



Metodología del Diseño Muestral y Factores de Expansión Encuesta de Caracterización Socioeconómica Nacional

OBSERVATORIO SOCIAL

**Documentos
Metodológicos, N°1**
13 de Septiembre de 2012



Contenido

I.	Presentación	5
II.	Síntesis	6
	1. OBJETIVOS	6
	2. MÉTODOS	6
	3. RESULTADOS	6
III.	Resumen Ejecutivo	7
IV.	Antecedentes del Diseño 2011	10
	1. CAMBIOS EN DISEÑO MUESTRAL 1987-2009	10
	2. CAMBIOS EN DISEÑO MUESTRAL 2011.....	12
	2.1. <i>Evaluación del diseño muestral vigente</i>	12
	2.2. <i>Principales características del diseño muestral 2011</i>	13
V.	Diseño Muestral.....	15
	1. CARACTERÍSTICAS DEL MARCO MUESTRAL	15
	1.1. <i>Cobertura del Marco Muestral</i>	18
	1.2. <i>Estratificación del Marco Muestral</i>	20
	2. ESTIMACIÓN Y DISTRIBUCIÓN DEL TAMAÑO MUESTRAL	21
	2.1. <i>Estimación del Tamaño Muestral por Región</i>	21
	2.2. <i>Distribución de la Muestra entre Comunas al interior de la Región</i>	26
	3. SELECCIÓN DE CONGLOMERADOS.....	29
	3.1. <i>Selección de Conglomerados desde el Marco Urbano</i>	29
	3.2. <i>Selección de Conglomerados desde el Marco Rural</i>	31
	4. SELECCIÓN DE VIVIENDAS	32
	4.1. <i>Enumeración de los Conglomerados</i>	32
	4.2. <i>Selección de las Viviendas</i>	34
	5. REDISTRIBUCIÓN DE CONGLOMERADOS ENTRE LAS SUBMUESTRAS 1 Y 2	37
VI.	Desarrollo de factores de Expansión.....	39
	1. VISIÓN GENERAL.....	39
	2. PONDERADOR DE SELECCIÓN.....	42
	2.1. <i>Probabilidades de selección de los Conglomerados</i>	42
	2.2. <i>Inverso de las Probabilidades de selección</i>	50
	2.3. <i>Ajuste por omisión de Conglomerados</i>	54
	3. PONDERACIÓN DE SELECCIÓN DE VIVIENDAS.....	59
	3.1. <i>Probabilidad Condicional de selección de una vivienda</i>	59
	3.2. <i>Ponderador de selección de viviendas</i>	59
	4. PONDERACIÓN POR ELEGIBILIDAD	63
	4.1. <i>Ajuste por Elegibilidad Desconocida</i>	63
	4.2. <i>Ajuste por No Elegibilidad</i>	67
	5. PONDERACIÓN POR NO RESPUESTA	67
	6. PONDERACIÓN DE CALIBRACIÓN	73

6.1.	<i>Ponderador de Calibración Regional</i>	74
6.2.	<i>Ponderador de Calibración Comunal</i>	78
VII.	Estimación de Varianza	82
1.	EL ESTIMADOR	82
2.	VARIANZA DEL ESTIMADOR	84
3.	VARIABLES QUE IDENTIFICAN EL DISEÑO MUESTRAL COMPLEJO	86
3.1.	<i>Creación de pseudo-estratos (varstrat)</i>	87
3.2.	<i>Creación de pseudo-conglomerados (varunit)</i>	88
4.	ESTIMACIÓN DE VARIANZAS EN MUESTRA COMPLEJA CASEN 2011	89
5.	RESULTADOS NIVEL NACIONAL	91
6.	RESULTADOS NIVEL REGIONAL.....	93
7.	MÁRGENES DE ERROR PROYECTADOS, EFECTIVOS E HISTÓRICOS.....	95
VIII.	Anexos.....	99

I. Presentación

Este informe describe el diseño muestral de la Encuesta de Caracterización Socioeconómica Nacional (Casen) 2011 e informa las implicancias para la investigación y análisis de innovaciones en distintos aspectos del mismo. En otros reportes se presentarán resultados acerca de las tasas de respuesta, errores de medición, resultados generales del trabajo de campo y guías para el análisis a partir de los datos de la encuesta.

La encuesta Casen 2011 fue diseñada por el Ministerio de Desarrollo Social y los datos fueron recolectados y procesados, en parte, por su contratista, el Centro de Microdatos del Departamento de Economía de la Universidad de Chile. El diseño y selección de la muestra, así como el cálculo de los factores de expansión y de márgenes de error fueron realizados por el Instituto Nacional de Estadísticas (INE) en conjunto con el Ministerio. La corrección y ajuste de ingresos a cuentas nacionales, estuvieron a cargo de la Comisión Económica para América Latina (CEPAL). Las estimaciones de pobreza comunal, utilizando la metodología de Estimación para Áreas Pequeñas, serán realizadas bajo la asistencia técnica del Programa de Naciones Unidas para el Desarrollo (PNUD).

Este informe fue preparado por el Ministerio de Desarrollo Social en conjunto con el Instituto Nacional de Estadísticas.

II. Síntesis

1. Objetivos

La Encuesta de Caracterización Socioeconómica Nacional (Casen) recolecta datos acerca de características sociales y económicas de una muestra representativa de la población que reside en hogares particulares en Chile. Este reporte describe el diseño muestral de la nueva estrategia de recolección de datos implementado a partir de Casen 2011 y los efectos de ese diseño en el desarrollo de factores de expansión y procedimientos para la estimación de la varianza. El conocimiento de estos aspectos del diseño es fundamental para los investigadores que planean analizar los datos de Casen 2011.

Se planea liberar dos bases de datos. La primera, denominada "Casen 2011-S2", correspondiente a la submuestra 2, cubre las 8 semanas de recolección de datos comprendidas entre el 26 de noviembre de 2011 y el 22 de enero de 2012. Esta base abarca un periodo de tiempo que en promedio tiende a coincidir con los periodos de recolección de datos de versiones anteriores de la encuesta. Esta es la base que por lo tanto se utiliza para generar las estadísticas oficiales de pobreza y distribución de ingresos del 2011¹. La segunda base, denominada simplemente "Casen 2011", cubrirá las 13 semanas de recolección de datos entre el 17 de octubre y el 22 de enero de 2012. Esta última base incluye datos adicionales recolectados durante las 5 semanas comprendidas entre el 17 de octubre y el 22 de noviembre de 2011, periodo en el cual generalmente no se recolectaron datos en versiones anteriores de la Encuesta. Junto con la liberación de la base Casen 2011 se entregará un set de factores de expansión que permita el análisis de los datos combinados de la Encuesta completa.

2. Métodos

El diseño de la Encuesta Casen 2011 se basa en dos muestras probabilísticas, independientes y que comparten el mismo diseño muestral (submuestra 1 y submuestra 2). El trabajo de campo fue implementado por el Centro de Microdatos del Departamento de Economía de la Universidad de Chile bajo contrato con el Ministerio de Desarrollo Social. Las entrevistas se aplicaron en modalidad cara-a-cara, por encuestadores profesionales, utilizando cuestionario en papel.

3. Resultados

La tasa de no respuesta a las preguntas de la encuesta fue en general baja², sin embargo y al igual que en versiones anteriores, se realizó un proceso de imputación para los datos faltantes en variables consideradas clave para las estimaciones de pobreza y distribución de ingresos³. La imputación simple de los datos se realiza a través de los métodos de medias condicionadas y hot-deck. En forma posterior a la imputación, y también como en versiones anteriores, se

¹ Se dice "en promedio" dado que las encuestas Casen anteriores presentan algunas diferencias en el período de recolección de los datos.

² En la versión 2011, la no respuesta a las principales preguntas del ingreso fluctuó entre 2 y 10%.

³ Se entiende como "datos faltantes" los casos donde, habiéndose declarado la existencia de una actividad generadora de ingresos, no se reporta el monto de los mismos. Hay tres situaciones principales a considerar a este respecto: ocupados que trabajan con remuneración y que no reportan los ingresos del trabajo; pensionados que no reportan sus pagos; personas que viven en una vivienda propia y que no declaran una estimación del valor del arriendo imputado.

realiza el proceso de ajuste de las partidas más importantes de ingresos⁴ con el objetivo de calibrar los resultados de la Encuesta a sus referentes en las Cuentas Nacionales. Tanto el procedimiento de imputación como el de ajuste de los ingresos son realizados, a requerimiento del Ministerio, por CEPAL. Las variables de ingresos incluidas en las bases de datos Casen puestas a disposición del público, corresponden a las partidas de ingresos ya imputadas y ajustadas.

III. Resumen Ejecutivo

La Encuesta de Caracterización Socioeconómica Nacional (Casen) recolecta datos acerca de características sociales y económicas de una muestra representativa de la población que reside en hogares particulares en Chile. Este informe describe los procedimientos utilizados para seleccionar la muestra, desarrollar los factores de expansión y estimar los errores muestrales⁵. Se recomienda considerar esta información a quienes realicen investigaciones a partir de los datos de la Encuesta Casen, o bien, que deseen comparar los procedimientos de otras encuestas con los empleados en Casen.

La Encuesta Casen es diseñada y administrada por el Ministerio de Desarrollo Social (ex Mideplan) - institución responsable de evaluar la situación socioeconómica, así como el impacto de los programas sociales en las condiciones de vida de la población, para obtener periódicamente datos útiles para el diseño y evaluación de la política social existente en el país. En este contexto, el objetivo de la Encuesta es producir estimaciones confiables sobre:

- La situación socioeconómica de los hogares y de la población. En particular, estimar la magnitud de la pobreza y la distribución del ingreso.
- Aspectos demográficos, de educación, salud, vivienda, trabajo e ingresos de aquellos grupos definidos como prioritarios por la política social (infancia, juventud, adultos mayores, mujeres, pueblos indígenas, personas con discapacidad).
- La cobertura, la focalización y la distribución del gasto fiscal de los principales programas sociales de alcance nacional entre los hogares y la población.

La Encuesta Casen 2011 fue levantada por el Centro de Microdatos del Departamento de Economía de la Universidad de Chile, bajo contrato con el Ministerio de Desarrollo Social. Las entrevistas fueron aplicadas en modalidad cara-a-cara, por encuestadores profesionales, en cuestionarios de papel. El cuestionario fue diseñado para demorar aproximadamente 60 minutos en hogares de 4 a 5 personas. La revisión del cuestionario 2009, en preparación para el desarrollo de la versión 2011, fue desarrollada por el Ministerio de Desarrollo Social a través de un proceso de consulta que contó con la participación de 35 instituciones de los sectores público y de organismos internacionales⁶.

⁴ El procedimiento se realiza para las partidas de sueldos y salarios, ingresos del trabajo independiente, pensiones, renta imputada por la vivienda propia e ingresos del capital.

⁵ El foco de este documento es la descripción del diseño, selección de la muestra y desarrollo de los factores de expansión. La descripción del proceso de corrección y ajuste de ingresos se encuentra en el documento "La medición de los ingresos en la encuesta Casen (Cepal, 2012)".

⁶ Talleres Intersectoriales Casen 2011. Santiago, 2 al 9 de Mayo de 2011. Reuniones bilaterales con equipos técnicos de Ministerios y organismos internacionales entre junio y agosto de 2011.

La muestra 2011 es la más grande que se ha levantado en la historia de la Encuesta. El aumento es mayor en las regiones extremas del país, donde históricamente se ha concentrado menor proporción de la muestra. Esto se explica porque el tamaño muestral se determinó, por primera vez, con el objetivo de mejorar el nivel de precisión⁷ de las estimaciones de la tasa de pobreza a nivel regional.

Otras dos innovaciones incluyen la extensión del periodo de recolección de datos (desde mediados de octubre hasta mediados de enero) y la utilización del sistema de submuestras para la recolección de los datos.

La extensión del periodo de recolección de datos de la Encuesta Casen es parte de la estrategia de rediseño de la Encuesta que busca mejorar la calidad de la recolección de datos. Este diseño busca maximizar el uso de recursos y obtener respuestas de mejor calidad a partir de (1) la producción de un flujo de trabajo continuo, (2) utilizando un equipo más pequeño de encuestadores mejor entrenados⁸, y (3) el monitoreo continuo del proceso de recolección de datos que permita, de ser necesario, optimizar las estrategias durante el trabajo de campo. Esta estrategia de recolección de datos es utilizada para encuestas de hogares en EEUU, Brasil, Uruguay, Colombia y El Salvador⁹. En Chile, la Encuesta Nacional de Empleo (ENE y NENE) también comparte este tipo de diseño.

Dado que implementar una muestra única hubiera generado problemas de comparabilidad con la serie completa Casen (porque los períodos muestrales son distintos) es que se optó por dividir el período completo en 2 submuestras, cada una con representatividad regional y nacional, siendo la segunda submuestra levantada en el mismo período que tradicionalmente se ha aplicado la encuesta Casen¹⁰.

La introducción de un sistema de submuestras persigue mejorar el control de la distribución de la muestra a lo largo del periodo de levantamiento. La estrategia permite generar estimaciones nacionales a partir de cada submuestra o bien combinar ambas submuestras para mejorar la precisión de las estimaciones regionales. La muestra 2011 se recolectó sobre la base de dos muestras independientes que comparten el mismo diseño muestral. La primera, denominada "Submuestra 1" (Casen 2011-S1), se levantó por 5 semanas entre el 17 de octubre y el 22 de noviembre de 2011. La segunda, denominada "Submuestra 2" (Casen 2011-S2), se levantó durante 8 semanas desde el 22 de noviembre de 2011 hasta el 22 de enero de 2012. Esta última submuestra cubre un periodo de tiempo que en promedio tiende a coincidir con los periodos de recolección de datos de versiones anteriores de la Encuesta. La "Submuestra 1",

⁷ Precisión (estadística): "Cercanía de los estimadores a los valores verdaderos. Resultado de errores aleatorios y errores sistemáticos". Ehling, M., and T. Körner (2007).

⁸ En las últimas versiones se ha evidenciado la creciente complejidad de coordinar la implementación de una encuesta de la magnitud de Casen. En la versión 2009, por ejemplo, se contrató a cerca de 1.500 encuestadores para recolectar 70 mil encuestas. En contraste, la versión 2011 mantuvo un máximo de 1.000 encuestadores para recolectar 87 mil encuestas.

⁹ En EEUU, encuestas de hogares que utilizan esta estrategia de recolección de datos incluyen: American Community Survey (ACS), National Health Immunization Survey (NHIS) y National Survey of Family Growth (NSFG), entre otras. En América Latina, esta estrategia la siguen las encuestas de hogares de Brasil (Encuesta Nacional por Muestra de Viviendas, PNAD Continua), Uruguay (Encuesta Nacional de Hogares Ampliada, ENHA, Encuesta Continua de Hogares), Colombia (Gran Encuesta Integrada de Hogares, GEIH) y El Salvador (Encuesta de Hogares de Propósitos Múltiples).

¹⁰ Históricamente, la encuesta Casen se recolecta en un periodo de aproximadamente 8 semanas entre los meses de noviembre y enero, con el 50-60% de la muestra entrevistada durante el mes de diciembre.

por otra parte, cubre un periodo de tiempo en que generalmente no se han recolectado datos en versiones anteriores de la Encuesta.

Para efectos de comparación de la serie de tiempo de las estadísticas de trabajo, pobreza y distribución de ingresos, se recomienda utilizar exclusivamente la "Submuestra 2", puesto que podría haber factores estacionales necesarios de considerar al hacer comparaciones con versiones anteriores de la encuesta.

Casen es una encuesta de hogares que selecciona viviendas en la muestra con distintas probabilidades de selección en distintas áreas geográficas del país. Por lo tanto, para producir estimaciones insesgadas, se deben utilizar los factores de expansión apropiados de manera de compensar por las distintas probabilidades de selección y probabilidades de responder la encuesta. Para generar estimaciones de los errores de muestreo se deben utilizar paquetes estadísticos que permitan incorporar el efecto de los factores de expansión, estratos, conglomerados y fracciones de muestreo del diseño muestral 2011. Paquetes que incorporan estas características son Stata, SPSS y SAS entre otros.

El resto de este informe describe la forma en que la muestra 2011 fue diseñada y seleccionada, cómo los factores de expansión fueron desarrollados y ajustados para compensar por la falta de respuesta y controles de población total, y cómo se pueden obtener estimaciones del margen de error de las variables de interés incorporando información sobre el diseño complejo de la encuesta. El reporte se puede leer en tres niveles de detalle:

- Primero, el resumen ejecutivo es para quienes requieren de una visión más general y no técnica de los procedimientos de la encuesta.
- Segundo, se proveen resúmenes para cada gran sección del reporte, los que entregan información más detallada de cada aspecto del diseño.
- Tercero, en el resto del informe se ofrecen más detalles técnicos. Quienes lean el informe completo encontrarán algunos temas descritos brevemente en los resúmenes y luego más profusamente en el resto de cada sección.

IV. Antecedentes del Diseño 2011

A lo largo de los años, la Encuesta Casen ha evolucionado en varios aspectos relacionados con la forma en que es planeada, diseñada y ejecutada. Aquí se revisan los cambios más importantes en relación al diseño de la muestra Casen.

1. Cambios en diseño muestral 1987-2009

A lo largo de los años, el Ministerio ha solicitado a diversos expertos recomendaciones para mejorar el diseño muestral de la Encuesta. El año 1996 convoca al grupo de trabajo integrado por INE, el Centro Interamericano de Enseñanza Estadística (CIENES) y la Universidad de Chile para la definición del diseño de la muestra. El diseño y los lineamientos desarrollados por este grupo se aplican en las Encuestas Casen entre 1996 y 2003. En 2008 el Ministerio solicita a la Comisión Económica para América Latina (CEPAL) una evaluación estadística de la Encuesta. En 2009 el Ministerio conforma una comisión para evaluar la propuesta de cambio de marco muestral recomendada por el INE. A continuación se listan los cambios más relevantes implementados entre 1987 y 2009 en relación al diseño de la muestra.

Marco muestral

- Entre 1987 y 1994, se utilizó el marco desarrollado a partir de la información cartográfica del Censo de Población y Vivienda del año 1982, la Fuerza Aérea de Chile, el instituto de Recursos Naturales y la Corporación de Fomento para la Producción (CORFO).
- Entre 1996 y 2000, se hizo uso del marco del Censo de Población y Vivienda del año 1992. Mientras que para el 2003 se utilizó el marco desarrollado por el INE a partir de la información cartográfica del Pre-Censo 2002.
- En 2006 y 2009, se utilizó el marco "Programa Integrado de Encuestas de Hogares" (PIDEH) desarrollado por el INE a partir de la información cartográfica del Censo de Población y Vivienda del año 2002.

Instituciones responsables del diseño muestral

- Entre 1987 y 1994 la institución responsable del diseño de muestral es el Departamento de Economía de la Universidad de Chile.
- Entre 1996 y 2003 el diseño es definido bajo criterios establecidos por un grupo de trabajo compuesto por Mideplan, INE, Centro Interamericano de Enseñanza Estadística (CIENES) y Universidad de Chile.
- A partir de 2006 el diseño de la muestra es definido por Mideplan e INE.

Diseño muestral

- Entre 1987 y 1994 el diseño muestral se caracteriza por ser estratificado y trietápico en todas las comunas del país¹¹.
- Entre 1996 y 2003 el diseño es estratificado bietápico en las comunas autorepresentadas, y estratificado trietápico en las no autorepresentadas¹².
- A partir de 2006 el diseño es estratificado y bietápico en todas las comunas del país¹³.

¹¹ Número de comunas autorepresentadas: 1987 (n=48), 1990 (n=72), 1992 (n=138) y 1994 (n=209).

¹² Número de comunas autorepresentadas: 1996 (n=124), 1998 (n=196), 2000 (n=286) y 2003 (n=302).

Selección y actualización de la Muestra

- Entre 1987 y 2003 la institución responsable de la selección y actualización de la muestra es el Departamento de Economía de la Universidad de Chile.
- A partir de 2006 la institución a cargo de la selección y actualización de la muestra es el INE.

Instituciones responsables del cálculo de Factores de Expansión

- Entre 1987 y 2006 el cálculo de los factores de expansión está a cargo del Departamento de Economía de la Universidad de Chile.
- En 2009 este cálculo fue realizado por Mideplan.

Cálculo de factores de expansión

- Entre 1987 y 1994 los factores de expansión se determinan como el cociente entre las proyecciones de población y la muestra lograda en cada comuna.
- A partir de 1996, se estiman como el inverso de la probabilidad de selección de las viviendas, ajustadas por la proyección de población en hogares particulares en cada comuna (al 30 de noviembre del año de la Encuesta).
- En 2005, se desarrollan nuevos factores de expansión para las encuestas Casen 1996, 1998 y 2000, en atención a la disponibilidad de las proyecciones de población basadas en el Censo de Población y Vivienda del año 2002.

Tamaño muestral efectivo

- Hogares entrevistados entre 1987 y 1998: 22.793 en 1987, 25.793 en 1990, 35.948 en 1992, 45.379 en 1994, 33.636 en 1996, y 48.107 en 1998).
- Hogares entrevistados entre 2000 y 2009: 65.036 en 2000, 68.153 en 2003, 73.720 en 2006 y 71.460 en 2009.

¹³ Número de comunas autorepresentadas: 2006 (n=335) y 2009 (n=334).

2. Cambios en diseño muestral 2011

2.1. Evaluación del diseño muestral vigente

En relación a la muestra, expertos han señalado que (1) la determinación del tamaño muestral requiere considerar la variabilidad de las características bajo estudio, (2) el tamaño de la muestra no es suficiente para producir estadísticas confiables al nivel comunal, pero a la vez (3) es demasiado grande y genera problemas operativos que afectan la calidad de los datos. Ver informe CEPAL (2008, pág. 6)¹⁴ e informe Comisión Técnicos Casen (2010, págs. 6 y 7)¹⁵.

Como se explicará más tarde, la determinación del tamaño de una muestra es función del nivel de precisión que se requiere para la producción de estadísticas para ciertos grupos de la población denominados dominios de estudio. Hasta 2009, el Ministerio definió las comunas como dominios de estudio, por lo tanto la muestra consideraba un número mínimo de viviendas seleccionadas en cada comuna. Esto, ya que el objetivo final era producir estimaciones de la tasa de pobreza para cada una de las comunas en la muestra a partir de los datos de la encuesta.

Lo que no se explicita en los documentos metodológicos, sin embargo, es cuál fue el nivel de precisión utilizado para determinar el tamaño muestral requerido para cada comuna. Es decir, si bien se planeaba contar con entrevistas y estimaciones en todas las comunas, no se explicitaba en nivel de confiabilidad estadística (nivel de precisión) que se esperaba en las estimaciones. Las observaciones de la Comisión Técnicos Casen en relación a este punto son ilustrativas. Para algunas comunas en 2009 el margen de error (nivel de precisión) para la estimación de la tasa de pobreza es tan grande que los resultados se tornan poco informativos (nivel de precisión bajo) para la toma de decisiones.

En este contexto, el Ministerio opta por remplazar las comunas como dominio de estudio por otros que estén más en línea con los objetivos de la encuesta – producir información confiable (de alta precisión) para informar la toma de decisiones en la esfera de las políticas públicas. Se opta entonces por definir como nuevos dominios de estudio a las regiones, en atención a la creciente demanda de información con que se requiere a este nivel territorial.

¹⁴ "Se argumenta la necesidad de recalcularse el tamaño de la muestra, redefiniendo los objetivos temáticos y acotando los dominios de estudio. Así mismo, se recomienda encarar el cálculo del número de hogares que se deben visitar, a partir de las tasas de pobreza e indigencia, los errores de muestreo, el efecto de diseño, la precisión deseada y la tasa de respuesta esperada observada en la Casen 2006. Es decir, utilizando la evidencia vigente"(Informe CEPAL 2008, Pág. 7).

¹⁵ "... el total de hogares en la encuesta Casen es demasiado grande y origina problemas de calidad que afectan a todo el instrumento por las dificultades inherentes a organizar un trabajo de campo en donde se logren entrevistar cerca de 70.000 hogares repartidos en todas las comunas del país en un período de tiempo que no debiera exceder a un mes. Por otra parte, es un tamaño de muestra demasiado pequeño para la estimación de estadísticas confiables a nivel comunal, especialmente de características de baja frecuencia (como el porcentaje de extrema pobreza). En el documento metodológico de la Casen 2006 (ver página web de Mideplan) se entrega el error muestral absoluto del dato comunal, el cual fluctúa entre 6,0 y 7,0 para la gran mayoría de las comunas. Esto significa que la estimación de un indicador como el porcentaje de pobreza de la comuna está afecto a un intervalo de confianza que las más de las veces torna no significativo los cambios del indicador en el tiempo. En suma, se trata de dos objetivos distintos que no se pueden conciliar en un único instrumento."(Comisión Técnicos casen 2010, Pág. 6).

Re-definir los dominios de estudio para la encuesta Casen implica re-definir el diseño de la muestra. El diseño 2011 se desarrolló siguiendo prácticas estándar para el diseño de muestras para encuestas de hogares, en la línea de las consideraciones discutidas en el informe CEPAL (2008). La sección V de este informe describe en más detalle los criterios y los parámetros utilizados.

Es importante mencionar, sin embargo, que la eliminación de las comunas como dominios de estudio no implica la eliminación de las comunas de la muestra Casen 2011. En efecto, como se explica más adelante, la muestra Casen 2011 seleccionó viviendas en las 324 comunas incluidas en el marco muestral del INE. A diferencia de años anteriores, sin embargo, las estimaciones de la tasa de pobreza comunal no serán desarrolladas exclusivamente a partir de los datos de la encuesta, sino que se serán desarrolladas a partir de una metodología de estimación denominada Estimación para Áreas Pequeñas (en inglés Small Area Estimation, SAE). Cabe mencionar que la metodología SAE combina información de encuestas y otras fuentes usando modelos predictivos estadísticos. Esta metodología es utilizada por gobiernos en EEUU, Reino Unido, Australia y Alemania para la asignación de recursos en programas públicos y el desarrollo de mapas de pobreza. En Chile, el Ministerio se encuentra desarrollando la implementación de la metodología a partir de 2011 con la asistencia técnica del Programa de Naciones Unidas para el Desarrollo (PNUD).

2.2. Principales características del diseño muestral 2011

El diseño muestral de la Encuesta Casen, en la mayoría de sus versiones, corresponde a una muestra probabilística, estratificada geográficamente y por tamaño poblacional, tanto en las áreas urbanas como rurales. La selección de la muestra se realiza en dos o tres etapas en las áreas urbanas y rurales. Las unidades primarias de muestreo (UPM) son seleccionadas con probabilidad proporcional al tamaño y las unidades de última etapa (viviendas) son seleccionadas con igual probabilidad al interior de cada UPM.

