas a seleccionar por manzana, se calcula el total de manzanas a seleccionar a partir del cociente entre

⁴² Es posible que para alguna comuna no existan manzanas de cada grupo de tamaño, en este caso se distribuye entre los grupos de tamaño que tengan manzanas asociadas.

⁴³ Corresponde al promedio ponderado de viviendas a encuestar por manzana dentro del grupo. Para calcularlo, primero se suma el número de viviendas a encuestar por conglomerado (aproximadamente igual a 25% de las viviendas disponibles de la manzana, con un tope de treinta) y segundo, se divide esta suma ponderada por el total de manzanas que contiene el grupo.

la cantidad de viviendas asignadas al grupo de tamaño y el número promedio de viviendas a encuestar por manzana⁴⁴.

En este sentido, el promedio de viviendas que pueden ser seleccionadas por manzana, depende del grupo de tamaño al que ésta pertenece, y por ende, del número total de viviendas que tiene la manzana en el marco. En la Tabla III.8 se muestran los mínimos y los máximos de viviendas a encuestar por grupo de tamaño.

Tabla III.8. Grupos de tamaño de manzanas y promedio de viviendas a encuestar según grupo.

Grupo Tamaño	Manzanas con Número de viviendas entre		Viviendas a Encuestar (aprox. 25% con tope de 31)		Tamaño promedio de viviendas a encuestar por manzana dentro del grupo de tamaño
	Mínimo	Máximo	Mínimo	Máximo	
1	8	23	3	5	4
2	24	44	8	13	10
3	45	81	12	18	15
4	82	154	16	23	20
5	155	1.430	20	31	25

Elaborado por el Instituto Nacional de Estadísticas.

Determinado el número de manzanas a seleccionar por grupo de tamaño, éstas se seleccionan de forma sistemática y con igual probabilidad dentro de cada grupo, para lo cual, se ordenan las manzanas por el distrito censal, la zona censal y el número de viviendas en el marco.

La Tabla III.9 resume la distribución de la muestra de manzanas seleccionadas del MM2008 y viviendas a nivel regional por grupo de tamaño.

⁴⁴ Por ejemplo, si en el grupo de tamaño 2 de la Tabla III.8 corresponde levantar 94 viviendas con sobremuestreo, de acuerdo a la distribución según proporción de viviendas que contiene este grupo de tamaño, el número de manzanas a seleccionar en este grupo, será aproximadamente de $94/10 = 9,4 \approx 9$ manzanas.

Tabla III.9. Número de manzanas seleccionadas del MM2008 y muestra sobredimensionada de viviendas, según región y grupo de tamaño, Casen 2017.

Región	Nro. de manzanas en la muestra						Nro. de viviendas en la muestra					
	Total	Grupo de Tamaño					Total	Grupo de Tamaño				
		1	2	3	4	5		1	2	3	4	5
País	9.179	5.307	2.289	678	382	523	74.743	21.081	22.767	10.179	7.586	13.130
		(57,8%)	(24,9%)	(7,4%)	(4,2%)	(5,7%)		(28,2%)	(30,5%)	(13,6%)	(10,1%)	(17,6%)
Arica y Parinacota	384	263	74	23	9	15	2.702	1.055	741	345	171	390
Tarapacá	406	177	138	43	16	32	3.861	707	1.377	645	339	793
Antofagasta	466	283	119	36	12	16	3.465	1.130	1.183	534	234	384
Atacama	446	330	87	16	7	6	2.722	1.324	869	246	133	150
Coquimbo	430	277	119	19	9	6	2.903	1.100	1.182	288	177	156
Valparaíso	972	600	209	72	43	48	7.572	2.368	2.054	1.087	852	1.211
Metropolitana	2.000	824	524	206	151	295	21.958	3.258	5.227	3.087	2.988	7.398
O'Higgins	469	276	111	36	23	23	3.774	1.088	1.086	553	469	578
Maule	458	274	125	35	12	12	3.384	1.090	1.230	515	241	308
Biobío (Antigua)	1.092	655	278	81	42	36	8.326	2.605	2.758	1.219	841	903
Biobío	876	524	215	69	36	32	6.786	2.091	2.135	1.041	716	803
Ñuble	216	131	63	12	6	4	1.540	514	623	178	125	100
La Araucanía	535	371	121	21	13	9	3.496	1.467	1.212	315	262	240
Los Ríos	391	241	96	25	17	12	2.937	965	958	382	337	295
Los Lagos	410	249	108	31	13	9	3.008	989	1.089	454	257	219
Aysén	277	184	74	15	3	1	1.770	725	734	224	57	30
Magallanes	443	303	106	19	12	3	2.865	1.210	1.067	285	228	75

Elaborado por el Instituto Nacional de Estadísticas.

III.5.2. Selección de secciones desde el MS2002

El marco muestral de secciones está particionado en 6.164⁴⁵ secciones en la parte Rural (R) y 928 secciones en la parte RAU. El número de secciones a seleccionar tanto en la parte rural como en la RAU, dentro de cada comuna, se obtiene considerando un promedio de 20 viviendas por sección.

La selección preliminar de las secciones se realiza al interior de cada comuna en forma proporcional al total de viviendas ocupadas que éstas contienen, según información del Censo 2002. Esta selección se realiza en forma sistemática⁴⁶ proporcional al tamaño⁴⁷, mediante un

⁴⁵ Previo a la selección se realizan las exclusiones correspondientes, por ello el total de secciones difiere entre las versiones de la encuesta Casen 2011, 2013, 2015 y 2017.

⁴⁶ Las secciones son ordenadas por el número de viviendas que éstas contienen.

⁴⁷ En la práctica, mediante el módulo de muestras complejas de SPSS u otro software estadístico, la selección aleatoria en forma sistemática es muy fácil de implementar, y ciertamente el software aplica el mismo algoritmo de selección.

algoritmo de acumulación de viviendas en el estrato o comuna, en el que se elige un período y arranque aleatorio, como sigue:

Paso 1

Se construyen Q intervalos (equivalente al número de secciones del Estrato o comuna) de la manera como se muestra en la Tabla III.10:

Tabla III.10. Método de selección de unidades primarias secciones, Casen 2017.

Sección	N° Viviendas	Acumulado del número de viviendas	Límite inferior	Límite superior
i	M_i	ΣM_i	L_{inf}	L_{sup}
1	M_1	$\Sigma_1 = M_1$	1	Σ_1
2	M_2	$\Sigma_2 = M_1 + M_2$	$\Sigma_1 + 1$	Σ_2
3	M_3	$\Sigma_3 = M_1 + M_2 + M_3$	$\Sigma_2 + 1$	Σ_3
⋮	⋮	⋮	⋮	⋮
Q-1	M_{Q-1}	$\Sigma_{Q-1} = M_1 + M_2 + \dots + M_{Q-1}$	$\Sigma_{Q-2} + 1$	Σ_{Q-1}
Q	M_Q	$\Sigma_Q = M_1 + M_2 + \dots + M_Q$	$\Sigma_{Q-1} + 1$	Σ_Q

Elaborado por el Instituto Nacional de Estadísticas.

Paso 2

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

Paso 3

Se genera un número aleatorio⁴⁸ entero A entre 1 y el período K .

Paso 4

Enseguida, se va sumando sucesivamente el período K para obtener distintos valores que pertenecerán a los distintos intervalos que indican qué sección es la seleccionada. De acuerdo a

⁴⁸ Para la generación de este número aleatorio, se puede introducir una semilla para conservar la selección de los conglomerados y no cambien en caso que se replique el procedimiento.

este algoritmo, las secciones seleccionadas son aquellas cuyo intervalo asociado contiene a los valores dados por: $A, A + K, A + 2K, A + 3K, \dots, A + (n - 1)K$.

El procedimiento anterior, para secciones homogéneas en cuanto al tamaño (de una cantidad de viviendas similar), no permite repeticiones. Por el contrario, cuando las secciones son de tamaño muy distinto, llegando a ser algunas el doble y más del doble de las demás, el procedimiento o algoritmo selecciona menos de las requeridas, porque esencialmente puede seleccionar dos o más veces una misma sección, si ésta es demasiado grande respecto a las demás.

Finalmente, 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 $n \cdot M_i / \Sigma Q$.

La Tabla III.11 resume la distribución de la muestra de secciones seleccionadas del MS2002 y viviendas a nivel regional por área.

Tabla III.11. Número de secciones seleccionadas del MS2002 y muestra sobredimensionada de viviendas, según región y área, Casen 2017.

Región	Nro. de secciones en la muestra			Nro. de viviendas en la muestra		
	Total	Área		Total	Área	
		RAU	Rural		RAU	Rural
País	1.050	274 (26,1%)	776 (73,9%)	20.416	5.249 (25,7%)	15.167 (74,3%)
Arica y Parinacota	19	3	16	364	54	310
Tarapacá	17	3	14	345	65	280
Antofagasta	14	3	11	281	68	213
Atacama	17	0	17	326	0	326
Coquimbo	49	9	40	938	172	766
Valparaíso	87	16	71	1.726	327	1.399
Metropolitana	54	5	49	1.093	121	972
O'Higgins	121	31	90	2.420	608	1.812
Maule	130	49	81	2.494	909	1.585
Biobío (Antigua)	188	61	127	3.616	1.181	2.435
Biobío	96	24	72	1.803	447	1.356
Ñuble	92	37	55	1.813	734	1.079
La Araucanía	130	41	89	2.437	746	1.691
Los Ríos	75	9	66	1.482	173	1.309
Los Lagos	101	30	71	1.930	546	1.384
Aysén	36	11	25	736	219	517
Magallanes	12	3	9	228	60	168

Elaborado por el Instituto Nacional de Estadísticas.

III.6. Selección de viviendas de la muestra

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

A partir de los resultados del Precenso realizado durante el año 2016, se dispone de información actualizada acerca del total de viviendas existentes en cada manzana o sección presente en su marco correspondiente. Además, se ha actualizado la cartografía asociada a estos conglomerados, incluyendo las direcciones y el número que identifica cada vivienda.

En Casen 2017, una vez seleccionada la manzana o sección, ésta será actualizada mediante el proceso de verificación de direcciones en función de la información obtenida del Precenso 2016, permitiendo la obtención de una muestra probabilística de viviendas. No obstante, se dan casos, en su mayoría en zonas rurales, donde existen viviendas sin número, lo cual implica que, de ser seleccionada una manzana o sección para la cual no sea posible realizar el proceso de verificación de direcciones, debe ser actualizada su información correspondiente mediante la salida a terreno, llevando a cabo el procedimiento de enumeración de conglomerados.

A continuación, se resumen los procedimientos de verificación de direcciones a partir el Precenso, enumeración de los conglomerados y selección de las viviendas, sin embargo, el detalle de estas actividades se presenta más adelante en el capítulo IV.

III.6.1. Verificación de direcciones en los conglomerados de la muestra

El objetivo de este procedimiento, es la disminución de los errores no muestrales, detectando de manera oportuna, aquellos casos de la base de registros de Precenso 2016, que no cumplen con la coherencia y completitud necesaria; tanto para la totalidad de la información contenida en cada unidad primaria de muestreo seleccionada, como también de los antecedentes necesarios para la ubicación precisa de las viviendas que resulten en la selección de la muestra.

Así mismo, esta instancia de gabinete aprobará todas aquellas UPM que cumplan con las exigencias ya descritas y verificará en trabajo de campo⁴⁹ toda aquella información que no es posible de resolver con los antecedentes contenidos en la base de registros.

Los objetivos específicos de la verificación son:

1. Analizar y aprobar la coherencia y completitud de la información.
2. Aprobar en gabinete las UPM que cumplan con la identificación y ubicación de cada vivienda contenida en los registros.
3. Evitar la omisión de viviendas.
4. Identificar información duplicada.
5. Corrección de direcciones mal registrada.
6. Clasificación de vivienda y hogar.
7. Descripción de viviendas sin identificación domiciliaria.
8. Clasificar la carga a verificar en terreno.
9. Otras acciones orientadas a la calidad de la información.

III.6.2. Enumeración de los conglomerados de la muestra

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

La enumeración es una operación de registro de viviendas en un área geográfica específica, que tiene como finalidad identificar el universo en el cual se realizará la selección de viviendas. Para realizar esta tarea, cada enumerador debe cumplir con las etapas y objetivos del referido estudio. La enumeración realizada para la muestra de la encuesta Casen 2017, contempló áreas urbanas y rurales.

III.6.2.1. Material de trabajo del enumerador:

A cada enumerador se entrega el siguiente set de materiales:

⁴⁹ En el trabajo de terreno se deberá seguir los protocolos y procedimientos que se señalan en el proceso de enumeración y actualización de viviendas (Ver sección III.6.2. Enumeración de los conglomerados de la muestra).

1. Plano de ubicación comunal: Plano de la comuna donde se encuentran identificadas todas las manzanas y secciones seleccionadas en la muestra.
2. Croquis de la manzana o sección: Plano a menor escala que identifica sólo la manzana o sección seleccionada.
3. Formulario de enumeración urbano o rural: En él se registran todas las construcciones encontradas en la manzana o sección según el recorrido (orden en que fueron encontradas en terreno). Ver Formulario de Enumeración en Áreas Urbanas en Figura III.2 (los formularios de áreas rurales son similares).

III.6.2.2. Recorrido y enumeración urbana:

El enumerador debe confirmar que se encuentra en la manzana o sección seleccionada, verificándolo en el plano de ubicación y croquis respectivo. Una vez realizado esto, 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 unidades destinadas a otros usos, registrando los datos asociados a cada uno de ellos en los campos correspondientes del Formulario de Enumeración.

III.6.2.3. Recorrido y Enumeración Rural:

Al recorrer la localidad o entidad de la sección, se 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, se procede a anotar el número de orden de dicha vivienda y luego se marca un punto en el plano de ubicación de la vivienda, y junto a él se anota el mismo número de orden que se registró en el Formulario de Enumeración. La enumeración de las viviendas se debe realizar en forma correlativa dentro de cada localidad o entidad. Es decir, la numeración se realiza de 1 a N para cada una de las localidades que encuentre en el área rural definida en el respectivo plano.

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.

En la Enumeración Urbana y Rural, y cuando sea posible, el enumerador debe incorporar en el Formulario de Enumeración el número total de personas que compone cada hogar.

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

1. El conglomerado puede estar temporalmente destruido porque se van a levantar nuevas edificaciones, ya sean casas nuevas o edificios;
2. La mayoría de las viviendas cambiaron de uso o destino a otro que no es de vivienda;
3. Es imposible ingresar a las viviendas del conglomerado porque están dentro de un edificio que tiene un conserje que les impide el ingreso;
4. El conglomerado no tiene el número de viviendas mínimo requerido para la selección de las viviendas;
5. La estructura en terreno del conglomerado difiere de la enviada en el croquis, debido a que ésta pudo haberse fusionado con otra o dividido debido al cambio de límites por cierre o construcción de nuevas calles. Siendo este caso el más frecuente;
6. El conglomerado se encuentra afectado por algún tipo de catástrofe natural como aluviones, terremotos u otros.

En total, los motivos descritos incidieron en el reemplazo de 10,3% de las manzanas y 0,9% de las secciones, para un reemplazo total de 9,3% de unidades primarias originalmente seleccionadas, porcentaje que se ha incrementado con relación a versiones anteriores de la encuesta, principalmente, por el envejecimiento del marco muestral y agotamiento de las unidades primarias de muestreo que lo componen. Para reemplazo se busca, dentro del entorno geográfico, un conglomerado con características similares en cuanto al tamaño (número de viviendas) y en cuanto a la ubicación (la misma comuna, distrito y zona, o la misma comuna y distrito, o la misma comuna, dependiendo si al nivel más desagregado se encuentre una de tamaño similar). El reemplazo de manzanas/secciones en esta etapa del proceso es práctica habitual en muestras para encuestas de hogares desarrolladas por el INE.

Figura III.2. Formulario de enumeración en áreas urbanas, Casen 2017 (Formato impreso).



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

IDENTIFICACIÓN GEOGRÁFICA

Hoja N° ___ de ___

Cód. Comuna (RPC)

Distrito

Zona

Manzana

ID Manzana

Nombre de Comuna

N° de línea	N° de orden de la edificación	DIRECCIÓN				USO DE LA EDIFICACIÓN				N° DE ORDEN		N° DE PERSONAS EN EL HOGAR	ANOTACIONES COMPLEMENTARIAS Y OBSERVACIONES				
		Cód.	Nombre de dirección	N° de domicilio (Sin número SN)	N° o letra de block o torre	N° de piso	N° o letra del depto. o plaza	1.- Moradores ausentes 2.- Desocupada 3.- Temporada 4.- Colectiva 5.- Sin edificación 6.- Otro uso	Código otro uso 1.- Servicio público 2.- Establecimiento educacional 3.- Establecimiento de salud 4.- Establecimiento deportivo, recreacional, social y/o cultural 5.- Establecimiento religioso 6.- Establecimiento productivo, comercial y/o de servicios 7.- Otro	Código sin Viviendas B= Bandedón P= Plaza, plazoleta o parque S= Sitio errazo O= Otro	De la Vivienda		Del Hogar	Total (Incluir niños, bebés y ancianos)	Cód.	Descripción y/u observación	
1																	
2																	
3																	
4																	
5																	
6																	

Figura III.3. Formulario de enumeración en áreas rurales, Casen 2017 (Dispositivo móvil).

Nueva Registro Rural

Latitud: -3837577.75
 Longitud: -7934226.50

Localidad: CATAPILCO

Entidad: POBLACION LUIS MENA

Categoría: 123 Caserío

Orden Edificación: 32 VER VIVIENDA EN MAPA Orden de Vivienda: 77

Edificios: Edificio CATAPILCO

Tipo de Calle: Avenida

Nombre de Calle: LUIS PORTO SEGURO

Número de Domicilio: 466 Block: Block

Nueva Registro Rural

N° Domicilio: [] Block: []

Piso: [] Departamento: N° Depto []

Uso o destino de Edificación: Moradores presentes

Código Anotación: 123 Sin anotaciones

Orientación de Vivienda en Ruta: Izquierda Derecha

Ubicación Horizontal: Izquierda Ubicación Vertical: Adelante

Descripción u Observación: Ingrese Observación o Descripción SUGERENCIAS

Agregar Hogares

Hogar 1: [] 1 [X]

Hogar 2: [] 3 [X]

III.6.3. 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 y secciones, la selección de viviendas se realiza al interior de cada una de las seleccionadas previamente, en forma sistemática, con igual probabilidad. Específicamente, al algoritmo de selección sistemático de viviendas, que es aplicado también en otras encuestas, se introduce una semilla de aleatorización al igual que para la selección de manzanas, para que, independientemente de quien aplique el mismo algoritmo, bajo las mismas condiciones de ordenamiento previo⁵⁰, seleccione las mismas viviendas.

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

Paso 1

Se calcula el período $K = M_i / m_i$.

Paso 2

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

Paso 3

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

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

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 $A + (m_i - 1)K$.

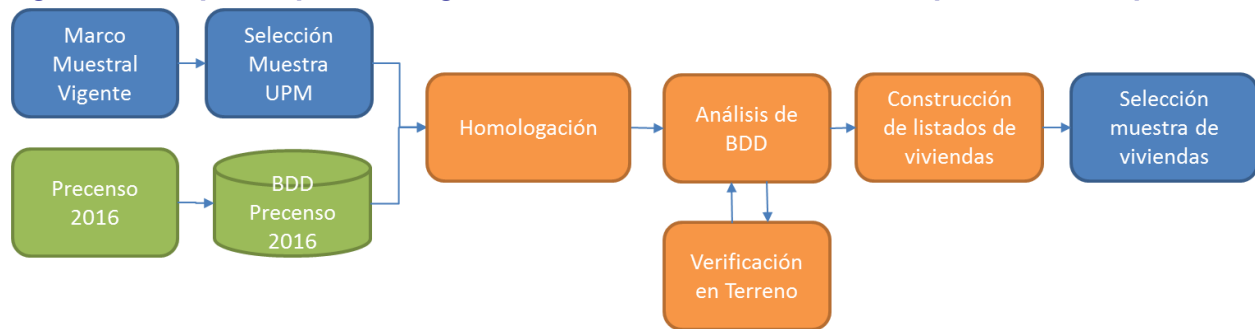
⁵⁰ Tanto en manzanas como en secciones las viviendas son ordenadas según la variable "orden vivienda".

IV. VERIFICACIÓN Y ACTUALIZACIÓN DE UNIDADES DE PRIMERA ETAPA DE LA MUESTRA

Una vez realizada la selección de la muestra de unidades de primera etapa (manzanas y secciones) desde el marco de manzanas y secciones, es necesario realizar la selección de la muestra de viviendas. Sin embargo, previamente se elabora y/o actualiza el listado de viviendas correspondientes a cada manzana y sección.

En esta versión de Casen 2017 se utilizó para la construcción de los listados de viviendas, así como para la actualización de la cartografía, la información precedente del Precenso 2016, la cual fue homologada, analizada y complementada con antecedentes de terreno para ajustarse al marco muestral de primera etapa de selección y a los requerimientos de identificación de viviendas para la selección de muestras. Una vez realizadas estas actividades, se construyeron los listados de viviendas particulares ocupadas y se realizó la selección de la muestra de viviendas. La Figura IV.1 muestra un detalle de las principales etapas realizadas para la elaboración de los listados de viviendas, las cuales se describen a continuación.

Figura IV.1. Etapas del proceso de generación de los listados de viviendas particulares ocupadas.



Elaborado por el Instituto Nacional de Estadísticas.

IV.1. Homologación

El Precenso 2016 capturó información utilizando definiciones e información cartográfica actualizada, no obstante, para efectos de este proyecto, deben primar necesariamente las definiciones del marco vigente, a modo de asegurar la comparabilidad del estudio con sus versiones anteriores. De esta manera, es necesario evaluar mediante procesos semi automatizados si las unidades de primera etapa del marco muestral son identificables en la cartografía actualizada

de Precenso, a modo de establecer códigos de homologación que permitan vincular las viviendas de Precenso a una manzana o sección seleccionada dentro de la muestra, así como establecer con precisión los casos en los cuales esto no es posible y que requerirán obligatoriamente visitas a terreno para capturar la información complementaria necesaria.

La Figura IV.2 muestra algunas de las diferencias entre las definiciones del marco muestral actualmente vigente y Precenso 2016, que obligan a realizar un procesamiento para utilizar la base de datos.

Figura IV.2. Comparativa entre actualización de definiciones marco muestral vigente y Precenso 2016.



Elaborado por el Instituto Nacional de Estadísticas.

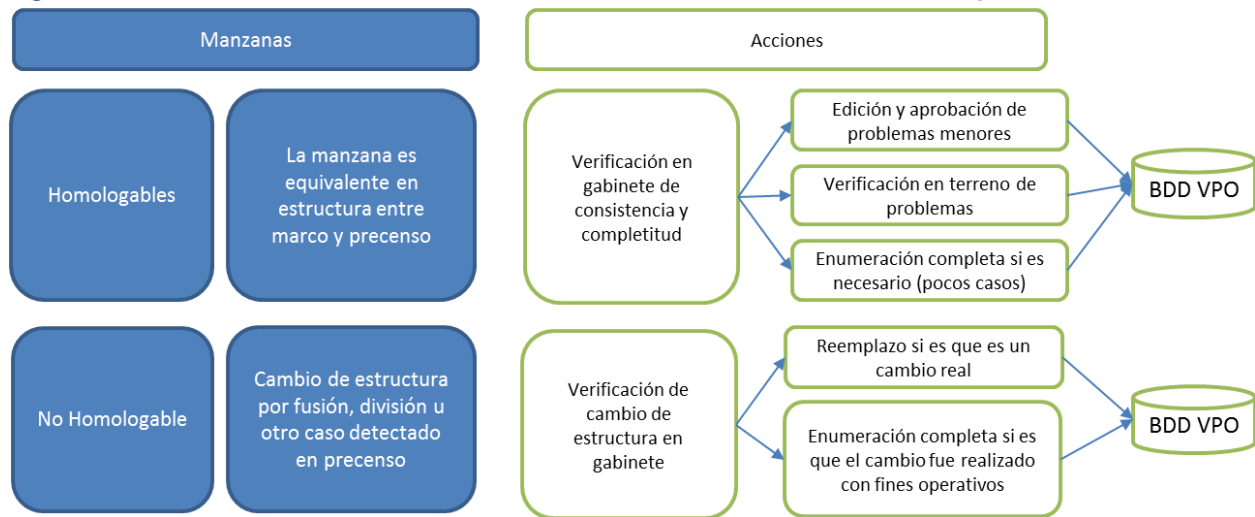
IV.1.1. Manzanas

En el caso de las manzanas, se consideró homologables los casos en que la cartografía de Precenso 2016 y la estructura de las manzanas definida en el marco muestral tuviesen una coincidencia que permitiera realizar una vinculación univoca de 1 manzana de Precenso a 1 manzana de marco muestral. Se definieron como no homologables los casos en que fuera factible asociar 2 o más manzanas de Precenso a una del marco muestral o bien, los casos en que una manzana de Precenso coincidiese con varias manzanas en el marco muestral. En estos casos, se evaluó si estas diferencias correspondían a cambios de estructura (fusiones o divisiones) o si correspondían a diferencias en la aplicación de criterios para la definición de manzanas. En caso de fusiones y

divisiones las manzanas fueron reemplazadas por otras, siguiendo el procedimiento habitual. En caso de diferencias de criterio, se determinó que las manzanas serían enumeradas nuevamente de acuerdo a la cartografía del marco muestral.

Las manzanas homologables pasaron directamente a la etapa de análisis.

Figura IV.3. Resumen actividades de verificación de información de Precenso para zonas urbanas.



Elaborado por el Instituto Nacional de Estadísticas.

IV.1.2. Secciones en Resto de Áreas Urbanas (RAU)

En el caso de las secciones urbanas, se realizó un trabajo de digitalización de la cartografía análoga del marco de secciones a modo de poder realizar la homologación. Un aspecto relevante es que las manzanas existentes dentro de las secciones RAU no cuentan con códigos de identificación únicos, siendo enumeradas correlativamente de 1 a n siguiendo un orden de recorrido (serpenteante), en tanto, las manzanas levantadas en Precenso sí cuentan con estos códigos de identificación, pero no cuentan con las capas de información que permitan agruparlas de acuerdo a la estructura de sección del marco muestral, por cuanto existen solo como unidades independientes. De esta manera:

1. Se consideraron homologables a aquellas secciones que hubiesen sido levantadas con protocolos urbanos en Precenso. En este tipo de casos fue factible homologar la cartografía, no obstante, por los tiempos necesarios para vincular los listados de viviendas de cada manzana de Precenso a una sección, lo que debía realizarse manualmente dada la inexistencia de códigos únicos en el marco muestral, se optó por enumerar nuevamente a

modo de no retrasar los trabajos de terreno, ya que, se estimó que de todas formas era muy probable que debiesen visitarse una parte importante de las secciones a modo de verificar antecedentes y recolectar información complementaria. La enumeración de estos casos se realizó con dispositivos móviles, lo que permitió georreferenciar las viviendas.

2. Se consideraron no homologables a aquellas secciones que hubiesen sido levantadas total o parcialmente con protocolos rurales en Precenso, es decir, careciesen total o parcialmente de manzanas definidas. En este tipo de situaciones, se generó una capa de información para identificar las manzanas de la sección y diferenciar aquellas que habían sido levantadas como manzanas urbanas en Precenso, de aquellas que habían sido levantadas como rurales, determinando que toda aquella información faltante debía ser levantada en terreno, por cuanto todas las áreas enumeradas como rurales en Precenso fueron enumeradas nuevamente con protocolos urbanos. En estos casos, el trabajo necesario para identificar zonas levantadas como rurales, incluyó la generación de una capa adicional de información, la que facilitó la homologación de las bases de datos, por lo que fue factible importar con reducidos niveles de esfuerzo las bases de datos de Precenso. La enumeración de estos casos se realizó con dispositivos móviles, lo que permitió georreferenciar las viviendas.

IV.1.3. Secciones Rurales

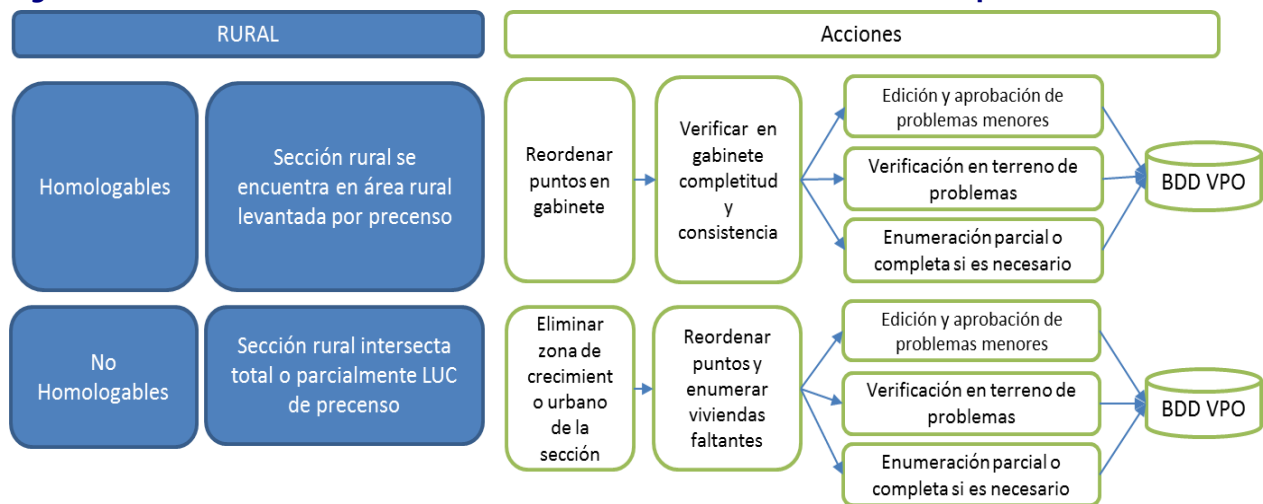
En el caso de las secciones rurales, se realizó un trabajo de digitalización de la cartografía análoga del marco de secciones a modo de poder realizar la homologación. La homologación en este caso enfrenta la dificultad de que Precenso no considera en su cartografía la definición de sección, sino que realiza el empadronamiento de las viviendas por zonas, localidades y entidades, numerando las viviendas de 1 a n por entidad. Pese a esta dificultad, existe la ventaja de que Precenso utilizó dispositivos móviles y georreferenció las viviendas y puntos de interés disponibles en cada sección. De esta manera, al vincular la geometría de la sección con los puntos capturados en Precenso, es factible filtrar y visualizar las viviendas que se encuentran dentro de la sección, cortando a su vez zonas, localidades y entidades levantadas en Precenso y siendo necesario reordenar los puntos visibles de acuerdo a la lógica de recorrido de la sección rural de 1 a n dentro de la sección. De esta manera:

1. Se consideraron homologables a aquellas secciones que hubiesen sido levantadas con protocolos rurales en Precenso. Estos casos pasaron directo a la etapa de análisis.

- Se consideraron no homologables a aquellas secciones que hubiesen sido levantadas total o parcialmente con protocolos urbanos en Precenso, es decir, careciesen total o parcialmente de puntos georreferenciados que permitieran saber si estaban dentro o fuera de la sección. En este tipo de situaciones, se generó una capa de información para identificar las zonas donde Precenso enumeró con protocolos urbanos y no existían puntos georreferenciados para las viviendas. En estos casos se procedió a definir zonas de levantamiento, las cuales fueron enumeradas nuevamente con dispositivos móviles para capturar cualquier vivienda o punto de interés existente y asignarle un punto georreferenciado. Si existen áreas de crecimiento urbano evidente en la sección, éstos fueron descartados (siguiendo protocolo de versiones anteriores).

Con la información capturada, estas secciones pasaron a la etapa de análisis.

Figura IV.4. Resumen actividades de verificación de información de Precenso para zonas rurales.



Elaborado por el Instituto Nacional de Estadísticas.

IV.2. Análisis

La etapa de análisis tiene por objeto realizar una evaluación y depuración a nivel de microdatos de la información disponible en la base de datos. Para el desarrollo de esta actividad se constituyeron manuales que sirvieron de apoyo al equipo de analistas, mediante los cuales se establecieron los protocolos que facilitaron la homologación de criterios dentro del equipo.

A continuación, se detallan los principales criterios utilizados para este análisis, de acuerdo a si la unidad de primera etapa pertenecía a uno u otro estrato.

IV.2.1. Manzanas

El análisis de las bases de datos de viviendas asociadas a cada manzana se realizó de manera previa a la realización de visitas a terreno para verificar, depurar y editar los siguientes aspectos:

1. Coherencia y completitud de base de datos, comparando la base de datos de cada manzana con el croquis F1⁵¹ actualizado de cada manzana enumerada en Precenso. Esto para verificar la correcta homologación, así como la coherencia y cambios en los nombres de las calles, ordenamiento de las viviendas dentro de la manzana (de acuerdo al orden de recorrido estandarizado), la enumeración de todas las caras de la manzana, entre otros.
2. Completitud de información para identificar viviendas con igual número domiciliario (ubicadas en el mismo sitio), sin identificación domiciliaria o con identificación no visible.
3. Completitud de información para identificar casos de numeraciones irregulares.
4. Casos de viviendas con elevado número de hogares.
5. Coherencia de registro de cada vivienda.

Como resultado de esta etapa, se determinó que aquellas manzanas con alguno de estos aspectos que no pudiese ser depurado con la información disponible de la base de datos, se visitarán en terreno para verificar o complementar los antecedentes faltantes.

Una vez que las bases de datos de cada manzana lograron cumplir con todos los criterios definidos se consideraron aptas para pasar a la construcción de los listados de viviendas. Adicionalmente, como resultado de esta revisión, se determinaron casos que requiriesen algún tipo de actualización de la cartografía asociada.

⁵¹ Corresponde a un plano que señala el área de cada manzana seleccionada, es decir, de las calles u otros elementos que la componen (ríos, esteros, etc.) En él se indica el punto de inicio del recorrido, además de algunas observaciones registradas por los enumeradores de Precenso, respecto a la cantidad de viviendas, alturas domiciliarias y cantidad de personas, entre otras. Ver Anexo N°3 con ejemplo del croquis de la manzana (F1).

IV.2.2. Secciones Resto de Áreas Urbanas (RAU)

El análisis de estas secciones se realizó una vez realizadas las visitas a terreno para recolectar la información faltante y consta de dos etapas.

Etapas 1: Se cargó en un visualizador la geometría de la sección, manzanas, mapas base, los puntos georreferenciados y dibujos. Se ordenaron las manzanas de la sección de acuerdo a un orden de recorrido por serpenteo, dividiendo o fusionando aquellas manzanas que según los antecedentes de terreno tuviesen modificaciones. Una vez que las manzanas fueron organizadas, se asociaron los puntos correspondientes a las viviendas levantadas a una manzana y se verificó su correcto ordenamiento de las viviendas de 1 a n de acuerdo al recorrido, procediendo posteriormente a cargar las bases de datos de viviendas de Precenso asociadas a cada manzana dentro de la sección.

Etapas 2: Se procedió al análisis de cada manzana, siguiendo criterios equivalentes a los expuestos para el análisis de las manzanas. Adicionalmente, se procedió a limpiar casos que pudiesen duplicarse al cargar la base de datos de Precenso y la base de datos de enumeración nueva.

IV.2.3. Secciones Rurales

El análisis de estas secciones se realizó en tres etapas.

Etapas 1: Se realizó un análisis de la base de datos para identificar viviendas sin observaciones de terreno que permitiesen facilitar su identificación. Las secciones con mayor proporción de observaciones faltantes se visitaron en terreno hasta el límite de la capacidad instalada disponible, con la instrucción de verificar todos los puntos levantados en la sección.

Las siguientes etapas de análisis se realizaron una vez realizadas las visitas a terreno para recolectar la información faltante.

Etapas 2: Se cargó en un visualizador la geometría de la sección, localidades, mapas base y los puntos georreferenciados y se ordenaron localidades de 1 a n dentro de la sección, luego de lo cual se ordenó la numeración de las viviendas de 1 a n dentro de la localidad, de acuerdo al orden de recorrido que los enumeradores deberían haber seguido. Este ordenamiento permite posteriormente ordenar las viviendas de la sección de 1 a n. En la práctica, debieron editarse todos los órdenes procedentes de terreno, dado que los puntos procedentes de Precenso contaban con

órdenes establecidos de 1 a n dentro de entidades, muchas de las cuales fueron cortadas al realizar el cruce entre los puntos y el polígono de la sección.

Etapas 3: Se procedió al análisis de la base de datos de cada sección para depurar y editar casos con problemas de completitud de información para identificar viviendas con igual número domiciliario (ubicadas en el mismo sitio), sin identificación domiciliaria o con identificación no visible, así como para estandarizar observaciones.

IV.3. Verificación en terreno

Para la conformación de los listados de viviendas, se contempló que en los casos en los que de acuerdo a los criterios antes expuestos, fuese necesario verificar en terreno o enumerar nuevamente, esta actividad sería ejecutada por equipos contratados en cada Dirección Regional del INE. Para el desarrollo de esta actividad se constituyeron manuales que sirvieron de apoyo al equipo de verificadores, mediante los cuales se establecieron los protocolos que facilitaron la homologación de criterios dentro del equipo.

A continuación, se detallan los principales aspectos relativos a las actividades de terreno realizadas.

IV.3.1. Manzanas

Se visitaron en terreno todas las manzanas que no pudieron ser homologadas, las cuales fueron enumeradas nuevamente. Adicionalmente, se visitaron todas las manzanas que de acuerdo a la evaluación inicial requerían de información complementaria para la identificación de las viviendas.

Para la realización de estas actividades, se proporcionó a cada verificador el siguiente set de materiales:

1. Plano de ubicación comunal: Plano de la comuna donde se encuentran identificadas todas las manzanas y secciones seleccionadas en la muestra.
2. Croquis de la manzana o sección F1 utilizado en Precenso: Plano a menor escala que identifica sólo la manzana o sección seleccionada.

3. Formulario de enumeración y verificación urbano: En él se identificaban todas las viviendas enumeradas en Precenso, identificando aquellas viviendas que requerían antecedentes complementarios.

IV.3.2. Secciones RAU

Se visitaron en terreno todas las secciones RAU, algunas de las cuales fueron enumeradas nuevamente de forma completa y otras solo en aquellas zonas con información levantada de acuerdo a protocolos rurales, según lo expuesto previamente.

Para la realización de estas actividades, se proporcionó a cada verificador el siguiente set de materiales:

1. Tablet: Contenía una aplicación en la cual era posible visualizar la cartografía digital, los límites de la sección y áreas de levantamiento (sin información), así como incorporar y editar sobre dicha cartografía, edificaciones, viviendas y otros usos identificados en terreno.
2. Carcasa protectora, arnés de soporte y batería adicional: Se consideraron estos elementos para facilitar la manipulación y seguridad en el uso del dispositivo.

IV.3.3. Secciones Rurales

Se visitaron en terreno todas las secciones rurales con problemas en su homologación, para el empadronamiento de las viviendas sin punto georreferenciado disponible. De igual forma, se visitaron todas aquellas secciones rurales con ausencia de observaciones de ubicación de las viviendas o con una proporción de viviendas con observaciones para su identificación inferior al 50% de las viviendas disponibles.

Para la realización de estas actividades a cada verificador se entregaron los siguientes materiales:

1. Tablet: Contenía una aplicación en la cual era posible visualizar la cartografía digital, incluyendo los límites de la sección, localidades, puntos georreferenciados y áreas de levantamiento (sin puntos georreferenciados), así como incorporar y editar sobre dicha cartografía, edificaciones, viviendas y otros usos identificados en terreno.

2. Carcasa protectora, arnés de soporte y batería adicional: Se consideraron estos elementos para facilitar la manipulación y seguridad en el uso del dispositivo.

IV.3.4. Conformación de los listados de viviendas

Una vez finalizado el análisis de todas las unidades de primera etapa, se procede a la exportación y envío a selección de la muestra de viviendas de una base de datos que contiene la totalidad de las viviendas particulares ocupadas y que excluye los registros asociados a viviendas desocupadas, de temporada, colectivas, edificaciones de otros usos, edificaciones en construcción y sitios sin edificación.

IV.4. Resumen operativo de las actividades de enumeración, verificación y recopilación de información complementaria en terreno

IV.4.1. Muestra de manzanas

Como se muestra en la Tabla IV.1, se visitó en terreno a nivel nacional 48,1% de las manzanas pertenecientes a la muestra total, las cuales también fueron analizadas y trabajadas en gabinete. El restante 51,9% se trabajó solo en gabinete, verificando, depurando y editando la información de Precenso 2016. A nivel de regiones, la cantidad gestionada en terreno fue heterogénea, siendo las regiones de Tarapacá (67,2%), Los Lagos (64,2%) y Biobío (59,9%) las más visitadas, en tanto las regiones donde fue visitada una menor proporción de manzanas fueron Atacama (24,9%), Arica y Parinacota (32,3%) y Coquimbo (36,5%).

Tabla IV.1. Distribución de las manzanas gestionadas en terreno, por tipo de trabajo de campo, Casen 2017.

Región	Muestra Gestionada	Casos de manzanas enviadas a terreno					
		Total gestionado en terreno y gabinete		Verificación de datos en terreno		Enumeradas desde cero	
		Cantidad	Porcentaje del total de la muestra gestionada	Cantidad	Porcentaje del total de la muestra gestionada	Cantidad	Porcentaje del total de la muestra gestionada
País	9.207	4.428	48,1%	4.108	44,6%	320	3,5%
Arica y Parinacota	384	124	32,3%	119	31,0%	5	1,3%
Tarapacá	405	272	67,2%	268	66,2%	4	1,0%
Antofagasta	465	266	57,2%	254	54,6%	12	2,6%
Atacama	450	112	24,9%	96	21,3%	16	3,6%
Coquimbo	430	157	36,5%	152	35,3%	5	1,2%
Valparaíso	975	471	48,3%	406	41,6%	65	6,7%
Metropolitana	2.005	872	43,5%	838	41,8%	34	1,7%
O'Higgins	471	180	38,2%	141	29,9%	39	8,3%
Maule	462	190	41,1%	159	34,4%	31	6,7%
Biobío (Antigua)	1.091	653	59,9%	607	55,6%	46	4,2%
La Araucanía	537	278	51,8%	265	49,3%	13	2,4%
Los Ríos	397	226	56,9%	212	53,4%	14	3,5%
Los Lagos	411	264	64,2%	246	59,9%	18	4,4%
Aysén	279	133	47,7%	124	44,4%	9	3,2%
Magallanes	445	230	51,7%	221	49,7%	9	2,0%

Elaborado por el Instituto Nacional de Estadísticas.

El 48,1% de manzanas visitadas, se compone de 44,6% de manzanas visitadas para la verificación de datos y recolección de información complementaria y el restante 3,5% corresponde a manzanas enumeradas nuevamente.

Los casos de verificación de manzanas en terreno y recolección de información complementaria, corresponden en 96,1% de las oportunidades, a casos de viviendas sin observaciones complementarias en circunstancias en que es requerido, tales como viviendas sin numeración y viviendas ubicadas dentro del mismo sitio con una idéntica numeración domiciliar producto de esta situación, siendo necesario una observación que ubique y describa la vivienda. Las regiones con una mayor proporción de manzanas visitadas para verificar y recolectar información complementaria, respecto al total de la muestra, fueron Tarapacá (66,2%), Los Lagos (59,9%) y Biobío (55,6%). Por el contrario, las regiones donde fueron visitadas una menor proporción de manzanas fueron las regiones de Atacama (21,3%), O'Higgins (29,9%) y Arica y Parinacota (36,5%).

Los casos en que se requirió enumerar corresponden en su totalidad a casos con dificultades para la homologación de las manzanas entre el marco muestral vigente y la cartografía de Precenso

2016, debido a que, una manzana del marco correspondía a dos o más manzanas de Precenso, o por el contrario, una manzana de Precenso correspondía a varias del marco muestral vigente, siendo estas diferencias producidas por diferencias en la aplicación de ciertos criterios, no obstante la situación física de las manzanas en cuanto a estructura no había sufrido alteraciones en el tiempo transcurrido entre las fechas de actualización de ambas cartografías. De esta manera, se optó por enumerar nuevamente estos casos aplicando el criterio de definición de las manzanas de acuerdo a lo vigente en el marco muestral. Las regiones con una mayor proporción de manzanas visitadas para enumerar nuevamente, respecto al total de la muestra, fueron O'Higgins (8,3%), Maule (6,7%) y Valparaíso (6,7%). En cambio, las regiones donde fueron visitadas una menor proporción de manzanas, fueron Tarapacá (1,0%), Coquimbo (1,2%) y Arica y Parinacota (1,3%).

IV.4.2. Muestra de secciones

Como se muestra en la Tabla IV.2, se visitó en terreno 66,9% de las secciones, considerando tanto secciones RAU como Rurales, las cuales posteriormente fueron analizadas en gabinete, depuradas y editadas según fuese pertinente. En tanto el restante 33,1% de las secciones solo se trabajó en gabinete.

Las regiones donde se visitó en terreno una mayor proporción de secciones fueron las regiones de Maule (88,5%), Magallanes (83,3%) y Araucanía (80,0%), en cambio, aquellas regiones donde fue visitada una menor proporción de secciones fueron Coquimbo (36,7%), Tarapacá (47,1%) y Los Ríos (49,3%).

Tabla IV.2. Distribución de las secciones gestionadas, por tipo de trabajo de campo, Casen 2017.

Región	Secciones trabajadas				
	Muestra gestionada	Solo trabajadas en gabinete		Trabajadas en terreno y en gabinete	
		Cantidad	Porcentaje	Cantidad	Porcentaje
País	1.050	348	33,1%	702	66,9%
Arica y Parinacota	19	7	36,8%	12	63,2%
Tarapacá	17	9	52,9%	8	47,1%
Antofagasta	14	2	14,3%	12	85,7%
Atacama	17	8	47,1%	9	52,9%
Coquimbo	49	31	63,3%	18	36,7%
Valparaíso	87	33	37,9%	54	62,1%
Metropolitana	54	27	50,0%	27	50,0%
O'Higgins	121	61	50,4%	60	49,6%
Maule	130	15	11,5%	115	88,5%
Biobío (Antigua)	188	48	25,5%	140	74,5%
La Araucanía	130	26	20,0%	104	80,0%
Los Ríos	75	38	50,7%	37	49,3%
Los Lagos	101	24	23,8%	77	76,2%
Aysén	36	17	47,2%	19	52,8%
Magallanes	12	2	16,7%	10	83,3%

Elaborado por el Instituto Nacional de Estadísticas.

Las secciones que se visitaron recibieron diferentes gestiones según fuese necesario, las cuales se detallan en la Tabla IV.3.

Como es posible observar, a nivel nacional se enumeró nuevamente 7,8% de las secciones, siendo todos estos casos correspondientes a (i) secciones de cualquier tipo con áreas afectadas por incendios, o (ii) secciones RAU de las cuales en Precenso se enumeraron completamente con protocolos urbanos, no obstante, debido a que el proceso de homologación debía realizarse manualmente, se optó por enumerar nuevamente con el objeto de ahorrar tiempo. Las regiones donde se enumeró nuevamente una mayor proporción de secciones fueron las de Magallanes (16,7%), Tarapacá (11,8%) y Maule (11,5%). Por el contrario, aquellas regiones donde se enumeró nuevamente menos secciones o bien no se visitaron por esta razón, fueron las regiones de Atacama (0,0%), Coquimbo (2,0%) y Los Ríos (2,7%).

Adicionalmente, se visitó 25,9% de las secciones con el objeto de enumerar áreas de levantamiento específicas en las cuales no se contaba con información por alguna de las siguientes razones: (i) Una sección RAU en el marco muestral fue enumerada en una o más áreas durante Precenso siguiendo protocolo rural, por cuanto, la información no seguía el orden de recorrido urbano, o (ii) Una sección rural fue enumerada en una o más áreas siguiendo protocolo urbano, por lo cual

no contaba con información de viviendas georreferenciada, necesaria para la creación de los planos. Las regiones en las cuales se visitó una mayor proporción de secciones por esta causa fueron las regiones de Aysén (36,1%), O'Higgins (35,5%) y Valparaíso (31,0%), mientras que aquellas en las que se debió visitar una menor proporción de secciones por este motivo fueron las regiones de Atacama (5,9%), Arica y Parinacota (10,5%) y Tarapacá (11,8%).

Tabla IV.3. Distribución de las secciones gestionadas en terreno, por tipo de trabajo de campo, Casen 2017.

Región	Secciones trabajadas en terreno y en gabinete								
	Total visitado en terreno	Secciones enumeradas completamente		Secciones enumeradas en área de levantamiento		Sección verificada completamente		Secciones enumeradas en área de levantamiento y verificadas completamente	
		Cantidad	%	Cantidad	%	Cantidad	%	Cantidad	%
País	702	78	7,4%	272	25,9%	276	26,3%	76	7,2%
Arica y Parinacota	12	1	5,3%	2	10,5%	8	42,1%	1	5,3%
Tarapacá	8	2	11,8%	2	11,8%	3	17,6%	1	5,9%
Antofagasta	12	1	7,1%	2	14,3%	6	42,9%	3	21,4%
Atacama	9	0	0,0%	1	5,9%	7	41,2%	1	5,9%
Coquimbo	18	1	2,0%	10	20,4%	5	10,2%	2	4,1%
Valparaíso	54	4	4,6%	27	31,0%	13	14,9%	10	11,5%
Metropolitana	27	2	3,7%	14	25,9%	7	13,0%	4	7,4%
O'Higgins	60	6	5,0%	43	35,5%	10	8,3%	1	0,8%
Maule	115	15	11,5%	31	23,8%	50	38,5%	19	14,6%
Biobío (Antigua)	140	20	10,6%	51	27,1%	51	27,1%	18	9,6%
La Araucanía	104	14	10,8%	26	20,0%	53	40,8%	11	8,5%
Los Ríos	37	2	2,7%	17	22,7%	18	24,0%	0	0,0%
Los Lagos	77	6	5,9%	30	29,7%	37	36,6%	4	4,0%
Aysén	19	2	5,6%	13	36,1%	4	11,1%	0	0,0%
Magallanes	10	2	16,7%	3	25,0%	4	33,3%	1	8,3%

Elaborado por el Instituto Nacional de Estadísticas.

También, se visitó 26,3% de los casos con el objeto de verificar las secciones completamente, lo cual corresponde principalmente a secciones rurales en el marco muestral enumeradas como rurales en Precenso, pero que 50% o más de las viviendas presentan ausencia de observaciones de ubicación y/o descripción de las viviendas, por lo que se optó por visitar y recorrerlas completamente para recopilar ese antecedente complementario. Adicionalmente, una proporción menor de estos casos corresponde a secciones RAU, que se visitaron por un motivo equivalente. Las regiones en las que se visitó una proporción más elevada de casos por esta razón fueron las regiones de Antofagasta (42,9%), Arica y Parinacota (42,1%) y Atacama (41,2%). Por el contrario,

las regiones en que se visitó una menor proporción de secciones por este motivo fueron las regiones de O'Higgins (8,3%), Aysén (11,1%) y Tarapacá (17,6%).

Por último, se visitó 7,2% de los casos con el objeto de verificar las secciones completamente y enumerar un área de levantamiento específica dentro de ellas. Las regiones donde se visitó una mayor proporción de manzanas por este concepto fueron las regiones de Antofagasta (21,4%), Maule (14,6%) y Valparaíso (11,5%), en tanto las regiones menos visitadas por este motivo fueron Los Ríos (0,0%), Aysén (0,0%) y O'Higgins (0,8%).

IV.5. Conformación definitiva de la muestra

Como se ha comentado en la sección III.6 sobre la selección de viviendas de la muestra, ésta se encuentra sujeta a los conglomerados que finalmente conforman la muestra, los que, producto del proceso de verificación y enumeración son susceptibles de ser reemplazados por otros de características similares. Adicionalmente, el tamaño de muestra de viviendas a seleccionar dentro de cada manzana o sección, está condicionado al número de viviendas que se encuentran efectivamente disponibles, en consecuencia, la distribución definitiva de la muestra dentro de cada región-área se ajustó según las situaciones presentadas.

Las tablas siguientes muestran la distribución regional de la muestra definitiva en comparación con la inicialmente propuesta.

Tabla IV.4. Comparación de la distribución inicial y la definitiva del número de manzanas seleccionadas del MM2008, según región y grupo de tamaño, Casen 2017.

Región	Nro. de manzanas en la muestra											
	Distribución Inicial						Distribución Definitiva					
	Total	Grupo de Tamaño					Total	Grupo de Tamaño				
1		2	3	4	5	1		2	3	4	5	
País	9.179	5.307	2.289	678	382	523	9.207	5.300	2.299	741	382	485
	57,8%	24,9%	7,4%	4,2%	5,7%	57,6%	25,0%	8,0%	4,1%	5,3%		
Arica y Parinacota	384	263	74	23	9	15	384	263	74	23	9	15
Tarapacá	406	177	138	43	16	32	405	176	138	43	16	32
Antofagasta	466	283	119	36	12	16	465	282	119	36	12	16
Atacama	446	330	87	16	7	6	450	330	87	24	7	2
Coquimbo	430	277	119	19	9	6	430	276	119	21	9	5
Valparaíso	972	600	209	72	43	48	975	599	210	79	43	44
Metropolitana	2.000	824	524	206	151	295	2.005	824	524	219	150	288
O'Higgins	469	276	111	36	23	23	471	274	113	42	23	19
Maule	458	274	125	35	12	12	462	274	127	39	15	7
Biobío (Antigua)	1.092	655	278	81	42	36	1.091	654	278	83	42	34
Biobío	876	524	215	69	36	32	875	523	215	71	36	30
Ñuble	216	131	63	12	6	4	216	131	63	12	6	4
La Araucanía	535	371	121	21	13	9	537	370	122	26	13	6
Los Ríos	391	241	96	25	17	12	397	242	96	36	16	7
Los Lagos	410	249	108	31	13	9	411	250	108	32	12	9
Aysén	277	184	74	15	3	1	279	183	78	15	3	0
Magallanes	443	303	106	19	12	3	445	303	106	23	12	1

Elaborado por el Instituto Nacional de Estadísticas.

Tabla IV.5. Comparación de la distribución inicial y la definitiva de la muestra sobredimensionada de viviendas, según región y grupo de tamaño, Casen 2017.

Región	Nro. de viviendas en la muestra											
	Distribución Inicial						Distribución Definitiva					
	Total	Grupo de Tamaño					Total	Grupo de Tamaño				
1		2	3	4	5	1		2	3	4	5	
País	74.743	21.081	22.767	10.179	7.586	13.130	74.743	21.100	22.872	10.996	7.584	12.191
		28,2%	30,5%	13,6%	10,1%	17,6%		28,2%	30,6%	14,7%	10,1%	16,3%
Arica y Parinacota	2.702	1.055	741	345	171	390	2.702	1.055	741	345	171	390
Tarapacá	3.861	707	1.377	645	339	793	3.861	707	1.377	645	339	793
Antofagasta	3.465	1.130	1.183	534	234	384	3.465	1.130	1.182	535	234	384
Atacama	2.722	1.324	869	246	133	150	2.722	1.325	868	346	133	50
Coquimbo	2.903	1.100	1.182	288	177	156	2.903	1.100	1.182	312	177	132
Valparaíso	7.572	2.368	2.054	1.087	852	1.211	7.572	2.368	2.067	1.167	853	1.117
Metropolitana	21.958	3.258	5.227	3.087	2.988	7.398	21.958	3.263	5.233	3.253	2.966	7.243
O'Higgins	3.774	1.088	1.086	553	469	578	3.774	1.088	1.105	642	469	470
Maule	3.384	1.090	1.230	515	241	308	3.384	1.094	1.244	563	309	174
Biobío (Antigua)	8.326	2.605	2.758	1.219	841	903	8.326	2.605	2.765	1.242	841	873
Biobío	6.786	2.091	2.135	1.041	716	803	6.786	2.091	2.142	1.064	716	773
Ñuble	1.540	514	623	178	125	100	1.540	514	623	178	125	100
La Araucanía	3.496	1.467	1.212	315	262	240	3.496	1.469	1.228	385	262	152
Los Ríos	2.937	965	958	382	337	295	2.937	969	958	523	318	169
Los Lagos	3.008	989	1.089	454	257	219	3.008	992	1.091	479	227	219
Aysén	1.770	725	734	224	57	30	1.770	725	764	224	57	0
Magallanes	2.865	1.210	1.067	285	228	75	2.865	1.210	1.067	335	228	25

Elaborado por el Instituto Nacional de Estadísticas.

Tabla IV.6. Comparación de la distribución inicial y la definitiva de secciones seleccionadas del MS2002, según región y área, Casen 2017.

Región	Nro. de secciones en la muestra					
	Distribución Inicial			Distribución Definitiva		
	Total	Área		Total	Área	
RAU		Rural	RAU		Rural	
País	1.050	274 26,1%	776 73,9%	1.045	274 26,2%	771 73,8%
Arica y Parinacota	19	3	16	19	3	16
Tarapacá	17	3	14	17	3	14
Antofagasta	14	3	11	14	3	11
Atacama	17	0	17	17	0	17
Coquimbo	49	9	40	49	9	40
Valparaíso	87	16	71	87	16	71
Metropolitana	54	5	49	53	5	48
O'Higgins	121	31	90	121	31	90
Maule	130	49	81	130	49	81
Biobío (Antigua)	188	61	127	185	61	124
Biobío	96	24	72	96	24	72
Ñuble	92	37	55	89	37	52
La Araucanía	130	41	89	130	41	89
Los Ríos	75	9	66	75	9	66
Los Lagos	101	30	71	101	30	71
Aysén	36	11	25	36	11	25
Magallanes	12	3	9	11	3	8

Elaborado por el Instituto Nacional de Estadísticas.

Tabla IV.7. Comparación de la distribución inicial y la definitiva de la muestra sobredimensionada de viviendas, según región y área, Casen 2017.

Región	Nro. de viviendas en la muestra					
	Distribución Inicial			Distribución Definitiva		
	Total	Área		Total	Área	
RAU		Rural	RAU		Rural	
País	20.416	5.249 25,7%	15.167 74,3%	20.416	5.249 25,7%	15.167 74,3%
Arica y Parinacota	364	54	310	364	54	310
Tarapacá	345	65	280	345	65	280
Antofagasta	281	68	213	281	68	213
Atacama	326	0	326	326	0	326
Coquimbo	938	172	766	938	172	766
Valparaíso	1.726	327	1.399	1.726	327	1.399
Metropolitana	1.093	121	972	1.093	121	972
O'Higgins	2.420	608	1.812	2.420	608	1.812
Maule	2.494	909	1.585	2.494	909	1.585
Biobío (Antigua)	3.616	1.181	2.435	3.616	1.181	2.435
Biobío	1.803	447	1.356	1.859	447	1.412
Ñuble	1.813	734	1.079	1.757	734	1.023
La Araucanía	2.437	746	1.691	2.437	746	1.691
Los Ríos	1.482	173	1.309	1.482	173	1.309
Los Lagos	1.930	546	1.384	1.930	546	1.384
Aysén	736	219	517	736	219	517
Magallanes	228	60	168	228	60	168

Elaborado por el Instituto Nacional de Estadísticas.

V. DISEÑO FACTORES DE EXPANSIÓN

V.1. Introducción

Los análisis y estudios basados en encuestas de hogares con diseño muestral probabilístico, utilizan un ponderador en la estimación de las variables de interés del estudio para que tengan validez sobre la población objetivo, que guarda relación con las probabilidades de selección de las distintas unidades de muestreo y que da cuenta del número de personas de la población que representa cada individuo que participa en dicha encuesta. Este ponderador es conocido 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 estimación de población regional (expr) y otro que expande a la población comunal (expc).

Para Casen 2015, se desarrollaron factores para expandir a la población regional, provincial y comunal, lo que se mantiene para Casen 2017. Sin embargo, en esta versión de la encuesta tanto el factor de expansión provincial como el comunal se calculan solo para fines de investigación, dado que, los niveles de estimación o dominios de estudio para los cuales la muestra es representativa son: nacional, nacional urbano, nacional rural y regional.

Los factores regionales se calculan para todas las regiones del país, incluyendo la antigua región de Biobío y las nuevas regiones de Biobío y Ñuble.

Por otra parte, los factores de expansión provinciales para fines de investigación se asumen válidos para todas las provincias que no sean ADA y que tengan todas las comunas presentes en el marco muestral del INE o, para aquellas provincias cuyas comunas faltantes no representen más del 10% de la población de personas total de la provincia, según estimaciones poblacionales utilizadas para el mismo cálculo del factor.