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 definida por INE. Al interior de cada vivienda seleccionada, se entrevista a todos los hogares y a las personas que se declaran residentes habituales. El informante idóneo para la Encuesta Casen es el jefe de cada hogar.

El diseño muestral 2011 mantiene las características fundamentales de las versiones anteriores, sin embargo presenta algunas diferencias. A continuación de identifican las diferencias más importantes. Las siguientes secciones discuten estos puntos en mayor detalle:

- a) Marco Muestral Maestro. Las muestras 2006 y 2009 fueron seleccionadas a partir del marco de secciones (MS2002) en áreas urbanas y rurales. La muestra Casen 2011 fue seleccionada a partir del marco de manzanas en las áreas urbanas (MM2008) y del marco de secciones en las áreas rurales (MS 2002).
- b) Dominios de estudio. Las encuestas 2006 y 2009 definieron como dominios de representación los siguientes niveles de la división político-administrativa: país, regiones, y comunas, según área urbano/rural. Para 2011 se ha definido como dominios de representación geográfica el nivel país y sus 15 regiones, según área urbano/rural.

- c) Objetivos de precisión y tamaño de muestra. Hasta 2009, la muestra nacional se definió a partir de tamaños fijos para cada comuna en la muestra, de esta forma, regiones con más comunas concentraban más muestra, que regiones con menos comunas. Para Casen 2011, en cambio, se definió el tamaño muestral en función de la optimización de la tasa de pobreza regional, bajo el criterio del error absoluto entre 1 y 4 puntos porcentuales a nivel regional. La distribución de la muestra entre comunas al interior de cada región se realiza en forma proporcional al tamaño de la población de cada comuna.
- d) Selección de dos muestras independientes. Con la finalidad de ampliar el periodo de recolección de datos, pero a la vez no sacrificar la comparabilidad con la serie histórica Casen, se seleccionó la muestra completa (90 mil viviendas) en 2 submuestras representativas de la población regional y nacional.
- La submuestra 1 se levanto entre oct11-nov11 y sirvió como base para la selección de la Encuesta EANNA 2012 (n=10.000). La encuesta EANNA se aplicó entre feb12-abr12 a hogares con niños, niñas y adolescentes entre 5 y 17 años.
 - La submuestra 2 se levantó entre nov11-ene12 y sirvió como base para el desarrollo de las estimaciones oficiales de pobreza para Casen 2011. Esta muestra es comparable, tanto en tamaño como en el periodo de recolección de datos de encuestas Casen anteriores¹⁶.
 - La muestra completa servirá como línea de base para las estimaciones que se realizaran, en base a los datos del trimestre oct-nov-dic, posteriormente.

Para lograr estos objetivos, en 2011 el Ministerio estableció un convenio con el Instituto Nacional de Estadísticas (INE) que amplía significativamente los espacios de colaboración entre ambas instituciones. En el marco de Casen 2011, el INE desarrolló, a solicitud del Ministerio, la simulación de tamaños óptimos muestra, la selección y actualización de la muestra, el cálculo de los factores de expansión, la estimación de los errores de muestreo bajo muestreo complejo, y en conjunto con el Observatorio Social del Ministerio, este informe del diseño muestral de la Encuesta. A continuación se describe en detalle el procedimiento utilizado para la determinación y selección de la muestra 2011.

¹⁶ Al igual que la mayoría de sus predecesoras, la encuesta Casen 2011 concentra cerca del 60% de las entrevistas en el mes de diciembre (2011).

V. 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 2011 es todo el territorio nacional, excluyendo aquellos sectores identificados por el INE como áreas de difícil acceso.

El diseño muestral de la Encuesta Casen 2011 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 en tres etapas (muestreo trietápico) en las áreas urbanas. Las unidades primarias de muestreo se seleccionan con probabilidad proporcional al tamaño (en cuanto al total de viviendas), mientras que las unidades últimas (las viviendas) se seleccionan con igual probabilidad.

La unidad de selección de la muestra es la vivienda. Al interior de la vivienda se identifican hogares y las personas miembros de cada hogar. La entrevista se realiza con un informante por hogar¹⁷ que corresponde generalmente al jefe de hogar¹⁸. A través del informante se recolectan datos de todos los miembros del hogar que son residentes habituales de la vivienda¹⁹. Al interior de la vivienda se recolectan datos de todos los residentes habituales y los diferentes hogares que la conforman.

1. Características del Marco Muestral

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

La población objetivo de la encuesta Casen son los hogares que residen en viviendas particulares ocupadas. El INE no mantiene un listado con las direcciones de todas las viviendas particulares del país, sin embargo, mantiene un marco de áreas geográficas que sirven de base para la selección de viviendas, requeridas para las muestras de las encuestas de hogares más importantes del país.

¹⁷ "Se consideran miembros de un hogar a todas aquellas personas que, siendo residentes de una misma vivienda, pueden tener (o pueden no tener) vínculos de parentesco y habitualmente hacen vida en común, es decir, se alojan y se alimentan juntas. Dicho de otra forma -- habitan en la misma vivienda y tienen presupuesto de alimentación común. No se consideran miembros del hogar las personas que pagan pensión (por lo tanto forman otro hogar), en caso de cumplir con los requisitos de permanencia. La pregunta h9 en este módulo permiten identificar a los miembros del hogar que comen juntos o bien comparten los gastos de alimentación, facilitando de este modo la identificación del concepto de hogar." Manual del Trabajo de Campo 2011 (pág. 23).

¹⁸ En la encuesta Casen el jefe de hogar es identificado por los propios entrevistados. No hay instrucciones pre-establecidas para determinar al jefe de hogar (ej. que sea el residente de mayor edad, que sea el dueño de la propiedad, etc.). Si no es el jefe de hogar el entrevistado podría ser cualquier mayor de 18 años.

¹⁹ Hasta 2009, todas las preguntas en el cuestionario Casen pueden ser respondidas por un tercero (el informante). En el cuestionario 2011, sin embargo, las preguntas sobre satisfacción con la vida (r20) a inclusión financiera (y29, y30) sólo pueden ser respondidas por las personas que están físicamente presentes durante la entrevista. Es decir, para estas preguntas no se acepta el reporte de terceros.

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

- Ciudad (CD): Es un gran centro urbano conformado por uno o un conjunto de centros urbanos adyacentes con 40.000 ó más habitantes.
- Resto de Área Urbana (RAU): Conformado por un conjunto de centros urbanos que totalizan menos de 40.000 y más de 2.000 habitantes. Esta clasificación se da cuando en una comuna existe una ciudad (CD) y entonces todos los centros urbanos restantes, si es que existen, se denominan resto de área urbana (RAU).
- Urbano (U): Es el centro urbano con menos de 40.000 y más de 2.000 habitantes. Esta clasificación se da cuando en la comuna no existe una ciudad (CD), por lo que cada uno de sus centros urbanos se denominan simplemente como urbanos (U).
- Rural (R): Conformado por el conjunto de entidades clasificadas como rurales de acuerdo a un tamaño poblacional menor a 1.000 habitantes o entre 1.001 y 2.000 habitantes con predominio de población económicamente activa (PEA) dedicada a actividades primarias.²⁰

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

Para cumplir con los objetivos de investigación, el marco de muestreo "ideal" debe ser completo, preciso y actualizado. A mediados del periodo intercensal, investigaciones realizadas por el INE indican que el marco muestral en uso (MM2002) presentaba deficiencias en estos aspectos, por lo tanto a partir de 2008 el Instituto desarrolló un nuevo marco de muestreo, definiciones para la conformación de unidades de muestreo y procedimientos para la selección de unidades de muestreo en las áreas urbanas del país²¹.

Bajo el nuevo marco de muestreo urbano, se mantiene la división político-administrativa, pero se cambia la conformación de las unidades primarias de muestreo. En el nuevo marco, las manzanas censales (predios urbanos delimitados por calles) remplazan a las secciones de empadronamiento censal en las áreas urbanas²². El que permite actualizar más rápidamente las unidades muestrales que el marco antiguo por medio del plan municipal de edificaciones que se entrega por convenio al Departamento de Gestión del Marco, logrando

²⁰ Se entiende por Actividad Primaria a toda aquella actividad relacionada con la extracción de recursos naturales. (agricultura, caza, pesca, minería, etc.).

²¹ Para mayor información ver documento: "Actualización del diseño muestral Nueva Encuesta Nacional de Empleo", Julio 2006, Instituto Nacional de Estadísticas. Disponible en la página web: http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/metodologia.php.

²² 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.

captar nuevos desarrollos urbanos tanto en áreas urbanas como en áreas previamente catalogadas como agrícolas (por ejemplo, nuevos desarrollos en comunas como Peñalolén, Lampa o Chicureo en la Región Metropolitana).

El INE comenzó a seleccionar las muestras urbanas utilizando el nuevo marco de muestreo a partir del año 2008. 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, la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) en 2008, y la Encuesta de Uso de Drogas en Población General en 2008. En las zonas rurales, sin embargo, el INE ha seguido utilizando el marco de secciones (MMS2002) para la selección de muestras.

La selección de la muestra 2011 se realizó sobre el marco muestral que el INE mantiene vigente al año 2011, el cual comprende el uso de dos marcos de muestreo que son mutuamente excluyentes:

- En el área rural, el marco de muestreo corresponde al generado a partir del Censo de Población y Vivienda del año 2002. Las unidades primarias de muestreo se denominan secciones y corresponden al área que puede recorrer un censista el día del censo. En adelante, denominaremos este marco muestral como marco de secciones (MS2002).
- En el área urbana, el marco de muestreo es aquel generado a partir de cartografía digital de alta validez métrica, actualizado al segundo semestre del año 2008. La información en el marco se actualiza con información anexa de Mapcity, Dmapas y los registros administrativos asociados a nuevas construcciones, otorgados por los municipios. Las unidades primarias de muestreo en el área urbana se denominan manzanas y corresponden a delimitaciones geográficas fijas. En adelante, denominaremos a este marco como marco de manzanas (MM2008).

El marco de manzanas clasifica las manzanas en 30 grandes grupos de tamaño de acuerdo al número total de viviendas en la manzana. Más de la mitad de las manzanas en el marco tiene entre 8 y 44 viviendas, mientras que el grupo de tamaño más grande corresponde a manzanas con entre 155 y 1.043 viviendas, las cuales representan el 17,3% del total de las manzanas en el marco (ver Tabla V.1).

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

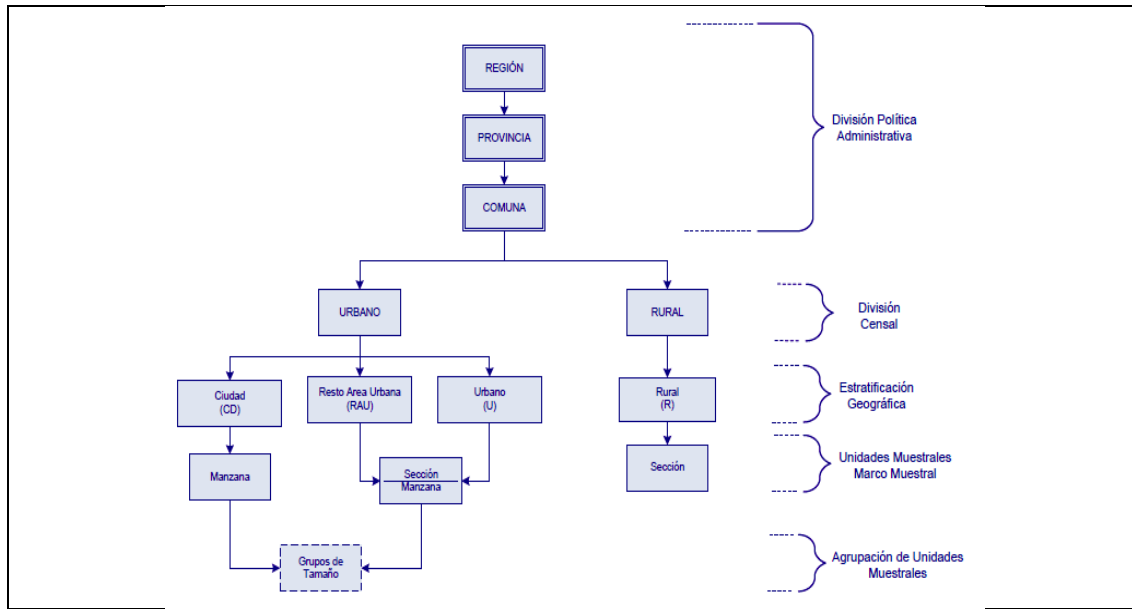


Tabla V.1: Número de manzanas y viviendas en el Marco Muestral de Manzanas, según grupo de tamaño de manzanas.

Grupo de tamaño	Rango de viviendas	Total de Manzanas	Dist. de manzanas	Total de Viviendas	Dist. De viviendas
Grupo 0	1 a 7	13.894	10,4%	53.578	1,3%
Grupo 1 a 4	8 a 23	63.319	47,5%	994.825	24,9%
Grupo 5 a 9	24 a 44	39.267	29,4%	1.216.764	30,5%
Grupo 10 a 19	45 a 81	10.334	7,7%	591.573	14,8%
Grupo 20 a 28	82 a 154	3.990	3,0%	439.327	11,0%
Grupo 29 y 30	155 a 1.043	2.545	1,9%	691.503	17,3%
Total		133.349	100%	3.987.570	100%

Fuente: Elaborado por el Instituto Nacional de Estadística.

1.1. Cobertura del Marco Muestral

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

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

Es importante distinguir la falta de cobertura (fallas no intencionadas), de las exclusiones que realiza el investigador en forma intencionada. Se pueden

identificar tres tipos de exclusiones de este último tipo en el proceso de selección de la muestra Casen 2011:

- las 22 áreas geográficas que han sido catalogadas por INE como áreas de difícil acceso (ADA)²³ ;
- las manzanas y secciones incluidas en otras muestras seleccionadas por el INE para el periodo de recolección de datos de Casen 2011;
- las manzanas con menos de 8 viviendas.

Las áreas de difícil acceso corresponden a zonas geográficas no incluidas en el marco muestral del INE. Estas áreas no están presentes en las muestras de ninguna de las encuestas de hogares seleccionadas por el INE. En total, corresponden al 0,4% de la población de viviendas y al 0,4% de la población de personas del país²⁴. La Tabla V.2 muestra las 22 áreas de difícil acceso excluidas de la muestra Casen 2011.

Por otra parte, la Encuesta Nacional del Empleo (NENE) 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 2011. Para evitar complicaciones en el trabajo de campo de ambas encuestas, el 2011 el INE excluyó del marco de muestreo, previo a la selección de la muestra Casen 2011, todas aquellas manzanas y secciones que forman parte de la muestra ENE. En total, se excluye aproximadamente 12,8% de las manzanas (3,8% de viviendas) y 8,3% de las secciones en el marco muestral.

En las zonas urbanas, el INE excluye del marco muestral aquellas manzanas con 7 o menos viviendas debido a potenciales problemas operativos y de costos. En total, estas manzanas representan el 10,4% del total de manzanas (1,3% de viviendas) del marco.

En total, el marco muestral a partir del cual se selecciona la muestra de la Encuesta Casen 2011, excluye aproximadamente un 4,4% 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 rurales) 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.

²³ Ver más antecedentes acerca del marco muestral maestro del INE en: http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/pdf/ene_espanol.pdf.

²⁴ Estimaciones propias del Instituto Nacional de Estadísticas.

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

Región	Provincia	Comuna
Arica y Parinacota	Parinacota	General Lagos
Tarapacá	Tamarugal	Colchane
Antofagasta	El Loa	Ollagüe
Valparaíso	Valparaíso	Juan Fernández
	Isla de Pascua	Isla de Pascua
Los Lagos	Llanquihue	Cochamó
	Palena	Chaitén
		Futaleufú
		Hualaihué
		Palena
Aisén del Gral. Carlos Ibáñez del Campo	Coihaique	Lago Verde
	Aisén	Guaitecas
	Capitán Prat	O'Higgins
		Tortel
Magallanes y de la Antártica Chilena	Magallanes	Laguna Blanca
		Río Verde
		San Gregorio
	Antártica Chilena	Cabo de Hornos (Ex - Navarino)
		Antártica
	Tierra el Fuego	Primavera
		Timaukel
	Última Esperanza	Torres del Paine

1.2. 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: 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/rural²⁵. A partir de 1998 los estratos se constituyen a partir de las comunas, según área urbano/rural²⁶. Para Casen 2011 se conformaron h=584 estratos, a partir de la interposición de la división política-administrativa (a nivel de comunas) y la división censal (urbano-rural). De éstos, 312 estratos se encuentran en zonas urbanas y 273 en zonas rurales.

²⁵ En 1994 se conforman 146 estratos y en 1996, 246 estratos.

²⁶ En 1998 se conformaron 358 estratos. En 2000, 2003, 2006 y 2009 se conformaron 529, 553, 605 y 602 estratos respectivamente.

2. Estimación y Distribución del Tamaño Muestral

El tamaño y distribución de la muestra 2011 tiene tres características que la hacen diferente a sus predecesoras: es más grande a nivel nacional, está distribuida en mayor proporción en las regiones del extremo norte y sur del país, y está distribuida en forma proporcional al tamaño de la población entre comunas al interior de cada región del país. A continuación se describen los procedimientos que se siguieron para determinar el tamaño y distribución territorial de la muestra Casen 2011.

2.1. Estimación del Tamaño Muestral por Región

Uno de los objetivos del diseño 2011 es alcanzar mejores niveles de precisión para las estimaciones regionales de la tasa de pobreza. Para conseguir este objetivo, el tamaño muestral fue determinado en forma independiente para cada una de las regiones del país. La determinación del tamaño muestral se realizó en etapas. La Tabla V.3 muestra los resultados del proceso completo, donde algunas columnas contienen los parámetros estimados a partir de los datos Casen 2009 (tasa de pobreza, efecto diseño, tasa de respuesta), los niveles de precisión requeridos (error absoluto y error relativo) y los tamaños muestrales en cada etapa del proceso (bajo muestreo aleatorio simple, bajo diseño complejo, corregido por finitud y sobredimensionado por no respuesta). La descripción detallada de cada etapa del proceso se realiza en los párrafos que sigue, donde se referencia cada una de las columnas de la Tabla V.3.

Tabla V.3: Tamaño muestral Casen 2011 y parámetros, por región.

Tabla 3: Tamaño muestral CASEN 2011 y parámetros, por región.											
Total	Población Viviendas año 2009 M	Ratio Estimate Pi	Error Absoluto Propuesto (%) e ₀	Error Relativo Propuesto (%) e _r	Tamaños Propuestos Muestreo Aleatorio Simple (M.A.S) m ₀	Design Effect Deff	Tamaños Propuestos Muestreo Bi-etápico con efecto de diseño m ₁	Fracción de Muestreo m ₁ / M	Tamaños Propuestos (M.A.S), corregido por finitud m ₂	Tasa de No Respuesta Tnr	Tamaños corregidos por tasa de No-Respuesta m ₃
Pais	4.595.725	0,151	0,67	4,45	26.882	4,29	93.885	2,0%	90.122	20,7%	112.938
I Region	71.788	0,158	2,4	15,2	1.435	4,76	6.835	9,5%	6.241	19,3%	7.734
II Region	123.376	0,080	1,6	20,0	1.374	4,68	6.426	5,2%	6.108	18,4%	7.486
III Region	68.332	0,175	2,7	15,4	1.530	3,06	4.683	6,9%	4.382	33,7%	6.610
IV Region	198.469	0,166	2,5	15,1	1.216	3,17	3.854	1,9%	3.780	16,2%	4.511
V Region	491.287	0,151	1,9	12,6	1.881	4,36	8.198	1,7%	8.064	16,8%	9.692
VI Region	244.622	0,128	1,9	14,9	2.795	2,00	5.595	2,3%	5.469	17,5%	6.630
VII Region	288.879	0,207	3,2	15,4	1.206	6,35	7.662	2,7%	7.464	9,1%	8.211
VIII Region	553.948	0,210	2,1	10,0	2.599	3,38	8.789	1,6%	8.652	17,5%	10.487
IX Region	267.546	0,271	2,8	10,3	1.951	3,03	5.909	2,2%	5.782	16,5%	6.924
X Region	229.140	0,142	2,7	19,0	1.357	4,85	6.575	2,9%	6.391	18,4%	7.832
XI Region	28.211	0,151	2,3	15,2	3.262	1,40	4.581	16,2%	3.941	29,2%	5.566
XII Region	45.284	0,091	1,8	19,9	1.481	1,65	2.440	5,4%	2.315	39,5%	3.826
RM Metrop	1.833.728	0,115	1,0	8,7	2.552	4,52	11.533	0,6%	11.461	26,0%	15.488
XIV Region	107.635	0,204	4,0	19,6	799	8,76	7.004	6,5%	6.576	16,7%	7.895
XV Region	43.480	0,128	2,0	15,6	1.443	2,63	3.801	8,7%	3.495	13,6%	4.046

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 (i.e. manzanas o secciones); j es la unidad secundaria de muestreo (i.e. viviendas) y k corresponde a la persona seleccionada.

Considérese,

- p : prevalencia de la variable cualitativa de interés en el estudio, corresponde a la tasa de pobreza.
- n : número de conglomerados o unidades primarias de muestreo.
- \bar{m} : número promedio de viviendas a encuestar por unidad primaria de muestreo.
- $n \cdot \bar{m} = m$: número de viviendas a encuestar.
- M : número de viviendas en la población.
- $Deff(p)_r$: Efecto del diseño asociado a la tasa de pobreza (p) en la región r , que se puede interpretar como el aumento o disminución en la varianza, debido a considerar un muestreo complejo (e.g. estratificado, bietapico, por conglomerados) en vez de un muestreo aleatorio simple de viviendas.
- $SE(p)_r^2$: 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 $(M/(M-1)) \cdot (P \cdot Q)$, siendo P la Pobreza Poblacional, con $Q = (1 - P)$ y M el total de viviendas en la población.
- $Z_{1-\frac{\alpha}{2}}$: percentil de nivel $(1-\alpha/2)$ de la distribución Normal, correspondiente a una estimación intervalar de $(1-\alpha)$ de confianza.

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

Paso 1

Se obtienen estimaciones, a nivel regional, de la tasa de pobreza, el error estándar asociado a la tasa de pobreza y el efecto del diseño asociado a este indicador, usando Casen 2009. Estas estimaciones se presentan en las columnas

2 a 4 de la Tabla 6. La cuasi-varianza poblacional $S(p)_r^2$ se deriva a partir del error estándar de la tasa de pobreza 2009 utilizando la siguiente fórmula:

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

Paso 2

Se establecen parámetros para los errores absolutos "objetivo" por región (e_{0r}) que varían entre 1 y 4 puntos porcentuales. También se establece el nivel de

confianza al 95% para obtener el percentil de la distribución normal $\left(Z_{1-\frac{\alpha}{2}} \right)$. Con estos y la cuasi-varianza de la tasa de pobreza derivada en (1), se obtiene un tamaño muestral considerando un muestreo aleatorio simple mono-etápico m_{0r} .

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

Paso 3

Se estima el tamaño muestral considerando el efecto del diseño al aplicar un muestreo en dos etapas. El efecto del diseño representa el ajuste del tamaño anterior debido al cambio que experimentará la varianza al pasar de un muestreo mono-etápico a uno bi-etápico, y con el objetivo de mantener el error absoluto inicial. Para esto se utilizan el efecto del diseño por región asociado a la tasa de pobreza $Deff(p)_r$ (ver columna 7 en Tabla V.3). Los efectos de diseño asociados a las tasas de pobreza regional de Casen 2009 son bastantes disímiles entre las regiones. El Deff más alto se observa para la región de Los Ríos (8,76) y el más bajo en Magallanes (1,4). La fórmula utilizada para el cálculo del tamaño muestral m_{1r} , incorporando el efecto del diseño se presenta a continuación:

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

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. Estos valores se presentan en las columnas 6 y 8 de Tabla V.3.

Paso 4

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

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

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

Paso 5

Una vez definido el total de unidades muestrales a encuestar para estimar la tasa de pobreza, se debe tener en consideración que es posible no lograr el total de unidades por diversas razones, tales como rechazos, moradores ausentes, etc. Con la finalidad de salvaguardar la precisión de la estimación de la tasa de pobreza, se aplica un factor de corrección a la fórmula (4) en base a una estimación de la tasa de no respuesta (tnr_r) regional.

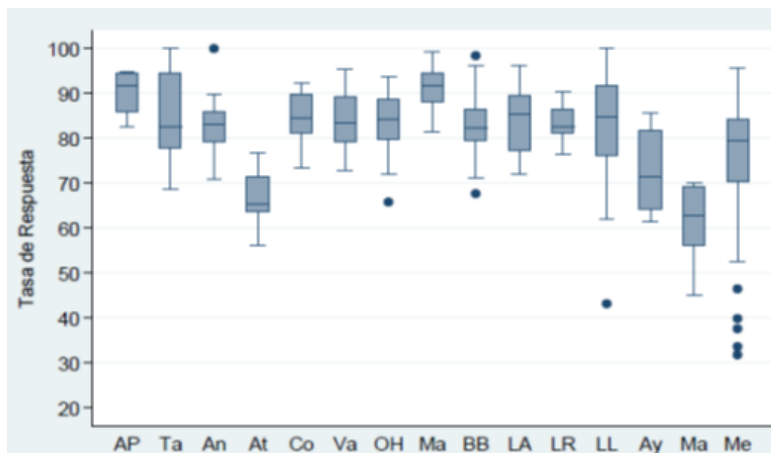
En años anteriores se usó un factor de ajuste por no respuesta parejo para todas las áreas geográficas de aproximadamente 25%. El año 2011, se utiliza la tasa de respuesta regional, estimada a partir de los resultados del trabajo de campo de la Encuesta Casen 2009 (ver figura V.2). Las tasas de no respuesta más altas se observan en las regiones de Magallanes (40%), Atacama (34%) y Metropolitana (26%).