De las 56 Provincias del País, Casen no cuenta con tres⁵² de ellas por estar excluidas del marco del INE debido a que fueron provincias catalogadas como ADA. De las restantes 53 provincias, existen 6 provincias que cuentan con sólo una comuna en la muestra. De éstas, de acuerdo al criterio establecido, se generaron factores de expansión provinciales sólo para Coyhaique, Magallanes y

⁵² Las provincias definidas como ADA son Isla de Pascua, Palena y Antártica Chilena.

Última Esperanza, quedando Parinacota, Capitán Prat y Tierra del Fuego como las únicas provincias del marco a las cuales no se les calcularon factores de expansión⁵³.

La metodología de cálculo de los factores de expansión consiste en la obtención 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, RAU o rural de una comuna) en la que se localiza. Adicionalmente, se realiza un ajuste a este ponderador, con lo que se busca compensar la falta de cobertura provocada por la exclusión de conglomerados por razones operativas y de costos.
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 2017 es la misma utilizada en Casen 2015, Casen 2013 y Casen 2011, con la cual se busca desarrollar grupos homogéneos (en relación a responder la encuesta) al interior de los cuales se calcula un ajuste de razón que permite corregir los ponderadores de selección.
4. Los ponderadores de calibración regional, provincial y comunal buscan cuadrar las estimaciones poblacionales obtenidas a partir de la encuesta con las estimaciones poblacionales desarrolladas por el Sub Departamento de Demografía y Vitales del INE, las que son utilizadas actualmente por el INE en todas las encuestas de hogares, mediante la corrección de fallas en la cobertura de subpoblaciones de interés. Este ajuste se conoce también como "corrección de población total".

⁵³ Más detalles en sección V.6.3 Ponderador de Calibración Provincial.

El ajuste "comunal", se obtiene como la razón entre el total de personas en la comuna según las estimaciones demográficas del INE y el total de personas residentes en viviendas particulares estimadas para cada comuna a partir de la encuesta.

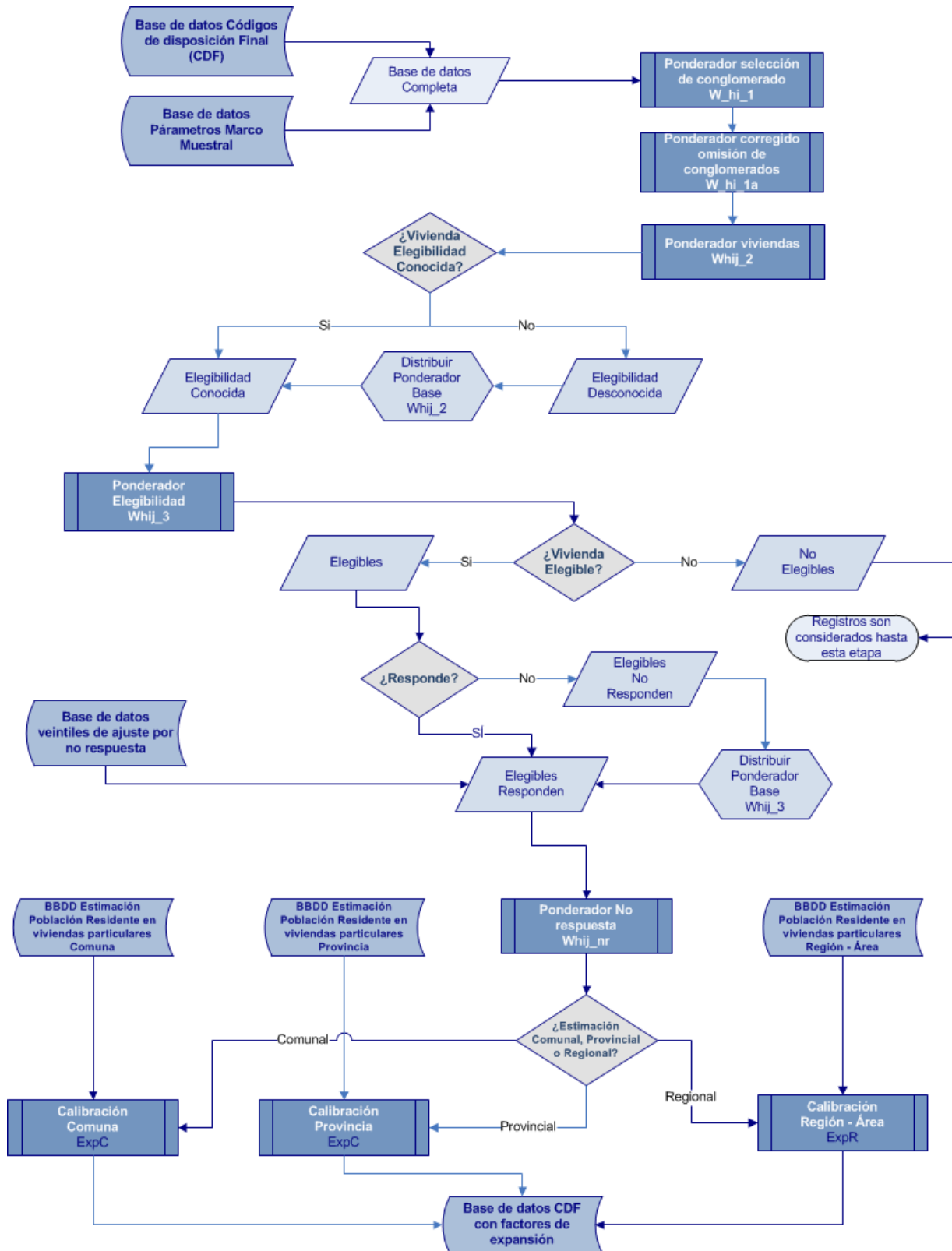
Asimismo, el ajuste "provincial", se obtiene como la razón entre el total de personas en este grupo poblacional según las estimaciones demográficas del INE y el total de personas residentes en viviendas particulares estimado para cada provincia a partir de la encuesta.

Por último, el ajuste "regional", se realiza separadamente en el área urbana y rural, a través de la razón entre el total de la población residente en viviendas particulares según las estimaciones poblacionales desarrolladas por el Sub Departamento de Demografía y Vitales del INE y el total de personas estimado a partir de la encuesta. La metodología de las estimaciones poblacionales utilizada para la calibración regional, provincial y comunal, es la misma utilizada para la encuesta Casen 2015.

En esta versión de la encuesta, la calibración regional incluye las nuevas regiones de Biobío y Ñuble, tanto para el área urbana como rural. En este sentido fue necesario calcular las estimaciones poblacionales en viviendas particulares para estas nuevas áreas, lo que se realizó a partir de la población llevada a noviembre 2017 en 15 regiones del país elaborada por el Sub Departamento de Demografía y Vitales del INE. En el Anexo N°5 se detalla el procedimiento que se siguió para la obtención de las estimaciones poblacionales a noviembre 2017 en las nuevas regiones de Biobío y Ñuble, a partir de la estimación de población de la antigua Biobío. Adicionalmente, se presenta el ajuste realizado para determinar la estimación poblacional en viviendas particulares ocupadas.

La Figura V.1 resume el proceso del cálculo de los factores de expansión para Casen 2017. 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 mencionado.

Figura V.1. Diagrama de procesos de desarrollo de factor de expansión Casen 2017.



Elaborado por el Instituto Nacional de Estadísticas.

V.2. Ponderador de selección de conglomerados

Casen 2017, posee un diseño probabilístico, estratificado según división política administrativa, área geográfica y tamaño poblacional en cuanto al número de viviendas. La muestra de viviendas fue seleccionada a partir de dos marcos muestrales, el Marco de Secciones (MS2002) y el Marco de Manzanas (MM2008), los cuales son mutuamente excluyentes y de cobertura nacional.

En ambos marcos muestrales la selección de conglomerados, manzanas o secciones, se realizó independientemente en cada estrato geográfico definido por el área (urbano, RAU o rural) de una comuna. Para la encuesta Casen 2017 se seleccionaron unidades muestrales en 1.124 estratos conformados por el cruce de comuna- área (urbano, RAU o rural) y grupo tamaño para el caso del área urbana los 1.558 estratos 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.

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

Un elemento importante de recordar, es que para la selección de los conglomerados de la muestra Casen 2017 fueron excluidas las secciones ya seleccionadas en la muestra de la Encuesta Nacional de Empleo, cuyo periodo de levantamiento coincidía con Casen 2017⁵⁴. El número de secciones excluidas⁵⁵ por este motivo corresponde a 543 y representan 6,5% del MS2002 (6,5% en cuanto a viviendas del MS2002).

Además, se excluyeron secciones pertenecientes al área rural de 11 comunas⁵⁶, en las cuales no se asignó muestra debido al agotamiento de las secciones, lo que implica la poca disponibilidad de

⁵⁴Estas unidades se excluyen de la muestra para prevenir el agotamiento de las mismas debido a la posibilidad que puedan ser seleccionadas en múltiples encuestas. Este procedimiento es el mismo utilizado para la encuesta Casen 2015 y versiones anteriores.

⁵⁵ Ver distribución de secciones excluidas en Tabla III.2 de la sección 2 del capítulo III.

⁵⁶ Las 11 comunas, en su área rural, donde no se asignó muestra son: Antofagasta, Mejillones, María Elena, Caldera, Chañaral, Concón, Papudo, El Quisco, El Tabo, San Pedro de la Paz y Hualpén. La exclusión de secciones por este criterio también se aplicó para la encuesta Casen 2015 y versiones anteriores.

viviendas al interior de éstas. Por último, se excluyen secciones RAU que pertenecen a áreas contenidas en el MM2008, al igual que en Casen 2015 y versiones anteriores.

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

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

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

$$P_h(i) = \frac{n_h M_{hi}}{M_h} \quad (8)$$

Donde:

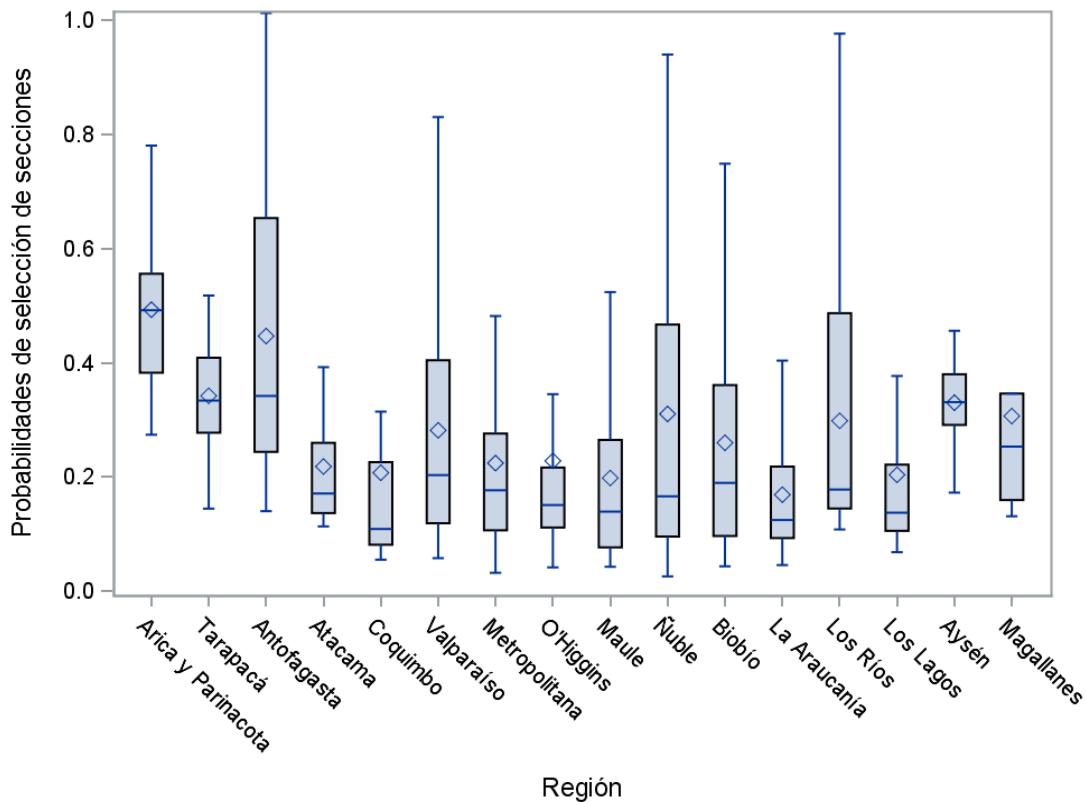
- i Es el índice de la unidad primaria de muestreo o sección.
- h Es el índice del estrato de muestreo, o área (urbana, RAU o rural) de una comuna.
- n_h Es el número de secciones seleccionadas del estrato h .
- M_{hi} Es el número de viviendas en la sección i del estrato h según el MS2002.
- M_h Es el número de viviendas en el estrato h según el MS2002 previamente acotado para la selección de Casen 2017⁵⁷.

El Gráfico V.1 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 con mayor concentración de población (en torno a 20%) y relativamente más alta para las regiones en los extremos con menor población.

⁵⁷ 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 2017 y a el área rural de 11 comunas excluidas por el agotamiento de secciones, descritas anteriormente.

Las regiones que presentan las menores medianas respecto a la probabilidad de selección de las secciones son la Región de Coquimbo (0,11), Región de La Araucanía (0,12) y Los Lagos (0,14). Por otro lado, las regiones que presentan las medianas más altas son Arica y Parinacota con 0,49, mientras Antofagasta y Atacama presentan un 0,34 y 0,33 respectivamente. La mediana nacional para las probabilidades de selección de secciones se ubica en 0,16.

Gráfico V.1. Distribuciones regionales de las probabilidades de selección de las secciones, por Región, Casen 2017.



Elaborado por el Instituto Nacional de Estadísticas.

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

Al igual que en el marco de secciones, en el marco de manzanas, para la selección de los conglomerados de la muestra Casen 2017, fueron excluidas las manzanas ya seleccionadas en la muestra de la Encuesta Nacional de Empleo⁵⁸, cuyo periodo de levantamiento coincidía con Casen

⁵⁸ Este procedimiento es el mismo utilizado en la encuesta Casen 2015 y anteriores y busca evitar el agotamiento de las unidades muestrales debido a la posibilidad de éstas de ser seleccionadas en diversas encuestas.

2017, además de las manzanas que cuentan con 7 viviendas o menos (grupo tamaño 0). El número de manzanas excluidas por pertenecer a la muestra ENE corresponde a 3.073 y representan el 2,3% del MM2008, concentrando un 2,3% en cuanto a viviendas, mientras que en el caso de las manzanas del grupo tamaño 0, éstas representan 53.425 viviendas, correspondientes a un 1,3% del MM2008.

Además, durante el proceso de distribución de la muestra, por razones operativas y de costo, en algunos estratos no se asignan viviendas, como por ejemplo, en manzanas grandes donde proporcionalmente se asignaría una o dos viviendas, no se asigna muestra. Este procedimiento también se utilizó en Casen 2015 y anteriores.

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

En una primera etapa se seleccionaron los conglomerados o “manzanas” del MM2008 (Unidades Primarias de Muestreo) con igual probabilidad. Las manzanas fueron seleccionadas todas a la vez de forma sistemática dentro de cada grupo de tamaño y con igual probabilidad.

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

$$P_h(i) = \frac{n_{hg}}{N_{hg}} \quad (9)$$

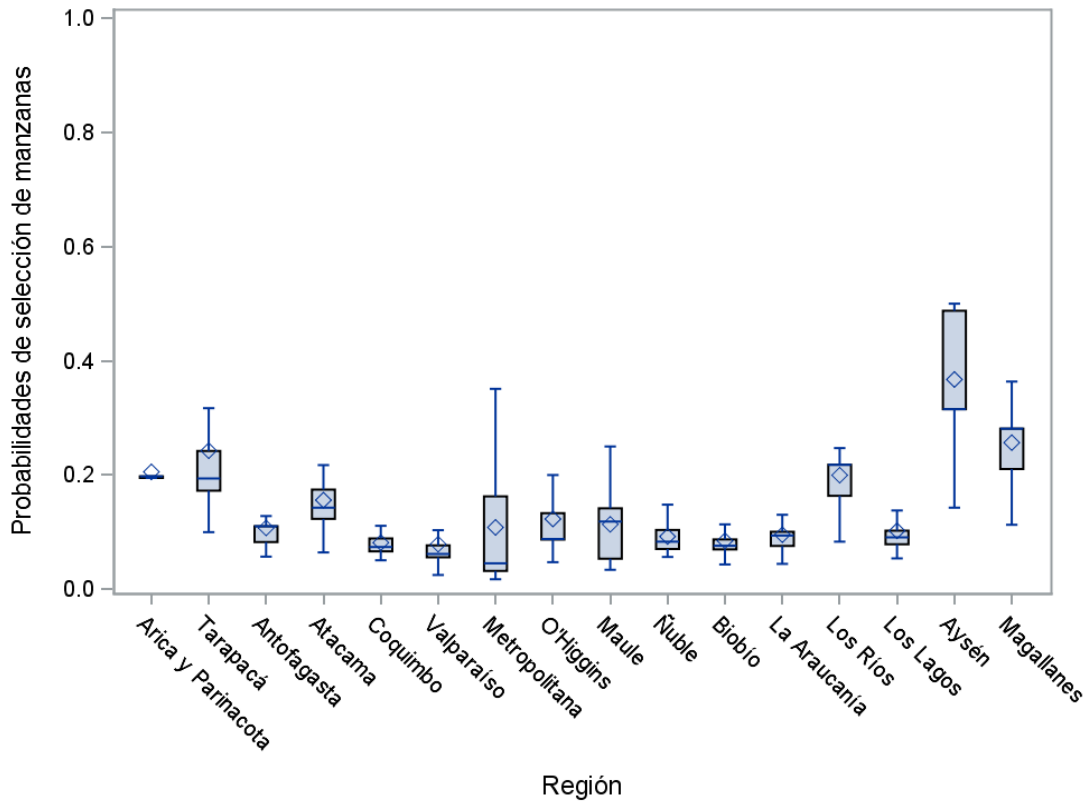
Donde,

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

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

El Gráfico V.2 presenta la distribución regional de probabilidades de selección de manzanas. Las regiones que presentan las menores medianas en la probabilidad de selección de las manzanas, son la Región Metropolitana con 0,04, seguidas de Valparaíso (0,06) y Coquimbo con 0,07. Por otra parte, las regiones con las medianas más altas son Aysén con 0,31, Magallanes (0,28) y Los Ríos con 0,22. Además, la mediana nacional de la probabilidad de selección de manzanas es de 0,09.

Gráfico V.2. Distribuciones regionales de las probabilidades de selección de las manzanas, por Región, Casen 2017.



Elaborado por el Instituto Nacional de Estadísticas.

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

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

$$w_{hi} = \frac{1}{P_h(i)} = \begin{cases} \frac{M_h}{n_h \cdot M_{hi}} & , \text{ Si unidad } i \text{ es una sección} \\ \frac{N_{hg}}{n_{hg}} & , \text{ Si unidad } i \text{ es una manzana} \end{cases} \quad (10)$$

Donde,

w_{hi}	Es el ponderador de selección base de la unidad primaria de muestreo o conglomerado (sección o manzana).
i	Es el índice de la unidad primaria de muestreo o conglomerado (manzana o sección).
h	Es el índice del estrato de muestreo, o área (urbana, RAU o rural) de una comuna.
n_h	Es el número de secciones seleccionadas en el estrato h .
M_{hi}	Es el número de viviendas en la sección i del estrato h según el MS2002.
M_h	Es el número de viviendas en el estrato h según el MS2002 previamente acotado para la selección de Casen 2017.
N_{hg}	Es el número de manzanas según el MM2008, en el grupo de tamaño g del estrato h .
n_{hg}	Es el número de manzanas en la muestra, seleccionadas en el grupo de tamaño g del estrato h .

El ponderador de selección de conglomerados se interpreta como “el total de unidades de su tipo que representa en la población”. 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. Al realizar este ejercicio se obtiene una estimación de 122.527 conglomerados a nivel país (secciones y manzanas).

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}).

La Tabla V.1 muestra la distribución regional de manzanas y secciones en cada marco muestral y las correspondientes estimaciones desarrolladas a partir del ponderador de manzanas y secciones. Se observa una subestimación del total de manzanas y secciones en todas las regiones del país. Esto se explica porque previa a la selección de la muestra Casen 2017, se acotó el marco muestral, excluyendo las secciones y manzanas ya seleccionadas en la muestra de la Encuesta Nacional de Empleo, cuyo periodo de levantamiento coincidía con Casen 2017 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 V.1. Distribución de frecuencias regionales de manzanas y secciones en el Marco Muestral y estimadas en Casen 2017 con ponderador de selección base.

Región	Marco Muestral		Estimadas en Casen 2017	
	Nº de Manzanas	Nº de Secciones	Nº de Manzanas	Nº de Secciones
Total País	133.757	8.324	115.386	7.141
Arica y Parinacota	2.424	64	1.990	42
Tarapacá	2.295	71	2.015	56
Antofagasta	5.698	53	4.832	45
Atacama	3.966	116	3.114	94
Coquimbo	6.824	568	5.710	429
Valparaíso	19.346	655	15.463	515
Metropolitana	46.017	509	42.322	424
O'Higgins	5.789	994	4.800	835
Maule	6.701	1.245	5.833	1.138
Biobío (Antigua)	16.492	1.409	13.960	1.253
Biobío	13.460	775	11.371	659
Ñuble	3.032	634	2.589	594
La Araucanía	7.202	1.173	6.049	1.057
Los Ríos	2.569	441	2.181	370
Los Lagos	5.332	821	4.487	718
Aysén	1.029	141	813	117
Magallanes	2.073	64	1.817	48

Elaborado por el Instituto Nacional de Estadísticas.

La Tabla V.2 y los Gráficos V.3 y V.4 presentan las distribuciones regionales de los ponderadores de selección de secciones y manzanas respectivamente, donde se observa que la mediana de los ponderadores de selección de las manzanas es superior al de las secciones entre 1,1 y 3,8 veces, a excepción de las regiones de Los Ríos y Magallanes, donde la mediana de los ponderadores de secciones es mayor.

Por otro lado, la media se comporta de manera similar, mostrando diferencias con medias superior entre 1,1 y 3,2 veces, exceptuando las regiones de Aysén y Magallanes.

Para más detalles estadísticos de estos ponderadores de selección, ver la Tabla V.2.

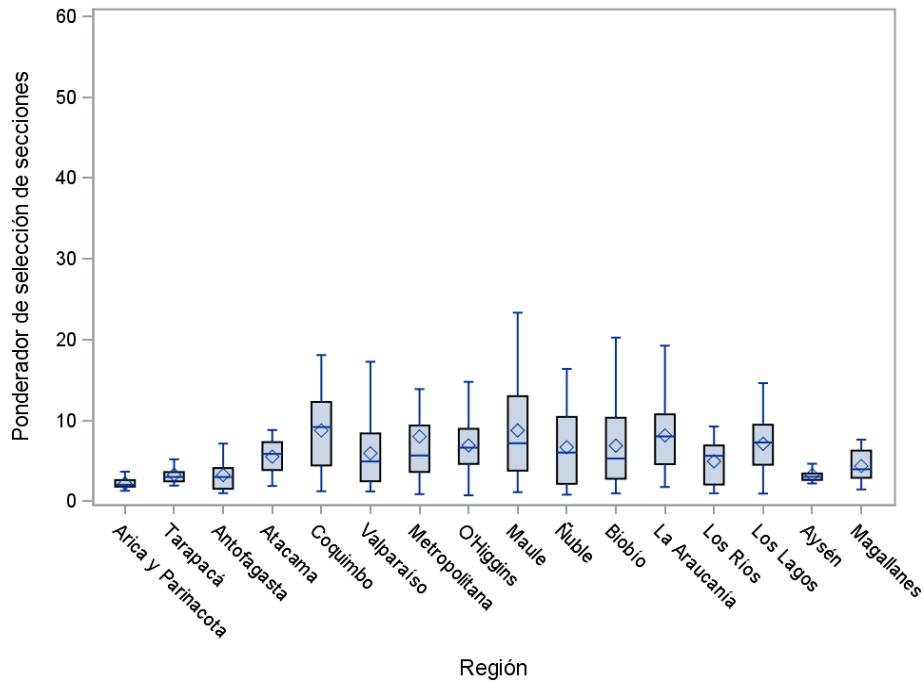
Tabla V.2. Estadísticos de ponderadores de UPM, manzanas y secciones, según MM2008 y MS2002, Casen 2017.

Región	Mínimo		Máximo		Media		Mediana		Desviación estándar	
	MM2008	MS2002 ⁵⁹	MM2008	MS2002	MM2008	MS2002	MM2008	MS2002	MM2008	MS2002
Total País	1,2	0,7	57,0	38,4	12,5	6,8	10,6	6,0	9,3	4,6
Arica y Parinacota	1,7	1,3	6,5	3,6	5,2	2,2	5,1	2,0	0,9	0,6
Tarapacá	1,2	1,9	10,0	6,9	5,0	3,3	5,2	3,0	1,7	1,3
Antofagasta	2,5	1,0	17,5	7,1	10,4	3,2	9,1	3,0	2,6	1,9
Atacama	3,2	1,9	15,5	8,8	6,9	5,5	7,0	5,8	1,9	2,1
Coquimbo	4,5	1,2	19,7	18,1	13,3	8,8	13,5	9,2	3,3	4,6
Valparaíso	2,0	1,2	40,0	17,3	15,9	5,9	16,2	4,9	6,3	4,0
Metropolitana	1,6	0,9	57,0	30,9	21,1	8,0	22,0	5,7	14,1	7,1
O'Higgins	2,5	0,7	21,1	23,9	10,2	6,9	11,4	6,6	3,9	3,8
Maule	2,2	1,1	29,3	23,3	12,6	8,8	8,4	7,2	6,9	5,7
Biobío (Antigua)	4,0	0,8	25,7	38,4	12,8	6,8	13,1	5,6	3,7	5,2
Biobío	4,0	1,0	25,7	22,9	13,0	6,9	13,1	5,3	3,7	5,0
Ñuble	5,4	0,8	17,7	38,4	12,0	6,7	12,0	6,0	3,5	5,5
La Araucanía	2,5	1,8	33,0	21,8	11,3	8,1	10,6	8,0	2,8	4,1
Los Ríos	3,0	1,0	12,0	9,2	5,5	4,9	4,6	5,6	1,8	2,4
Los Lagos	3,5	0,9	18,5	14,6	10,9	7,1	11,0	7,3	3,1	3,2
Aysén	2,0	2,2	7,0	5,8	2,9	3,2	3,2	3,0	0,8	0,9
Magallanes	2,8	1,5	8,9	7,6	4,1	4,3	3,6	3,9	1,0	2,1

Elaborado por el Instituto Nacional de Estadísticas.

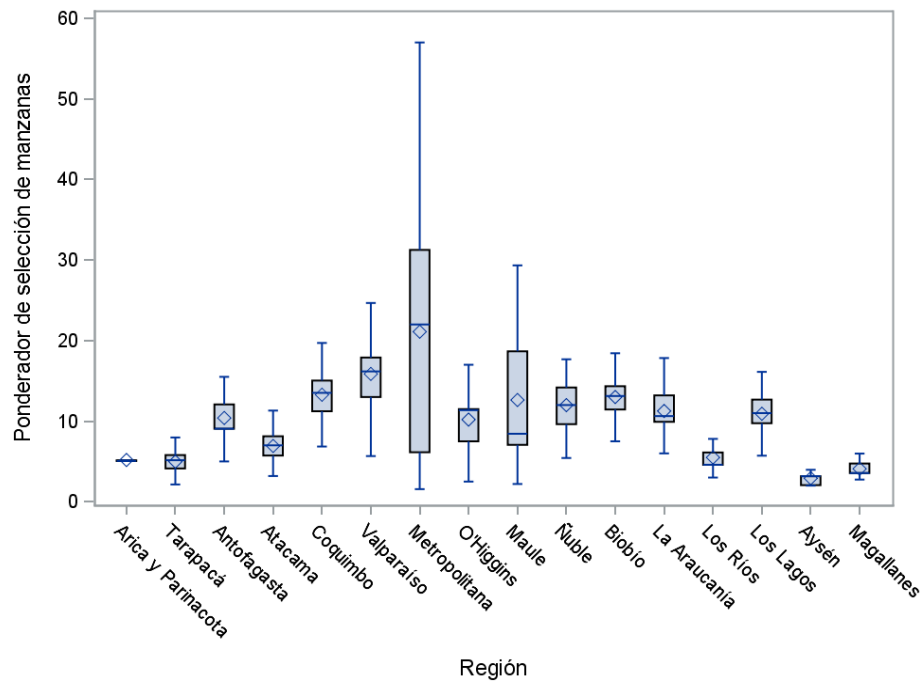
⁵⁹ Los ponderadores de secciones menores que 1 se justifican por el hecho de que la selección se realiza con probabilidad proporcional al tamaño, cuyo inverso de ésta puede ser menor que 1.

Gráfico V.3. Distribuciones regionales del ponderador de selección de secciones en Casen 2017.



Elaborado por el Instituto Nacional de Estadísticas.

Gráfico V.4. Distribuciones regionales del ponderador de selección de manzanas en Casen 2017.



Elaborado por el Instituto Nacional de Estadísticas.

V.2.4. Ajuste por omisión de conglomerados

El ajuste por omisión de conglomerados, tiene como finalidad corregir los ponderadores de selección de los conglomerados (manzanas y secciones) según la cantidad de viviendas que se excluyeron por estrato de muestreo en Casen 2017 por motivos operativos.

Previo a la selección de la muestra de Casen 2017, se descartaron *a priori* del marco muestral de Manzanas (MM2008) los conglomerados que contienen 7 viviendas o menos, aquellos seleccionados en la ENE y manzanas grandes donde proporcionalmente se asignaría una muestra pequeña.

Análogamente, se excluyeron del marco muestral de secciones (MS2002) las secciones previamente seleccionadas para la muestra de la Encuesta Nacional de Empleo, el área rural de las 11 comunas nombradas anteriormente, excluidas debido al agotamiento de sus secciones y las secciones pertenecientes a áreas RAU contenidas en el MM2008.

El número de manzanas y secciones descartadas a priori de ambos marcos muestrales equivale a un total de 424.786 viviendas, lo que corresponde a 9.0% de la población de viviendas pertenecientes a los estratos Casen⁶⁰.

Tabla V.3. Total de viviendas según marco de muestreo, con y sin exclusión de conglomerados.

Total de viviendas MM	Marco de Muestreo		Total Viviendas
	MM 2008	MS2002	
Marco con exclusión de unidades	3.741.864	559.563	4.301.427
Exclusiones Muestra ENE	94.735	42.985	137.720
Manzanas con 7 viviendas o menos	53.425	-	53.425
Manzanas excluidas por razones operativas	174.487	-	174.487
Áreas RAU seleccionadas del MM2008	-	55.278	55.278
Áreas excluidas por agotamiento unidades en 11 comunas	-	1.356	1.356
Áreas de Difícil Acceso presentes en Marco (ADA)	-	2.520	2.520
Marco sin exclusión de unidades	4.064.511	661.702	4.726.213

Elaborado por el Instituto Nacional de Estadísticas.

Utilizar el ponderador de selección de los conglomerados sin ajustar por omisión de conglomerados para la estimación de totales de conglomerados y viviendas, resultaría en una subestimación del número total de viviendas por conglomerado. Para evitar esto, se realizó un

⁶⁰ El total de viviendas en el marco muestral de Casen 2017 es de 4.726.213 viviendas.

ajuste, que consistió en aumentar los ponderadores de selección de los conglomerados proporcionalmente a la cantidad de viviendas excluidas por estrato de muestreo.

De este modo, la estimación del número de viviendas resultó consistente (4.722.337 a nivel nacional) con el número de viviendas de la población, cuya diferencia con el marco muestral sin exclusiones, corresponde 3.876 viviendas, equivalente a la suma de las viviendas pertenecientes a las secciones contenidas en las ADA's y al área rural de 11 comunas donde no se asignó muestra por el agotamiento de éstas⁶¹. Sin embargo, las dos últimas exclusiones son consideradas en el ajuste por estimación poblacional.

Para determinar el ajuste R_h por estrato h , según el marco muestral que corresponda, se calculó la razón entre M_h , el número total de viviendas del estrato h y el número total de viviendas estimadas con el ponderador de selección del estrato h ,

$$\hat{R}_h = \frac{M_h}{\sum_{i \in \Omega_h} w_{hi} \cdot M_{hi}} \quad (11)$$

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,

$$w'_{hi} = \hat{R}_h \cdot w_{hi} \quad (12)$$

V.3. Ponderador 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.

⁶¹ Las 11 comunas, en su área rural, donde no se asignó muestra son: Antofagasta, Mejillones, María Elena, Caldera, Chañaral, Concón, Papudo, El Quisco, El Tabo, San Pedro de la Paz y Hualpén. Este procedimiento también fue utilizado en Casen 2015 y anteriores.

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

V.3.1. Probabilidad Condicional de selección de viviendas

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

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

$$P_{hi}(j|i) = \frac{m_{hi}}{M'_{hi}} \quad (13)$$

Donde,

j Es el índice de la vivienda.

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

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

Al aplicar la definición de la probabilidad condicional, se puede determinar la probabilidad $P_{hi}(j)$ de selección de la vivienda j ubicada en el conglomerado (manzana o sección) i del estrato h en la muestra Casen 2017:

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

Donde,

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

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

V.3.2. Probabilidad de selección de viviendas como inverso de la probabilidad de selección

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

$$w_{hij} = \frac{1}{P_h(i) \cdot P_{hi}(j|i)} = \frac{1}{P_h(i)} \cdot \frac{1}{P_{hi}(j|i)} = w_{hi} \cdot \frac{1}{P_{hi}(j|i)} \quad (15)$$

Considerando que el ponderador de selección de conglomerado w_{hi} tuvo un ajuste por omisión descrito en el punto V.2.4 y expresado como el ponderador w'_{hi} , se puede reescribir el ponderador de selección de la vivienda, incorporando el ajuste por omisión de conglomerado. La expresión queda:

$$w_{hij} = w'_{hi} \cdot \frac{1}{P_{hi}(j|i)} = w_{hi} \cdot \hat{R}_h \cdot \frac{1}{P_{hi}(j|i)} = w_{hi} \cdot \hat{R}_h \cdot \frac{M'_{hi}}{m_{hi}} \quad (16)$$

Donde,

- h, i, j Es el índice del estrato, conglomerado (sección o manzana) y vivienda, respectivamente.
- w_{hij} Es el ponderador de selección de la vivienda j ubicada en el conglomerado i del estrato h .
- w'_{hi} Es el ponderador de selección del conglomerado i del estrato h , corregido por omisión de conglomerados (descartados a priori de los respectivos marcos muestrales).
- $P_{hi}(j|i)$ Probabilidad de selección de la vivienda j ubicada en el conglomerado (manzana o sección) i del estrato h en la muestra (definitiva) de Casen 2017 (fórmula 13).
- w_{hi} Es el ponderador de selección del conglomerado i del estrato h .
- \hat{R}_h Es el factor de ajuste por omisión de conglomerados (secciones o manzanas).
- M'_{hi} Es el número de viviendas ocupadas en la manzana o sección i , del estrato h , según enumeración previa a la recolección de datos.
- m_{hi} Es el número de viviendas seleccionadas en la manzana o sección i , del estrato h .

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

$$w_{hij} = \begin{cases} \frac{M_h}{n_h \cdot M_{hi}} \cdot \hat{R}_h \cdot \frac{M'_{hi}}{m_{hi}} & , \text{ Si unidad } i \text{ es una sección} \\ \frac{N_{hg}}{n_{hg}} \cdot \hat{R}_h \cdot \frac{M'_{hi}}{m_{hi}} & , \text{ Si unidad } i \text{ es una manzana} \end{cases} \quad (17)$$

Donde se agrega la nomenclatura:

M_h Es el número de viviendas en el estrato h según el MS2002 previamente acotado para la selección de Casen 2017.

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

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

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

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

Este ponderador puede ser interpretado como el número de viviendas en la población que representan las viviendas seleccionadas en la muestra Casen 2017. Al interior de cada conglomerado, la ponderación es la misma para todas las viviendas.

V.4. Ponderación por elegibilidad

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

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

Por ejemplo, fallas en el proceso de enumeración y/o verificación pueden derivar en el registro de inmuebles "no elegibles" (ej. casas de veraneo, negocios, etc.) en el marco. Por otra parte, también pueden producirse cambios en el uso de la vivienda entre el momento de enumeración y la aplicación 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 que para el caso de otras viviendas se desconoce su estado de elegibilidad.

Al término del trabajo de campo, todas las viviendas seleccionadas inicialmente por el INE son 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 completas, como no entrevistadas).
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.

V.4.1. Ajuste por elegibilidad desconocida

Inicialmente, todas las viviendas tienen un estatus conocido de elegibilidad: son o no viviendas particulares ocupadas. La muestra de viviendas seleccionadas incluye, en principio solo viviendas elegibles, según información disponible con posterioridad de procesos de enumeración y verificación en el caso de Casen 2017.

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

Los pesos de las edificaciones consideradas en un principio, como viviendas con Elegibilidad Conocida (elegibles y no elegibles), pero que finalmente terminan siendo clasificadas como edificaciones de "Elegibilidad Desconocida", son redistribuidos proporcionalmente al interior de cada estrato.

La proporción de los pesos a redistribuir $R_{h,known}$ se calcula como la razón entre dos estimaciones realizadas en base al ponderador de selección de viviendas.

El numerador corresponde a la estimación del total de viviendas en la población y el denominador corresponde a la estimación del total de viviendas con elegibilidad conocida (viviendas elegibles y no elegibles). Ambas estimaciones se realizan con el ponderador de selección de viviendas, es decir,

$$\hat{R}_{h,known} = \frac{\sum_{i \in \Omega_h} \sum_{j \in \Theta_i} w_{hij}}{\sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,known}} w_{hij}} \quad (18)$$

Donde,

- Θ_i Es el conjunto de viviendas j seleccionadas en la manzana o sección i .
- $\Theta_{i,known}$ Es el conjunto de viviendas j seleccionadas en la en la manzana o sección i y clasificadas como elegibles o no elegibles.
- Ω_h Es el conjunto de estratos h (áreas geográficas urbana, RAU o rural de las comunas).
- w_{hij} Es el ponderador de selección de la vivienda j en el conglomerado i del estrato h .

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

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

$$w'_{hij} = \hat{R}_{h,known} \cdot w_{hij} \quad (19)$$

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

V.4.2. Ajuste por no elegibilidad

Originalmente, la muestra definitiva Casen 2017 contemplaba 95.159 viviendas incluyendo sobremuestreo. De ese total 937 resultaron de "elegibilidad desconocida" una vez finalizado el trabajo de campo. Entre las 94.222 viviendas con elegibilidad conocida, hubo 90.670 que correspondían a viviendas "elegibles" y 3.552 que correspondieron a viviendas "no elegibles" (ver Tabla V.4).

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.).

Este ajuste consiste en asignar un valor blanco (“missing”), en el ponderador de selección de viviendas corregido por elegibilidad, a las viviendas con clasificación “no elegible”.

Entre las 90.670 viviendas elegibles, hubo 68.466 viviendas para las cuales se logró concretar una o más entrevistas⁶² y 22.204 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, desde Casen 2011 incluida Casen 2017.

Tabla V.4. Distribución de viviendas seleccionadas por el INE en Casen 2017 por elegibilidad en el trabajo de campo.

Vivienda Elegible	Frecuencia	Porcentaje
No elegible	3.552	3,7
Elegible	90.670	95,3
Elegibilidad desconocida	937	1,0
Total	95.159	100,0

Elaborado por el Instituto Nacional de Estadísticas.

V.5. Ponderación por no respuesta.

En esta etapa sólo son consideradas las viviendas elegibles, pues son éstas las que cumplen con las características necesarias para participar en la encuesta, es decir, 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 o simplemente no puedan ser contactados.

⁶² Bajo el protocolo de trabajo de campo de Casen, al interior de una vivienda seleccionada se completan tantas entrevistas como hogares sean identificados, por tanto, el encuestador debe intentar realizar una entrevista para cada hogar identificado.

En Casen 2017 de las 95.159 viviendas elegibles seleccionadas en la muestra, se lograron concretar 68.466 entrevistas. Es decir, la tasa de respuesta de la encuesta fue de 75,5%⁶³.

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, como el nivel socioeconómico, no ajustar por no respuesta, podría generar una subestimación (o sobreestimación) de algunos estadísticos de interés, ya que en la encuesta se 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 2017, se realiza un ajuste para compensar la pérdida de las unidades, denominado “propensity score”.

Este método de ajuste consiste en estimar la probabilidad de responder de las viviendas con hogares en la muestra. 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 estimación de la tasa de respuesta por unidad primaria de muestreo (sección o manzana).

⁶³ 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. Esta tasa es de 74,7%.

El ajuste consistió, al igual que en Casen 2015 y Casen 2013, en implementar un modelo de regresión lineal para predecir la participación en la encuesta (tasa de respuesta) de las unidades de muestreo (secciones y manzanas) utilizando un conjunto de variables auxiliares con información de las unidades mismas (total de viviendas según el Marco de muestreo, tipo de unidad - manzana o sección - entre otras), así como también información de la comuna, proveniente de otras encuestas o incluso de registros administrativos, con el objeto de predecir la tasa de respuesta al interior de las unidades, para posteriormente ordenarlas en orden creciente.

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

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

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

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

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

La determinación de la razón de ajuste viene dada por la ecuación (20) 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,

$$\hat{R}_{g,R} = \frac{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i,eleg}} w'_{hij}}{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i,eleg,R}} w'_{hij}} \quad (20)$$

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 V.5 presenta la composición del ponderador de no respuesta (factor de ajuste) estimado para cada veintil. 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 (razón de ajuste 1).

Tabla V.5. Ponderador de no respuesta por celda de ajuste.

Celda de ajuste de no respuesta o Veintiles	Componentes del ponderador de corrección por no respuesta			$\hat{R}_{g,R}$
	Viviendas Seleccionadas o elegibles	Viviendas que responden	Viviendas que no responden	
Total	90.670	68.466	22.204	
1	6.535	2.121	4.414	2,9
2	5.339	2.426	2.913	2,0
3	4.368	2.550	1.818	1,7
4	4.778	3.079	1.699	1,5
5	4.711	3.245	1.466	1,4
6	4.302	3.084	1.218	1,4
7	4.691	3.458	1.233	1,3
8	4.248	3.306	942	1,3
9	3.114	2.432	682	1,3
10	4.614	3.687	927	1,3
11	3.655	2.902	753	1,3
12	4.726	3.923	803	1,2
13	3.740	3.243	497	1,2
14	3.918	3.244	674	1,2
15	3.663	3.213	450	1,1
16	3.555	3.202	353	1,1
17	4.489	4.104	385	1,1
18	4.843	4.429	414	1,1
19	4.711	4.441	270	1,1
20	6.670	6.377	293	1,0

Fuente: Veintiles desarrollados por MDS.

Elaborado por el 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,

$$w_{hij}^{NR} = \hat{R}_{g,R} \cdot w'_{hij} \quad (21)$$

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

V.6. Ponderación de calibración

La muestra Casen es una muestra de viviendas y, hasta este punto, los factores de expansión calculados sirven para llevar estas unidades muestrales a sus respectivos stocks poblacionales (totales de viviendas). Sin embargo, como las principales unidades de interés analítico de la encuesta son las personas que residen en las viviendas seleccionadas se requiere incorporar información acerca de este stock poblacional a los factores de expansión.

Con este objetivo, el INE preparó estimaciones poblacionales, basadas en el Censo 2002⁶⁴, realizadas por el Sub Departamento de Demografía y Vitales del INE, las cuales son utilizadas por todas las encuestas de hogares que desarrolla el INE, con estimaciones al 30 de noviembre 2017 para todas las comunas de la muestra y además para el cruce dado por región y área (urbano, rural) como se observa en la Tabla V.6.

⁶⁴ No incluye información más reciente o actualizada.

Tabla V.6. Estimaciones poblacionales basadas en el Censo 2002, al 30 de noviembre de 2017.

Región	Estimación de población total			Estimación de población residente en viviendas particulares		
	Urbano	Rural	Total	Urbano	Rural	Total
Total país	15.853.320	2.341.063	18.194.383	15.550.716	2.256.639	17.807.355
Arica y Parinacota	149.620	21.028	170.648	145.331	15.821	161.152
Tarapacá	338.267	29.087	367.354	326.933	20.266	347.199
Antofagasta	604.551	16.528	621.079	576.762	9.578	586.340
Atacama	266.169	28.981	295.150	261.089	23.879	284.968
Coquimbo	648.739	144.285	793.024	636.376	139.130	775.506
Valparaíso	1.730.145	161.113	1.891.258	1.698.670	157.820	1.856.490
Metropolitana	7.064.808	257.494	7.322.302	6.975.088	248.072	7.223.160
O'Higgins	673.110	270.990	944.100	662.856	266.739	929.595
Maule	721.048	343.651	1.064.699	707.537	339.480	1.047.017
Biobío (Antigua)	1.794.298	331.309	2.125.607	1.761.239	324.629	2.085.868
Biobío	1.461.558	194.110	1.655.668	1.435.927	189.307	1.625.234
Ñuble	332.740	137.199	469.939	325.312	135.322	460.634
La Araucanía	699.029	328.847	1.027.876	677.866	321.478	999.344
Los Ríos	264.479	120.983	385.462	254.131	116.983	371.114
Los Lagos	650.895	260.829	911.724	629.513	253.449	882.962
Aysén	96.903	15.122	112.025	92.012	12.927	104.939
Magallanes	151.259	10.816	162.075	145.313	6.388	151.701

Elaborado por el Instituto Nacional de Estadísticas.

Las estimaciones fueron desarrolladas a partir de modelos⁶⁵ para toda la población, sin diferenciar entre aquellos que residen en viviendas particulares y aquellos que se encuentran en viviendas colectivas. Esta metodología para la obtención de las estimaciones poblacionales fue la misma utilizada para Casen 2015.

Como la población de inferencia de la Encuesta Casen está restringida a las personas que residen en viviendas particulares ocupadas, las estimaciones 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. En Casen 2015, este factor se calculó con base en los resultados del Censo 2002, para cada estrato (Comuna-Área), obteniendo por agregación la proporción de viviendas particulares ocupadas a nivel de región. En esta versión de Casen, este factor se estima para cada

⁶⁵ Estas estimaciones utilizan modelos demográficos a nivel país y región, y un modelo mixto (demográfico y matemático) a nivel de comuna.

comuna, así como para cada desagregación Región-Área, dado que las estimaciones poblacionales solo están disponibles a esos niveles.

Por ejemplo, la estimación de población nacional al 30 de noviembre de 2017 era de 18.194.383 personas⁶⁶. Según el Censo 2002, la proporción de personas residentes en viviendas particulares era de 97,9% aproximadamente, por lo cual se estima que el total de la población que reside en viviendas particulares es 17.807.355 personas⁶⁷.

Las estimaciones poblacionales al 30 de noviembre de 2017 permiten construir ponderadores regionales, provinciales y comunales, sin embargo, dado que la encuesta Casen 2017 es representativa a nivel nacional, nacional urbano, nacional rural y regional, el ponderador con el cual se calculan los indicadores oficiales es el regional, considerando los ponderadores provinciales y comunales para fines de investigación.

A continuación, se describen los procedimientos para la última etapa de desarrollo de los factores de expansión de Casen 2017.

V.6.1. Ponderador de Calibración Regional

El Ponderador de Calibración Regional, también conocido como “Factor de Expansión Regional” (la variable se denomina “expr” en la base de datos Casen), 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 estimaciones de población y estimar el total de personas en viviendas particulares. Se solicitó al INE las estimaciones poblacionales para los 32 estratos formados por el cruce de región y área (urbano, rural, incluyendo la Región de Ñuble). Las estimaciones 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 región y área, obteniendo el total de personas residentes en viviendas particulares (N_{RA}) en los 30 y

⁶⁶ Con fines analíticos, el departamento de Demografía del INE genera estimaciones 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.

⁶⁷ Las estimaciones poblacionales 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. Consultar Anexo N°5 para el detalle del cálculo de población en viviendas particulares.

32 estratos formados por el cruce de región-área, es decir, se obtuvieron estimaciones tanto para 15 como para 16 regiones (Ver Anexo N°5).

2. Estimar el total de personas en viviendas particulares a partir de Casen 2017. 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 estimaciones poblacionales (N_{RA} , en paso 1) y la estimación de la Encuesta Casen 2017 (paso 2).

$$\hat{R}_{RA} = \frac{N_{RA}}{\sum_{h \in RA} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,elg,R}} N_{hij} \cdot w_{hij}^{NR}} \quad R = 1, \dots, 16; \quad A = \text{Urbano}, \text{Rural} \quad (22)$$

Donde,

N_{RA} Es el número de personas residentes en viviendas particulares, en la región R y área A , según estimaciones poblacionales desarrolladas por INE al 30 de noviembre de 2017.

N_{hij} Es el número de personas residentes en la vivienda j , del conglomerado i , en el estrato h , según población estimada por la encuesta Casen 2017.

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:

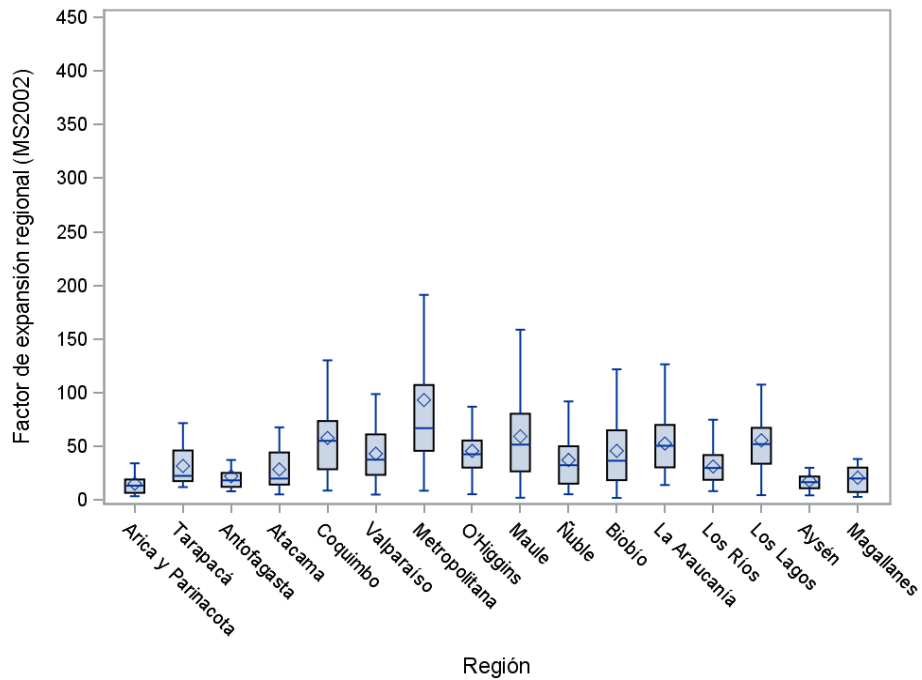
$$w_{hijk}^{RA} = \hat{R}_{RA} \cdot w_{hij}^{NR} \quad (23)$$

Los Gráficos V.5 y V.6 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 en general, las unidades provenientes del MM2008 son las que poseen mayor variabilidad, lo que se explica, en parte, porque en la parte urbana se concentra

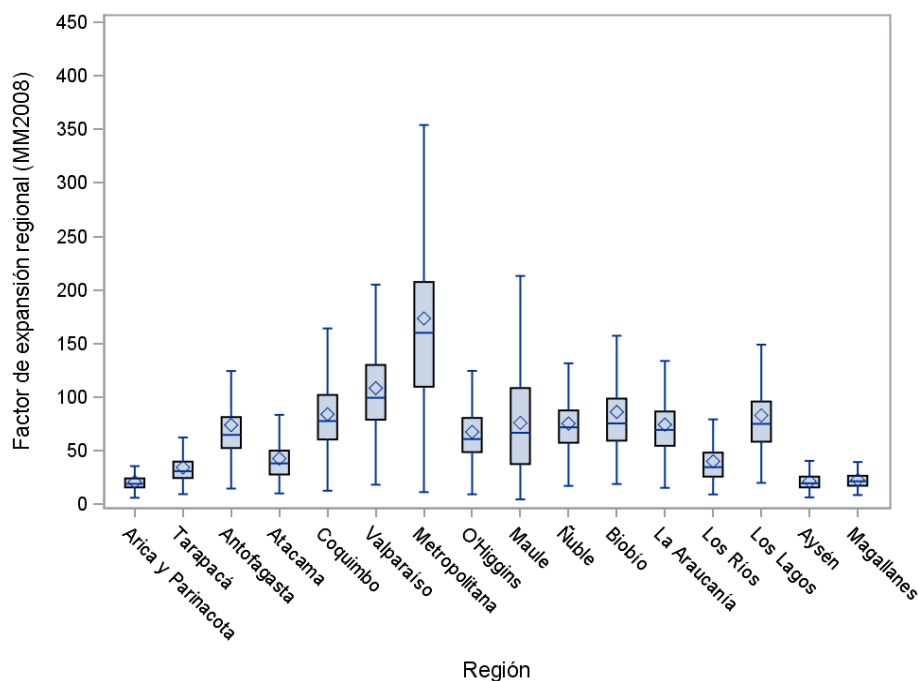
gran parte de la población total, así como también en ellas existe gran diversidad en estructura de sus unidades (manzanas grandes y chicas).

Gráfico V.5. Factor de expansión calibración Región-Área (RAU-Rural).



Elaborado por el Instituto Nacional de Estadísticas.

Gráfico V.6. Factor de expansión calibración Región-Área (Urbano).



Elaborado por el Instituto Nacional de Estadísticas.

V.6.2. Ponderador de Calibración Comunal

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

1. Ajustar las estimaciones poblacionales y estimar el total de personas en hogares particulares. Se solicitó al INE las estimaciones poblacionales para las 324 comunas presentes en la muestra Casen 2017. Las estimaciones fueron ajustadas, según la proporción de personas residentes en viviendas particulares reportadas por el Censo de Población y Vivienda del año 2002. De este modo se estimó el total de personas residentes en viviendas particulares (N_c) en las 324 comunas con presencia en la muestra 2017. (Ver Anexo N°5.)
2. Estimar el total de personas en hogares particulares a partir de Casen 2017. 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 2017.

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 estimaciones poblacionales (N_C , en paso 1) y el número de personas estimadas en la comuna por la encuesta Casen 2017 (paso 2),

$$\hat{R}_C = \frac{N_C}{\sum_{h \in C} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,elg,R}} N_{hij} \cdot w_{hij}^{NR}} \quad (24)$$

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:

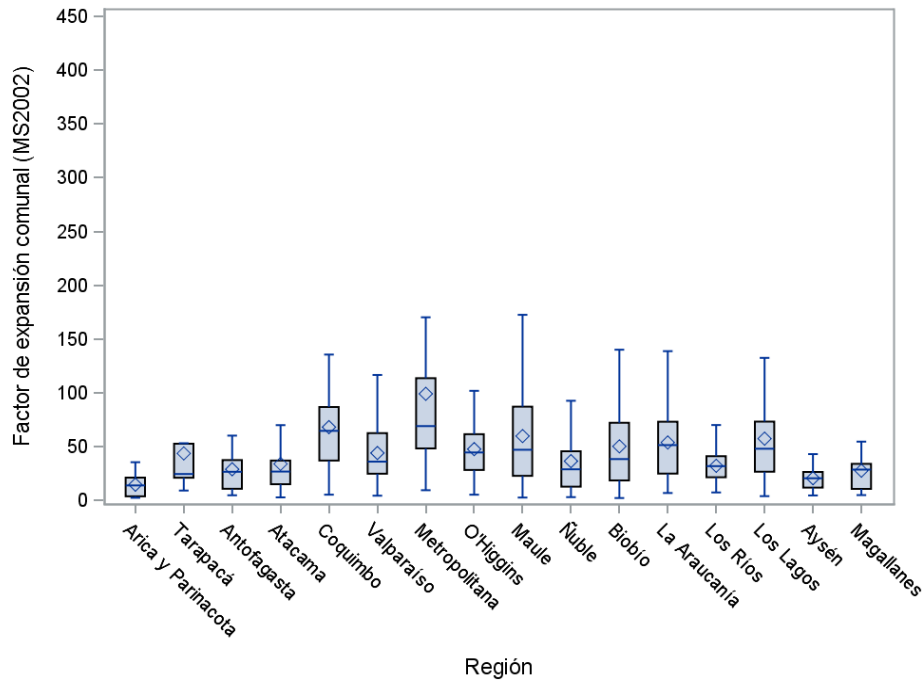
$$w_{hijk}^C = \hat{R}_C \cdot w_{hij}^{NR} \quad (25)$$

Los Gráficos V.7 e V.8 muestran la distribución regional de los ponderadores de calibración comunal para las unidades en los marcos rural y urbano, respectivamente.

De la comparación entre los Gráficos V.7 y V.8 se desprende que las unidades provenientes del marco de manzanas, en general, poseen mayor variabilidad, destacándose nuevamente la Región Metropolitana con la mayor variabilidad, seguida por la Región de Maule.

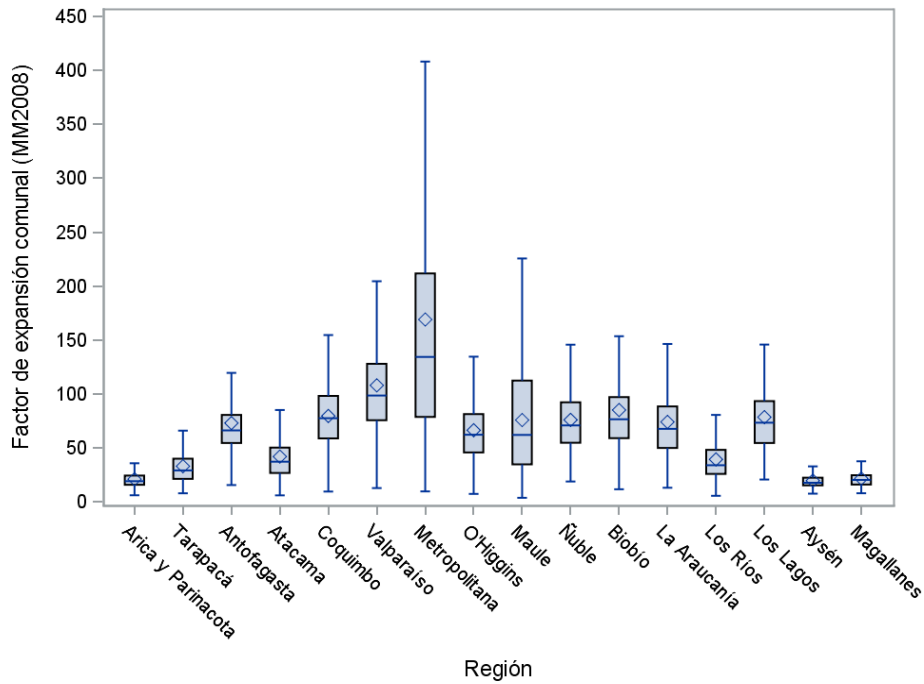
La Región de Arica y Parinacota en el Marco de secciones posee los ponderadores más pequeños y con menor variabilidad, mientras que en el Marco de manzanas, la Región de Aysén posee los factores más pequeños y de menor variabilidad.

Gráfico V.7. Distribución regional del Ponderador de Calibración Comunal de secciones (RAU-Rural).



Elaborado por el Instituto Nacional de Estadísticas.

Gráfico V.8. Distribución regional del Ponderador de Calibración Comunal de manzanas (Urbano).



Elaborado por el Instituto Nacional de Estadísticas.

V.6.3. Ponderador de Calibración Provincial

Como se planteó en capítulos precedentes, el criterio para desarrollar el factor provincial, es que se asumen válidos para todas las provincias que no sean ADA y que tengan todas las comunas presentes en el marco muestral del INE o, para aquellas provincias cuyas comunas faltantes no representen más del 10% de la población de personas total de la provincia, según estimaciones poblacionales.

En total de las 56 provincias del país, solo tres de éstas son consideradas ADA's, las cuales son Isla de Pascua, Palena y Antártica Chilena. Por otro lado, existen 6 provincias que cuentan con una sola comuna en la muestra de Casen 2017, las que según los criterios para el cálculo de los factores de expansión provincial, estos se generan solo para Coyhaique, Magallanes y Última Esperanza, mientras Parinacota, Capitán Prat y Tierra del Fuego se excluyen del cálculo, como se observa en la Tabla V.7.