La fórmula que define el tamaño muestral final de la Encuesta Casen 2011, que incorpora finalmente los ajustes conjuntos de finitud, efecto de diseño y tasa de no-respuesta, está dada por:

$$m_{3r} = \frac{Z^2_{1-\frac{\alpha}{2}} \cdot S(p)_r^2}{e^2_{0r} + Z^2_{1-\frac{\alpha}{2}} \cdot Deff(p)_r \cdot \frac{S(p)_r^2}{M}} \cdot \frac{Deff(p)_r}{(1-tnr_r)} = \frac{m_{2r}}{(1-tnr_r)} \tag{5}$$

Donde m_{3r} representa el número de viviendas final a encuestar en la región r con los 3 ajustes mencionados. Ver columnas 11 y 12 de Tabla V.3 para el ajuste por no-respuesta.

Figura V.2: Distribución de tasas de respuesta comunal, según región. Casen 2009.



Fuente: Elaborado por el Ministerio de Desarrollo Social.

Las dos primeras columnas de la Tabla V.4 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 la encuesta Casen 2009²⁷ y 2011. Al comparar las dos muestras, se observa que la distribución de la muestra 2009 se aproxima más a la distribución de la población de viviendas. Esto es consistente con el diseño muestral empelado hasta 2009, que asignaba más muestra a las regiones con más comunas (las cuales tienen mayor población de viviendas). La muestra 2011, sin embargo, fue diseñada para producir estimaciones con cierto nivel de precisión a nivel regional, por esta razón, requiere de tamaños muestrales relativamente similares para lograr similares niveles de precisión en cada región.

Tabla V.4: Distribución de la muestra objetivo para las encuestas Casen 2009 y 2011.

Región	Marco Muestral de Viviendas 2002		Casen 2009		Casen 2011	
	Población	%	Muestra	%	Muestra	%
Tarapacá	71.684	1,6	1.418	1,9	6.241	6,9
Antofagasta	123.340	2,7	1.995	2,7	6.108	6,8
Atacama	68.332	1,5	1.953	2,6	4.382	4,9
Coquimbo	198.469	4,3	3.255	4,4	3.780	4,2
Valparaíso	491.287	10,7	8.045	10,8	8.064	8,9
L. G. B. O'Higgins	244.622	5,3	6.874	9,2	5.469	6,1
Maule	288.879	6,3	6.430	8,6	7.464	8,3
Bío Bío	553.948	12,1	11.850	15,9	8.652	9,6
La Araucanía	267.546	5,8	6.760	9,1	5.782	6,4
Los Lagos	228.440	5,0	5.825	7,8	6.391	7,1
Aisén	28.211	0,6	1.300	1,7	3.941	4,4
Magallanes	45.118	1,0	1.210	1,6	2.315	2,6
Metropolitana	1.833.728	39,9	14.125	19	11.461	12,7
Los Ríos	107.635	2,3	2.535	3,4	6.576	7,3
Arica y Parinacota	43.429	0,9	764	1	3.495	3,9
País	4.594.668	100,0	74.339	100,0	90.122	100,0

Fuente: Elaborado por el Ministerio de Desarrollo Social.

²⁷ La muestra objetivo para Casen 2009 fue de 74.339 viviendas. Al término del trabajo de campo se logró completar entrevistas con 71.460 hogares.

2.2. Distribución de la Muestra entre Comunas al interior de la Región

Una vez establecido el total de viviendas en cada región r , se realiza la distribución de las viviendas por comuna y área geográfica de las 90.122 viviendas objetivo (m_{2r}). A continuación se describe paso a paso el protocolo de distribución de la muestra Casen 2011.

Paso 1

Al interior de cada región, la muestra de viviendas se distribuye en forma proporcional al peso de la población de viviendas de cada comuna, en relación al tamaño de la población total de viviendas en la región, según el Censo 2002.

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

Cabe mencionar que se ha determinado un par de restricciones adicionales dentro del protocolo de distribución de la muestra 2011:

- Se establece un número mínimo de viviendas por estrato: 50 en estratos urbanos y 30 en estratos rurales²⁸.
- Se establece un mínimo de 2 manzanas (o secciones) por estrato, tanto para estratos urbanos como rurales.

La primera restricción, que implica fijar un mínimo de viviendas por estrato, busca estabilizar las estimaciones de varianza. Dicho de otro modo, reducir la variabilidad de las estimaciones de la varianza. La segunda restricción, que implica fijar un mínimo de 2 manzanas (o secciones) por estrato, busca hacer factible la estimación de varianza bajo muestreo complejo al evitar, por diseño, el caso de un solo conglomerado por estrato.

En términos concretos, la aplicación de estas restricciones sólo implica la redistribución de la muestra al interior de una región. Por ejemplo, en la comuna de San Rosendo son 95 viviendas en la muestra, pero al revisar la distribución poblacional de las viviendas en el Marco muestral se aprecia que el 2,8% de las viviendas se concentra en el área rural, esto significaría que en dicha área sólo 3 viviendas pertenecerían al área rural de San Rosendo, sin embargo al aplicar la regla del mínimo de viviendas el total de viviendas a seleccionar queda en 30.

Paso 2

El siguiente paso consiste en sobredimensionar la muestra objetivo ($n=90.122$) para contrarrestar las pérdidas que se originan producto de distintas formas de no-respuesta (ej. unidades no elegibles, falta de contacto, rechazo a la entrevista, etc.). El diseño 2011 toma un ajuste diferenciado por región y comuna, según la no-respuesta observada en el proceso de encuestaje Casen

²⁸ Con 50 viviendas en el área urbana, se asegura un mínimo de 6 manzanas (tomando en cuenta que en promedio se encuestan 8 viviendas por manzana. En el área rural, se asegura un mínimo de dos secciones, tomando también en cuenta que en esta área se encuestan en promedio 15 viviendas por sección).

2009. Las tasas de respuesta se pueden interpretar como una estimación de la probabilidad de responder de los hogares en las comunas seleccionadas²⁹. Las tasas de respuesta a nivel comunal presentan alta variabilidad en cada región. En atención a esto, se optó por definir cinco grandes categorías para las tasas de respuesta y utilizar esta aproximación (tasas de respuesta "suavizadas") en vez de las estimaciones directas de las tasas de respuesta comunales.

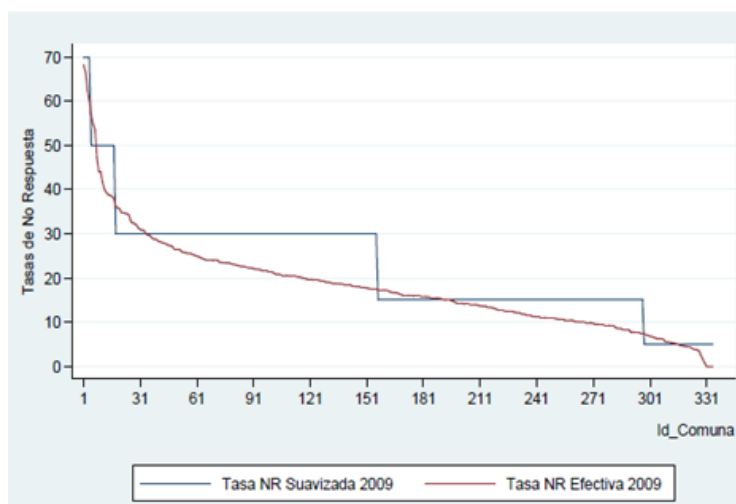
El Tabla V.5 muestra que las tasas de respuesta suavizadas están "cargadas" hacia los extremos superiores de cada intervalo, con el objeto de aplicar un criterio conservador y asegurar el tamaño de muestra suficiente al corregir por este factor.

Tabla V.5: Tasas de No-Respuesta suavizadas

Tasas de No-Respuesta efectiva		Tasa de No-Respuesta Suavizada
Límite Inferior	Límite Superior	
≤ 7,5		5
7,5	17,5	15
17,5	37,5	30
37,5	60	50
> 60		70

Fuente: Elaborado por el Instituto Nacional de Estadística.

Figura V.3: Tasas de No-Respuesta suavizadas utilizadas para sobredimensionar la muestra Casen 2011 en cada comuna.



Fuente: Elaborado por el Ministerio de Desarrollo Social.

²⁹ Ver más detalles del marco conceptual de la no respuesta en encuestas de hogares en "Nonresponse in Household Surveys" (Groves and Couper, 1998).

Paso 3

Finalmente, la sobredimensión de la muestra generada por la no-respuesta a nivel comunal es ajustada para coincidir con la no-respuesta a nivel regional. Esto, dado que la muestra ajustada a nivel comunal es más grande que la ajustada a nivel regional, por lo que hay que disminuirla hasta el nivel regional, que es el fijado inicialmente. Aquí se utiliza un simple ajuste de razón que redistribuye la sobredimensión en forma proporcional al peso de cada comuna en la región.

Con esta disminución forzada, algunas de las tasas iniciales de no-respuesta disminuyen a cero y en algunos casos pasan a ser negativas, por lo que se aplica la restricción que la tasa mínima de no-respuesta sea aproximadamente del 2,5%. Las tasas (efectivas) de no-respuesta final aplicadas en cada comuna están en el Anexo. Sin embargo en la Tabla V.6 se visualiza cómo funciona el mecanismo descrito para el caso de la región de Tarapacá:

Para la región de Tarapacá, se determinó un tamaño regional para la muestra objetivo de 7.734 viviendas. A nivel de comuna, este total se redistribuyó en forma proporcional al tamaño de la población entre sus 6 comunas (columna 1). Para determinar la muestra a seleccionar, se aplicó una sobredimensión de 5% (Pozo Almonte y Camiña), 15% (Pica) y 30% (Iquique, Alto Hospicio y Huara) (columna (2)), lo cual lleva a exceder el total regional objetivo (columna (3)). Al reajustar el total regional (columna (4)), se recalcula la tasa de no respuesta y las comunas de Pozo Almonte y Camuña quedan con tasas negativas (columna (5)). Para corregir este problema, se fija un piso de 2,5% de tasa de respuesta para estas comunas y se redistribuye la muestra entre el resto de las comunas en la región. El tamaño final de la muestra a seleccionar en cada comuna se despliega en la columna (6) y las tasas de sobredimensión (no respuesta) efectivas en la columna (7).

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

Tabla V.6: Ejemplo de distribución de tamaño de muestra para Región de Tarapacá y sus comunas.

REGION Comuna	Tamaño Inicial Distribuido (1)	Tasas de No-Respuesta Suavizadas Comunales (2)	Tamaño Inicial con Ajuste por No-Respuesta Regional (3)	Tamaño Final con Ajuste por No-Respuesta Comunal (4)	Tasa Inicial Efectiva de No-Respuesta (5)	Tamaño Definitivo Viviendas Ajuste a Región (6)	Tasa Final Efectiva de No-Respuesta (7)
TARAPACA	6.241		8.740	7.734	0,193	7.734	0,193
Iquique	3.822	0,30	5.460	4.831	0,209	4.810	0,205
Alto Hospicio	1.711	0,30	2.444	2.163	0,209	2.153	0,205
Pozo Almonte	275	0,05	289	256	-0,074	282	0,025
Camiña	65	0,05	68	61	-0,066	67	0,030
Huara	180	0,30	257	228	0,211	227	0,207
Pica	188	0,15	221	195	0,041	195	0,036

Fuente: Elaborado por el Instituto Nacional de Estadística.

3. Selección de Conglomerados

Como se mencionó en la sección V apartado 1, el INE no mantiene un listado con las direcciones de todas las viviendas particulares del país, sin embargo, mantiene un marco de áreas que contiene un listado de unidades (que denominaremos conglomerados) que sirven de base para la identificación y posterior selección de las viviendas – que son las unidades últimas de selección de la muestra.

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

3.1. Selección de Conglomerados desde el Marco Urbano

El marco muestral urbano está particionado en 119.455 manzanas y estratificado en 30 estratos, que reflejan el tamaño de las manzanas en el marco (ver Tabla en Anexo). La selección de las manzanas se realiza en cada estrato de la muestra de forma independiente, aleatoria y proporcional al tamaño en cuanto al total de viviendas y manzanas por cada grupo de tamaño³⁰. La selección se implementa en base a un algoritmo de selección que considera las dos etapas, que se describen a continuación. El procedimiento descrito permite seleccionar en la muestra manzanas grandes y pequeñas.

³⁰ Los grupos de tamaño son conglomerados de manzanas censales. Los grupos de tamaño son las unidades primarias de muestreo que se utilizan en el procedimiento de selección implementado en el marco de manzanas de INE.

Etapa 1

Selección de un grupo de tamaño. Las manzanas se agrupan en i grupos de tamaño a nivel del marco de manzanas. Esta información se cruza con las comunas, de modo que en cada comuna puede existir a lo más 30 grupos ($i = 1, 2, \dots, z$ donde $z \geq 30$). El procedimiento aplicado es un proceso iterativo³¹ en que se escoge en forma aleatoria un grupo de tamaño i en forma proporcional a su tamaño. Este último medido en términos de W_i que corresponde al producto entre el número total de manzanas y el número total de viviendas de cada grupo i ($W_i = N_i * M_i$).

Tabla V.7: Método de selección del grupo de tamaño y manzanas.

Grupo de Tamaño G_i	Número de Manzanas N_i	Número de Viviendas M_i	Peso del Grupo de Tamaño $W_i = N_i * M_i$	Ponderación $P_i = 1/W_i$
1	N1	M1	$N1 * M1$	P1
2	N2	M2	$N2 * M2$	P2
3	N3	M3	$N3 * M3$	P3
:	:	:		
:	:	:		
:	:	:		
z	Nz	Mz	$Nz * Mz$	Pz
Total	Nh	Mh	$\sum W_i$	1

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Donde,

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

Etapa 2

Selección de una manzana. Una vez escogido el grupo de tamaño, la manzana se elige con probabilidad igual³².

Enseguida se descuenta la manzana (disminuye en uno el número de manzanas en el grupo y en el total de la comuna y/o estrato) y también se descuenta el número de viviendas³³ (disminuye el número de viviendas en el grupo seleccionado y en el total de la comuna y/o estrato), para finalmente actualizar tanto el número de manzanas y viviendas en cada grupo de tamaño y repetir el procedimiento hasta seleccionar todas las manzanas que completan el número de viviendas requeridas.

³¹ En el programa R se aplica la función SAMPLE con Probabilidad = P_i .

³² Otra vez se aplica en R la función SAMPLE, pero ahora con Probabilidad = $1/N_i$, donde N_i es el número de manzanas que contiene el grupo i .

³³ De acuerdo al número total de viviendas que se obtiene de la enumeración, cada manzana tiene predeterminado el total de viviendas que pueden ser seleccionadas para una muestra. Este total es consistente con una fracción de muestreo de aproximadamente 25% (ver tabla C en los Anexos). Para efectos de selección, sin embargo, se descuenta el número de viviendas enumeradas con el fin de recalcular las probabilidades de selección P_i .

La Tabla V.9 muestra que, a partir del marco urbano, se seleccionaron inicialmente un total de 3.636 manzanas para la submuestra 1 y 5.600 manzanas para la submuestra 2.

3.2. Selección de Conglomerados desde el Marco Rural

El marco muestral rural está particionado en 7.505 secciones en la parte Urbana (U) y Resto de Área Urbana (RAU), y en 6.679 secciones en la parte Rural (R). La selección preliminar de las secciones se realiza al interior de cada estrato Urbano, RAU o Rural de la comuna en forma proporcional al total de viviendas ocupadas que éstas contienen, según información del Censo 2002. Esta selección se realiza en forma sistemática proporcional al tamaño, mediante un algoritmo de acumulación de viviendas en el estrato o comuna, en el que se elige un período y arranque aleatorio, como sigue:

Paso 1

Se construyen Q intervalos (equivalente al número de secciones del Estrato o comuna) de la manera como se muestra en la Tabla V.8:

Tabla V.8: Método de selección de unidades primarias secciones

Sección	Nº Viviendas	Acumulado del número de viviendas	Límite inferior	Límite superior
i	M _i	ΣM _i	L _{inf}	L _{sup}
1	M ₁	Σ ₁ = M ₁	1	Σ ₁
2	M ₂	Σ ₂ = M ₁ + M ₂	Σ ₁ + 1	Σ ₂
3	M ₃	Σ ₃ = M ₁ + M ₂ + M ₃	Σ ₂ + 1	Σ ₃
□	□	□	□	□
Q-1	M _{Q-1}	Σ _{Q-1} = M ₁ + M ₂ + . . . + M _{Q-1}	Σ _{Q-2} + 1	Σ _{Q-1}
Q	M _Q	Σ _Q = M ₁ + M ₂ + . . . + M _Q	Σ _{Q-1} + 1	Σ _Q

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

El procedimiento anterior, para secciones homogéneas en cuanto al tamaño (de una cantidad de viviendas similar), no permite repeticiones. Por el contrario, cuando las secciones son de tamaño muy distinto, llegando a ser algunas el doble y más del doble de las demás, el procedimiento o algoritmo selecciona menos de las requeridas, porque esencialmente puede seleccionar dos o más veces una misma sección, si ésta es demasiado grande respecto a las demás. Se puede demostrar que para secciones homogéneas en cuanto al tamaño, la probabilidad de inclusión de una sección con M_i viviendas está dada por $n \cdot M_i / \sum Q$.

La Tabla V.9 muestra que, a partir del marco de secciones, se seleccionaron inicialmente un total de 507 secciones para la submuestra 1. De estas 507 secciones, 398 corresponden a entidades catalogadas como rural (R) mientras que 109 corresponden a entidades RAU o U. Para la submuestra 2 se seleccionaron 736 secciones, donde 548 de ellas corresponden a áreas rurales y 188 corresponden a áreas RAU o U.

4. Selección de Viviendas

Como se desprende de la argumentación anterior, la selección de conglomerados (manzanas y secciones) se realiza en dos etapas en el marco de manzanas³⁴ y en una etapa en el marco de 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. El INE es el responsable de realizar estas actividades. Para Casen 2011, estas actividades fueron realizadas en paralelo al proceso de levantamiento del Pre Censo 2012.

4.1. Enumeración de los Conglomerados

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 2011, contempló áreas urbanas y rurales.

³⁴ En el marco urbano, la primera unidad de selección corresponde a un *grupo de tamaño* y la segunda unidad de selección corresponde a una *manzana* dentro del *grupo de tamaño* seleccionado.

Material de trabajo del enumerador. A cada enumerador se entrega el siguiente set de materiales:

- Plano de ubicación comunal: Plano de la comuna donde se encuentran identificadas todas las manzanas y secciones seleccionadas en la muestra.
- Croquis de la manzana o sección: Plano a menor escala que identifica sólo la manzana o sección seleccionada.
- 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 V.4 (los formularios de áreas rurales son similares).

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.

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.

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

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

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

Figura V.4 Formulario de Enumeración en Áreas Urbanas, Casen 2011.

ENUMERACIÓN DE VIVIENDAS EN ÁREAS GEOGRÁFICAS URBANAS CASEN 2011 <small>LOS DATOS QUE EL INE SOLICITA SON CONFIDENCIALES. EL ARTÍCULO 29 LEY 17.374 ESTABLECE EL SECRETO ESTADÍSTICO</small>										Hoja Nº ____ de ____			
 <small>INSTITUTO NACIONAL DE ESTADÍSTICAS SUBDIRECCIÓN DE OPERACIONES Depto. Estadísticas Sociales y Proyectos Especiales Avda. Bulnes 205-Santiago Fono: 02-7962301 02-7962328 02-6711316 02-7962323</small>		IDENTIFICACIÓN GEOGRÁFICA Uso Exclusivo de Secciones:					Estrato			Sección			
		Comuna		Distrito		Zona		Manzana		Nombre de comuna			
Nº de orden de la edificación	Dirección					Uso o destino de la edificación 1.- Moradores Ausentes 2.- Desocupada 3.- Temporada 4.- Colectiva 5.- Sin Vivienda 6.- Otro Uso	Nº de Orden de la Vivienda	Nº de Orden del Hogar	Número Total de Personas en el Hogar (Incluir menores de 15 años)	Anotaciones complementarias. Anote el número del código en la columna correspondiente y describa en la columna de la derecha.			
	Tipo de Dirección 1.- Calle 2.- Pje 3.- Avda 4.- Cam 5.- Paseo 6.- Otro	Nombre de Calle, Pje, Avda, Camino u Otro (Indique nombre si corresponde a Otro Tipo)								Número domicilio	Block o Torre	Nº de Piso	Nº de letra, Depto. o pieza
											3.- Rechazo 4.- Describir tipo de vivienda colectiva 5.- Observaciones sin vivienda 6.- Describir actividad de otro uso de la edificación.	Código	Descripción y/u Observación

4.2. Selección de las Viviendas

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

En el marco de manzanas la fracción de muestreo que se aplica al interior de cada manzana es de 25% aproximadamente, con un mínimo de 2 viviendas a encuestar. Para no concentrar la muestra en manzanas grandes (en tamaño de viviendas), se establece una cota superior de 31 viviendas a encuestar. Esto permite que manzanas de todos los tamaños estén representadas en la muestra en forma proporcional, a través de un número fijo de viviendas, según el grupo al que pertenezca. Este criterio de selección se desarrolló originalmente para la Encuesta Nacional de Empleo y el INE la aplica para la selección de muestras urbanas de todas las encuestas de hogares. La Tabla VIII.3 en el Anexo presenta el total de viviendas a seleccionar en cada manzana según su grupo de tamaño.

En el marco de secciones la selección de viviendas se realiza al interior de cada sección, en forma sistemática, con igual probabilidad. Específicamente, el algoritmo de selección sistemático de viviendas, que es aplicado también en otras encuestas, se estandariza para que independientemente de quien aplique el mismo algoritmo, bajo las mismas condiciones de ordenamiento previo ³⁵, seleccione las mismas viviendas. Este procedimiento es análogo a introducir una semilla de aleatorización para selecciones aleatorias, las cuales arrojan también las mismas selecciones.

³⁵ Si se trata de manzanas, el ordenamiento es comuna-distrito-zona-manzana-orden_vivienda. Si se trata de secciones, el ordenamiento es comuna-estrato-sección-manzana-orden_vivienda.

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

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

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

Podría pensarse que se resta aleatoriedad al partir la selección siempre con la mitad entre uno y el período, pero el arranque es tan impredecible debido a que se pre-seleccionan sólo las elegibles, generando que la primera vivienda elegible no corresponda con la primera vivienda de acuerdo al orden de la vivienda según recorrido de la enumeración. La Tabla V.9 muestra la distribución total de las unidades primarias y secundarias de muestreo, según submuestra y área urbana-rural. La Tabla V.10 revela la distribución de las unidades secundarias de muestreo, por región y submuestra.

Tabla V.9: Total de conglomerados y viviendas, según submuestra y área.

Unidades de muestreo	Total	Submuestra 1			Submuestra 2		
	(1)	Urbano (2)	Rural (3)	Total (4)	Urbano (5)	Rural (6)	Total (7)
Estratos	585	312	273	585	311	273	584
Conglomerados	10.479	3.745	398	4.143	5.788	548	6.336
Manzanas	9.236	3.636	0	3.636	5.600	0	5.600
Secciones	1.243	109	398	507	188	548	736
Viviendas	112.871	36.477	8.777	45.254	54.622	12.995	67.617
En Manzanas	84.410	33.765	0	33.765	50.645	0	50.645
En Secciones	28.461	2.712	8.777	11.489	3.977	12.995	16.972

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Tabla V.10: Total de viviendas por región, según submuestra.

Región	Submuestra 1		Submuestra 2		Total	
	Viviendas	% en muestra	Viviendas	% en muestra	Viviendas	% en muestra
1	3.096	39,9	4.667	60,1	7.763	100,0
2	2.995	39,9	4.520	60,1	7.515	100,0
3	2.639	40,1	3.945	59,9	6.584	100,0
4	1.818	40,2	2.700	59,8	4.518	100,0
5	3.892	40,1	5.803	59,9	9.695	100,0
6	2.657	40,2	3.953	59,8	6.610	100,0
7	3.309	40,2	4.919	59,8	8.228	100,0
8	4.223	40,2	6.291	59,8	10.514	100,0
9	2.796	40,3	4.146	59,7	6.942	100,0
10	3.156	40,2	4.691	59,8	7.847	100,0
11	2.215	39,7	3.368	60,3	5.583	100,0
12	1.510	40,0	2.263	60,0	3.773	100,0
13	6.204	40,1	9.255	59,9	15.459	100,0
14	3.166	40,1	4.732	59,9	7.898	100,0
15	1.578	40,0	2.364	60,0	3.942	100,0
Total	45.254	40,1	67.617	59,9	112.871	100,0

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Tabla V.11: Distribución del total de viviendas, manzanas y secciones, por región, según submuestra. Encuesta Casen 2011.

Región	Submuestra 1			Submuestra 2		
	Vivienda	Manzana	Sección	Vivienda	Manzana	Sección
1	39,9%	38,4%	41,9%	60,1%	61,6%	58,1%
2	39,9%	39,2%	44,4%	60,1%	60,8%	55,6%
3	40,1%	39,2%	41,9%	59,9%	60,8%	58,1%
4	40,2%	41,4%	41,9%	59,8%	58,6%	58,1%
5	40,1%	40,7%	45,3%	59,9%	59,3%	54,7%
6	40,2%	39,7%	40,3%	59,8%	60,3%	59,7%
7	40,2%	39,8%	39,1%	59,8%	60,2%	60,9%
8	40,2%	40,3%	41,0%	59,8%	59,7%	59,0%
9	40,3%	38,7%	37,8%	59,7%	61,3%	62,2%
10	40,2%	41,1%	40,1%	59,8%	58,9%	59,9%
11	39,7%	35,2%	39,7%	60,3%	64,8%	60,3%
12	40,0%	40,3%	47,1%	60,0%	59,7%	52,9%
13	40,1%	39,4%	45,8%	59,9%	60,6%	54,2%
14	40,1%	36,3%	41,1%	59,9%	63,7%	58,9%
15	40,0%	40,3%	35,7%	60,0%	59,7%	64,3%
Total	40,1%	39,4%	40,8%	59,9%	60,6%	59,2%

Fuente: Elaborado por el Instituto nacional de Estadísticas.

5. Redistribución de Conglomerados entre las Submuestras 1 y 2

Uno de los objetivos del diseño 2011 ha sido implementar una estrategia de muestreo y recolección de datos que permita extender el periodo de trabajo de campo sin sacrificar comparabilidad con la serie histórica Casen. Para lograr este objetivo, se implementa el sistema de selección de la muestra 2011 en dos submuestras independientes, cada una compartiendo el mismo diseño de la muestra completa. Esta estrategia da origen a dos submuestras que permiten el desarrollo de estadísticas independientes a partir de cada una de ellas, así como también de la combinación de ambas.

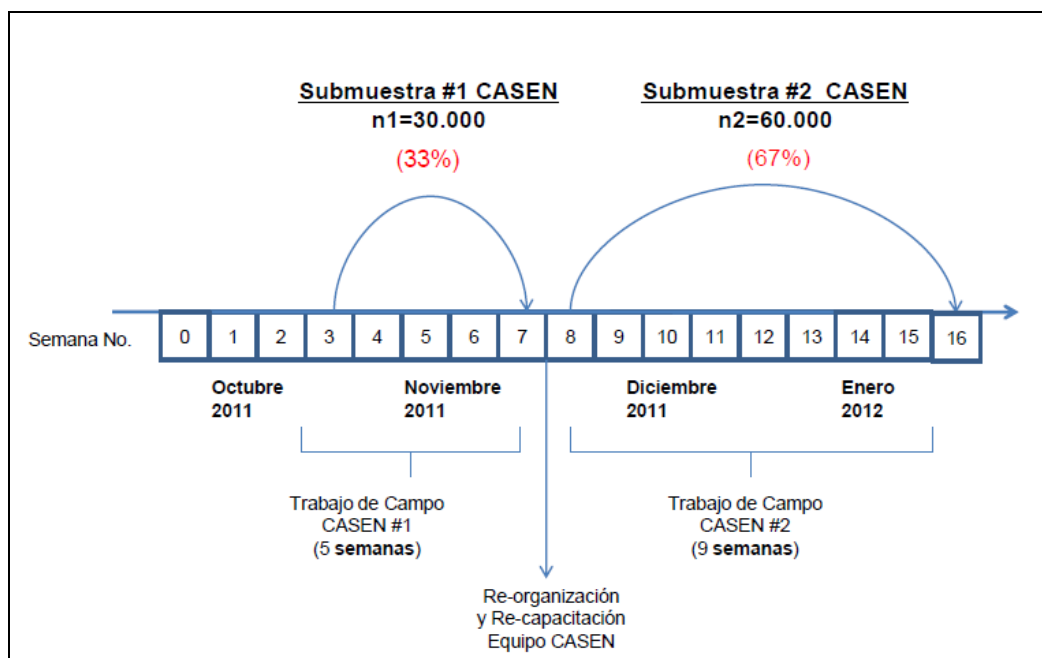
La partición de la muestra en las submuestras independientes 1 y 2 fue establecida inicialmente, en cuanto al número de conglomerados, en una proporción 40% y 60% respectivamente, tal como se visualiza en la Tabla V.11. Debido a retrasos inesperados en el proceso de licitación y desarrollo de los instrumentos de recolección de datos, el Ministerio determinó, en conjunto con el contratista Casen, la necesidad de aplazar en 2 semanas el inicio del encuestaje. Esta decisión fue tomada a pocas semanas del inicio programado del trabajo de campo. Para no afectar las fechas planeadas de aplicación de la submuestra 2, que corresponde a la muestra que se utilizaría para las estimaciones comparables con la serie histórica Casen, se optó por materializar el ajuste a través de la reducción del tamaño muestral y el periodo de recolección de la submuestra 1 y el consiguiente aumento en el tamaño de la submuestra 2.

La disminución del tamaño de la submuestra 1 se realizó a partir de la selección al azar del 15% de sus conglomerados. La selección de este 15% se hizo a través de un algoritmo de selección sistemático, con arranque aleatorio, sobre el listado de conglomerados previamente ordenados según región, estrato Casen (cruce de comuna y área urbana-rural) y el tamaño de viviendas total que poseen los distintos conglomerados. El ordenamiento, previo a la selección, asegura que se elija aproximadamente la misma fracción de muestreo en cada subdivisión de área geográfica y sean seleccionados conglomerados de todos los tamaños de acuerdo a la proporción de ellos en la submuestra 1.

Si bien es cierto la redistribución de la muestra fue un hecho inesperado, la detección del problema se realizó en forma oportuna lo cual permitió: (1) convenir una estrategia de ajuste con el consultor internacional; (2) implementar la estrategia de selección y posterior distribución del material en forma oportuna por parte del contratista. La Tabla V.12 y la Figura V.5 muestran la distribución del total de viviendas entre las submuestras 1 y 2. Las submuestras pasaron desde un 40% y 60%, antes del traslado a un 33% y 67% después del traslado.

Tabla V.12: Distribución de viviendas por región, según submuestra, antes y después del traslado de conglomerados.

REGION	Antes del traslado					Después del traslado				
	Sub-muestra 1	%	Sub-muestra 2	%	TOTAL	Sub-muestra 1	%	Sub-muestra 2	%	TOTAL
1	2.988	38,5	4.775	61,5	7.763	2.541	32,7	5.222	67,3	7.763
2	2.570	34,2	4.936	65,8	7.506	2.182	29,1	5.324	70,9	7.506
3	2.639	40,1	3.945	59,9	6.584	2.227	33,8	4.357	66,2	6.584
4	1.818	40,2	2.700	59,8	4.518	1.568	34,7	2.950	65,3	4.518
5	3.839	39,6	5.856	60,4	9.695	3.224	33,3	6.471	66,7	9.695
6	2.594	39,2	4.016	60,8	6.610	2.239	33,9	4.371	66,1	6.610
7	3.309	40,2	4.919	59,8	8.228	2.776	33,7	5.452	66,3	8.228
8	4.170	39,8	6.307	60,2	10.477	3.637	34,7	6.840	65,3	10.477
9	2.706	39,0	4.236	61,0	6.942	2.274	32,8	4.668	67,2	6.942
10	3.156	40,2	4.691	59,8	7.847	2.663	33,9	5.184	66,1	7.847
11	2.216	39,7	3.367	60,3	5.583	1.898	34,0	3.685	66,0	5.583
12	1.510	40,0	2.263	60,0	3.773	1.295	34,3	2.478	65,7	3.773
13	5.870	38,0	9.589	62,0	15.459	4.985	32,2	10.474	67,8	15.459
14	3.106	39,4	4.784	60,6	7.890	2.613	33,1	5.277	66,9	7.890
15	1.578	40,0	2.364	60,0	3.942	1.341	34,0	2.601	66,0	3.942
TOTAL	44.069	39,1	68.748	60,9	112.817	37.463	33,2	75.354	66,8	112.817

Figura V.5 Distribución del trabajo de campo para submuestras 1 y 2.

VI. Desarrollo de factores de Expansión

1. Visión General

En el caso de encuestas cuyas unidades de observación son seleccionadas mediante un muestreo aleatorio simple, cada unidad incluida en la muestra tiene igual probabilidad de participar, y por tanto cada una de ellas representa en la misma proporción a la totalidad de la población. Sin embargo, en encuestas donde el diseño muestral es complejo (estratificado y multi-etápico), la probabilidad de inclusión de cada unidad de observación es disímil según su procedencia, en cuanto al estrato y a la unidad de muestreo, por tanto el número de unidades que representan en la población es distinto.

Históricamente para las encuestas Casen, se han desarrollado dos factores de expansión³⁶ para cada persona entrevistada: un factor que expande a la proyección de población regional (*expr*) y otro que expande a la población comunal (*expc*)³⁷. Entre 1987 y 1994 los factores de expansión se determinaron como el cociente entre las proyecciones de población y la muestra lograda en cada comuna. Entre 1996 y 2009 los factores de expansión se estiman como el inverso de la probabilidad de selección de las viviendas ajustadas por la proyección de población en hogares particulares en cada comuna.

A partir de 2011, siguiendo las recomendaciones del experto internacional, se implementó una metodología de cálculo de los factores de expansión que involucra el desarrollo de cuatro ponderadores que vienen a componer los factores de expansión regional y comunal. El diagrama de flujo en la Figura VI.1 resume los pasos y los insumos involucrados en la metodología de cálculo 2011.

El primer ponderador, denominado Ponderador de Selección, se construye en base a la probabilidad de seleccionar a un conglomerado (*whi_1*) y a la vivienda a partir del conglomerado (*whij_2*). Intuitivamente indica a cuantas viviendas en la población representa la vivienda seleccionada en la muestra. Este se calcula, en términos generales, como el inverso de la probabilidad de selección de la vivienda, condicional a la selección del conglomerado (sección o manzana) que la contiene.

El segundo ponderador, denominado Ponderador Corregido por Elegibilidad (*whij_3*), 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.

El ajuste que sigue da origen al Ponderador Corregido por No Respuesta. El ajuste se aplica con el objetivo de reducir el sesgo que se pueda producir debido a la falla en conseguir la cooperación de los hogares seleccionados en la muestra. La técnica implementada se denomina "propensity score stratification", la cual fue aplicada para desarrollar grupos homogéneos (en relación a la probabilidad de responder la encuesta) al interior de los cuales se calcula un ajuste de razón

³⁶ En forma excepcional, en 2006 y 2009 se desarrolló un factor de expansión provincial (*expp*) que suma a la proyección de población en la provincia.

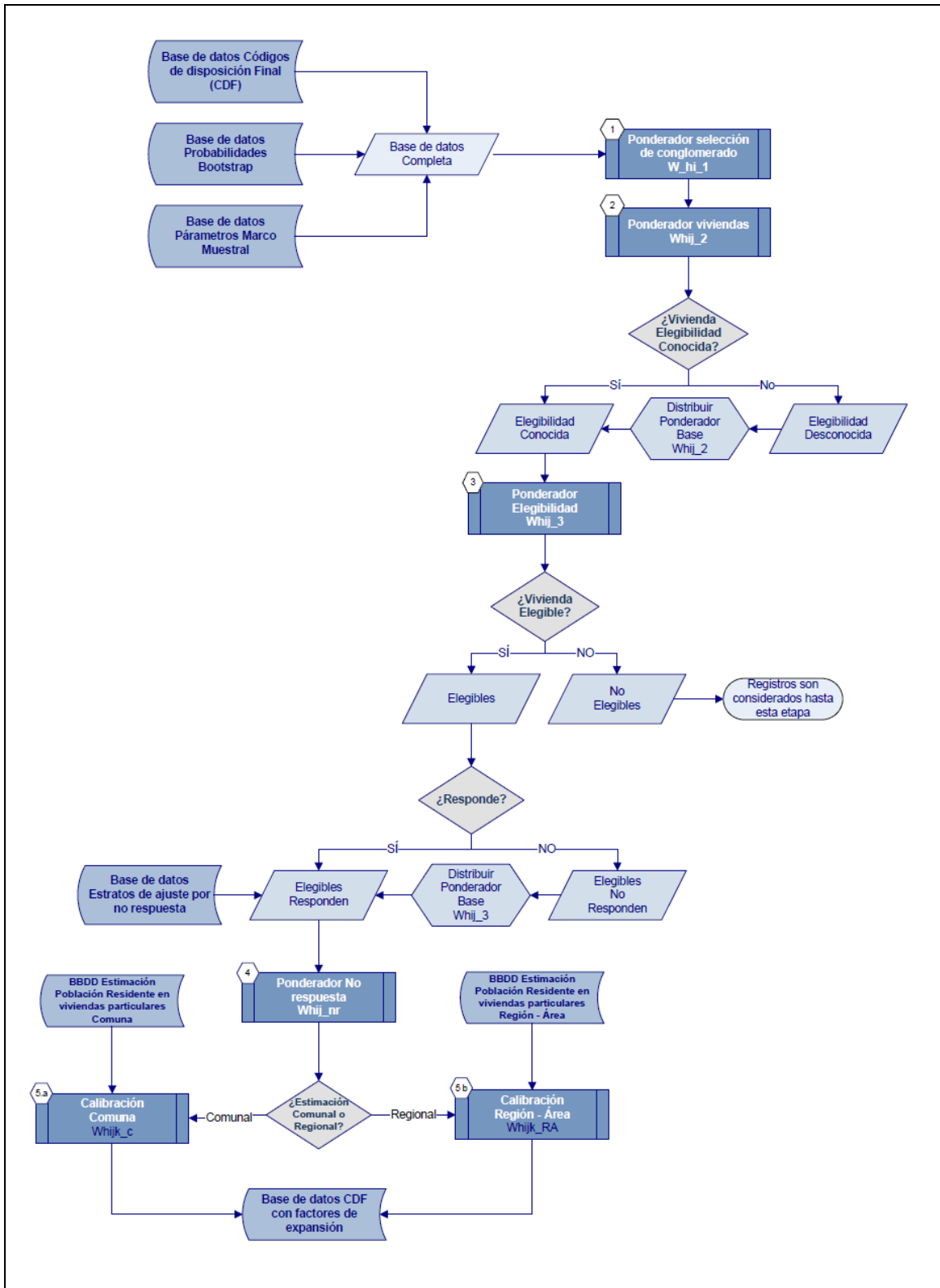
³⁷ La suma de los factores regionales es levemente mayor a la suma de los factores comunales ($\sum expr > \sum expc$), ya que por construcción el factor regional incluye la población en las comunas no incluidas en la muestra.

que permite corregir (en función del poder explicativo del modelo) el potencial efecto del sesgo de no respuesta.

El último ajuste da origen a los Ponderadores de Calibración Regional y Comunal. Estos ajustes buscan corregir por fallas en la cobertura de subpoblaciones de interés y hacer conformar las estimaciones poblacionales obtenidas a partir de la encuesta con stock poblacionales externos. A este ajuste se reconoce también como "corrección de población total". El ajuste "comunal", se computa como la razón entre el total de personas estimado para cada comuna a partir de la encuesta y el total de personas según las proyecciones de población del INE. El ajuste "regional", se computa la razón entre el total de personas estimado a partir de la encuesta y el total de la población según las proyecciones de población del INE en cada región-área.

La Figura VI.1 resume el proceso del cálculo de los factores de expansión para Casen 2011. Los capítulos que siguen describen los detalles del proceso de construcción de los factores de la muestra completa (n=86.854 hogares). Cada apartado hace referencia a las rutinas respectivas utilizadas para su implementación, así como también a los resultados necesarios para la validación y comprobación de los cálculos. Las variaciones implementadas para el cálculo de los factores de expansión de las submuestras 1 (n=27.770 hogares) y submuestra 2 (n=59.854 hogares) se describen en el Anexo.

Figura VI.1 Diagrama proceso de desarrollo de factores de expansión Casen 2011.



2. Ponderador de Selección

La Encuesta Casen 2011 posee un diseño probabilístico, estratificado según área geográfica y tamaño poblacional. La muestra fue seleccionada a partir de dos marcos muestrales, mutuamente excluyentes, que cubren la totalidad del territorio nacional. En el área rural corresponde al Marco de Secciones (MS2002) y en el área urbana corresponde al Marco de Manzanas (MM2008).

La muestra fue seleccionada en dos etapas en el área rural y tres etapas en el área urbana. En el área rural los conglomerados (unidades de muestreo), denominados secciones, fueron seleccionadas con probabilidad proporcional al tamaño, con respecto al total de viviendas particulares que estas contenían al Censo de Población y Vivienda realizado el año 2002, mientras que las viviendas, contenidas en su interior y que se encontraban elegibles al momento del empadronamiento, fueron seleccionadas con igual probabilidad.

Por su parte, en el área urbana los conglomerados, denominados manzanas, fueron agrupadas en 30 "grupos de tamaño" según la cantidad de viviendas que éstas contienen, basado en información proveniente del MM2008. La selección de grupos de tamaño y manzanas fue realizada mediante un método iterativo y sin remplazo, con probabilidad proporcional al tamaño, definido según el total de manzanas y viviendas que dicho grupo contiene. Al interior de cada manzana seleccionada se escogen viviendas con la misma probabilidad.

En ambos marcos muestrales la selección de conglomerados, manzanas o secciones, fue realizada de forma independiente en cada estrato, el cual es definido como la intersección entre comuna y área (urbano o rural). Para la encuesta Casen 2011 se seleccionaron unidades muestrales en 585 estratos de los 610 que posee el INE en sus Marcos de selección.

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

2.1. Probabilidades de selección de los Conglomerados

Las probabilidades de selección de los conglomerados están diferenciadas según el Marco de procedencia, ya que tanto las unidades de muestreo como los métodos de selección implementados en cada Marco son distintos.

Las unidades primarias de muestreo provenientes del Marco de Secciones (MS2002) fueron seleccionadas todas a la vez de forma sistemática y con probabilidades definidas previamente. El procedimiento consistente en usar la estrategia de selección del muestreo bietápico con probabilidad proporcional al tamaño y, por consiguiente, se utilizan formulas de texto comúnmente aceptadas para la estimación de las probabilidades de selección asociadas a la selección de las unidades primarias de muestreo (las secciones).

Las unidades de muestreo provenientes del Marco de Manzanas (MM2008) fueron seleccionadas mediante un método recursivo que consiste en seleccionar primero un "grupo de tamaño", para posteriormente en su interior seleccionar una manzana. La siguiente etapa consiste en recalcular las probabilidades de

selección de los n grupos de tamaño n , descontando el total de viviendas de la manzana seleccionada, para volver a realizar la selección de un "grupo de tamaño" y reiterar la selección en su interior de una manzana. El proceso se realiza sucesivamente hasta alcanzar el total predeterminado de viviendas en una comuna. El procedimiento empleado no tiene asociado una fórmula conocida explícita para el cálculo de las probabilidades de selección de las manzanas, por lo tanto, éstas probabilidades deben ser estimadas a través de un método numérico.

A continuación se describe la metodología de estimación de las probabilidades de selección de los conglomerados (manzanas y secciones) bajo los dos marcos muestrales.

2.1.1. Probabilidades de selección de Secciones (Marco 2002)

Las secciones, unidades primarias de muestreo del Marco 2002, fueron seleccionadas de forma independiente, por comuna y área, mediante un método de selección sistemático y con probabilidad proporcional al tamaño, en cuanto al total de viviendas particulares que ellas contienen según Censo de Población y Vivienda del año 2002.

Previo a la selección de la muestra 2011, se excluyeron del marco las secciones pertenecientes a muestras asociadas a otras encuestas de hogares que el INE seleccionó para ser entrevistadas en el mismo periodo que el trabajo de campo Casen (es decir, entre octubre 2011 y enero 2012). En particular, secciones incluidas en la Nueva Encuesta Nacional de Empleo (NENE) 2011 y Encuesta Nacional de Calidad de Vida y Salud (ENCAVI) 2011³⁸. La Tabla VI.1 presenta, para cada región del país, el total de secciones y viviendas en el marco de secciones (MS2002) y el total de secciones y viviendas excluidas del marco para la selección de la muestra Casen 2011.

Luego, la probabilidad de seleccionar una unidad primaria, sección, puede ser calculada como,

$$P_h(i) = S_h \cdot \frac{M_{hi_2000}}{U_h} \quad (1)$$

donde:

i : Representa el índice de la unidad de muestreo, correspondiente a una sección.

h : Representa el índice del estrato, correspondiente a la intersección de comuna y área urbano-rural.

S_h : Es el número de secciones seleccionadas en el estrato h .

M_{hi_2000} : Corresponde al total de viviendas en la sección i del estrato h , según el MS2002.

38 Al momento de selección de la muestra Casen se tenía planeado el levantamiento de la encuesta Encuesta Nacional de Calidad de Vida y salud 2011, sin embargo, esta última no fue levantada finalmente.

U_h : Es el total de viviendas en el estrato h según el MS2002 previamente "ajustado" para la selección de Casen 2011³⁹.

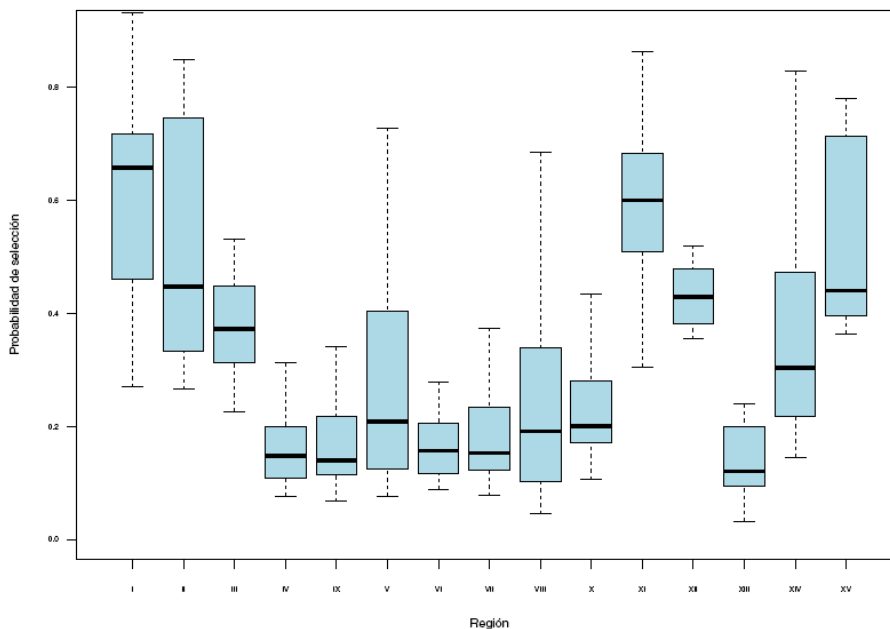
El Figura VI.2 presenta la distribución de la probabilidad de selección de las secciones en cada región del país. En general, la mediana de las probabilidades de selección es baja (línea negra al medio de cada caja) para las regiones en el centro del país (en torno al 20%) y relativamente más alta para las regiones en los extremos del país, las cuales tienen menos población. La región con menor variabilidad es la región de Magallanes, donde el 50% de sus secciones tienen una probabilidad superior o igual al 45% de ser seleccionadas. En la Región Metropolitana se concentran las secciones con menor probabilidad de selección lo que se explica, en parte, porque el total de viviendas que concentran las secciones es pequeño en comparación al total de viviendas que posee el estrato de pertenencia. Mientras que en aquellas áreas de menor concentración de la población (principalmente rural), como las regiones I, II, XI, XII y XV, más del 50% de sus secciones poseen probabilidad de selección superior al 40%, que en parte puede ser explicado porque el número de viviendas que concentran las secciones es alto en comparación al total de viviendas que posee el estrato.

Tabla VI.1: Total de unidades contenidas en el MM2002 y total de unidades excluidas

Región	Unidades MM2002		Unidades MM2002 excluidas	
	N° Secciones	N° Viviendas	N° Secciones	N° Viviendas
Tarapacá	63	3.268	12	672
Antofagasta	46	2.347	5	265
Atacama	103	5.192	15	777
Coquimbo	501	39.412	80	6.349
Valparaíso	538	38.768	43	3.141
Libertador General Bernardo OHiggins	883	67.371	73	5.346
Maule	1.192	96.332	84	6.826
Biobío	1.293	100.123	51	3.823
Araucanía	1.096	89.333	56	4.637
Los Lagos	791	64.581	95	7.699
Aysén del General Carlos Ibañez del Campo	129	6.951	18	988
Magallanes y La Antártica Chilena	47	2.630	6	343
Metropolitana de Santiago	441	45.434	27	2.697
Los Ríos	395	31.685	38	3.108
Arica y Parinacota	56	2.913	23	1.172
TOTAL	7.574	596.340	626	47.843

39 El total de viviendas en el universo (U_h) excluye las viviendas en los conglomerados seleccionados para NENE y ENCAVI (encuestas cuyo periodo de levantamiento coincidía con el de la Encuesta Casen 2011).

Figura VI.2: Distribución de las probabilidades de selección de las Secciones, por Región, Casen 2011.



2.1.2. Probabilidades de selección de Manzanas (Marco 2008)

Previo a la selección de las manzanas, se excluyeron del marco todas aquellas manzanas que habían sido seleccionadas en la muestra de la Encuesta Nacional de Empleo cuyo periodo de trabajo de campo fuese el mismo que el de la encuesta Casen (es decir, entre octubre 2011 y enero 2012). Además se excluyeron aquellas manzanas que tienen siete o menos viviendas. En total, se excluyó un 12,8% de las manzanas y un 3,8% de las viviendas del marco.

La Tabla VI.2 presenta, para cada región del país, el total de manzanas y viviendas en el marco de manzanas (MM2008) y el total de manzanas y viviendas excluidas del marco para la selección de la muestra Casen 2011.

El método de selección de las manzanas, a partir del marco MM2008, corresponde a un método complejo donde las probabilidades de selección no permiten el desarrollo de una fórmula explícita de cálculo.

En concreto, las manzanas fueron seleccionadas mediante un método recursivo que consiste en seleccionar, en una primera etapa, un "grupo de tamaño" y, en una segunda etapa, seleccionar una manzana al interior del "grupo de tamaño" previamente seleccionado. Luego se recalculan las probabilidades de selección del grupo de tamaño, descontando el total de viviendas en la manzana seleccionada, así como también reduciendo el total de manzanas que contiene el grupo desde el cual fue extraída; así posteriormente realizar nuevamente la selección de un grupo de tamaño y reiterar la selección en su interior de una manzana. El proceso se realiza sucesivamente hasta acumular el total de viviendas predeterminadas en el cálculo del tamaño muestral.

Tabla VI.2: Total de manzanas y viviendas MM2008, completo y con exclusiones

Región	Undidades MM2008		Undidades MM2008 excluidas	
	N° Manzanas	N° Viviendas	N° Manzanas	N° Viviendas
Tarapacá	2.198	70.162	283	3.947
Antofagasta	5.614	131.946	843	6.057
Atacama	4.000	74.686	854	5.438
Coquimbo	6.817	152.956	1.037	7.161
Valparaíso	19.259	507.778	3.517	27.963
Libertador General Bernardo OHiggins	5.881	159.390	927	6.812
Maule	6.703	185.055	772	6.722
Biobío	16.492	445.810	2.374	22.339
Araucanía	7.158	170.383	1.019	7.468
Los Lagos	5.221	136.763	748	6.257
Aysén del General Carlos Ibañez del Campo	990	19.782	229	2.861
Magallanes y La Antártica Chilena	2.038	42.374	259	2.302
Metropolitana de Santiago	46.027	1.784.022	3.439	39.582
Los Ríos	2.547	66.790	346	3.130
Arica y Parinacota	2.415	52.865	453	4.390
TOTAL	133.360	4.000.762	17.100	152.429

Dado que las probabilidades de selección no tienen fórmula explícita, es que en encuestas anteriores que utilizan el mismo diseño muestral que la Encuesta Casen 2011, el INE ha optado por aproximar las probabilidades de selección de las manzanas bajo el supuesto de que corresponden a las calculadas para un diseño probabilístico, estratificado, bi-etápico, y donde sus unidades primarias son extraídas con probabilidad proporcional al tamaño, medido en cuanto al total de viviendas particulares que estas contienen. Es decir, bajo el mismo procedimiento con el cual se seleccionan las secciones en el marco MS2002.