Tabla V.7. Aplicación de criterio para factibilidad de cálculo de factor de expansión provincial.

Región	Provincia	Tipo	N° Comunas en Provincia		Comuna Única	Estimaciones*		% de Comuna en Provincia	Provincia con Factor
			Total	Muestra		Provincial	Comunal		
Arica y Parinacota	Parinacota		2	1	Putre	1.348	574	42,6%	No
Tarapacá	Tamarugal		5	4		39.302	37.922	96,5%	Si
Antofagasta	El Loa		3	2		155.308	155.168	99,9%	Si
Valparaíso	Valparaíso		7	6		690.998	690.065	99,9%	Si
	Isla de Pascua	ADA	1	0		5.634			
Los Lagos	Llanquihue		9	8		432.657	428.892	99,1%	Si
	Palena	ADA	4	0		17.095			
Aysén	Coyhaique		2	1	Coyhaique	59.785	59.065	98,8%	Si
	Aysén		3	2		34.273	32.436	94,6%	Si
	Capitán Prat		3	1	Cochrane	3.745	2.492	66,5%	No
Magallanes	Magallanes		4	1	Punta Arenas	122.521	121.793	99,4%	Si
	Antártica Chilena	ADA	2	0		2.392			
	Tierra del Fuego		3	1	Porvenir	5.628	4.830	85,8%	No
	Última Esperanza		2	1	Natales	21.160	20.551	97,1%	Si

*Estimación de población en viviendas particulares ocupadas al 30 de noviembre de 2017⁶⁸, INE. Elaborado por el Instituto Nacional de Estadísticas.

⁶⁸ Estimaciones Poblacionales desarrolladas por el Sub departamento de Demografía y Vitales del INE, las que son utilizadas actualmente por el INE en todas las encuestas de hogares.

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

1. Ajustar las estimaciones poblacionales y estimar el total de personas en hogares particulares. Las estimaciones fueron ajustadas, según la proporción de personas residentes en viviendas particulares reportadas por el Censo de Población y Vivienda del año 2002, por comuna y luego se obtuvo el total de población en cada provincia, a través de la suma sobre todas las comunas que conforman la provincia. De este modo se estimó el total de personas residentes en viviendas particulares (N_p) en cada provincia con presencia en la muestra 2017. (Ver Anexo N°5).
2. Estimar el total de personas en hogares particulares a partir de Casen 2017. 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 provincia en la muestra 2017.
3. Construir el ajuste a población total. Se construyó la razón \hat{R}_p , a partir del cociente entre la estimación del total de personas residentes en viviendas particulares desarrollado a partir de las estimaciones poblacionales (N_p , en paso 1) y el número de personas estimadas en la Provincia por la encuesta Casen 2017 (paso 2),

$$\hat{R}_p = \frac{N_p}{\sum_{h \in P} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,elg,R}} N_{hij} \cdot w_{hij}^{NR}} \quad (26)$$

Donde,

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

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

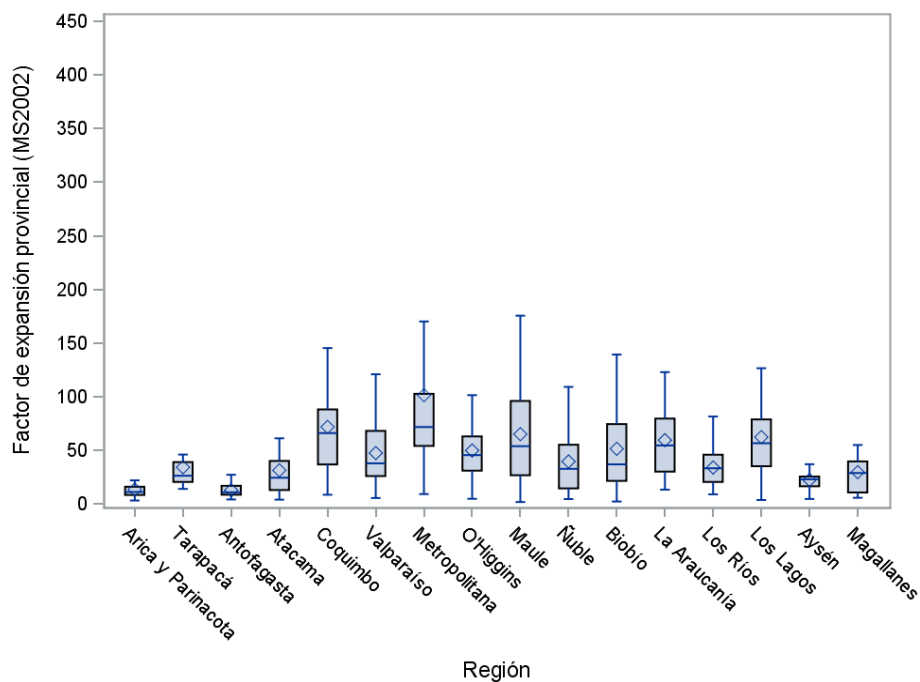
4. Construir el Ponderador de Calibración provincial. Finalmente se multiplica el Ponderador corregido por no respuesta con el ajuste de población total (paso3).

El factor de expansión provincial tiene la siguiente expresión:

$$w_{hijk}^P = \hat{R}_p \cdot w_{hij}^{NR} \quad (27)$$

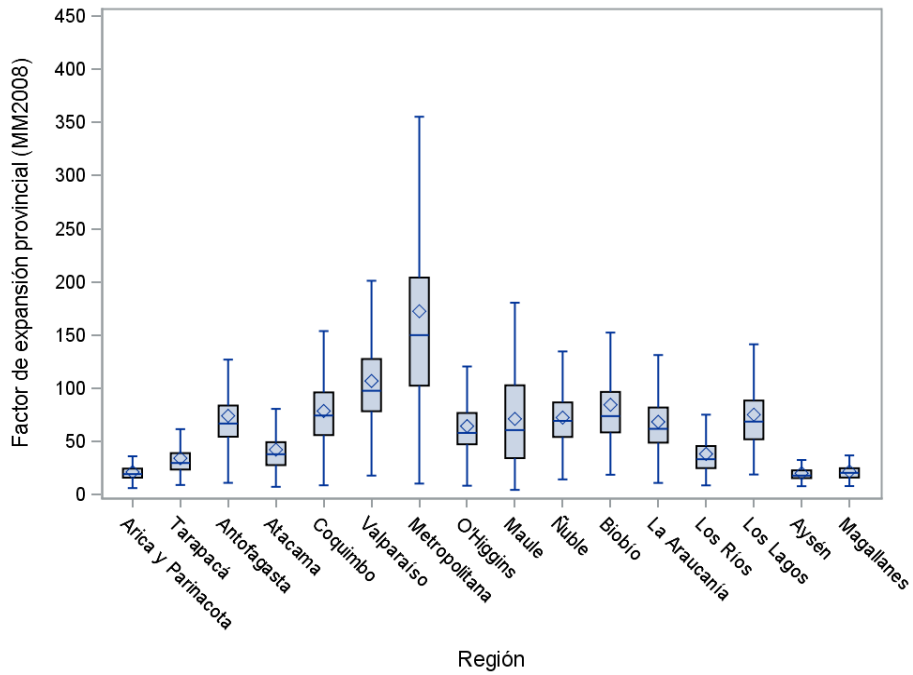
Los Gráficos V.9 y V.10 muestran la distribución regional de los ponderadores de calibración provincial para las unidades en los marcos rural y urbano, respectivamente.

Gráfico V.9. Distribución regional del Ponderador de Calibración Provincial de secciones (RAU-Rural).



Elaborado por el Instituto Nacional de Estadísticas.

Gráfico V.10. Distribución regional del Ponderador de Calibración Provincial de manzanas (Urbano).



Elaborado por el Instituto Nacional de Estadísticas

De la comparación entre los Gráficos V.9 y V.10 se desprende que las unidades provenientes del marco de manzanas, en general, poseen mayor variabilidad. La distribución de los factores de expansión en el Marco de Manzanas se encuentra entre 1 y 370 unidades, mientras en el marco de secciones entre 1 y 200 unidades, aproximadamente (gráficamente se excluyen los valores atípicos).

VI. ESTIMACIÓN DE LA VARIANZA COMPLEJA

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

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

VI.1. Métodos convencionales para estimación de la varianza

Existen principalmente cuatro métodos convencionales para estimar las varianzas o errores muestrales para estimaciones basadas en datos de una encuesta⁶⁹:

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

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

VI.1.1. Métodos exactos

Como se ha mencionado, los métodos exactos de estimación de la varianza, cuando son aplicables, son la mejor forma de estimar la varianza. Sin embargo, los diseños muestrales empleados en la

⁶⁹ Para más información ver Kish y Frankel (1974), Wolter (1985) y Lehtonen y Pahkinen (1995).

mayoría de las encuestas de hogares son más complejos que el muestreo aleatorio simple o estratificado.

En este sentido, las estimaciones de interés no son funciones lineales simples de los valores observados, por tanto la varianza de muestreo no puede expresarse por una fórmula de forma cerrada, como la varianza de la media muestral en el caso del muestreo aleatorio simple o el muestreo estratificado.

Los métodos exactos dependen del diseño de la muestra en cada caso, de la estimación de interés y de los procedimientos de ponderación empleados.

VI.1.2. Método del conglomerado último

El método del conglomerado último para estimar la varianza (véase Hansen, Hurwitz y Madow, 1953, págs. 257-259) se emplea para estimar las varianzas de las estimaciones basadas en una muestra obtenida a partir de un diseño muestral complejo, generalmente polietápico o polifásico.

Según este método, el conglomerado último consiste en considerar a la unidad primaria de muestreo UPM como la unidad de información última que conglojera la información de todas las unidades de muestreo de etapas posteriores dentro de ella.

Por ejemplo, si la unidad primaria de muestreo corresponde a un conglomerado que contiene unidades secundarias de muestreo como las viviendas, y a su vez estas unidades secundarias de muestreo contienen unidades terciarias como las personas, y se está interesado en alguna variable de interés como la ocupación, entonces cada UPM tendrá asociado el total de ocupados, como la sumatoria sobre todas las personas de todas las viviendas contenidas en la UPM.

Las estimaciones de la varianza se calculan utilizando sólo totales entre UPM, sin tener que calcular los componentes de la varianza en cada etapa de selección.

VI.1.3. Método de linealización

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

Se aplica la linealización generalmente para la estimación de razones o cocientes de dos variables las que se desarrollan en una serie de Taylor infinita centrada en el valor esperado (estimado) del numerador y el valor esperado (estimado) del denominador.

Seguidamente, el estimador no lineal se somete a aproximación algebraica conservando sólo los primeros términos de la serie infinita de Taylor, lo que produce una expresión algebraica que es una función lineal de los datos muestrales, es decir, el estimador no lineal cociente se ha "linealizado". Ahora es posible obtener directamente la varianza estimada de la función linealizada (incluidos los términos de covarianza pertinentes).

Una vez linealizada, la varianza de la estimación no lineal puede aproximarse mediante los métodos exactos descritos anteriormente (véanse Cochran, 1997, y Lohr, 1999, para consultar información técnica detallada sobre el proceso de linealización ilustrada con ejemplos).

VI.1.4. Método de replicación

Este método consiste en tomar submuestras o réplicas de la muestra total. Se estima con la muestra total la variable de interés Y , de acuerdo al diseño probabilístico, sea cual fuere. Enseguida se generan réplicas o se seleccionan submuestras de la muestra total, de manera que cada réplica refleje el plan de muestreo, los ajustes y procedimientos de ponderación de la muestra completa, de modo que con cada una de ellas se llegue a la misma estimación que con la muestra completa.

Existen varias técnicas para hacerlo y en algunos software ya vienen incluidos los módulos de selección de réplicas, entre los cuales los más utilizadas son los grupos aleatorios; la replicación repetida equilibrada (BRR); la replicación jackknife (JK1, JK2, y JK n); el bootstrap.

Por ejemplo, si se crean K réplicas a partir de una muestra, cada una con estimaciones $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K$ de un parámetro $\hat{\theta}$, y además la estimación basada en la muestra completa es $\hat{\theta}_0$. La estimación de la varianza basada en la replicación viene dada por: $V(\hat{\theta}) = \frac{1}{c} \sum_{r=1}^K (\hat{\theta}_r - \hat{\theta}_0)^2$, donde c es una constante que depende del método de estimación.

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

de los principales estadísticos asociados a alguna variable de interés y también la estimación de sus varianzas.

VI.2. Varianza en muestras complejas

Un diseño de muestreo complejo se origina cuando interviene más de una etapa de selección. Esta selección en más de una etapa suele ser 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 los parámetros de interés.

La encuesta Casen 2017, como también sus predecesoras, presenta un diseño muestral que se puede caracterizar como probabilístico, estratificado y polietápico. En el muestreo estratificado la estimación de la varianza total se obtiene mediante la suma de las varianzas estimadas en cada estrato y la varianza en cada estrato se estima a partir de la variabilidad de la característica de interés entre los conglomerados.

Una aproximación para la estimación de varianza comúnmente utilizada en encuestas complejas, de muestreo polietápico por conglomerados como Casen, corresponde al método de estimación de varianza por conglomerado último (EVCU) o por reemplazo (WR por sus siglas en inglés)⁷⁰. Éste método permite aproximarse a la varianza de los estimadores considerando aquellos diseños muestrales que contemplan la estratificación de las UPM de la población antes del muestreo en la primera etapa, además de la estratificación de las unidades de la segunda y posteriores etapas (en una UPM seleccionada) lo que puede tener lugar antes del muestreo en estas etapas.

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

⁷⁰ En adelante, para referirnos a éste método de estimación de varianzas con diseños complejos, se utilizará la nomenclatura “EVCU”.

Estas estimaciones con EVCU, pueden efectuarse indistintamente por linealización mediante el desarrollo en serie de Taylor o con una técnica de replicación. Con este enfoque, las UPM se tratan como si se hubieran seleccionado con reemplazo entre los estratos de la primera etapa. En tal caso, cada elemento o individuo sólo tiene que identificarse por el estrato de la primera etapa y por la UPM (dentro del estrato) de la que se haya seleccionado.

Para efectos de la estimación de varianzas no se necesita información sobre las etapas de muestreo por debajo del nivel de las UPM. De este modo la descripción del plan de muestreo, necesario para la estimación de varianzas con diseños complejos, se simplifica de forma que se asemeja a un muestreo por conglomerados estratificados en una etapa, esto es, una muestra estratificada de conglomerados finales completamente enumerados. Este enfoque por conglomerados finales produce una buena aproximación para la estimación de la varianza, siempre que sea razonable partir de la hipótesis de una primera etapa con reemplazo.

Así, cuando el plan de muestreo se describe como EVCU, sólo se necesitan tres variables del diseño de la encuesta para la estimación de varianzas:

- a. La variable de ponderación de la muestra o factor de expansión (necesaria también para las estimaciones) en Casen 2017 designado *expr* o *expc* o *expp*, según el nivel territorial de expansión: regional, comunal o provincial, respectivamente.
- b. La variable de estratificación (o variable de pseudo-estratificación) en Casen 2017 designado *varstrat*.
- c. La variable UPM (o pseudo-conglomerado) en Casen 2017 designada como *varunit*.

El empleo de EVCU para la estimación de varianzas considerando el diseño complejo puede dar lugar a una ligera sobreestimación. Sin embargo, los analistas de datos de encuestas no tienen, por lo general, problemas en aceptar un cierto grado de sobreestimación por la relativa simplicidad de la aproximación mediante EVCU. No obstante, hay que tener en cuenta que la sobreestimación puede ser apreciable si hay varios estratos en los que el muestreo en la primera etapa se hace sin reemplazo y con grandes fracciones de muestreo. En este caso, puede ser preferible optar por un software que cuente con la opción de incorporar los factores de corrección por población finita *cpf* o corrección por finitud para la primera etapa.

VI.3. Algoritmo de cálculo en muestras complejas

En el cálculo de varianzas con muestras complejas, tanto en STATA como SPSS se utiliza el método de conglomerado último.

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

Si una vivienda se incluye en la muestra con una probabilidad $P_{hi}(j)$, su ponderación básica, viene dada por $w_{hij} = 1/P_{hi}(j)$, es decir, el inverso de la probabilidad de selección y se denomina factor de expansión teórico, de acuerdo al diseño probabilístico. Este ponderador es igual para todas las viviendas dentro del conglomerado (manzana o sección) y, por lo tanto, para todas las personas dentro de ellas.

En la encuesta Casen, posteriormente, este factor se ajustó por omisión de conglomerados, elegibilidad desconocida y también por no-respuesta de las viviendas, además se calibró para alcanzar los stocks poblacionales de personas según las proyecciones de población.

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

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

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

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

n_h es el número de UPM (manzanas o secciones) seleccionadas en el estrato h en la muestra Casen 2017.

N_h es el número de UPM (manzanas o secciones) en el estrato h , contenidas en el marco muestral utilizado en Casen 2017.

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

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

VI.3.1. Algoritmo de cálculo de los estimadores

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

El Total estimado en el estrato h de la variable de interés Y es:

$$\hat{Y}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} z_{hij} = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \quad (28)$$

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

$$\hat{Y} = \sum_{h=1}^H \hat{Y}_h \quad (29)$$

La estimación del número de unidades en la población en el estrato h es:

$$\hat{N}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \quad (30)$$

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

$$\hat{N} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \quad (31)$$

La media estimada de la variable de interés Y , en el estrato h es:

$$\hat{\bar{Y}}_h = \frac{1}{\hat{N}_h} \hat{Y}_h \quad (32)$$

La media estimada de la variable de interés Y , en el nivel de estimación requerido es:

$$\hat{\bar{Y}} = \frac{1}{\hat{N}} \hat{Y} \quad (33)$$

El estimador de razón entre las variables Y y X , en el estrato h es:

$$\hat{R}_h = \frac{\hat{Y}_h}{\hat{X}_h} = \frac{\sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij}}{\sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij}} = \frac{\hat{Y}_h}{\hat{X}_h} \quad \text{es decir, } R_h = \frac{Y_h}{X_h} = \frac{\bar{Y}_h}{\bar{X}_h} \quad (34)$$

El estimador de razón entre las variables Y y X , en el nivel de estimación requerido del estudio es:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij}}{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij}} = \frac{\hat{Y}}{\hat{X}} \quad (35)$$

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

Las estimaciones del Total y la Media de la variable Y en el estrato h , obtenidas mediante el estimador de razón se expresan:

$$\hat{Y}_{R_h} = \hat{R}_h \cdot X_h \quad \text{y también} \quad \hat{\bar{Y}}_{R_h} = \hat{R}_h \cdot \hat{X}_h \quad (36)$$

Las estimaciones del Total y la Media de la variable Y a nivel agregado, obtenidas mediante el estimador de razón se expresan:

$$\hat{Y}_R = \hat{R} \cdot X \quad \text{y también} \quad \hat{\bar{Y}}_R = \hat{R} \cdot \hat{X} \quad (37)$$

Para las fórmulas (34), (35), (36) y (37), se asume que el Total X o la media \bar{X} son conocidas o dadas como fijas⁷¹. Pero, generalmente se desconoce el valor poblacional y entonces son estimados por \hat{X} y $\hat{\bar{X}}$ en la expresión (34), o son estimadas por \hat{X}_h y $\hat{\bar{X}}_h$ en la expresión (35).

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

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

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

$$\hat{V}(\hat{Y}) = \sum_{h=1}^H \hat{V}(\hat{Y}_h) = \sum_{h=1}^H S_h^2 = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (z_{hi} - \bar{z}_h)^2 \quad (38)$$

⁷¹ Por ejemplo, en la calibración a las proyecciones de población de personas, los stocks poblacionales son estimaciones obtenidas mediante modelos son considerados fijos, como los verdaderos valores en la población. Además si la variable es dicotómica, X coincide con N como asimismo \hat{X} coincide con \hat{N} .

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

$$\hat{V}(\hat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\sum_{j=1}^{m_{hi}} z_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} z_{hi} \right)^2 \quad (39)$$

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

$$\hat{V}(\hat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right)^2 \quad (40)$$

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

La varianza o cuasivarianza estimada de la variable de interés expandida en el estrato h es:

$$\hat{V}(\hat{Y}_h) = \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right)^2 \quad (41)$$

La estimación de la varianza del Total de la variable de interés Y , en el nivel de estimación requerido es:

$$V(\hat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right)^2 \quad (42)$$

La estimación de la varianza en el estrato h , de la Media de la variable de interés Y es:

$$V(\hat{Y}_h) = \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[\left(\sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right) / \hat{N}_h \right]^2 \quad (43)$$

La estimación de la varianza de la Media de la variable de interés Y , según el nivel de estimación requerido es:

$$V(\hat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[\left(\sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} \right) / \hat{N} \right]^2 \quad (44)$$

La varianza del estimador de razón entre las variables Y y X , en el estrato h es:

$$\hat{V}(\hat{R}_h) = \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[\left(\sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \hat{R}_h \cdot \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij} \right) / \hat{X}_h \right]^2 \quad (45)$$

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

La varianza del estimador de razón entre las variables Y y X , en el nivel de estimación requerido es:

$$\hat{V}(\hat{R}) = \sum_{h=1}^H \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} \left[\left(\sum_{j=1}^{m_{hi}} w_{hij} \cdot y_{hij} - \hat{R} \cdot \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij} \right) / \hat{X} \right]^2 \quad (46)$$

La estimación de la varianza para el Total y la Media respectivamente de la variable Y en base a la razón entre Y y X o la razón entre \bar{Y} y \bar{X} (ver expresiones (34), (35), (36) y (37) se expresa mediante las siguientes notaciones:

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

$$\hat{V}(\hat{Y}_{R_h}) = \hat{V}(\hat{R}_h \cdot X_h) \doteq \hat{V}(\hat{R}_h) \cdot \hat{X}_h^2 \quad ; \quad \hat{V}(\hat{Y}_{\hat{R}_h}) = \hat{V}(\hat{R}_h \cdot \bar{X}_h) \doteq \hat{V}(\hat{R}_h) \cdot \hat{X}_h^2 \quad (47)$$

La estimación de la varianza del Total y la Media de la variable Y a nivel agregado, (48) mediante el estimador de razón se expresa:

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

VI.4. Variables que identifican el diseño muestral complejo en Casen 2017

En la muestra de la encuesta Casen 2017, 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 de Casen 2017 se consideró 1.124 estratos y 10.252 segmentos (9.207 manzanas y 1.045 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 2017 estableció, para la muestra, 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 2017 y producir estimaciones de varianza más estables.

Las estimaciones oficiales de la varianza compleja de Casen 2017 han sido desarrolladas utilizando pseudo-estrato y pseudo-conglomerado. A continuación, se describen los procedimientos seguidos para la construcción de estas variables.

VI.4.1. Creación de pseudo-estratos (varstrat) en Casen 2017

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 1.124 estratos de muestreo originales⁷², 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.

⁷² Los estratos están conformados por las agrupaciones de comuna-área (urbana-RAU-rural) y grupo tamaño, para el caso del área urbana. De éstos se lograron viviendas en 1.123 estratos.

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 1.124 estratos muestrales se crearon 342 varstrat, los que poseen al menos dos conglomerados. La Tabla V.1 presenta el total de estratos y varstrat por región.

VI.4.2. Creación de pseudo-conglomerados (varunit) en Casen 2017

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 9.979⁷³ conglomerados en los que se lograron viviendas, es decir, al menos una vivienda al interior de éstos respondió la encuesta Casen 2017, donde cada uno contiene un número mínimo de viviendas.

A partir de un total de 9.979 conglomerados en los que se lograron viviendas de la muestra Casen 2017, se crearon 1.637 varunit, los cuales agrupan entre 22 y 53 viviendas entrevistadas (cifras menor y mayor a los límites mínimos y máximos establecidos inicialmente de 30 y 50 viviendas por varstrat). La Tabla VI.1 presenta el total de conglomerados (Manzanas/Secciones) y varunit por región.

⁷³ El número original de conglomerados seleccionados en la muestra de Casen 2017 es de 10.252 manzanas o secciones.

Tabla VI.1. Total de estratos, conglomerados, varstrat y varunit, según región por área geográfica, para la muestra lograda de Casen 2017.

Región	Tipo	N° Estratos Casen 2017*	N° Manzanas o Secciones**	N° de Viviendas***	VarStrat	VarUnit
Total País	Urbano	847	9.210	55.224	207	1.289
	Rural	276	769	13.242	135	348
Arica y Parinacota	Urbano	6	386	2.165	1	48
	Rural	3	16	282	2	8
Tarapacá	Urbano	13	401	2.556	3	56
	Rural	5	13	209	3	6
Antofagasta	Urbano	21	434	2.293	4	51
	Rural	5	10	121	2	4
Atacama	Urbano	20	434	1.806	6	42
	Rural	7	17	266	3	7
Coquimbo	Urbano	31	429	2.282	7	52
	Rural	15	40	696	8	19
Valparaíso	Urbano	102	965	5.322	21	124
	Rural	31	71	1.108	11	29
Metropolitana	Urbano	211	1.874	12.102	45	282
	Rural	19	48	670	6	16
O'Higgins	Urbano	86	497	3.443	21	84
	Rural	33	90	1.666	18	43
Maule	Urbano	62	508	3.510	19	83
	Rural	30	81	1.518	16	38
Biobío (Antigua)	Urbano	128	1.144	7.602	35	186
	Rural	50	124	2.186	22	57
Biobío	Urbano	93	891	5.705	24	139
	Rural	29	72	1.271	13	33
Ñuble	Urbano	35	253	1.897	11	47
	Rural	21	52	915	9	24
La Araucanía	Urbano	64	574	3.481	20	84
	Rural	32	89	1.629	17	44
Los Ríos	Urbano	31	400	2.202	8	52
	Rural	12	66	1.158	8	31
Los Lagos	Urbano	53	433	2.818	13	66
	Rural	25	71	1.302	15	35
Aysén	Urbano	10	289	1.480	2	33
	Rural	6	25	309	3	8
Magallanes	Urbano	9	442	2.162	2	46
	Rural	3	8	122	1	3

*Estratos de diseño en los que se logró muestra según Casen 2017, desagregación Comuna-área (Urbana, RAU, Rural) y Grupo de tamaño en el caso del área urbana.

** Numero de conglomerados en los que se logró muestra para Casen 2017

*** Número de viviendas que respondió Casen 2017

Elaborado por el Instituto Nacional de Estadísticas.

Los varunit fueron construidos 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 solo 2 varunit quedaron por debajo del criterio establecido (menos de 30 viviendas), mientras que 4 quedaron sobre el límite superior de 50 viviendas. Ver Tabla VI.2 siguiente.

Por otro lado, en la tabla VI.3 se observa que el máximo de conglomerados que componen los varunit son 23 conglomerados, lo que se explica por la alta tasa de no respuesta al interior de las manzanas que componen éste varunit. Estos conglomerados corresponden a la comuna de Vitacura.

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

Tabla VI.2. Frecuencia de varunit según número de viviendas que lo componen, Casen 2017.

N° de viviendas que responde	Frecuencia	Porcentaje	Porcentaje Acumulado
Total	1.637	100	100
22	1	0,06	0,06
29	1	0,06	0,12
30	63	3,85	3,97
31	49	2,99	6,96
32	56	3,42	10,38
33	60	3,67	14,05
34	60	3,67	17,72
35	58	3,54	21,26
36	72	4,40	25,66
37	74	4,52	30,18
38	69	4,22	34,39
39	67	4,09	38,49
40	56	3,42	41,91
41	63	3,85	45,75
42	44	2,69	48,44
43	58	3,54	51,99
44	64	3,91	55,89
45	85	5,19	61,09
46	84	5,13	66,22
47	104	6,35	72,57
48	141	8,61	81,19
49	133	8,12	89,31
50	171	10,45	99,76
51	3	0,18	99,94
53	1	0,06	100,00

Elaborado por el Instituto Nacional de Estadísticas.

Tabla VI.3. Distribución del número de conglomerados que componen los varunit, Casen 2017.

N° de Manzanas o Secciones	Frecuencia	Porcentaje	Porcentaje Acumulado
Total	1.637	100	100
2	355	21,69	21,69
3	133	8,12	29,81
4	103	6,29	36,10
5	143	8,74	44,84
6	164	10,02	54,86
7	153	9,35	64,20
8	182	11,12	75,32
9	160	9,77	85,09
10	96	5,86	90,96
11	53	3,24	94,20
12	34	2,08	96,27
13	36	2,20	98,47
14	15	0,92	99,39
15	4	0,24	99,63
16	5	0,31	99,94
23	1	0,06	100,00

Elaborado por el Instituto Nacional de Estadísticas.

Tabla VI.4. Distribución del número de varunit que componen los varstrat, Casen 2017.

N° de varunit	Frecuencia	Porcentaje	Porcentaje Acumulado
Total	342	100	100
2	137	40,06	40,06
3	83	24,27	64,33
4	45	13,16	77,49
5	18	5,26	82,75
6	17	4,97	87,72
8	4	1,17	88,89
9	3	0,88	89,77
10	2	0,58	90,35
11	3	0,88	91,23
12	1	0,29	91,52
13	1	0,29	91,81
14	4	1,17	92,98
15	3	0,88	93,86
16	4	1,17	95,03
17	1	0,29	95,32
18	1	0,29	95,61
19	2	0,58	96,20
20	2	0,58	96,78
22	3	0,88	97,66
24	1	0,29	97,95
25	2	0,58	98,54
28	1	0,29	98,83
32	1	0,29	99,12
38	1	0,29	99,42
41	1	0,29	99,71
48	1	0,29	100,00

Elaborado por el Instituto Nacional de Estadísticas.

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

En Casen 2017, al igual que en Casen 2015, Casen 2013 y Casen 2011, las estimaciones y márgenes de error de la situación de pobreza por ingresos, fueron determinadas utilizando los parámetros varstrat, varunit y expr (factor de expansión regional), considerando la utilización de pseudo-estratos y pseudo-conglomerados, siendo éstos implementados mediante cálculos realizados en paralelo utilizando los programas SPSS y STATA.

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

En la Tabla VI.5 se presentan las tasas de pobreza por ingresos para las personas y los hogares para la encuesta Casen 2017, a nivel nacional, por área geográfica urbana y rural y a nivel regional. Al igual, se calcularon los márgenes de error al 95% de confianza, asociados a estas estimaciones.