A requerimiento del experto internacional, el Ministerio de Desarrollo Social, solicitó al INE desarrollar un estudio de simulaciones que pretendía verificar, para la muestra Casen 2011, la pertinencia de aproximar la estimación de las probabilidades de selección de las manzanas utilizando las fórmulas del muestreo bietápico proporcional al tamaño. Para estos efectos, se realizaron 2.000 simulaciones de selección de manzanas, en las 223 comunas pertenecientes al MM2008, con la finalidad de guardar las probabilidades de selección empírica. Posteriormente, mediante dos métodos distintos se estimó la probabilidad de selección de las manzanas; el primer método consistió en una aproximación lineal, fundamentada en las probabilidades empíricas obtenidas de las simulaciones; el segundo método consiste en obtener la proporción de muestras en las que fue seleccionada una manzana. Este método se denomina probabilidad Bootstrap (Valliant *et al.*, 2008; Valliant & Rust, 2010); Sitter, 1992).

Luego, para comparar la calidad predictiva de ambos métodos se obtuvo una estimación del total de viviendas por comuna, mediante la utilización de factores de expansión en cada una de las 2.000 replicas. Los factores de expansión fueron contruidos con los inversos de las probabilidades de selección empíricas y bootstrap. Posteriormente, se ponderó el total de viviendas por manzana, reportadas en el MM2008, con su respectivo factor de expansión. Finalmente, se elaboraron gráficos de densidad para comparar el comportamiento de ambas expansiones (inverso probabilidad empírica e inverso probabilidad bootstrap), respecto al total de viviendas reportadas en el MM2008.

Como resultado de este estudio, se verificó que las estimaciones desarrolladas a partir de las probabilidades estimadas con el método bootstrap presentaban, para la mayoría de las comunas en la muestra, mejores propiedades estadísticas (menor error cuadrático medio) que las estimaciones a partir del método que asume muestreo bietápico. En consecuencia, se desestimó la opción de utilizar la fórmula del muestreo bietápico para selecciones proporcionales al tamaño y se procedió a estimar las probabilidades de selección de las manzanas mediante un procedimiento de re-muestreo no paramétrico conocido como Bootstrap.

Para la implementación del cálculo de las probabilidades de selección de las manzanas, se desarrollaron inicialmente 2.000 simulaciones de muestras de manzanas en 223 estratos urbanos de la muestra. Posteriormente, se agregaron 2.000 simulaciones en 82 estratos cuyos resultados fueron considerablemente más volátiles para producir estimaciones confiables de la probabilidad de selección de las manzanas.

Inicialmente, el proceso de simulación consideró el remuestreo de manzanas asociadas a una muestra de 36.600 viviendas aproximadamente. Para el objetivo inicial del estudio, este tamaño se consideró suficiente. Dado que la muestra Casen tiene distintos tamaños de muestra para la submuestra 1, submuestra 2 y muestra completa, las probabilidades de selección estimadas a partir del método bootstrap (π_{hi}) deben ser "ajustadas" para expandir en forma consistente con los tamaños de muestrales de cada submuestra. De esta forma, la probabilidad de que una manzana pertenezca a la muestra Casen 2011, puede ser estimada como:

$$P_h(i) = (f_1 + f_2) \cdot \pi_{hi} \quad (2)$$

donde

f_1 : Corresponde a la razón entre el total de viviendas asignadas a la submuestra 1 en el área urbana (27.766) vs. tamaño de la muestra con el cual se realizaron las simulaciones (32.616).

f_2 : Corresponde a la razón entre el total de viviendas asignadas a la submuestra 2 en el área urbana (56.644) vs. tamaño de la muestra con el cual se realizaron las simulaciones (32.616).

π_{hi} : corresponde a la probabilidad de seleccionar una manzana para una muestra de 32.616 viviendas, estimada mediante bootstrap.

La figura VI.3 presenta la distribución de la probabilidad de selección de las manzanas en cada región del país. En general, la mediana de las probabilidades de selección es baja (línea negra al medio de cada caja) para las regiones en el centro del país (en torno al 5%) y relativamente más alta para las regiones en los extremos del país, las cuales tienen menos población urbana. Se puede apreciar que las probabilidades estimadas son relativamente homogéneas al interior de cada región (cajas pequeñas), sin embargo, en la Región de Aysén se observa gran heterogeneidad, y que el 50% de las manzanas poseen probabilidad de selección superior a 0.5. Esto último se puede explicar por la gran cantidad de manzanas seleccionadas en la muestra en comparación al total de unidades existentes en los estratos de la región.

Figura VI. 3: Estimación de la probabilidad de selección de las manzanas mediante Bootstap, Casen 2011.

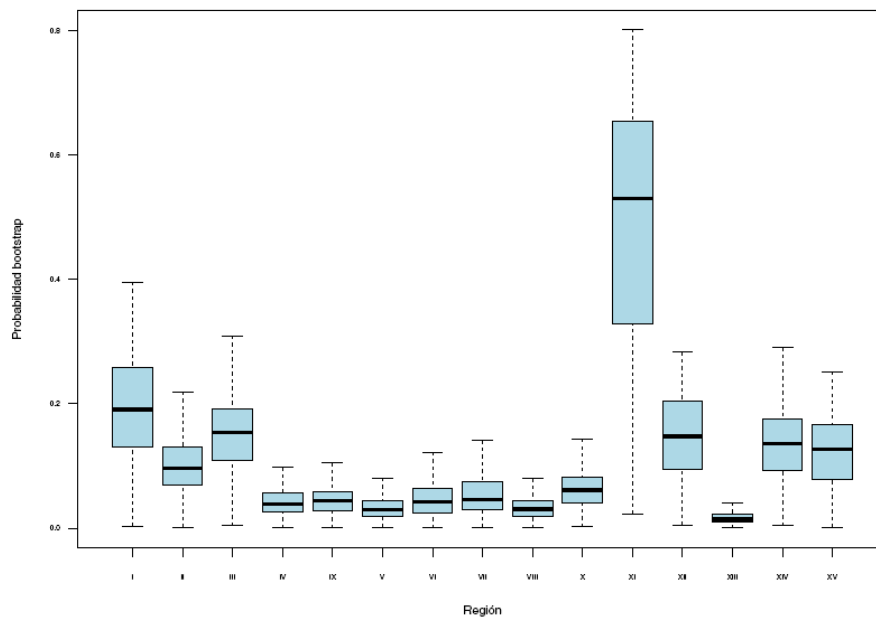


Tabla VI.3: Estadísticas descriptivas estimación de probabilidades de las manzanas mediante método bootstrap

Región	Prob. de selección Manzanas Bootstrap											
	Recuento	Mínimo	Percentil 05	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Moda	Media
Tarapacá	733	0,00	0,05	0,09	0,13	0,19	0,26	0,30	0,31	0,40	0,13	0,19
Antofagasta	839	0,00	0,02	0,04	0,07	0,10	0,13	0,16	0,18	0,22	0,08	0,10
Atacama	940	0,00	0,05	0,07	0,11	0,15	0,19	0,22	0,24	0,37	0,13	0,15
Coquimbo	410	0,00	0,01	0,02	0,03	0,04	0,06	0,08	0,09	0,15	0,03	0,05
Valparaíso	804	0,00	0,01	0,01	0,02	0,03	0,04	0,06	0,07	0,20	0,02	0,03
Libertador General Bernardo OHiggins	430	0,00	0,01	0,02	0,02	0,04	0,06	0,10	0,13	0,17	0,02	0,05
Maule	462	0,00	0,01	0,02	0,03	0,05	0,07	0,11	0,12	0,16	0,04	0,05
Biobio	729	0,00	0,01	0,01	0,02	0,03	0,04	0,06	0,07	0,16	0,03	0,03
Araucanía	448	0,00	0,01	0,02	0,03	0,04	0,06	0,08	0,09	0,14	0,05	0,05
Los Lagos	499	0,00	0,02	0,02	0,04	0,06	0,08	0,11	0,12	0,20	0,07	0,06
Aysén del General Carlos Ibañez del Campo	390	0,02	0,16	0,17	0,33	0,53	0,65	0,70	0,73	0,80	0,55	0,49
Magallanes y La Antártica Chilena	496	0,01	0,04	0,07	0,10	0,15	0,20	0,24	0,25	0,28	0,13	0,15
Metropolitana de Santiago	1.090	0,00	0,00	0,01	0,01	0,02	0,02	0,04	0,06	0,19	0,02	0,02
Los Ríos	639	0,00	0,02	0,05	0,09	0,14	0,18	0,22	0,23	0,29	0,09	0,13
Arica y Parinacota	504	0,00	0,03	0,04	0,08	0,13	0,17	0,19	0,21	0,25	0,12	0,13
Total	9.413	0,00	0,01	0,01	0,03	0,07	0,14	0,21	0,27	0,80	0,02	0,10

2.2. Inverso de las Probabilidades de selección

Una vez definidas las probabilidades de selección se calcula el Ponderador de Selección del Conglomerado como el inverso de las probabilidades de selección de manzanas y secciones respectivamente,

$$w_{hi} = \frac{1}{P_h(i)} \quad (3)$$

$$w_{hi} = \begin{cases} \frac{U_h}{S_h \cdot M_{hi,2000}}, & \text{Si la unidad de muestreo } i \text{ es una sección} \\ \frac{1}{(f_1 + f_2) \cdot \pi_{hi}}, & \text{Si la unidad de muestreo } i \text{ es una manzana} \end{cases} \quad (4)$$

Para cada manzana en la muestra, por ejemplo, su ponderador de selección puede ser interpretado como "el total de manzanas en la población que ella representa". Del mismo modo, la suma de los ponderadores de selección de todas las manzanas en la muestra puede ser interpretada como una estimación del total de manzanas existentes en la población.

Para realizar dichas estimaciones se debe utilizar la base de datos a nivel de manzanas y secciones. Para obtener la estimación del total de manzanas y secciones, basta sumar el Ponderador de selección de todas las unidades en la muestra. Mientras que para estimar el total de viviendas de la población, se debe ponderar previamente el inverso de la probabilidad de selección (w_{hi}) por el total de viviendas por conglomerado según información del Marco de muestreo ($M_{hi,2000}$). Al realizar este ejercicio se obtiene una estimación de 135.738 conglomerados en el total del país (secciones y manzanas). En la Tabla VI.4 se pueden observar las estadísticas descriptivas y estimación del total de manzanas y secciones por región, según el marco de muestreo.

Tabla VI.4: Estadísticas descriptivas del Ponderador de Selección de Conglomerados, según Marco

Marco	Región	Factor Base FULL												
		Recuento	Mínimo	Percentil 05	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Media	Moda	Suma
Manzanas	Tarapacá	733	1,0	1,2	1,3	1,5	2,0	3,0	4,1	8,6	110,4	3,0	3,0	2.200
	Antofagasta	839	1,8	2,2	2,4	2,9	4,0	5,6	10,0	16,8	257,6	6,2	4,9	5.182
	Atacama	940	1,0	1,6	1,8	2,0	2,5	3,6	5,9	7,9	85,9	3,7	2,3	3.462
	Coquimbo	410	2,6	4,2	5,0	6,7	9,8	14,9	22,1	35,1	772,8	15,8	14,9	6.474
	Valparaíso	804	1,9	5,7	7,0	8,9	12,9	20,9	35,1	55,2	257,6	18,8	11,7	15.140
	Libertador General Bernardo OHiggins	430	2,3	3,1	3,7	6,1	9,1	15,8	25,2	35,1	772,8	18,3	23,4	7.855
	Maule	462	2,4	3,2	3,6	5,2	8,3	12,9	22,1	28,6	257,6	12,1	10,4	5.598
	Biobio	729	2,4	5,3	6,5	8,8	12,5	20,3	35,1	55,2	772,8	21,7	12,9	15.807
	Araucanía	448	2,7	4,5	5,1	6,5	8,7	13,4	20,9	35,7	772,8	16,2	8,3	7.241
	Los Lagos	499	1,9	3,1	3,6	4,7	6,3	9,6	16,9	23,4	154,6	9,4	5,9	4.713
	Aysén del General Carlos Ibañez del Campo	390	0,5	0,5	0,5	0,6	0,7	1,2	2,2	2,4	16,4	1,2	0,7	478
	Magallanes y La Antártica Chilena	496	1,4	1,6	1,6	1,9	2,6	4,1	5,2	10,4	77,3	3,8	2,9	1.894
	Metropolitana de Santiago	1.090	2,1	6,2	9,5	17,6	25,8	40,7	77,3	110,4	772,8	43,3	24,9	47.183
	Los Ríos	639	1,3	1,7	1,8	2,2	2,8	4,1	7,5	15,6	92,7	5,1	4,2	3.242
	Arica y Parinacota	504	1,5	1,9	2,0	2,3	3,0	4,9	9,8	11,9	257,6	4,8	3,2	2.440
Secciones	Tarapacá	27	1,1	1,1	1,2	1,4	1,5	2,4	2,9	3,2	3,7	1,8	1,2	49
	Antofagasta	18	1,2	1,2	1,2	1,3	2,2	3,0	3,3	3,8	3,8	2,2	2,4	40
	Atacama	30	1,9	1,9	2,0	2,2	2,7	3,2	3,9	4,4	4,4	2,8	2,8	85
	Coquimbo	61	1,2	1,4	1,5	5,0	6,8	9,2	11,3	12,0	13,0	6,9	5,9	421
	Valparaíso	86	1,4	1,5	1,7	2,5	4,8	8,1	10,5	10,9	13,0	5,5	1,5	470
	Libertador General Bernardo OHiggins	124	1,2	1,5	2,0	4,9	6,4	8,6	9,8	10,2	11,3	6,3	1,6	783
	Maule	174	1,1	1,5	2,3	4,3	6,5	8,2	10,0	10,6	12,9	6,3	2,1	1.102
	Biobio	182	1,0	1,4	1,5	2,9	5,2	9,8	12,2	16,0	21,9	6,6	11,1	1.198
	Araucanía	148	1,8	2,4	3,1	4,6	7,2	8,7	11,1	12,4	14,8	7,0	2,9	1.032
	Los Lagos	142	1,0	2,0	2,6	3,6	5,0	5,8	6,8	7,6	9,4	4,8	6,8	681
	Aysén del General Carlos Ibañez del Campo	58	1,2	1,2	1,3	1,5	1,7	2,0	2,8	2,9	3,3	1,8	1,7	106
	Magallanes y La Antártica Chilena	17	1,9	1,9	2,0	2,1	2,3	2,6	2,8	2,8	2,8	2,4	2,8	40
	Metropolitana de Santiago	48	1,7	2,1	2,5	5,0	8,3	10,5	13,3	26,3	31,0	8,9	4,9	427
	Los Ríos	105	1,0	1,3	1,4	2,1	3,3	4,6	5,6	6,0	6,9	3,4	1,6	360
	Arica y Parinacota	16	1,3	1,3	1,3	1,4	2,3	2,5	2,7	2,7	2,7	2,1	1,3	33
Total País		10.649	0,5	1,4	1,8	2,7	5,6	11,9	24,9	38,6	772,8	23,4	12,7	135.737,8

Las figuras VI.4 y VI.5 presentan la distribución de los Ponderadores de Selección de los Conglomerados Rurales (secciones) y Urbanos (manzanas) respectivamente, en cada región del país. En general, los ponderadores de selección de las secciones son más pequeños que los ponderadores de las manzanas. La mayor homogeneidad de los ponderadores de las secciones (que las manzanas) se explica por la diferencia en la composición de las unidades, pues las secciones fueron diseñadas para ser homogéneas en cuanto al total de viviendas, mientras que las manzanas son unidades definidas por límites de calles y por lo tanto el total de viviendas que poseen oscila entre 8 y más de 1.000 unidades.

La figura VI.4 muestra que el comportamiento de los Ponderadores de selección de las secciones es homogéneo en casi todas las regiones del país, lo que se debe principalmente a que las secciones poseen tamaños, en cuanto al total de viviendas, muy parecidos entre sí al interior de cada región. Sin embargo, aquellas regiones donde su población rural es pequeña, se observa que en promedio el inverso de la probabilidad no es superior a las 10 unidades, esto debido a que representan una población pequeña. El valor extremo máximo observado se encuentra en la Región Metropolitana, donde una sección representa a 31 unidades (ver Tabla VIII.5 en el anexo).

En la figura VI.5 se observa que en la Región Metropolitana el Ponderador de selección de las manzanas tiene mayor variabilidad, pues sus manzanas representan entre 2 y 386 unidades en la población. Por otra parte, en consistencia con las distribuciones en la figura VI.3 (probabilidades de selección), se observa que las manzanas de la Región de Aysén (XI) poseen pequeños ponderadores (inversos de la probabilidad de selección) y corresponden a los más homogéneos del país.

La Tabla VI.5 muestra el total de manzanas y secciones en cada marco (totales poblacionales) y las estimaciones desarrolladas a partir del ponderador de manzanas y secciones respectivamente (estimaciones muestrales). Se observa una subestimación del total de manzanas y secciones en todas las regiones del país. Esto se explica porque, previo a la selección de la muestra Casen 2011, se excluyeron unidades de muestreo de ambos marcos muestrales. Además, el marco urbano excluye del proceso de selección a las manzanas con menos de 8 viviendas. Por esta razón, es necesario ajustar el ponderador de selección para corregir las exclusiones deliberadas de conglomerados implementadas al momento de la selección de la muestra. En la sección que sigue se describe el ajuste implementado.

Figura VI.4: Distribución del Ponderador de Selección de Conglomerados Rurales (Secciones)

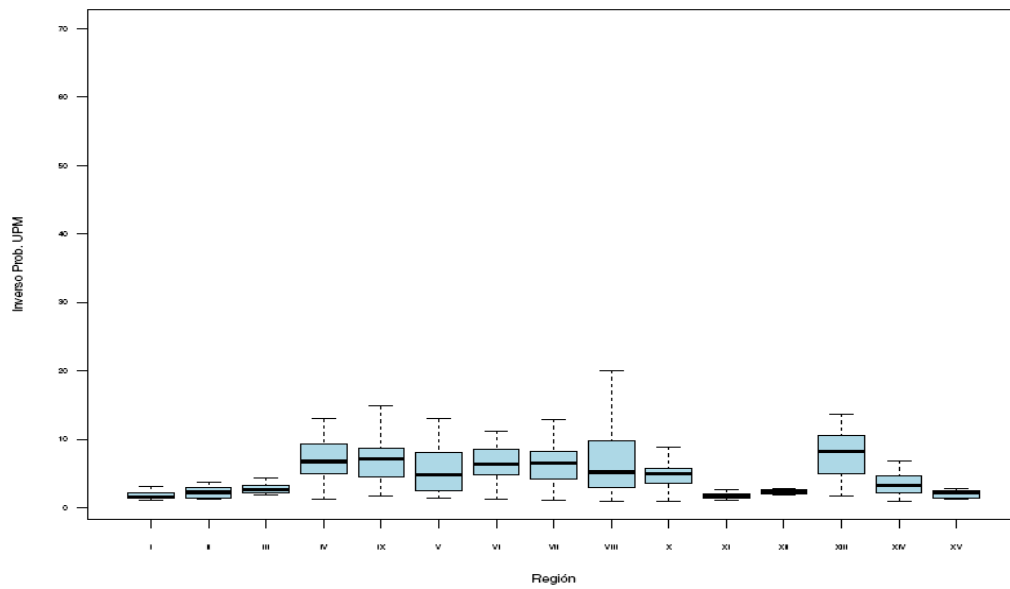


Figura VI.5: Distribución del Ponderador de Selección de Conglomerados Urbanos (Manzanas)

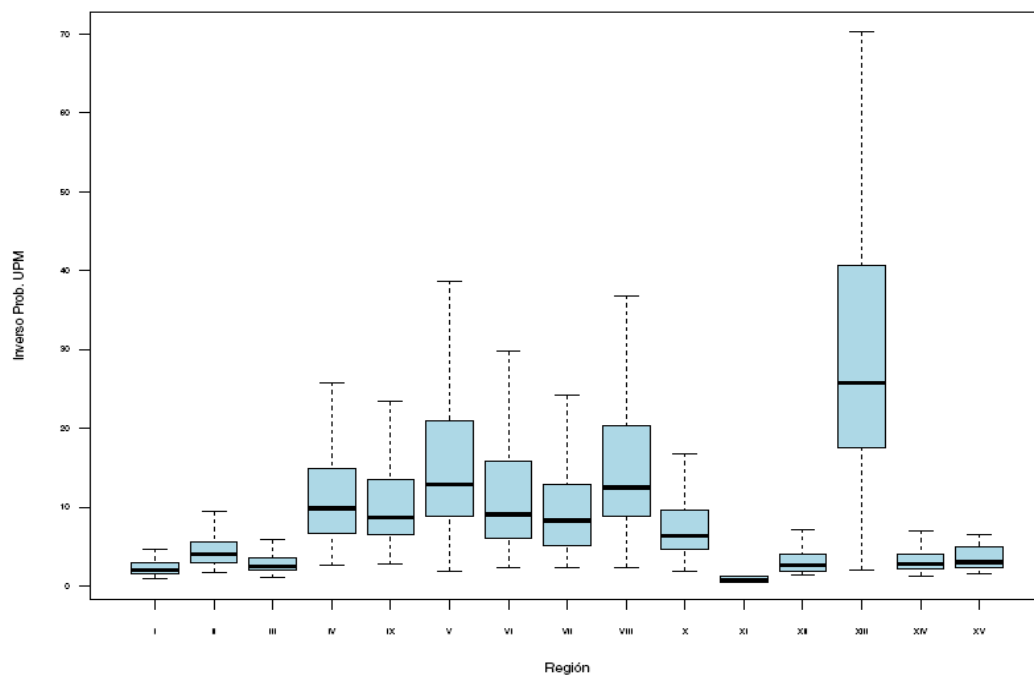


Tabla VI.5: Total de Manzanas y secciones estimadas con la muestra y según marco muestral

Región	Marco		Estimación Muestra	
	N° Secciones	N° Manzanas	N° Secciones	N° Manzanas
Tarapacá	63	2.198	49	2.200
Antofagasta	46	5.614	40	5.182
Atacama	103	4.000	85	3.462
Coquimbo	501	6.817	421	6.474
Valparaíso	538	19.259	470	15.140
Libertador General Bernardo OHiggins	883	5.881	783	7.855
Maule	1.192	6.703	1.102	5.598
Biobío	1.293	16.492	1.198	15.807
Araucanía	1.096	7.158	1.032	7.241
Los Lagos	791	5.221	681	4.713
Aysén del General Carlos Ibañez del Campo	129	990	106	478
Magallanes y La Antártica Chilena	47	2.038	40	1.894
Metropolitana de Santiago	441	46.027	427	47.183
Los Ríos	395	2.547	360	3.242
Arica y Parinacota	56	2.415	33	2.440
TOTAL	7.574	133.360	6.827	128.910

2.3. Ajuste por omisión de Conglomerados

Con la finalidad de reducir problemas operativos, el MM2008, excluye todas aquellas manzanas que contienen 7 o menos viviendas, así como también todas aquellas manzanas previamente seleccionadas para otra encuesta con periodo de recolección de datos coincidente con el de la Encuesta Casen 2011. El MS2002 por su parte, excluye solamente aquellas secciones que habían sido seleccionadas en otra encuesta y cuyo trabajo de campo era el mismo que el de la Encuesta Casen 2011. El total de manzanas y secciones excluidas desde ambos marcos muestrales equivale a un total de 200.272 viviendas, lo que corresponde a un 4,4% de la población de viviendas pertenecientes a los estratos Casen.

Tabla VI.6: Total de viviendas según MM, con y sin exclusión de conglomerados.

Total de viviendas MM	Marco de Muestreo		Total Viviendas
	MM 2008	MS 2002	
- Exclusión unidades*	3.848.333	548.497	4.396.830
- Todas las unidades*	4.000.762	596.340	4.597.102

* Total Estratos CASEN

En este contexto, utilizar los Ponderadores de Selección de Conglomerados para estimar totales asociados a fenómenos de interés implica subestimar dichos totales. Para corregir este problema, se realizó un ajuste que permite corregir las omisiones explícitas en los Ponderadores de Selección de Conglomerados. El ajuste consiste en abultar el Ponderador de Selección de Conglomerados según la cantidad de viviendas excluidas por estrato, con la finalidad que el total de

viviendas estimadas con este nuevo ponderador sea consistente con el total de viviendas de la población (4.597.102 a nivel nacional).

Para realizar el ajuste se obtuvo la razón entre el total de viviendas poblacional, por estrato, y el total de viviendas estimadas con el Ponderador de Selección (de manzanas y secciones respectivamente), es decir,

$$\hat{R}_h = \frac{M_{h_2000}}{\sum_{i \in \Omega_h} w_{hi} \cdot M_{hi_2000}}, \quad (4)$$

donde,

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

M_{hi_2000} : Corresponde al total de viviendas en el conglomerado i del estrato h, según el Marco de secciones o Marco de manzanas.

M_{h_2000} : Es el total de viviendas en el estrato h según el marco de muestreo correspondiente.

De este modo, el Ponderador de Selección de Conglomerados Corregido por Omisión (en adelante, simplemente "Ponderador de Selección de Conglomerado Corregido") viene dado por:

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

La forma de cálculo se ilustra mediante el siguiente ejemplo: considere el estrato "San Pedro de la Paz - urbano", que en la población posee 32.644 viviendas. Sin embargo, al estimar el total de viviendas con el Ponderador de Selección de Conglomerado en la muestra se obtiene un total de 26.635 unidades, por lo tanto la proporción de unidades que debe aumentar cada Ponderador de Selección de Conglomerado es 1,23 (32.644/26.635) unidades. Luego el ponderador ajustado al total de viviendas de la manzana "A", perteneciente al estrato en cuestión está dado por $13,8 \cdot 1,23 = 16,91$. En consecuencia, el conglomerado "A" ahora representa a 16,91 manzanas de la población de conglomerados de San Pedro de la Paz.

Las figuras VI.6 y VI.7 muestran la distribución del Ponderador de Selección de Conglomerados Corregido por Omisión de Secciones y Manzanas respectivamente. En la figura VI.6 se puede observar que las distribuciones tienen poca variabilidad en el marco de secciones, con valores medios del ponderador que oscilan entre 2,3 y 30 unidades. En contraste, el ponderador presenta gran variabilidad en el marco de manzanas, donde destaca la Región Metropolitana. Este comportamiento desigual entre los ponderadores asociados a secciones y manzanas refleja nuevamente la heterogeneidad en la construcción de estas unidades⁴⁰.