Tabla VI.5. Estimación de porcentaje de la población en situación de pobreza por ingresos, según zona y región, Casen 2017.

Región	Situación de pobreza por ingresos	Estimación	Error estándar	95% de intervalo de confianza		Coeficiente de Variación
				Inferior	Superior	
Nacional	Pobreza extrema	2,3%	0,1%	2,2%	2,5%	3,8%
	Pobreza no extrema	6,3%	0,1%	6,0%	6,6%	2,4%
	Total Pobreza	8,6%	0,2%	8,2%	9,0%	2,2%
	No pobre	91,4%	0,2%	91,0%	91,8%	0,2%
Urbano	Pobreza extrema	2,0%	0,1%	1,8%	2,2%	4,6%
	Pobreza no extrema	5,4%	0,2%	5,1%	5,7%	2,9%
	Total Pobreza	7,4%	0,2%	7,1%	7,8%	2,7%
	No pobre	92,6%	0,2%	92,2%	92,9%	0,2%
Rural	Pobreza extrema	4,4%	0,3%	3,9%	5,0%	6,2%
	Pobreza no extrema	12,1%	0,4%	11,3%	13,0%	3,5%
	Total Pobreza	16,5%	0,5%	15,5%	17,6%	3,2%
	No pobre	83,5%	0,5%	82,4%	84,5%	0,6%
Arica y Parinacota	Pobreza extrema	2,4%	0,4%	1,7%	3,4%	17,6%
	Pobreza no extrema	6,0%	0,7%	4,8%	7,5%	11,5%
	Total Pobreza	8,4%	0,9%	6,8%	10,4%	10,9%
	No pobre	91,6%	0,9%	89,6%	93,2%	1,0%
Tarapacá	Pobreza extrema	1,7%	0,3%	1,2%	2,3%	16,1%
	Pobreza no extrema	4,7%	0,4%	4,0%	5,7%	9,1%
	Total Pobreza	6,4%	0,6%	5,4%	7,6%	8,8%
	No pobre	93,6%	0,6%	92,4%	94,6%	0,6%
Antofagasta	Pobreza extrema	1,6%	0,4%	1,0%	2,6%	25,2%
	Pobreza no extrema	3,5%	0,5%	2,6%	4,8%	15,4%
	Total Pobreza	5,1%	0,7%	4,0%	6,6%	12,8%
	No pobre	94,9%	0,7%	93,4%	96,0%	0,7%
Atacama	Pobreza extrema	2,6%	0,5%	1,8%	3,6%	17,8%
	Pobreza no extrema	5,4%	0,6%	4,2%	6,8%	12,0%
	Total Pobreza	7,9%	1,0%	6,2%	10,0%	12,0%
	No pobre	92,1%	1,0%	90,0%	93,8%	1,0%
Coquimbo	Pobreza extrema	3,0%	0,4%	2,3%	4,0%	14,4%
	Pobreza no extrema	8,9%	0,8%	7,5%	10,5%	8,8%
	Total Pobreza	11,9%	0,8%	10,4%	13,6%	7,0%
	No pobre	88,1%	0,8%	86,4%	89,6%	0,9%
Valparaíso	Pobreza extrema	1,6%	0,2%	1,2%	2,1%	13,3%
	Pobreza no extrema	5,5%	0,4%	4,8%	6,3%	7,0%
	Total Pobreza	7,1%	0,4%	6,3%	8,0%	5,8%
	No pobre	92,9%	0,4%	92,0%	93,7%	0,4%
Metropolitana	Pobreza extrema	1,5%	0,2%	1,3%	1,9%	10,2%
	Pobreza no extrema	3,8%	0,3%	3,4%	4,4%	6,7%
	Total Pobreza	5,4%	0,3%	4,8%	6,0%	6,1%
	No pobre	94,6%	0,3%	94,0%	95,2%	0,3%

Región	Situación de pobreza por ingresos	Estimación	Error estándar	95% de intervalo de confianza		Coeficiente de Variación
				Inferior	Superior	
O'Higgins	Pobreza extrema	2,2%	0,3%	1,7%	2,9%	14,6%
	Pobreza no extrema	7,9%	0,6%	6,8%	9,1%	7,4%
	Total Pobreza	10,1%	0,7%	8,7%	11,6%	7,2%
	No pobre	89,9%	0,7%	88,4%	91,3%	0,8%
Maule	Pobreza extrema	3,0%	0,3%	2,5%	3,7%	10,3%
	Pobreza no extrema	9,7%	0,6%	8,5%	11,0%	6,4%
	Total Pobreza	12,7%	0,7%	11,5%	14,1%	5,3%
	No pobre	87,3%	0,7%	85,9%	88,5%	0,8%
Ñuble	Pobreza extrema	4,6%	0,5%	3,8%	5,6%	10,1%
	Pobreza no extrema	11,5%	0,7%	10,1%	13,1%	6,5%
	Total Pobreza	16,1%	0,9%	14,5%	17,9%	5,5%
	No pobre	83,9%	0,9%	82,1%	85,5%	1,1%
Biobío	Pobreza extrema	3,7%	0,3%	3,1%	4,5%	9,3%
	Pobreza no extrema	8,6%	0,5%	7,6%	9,7%	6,2%
	Total Pobreza	12,3%	0,7%	11,0%	13,8%	5,8%
	No pobre	87,7%	0,7%	86,2%	89,0%	0,8%
La Araucanía	Pobreza extrema	4,6%	0,4%	3,9%	5,4%	8,5%
	Pobreza no extrema	12,6%	0,6%	11,4%	13,8%	4,9%
	Total Pobreza	17,2%	0,8%	15,7%	18,7%	4,5%
	No pobre	82,8%	0,8%	81,3%	84,3%	0,9%
Los Ríos	Pobreza extrema	3,3%	0,5%	2,6%	4,3%	13,5%
	Pobreza no extrema	8,8%	0,7%	7,5%	10,3%	8,3%
	Total Pobreza	12,1%	0,9%	10,5%	14,0%	7,3%
	No pobre	87,9%	0,9%	86,0%	89,5%	1,0%
Los Lagos	Pobreza extrema	3,4%	0,4%	2,7%	4,2%	11,3%
	Pobreza no extrema	8,3%	0,6%	7,2%	9,6%	7,3%
	Total Pobreza	11,7%	0,8%	10,2%	13,4%	7,1%
	No pobre	88,3%	0,8%	86,6%	89,8%	0,9%
Aysén	Pobreza extrema	1,1%	0,3%	0,7%	1,7%	23,6%
	Pobreza no extrema	3,5%	0,7%	2,3%	5,3%	20,8%
	Total Pobreza	4,6%	0,8%	3,3%	6,3%	16,4%
	No pobre	95,4%	0,8%	93,7%	96,7%	0,8%
Magallanes	Pobreza extrema	0,7%	0,2%	0,4%	1,3%	31,3%
	Pobreza no extrema	1,4%	0,3%	0,9%	2,2%	22,3%
	Total Pobreza	2,1%	0,4%	1,5%	3,0%	18,2%
	No pobre	97,9%	0,4%	97,0%	98,5%	0,4%

Elaborado por el Instituto Nacional de Estadísticas.

Tabla VI.6. Estimación de porcentaje de la hogares en situación de pobreza por ingresos, según zona y región, Casen 2017.

Región	Situación de pobreza por ingresos	Estimación	Error estándar	95% de intervalo de confianza		Coeficiente de Variación
				Inferior	Superior	
Nacional	Pobreza extrema	2,2%	0,1%	2,0%	2,4%	4,0%
	Pobreza no extrema	5,4%	0,1%	5,1%	5,6%	2,3%
	Total Pobreza	7,6%	0,2%	7,3%	7,9%	2,2%
	No pobre	92,4%	0,2%	92,1%	92,7%	0,2%
Urbano	Pobreza extrema	2,0%	0,1%	1,8%	2,2%	4,8%
	Pobreza no extrema	4,6%	0,1%	4,4%	4,9%	2,9%
	Total Pobreza	6,6%	0,2%	6,2%	6,9%	2,7%
	No pobre	93,4%	0,2%	93,1%	93,8%	0,2%
Rural	Pobreza extrema	4,0%	0,2%	3,5%	4,4%	6,0%
	Pobreza no extrema	10,7%	0,4%	10,0%	11,4%	3,3%
	Total Pobreza	14,6%	0,5%	13,7%	15,6%	3,2%
	No pobre	85,4%	0,5%	84,4%	86,3%	0,5%
Arica y Parinacota	Pobreza extrema	2,2%	0,3%	1,6%	3,0%	15,5%
	Pobreza no extrema	4,6%	0,5%	3,6%	5,8%	12,1%
	Total Pobreza	6,7%	0,7%	5,4%	8,3%	11,0%
	No pobre	93,3%	0,7%	91,7%	94,6%	0,8%
Tarapacá	Pobreza extrema	1,9%	0,2%	1,5%	2,4%	12,8%
	Pobreza no extrema	4,3%	0,4%	3,6%	5,2%	9,7%
	Total Pobreza	6,2%	0,5%	5,2%	7,4%	8,8%
	No pobre	93,8%	0,5%	92,6%	94,8%	0,6%
Antofagasta	Pobreza extrema	1,4%	0,3%	0,9%	2,1%	21,5%
	Pobreza no extrema	2,6%	0,4%	1,9%	3,6%	15,7%
	Total Pobreza	4,0%	0,5%	3,1%	5,2%	12,9%
	No pobre	96,0%	0,5%	94,8%	96,9%	0,5%
Atacama	Pobreza extrema	2,1%	0,3%	1,5%	2,9%	16,5%
	Pobreza no extrema	4,3%	0,4%	3,5%	5,2%	10,2%
	Total Pobreza	6,4%	0,7%	5,2%	7,8%	10,2%
	No pobre	93,6%	0,7%	92,2%	94,8%	0,7%
Coquimbo	Pobreza extrema	2,5%	0,3%	1,9%	3,2%	13,0%
	Pobreza no extrema	7,4%	0,6%	6,3%	8,7%	8,3%
	Total Pobreza	9,9%	0,7%	8,6%	11,3%	6,9%
	No pobre	90,1%	0,7%	88,7%	91,4%	0,8%
Valparaíso	Pobreza extrema	1,6%	0,2%	1,2%	2,0%	12,8%
	Pobreza no extrema	4,6%	0,3%	4,0%	5,2%	6,7%
	Total Pobreza	6,2%	0,4%	5,5%	6,9%	5,7%
	No pobre	93,8%	0,4%	93,1%	94,5%	0,4%
Metropolitana	Pobreza extrema	1,5%	0,2%	1,2%	1,8%	10,3%
	Pobreza no extrema	3,2%	0,2%	2,9%	3,7%	6,6%
	Total Pobreza	4,7%	0,3%	4,2%	5,3%	6,1%
	No pobre	95,3%	0,3%	94,7%	95,8%	0,3%

Región	Situación de pobreza por ingresos	Estimación	Error estándar	95% de intervalo de confianza		Coeficiente de Variación
				Inferior	Superior	
O'Higgins	Pobreza extrema	1,9%	0,2%	1,5%	2,5%	12,0%
	Pobreza no extrema	6,7%	0,5%	5,8%	7,9%	7,8%
	Total Pobreza	8,7%	0,6%	7,5%	10,0%	7,4%
	No pobre	91,3%	0,6%	90,0%	92,5%	0,7%
Maule	Pobreza extrema	2,9%	0,3%	2,4%	3,6%	9,9%
	Pobreza no extrema	8,4%	0,5%	7,5%	9,4%	5,5%
	Total Pobreza	11,3%	0,5%	10,3%	12,4%	4,7%
	No pobre	88,7%	0,5%	87,6%	89,7%	0,6%
Ñuble	Pobreza extrema	4,2%	0,4%	3,5%	5,1%	9,3%
	Pobreza no extrema	10,9%	0,6%	9,8%	12,3%	5,8%
	Total Pobreza	15,2%	0,8%	13,7%	16,8%	5,2%
	No pobre	84,8%	0,8%	83,2%	86,3%	0,9%
Biobío	Pobreza extrema	3,7%	0,4%	3,0%	4,6%	11,3%
	Pobreza no extrema	7,5%	0,5%	6,6%	8,4%	6,1%
	Total Pobreza	11,2%	0,7%	9,9%	12,6%	6,1%
	No pobre	88,8%	0,7%	87,4%	90,1%	0,8%
La Araucanía	Pobreza extrema	4,0%	0,3%	3,4%	4,7%	8,1%
	Pobreza no extrema	10,5%	0,5%	9,5%	11,5%	5,0%
	Total Pobreza	14,5%	0,7%	13,2%	15,9%	4,7%
	No pobre	85,5%	0,7%	84,1%	86,8%	0,8%
Los Ríos	Pobreza extrema	2,8%	0,4%	2,1%	3,7%	13,6%
	Pobreza no extrema	7,6%	0,6%	6,5%	8,9%	8,0%
	Total Pobreza	10,4%	0,8%	9,0%	12,1%	7,6%
	No pobre	89,6%	0,8%	87,9%	91,0%	0,9%
Los Lagos	Pobreza extrema	3,2%	0,4%	2,6%	4,0%	11,2%
	Pobreza no extrema	7,0%	0,5%	6,1%	8,0%	6,9%
	Total Pobreza	10,2%	0,7%	9,0%	11,7%	6,8%
	No pobre	89,8%	0,7%	88,3%	91,0%	0,8%
Aysén	Pobreza extrema	0,9%	0,2%	0,6%	1,4%	23,8%
	Pobreza no extrema	2,8%	0,5%	2,0%	4,0%	18,2%
	Total Pobreza	3,7%	0,6%	2,7%	5,1%	16,1%
	No pobre	96,3%	0,6%	94,9%	97,3%	0,6%
Magallanes	Pobreza extrema	0,6%	0,2%	0,4%	1,1%	27,1%
	Pobreza no extrema	1,2%	0,3%	0,8%	2,0%	24,6%
	Total Pobreza	1,9%	0,3%	1,3%	2,7%	18,0%
	No pobre	98,1%	0,3%	97,3%	98,7%	0,3%

Elaborado por el Instituto Nacional de Estadísticas.

Anexo N°1. Composición de las nuevas regiones de Biobío y Ñuble

VIII Región de Biobío	XVI Región de Ñuble
<p>81 Concepción</p> <p>8101 Concepción 8102 Coronel 8103 Chiguayante 8104 Florida 8105 Hualqui 8106 Lota 8107 Penco 8108 San Pedro de la Paz 8109 Santa Juana 8110 Talcahuano 8111 Tomé 8112 Hualpén</p> <p>82 Arauco</p> <p>8201 Lebu 8202 Arauco 8203 Cañete 8204 Contulmo 8205 Curanilahue 8206 Los Álamos 8207 Tirúa</p> <p>83 Biobío</p> <p>8301 Los Ángeles 8302 Antuco 8303 Cabrero 8304 Laja 8305 Mulchén 8306 Nacimiento 8307 Negrete 8308 Quilaco 8309 Quilleco 8310 San Rosendo 8311 Santa Bárbara 8312 Tucapel 8313 Yumbel 8314 Alto Biobío</p>	<p>161 Diguillín</p> <p>16101 Chillán 16102 Bulnes 16103 Chillán Viejo 16104 El Carmen 16105 Pemuco 16106 Pinto 16107 Quillón 16108 San Ignacio 16109 Yungay</p> <p>162 Itata</p> <p>16201 Quirihue 16202 Cobquecura 16203 Coelemu 16204 Ninhue 16205 Portezuelo 16206 Ránquil 16207 Treguaco</p> <p>163 Punilla</p> <p>16301 San Carlos 16302 Coihueco 16303 Ñiquén 16304 San Fabián 16305 San Nicolás</p>

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Arica y Parinacota	61.313	2.408	2.155	253	21,5%	3.066	2.756	310	384	3	16
Arica	60.396	2.293	2.105	188	22,1%	2.943	2.702	241	384	0	12
Camarones	351	32	0	32	3,6%	33	0	33	0	0	2
Putre	566	83	50	33	8,2%	90	54	36	0	3	2
Tarapacá	81.824	2.974	2.733	241	29,3%	4.206	3.926	280	406	3	14
Iquique	50.075	1.845	1.807	38	34,9%	2.834	2.776	58	274	0	3
Alto Hospicio	26.649	819	819	0	20,3%	1.027	1.027	0	121	0	0
Pozo Almonte	2.913	92	53	39	8,8%	101	58	43	11	0	2
Camiña	389	38	0	38	3,7%	39	0	39	0	0	2
Huara	810	90	0	90	7,0%	97	0	97	0	0	5
Pica	988	90	54	36	16,7%	108	65	43	0	3	2
Antofagasta	144.072	2.511	2.348	163	33,0%	3.746	3.533	213	466	3	11
Antofagasta	85.651	1.509	1.509	0	35,0%	2.321	2.321	0	294	0	0
Mejillones	2.649	50	50	0	26,7%	68	68	0	11	0	0
Sierra Gorda	336	33	0	33	15,8%	39	0	39	0	0	2
Taltal	3.118	80	50	30	15,6%	95	59	36	10	0	2
Calama	40.364	560	526	34	35,2%	863	811	52	117	0	3
San Pedro de Atacama	1.149	84	50	34	26,0%	114	68	46	0	3	2
Tocopilla	8.190	141	109	32	19,1%	175	135	40	20	0	2
María Elena	2.615	54	54	0	24,1%	71	71	0	14	0	0

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Atacama	81.301	2.331	2.054	277	23,5%	3.048	2.722	326	446	0	17
Copiapó	40.471	1.234	1.195	39	29,7%	1.755	1.700	55	244	0	3
Caldera	5.923	144	144	0	20,2%	180	180	0	35	0	0
Tierra Amarilla	3.859	107	65	42	14,4%	125	76	49	14	0	2
Chañaral	4.442	67	67	0	20,4%	84	84	0	18	0	0
Diego de Almagro	6.238	171	141	30	16,3%	204	168	36	39	0	2
Vallenar	14.458	379	337	42	15,4%	448	398	50	72	0	3
Alto del Carmen	1.264	48	0	48	8,4%	52	0	52	0	0	3
Freirina	1.692	92	50	42	8,3%	101	55	46	11	0	2
Huasco	2.954	89	55	34	10,3%	99	61	38	13	0	2
Coquimbo	199.142	3.028	2.390	638	21,2%	3.841	3.075	766	430	9	40
La Serena	54.162	870	824	46	26,8%	1.189	1.126	63	147	0	3
Coquimbo	55.830	697	651	46	23,3%	909	849	60	123	0	3
Andacollo	3.516	80	50	30	12,2%	91	57	34	12	0	2
La Higuera	1.102	84	50	34	5,7%	89	53	36	0	3	2
Paiguano	1.249	34	0	34	2,6%	35	0	35	0	0	2
Vicuña	8.591	98	50	48	6,6%	105	54	51	10	0	3
Illapel	9.056	128	81	47	11,8%	145	92	53	18	0	3
Canela	2.796	83	50	33	15,2%	98	59	39	0	3	2
Los Vilos	6.622	165	133	32	22,1%	212	171	41	30	0	2
Salamanca	7.564	85	53	32	15,0%	100	62	38	10	0	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Ovalle	30.015	389	297	92	18,6%	478	365	113	55	0	6
Combarbalá	4.869	84	50	34	25,2%	112	67	45	13	0	2
Monte Patria	9.536	115	51	64	14,4%	135	60	75	12	0	4
Punitaqui	2.737	83	50	33	16,1%	99	60	39	0	3	2
Río Hurtado	1.497	33	0	33	24,3%	44	0	44	0	0	2
Valparaíso	571.793	6.717	5.621	1.096	27,8%	9.298	7.899	1.399	972	16	71
Valparaíso	96.405	988	958	30	30,2%	1.415	1.372	43	154	0	2
Casablanca	8.099	85	55	30	13,4%	98	63	35	12	0	2
Concón	12.448	107	107	0	40,9%	181	181	0	20	0	0
Puchuncaví	9.939	80	50	30	13,9%	93	58	35	7	0	2
Quintero	11.596	80	50	30	17,8%	97	61	36	12	0	2
Viña del Mar	108.927	1.127	1.127	0	40,0%	1.878	1.878	0	199	0	0
Los Andes	19.725	293	262	31	31,3%	426	381	45	55	0	2
Calle Larga	3.626	81	50	31	32,8%	120	74	46	8	0	2
Rinconada	2.774	80	50	30	17,7%	97	61	36	15	0	2
San Esteban	5.827	81	50	31	22,4%	104	64	40	7	0	2
La Ligua	9.761	112	81	31	16,2%	134	97	37	19	0	2
Cabildo	5.613	81	50	31	15,5%	96	59	37	11	0	2
Papudo	1.151	50	50	0	12,6%	57	57	0	0	3	0
Petorca	2.556	82	50	32	20,2%	103	63	40	0	3	2
Zapallar	1.462	83	52	31	31,6%	121	76	45	0	4	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Quillota	24.258	278	248	30	25,7%	374	334	40	40	0	2
Calera	15.061	160	130	30	14,2%	187	152	35	20	0	2
Hijuelas	4.839	81	50	31	11,7%	92	57	35	9	0	2
La Cruz	5.088	81	50	31	14,8%	95	59	36	8	0	2
Nogales	6.482	80	50	30	16,6%	96	60	36	12	0	2
San Antonio	27.395	297	267	30	19,3%	368	331	37	49	0	2
Algarrobo	11.666	81	51	30	25,8%	109	69	40	9	0	2
Cartagena	12.785	83	53	30	26,8%	113	72	41	14	0	2
El Quisco	13.161	50	50	0	26,2%	68	68	0	14	0	0
El Tabo	7.475	50	50	0	28,4%	70	70	0	14	0	0
Santo Domingo	1.858	83	52	31	15,5%	99	62	37	0	3	2
San Felipe	21.793	218	187	31	35,4%	338	290	48	39	0	2
Catemu	3.674	81	50	31	23,1%	105	65	40	13	0	2
Llaillay	7.390	81	50	31	16,3%	97	60	37	10	0	2
Panquehue	1.680	81	50	31	28,0%	112	69	43	0	3	2
Putendo	4.461	80	50	30	34,1%	121	76	45	14	0	2
Santa María	4.304	82	50	32	30,8%	118	72	46	8	0	2
Quilpué	46.189	561	531	30	22,2%	722	683	39	87	0	2
Limache	12.835	295	117	178	15,9%	351	139	212	18	0	11
Olmué	5.390	81	50	31	17,4%	99	61	38	9	0	2
Villa Alemana	34.100	423	393	30	22,1%	544	505	39	66	0	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Metropolitana	1.842.153	12.954	12.227	727	43,8%	23.051	22.079	972	2.000	5	49
Santiago	112.526	765	765	0	55,8%	1.732	1.732	0	105	0	0
Cerrillos	21.577	114	114	0	20,1%	143	143	0	17	0	0
Cerro Navia	35.165	193	193	0	18,4%	237	237	0	29	0	0
Conchalí	32.901	187	187	0	22,9%	243	243	0	31	0	0
El Bosque	43.372	240	240	0	24,2%	317	317	0	36	0	0
Estación Central	33.549	230	230	0	24,8%	306	306	0	36	0	0
Huechuraba	20.505	107	107	0	21,4%	136	136	0	17	0	0
Independencia	21.682	134	134	0	29,1%	189	189	0	16	0	0
La Cisterna	24.910	156	156	0	35,6%	242	242	0	22	0	0
La Florida	102.735	632	632	0	38,3%	1.024	1.024	0	110	0	0
La Granja	31.386	209	209	0	17,8%	254	254	0	29	0	0
La Pintana	45.819	246	246	0	21,0%	311	311	0	33	0	0
La Reina	26.563	180	180	0	48,1%	347	347	0	34	0	0
Las Condes	101.078	856	856	0	60,7%	2.178	2.178	0	154	0	0
Lo Barnechea	20.790	291	291	0	68,4%	922	922	0	73	0	0
Lo Espejo	24.763	127	127	0	21,8%	162	162	0	18	0	0
Lo Prado	26.046	156	156	0	13,3%	180	180	0	20	0	0
Macul	32.467	208	208	0	30,8%	301	301	0	28	0	0
Maipú	140.885	887	887	0	30,5%	1.276	1.276	0	144	0	0
Ñuñoa	67.647	550	550	0	57,3%	1.288	1.288	0	91	0	0

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Pedro Aguirre Cerda	27.638	153	153	0	19,9%	191	191	0	23	0	0
Peñalolén	57.064	380	380	0	28,2%	529	529	0	56	0	0
Providencia	63.695	876	876	0	69,7%	2.893	2.893	0	161	0	0
Pudahuel	55.722	283	283	0	17,9%	345	345	0	38	0	0
Quilicura	49.796	235	235	0	18,7%	289	289	0	29	0	0
Quinta Normal	29.409	148	148	0	22,6%	191	191	0	19	0	0
Recoleta	39.553	249	249	0	29,1%	351	351	0	38	0	0
Renca	36.463	197	197	0	16,9%	237	237	0	25	0	0
San Joaquín	23.951	171	171	0	40,8%	289	289	0	35	0	0
San Miguel	28.977	172	172	0	44,5%	310	310	0	24	0	0
San Ramón	21.772	130	130	0	21,3%	165	165	0	22	0	0
Vitacura	26.993	410	410	0	70,1%	1.372	1.372	0	130	0	0
Puente Alto	152.016	727	727	0	20,3%	912	912	0	99	0	0
Pirque	4.465	82	50	32	27,0%	112	68	44	9	0	2
San José de Maipo	3.499	82	50	32	37,0%	130	79	51	12	0	3
Colina	23.106	213	164	49	29,9%	304	234	70	23	0	4
Lampa	15.487	209	97	112	25,3%	280	130	150	16	0	8
Tiltil	4.174	112	63	49	18,8%	138	78	60	14	0	3
San Bernardo	70.653	420	390	30	28,9%	590	548	42	62	0	2
Buín	19.107	108	78	30	25,3%	144	104	40	19	0	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Calera de Tango	4.621	81	50	31	36,7%	128	79	49	9	0	2
Paine	14.706	116	83	33	29,0%	163	117	46	15	0	2
Melipilla	28.382	169	117	52	24,6%	224	155	69	21	0	3
Alhué	1.211	81	50	31	9,8%	89	55	34	0	3	2
Curacaví	7.339	82	50	32	28,2%	115	70	45	9	0	2
María Pinto	2.725	80	50	30	23,7%	105	66	39	0	2	2
San Pedro	2.090	30	0	30	8,8%	33	0	33	0	0	2
Talagante	15.462	115	83	32	16,9%	138	100	38	11	0	2
El Monte	7.509	80	50	30	19,7%	99	62	37	7	0	2
Isla de Maipo	6.931	82	50	32	28,2%	115	70	45	10	0	2
Padre Hurtado	10.520	80	50	30	24,6%	106	66	40	9	0	2
Peñaflor	20.751	133	103	30	24,5%	176	136	40	12	0	2
O'Higgins	241.593	5.099	3.585	1.514	17,7%	6.194	4.382	1.812	469	31	90
Rancagua	67.740	1.180	1.138	42	20,9%	1.492	1.439	53	171	0	3
Codegua	3.014	84	50	34	11,7%	95	57	38	10	0	2
Coinco	1.735	84	50	34	15,7%	99	59	40	0	3	2
Coltauco	4.309	88	50	38	13,6%	102	58	44	0	3	2
Doñihue	5.523	116	86	30	8,4%	127	94	33	11	0	2
Graneros	8.647	240	206	34	16,0%	285	245	40	25	0	2
Las Cabras	6.114	118	50	68	14,3%	137	58	79	6	0	4
Machalí	10.333	184	149	35	23,2%	240	194	46	22	0	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Malloa	3.356	80	50	30	6,4%	85	53	32	0	3	2
Mostazal	6.129	133	98	35	14,0%	155	114	41	12	0	2
Olivar	3.400	84	50	34	8,8%	92	55	37	8	0	2
Peumo	4.158	103	67	36	13,7%	120	78	42	12	0	2
Pichidegua	5.175	84	50	34	9,6%	93	55	38	5	0	2
Quinta de Tilcoco	3.372	85	50	35	9,3%	94	55	39	7	0	2
Rengo	14.746	275	205	70	15,5%	326	243	83	28	0	4
Requínoa	6.076	108	50	58	12,6%	123	57	66	7	0	3
San Vicente	12.760	244	97	147	12,5%	279	111	168	16	0	8
Pichilemu	7.739	86	51	35	22,0%	110	65	45	14	0	2
La Estrella	843	85	50	35	21,8%	109	64	45	0	3	2
Litueche	2.003	85	50	35	20,1%	107	63	44	12	0	2
Marchihue	2.001	85	50	35	18,0%	104	61	43	11	0	2
Navidad	1.658	84	50	34	32,5%	124	74	50	0	4	3
Paredones	1.833	85	50	35	24,3%	112	66	46	0	3	2
San Fernando	21.003	395	286	109	20,6%	497	360	137	41	0	7
Chépica	4.274	85	50	35	23,0%	110	65	45	7	0	2
Chimbarongo	9.658	150	80	70	14,1%	174	93	81	15	0	4
Lolol	1.540	84	50	34	17,7%	102	61	41	0	3	2
Nancagua	4.392	85	55	30	18,7%	105	68	37	11	0	2
Palmilla	2.696	84	50	34	12,4%	96	57	39	0	3	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Peralillo	2.580	84	50	34	12,4%	96	57	39	0	3	2
Placilla	2.104	85	50	35	15,9%	101	59	42	0	3	2
Pumanque	905	35	0	35	16,0%	42	0	42	0	0	2
Santa Cruz	9.777	212	117	95	18,9%	261	144	117	18	0	6
Maule	285.612	5.007	3.610	1.397	14,8%	5.878	4.293	1.585	458	49	81
Talca	67.123	692	642	50	22,5%	893	828	65	107	0	3
Constitución	13.297	180	139	41	16,8%	216	167	49	16	0	2
Curepto	3.049	86	50	36	5,7%	91	53	38	0	3	2
Empedrado	1.212	80	50	30	11,8%	91	57	34	0	3	2
Maule	4.569	87	50	37	8,6%	95	55	40	0	3	2
Pelarco	1.949	80	50	30	4,4%	83	52	31	0	3	2
Pencahue	2.126	86	50	36	5,5%	91	53	38	0	3	2
Río Claro	3.341	87	50	37	6,9%	94	54	40	0	3	2
San Clemente	11.244	140	50	90	7,3%	151	54	97	10	0	5
San Rafael	2.042	80	50	30	7,2%	86	54	32	0	3	2
Cauquenes	14.017	256	169	87	12,6%	293	193	100	22	0	5
Chanco	2.597	131	72	59	14,7%	153	84	69	14	0	3
Pelluhue	1.885	118	72	46	18,3%	144	88	56	0	4	3
Curicó	40.933	1.105	1.029	76	19,2%	1.368	1.274	94	172	0	5
Hualañé	2.920	86	50	36	23,0%	112	65	47	13	0	2
Licantén	1.881	85	50	35	11,5%	96	56	40	0	3	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Molina	13.143	127	92	35	11,9%	144	104	40	17	0	2
Rauco	2.342	86	50	36	9,1%	95	55	40	0	3	2
Romeral	3.222	86	50	36	7,1%	93	54	39	0	3	2
Sagrada Familia	4.851	80	50	30	10,4%	89	56	33	0	3	2
Teno	6.790	118	50	68	12,7%	135	57	78	10	0	4
Vichuquén	1.424	86	50	36	20,9%	109	63	46	0	3	2
Linares	27.798	309	240	69	13,4%	357	277	80	32	0	4
Colbún	4.772	80	50	30	6,3%	85	53	32	0	3	2
Longaví	7.787	118	50	68	6,0%	125	53	72	10	0	4
Parral	12.337	143	81	62	15,2%	169	96	73	14	0	4
Retiro	5.187	91	50	41	6,4%	97	53	44	0	3	2
San Javier	12.560	137	74	63	4,3%	143	77	66	12	0	3
Villa Alegre	5.053	80	50	30	9,5%	88	55	33	9	0	2
Yerbas Buenas	4.161	87	50	37	5,6%	92	53	39	0	3	2
Biobío (Antigua)	555.327	9.735	7.676	2.059	18,5%	11.942	9.507	2.435	1.092	61	127
Concepción	66.291	1.000	930	70	33,6%	1.505	1.400	105	131	0	5
Coronel	28.927	448	390	58	17,4%	542	472	70	68	0	4
Chiguayante	23.956	467	467	0	21,2%	593	593	0	58	0	0
Florida	3.221	83	50	33	14,1%	96	58	38	12	0	2
Hualqui	6.825	88	55	33	16,1%	105	66	39	9	0	2
Lota	13.800	218	218	0	12,5%	249	249	0	45	0	0