40 Ver Manual Conceptual y Metodológico diseño muestral Encuesta Nacional de Empleo, Abril 2010 página 28.
http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/pdf/031110/manual_metodologico031110.pdf

En comparación los ponderadores “sin ajustar” (figuras VI.4 y VI.5) se aprecia que existe un aumento en dispersión de la distribución, el que se puede apreciar claramente en la Región Metropolitana.

Las Tablas VI.7 y VI.8 muestran las estimaciones de conglomerados y viviendas a partir de los Ponderadores de Selección de Conglomerados Corregidos. Con el ponderador corregido se estimó un total de 7.448 secciones (en vez de las 6.827 estimadas sin corregir, Tabla VI.5) y 138.659 manzanas (en vez de las 128.910 estimadas sin corregir, Tabla VI.5). En total, la estimación de conglomerados aumentó en 10.369 unidades (en comparación a la estimación sin corregir). Queda en evidencia, por lo tanto, que de no realizar dicho ajuste se sub-estima el total de manzanas, secciones y viviendas, lo que se puede apreciar al contrastar las estimaciones en el Tabla VI.5 con aquellas en las Tablas VI.7 y VI.8.

Figura VI.6: Distribución del Ponderador de Selección de Conglomerado Corregidas por Omisión de Secciones.

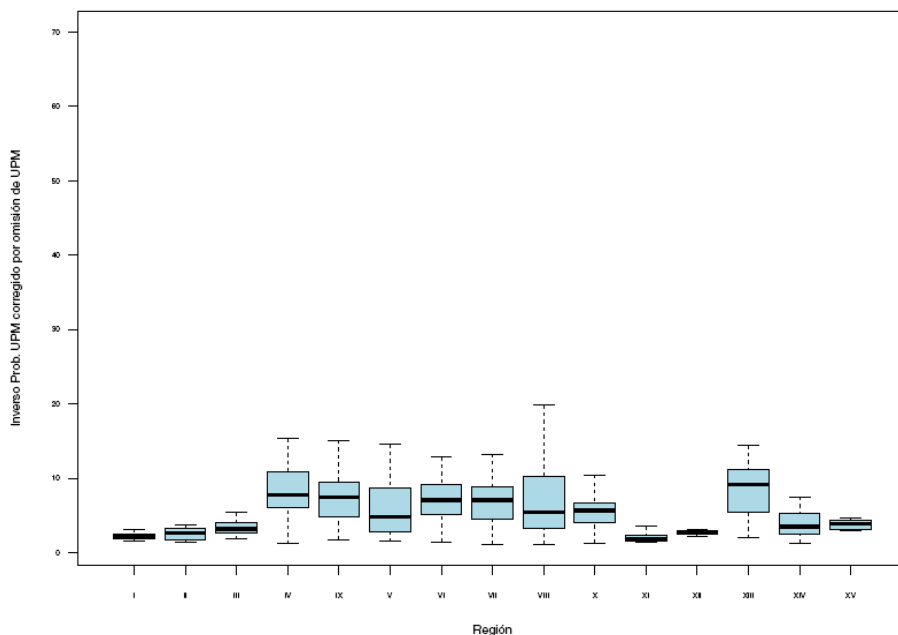


Figura VI.7: Distribución del Ponderador de Selección Corregido por Omisión de Manzanas

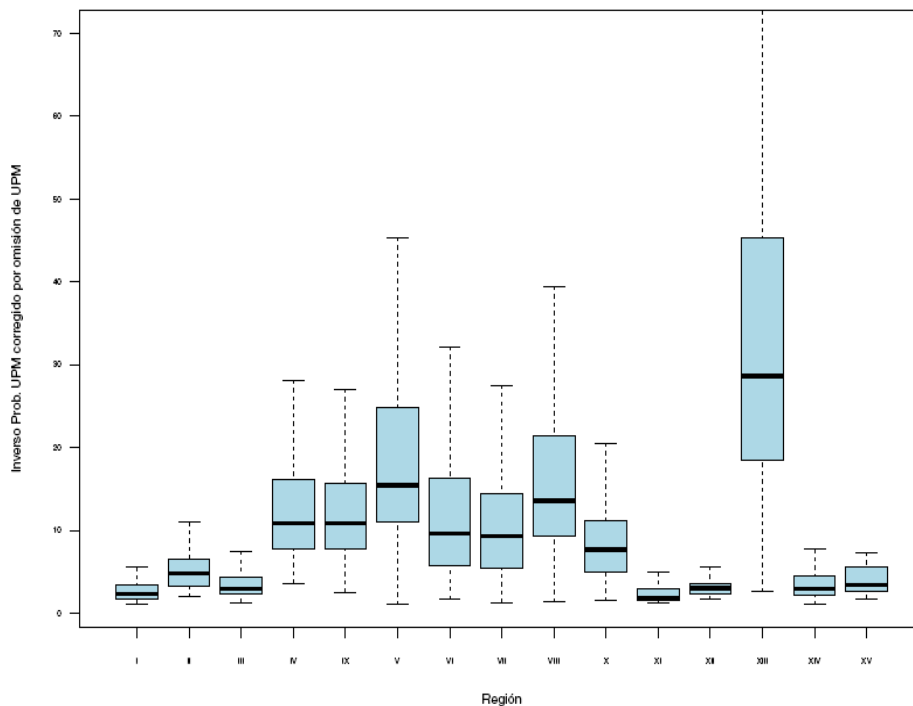


Tabla VI.7: Número de Manzanas y Secciones Estimadas con Ponderador de Selección Ajustado por Omisión de Conglomerados

Región	Estimación Muestra		Total
	N° Secciones	N° Manzanas	
Tarapacá	61	2.581	2.642
Antofagasta	45	6.039	6.083
Atacama	100	3.888	3.988
Coquimbo	502	7.033	7.534
Valparaíso	511	17.815	18.327
Libertador General Bernardo OHiggins	855	7.798	8.653
Maule	1.186	6.060	7.245
Biobio	1.248	16.489	17.737
Araucanía	1.088	7.887	8.976
Los Lagos	774	5.368	6.142
Aysén del General Carlos Ibañez del Campo	124	1.215	1.339
Magallanes y La Antártica Chilena	46	2.136	2.182
Metropolitana de Santiago	454	48.388	48.842
Los Ríos	400	3.221	3.621
Arica y Parinacota	56	2.741	2.797
Total	7.448	138.659	146.107

Tabla VI.7: Número de Viviendas Estimadas con Ponderador de Selección Ajustado por Omisión de Conglomerados

Región	Estimación Muestra		Total
	N° Secciones	N° Manzanas	
Tarapacá	3.268	70.162	73.430
Antofagasta	2.347	131.946	134.293
Atacama	5.192	74.686	79.878
Coquimbo	39.412	152.956	192.368
Valparaíso	38.768	507.778	546.546
Libertador General Bernardo OHiggins	67.371	159.390	226.761
Maule	96.332	185.055	281.387
Biobio	100.123	445.810	545.933
Araucanía	89.333	170.383	259.716
Los Lagos	64.581	136.763	201.344
Aysén del General Carlos Ibañez del Campo	6.951	19.782	26.733
Magallanes y La Antártica Chilena	2.630	42.374	45.004
Metropolitana de Santiago	45.434	1.784.022	1.829.456
Los Ríos	31.685	66.790	98.475
Arica y Parinacota	2.913	52.865	55.778
Total	596.340	4.000.762	4.597.102

3. Ponderación de selección de viviendas

Los pasos descritos anteriormente están vinculados con la probabilidad de selección de las manzanas y secciones. Sin embargo, la última unidad de selección de la muestra es la vivienda al interior de los conglomerados previamente seleccionados. En los apartados siguientes se detallan las probabilidades de selección de las viviendas, así como los ajustes y ponderadores asociados.

3.1. Probabilidad Condicional de selección de una vivienda

Al interior de cada conglomerado, indistintamente del marco de procedencia, las viviendas fueron seleccionadas bajo un algoritmo de selección sistemático, de modo que todas las viviendas particulares ocupadas en el conglomerado, registradas en el empadronamiento, tuvieran la misma probabilidad de ser seleccionadas. Luego, la probabilidad condicional 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 , viene dada por,

$$P_{hi}(j/i) = \frac{m_{hi}}{M_{hi}} \quad (6)$$

donde,

j : Corresponde al índice de la vivienda.

M_{hi} : Corresponde al total de viviendas ocupadas en la manzana o sección i , según empadronamiento realizado previo a la recolección de datos.

m_{hi} : Corresponde al total de viviendas seleccionadas en la manzana o sección i del estrato h .

De esta forma, siguiendo las reglas de probabilidad, se tiene que la probabilidad incondicional de que la j -ésima vivienda pertenezca a la muestra Casen 2011 puede ser expresada como,

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

donde,

$P_{hi}(j)$: Corresponde a la probabilidad de que la vivienda j -ésima pertenezca a la muestra.

$P_h(i)$: Corresponde a la probabilidad de que el conglomerado (manzana o sección) i -ésimo pertenezca a la muestra.

A continuación se describe el cálculo del Ponderador de Selección de Viviendas según la fórmula (6), donde $P_h(i)$ corresponde al inverso del Ponderador de Selección de Conglomerados Corregido por Omisión.

3.2. Ponderador de selección de viviendas

La fórmula (6) puede ser interpretada como la probabilidad de que una vivienda sea seleccionada desde el marco de muestreo. El inverso de esta probabilidad, en la fórmula (7) abajo, corresponde al Ponderador de Selección de Viviendas.

Este ponderador puede ser interpretado como el número de viviendas que representan en la población las viviendas seleccionadas⁴¹ en la muestra Casen 2011, donde todas las viviendas pertenecientes al mismo conglomerado les corresponde el mismo Ponderador de Selección de Viviendas.

$$\begin{aligned} w_{hij}'' &= w_{hi}' \cdot \frac{1}{P_{hi}(j/i)} \\ &= w_{hi}' \cdot \frac{M_{hi}}{m_{hi}} \end{aligned} \quad (8)$$

El Ponderador de Selección de Viviendas puede ser utilizado directamente para realizar estimaciones a nivel de viviendas. Por ejemplo, estimar el total de viviendas existentes por región, estimar el número de viviendas con determinadas características, etc. A medida que se aleja el momento del censo, las estimaciones de los stocks de viviendas comienzan a desactualizarse. Contar con estimaciones a partir de una encuesta grande como Casen permite identificar zonas del país donde pudiera haber más dinamismo en la construcción de viviendas y por lo tanto estar más proclives a desactualización del marco muestral. Las Figuras VI.8 y VI.9 ilustran la distribución del Ponderador de Selección de Viviendas para cada una de las regiones del país.

Al comparar la Figura VI.8 con la Figura VI.9, se puede apreciar que se mantiene el patrón de comportamiento de la distribución de los ponderadores de las manzanas y secciones entre las distintas regiones. Sin embargo, las brechas entre regiones aumentan, así como también la variabilidad al interior de cada región, con la mayor dispersión visible para la región Metropolitana. Este resultado podría sorprender, ya que por diseño se espera que las probabilidades de selección de las viviendas al interior de los conglomerados fuesen relativamente homogéneas. Esta homogeneidad, sin embargo, se ve interrumpida cuando hay desactualizaciones importantes en el marco muestral y el proceso de empadronamiento en terreno muestra que la cantidad esperada de viviendas en un conglomerado dista de lo observado en terreno. Las probabilidades de selección de las viviendas se ven entonces afectadas (aumentan/disminuyen) debido a los crecimientos/decrecimientos observados en el conglomerado, lo cual hace que las viviendas en la muestra deban representar a más/menos viviendas que lo definido según la fracción de selección.

Este resultado se observa en todas las regiones - tanto la dispersión como el valor medio del Ponderador de Selección de Vivienda aumenta. El fenómeno se observa más pronunciado en el marco de manzanas, ya que a la variabilidad asociada al proceso de empadronamiento se agrega la mayor variabilidad del Ponderador de Selección de Manzanas. El ponderador de vivienda está asociado a todas las viviendas seleccionadas, sin embargo, es posible que algunas de ellas no correspondan a viviendas particulares (ej. viviendas colectivas, negocios, etc.) o bien se desconozca su condición. El ajuste que se describe a continuación corrige esta situación.

41 El número de viviendas seleccionadas varía de acuerdo al Marco de procedencia. En el caso de las secciones pertenecientes a la muestra Casen 2011, éstas están constituidas por un total que oscila entre 41 y 176 viviendas.

Figura VI.8: Distribución del Ponderador de Selección de Viviendas, en áreas rurales (MS 2002).

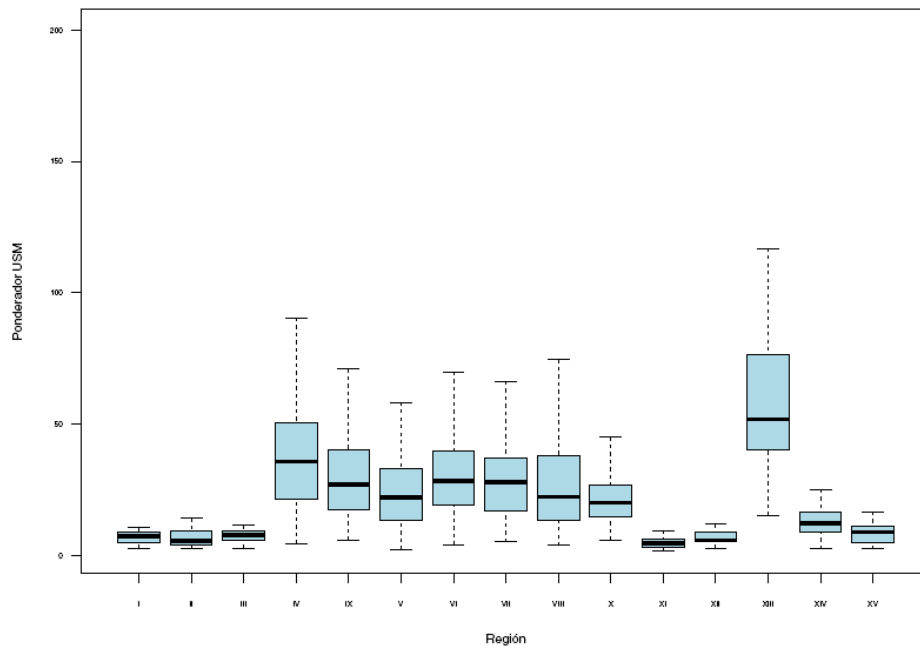


Figura VI.9: Distribución del Ponderador de Selección de Viviendas, en áreas urbanas (MM 2008).

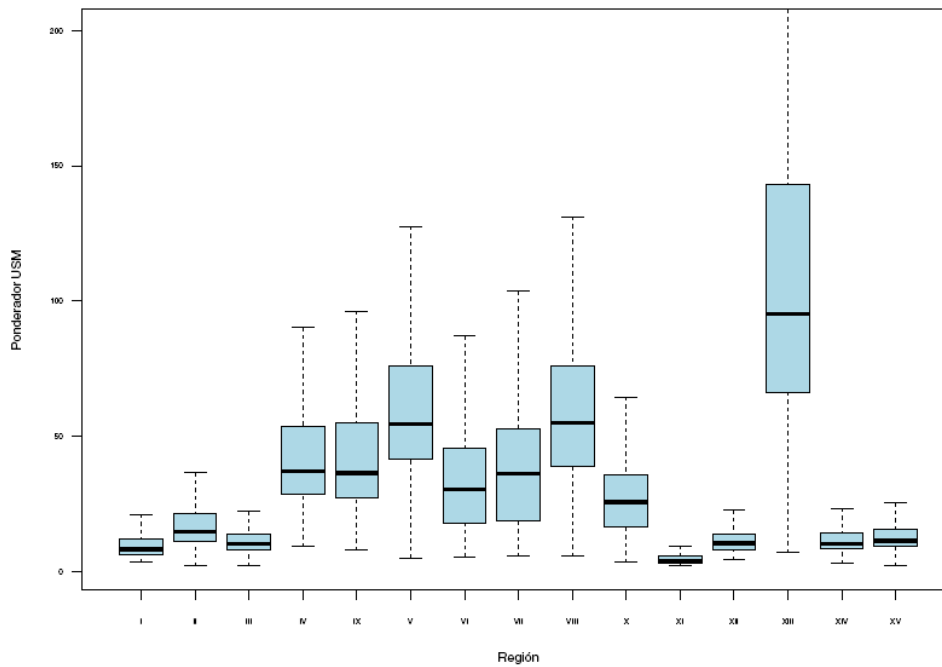


Tabla VI.9: Estadísticas descriptivas del Ponderador de Selección de Viviendas, según Marco y Región

Marco	Región	Ponderador de viviendas/hogares - FULL													
		Viviendas Muestra	Mínimo	Percentil 05	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Percentil 99	Máximo	Media	Moda	Viviendas Estimadas
Manzanas	Tarapacá	7.003	3,4	5,2	5,6	6,3	8,1	12,2	19,4	27,5	59,2	730,6	11,9	7,2	83.649
	Antofagasta	7.118	2,2	8,5	9,2	11,1	14,6	21,3	31,5	42,6	99,6	1.310,7	20,2	12,7	143.619
	Atacama	5.907	2,0	5,9	6,7	8,0	10,1	13,7	21,4	28,7	54,9	453,1	13,1	12,7	77.179
	Coquimbo	3.142	9,2	16,1	19,0	28,6	37,0	53,4	75,1	97,0	230,7	2.499,4	49,7	87,0	156.211
	Valparaíso	7.867	5,0	20,8	28,4	41,5	54,4	76,0	121,3	165,6	342,1	946,0	69,6	35,2	547.545
	Libertador General Bernardo OHiggins	3.830	5,5	7,6	11,1	17,7	30,5	45,5	68,6	99,9	257,7	2.930,8	43,0	49,3	164.765
	Maule	4.365	5,6	13,0	14,3	18,7	36,1	52,8	74,6	98,8	161,1	1.823,3	43,1	24,2	188.029
	Biobío	6.588	5,7	22,6	27,4	38,8	55,0	75,8	127,5	156,8	355,2	3.602,5	74,8	34,8	493.099
	Araucanía	3.630	7,9	13,9	20,1	27,2	36,3	55,0	79,3	105,7	239,9	2.367,7	50,4	65,2	182.794
	Los Lagos	4.465	3,4	8,5	11,2	16,5	25,6	35,7	53,9	77,2	185,3	874,5	32,6	9,7	145.461
	Aysén del General Carlos Ibañez del Campo	4.017	2,0	2,5	2,6	3,0	3,8	5,6	9,3	11,6	40,2	85,9	5,5	5,9	22.037
	Magallanes y La Antártica Chilena	3.351	4,4	6,3	6,8	8,0	10,5	13,8	22,9	27,8	70,6	421,4	14,0	9,0	46.805
	Metropolitana de Santiago	14.496	7,0	36,8	46,8	66,2	95,2	143,1	227,6	306,9	809,1	4.694,0	132,6	52,7	1.921.950
	Los Ríos	5.081	3,0	6,1	6,9	8,2	10,3	14,3	20,9	30,4	88,3	696,2	14,6	7,6	74.371
	Arica y Parinacota	3.550	2,2	7,3	7,9	9,2	11,4	15,7	24,9	37,1	75,9	868,2	15,9	7,3	56.584
Secciones	Tarapacá	760	2,7	2,7	3,4	5,0	7,4	8,7	9,7	10,5	35,2	35,2	7,6	2,7	5.757
	Antofagasta	388	2,4	2,4	2,8	3,8	5,5	9,4	14,1	18,7	18,7	18,7	7,1	2,4	2.773
	Atacama	677	2,8	3,3	5,4	5,9	7,8	9,5	18,5	20,8	24,7	24,7	9,0	5,6	6.068
	Coquimbo	1.376	4,5	7,2	10,3	21,3	35,6	50,4	59,3	67,2	90,3	90,3	35,9	28,4	49.405
	Valparaíso	1.828	2,3	4,5	6,7	13,5	22,1	33,1	50,5	76,8	79,8	137,1	26,6	31,2	48.622
	Libertador General Bernardo OHiggins	2.780	4,1	8,4	10,8	19,1	28,4	39,6	54,1	60,4	76,2	114,0	30,4	4,1	84.574
	Maule	3.863	5,2	8,9	10,1	16,9	27,8	37,2	45,4	53,5	81,2	558,4	30,9	16,9	119.211
	Biobío	3.889	3,7	5,5	7,8	13,3	22,3	38,1	53,9	60,0	99,6	135,4	27,5	5,4	107.101
	Araucanía	3.312	5,9	10,6	13,0	17,4	26,9	40,3	51,0	56,5	105,8	107,6	30,1	19,4	99.831
	Los Lagos	3.382	5,9	9,2	10,7	14,6	20,1	26,9	34,9	44,9	65,7	117,0	22,5	16,3	76.196
	Aysén del General Carlos Ibañez del Campo	1.566	1,9	2,3	2,6	3,2	4,7	6,1	13,9	16,2	20,9	20,9	6,1	4,8	9.579
	Magallanes y La Antártica Chilena	422	2,7	2,7	2,9	5,6	5,7	8,7	10,0	12,0	17,7	17,7	7,0	5,6	2.950
	Metropolitana de Santiago	963	15,3	19,2	27,7	40,1	51,9	76,4	106,0	131,2	149,2	149,2	59,5	149,2	57.320
	Los Ríos	2.809	2,5	4,6	6,6	8,8	12,3	16,7	21,9	25,1	34,3	37,1	13,2	12,3	37.188
	Arica y Parinacota	442	2,5	2,5	3,5	4,7	8,9	11,0	14,5	16,6	16,6	16,6	8,4	14,5	3.732
Total País		112.867	1,9	5,2	6,9	10,9	23,8	50,7	94,2	135,3	299,4	4.694,0	44,4	2,7	5.014.405

4. Ponderación por elegibilidad

La población de inferencia de la encuesta Casen 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 (y exclusión) en el marco. Por ejemplo, fallas en el proceso de enumeración pueden derivar en el registro de inmuebles no elegibles (ej. casas de veraneo, negocios) en el marco⁴². Cambios en el uso de la vivienda también pueden ocurrir entre el momento de enumeración y el momento de la entrevista. Por esta razón, es importante ajustar las probabilidades de selección de las viviendas seleccionadas para incorporar el hecho de que una fracción de las viviendas seleccionadas en la muestra es no elegible y para otras simplemente se desconoce su elegibilidad.

Al término del trabajo de campo, todas las viviendas seleccionadas inicialmente por el INE terminan siendo clasificadas en tres grandes grupos: (1) elegibles, (2) no elegibles, y (3) de elegibilidad desconocida⁴³. Al primer grupo pertenecen aquellas edificaciones en que el encuestador pudo determinar que se trataban de viviendas particulares ocupadas (incluye tanto entrevistas como no entrevistas). Al segundo grupo corresponden edificaciones identificadas como negocios, viviendas colectivas, viviendas deshabitadas, viviendas de veraneo, viviendas destruidas, etc. El tercer grupo de unidades son aquellas en que no se pudo determinar su estado. Este es el caso, por ejemplo, de unidades que nunca fueron enviadas a terreno, donde no se pudo llegar al lugar, donde no se pudo encontrar el lugar, y otros casos similares.

El Ponderador de Selección de Viviendas tiene valores válidos para las viviendas elegibles, no elegibles y de elegibilidad desconocida. En lo que sigue del proceso sólo se dejarán valores válidos para las viviendas elegibles. A continuación se describen los ajustes aplicados.

4.1. Ajuste por Elegibilidad Desconocida

En la población, todas las viviendas tienen un estatus conocido de elegibilidad – ellas son o no son viviendas particulares ocupadas. Durante el trabajo de campo, sin embargo, este estado no se puede verificar y algunas viviendas terminan clasificadas como “Elegibilidad Desconocida”. Antes de seguir avanzando, se debe corregir esta situación. Con este fin, el Ponderador de Selección de Viviendas asociado a viviendas clasificadas como de “Elegibilidad Desconocida” se redistribuye entre las viviendas con Elegibilidad Conocida (elegibles y no elegibles) al interior de cada estrato.

La proporción de los pesos que debe ser redistribuida 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), es decir,

42 El proceso de enumeración consiste en el registro de direcciones asociadas a las manzanas y secciones seleccionadas en la muestra. Este trabajo se realiza en terreno por personal del INE entre 3 y 6 meses antes del trabajo de campo.

43 Para una completa descripción de los Estados de Disposición Final de Casos y su protocolo de implementación ver Manual del Trabajo de Campo Casen 2011.

$$\hat{R}_{h,known} = \frac{\sum_{i \in \Omega_h} \sum_{j \in \Theta_i} w_{hij}''}{\sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,known}} w_{hij}''}, \quad (9)$$

donde

Θ_i : Corresponde al conjunto de viviendas seleccionadas en la manzana o sección i.

$\Theta_{i,known}$: Corresponde al conjunto de viviendas seleccionadas en la manzana o sección i y clasificadas como elegibles o no elegibles.

Luego, el Ponderador de Selección de Vivienda Corregido por Elegibilidad está dado por,

$$w_{hij}''' = \hat{R}_{h,known} \cdot w_{hij}'' \quad (10)$$

Las figuras VI.10 y VI.11 presentan la distribución del Ponderador de Selección de Viviendas corregido por Elegibilidad, para cada región y según el marco de procedencia. Este ajuste no afecta significativamente la magnitud y variabilidad de los ponderadores asociados al Marco de secciones o de manzanas. Esto, probablemente por la baja prevalencia de unidades de elegibilidad desconocida. De las 112.867 viviendas seleccionadas un 1,8% fueron clasificadas como elegibilidad desconocida (2.082 viviendas).

Algunos resultados del proceso de ajuste se presentan en la Tabla VI.10. En la muestra Casen 2011, hay 110.785 viviendas con Elegibilidad Conocida. El Ponderador de Selección de Viviendas Corregido por Elegibilidad oscila entre 2 y 4.736. La estimación del total del total de viviendas utilizando este ponderador es de 5.014.405 unidades, lo que corresponde al mismo total estimado en el apartado 3.2.

Figura VI.10: Distribución del Ponderador de Selección Corregido por Elegibilidad (MS2002)

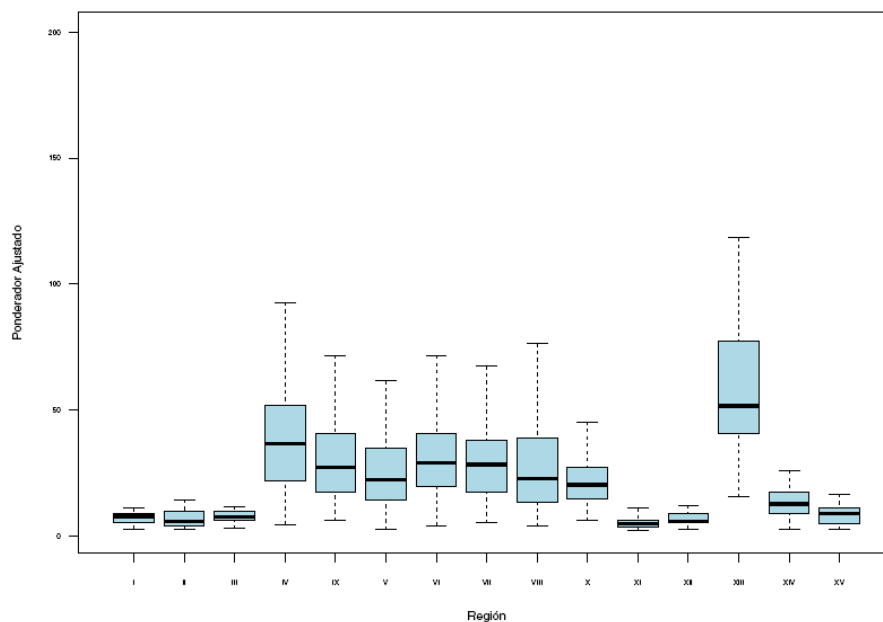


Figura VI.11: Distribución del Ponderador de Selección de Vivienda Corregido por Elegibilidad (MM2008)

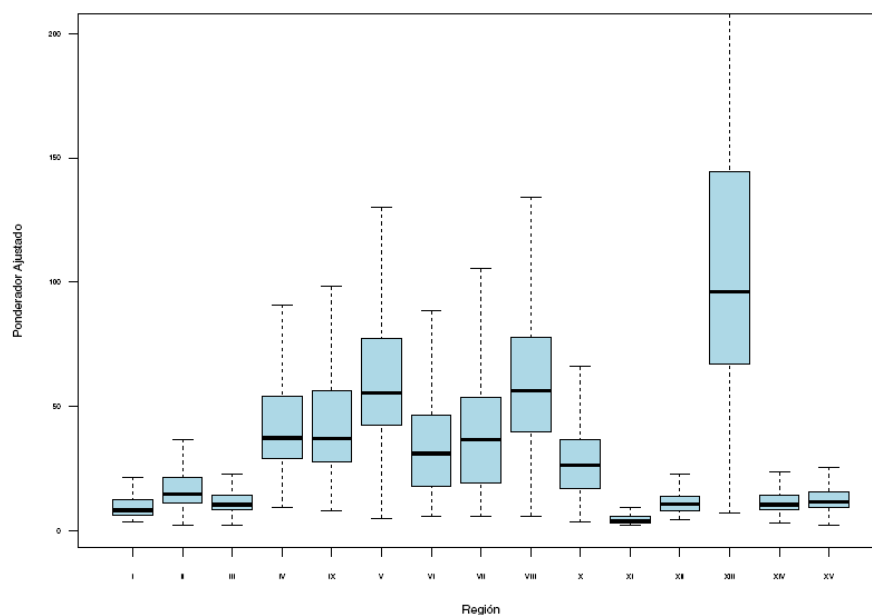


Tabla VI.10: Estadísticas descriptivas del Ponderador de Vivienda Corregido por Elegibilidad, según Marco y Región.

Marco	Región	Factor ajustado por elegibilidad desconocida - FULL													
		Sin ponderador	Viviendas muestra	Mínimo	Moda	Percentil 05	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Media	Viviendas estimadas
Manzanas	Tarapacá	165	7.003	3,4	7,3	5,3	5,7	6,4	8,3	12,5	19,9	28,1	747,0	12,2	83.727
	Antofagasta	99	7.118	2,2	12,9	8,6	9,3	11,2	14,8	21,5	32,0	43,3	1.330,7	20,5	143.611
	Atacama	215	5.907	2,1	13,2	6,1	7,0	8,3	10,5	14,2	22,1	29,8	471,9	13,6	77.179
	Coquimbo	29	3.142	9,2	87,7	16,3	19,1	28,9	37,3	53,9	75,7	99,9	2.519,2	50,2	156.222
	Valparaíso	135	7.867	5,1	35,9	21,2	29,0	42,3	55,5	77,5	123,6	168,8	964,4	70,9	547.996
	Libertador General Bernardo OHiggins	70	3.830	5,6	50,1	7,7	11,3	17,9	31,1	46,3	69,7	101,5	2.978,7	43,8	164.758
	Maule	68	4.365	5,7	24,6	13,2	14,5	19,0	36,7	53,7	75,8	100,4	1.852,7	43,8	188.331
	Biobío	167	6.588	5,9	38,1	23,4	28,2	39,7	56,2	77,6	130,4	160,8	3.685,5	76,8	493.075
	Araucanía	89	3.630	8,0	37,0	14,2	20,5	27,8	37,1	56,2	80,9	107,9	2.417,0	51,6	182.621
	Los Lagos	106	4.465	3,5	10,0	8,7	11,4	16,9	26,3	36,6	55,2	81,0	896,8	33,4	145.463
	Aysén del General Carlos Ibañez del Campo	43	4.017	2,0	6,0	2,5	2,7	3,0	3,8	5,7	9,4	11,7	86,9	5,5	22.019
	Magallanes y La Antártica Chilena	41	3.351	4,5	9,1	6,4	6,9	8,1	10,7	14,0	23,2	28,2	427,1	14,1	46.794
	Metropolitana de Santiago	72	14.496	7,1	53,1	37,1	47,2	66,8	96,0	144,3	229,6	309,6	4.735,7	133,2	1.921.933
	Los Ríos	64	5.081	3,0	7,7	6,2	7,0	8,4	10,5	14,5	21,2	30,7	706,3	14,8	74.330
Arica y Parinacota	20	3.550	2,2	7,4	7,3	7,9	9,3	11,5	15,8	24,9	37,3	873,7	16,0	56.592	
Secciones	Tarapacá	47	760	2,8	2,8	2,8	3,6	5,3	7,7	9,1	10,1	10,1	36,7	8,0	5.680
	Antofagasta	8	388	2,5	2,5	2,5	2,8	3,9	5,6	9,7	14,4	19,1	19,1	7,3	2.781
	Atacama	13	677	2,9	5,7	3,4	5,6	6,1	7,4	9,8	19,0	21,4	25,3	9,1	6.068
	Coquimbo	36	1.376	4,5	29,1	7,3	10,6	21,9	36,5	51,7	60,9	68,9	92,7	36,9	49.394
	Valparaíso	83	1.828	2,5	33,0	4,7	7,0	14,3	22,2	34,7	53,4	81,2	144,9	27,6	48.171
	Libertador General Bernardo OHiggins	76	2.780	4,2	4,2	8,5	10,9	19,6	28,9	40,5	55,4	61,8	116,6	31,3	84.581
	Maule	80	3.863	5,3	17,2	9,0	10,2	17,2	28,4	37,9	46,2	54,5	567,4	31,4	118.909
	Biobío	104	3.889	3,8	5,6	5,7	8,0	13,6	22,8	38,9	55,1	61,4	138,5	28,3	107.124
	Araucanía	52	3.312	6,1	19,6	10,7	13,2	17,5	27,2	40,7	51,4	57,0	109,9	30,7	100.003
	Los Lagos	48	3.382	6,1	22,8	9,4	10,8	14,8	20,2	27,2	35,2	45,3	118,1	22,9	76.194
	Aysén del General Carlos Ibañez del Campo	52	1.566	2,0	5,0	2,4	2,7	3,4	4,7	6,4	14,0	16,9	21,8	6,3	9.597
	Magallanes y La Antártica Chilena	6	422	2,8	5,6	2,8	3,0	5,6	5,8	8,8	10,2	12,2	18,0	7,1	2.962
	Metropolitana de Santiago	14	963	15,4	151,0	19,4	28,0	40,6	51,6	77,3	107,2	132,8	151,0	60,4	57.337
	Los Ríos	77	2.809	2,6	12,7	4,8	6,7	9,0	12,7	17,2	22,5	25,8	38,1	13,6	37.228
Arica y Parinacota	3	442	2,5	14,5	2,5	3,5	4,7	8,9	11,0	14,5	16,6	16,6	8,5	3.725	
Total País		2082	112.867	2,00	2,83	5,28	7,04	11,09	24,42	51,72	96,02	138,09	4735,74	45,26	5.014.405

4.2. Ajuste por No Elegibilidad

Entre las 110.785 viviendas con Elegibilidad Conocida, hay 105.560 que corresponden a viviendas Elegibles y 5.225 que corresponden a No Elegibles (ver Cuadro 14). 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 (No Elegibles, tales como empresas, viviendas abandonadas, viviendas de veraneo, viviendas demolidas, etc.).

Esto se implementa asignando un valor blanco (missing), en el Ponderador de Selección de Viviendas Corregido por Elegibilidad, a las viviendas con clasificación "No Elegible". A partir de este ajuste, la suma del ponderador corresponde a la estimación del total de viviendas elegibles (viviendas particulares ocupadas) y es de 4.794.266 unidades.

Entre las viviendas elegibles, hay 84.239 para las cuales se logró concretar una o más entrevistas⁴⁴ y 21.321 para las cuales no se logró concretar ninguna. El siguiente capítulo discute los problemas que esta situación conlleva y la forma de corregirlos implementada en Casen 2011.

Tabla VI.11: Total de viviendas elegibles

Vivienda elegible	Frecuencia	Porcentaje
No	5.225	4,6
Sí	105.560	93,5
Desconocida	2.082	1,8
Total	112.867	100,0

5. Ponderación por no respuesta

De las 105.560 viviendas elegibles seleccionadas en la muestra, se logró concretar entrevistas en 84.239 de ellas. Es decir, la tasa de respuesta de la encuesta fue de 80%. ¿Qué implicancias puede tener el hecho de no haber logrado entrevistas en 1 de cada 5 viviendas elegibles? La respuesta es "depende".

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 de las estimaciones de interés. Sin embargo, si hay alguna relación entre la no respuesta y la variable de interés, existe riesgo de sesgo en las estimaciones a partir de la muestra. En Casen 2011, existe evidencia de que la no respuesta es mayor en comunas de mayor nivel socioeconómico. Si no se corrige por este patrón de no respuesta, estadísticas como el coeficiente de Gini, podrían estar subestimadas (ya que en la encuesta estarían subrepresentadas las viviendas en comunas más afluentes). Otras estadísticas, como la participación en

44 Recordar que, bajo el protocolo de trabajo de campo Casen, al interior de una vivienda seleccionada se completan tanta entrevistas como hogares sean identificados en la vivienda.

organizaciones sociales, puede no verse afectada por la falta de respuestas en estos sectores.

La población en Chile está segregada geográficamente según nivel socioeconómico y el patrón de no respuesta lo está de la misma manera. Dado que los temas de mayor interés de la Encuesta Casen están fuertemente relacionados con variables socioeconómicas, el riesgo de sesgo por no respuesta es elevado. En este contexto, se hace necesario aplicar algún tipo de ajuste estadístico que trate de corregir este problema.

En la literatura existen métodos explícitos e implícitos para la corrección de la no respuesta. Hasta 2006, la metodología de construcción de los factores de expansión Casen consideraba ajustes "implícitos". En 2009, se implementó por primera vez un ajuste explícito. El ajuste consistió en re-ponderar los ponderadores de selección de vivienda por el inverso de la tasa de respuesta en cada conglomerado. Intuitivamente, este ajuste asume que la tasa de respuesta al interior de cada conglomerado es una buena aproximación de la probabilidad de responder de cada vivienda en el conglomerado. También asume que al interior de cada conglomerado, la no respuesta es aleatoria, por lo tanto los que responden pueden representar a los que no responden.

El enfoque implementado en 2009 es consistente con los métodos denominados "ajuste por medias condicionadas", donde la media corresponde a la tasa de respuesta y se condiciona en el identificador de conglomerados. Esta metodología va en la dirección correcta y es perfectible.

El método utilizado para el desarrollo del ajuste de no respuesta en 2011 se denomina "estratificación en base al propensity score" (Valliant et al. 2012). En términos concretos, el método requiere de: (1) la definición de las celdas de ajuste y (2) la determinación de la razón de ajuste.

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 probabilidad de responder de las viviendas elegibles.
2. Predicción de la probabilidad de responder de cada vivienda elegible en la muestra ($n=105.560$).
3. Ordenamiento de las probabilidades predichas, de mayor a menor, y creación de veintiles de la probabilidad de responder.

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 (11) 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, lo que puede ser expresado como,

$$\hat{R}_{g,R} = \frac{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i,eleg}} w_{hij}'''}{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i,eleg,resp}} w_{hij}'''} , \quad (11)$$

donde

g : Representa el indicador del veintil de no respuesta.

Φ_g : Corresponde al conjunto de manzanas o secciones asignadas al veintil g .

$\Theta_{i,eleg}$: Es el conjunto de viviendas pertenecientes al conglomerado i y catalogadas como elegibles.

$\Theta_{i,eleg,resp}$: Es el conjunto de viviendas pertenecientes al conglomerado i , catalogadas como elegibles y que responden la encuesta.

La Tabla VI.12 presenta el ponderador de no respuesta (factor de ajuste) estimado para cada veintil. Aquí se aprecia que el grupo que tiene una mayor ponderación es el grupo 1 incrementando los ponderadores de elegibilidad desconocida al doble, mientras que el ponderador de las viviendas pertenecientes al grupo 20 son los que tienen menor impacto por la corrección de no respuesta.

Tabla VI.12: Ponderador de no respuesta por celda de ajuste

Celda de ajuste de no respuesta	Ponderador de corrección por no respuesta		
	Sin Ponderador	Viviendas Muestra	R_nr
1	4.105	7.237	2,19
2	2.409	5.025	1,78
3	2.433	6.459	1,45
4	1.267	3.708	1,46
5	1.346	3.819	1,44
6	1.783	5.594	1,37
7	1.636	6.219	1,31
8	1.188	4.532	1,30
9	1.660	6.401	1,26
10	1.292	5.583	1,25
11	947	4.785	1,16
12	948	5.335	1,17
13	1.013	5.218	1,15
14	983	5.039	1,19
15	914	5.537	1,14
16	807	5.129	1,13
17	907	5.520	1,15
18	979	5.903	1,13
19	996	6.773	1,10
20	1.015	9.051	1,07

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

$$w_{hij}^{NR} = \hat{R}_{g,R} \cdot w_{hij}^m \quad (12)$$

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

Las Figuras VI.12 y VI.13 presentan la distribución del Ponderador Corregido por No Respuesta, diferenciado según Marco de procedencia. Para las unidades provenientes del marco de secciones (MS2002), se puede apreciar en la figura VI.12, que la variabilidad del Ponderador Corregido por No Respuesta tiene un comportamiento similar al Ponderador Corregido por Elegibilidad. Se observa que sólo cambia la escala, es decir los rangos entre los cuales oscilan los ponderadores cambian entre un ajuste y otro, pero no existe un cambio significativo en cuanto a la variabilidad del ponderador, debido al nuevo ajuste. La Región Metropolitana sigue concentrando la mayor variabilidad. El valor más extremo, sin embargo, se encuentra en la Región del Maule (655 unidades, valor extremo omitido de la figura).

Los ponderadores ajustados por no respuesta para las unidades provenientes del marco de manzanas (MM2008), visualmente no tienen cambios significativos en lo que respecta a variabilidad. Sin embargo, se debe señalar que el rango de los ponderadores al interior de cada región fue incrementado. El caso más relevante es la región Metropolitana, pues el 50% de los ponderadores de elegibilidad poseen valores inferiores a 96 unidades, mientras que ahora asciende a 137 unidades.

En la Tabla VI.13, se exponen las estadísticas descriptivas básicas del ponderador ajustado por no respuesta, donde se puede apreciar que la estimación del total de viviendas elegibles que responde, coincide con la estimación de viviendas elegibles realizados en el apartado 4.2, 4.794.266 unidades. Los ponderadores más grandes (superior a mil unidades) provienen del Marco de manzanas, con mayor énfasis en las regiones Metropolitana y Biobío.

Con este ponderador se pueden realizar análisis a nivel de viviendas, sin embargo las unidades de análisis corresponden a los hogares y personas residentes en las viviendas particulares presentes en la muestra. Si se realizan análisis de personas u hogares con este ponderador, es posible que la población sea subestimada o sobrestimada de forma desmesurada. Para resolver este tipo de inconvenientes se realiza un nuevo ajuste, apoyado por variables exógenas, que den cuenta del total de la personas del país. La sección que sigue describe el método de calibración que corrige esta situación.

Figura VI.12: Distribución del Ponderador Corregido por No Respuesta (MS2002)

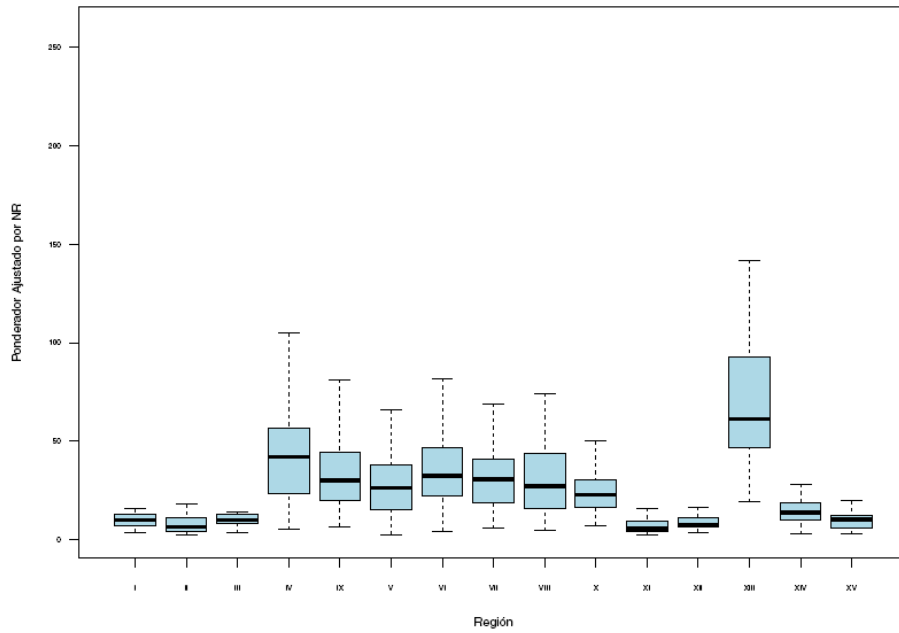


Figura VI.13: Distribución del Ponderador Corregido por No Respuesta (MM 2008)

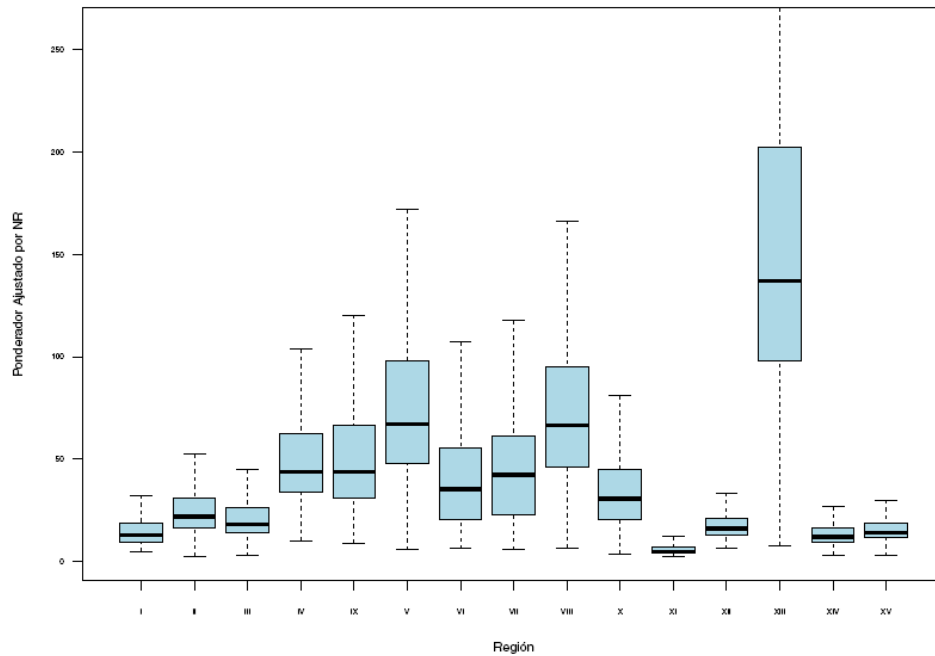


Tabla VI.13: Estadísticas descriptivas del Ponderador Corregido por No Respuesta.

Marco	Región	Factor ajustado por no respuesta - FULL													
		Sin ponderador	Viviendas muestra	Mínimo	Moda	Percentil 05	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Media	Suma
Manzanas	Tarapacá	2621	7.003	5,0	9,0	7,4	8,0	9,3	12,9	18,6	32,2	44,9	1.078,4	18,2	79.583
	Antofagasta	2468	7.118	2,6	25,1	11,7	13,5	16,8	22,1	31,0	45,3	61,6	1.921,2	30,3	140.675
	Atacama	2519	5.907	3,1	28,8	8,1	9,9	14,2	18,1	26,5	36,7	49,7	1.032,2	23,2	78.531
	Coquimbo	691	3.142	10,1	127,2	17,8	21,8	34,2	43,8	62,4	88,7	127,2	2.954,6	58,0	142.272
	Valparaíso	2698	7.867	5,8	44,9	22,7	31,8	48,1	67,2	97,8	146,4	207,8	1.258,8	87,0	449.622
	Libertador General Bernardo OHiggins	723	3.830	6,4	63,2	9,7	13,0	20,5	35,5	55,7	88,1	117,8	3.551,7	50,8	157.911
	Maule	722	4.365	6,0	73,9	14,9	16,6	23,0	42,3	61,1	89,3	115,7	2.111,0	50,5	184.072
	Biobio	1362	6.588	6,7	70,6	26,5	32,4	46,2	66,4	94,9	156,6	215,4	4.199,2	91,7	479.242
	Araucanía	830	3.630	9,1	43,8	15,8	23,1	31,0	43,8	66,8	95,2	134,5	3.027,1	61,4	171.836
	Los Lagos	977	4.465	3,8	25,7	10,4	13,0	20,4	30,7	44,7	64,6	95,5	1.123,2	40,2	140.287
	Aysén del General Carlos Ibañez del Campo	1085	4.017	2,6	3,3	3,3	3,5	4,0	4,9	7,3	11,3	15,2	112,7	7,0	20.450
	Magallanes y La Antártica Chilena	1267	3.351	6,5	16,2	9,3	10,2	12,8	16,3	21,2	32,5	40,6	619,2	21,5	44.889
	Metropolitana de Santiago	4423	14.496	8,1	76,4	55,5	70,1	97,9	137,2	202,2	324,6	452,3	5.931,1	186,6	1.879.634
	Los Ríos	846	5.081	3,2	7,7	6,8	7,9	9,6	12,0	16,5	24,5	35,2	798,3	17,0	71.849
	Arica y Parinacota	838	3.550	3,0	9,3	9,3	9,9	11,6	14,1	19,1	30,1	43,6	1.024,8	19,3	52.427
Secciones	Tarapacá	300	760	3,9	9,6	3,9	4,7	7,2	10,0	13,1	13,8	15,9	39,2	11,0	5.044
	Antofagasta	114	388	2,7	2,7	2,7	3,5	4,5	6,5	11,3	18,2	24,8	24,8	9,3	2.535
	Atacama	173	677	3,6	10,2	4,3	7,2	8,2	10,2	12,7	14,4	34,7	43,6	11,9	6.004
	Coquimbo	242	1.376	5,3	33,6	8,1	10,4	23,3	42,2	57,0	65,2	79,5	104,8	41,2	46.748
	Valparaíso	404	1.828	2,8	5,1	5,2	7,7	15,3	26,3	37,9	62,5	76,5	181,5	30,7	43.763
	Libertador General Bernardo OHiggins	476	2.780	4,6	26,4	10,8	12,8	22,3	32,4	46,7	60,7	68,6	135,3	35,6	81.928
	Maule	347	3.863	6,0	27,0	9,7	11,4	19,0	30,8	40,9	50,1	61,6	654,9	34,5	121.304
	Biobio	587	3.889	4,8	6,0	6,7	10,1	15,7	27,3	44,1	64,6	70,8	151,9	32,8	108.274
	Araucanía	238	3.312	6,9	20,9	12,0	14,1	20,1	30,3	44,7	56,9	62,5	125,2	34,2	105.277
	Los Lagos	246	3.382	7,3	13,1	10,4	12,1	16,3	22,9	30,4	38,2	50,0	138,5	25,8	81.030
	Aysén del General Carlos Ibañez del Campo	350	1.566	2,6	5,9	3,2	3,4	4,2	5,7	9,3	16,0	18,6	25,1	7,6	9.243
	Magallanes y La Antártica Chilena	118	422	3,8	7,5	3,8	4,1	7,3	7,5	11,1	13,1	16,7	26,3	9,4	2.861
	Metropolitana de Santiago	253	963	19,4	48,3	24,3	29,7	46,6	61,5	92,9	125,0	141,8	171,3	71,0	50.395
	Los Ríos	677	2.809	3,0	18,9	6,4	8,2	10,1	14,0	18,9	24,1	27,6	44,0	15,2	32.388
	Arica y Parinacota	33	442	3,1	16,3	4,2	4,4	5,9	10,3	12,5	16,9	19,8	19,8	10,2	4.189
Total País		28.628	112.867	2,6	5,9	6,5	9,2	15,0	29,8	61,5	123,7	180,4	5.931	56,9	4.794.266

6. Ponderación de calibración

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

Con este objetivo, para cada Encuesta Casen se solicita al INE que prepare proyecciones de población al 30 de noviembre del año de la Casen en curso. Para 2011, estas proyecciones se solicitaron para las 324 comunas en la muestra y además para las 30 divisiones dadas por el cruce de región y área (urbano, rural). Las proyecciones de población preparadas por el INE son desarrolladas a partir de modelos demográficos que modelan toda la población en el país, sin diferenciar entre aquellos que residen en viviendas particulares y aquellos que se encuentran en viviendas colectivas.