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Penco	12.156	199	169	30	16,5%	238	202	36	25	0	2
San Pedro de la Paz	32.705	324	324	0	22,7%	419	419	0	44	0	0
Santa Juana	3.415	83	50	33	21,5%	106	64	42	10	0	2
Talcahuano	44.822	657	627	30	22,2%	845	806	39	112	0	2
Tomé	15.007	216	186	30	13,6%	250	215	35	31	0	2
Hualpén	26.179	359	359	0	17,8%	437	437	0	64	0	0
Lebu	6.604	118	88	30	19,7%	147	110	37	18	0	2
Arauco	9.054	161	122	39	18,4%	198	150	48	19	0	2
Cañete	8.909	135	96	39	14,7%	159	113	46	19	0	2
Contulmo	1.142	84	50	34	19,0%	104	62	42	0	3	2
Curanilahue	8.502	203	169	34	17,1%	245	204	41	33	0	2
Los Álamos	5.000	99	69	30	16,8%	119	83	36	18	0	2
Tirúa	1.905	83	50	33	14,5%	97	58	39	0	3	2
Los Ángeles	48.820	702	515	187	14,7%	823	604	219	69	0	11
Antuco	1.128	83	50	33	5,0%	88	53	35	0	3	2
Cabrero	8.031	134	104	30	11,8%	152	118	34	21	0	2
Laja	7.316	80	50	30	11,1%	90	56	34	10	0	2
Mulchén	8.600	116	86	30	8,4%	127	94	33	15	0	2
Nacimiento	7.941	117	87	30	13,1%	135	100	35	14	0	2
Negrete	2.241	80	50	30	5,6%	85	53	32	0	3	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Quilaco	1.146	80	50	30	17,6%	97	61	36	0	3	2
Quilleco	2.980	80	50	30	3,8%	83	52	31	0	3	2
San Rosendo	1.058	80	50	30	11,0%	90	56	34	0	3	2
Santa Bárbara	4.007	80	50	30	9,9%	88	55	33	7	0	2
Tucapel	4.252	80	50	30	5,5%	85	53	32	11	0	2
Yumbel	6.224	84	50	34	22,7%	109	65	44	13	0	2
Alto Biobío	1.284	80	50	30	3,7%	83	52	31	0	3	2
Chillán	50.953	745	678	67	18,0%	909	827	82	99	0	4
Bulnes	6.195	92	62	30	15,9%	110	74	36	13	0	2
Cobquecura	1.484	86	50	36	15,8%	102	59	43	0	3	2
Coelemu	4.498	105	69	36	15,5%	125	82	43	16	0	2
Coihueco	6.369	117	50	67	13,6%	136	58	78	10	0	4
Chillán Viejo	7.803	140	72	68	17,7%	171	88	83	14	0	4
El Carmen	3.284	85	50	35	12,1%	97	57	40	0	3	2
Ninhue	1.636	86	50	36	11,8%	98	57	41	0	3	2
Ñiquén	3.185	99	53	46	12,2%	112	60	52	0	3	3
Pemuco	2.199	88	53	35	14,8%	103	62	41	0	3	2
Pinto	2.947	87	52	35	10,7%	97	58	39	0	3	2
Portezuelo	1.445	87	52	35	16,4%	104	62	42	0	3	2
Quillón	5.606	91	54	37	11,1%	103	61	42	13	0	2
Quirihue	3.311	80	50	30	11,2%	90	56	34	9	0	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Ránquil	1.659	86	50	36	13,5%	100	58	42	0	3	2
San Carlos	14.479	288	197	91	16,6%	345	236	109	34	0	5
San Fabián	889	111	65	46	13,4%	128	75	53	0	4	3
San Ignacio	4.355	92	50	42	16,5%	110	60	50	0	3	3
San Nicolás	2.735	102	57	45	12,1%	116	65	51	0	3	3
Treguaco	1.501	86	50	36	18,0%	105	61	44	0	3	2
Yungay	5.346	81	51	30	11,5%	92	58	34	8	0	2
La Araucanía	269.653	5.136	3.622	1.514	13,4%	5.933	4.242	1.691	535	41	89
Temuco	77.543	1.353	1.263	90	22,0%	1.735	1.620	115	217	0	6
Carahue	7.987	113	50	63	7,8%	122	54	68	6	0	3
Cunco	5.590	98	50	48	11,2%	110	56	54	11	0	3
Curarrehue	1.829	83	50	33	6,7%	89	54	35	0	3	2
Freire	7.164	139	50	89	6,9%	150	54	96	10	0	5
Galvarino	3.163	83	50	33	12,1%	95	57	38	0	3	2
Gorbea	4.909	82	50	32	7,7%	89	54	35	11	0	2
Lautaro	9.890	124	82	42	9,3%	136	90	46	18	0	2
Loncoche	8.481	147	99	48	20,7%	186	125	61	22	0	3
Melipeuco	1.494	82	50	32	9,4%	90	55	35	0	3	2
Nueva Imperial	9.303	132	69	63	8,5%	144	75	69	13	0	3
Padre Las Casas	17.379	332	222	110	11,1%	374	250	124	33	0	6
Perquenco	1.787	80	50	30	10,2%	89	56	33	0	3	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Pitrufquén	6.968	135	72	63	7,7%	146	78	68	15	0	3
Pucón	9.457	125	63	62	12,6%	143	72	71	9	0	4
Saavedra	3.722	80	50	30	5,2%	85	53	32	0	3	2
Teodoro Schmidt	4.501	100	50	50	8,3%	108	54	54	0	3	3
Toltén	3.142	80	50	30	4,4%	83	52	31	0	3	2
Vilcún	6.049	164	83	81	10,5%	183	93	90	0	5	5
Villarrica	14.864	240	169	71	8,0%	261	184	77	23	0	4
Cholchol	2.454	80	50	30	8,4%	88	55	33	0	3	2
Angol	14.978	288	253	35	9,3%	318	279	39	43	0	2
Collipulli	7.122	93	59	34	9,8%	103	65	38	13	0	2
Curacautín	5.856	94	64	30	6,0%	100	68	32	15	0	2
Ercilla	2.308	83	50	33	6,8%	89	54	35	0	3	2
Lonquimay	2.824	83	50	33	5,0%	88	53	35	0	3	2
Los Sauces	2.091	82	50	32	9,8%	90	55	35	0	3	2
Lumaco	2.952	83	50	33	8,3%	91	55	36	0	3	2
Purén	4.211	82	50	32	9,6%	90	55	35	12	0	2
Renaico	2.962	83	50	33	8,2%	90	54	36	10	0	2
Traiguén	6.139	127	94	33	12,2%	145	107	38	23	0	2
Victoria	10.534	186	130	56	16,6%	223	156	67	31	0	3
Los Ríos	103.351	3.624	2.487	1.137	18,0%	4.419	3.110	1.309	391	9	66
Valdivia	41.434	1.738	1.489	249	26,0%	2.348	2.012	336	245	0	17

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Corral	1.310	81	51	30	11,9%	92	58	34	0	3	2
Lanco	4.360	120	67	53	8,1%	131	73	58	13	0	3
Los Lagos	5.696	153	67	86	7,1%	165	72	93	13	0	5
Máfil	1.924	87	51	36	11,3%	99	58	41	0	3	2
Mariquina	5.107	222	109	113	10,1%	247	121	126	25	0	6
Paillaco	5.655	167	93	74	9,8%	185	103	82	19	0	4
Panguipulli	10.449	246	87	159	5,8%	261	92	169	14	0	8
La Unión	11.311	294	189	105	9,8%	326	210	116	25	0	6
Futroneo	4.368	115	64	51	12,7%	131	73	58	7	0	3
Lago Ranco	1.616	86	50	36	11,9%	98	57	41	0	3	2
Río Bueno	10.121	315	170	145	6,3%	336	181	155	30	0	8
Los Lagos	208.523	4.129	2.927	1.202	16,4%	4.938	3.554	1.384	410	30	71
Puerto Montt	53.914	862	711	151	21,2%	1.094	902	192	113	0	10
Calbuco	8.741	162	54	108	3,7%	168	56	112	9	0	6
Fresia	4.415	84	50	34	8,2%	91	54	37	10	0	2
Frutillar	6.306	84	50	34	14,0%	98	58	40	10	0	2
Los Muermos	4.126	93	50	43	6,1%	99	53	46	11	0	2
Llanquihue	4.641	88	52	36	7,0%	95	56	39	10	0	2
Maullín	4.504	108	50	58	6,7%	116	54	62	0	3	3
Puerto Varas	10.665	197	147	50	26,6%	268	200	68	24	0	3
Castro	11.511	221	154	67	18,2%	270	188	82	22	0	4

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

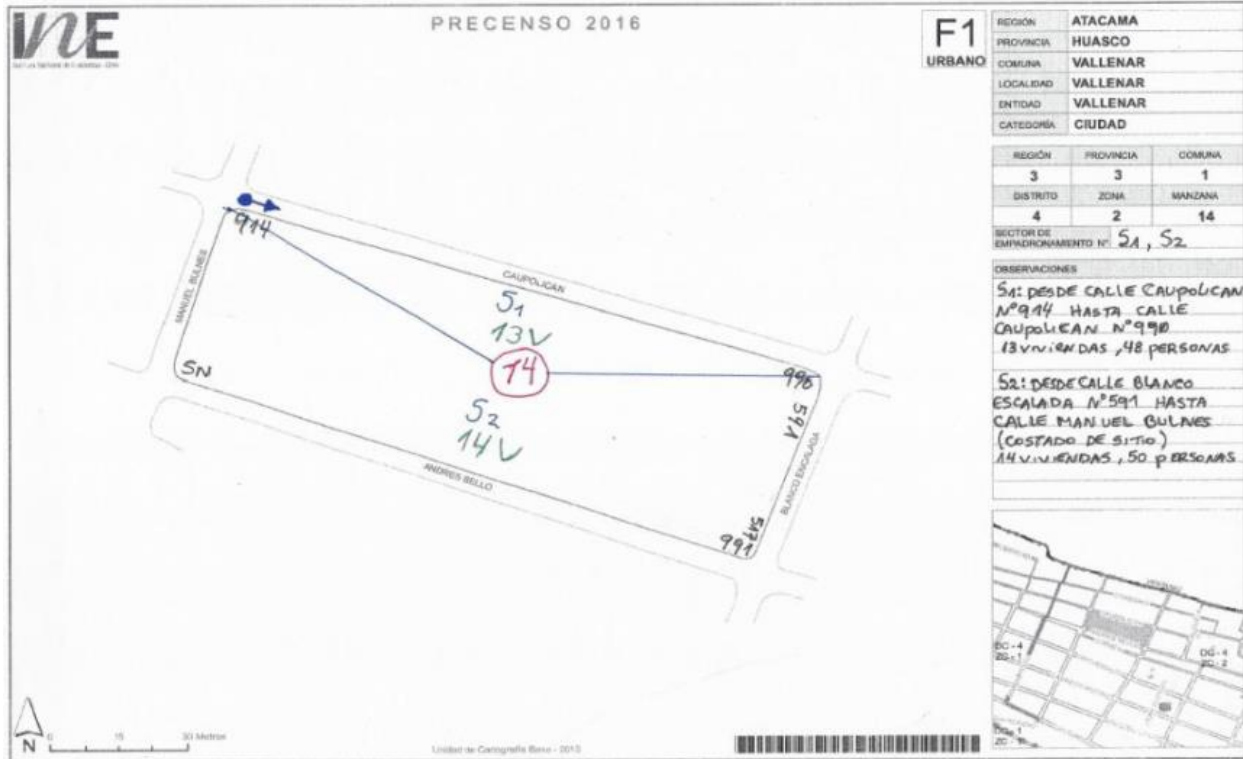
Región / Comuna	Población Viviendas en el Marco	Muestra Objetivo Casen 2017	Muestra Objetivo (Urbana) Casen 2017	Muestra Objetivo (Rural) Casen 2017	Tasa de No Resp.	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
	M_r	m_{2r}	$m_{U\ 2r}$	$m_{R\ 2r}$	tnr_r	m_{3r}	$m_{U\ 3r}$	$m_{R\ 3r}$	n	n_{RAU}	n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Ancud	12.237	206	139	67	14,9%	242	163	79	22	0	4
Chonchi	3.006	84	50	34	9,4%	93	55	38	0	3	2
Curaco de Vélez	890	30	0	30	3,1%	31	0	31	0	0	2
Dalcahue	2.621	84	50	34	7,0%	91	54	37	0	3	2
Puqueldón	1.018	34	0	34	3,6%	35	0	35	0	0	2
Queilén	1.419	84	50	34	6,5%	89	53	36	0	3	2
Quellón	6.835	160	110	50	17,4%	194	133	61	14	0	3
Quemchi	2.064	80	50	30	6,3%	85	53	32	0	3	2
Quinchao	2.233	85	50	35	15,0%	100	59	41	0	3	2
Osorno	45.469	835	769	66	21,5%	1.064	980	84	137	0	4
Puerto Octay	2.738	80	50	30	7,7%	87	54	33	0	3	2
Purranque	5.975	125	91	34	14,7%	147	107	40	18	0	2
Puyehue	3.121	85	50	35	7,7%	92	54	38	0	3	2
Río Negro	4.498	89	50	39	13,9%	103	58	45	10	0	2
San Juan de la Costa	2.613	85	50	35	8,8%	93	55	38	0	3	2
San Pablo	2.963	84	50	34	9,7%	93	55	38	0	3	2
Aysén	29.541	1.862	1.484	378	25,7%	2.506	1.989	517	277	11	25
Coyhaique	18.990	1.118	909	209	25,9%	1.508	1.226	282	177	0	14
Aysén	5.972	451	419	32	23,0%	586	544	42	100	0	2
Cisnes	1.239	80	50	30	24,2%	106	66	40	0	3	2

Anexo N°2. Distribución inicial de los tamaños muestrales de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Población Viviendas en el Marco M_r	Muestra Objetivo Casen 2017 m_{2r}	Muestra Objetivo (Urbana) Casen 2017 $m_{U 2r}$	Muestra Objetivo (Rural) Casen 2017 $m_{R 2r}$	Tasa de No Resp. tnr_r	Sobre-muestra Casen 2017 m_{3r}	Sobre-muestra (Urbana) Casen 2017 $m_{U 3r}$	Sobre-muestra (Rural) Casen 2017 $m_{R 3r}$	Muestra Manzanas n	Muestra Secciones RAU n_{RAU}	Muestra Secciones Rural n_R
Nivel País	4.723.693	69.816	57.087	12.729	26,6%	95.159	79.992	15.167	9.179	274	776
Cochrane	990	81	50	31	32,8%	120	74	46	0	4	2
Chile Chico	1.478	88	56	32	29,4%	124	79	45	0	4	2
Río Ibáñez	872	44	0	44	28,7%	62	0	62	0	0	3
Magallanes	48.495	2.301	2.168	133	25,6%	3.093	2.925	168	443	3	9
Punta Arenas	40.437	1.981	1.908	73	27,2%	2.722	2.622	100	396	0	5
Porvenir	1.458	84	54	30	10,0%	93	60	33	0	3	2
Natales	6.600	236	206	30	15,2%	278	243	35	47	0	2

Elaborado por el Instituto Nacional de Estadísticas.

Anexo N°3. Croquis de la manzana (F1).



Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Arica y Parinacota	3.066	2.756	310	384	3	16	3.066	2.756	310	384	3	16
Arica	2.943	2.702	241	384	0	12	2.943	2.702	241	384	0	12
Camarones	33	0	33	0	0	2	33	0	33	0	0	2
Putre	90	54	36	0	3	2	90	54	36	0	3	2
Tarapacá	4.206	3.926	280	406	3	14	4.206	3.926	280	405	3	14
Iquique	2.834	2.776	58	274	0	3	2.834	2.776	58	273	0	3
Alto Hospicio	1.027	1.027	0	121	0	0	1.027	1.027	0	121	0	0
Pozo Almonte	101	58	43	11	0	2	101	58	43	11	0	2
Camíña	39	0	39	0	0	2	39	0	39	0	0	2
Huara	97	0	97	0	0	5	97	0	97	0	0	5
Pica	108	65	43	0	3	2	108	65	43	0	3	2
Antofagasta	3.746	3.533	213	466	3	11	3.746	3.533	213	465	3	11
Antofagasta	2.321	2.321	0	294	0	0	2.321	2.321	0	294	0	0
Mejillones	68	68	0	11	0	0	68	68	0	11	0	0
Sierra Gorda	39	0	39	0	0	2	39	0	39	0	0	2
Taltal	95	59	36	10	0	2	95	59	36	10	0	2
Calama	863	811	52	117	0	3	863	811	52	117	0	3
San Pedro de Atacama	114	68	46	0	3	2	114	68	46	0	3	2
Tocopilla	175	135	40	20	0	2	175	135	40	20	0	2
María Elena	71	71	0	14	0	0	71	71	0	13	0	0
Atacama	3.048	2.722	326	446	0	17	3.048	2.722	326	450	0	17
Copiapó	1.755	1.700	55	244	0	3	1.755	1.700	55	248	0	3
Caldera	180	180	0	35	0	0	180	180	0	35	0	0
Tierra Amarilla	125	76	49	14	0	2	125	76	49	14	0	2
Chañaral	84	84	0	18	0	0	84	84	0	18	0	0
Diego de Almagro	204	168	36	39	0	2	204	168	36	39	0	2
Vallenar	448	398	50	72	0	3	448	398	50	72	0	3
Alto del Carmen	52	0	52	0	0	3	52	0	52	0	0	3

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Freirina	101	55	46	11	0	2	101	55	46	11	0	2
Huasco	99	61	38	13	0	2	99	61	38	13	0	2
Coquimbo	3.841	3.075	766	430	9	40	3.841	3.075	766	430	9	40
La Serena	1.189	1.126	63	147	0	3	1.189	1.126	63	147	0	3
Coquimbo	909	849	60	123	0	3	909	849	60	124	0	3
Andacollo	91	57	34	12	0	2	91	57	34	12	0	2
La Higuera	89	53	36	0	3	2	89	53	36	0	3	2
Paiguano	35	0	35	0	0	2	35	0	35	0	0	2
Vicuña	105	54	51	10	0	3	105	54	51	10	0	3
Illapel	145	92	53	18	0	3	145	92	53	18	0	3
Canela	98	59	39	0	3	2	98	59	39	0	3	2
Los Vilos	212	171	41	30	0	2	212	171	41	29	0	2
Salamanca	100	62	38	10	0	2	100	62	38	10	0	2
Ovalle	478	365	113	55	0	6	478	365	113	55	0	6
Combarbalá	112	67	45	13	0	2	112	67	45	13	0	2
Monte Patria	135	60	75	12	0	4	135	60	75	12	0	4
Punitaqui	99	60	39	0	3	2	99	60	39	0	3	2
Río Hurtado	44	0	44	0	0	2	44	0	44	0	0	2
Valparaíso	9.298	7.899	1.399	972	16	71	9.298	7.899	1.399	975	16	71
Valparaíso	1.415	1.372	43	154	0	2	1.415	1.372	43	155	0	2
Casablanca	98	63	35	12	0	2	98	63	35	12	0	2
Concón	181	181	0	20	0	0	180	180	0	20	0	0
Puchuncaví	93	58	35	7	0	2	94	59	35	8	0	2
Quintero	97	61	36	12	0	2	96	60	36	12	0	2
Viña del Mar	1.878	1.878	0	199	0	0	1.879	1.879	0	199	0	0
Los Andes	426	381	45	55	0	2	426	381	45	57	0	2
Calle Larga	120	74	46	8	0	2	120	74	46	8	0	2
Rinconada	97	61	36	15	0	2	97	61	36	15	0	2
San Esteban	104	64	40	7	0	2	104	64	40	7	0	2
La Ligua	134	97	37	19	0	2	134	97	37	19	0	2

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Cabildo	96	59	37	11	0	2	96	59	37	11	0	2
Papudo	57	57	0	0	3	0	57	57	0	0	3	0
Petorca	103	63	40	0	3	2	103	63	40	0	3	2
Zapallar	121	76	45	0	4	2	121	76	45	0	4	2
Quillota	374	334	40	40	0	2	374	334	40	40	0	2
Calera	187	152	35	20	0	2	188	153	35	20	0	2
Hijuelas	92	57	35	9	0	2	92	57	35	9	0	2
La Cruz	95	59	36	8	0	2	95	59	36	8	0	2
Nogales	96	60	36	12	0	2	96	60	36	12	0	2
San Antonio	368	331	37	49	0	2	368	331	37	49	0	2
Algarrobo	109	69	40	9	0	2	109	69	40	9	0	2
Cartagena	113	72	41	14	0	2	113	72	41	13	0	2
El Quisco	68	68	0	14	0	0	67	67	0	14	0	0
El Tabo	70	70	0	14	0	0	70	70	0	14	0	0
Santo Domingo	99	62	37	0	3	2	99	62	37	0	3	2
San Felipe	338	290	48	39	0	2	338	290	48	39	0	2
Catemu	105	65	40	13	0	2	104	64	40	13	0	2
Llailay	97	60	37	10	0	2	97	60	37	10	0	2
Panquehue	112	69	43	0	3	2	112	69	43	0	3	2
Putendo	121	76	45	14	0	2	121	76	45	14	0	2
Santa María	118	72	46	8	0	2	118	72	46	8	0	2
Quilpué	722	683	39	87	0	2	723	684	39	87	0	2
Limache	351	139	212	18	0	11	351	139	212	18	0	11
Olmué	99	61	38	9	0	2	99	61	38	9	0	2
Villa Alemana	544	505	39	66	0	2	544	505	39	66	0	2
Metropolitana	23.051	22.079	972	2.000	5	49	23.051	22.079	972	2.005	5	48
Santiago	1.732	1.732	0	105	0	0	1.732	1.732	0	103	0	0
Cerrillos	143	143	0	17	0	0	143	143	0	17	0	0
Cerro Navia	237	237	0	29	0	0	237	237	0	29	0	0
Conchalí	243	243	0	31	0	0	243	243	0	31	0	0

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
El Bosque	317	317	0	36	0	0	317	317	0	36	0	0
Estación Central	306	306	0	36	0	0	306	306	0	36	0	0
Huechuraba	136	136	0	17	0	0	136	136	0	17	0	0
Independencia	189	189	0	16	0	0	189	189	0	16	0	0
La Cisterna	242	242	0	22	0	0	242	242	0	22	0	0
La Florida	1.024	1.024	0	110	0	0	1.024	1.024	0	110	0	0
La Granja	254	254	0	29	0	0	254	254	0	29	0	0
La Pintana	311	311	0	33	0	0	311	311	0	33	0	0
La Reina	347	347	0	34	0	0	347	347	0	34	0	0
Las Condes	2.178	2.178	0	154	0	0	2.178	2.178	0	153	0	0
Lo Barnechea	922	922	0	73	0	0	922	922	0	78	0	0
Lo Espejo	162	162	0	18	0	0	162	162	0	18	0	0
Lo Prado	180	180	0	20	0	0	180	180	0	20	0	0
Macul	301	301	0	28	0	0	301	301	0	28	0	0
Maipú	1.276	1.276	0	144	0	0	1.276	1.276	0	144	0	0
Ñuñoa	1.288	1.288	0	91	0	0	1.288	1.288	0	91	0	0
Pedro Aguirre Cerda	191	191	0	23	0	0	191	191	0	23	0	0
Peñalolén	529	529	0	56	0	0	529	529	0	56	0	0
Providencia	2.893	2.893	0	161	0	0	2.893	2.893	0	161	0	0
Pudahuel	345	345	0	38	0	0	345	345	0	38	0	0
Quilicura	289	289	0	29	0	0	289	289	0	29	0	0
Quinta Normal	191	191	0	19	0	0	191	191	0	19	0	0
Recoleta	351	351	0	38	0	0	351	351	0	38	0	0
Renca	237	237	0	25	0	0	237	237	0	25	0	0
San Joaquín	289	289	0	35	0	0	289	289	0	35	0	0
San Miguel	310	310	0	24	0	0	310	310	0	24	0	0
San Ramón	165	165	0	22	0	0	165	165	0	22	0	0
Vitacura	1.372	1.372	0	130	0	0	1.372	1.372	0	130	0	0
Puente Alto	912	912	0	99	0	0	912	912	0	99	0	0

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Pirque	112	68	44	9	0	2	112	68	44	9	0	2
San José de Maipo	130	79	51	12	0	3	130	79	51	14	0	3
Colina	304	234	70	23	0	4	304	234	70	23	0	4
Lampa	280	130	150	16	0	8	280	130	150	16	0	8
Tiltil	138	78	60	14	0	3	138	78	60	14	0	3
San Bernardo	590	548	42	62	0	2	590	548	42	62	0	2
Buín	144	104	40	19	0	2	129	104	25	19	0	1
Calera de Tango	128	79	49	9	0	2	133	79	54	10	0	2
Paine	163	117	46	15	0	2	163	117	46	15	0	2
Melipilla	224	155	69	21	0	3	224	155	69	21	0	3
Alhué	89	55	34	0	3	2	89	55	34	0	3	2
Curacaví	115	70	45	9	0	2	115	70	45	9	0	2
María Pinto	105	66	39	0	2	2	105	66	39	0	2	2
San Pedro	33	0	33	0	0	2	33	0	33	0	0	2
Talagante	138	100	38	11	0	2	138	100	38	11	0	2
El Monte	99	62	37	7	0	2	99	62	37	7	0	2
Isla de Maipo	115	70	45	10	0	2	125	70	55	10	0	2
Padre Hurtado	106	66	40	9	0	2	106	66	40	9	0	2
Peñaflor	176	136	40	12	0	2	176	136	40	12	0	2
O'Higgins	6.194	4.382	1.812	469	31	90	6.194	4.382	1.812	471	31	90
Rancagua	1.492	1.439	53	171	0	3	1.492	1.439	53	169	0	3
Codegua	95	57	38	10	0	2	95	57	38	10	0	2
Coinco	99	59	40	0	3	2	99	59	40	0	3	2
Coltauco	102	58	44	0	3	2	102	58	44	0	3	2
Doñihue	127	94	33	11	0	2	127	94	33	11	0	2
Graneros	285	245	40	25	0	2	285	245	40	25	0	2
Las Cabras	137	58	79	6	0	4	137	58	79	6	0	4
Machalí	240	194	46	22	0	2	240	194	46	23	0	2
Malloa	85	53	32	0	3	2	85	53	32	0	3	2

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Mostazal	155	114	41	12	0	2	155	114	41	12	0	2
Olivar	92	55	37	8	0	2	92	55	37	8	0	2
Peumo	120	78	42	12	0	2	120	78	42	12	0	2
Pichidegua	93	55	38	5	0	2	93	55	38	6	0	2
Quinta de Tilcoco	94	55	39	7	0	2	94	55	39	7	0	2
Rengo	326	243	83	28	0	4	326	243	83	28	0	4
Requínoa	123	57	66	7	0	3	123	57	66	7	0	3
San Vicente	279	111	168	16	0	8	279	111	168	16	0	8
Pichilemu	110	65	45	14	0	2	110	65	45	14	0	2
La Estrella	109	64	45	0	3	2	109	64	45	0	3	2
Litueche	107	63	44	12	0	2	107	63	44	12	0	2
Marchihue	104	61	43	11	0	2	104	61	43	11	0	2
Navidad	124	74	50	0	4	3	124	74	50	0	4	3
Paredones	112	66	46	0	3	2	112	66	46	0	3	2
San Fernando	497	360	137	41	0	7	497	360	137	42	0	7
Chépica	110	65	45	7	0	2	110	65	45	7	0	2
Chimbarongo	174	93	81	15	0	4	174	93	81	15	0	4
Lolol	102	61	41	0	3	2	102	61	41	0	3	2
Nancagua	105	68	37	11	0	2	105	68	37	11	0	2
Palmilla	96	57	39	0	3	2	96	57	39	0	3	2
Peralillo	96	57	39	0	3	2	96	57	39	0	3	2
Placilla	101	59	42	0	3	2	101	59	42	0	3	2
Pumanque	42	0	42	0	0	2	42	0	42	0	0	2
Santa Cruz	261	144	117	18	0	6	261	144	117	19	0	6
Maule	5.878	4.293	1.585	458	49	81	5.878	4.293	1.585	462	49	81
Talca	893	828	65	107	0	3	893	828	65	106	0	3
Constitución	216	167	49	16	0	2	235	186	49	18	1	2
Curepto	91	53	38	0	3	2	91	53	38	0	3	2
Empedrado	91	57	34	0	3	2	72	38	34	0	2	2
Maule	95	55	40	0	3	2	95	55	40	0	3	2

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Pelarco	83	52	31	0	3	2	83	52	31	0	3	2
Pencahue	91	53	38	0	3	2	91	53	38	0	3	2
Río Claro	94	54	40	0	3	2	94	54	40	0	3	2
San Clemente	151	54	97	10	0	5	151	54	97	10	0	5
San Rafael	86	54	32	0	3	2	86	54	32	0	3	2
Cauquenes	293	193	100	22	0	5	293	193	100	23	0	5
Chanco	153	84	69	14	0	3	153	84	69	14	0	3
Pelluhue	144	88	56	0	4	3	144	88	56	0	4	3
Curicó	1.368	1.274	94	172	0	5	1.368	1.274	94	174	0	5
Hualañé	112	65	47	13	0	2	112	65	47	13	0	2
Licantén	96	56	40	0	3	2	96	56	40	0	3	2
Molina	144	104	40	17	0	2	144	104	40	17	0	2
Rauco	95	55	40	0	3	2	95	55	40	0	3	2
Romeral	93	54	39	0	3	2	93	54	39	0	3	2
Sagrada Familia	89	56	33	0	3	2	89	56	33	0	3	2
Teno	135	57	78	10	0	4	135	57	78	10	0	4
Vichuquén	109	63	46	0	3	2	109	63	46	0	3	2
Linares	357	277	80	32	0	4	357	277	80	32	0	4
Colbún	85	53	32	0	3	2	85	53	32	0	3	2
Longaví	125	53	72	10	0	4	125	53	72	10	0	4
Parral	169	96	73	14	0	4	169	96	73	14	0	4
Retiro	97	53	44	0	3	2	97	53	44	0	3	2
San Javier	143	77	66	12	0	3	143	77	66	12	0	3
Villa Alegre	88	55	33	9	0	2	88	55	33	9	0	2
Yerbas Buenas	92	53	39	0	3	2	92	53	39	0	3	2
Biobío (Antigua)	11.942	9.507	2.435	1.092	61	127	11.942	9.507	2.435	1.091	61	124
Concepción	1.505	1.400	105	131	0	5	1.526	1.400	126	130	0	5
Coronel	542	472	70	68	0	4	542	472	70	67	0	4
Chiguayante	593	593	0	58	0	0	593	593	0	58	0	0
Florida	96	58	38	12	0	2	103	58	45	12	0	2

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Hualqui	105	66	39	9	0	2	105	66	39	9	0	2
Lota	249	249	0	45	0	0	249	249	0	45	0	0
Penco	238	202	36	25	0	2	238	202	36	25	0	2
San Pedro de la Paz	419	419	0	44	0	0	419	419	0	44	0	0
Santa Juana	106	64	42	10	0	2	106	64	42	10	0	2
Talcahuano	845	806	39	112	0	2	852	806	46	112	0	2
Tomé	250	215	35	31	0	2	250	215	35	31	0	2
Hualpén	437	437	0	64	0	0	437	437	0	64	0	0
Lebu	147	110	37	18	0	2	147	110	37	18	0	2
Arauco	198	150	48	19	0	2	198	150	48	19	0	2
Cañete	159	113	46	19	0	2	159	113	46	19	0	2
Contulmo	104	62	42	0	3	2	104	62	42	0	3	2
Curanilahue	245	204	41	33	0	2	245	204	41	33	0	2
Los Álamos	119	83	36	18	0	2	119	83	36	18	0	2
Tirúa	97	58	39	0	3	2	97	58	39	0	3	2
Los Ángeles	823	604	219	69	0	11	823	604	219	70	0	11
Antuco	88	53	35	0	3	2	88	53	35	0	3	2
Cabrero	152	118	34	21	0	2	152	118	34	21	0	2
Laja	90	56	34	10	0	2	90	56	34	10	0	2
Mulchén	127	94	33	15	0	2	134	94	40	15	0	2
Nacimiento	135	100	35	14	0	2	135	100	35	14	0	2
Negrete	85	53	32	0	3	2	85	53	32	0	3	2
Quilaco	97	61	36	0	3	2	97	61	36	0	3	2
Quilleco	83	52	31	0	3	2	83	52	31	0	3	2
San Rosendo	90	56	34	0	3	2	90	56	34	0	3	2
Santa Bárbara	88	55	33	7	0	2	95	55	40	7	0	2
Tucapel	85	53	32	11	0	2	92	53	39	11	0	2
Yumbel	109	65	44	13	0	2	109	65	44	13	0	2
Alto Biobío	83	52	31	0	3	2	83	52	31	0	3	2

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Chillán	909	827	82	99	0	4	844	827	17	99	0	1
Bulnes	110	74	36	13	0	2	110	74	36	13	0	2
Cobquecura	102	59	43	0	3	2	102	59	43	0	3	2
Coelemu	125	82	43	16	0	2	125	82	43	16	0	2
Coihueco	136	58	78	10	0	4	136	58	78	10	0	4
Chillán Viejo	171	88	83	14	0	4	171	88	83	14	0	4
El Carmen	97	57	40	0	3	2	97	57	40	0	3	2
Ninhue	98	57	41	0	3	2	98	57	41	0	3	2
Ñiquén	112	60	52	0	3	3	112	60	52	0	3	3
Pemuco	103	62	41	0	3	2	103	62	41	0	3	2
Pinto	97	58	39	0	3	2	97	58	39	0	3	2
Portezuelo	104	62	42	0	3	2	104	62	42	0	3	2
Quillón	103	61	42	13	0	2	103	61	42	13	0	2
Quirihue	90	56	34	9	0	2	99	56	43	9	0	2
Ránquil	100	58	42	0	3	2	100	58	42	0	3	2
San Carlos	345	236	109	34	0	5	345	236	109	34	0	5
San Fabián	128	75	53	0	4	3	128	75	53	0	4	3
San Ignacio	110	60	50	0	3	3	110	60	50	0	3	3
San Nicolás	116	65	51	0	3	3	116	65	51	0	3	3
Treguaco	105	61	44	0	3	2	105	61	44	0	3	2
Yungay	92	58	34	8	0	2	92	58	34	8	0	2
La Araucanía	5.933	4.242	1.691	535	41	89	5.933	4.242	1.691	537	41	89
Temuco	1.735	1.620	115	217	0	6	1.735	1.620	115	218	0	6
Carahue	122	54	68	6	0	3	122	54	68	7	0	3
Cunco	110	56	54	11	0	3	110	56	54	11	0	3
Curarrehue	89	54	35	0	3	2	89	54	35	0	3	2
Freire	150	54	96	10	0	5	150	54	96	10	0	5
Galvarino	95	57	38	0	3	2	95	57	38	0	3	2
Gorbea	89	54	35	11	0	2	89	54	35	11	0	2
Lautaro	136	90	46	18	0	2	136	90	46	18	0	2

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Loncoche	186	125	61	22	0	3	186	125	61	22	0	3
Melipeuco	90	55	35	0	3	2	90	55	35	0	3	2
Nueva Imperial	144	75	69	13	0	3	144	75	69	12	0	3
Padre Las Casas	374	250	124	33	0	6	374	250	124	33	0	6
Perquenco	89	56	33	0	3	2	89	56	33	0	3	2
Pitrufquén	146	78	68	15	0	3	146	78	68	15	0	3
Pucón	143	72	71	9	0	4	143	72	71	9	0	4
Saavedra	85	53	32	0	3	2	85	53	32	0	3	2
Teodoro Schmidt	108	54	54	0	3	3	108	54	54	0	3	3
Toltén	83	52	31	0	3	2	83	52	31	0	3	2
Vilcún	183	93	90	0	5	5	183	93	90	0	5	5
Villarrica	261	184	77	23	0	4	261	184	77	23	0	4
Cholchol	88	55	33	0	3	2	88	55	33	0	3	2
Angol	318	279	39	43	0	2	318	279	39	44	0	2
Collipulli	103	65	38	13	0	2	103	65	38	13	0	2
Curacautín	100	68	32	15	0	2	100	68	32	15	0	2
Ercilla	89	54	35	0	3	2	89	54	35	0	3	2
Lonquimay	88	53	35	0	3	2	88	53	35	0	3	2
Los Sauces	90	55	35	0	3	2	90	55	35	0	3	2
Lumaco	91	55	36	0	3	2	91	55	36	0	3	2
Purén	90	55	35	12	0	2	90	55	35	12	0	2
Renaico	90	54	36	10	0	2	90	54	36	10	0	2
Traiguén	145	107	38	23	0	2	145	107	38	23	0	2
Victoria	223	156	67	31	0	3	223	156	67	31	0	3
Los Ríos	4.419	3.110	1.309	391	9	66	4.419	3.110	1.309	397	9	66
Valdivia	2.348	2.012	336	245	0	17	2.348	2.012	336	249	0	17
Corral	92	58	34	0	3	2	92	58	34	0	3	2
Lanco	131	73	58	13	0	3	131	73	58	13	0	3
Los Lagos	165	72	93	13	0	5	165	72	93	13	0	5
Máfil	99	58	41	0	3	2	99	58	41	0	3	2

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Mariquina	247	121	126	25	0	6	247	121	126	25	0	6
Paillaco	185	103	82	19	0	4	185	103	82	19	0	4
Panguipulli	261	92	169	14	0	8	261	92	169	14	0	8
La Unión	326	210	116	25	0	6	326	210	116	26	0	6
Futrono	131	73	58	7	0	3	131	73	58	7	0	3
Lago Ranco	98	57	41	0	3	2	98	57	41	0	3	2
Río Bueno	336	181	155	30	0	8	336	181	155	31	0	8
Los Lagos	4.938	3.554	1.384	410	30	71	4.938	3.554	1.384	411	30	71
Puerto Montt	1.094	902	192	113	0	10	1.094	902	192	113	0	10
Calbuco	168	56	112	9	0	6	168	56	112	9	0	6
Fresia	91	54	37	10	0	2	91	54	37	10	0	2
Frutillar	98	58	40	10	0	2	98	58	40	10	0	2
Los Muermos	99	53	46	11	0	2	99	53	46	11	0	2
Llanquihue	95	56	39	10	0	2	95	56	39	10	0	2
Mauñín	116	54	62	0	3	3	116	54	62	0	3	3
Puerto Varas	268	200	68	24	0	3	268	200	68	24	0	3
Castro	270	188	82	22	0	4	270	188	82	23	0	4
Ancud	242	163	79	22	0	4	242	163	79	22	0	4
Chonchi	93	55	38	0	3	2	93	55	38	0	3	2
Curaco de Vélez	31	0	31	0	0	2	31	0	31	0	0	2
Dalcahue	91	54	37	0	3	2	91	54	37	0	3	2
Puqueldón	35	0	35	0	0	2	35	0	35	0	0	2
Queilén	89	53	36	0	3	2	89	53	36	0	3	2
Quellón	194	133	61	14	0	3	194	133	61	14	0	3
Quemchi	85	53	32	0	3	2	85	53	32	0	3	2
Quinchao	100	59	41	0	3	2	100	59	41	0	3	2
Osorno	1.064	980	84	137	0	4	1.064	980	84	137	0	4
Puerto Octay	87	54	33	0	3	2	87	54	33	0	3	2
Purranque	147	107	40	18	0	2	147	107	40	18	0	2
Puyehue	92	54	38	0	3	2	92	54	38	0	3	2

Anexo N°4. Comparación de la distribución de los tamaños muestrales iniciales y definitivos de UPM y USM a nivel nacional, regiones y comunas, por área urbana y rural, Casen 2017.

Región / Comuna	Distribución inicial						Distribución definitiva					
	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural	Sobre-muestra Casen 2017	Sobre-muestra (Urbana) Casen 2017	Sobre-muestra (Rural) Casen 2017	Muestra Manzanas	Muestra Secciones RAU	Muestra Secciones Rural
Nivel País	95.159	79.992	15.167	9.179	274	776	95.159	79.992	15.167	9.207	274	771
Río Negro	103	58	45	10	0	2	103	58	45	10	0	2
San Juan de la Costa	93	55	38	0	3	2	93	55	38	0	3	2
San Pablo	93	55	38	0	3	2	93	55	38	0	3	2
Aysén	2.506	1.989	517	277	11	25	2.506	1.989	517	279	11	25
Coyhaique	1.508	1.226	282	177	0	14	1.508	1.226	282	177	0	14
Aysén	586	544	42	100	0	2	586	544	42	102	0	2
Cisnes	106	66	40	0	3	2	106	66	40	0	3	2
Cochrane	120	74	46	0	4	2	120	74	46	0	4	2
Chile Chico	124	79	45	0	4	2	124	79	45	0	4	2
Río Ibáñez	62	0	62	0	0	3	62	0	62	0	0	3
Magallanes	3.093	2.925	168	443	3	9	3.093	2.925	168	445	3	8
Punta Arenas	2.722	2.622	100	396	0	5	2.722	2.622	100	398	0	4
Porvenir	93	60	33	0	3	2	93	60	33	0	3	2
Natales	278	243	35	47	0	2	278	243	35	47	0	2

Elaborado por el Instituto Nacional de Estadísticas.

Anexo N°5. Ajuste de las estimaciones de población para la encuesta de Caracterización Socioeconómica Nacional Casen 2017

El Sub Departamento de Demografía y Vitales se encarga de diseñar y elaborar las estimaciones poblacionales de todas las encuestas de hogares desarrolladas por el INE, y así lo ha hecho para versiones anteriores de Casen, así como para Casen 2017. En esta ocasión fueron elaboradas para la población llevada a noviembre 2017 en 15 regiones del país, sin considerar, las nuevas regiones de Biobío y Ñuble.

A continuación, se detalla el procedimiento que se siguió, por parte del Departamento de Investigación y Desarrollo del INE, para la obtención de las estimaciones poblacionales a noviembre 2017 en las nuevas regiones de Biobío y Ñuble, a partir de la estimación de población de la antigua Biobío. Adicionalmente, se presenta el ajuste realizado para determinar la estimación de población en viviendas particulares ocupadas.

Para facilitar la lectura de las tablas, se omiten los decimales de los valores, por lo que los números absolutos no equivalen de manera exacta al porcentaje. Así mismo, se advierte que los cálculos, para cada paso, fueron realizados considerando todos los decimales y solo se redondeó, en el resultado definitivo.

1. Obtención de las estimaciones poblacionales de Biobío y Ñuble

1.1. Insumos para el cálculo

1. Estimaciones poblacionales de la antigua Biobío a noviembre 2015; nacional, urbano y rural.

Antigua Biobío Nov2015			Nueva Biobío Nov2015 (8101-8314)			Ñuble Nov2015 (8401-8421)		
Regional	Urbano	Rural	Regional	Urbano	Rural	Regional	Urbano	Rural
2.103.762	1.770.456	333.306	1.635.372	1.440.648	194.724	468.390	329.808	138.582
100%	84%	16%	100%	88%	12%	100%	70%	30%

2. Estimaciones poblacionales de la antigua Biobío a noviembre 2017; nacional, urbano y rural.

Antigua Biobío (Nov2017)		
Regional	Urbano	Rural
2.125.607	1.794.298	331.309
100%	84%	16%

3. Suma de las estimaciones poblacionales a noviembre 2017 de las 54 comunas que componen la antigua Biobío; a nivel nacional, para la nueva Biobío y Ñuble.

Suma de Estimaciones en Comunas Nov2017		
Regional (Antigua)	Nueva Biobío (8101-8314)	Ñuble (8401-8421)
2.125.563 100%	1.655.634 78%	469.929 22%

Cabe señalar, que los stock de las estimaciones a nivel regional y comunal de noviembre 2017 no coinciden porque éstas fueron realizadas sobre modelos distintos.

Paso 1

Se calcula la diferencia entre la estimación regional de la antigua Biobío y la suma de las estimaciones de las 54 comunas que comprenden la región. La diferencia observada es de 44 personas.

Antigua Biobío Nov2017			Suma de Estimaciones en Comunas Nov2017			Diferencia regional
Regional	Urbano	Rural	Regional (Antigua)	Nueva Biobío (8101-8314)	Ñuble (8401-8421)	
2.125.607 100%	1.794.298 84%	331.309 16%	2.125.563 100%	1.655.634 78%	469.929 22%	44 -

Paso 2

Se distribuyen las 44 personas en entre la nueva Biobío y Ñuble, proporcionalmente al peso de cada región, en función del número de personas.

Antigua Biobío Nov2017			Nuevas Regiones Nov2017			Diferencia regional
Regional	Urbano	Rural	Regional (Antigua)	Nueva Biobío (8101-8314)	Ñuble (8401-8421)	
2.125.607 100%	1.794.298 84%	331.309 16%	2.125.607 100%	1.655.668 78%	469.939 22%	0 -

Paso 3

Para cada nueva región, se redistribuye el total personas en cada área (Urbano-Rural), proporcionalmente al peso de éstas, en función del número de personas a noviembre 2015.

Nueva Biobío Nov2017 (8101-8314)			Ñuble Nov2017 (8401-8421)		
Regional	Urbano	Rural	Regional	Urbano	Rural
1.655.668	1.458.528	197.141	469.939	330.899	139.040
100%	88%	12%	100%	70%	30%

Paso 4

A raíz del paso 3, la suma de las nuevas regiones coinciden con la estimación regional de la antigua Biobío, sin embargo, para las áreas Urbano – Rural existen diferencias.

Antigua Biobío Nov2017			Sumas de Nueva Biobío y Ñuble Nov2017			Diferencias Urbano - Rural		
Regional	Urbano	Rural	Regional	Urbano	Rural	Regional	Urbano	Rural
2.125.607	1.794.298	331.309	2.125.607	1.789.426	336.181	0	4.872	-4.872
100%	84%	16%	100%	84%	16%	-	-	-

Dadas las diferencias por área Urbano-Rural, se ajustan estos niveles de tal manera que las sumas coincidan con las áreas Urbano-Rural de la antigua Biobío, no obstante, persisten diferencias en la suma de las nuevas regiones.

Nueva Biobío Nov2017 (8101-8314)			Ñuble Nov2017 (8401-8421)		
Regional	Urbano	Rural	Regional	Urbano	Rural
1.656.782	1.462.499	194.284	468.825	331.799	137.025
100%	88%	12%	100%	71%	29%

Paso 5

Las diferencias entre la suma de las nuevas regiones y el total de la antigua Biobío, es distribuida en las áreas Urbano-Rural, proporcionalmente al peso que tienen estas áreas en la estimación de la antigua Biobío a noviembre 2017.

El resultado de este último paso es la estimación de población para las nuevas regiones de Biobío y Ñuble por área Urbano-Rural, coincidiendo los totales con la estimación de la antigua Biobío por área a noviembre 2017.

Diferencias Nueva Biobío		Diferencias Ñuble		Nueva Biobío Nov2017 (8101-8314)			Ñuble Nov2017 (8401-8421)		
Regional	U + R	Regional	U + R	Regional	Urbano	Rural	Regional	Urbano	Rural
1.655.668	1.656.782	469.939	468.825	1.655.668	1.461.558	194.110	469.939	332.740	137.199
	-1.114		1.114	100%	88%	12%	100%	71%	29%

2. Ajuste de estimaciones poblacionales en viviendas particulares ocupadas

Las estimaciones poblacionales recibidas del Sub Departamento de Demografía y Vitales se refieren a la población total del país, sin embargo, la población objetivo de la encuesta Casen es la población en viviendas particulares ocupadas, razón por la que se aplica un factor de ajuste a las estimaciones poblacionales.

Para la encuesta Casen 2017 se calculan tres factores de expansión, Regional, Provincial y Comunal, lo que implica calibrar los factores por la estimación de población en viviendas particulares ocupadas para esos niveles.

En ese sentido se generaron estimaciones poblacionales para los niveles Región-Área y Comunal, para los que se obtienen factores de ajuste para la población en viviendas particulares ocupadas a nivel Regional-Área y Comunal. El factor de ajuste se obtiene a partir del Censo de Población y Vivienda del año 2002.

La siguiente tabla muestra la comparación entre las estimaciones regionales y la suma de las comunales por región, tanto para la población total como para la población en viviendas particulares ocupadas.

Tabla A2.1. Comparación entre las estimaciones regionales y la suma de las comunales por región

Región	Estimaciones poblacionales Regionales		Suma de estimaciones de población Comunales		Diferencias Regional - Comunal	
	Total	Viv. Particulares	Total	Viv. Particulares	Total	Viv. Particulares
País	18.194.383	17.807.355	18.193.051	17.816.788	1.332	-9.433
Arica y Parinacota	170.648	161.152	170.663	163.311	-15	-2.159
Tarapacá	367.354	347.199	367.297	340.713	57	6.486
Antofagasta	621.079	586.340	621.055	584.259	24	2.081
Atacama	295.150	284.968	295.141	286.033	9	-1.065
Coquimbo	793.024	775.506	793.030	776.158	-6	-652
Valparaíso	1.891.258	1.856.490	1.891.048	1.858.225	210	-1.735
Metropolitana	7.322.302	7.223.160	7.321.550	7.236.467	752	-13.307
O'Higgins	944.100	929.595	944.060	929.358	40	237
Maule	1.064.699	1.047.017	1.064.683	1.046.851	16	166
Ñuble	469.939	460.634	469.929	460.971	10	-337
Biobío	1.655.668	1.625.234	1.655.634	1.624.939	34	295
La Araucanía	1.027.876	999.344	1.027.788	999.621	88	-277
Los Ríos	385.462	371.114	385.426	371.136	36	-22
Los Lagos	911.724	882.962	911.694	883.262	30	-300
Aysén	112.025	104.939	112.005	104.423	20	516
Magallanes	162.075	151.701	162.048	151.061	27	640

Como se ha mencionado anteriormente, las estimaciones regionales no coinciden con la suma de las estimaciones comunales, diferencia que se incrementa al realizar el ajuste para la población en viviendas particulares ocupadas.

Las diferencias observadas a nivel regional, son distribuidas proporcionalmente entre las comunas que forman la región, de tal manera que se concilia la estimación de población en viviendas particulares de cada región con la suma de las comunas.

Finalmente, la estimación de las provincias es obtenida por agregación de las comunas.

Anexo N°6. Comandos y Sintaxis en SPSS y STATA.

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

Plan de muestreo estimación Tasa de Pobreza y Tasa de Pobreza Extrema por Ingresos, Casen 2017.

```
CSPLAN ANALYSIS
/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'
/PLANVARS ANALYSISWEIGHT=expr
/SRSESTIMATOR TYPE=WOR
/PRINT PLAN
/DESIGN STRATA=varstrat CLUSTER=varunit
/ESTIMATOR TYPE=WR.
```

Transformación de variables para identificar Personas y Hogares en situación de pobreza por ingresos.

VALUE LABELS pobreza

- 1 "Pobres extremos"
- 2 "Pobres no extremos"
- 3 "No pobres".

RECODE pobreza (3=0) (1 thru 2=1) INTO pobre.

VARIABLE LABELS pobre 'Categoría de pobreza por ingresos'.

EXECUTE.

FORMATS pobre (F8.0).

VALUE LABELS pobre

- 0 "No pobres"
- 1 "Pobres".

Cálculo para personas.

1. Porcentaje de personas en situación de pobreza extrema por ingresos, por región, año 2017.

CSTABULATE

```
/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'  
/TABLES VARIABLES=region_16 BY pobreza  
/CELLS ROWPCT  
/STATISTICS SE CV CIN(95) COUNT  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

2. Porcentaje de personas en situación de pobreza por ingresos, por región, año 2017.

CSTABULATE

```
/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'  
/TABLES VARIABLES=region_16 BY pobre  
/CELLS ROWPCT  
/STATISTICS SE CV CIN(95) COUNT  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

3. Porcentaje de personas en situación de pobreza extrema por ingresos, por zona, año 2017.

CSTABULATE

```
/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'  
/TABLES VARIABLES=zona BY pobreza  
/CELLS ROWPCT  
/STATISTICS SE CV CIN(95) COUNT  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

4. Porcentaje de personas en situación de pobreza por ingresos, por zona, año 2017.

CSTABULATE

```
/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'  
/TABLES VARIABLES=zona BY pobre  
/CELLS ROWPCT  
/STATISTICS SE CV CIN(95) COUNT  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

Cálculo para hogares.

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

USE ALL.

COMPUTE filter_\$=(nucleo ~= 0 & pco1 = 1).

VARIABLE LABELS filter_\$ 'nucleo ~= 0 (FILTER)'.
VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_\$ (f1.0).
FILTER BY filter_\$.
EXECUTE.

1. Porcentaje de hogares en situación de pobreza extrema por ingresos, por región, año 2017.

CSTABULATE

/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'

/TABLES VARIABLES=region_16 BY pobreza

/CELLS ROWPCT

/STATISTICS SE CV CIN(95) COUNT

/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.

2. Porcentaje de hogares en situación de pobreza por ingresos, por región, año 2017.

CSTABULATE

/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'

/TABLES VARIABLES=region_16 BY pobre

/CELLS ROWPCT

/STATISTICS SE CV CIN(95) COUNT

/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.

3. Porcentaje de hogares en situación de pobreza extrema por ingresos, por zona, año 2017.

CSTABULATE

/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'

/TABLES VARIABLES=zona BY pobreza

/CELLS ROWPCT

/STATISTICS SE CV CIN(95) COUNT

/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.

4. Porcentaje de hogares en situación de pobreza por ingresos, por zona, año 2017.

CSTABULATE

/PLAN FILE='Z:\Base CEPAL 2\CASEN_2017\PLAN SPSS.csaplan'

/TABLES VARIABLES=zona BY pobre

/CELLS ROWPCT

/STATISTICS SE CV CIN(95) COUNT

/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.

Comandos utilizados en STATA para estimar la tasa de pobreza por ingresos y errores muestrales.

Plan de muestreo estimación Tasa de Pobreza y Tasa de Pobreza Extrema por Ingresos, Casen 2017.

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

Transformación de variables para identificar Personas y Hogares en situación de pobreza por ingresos.

```
gen pobre = 1 if(pobreza==1 | pobreza==2)
replace pobre=0 if (pobreza==3)
label var pobre "pobreza ingreso"
label define pobre 1 "pobres" 0 "no pobres"
label values pobre pobre
```

```
gen pobreza1 = 1 if(pobreza==1)
replace pobreza1 = 2 if(pobreza==2)
replace pobreza1 = 3 if(pobreza==3)
label var pobreza1 "pobreza ingreso"
label define pobreza1 1 "pobres extremos" 2 "pobres no extremos" 3 "no pobres"
label values pobreza1 pobreza1
```

Cálculo para personas.

5. Porcentaje de personas en situación de pobreza extrema por ingresos, por región, año 2017.

```
svy : prop pobreza1 if nucleo !=0, level(95) over(region_16)
estat cv
```

6. Porcentaje de personas en situación de pobreza por ingresos, por región, año 2017.

```
svy : prop pobre if nucleo !=0, level(95) over(region_16)
estat cv
```

7. Porcentaje de personas en situación de pobreza extrema por ingresos, por zona, año 2017.

```
svy : prop pobreza1 if nucleo !=0, level(95) over(zona)
estat cv
```

8. Porcentaje de personas en situación de pobreza por ingresos, por zona, año 2017.

svy : prop pobre if nucleo !=0, level(95) over(zona)

estat cv

Cálculo para hogares.

1. Porcentaje de hogares en situación de pobreza extrema por ingresos, por región, año 2017.

svy : prop pobreza1 if nucleo !=0 & pco1==1, level(95) over(region_16)

estat cv

2. Porcentaje de hogares en situación de pobreza por ingresos, por región, año 2017.

svy : prop pobre if nucleo !=0 & pco1==1, level(95) over(region_16)

estat cv

3. Porcentaje de hogares en situación de pobreza extrema por ingresos, por zona, año 2017.

svy : prop pobreza1 if nucleo !=0 & pco1==1, level(95) over(zona)

estat cv

4. Porcentaje de hogares en situación de pobreza por ingresos, por zona, año 2017.

svy : prop pobre if nucleo !=0 & pco1==1, level(95) over(zona)

estat cv

BIBLIOGRAFÍA

1. INE Chile. (2006). Encuesta Nacional de Empleo. Diseño muestral.
2. Hansen, M; Hurwitz, W; Madow, W.(1953). Sample Survey Method and Theory. Volume II-Theory. John Wiley & Sons, Inc. New York.
3. Cochran, W. (1977). Sampling Techniques. Third Edition. John Wiley & Sons, Inc. New York.
4. Kish, L. (1965). Survey Sampling. John Wiley & Sons, Inc. New York.
5. Groves, R. (1989, 2004). Survey Errors and Survey Cost. John Wiley & Sons, Inc. Hoboken. New Jersey.
6. Levy, P; Lemeshow, S. (1999). Sampling of Populations: Methods and Applications. Third Edition. Wiley & Sons, Inc. Canada.
7. Lehtonen, R; Pahkinen, E. (2004). Practical Methods for Design and Analysis of Complex Surveys. Second Edition. John Wiley & Sons, Ltd. England.
8. Rao, P. (2000). Sampling Methodologies with Applications. Chapman & Hall/CRC.US.
9. Naciones Unidas. (2009). Encuesta de hogares en los países de desarrollo y transición. Departamento de Asuntos Económicos. División Estadística. New York.
10. 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.
11. Ehling, M., and T. Körner (2007), Handbook on Data Quality Assessment Methods and Tools, Wiesbaden: European Commission, Eurostat.
12. Sitter RR (1992) Comparing Three Bootstrap Methods for Survey Data. The Canadian Journal of Statistics / La Revue Canadienne de Statistique 20:135-154

- 13.** Valliant R, Brick JM, Dever JA (2008) Weight Adjustments for the Grouped Jackknife Variance Estimator. *Journal of Social Statistics* 24(3):469-488.
- 14.** Valliant, R. Dever, J.A., Kreuter, F. (2013, chapter 14).
- 15.** Valliant, R. Dever, J.A., Kreuter, F. (2013, section 6.5.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- 16.** Valliant R, Rust KF (2010) Degrees of Freedom Approximations and Rules-of-Thumb. *Journal of Social Statistics* 26:585-602.
- 17.** 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.co