La población de inferencia de la Encuesta Casen, sin embargo, está restringida a las personas que residen en viviendas particulares ocupadas. En atención a esta discrepancia, las proyecciones de población elaboradas por INE se ajustan por un factor que representa la proporción de personas que reside en viviendas particulares sobre el total de la población. Este factor se estima, para cada estrato, en base a los resultados del último Censo disponible. Por ejemplo, a nivel nacional, la proyección de población al 30 de noviembre de 2011 era de 17.312.726 personas. Según el Censo 2002, la proporción de personas residentes en viviendas particulares es de 97,9%, por lo tanto una estimación del total de la población que reside en viviendas particulares es 16.962.560 personas.

Hasta 2009, el diseño muestral de la Encuesta Casen tenía como objetivo producir estimaciones a nivel comunal y regional. Por lo tanto, se desarrollaron factores de expansión comunales y regionales para cumplir tales propósitos⁴⁵. El diseño de la muestra 2011 fue optimizado para la estimación de las tasas de pobreza a nivel regional. Por comparabilidad, sin embargo, se desarrollaron factores de expansión regionales y comunales.

A continuación se describen los procedimientos para la última etapa de desarrollo de los factores de expansión de Casen 2011. Primero se describe el desarrollo del Ponderador de Calibración Regional y luego el Ponderador de Calibración Comunal.

45 El año 2006 y 2009 se desarrollaron factores de expansión provinciales. El año 2009 se desarrollaron dos sets de factores de expansión, un set que expande a la población en hogares particulares (expr, expc, exp) y otro set que expande a la población total nacional (expr_p, expc_p, exp_p).

6.1. Ponderador de Calibración Regional

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

1. Ajustar las proyecciones de población y estimar el total de personas en hogares particulares. Se solicitó al INE las proyecciones de población para los 30 estratos formados por el cruce de región y área (urbano, rural). Las proyecciones fueron ajustadas, según la proporción de personas residentes en viviendas particulares reportadas por el Censo de Población y Vivienda del año 2002. Este ajuste fue realizado por comuna-área y luego se sumó el total de población en la región y área, utilizando la información disponible en todas las comunas de la región. De este modo se estimó el total de personas residentes en viviendas particulares (N_{RA}) en los 30 estratos formados por el cruce de región-área.
2. Estimar el total de personas en hogares particulares a partir de Casen 2011. En cada vivienda entrevistada se multiplicó en Ponderador Corregido por No Respuesta por el número de personas en la vivienda para obtener una estimación del total de personas residentes en viviendas particulares en cada región-área.
3. Construir el ajuste a población total. Se construyó la razón \hat{R}_{RA} , a partir del cociente entre la estimación del total de personas residentes en viviendas particulares desarrollado a partir de las proyecciones de población (N_{RA} , en paso 1) y la estimación de la Encuesta 2011 (paso 2).

Donde

$$\hat{R}_{RA} = \frac{N_{RA}}{\sum_{h \in RA} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i, \text{elcg, nsp}}} N_{hij} \cdot W_{hij}^{NR}}, \quad R=1, \dots, 15 \text{ y } A = \text{Urbano, Rural} \quad (13)$$

N_{RA} : Corresponde al total de personas residentes en viviendas particulares, en la región R y área A.

N_{hij} : Corresponde al total de personas residentes en la vivienda j, del conglomerado i, en el estrato h.

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

El factor de expansión regional tiene la siguiente expresión:

$$w_{hijk}^{RA} = \hat{R}_{RA} \cdot w_{hij}^{NR} \quad (14)$$

Las figuras VI.14 y VI.15 muestran la distribución de los Ponderadores de Calibración regional para el marco rural y urbano respectivamente. Se puede apreciar que los patrones observados en ponderadores preliminares se mantiene y que los mayores ponderadores se concentran en la región Metropolitana (tanto para el marco urbano como rural). Además, se observa que las unidades provenientes del MM2008 pertenecientes a las regiones Metropolitana, V y VIII son las que poseen mayor variabilidad, lo que se explica, en parte, porque en ellas se concentra gran parte de la población total, así como también existe gran diversidad en estructura de sus unidades (manzanas grandes y chicas).

En ambos gráficos se observa que, en comparación con los ponderadores Corregidos por No Respuesta, los Ponderadores de Calibración Regional de Valparaíso y Aysén se incrementan más que los de otras regiones. Esto se explica porque en estas regiones se concentra parte importante de las comunas no presentes en la muestra (Áreas de Difícil Acceso).

En él se aprecia que las unidades provenientes del MS 2002 sus ponderadores ajustados al stock regional oscila entre 2,2 y 604, mientras que para las unidades provenientes del MM2008 este valor varía entre 2,3 y 5.751.

Figura VI.14: Factor de expansión calibración Región – área – MS2002

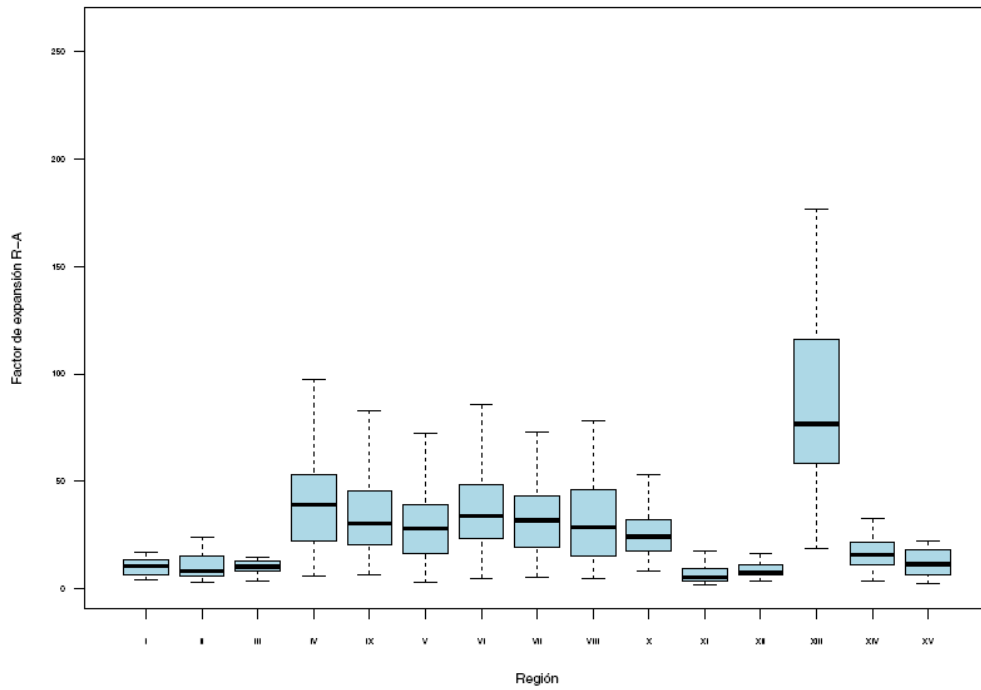


Figura VI.15: Factor de expansión calibración Región – área – MM2008

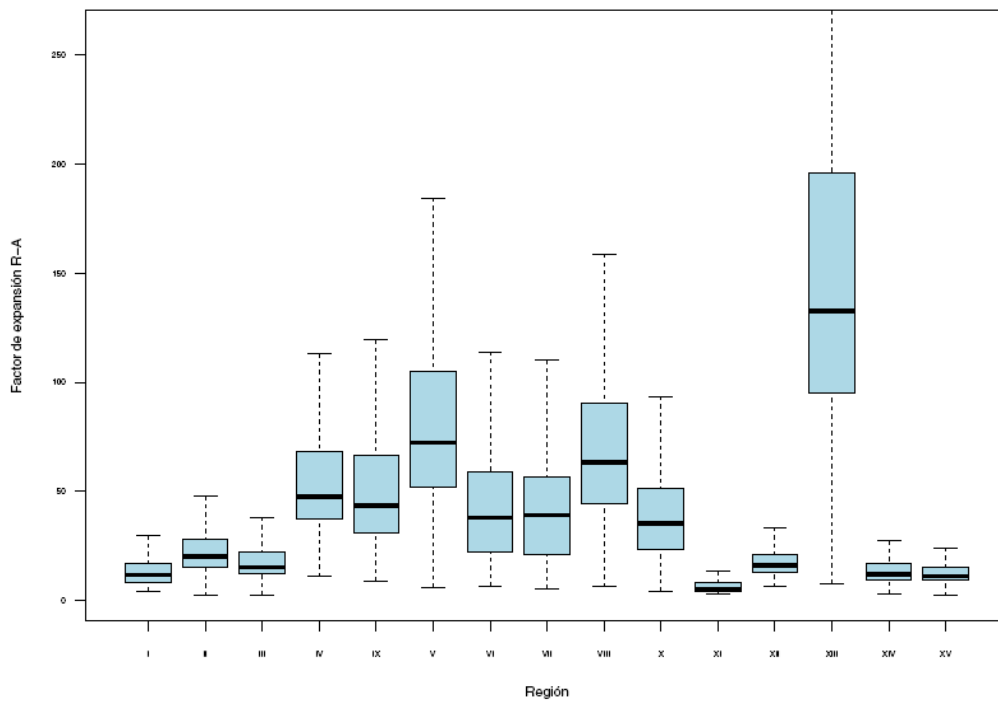


Tabla VI.14: Estadísticas descriptivas, Ponderador de Calibración Regional (Factor de Expansión regional), según marco.

Marco	Región	Factor ajustado a las PP Región-Area - FULL											
		Perdidos	Recuento	Mínimo	Moda	Percentil 10	Mediana	Percentil 90	Percentil 95	Percentil 99	Máximo	Media	Suma
Manzanas	Tarapacá	2.184	6.566	4,5	8,2	7,2	11,7	29,3	40,8	80,2	979,9	16,5	72.312,3
	Antofagasta	2.152	6.802	2,3	22,6	12,2	20,0	40,9	55,6	140,6	1.734,7	27,3	127.018,7
	Atacama	1.915	5.303	2,6	24,6	8,5	15,5	31,3	42,4	92,9	880,2	19,8	66.966,4
	Coquimbo	477	2.928	11,1	138,8	23,8	47,7	96,8	138,8	329,1	3.224,1	63,3	155.250,0
	Valparaíso	1.585	6.754	6,3	48,2	34,2	72,1	157,1	223,0	493,2	1.350,7	93,3	482.478,8
	Libertador General Bernardo OHiggins	493	3.600	6,9	68,2	14,1	38,3	95,2	127,3	301,9	3.838,1	54,9	170.644,3
	Maule	505	4.148	5,6	68,2	15,3	39,1	82,4	106,8	170,2	1.947,3	46,6	169.795,7
	Biobio	987	6.213	6,4	67,3	30,9	63,4	149,4	205,4	416,1	4.005,0	87,5	457.076,7
	Araucanía	604	3.404	9,0	43,5	22,9	43,5	94,5	133,5	295,3	3.004,7	60,9	170.560,9
	Los Lagos	669	4.157	4,4	29,6	15,0	35,4	74,5	110,2	272,1	1.295,6	46,4	161.827,9
	Aysén del General Carlos Ibañez del Cam	890	3.822	2,8	3,7	3,9	5,5	12,5	16,9	44,4	125,3	7,8	22.736,0
	Magallanes y La Antártica Chilena	1.063	3.147	6,5	16,1	10,2	16,2	32,4	40,5	113,3	617,6	21,5	44.769,6
	Metropolitana de Santiago	4.129	14.202	7,8	74,0	67,9	133,0	314,7	438,5	1.062,1	5.750,9	180,9	1.822.540,4
	Los Ríos	591	4.826	3,3	7,8	8,1	12,2	25,0	35,9	103,6	813,4	17,3	73.212,4
Arica y Parinacota	744	3.456	2,4	7,4	7,9	11,2	23,9	34,6	71,5	812,9	15,3	41.589,5	
Secciones	Tarapacá	75	535	4,1	10,2	5,0	10,7	14,7	16,9	41,6	41,6	11,4	5.262,2
	Antofagasta	75	349	3,2	3,6	3,6	8,5	24,3	33,1	33,1	33,1	11,6	3.173,8
	Atacama	131	635	3,6	10,3	7,3	10,3	14,5	34,8	41,8	43,8	12,0	6.028,9
	Coquimbo	103	1.237	5,8	31,3	11,4	39,2	60,6	74,0	97,5	97,5	38,7	43.941,5
	Valparaíso	124	1.548	3,1	5,6	8,5	28,3	68,5	83,9	109,3	198,9	33,6	47.859,7
	Libertador General Bernardo OHiggins	249	2.553	4,9	28,5	13,8	33,8	63,4	72,7	94,0	141,4	37,3	86.020,9
	Maule	153	3.669	5,5	28,4	11,2	31,9	52,7	63,2	93,0	604,1	35,2	123.810,0
	Biobio	284	3.586	4,6	5,8	9,7	28,8	68,1	74,6	122,3	160,1	34,0	112.426,5
	Araucanía	154	3.228	6,8	21,3	14,2	30,9	58,1	63,8	123,8	124,2	34,8	106.911,6
	Los Lagos	136	3.272	8,4	13,9	12,8	24,3	40,6	53,0	88,7	146,9	27,8	87.030,3
	Aysén del General Carlos Ibañez del Cam	224	1.440	2,2	4,8	2,7	5,5	14,3	16,2	20,6	20,6	6,9	8.366,3
	Magallanes y La Antártica Chilena	88	392	3,7	7,5	4,0	7,5	13,1	16,2	25,5	25,5	9,3	2.819,0
	Metropolitana de Santiago	208	918	18,8	60,3	35,4	76,7	155,9	176,9	213,6	213,6	87,7	62.295,9
	Los Ríos	308	2.440	3,4	21,8	9,2	16,0	27,8	31,9	41,2	50,8	17,3	36.964,3
Arica y Parinacota	21	430	2,5	18,2	4,7	11,6	18,9	22,1	22,1	22,1	11,3	4.611,2	
Total País		21.321	105.560	2,2	4,8	8,9	30,3	122,5	177,6	391,7	5.750,9	56,7	4.776.302

6.2. Ponderador de Calibración Comunal

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

- Ajustar las proyecciones de población y estimar el total de personas en hogares particulares. Se solicitó al INE las proyecciones de población para los 324 estratos formados por las comunas en la muestra Casen 2011. Las proyecciones fueron ajustadas, según la proporción de personas residentes en viviendas particulares reportadas por el Censo de Población y Vivienda del año 2002. Este ajuste fue realizado por comuna-área y luego se sumó el total de población en cada comuna. De este modo se estimó el total de personas residentes en viviendas particulares (N_c) en los 324 estratos formados por las comunas en la muestra 2011.
- Estimar el total de personas en hogares particulares a partir de Casen 2011. 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 2011.
- Construir el ajuste a población total. Se construyó la razón \hat{R}_c , a partir del cociente entre la estimación del total de personas residentes en viviendas particulares desarrollado a partir de las proyecciones de población (N_c , en paso 1) y la estimación de Casen 2011 (paso 2).
Donde:

$$\hat{R}_c = \frac{N_c}{\sum_{h \in C} \sum_{i \in \Omega_h} \sum_{j \in \Theta_{i, eleg, resp}} N_{hij} \cdot w_{hij}^{NR}} \quad (15)$$

N_c : Corresponde al total de personas residentes en viviendas particulares, en la comuna C.

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

- 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:

$$w_{hijk}^C = \hat{R}_c \cdot w_{hij}^{NR}, \quad (16)$$

Por la forma de construcción, se tiene que todos los individuos residentes en el mismo conglomerado (manzana o sección), comparten el mismo ponderador, el que es interpretado como el total de personas en la comuna de residencia que representa un individuo de la muestra. Por lo tanto, el ponderador puede ser utilizado para realizar análisis de personas y hogares, a nivel de comuna⁴⁶.

Al usar este ponderador, es importante recordar que este expande sólo al total de la población residente en las comunas que participan en la muestra Casen 2011. Esto, a diferencia del ponderador regional que expande a la población nacional (i.e. incluyendo a la población que reside en Áreas de Difícil Acceso).

Las figuras VI.16 y VI.17 muestran la distribución de los ponderadores de Calibración Comunal para las unidades en los marcos rural y urbano respectivamente. Se mantienen las tendencias observadas entre ambos conjuntos de figuras. La región de Aysén posee los ponderadores más pequeños y con menor variabilidad, para las manzanas y secciones. Mientras que en la región Metropolitana se observa mayor variabilidad y ponderadores más grandes. De las figuras VI.16 y VI.17 se desprende que las unidades provenientes del marco de manzanas, en general, poseen mayor variabilidad.

Es importante señalar que existen regiones donde se observa un leve decrecimiento, a nivel agregado, en los ponderadores de la Región de Tarapacá, Antofagasta y Atacama, lo que se debe a que las proyecciones de población determinan un total de personas inferior a lo estimado con la información de la muestra. En el resto de las regiones no se aprecia una diferencia significativa en los factores de expansión entre los percentiles 1 al 99. Sin embargo, en los casos extremos (outliers) se aprecia un fuerte incremento o caída. Por ejemplo, el ponderador de la región Metropolitana tenía un valor de 5.931 unidades previo a la calibración comunal, aumentando a 10.637 posteriormente. Si se comparan los valores extremos con los obtenidos al calibrar por región se evidencia un aumento importante en los rangos y en los valores más grandes, como se observa en la Tabla VI.15.

46 Si bien el ponderador puede ser utilizado a nivel de comuna es necesario antes de realizar el análisis considerar el error estándar asociado, ya que en algunas comunas el tamaño muestral no es suficiente para obtener estimaciones robustas.

Figura VI.16: Distribución del Ponderador de Calibración Regional (Factor de Expansión Comunal), según región (MS2002).

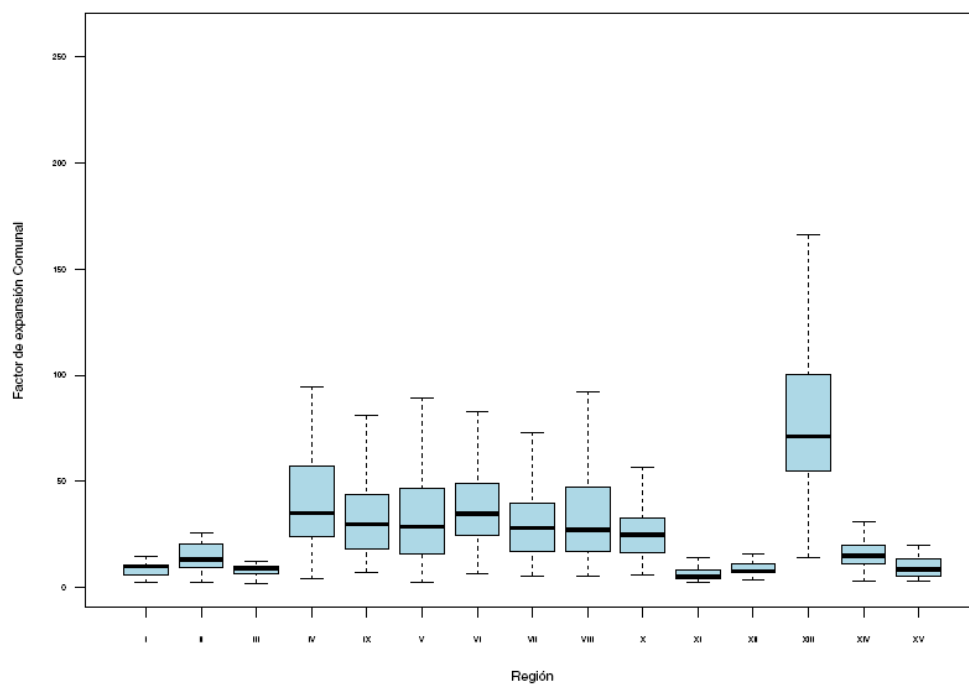


Figura VI.17: Distribución del Ponderador de Calibración Comunal (Factor de Expansión Comunal), según región (MM2008).

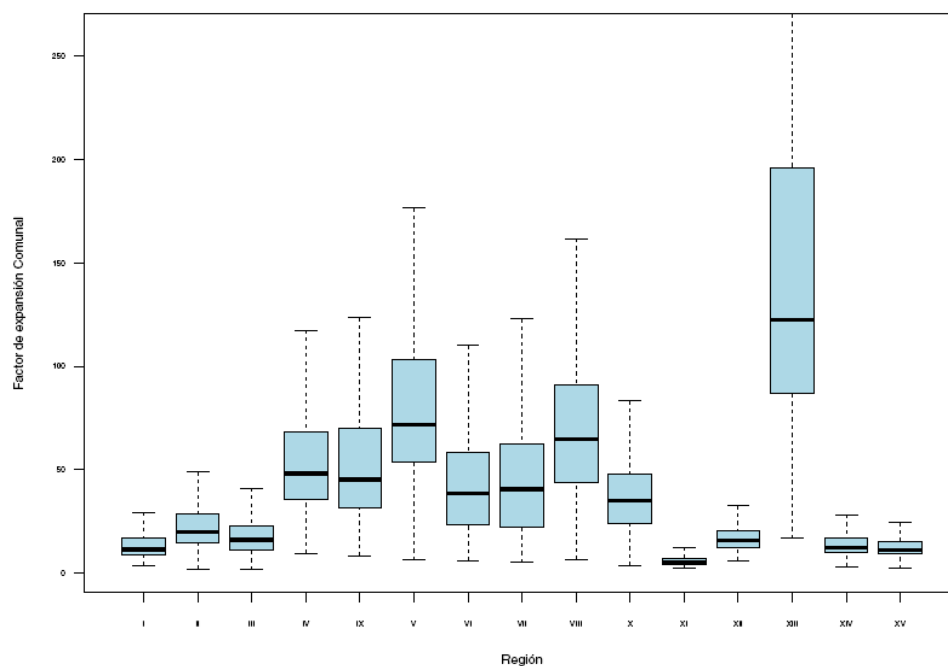


Tabla VI.15: Estadísticas descriptivas básicas del factor de expansión calibrado por comuna, según MM y región

Marco	Región	Factor ajustado a las PP comuna											
		Viviendas No responden	Total Viviendas Elegibles	Mínimo	Moda	Percentil 10	Mediana	Percentil 90	Percentil 95	Percentil 99	Máximo	Media	Total Viviendas Expandidas
Manzanas	Tarapacá	2.184	6.566	4,0	9,4	7,2	11,4	26,4	36,4	80,0	864,8	16,2	71.037,8
	Antofagasta	2.152	6.802	2,0	21,9	11,7	19,7	41,5	55,0	135,9	1.677,0	27,2	126.257,2 ¹
	Atacama	1.915	5.303	1,9	27,0	6,8	16,2	33,3	43,5	101,9	965,0	20,0	67.653,1
	Coquimbo	477	2.928	9,8	143,9	19,9	48,2	99,4	134,7	341,2	3.342,8	62,6	153.511,9
	Valparaíso	1.585	6.754	6,4	50,2	36,8	71,9	154,4	220,3	430,1	1.185,0	92,2	476.699,9
	Libertador General Bernardo OHiggins	493	3.600	6,1	65,1	13,5	38,7	94,8	118,9	324,7	3.663,4	54,5	169.378,5
	Maule	505	4.148	5,2	60,8	17,0	40,2	96,4	116,9	187,4	1.511,9	51,5	187.718,5
	Biobio	987	6.213	6,8	73,9	32,3	64,6	147,4	216,9	419,0	3.604,3	88,0	459.666,1
	Araucanía	604	3.404	8,5	44,6	22,9	45,0	96,8	138,3	308,2	3.081,9	63,3	177.116,4
	Los Lagos	669	4.157	4,0	23,1	15,2	35,0	71,1	102,7	250,3	1.181,0	44,6	155.722,5
	Aysén del General Carlos Ibañez del Campo	890	3.822	2,6	3,3	3,5	5,2	12,0	15,5	48,1	111,5	7,3	21.441,2
	Magallanes y La Antártica Chilena	1.063	3.147	6,3	15,4	9,8	15,9	31,1	39,3	108,6	591,8	20,8	43.379,8
	Metropolitana de Santiago	4.129	14.202	16,9	80,7	65,9	123,0	308,5	443,2	1.021,0	10.636,5	179,8	1.811.075,5
	Los Ríos	591	4.826	3,2	8,0	8,4	12,5	25,8	35,4	101,3	838,3	17,8	75.463,5
Arica y Parinacota	744	3.456	2,4	7,5	8,0	11,4	24,2	35,0	72,4	822,9	15,5	42.098,0	
Secciones	Tarapacá	75	535	2,3	9,8	3,5	10,3	30,0	46,8	57,5	57,5	13,0	5.989,4
	Antofagasta	75	349	2,5	2,5	5,0	13,3	25,7	38,0	38,0	38,0	15,1	4.137,9 ¹
	Atacama	131	635	2,2	9,6	5,5	9,0	12,7	25,1	38,9	40,7	9,8	4.927,9
	Coquimbo	103	1.237	4,0	32,0	10,0	34,1	73,7	91,4	122,4	122,4	40,5	45.969,6
	Valparaíso	124	1.548	2,6	4,7	7,1	28,6	71,6	87,5	132,9	205,6	35,7	50.860,1
	Libertador General Bernardo OHiggins	249	2.553	6,6	28,3	14,6	34,8	67,5	75,4	108,9	123,1	38,1	87.805,0
	Maule	153	3.669	5,7	28,9	11,4	28,1	49,2	55,2	79,4	266,1	30,4	106.710,9
	Biobio	284	3.586	5,2	8,9	9,1	27,4	66,6	76,4	99,1	152,5	34,0	112.224,8
	Araucanía	154	3.228	7,3	18,3	12,9	29,8	56,2	67,6	100,4	104,3	32,9	100.990,7
	Los Lagos	136	3.272	6,0	16,0	11,5	24,9	42,4	57,6	87,9	145,6	27,9	87.555,0
	Aysén del General Carlos Ibañez del Campo	224	1.440	2,6	7,1	3,3	5,4	13,6	14,4	24,0	24,0	7,2	8.707,9
	Magallanes y La Antártica Chilena	88	392	3,7	7,8	3,9	7,8	13,5	16,0	25,1	25,1	9,5	2.884,7
	Metropolitana de Santiago	208	918	14,2	45,4	23,5	71,1	132,2	166,0	195,7	195,7	78,7	55.849,7
	Los Ríos	308	2.440	3,1	18,7	8,4	15,2	25,1	28,1	43,7	47,4	16,3	34.765,6
Arica y Parinacota	21	430	3,4	13,1	4,7	8,6	15,9	15,9	19,8	19,8	9,6	3.907,5	
Total País		21.321	105.560	1,88	7,13	8,76	29,80	118,09	175,74	410,57	10636,55	56,41	4751506,38

VII. Estimación de Varianza

Casen es una encuesta con probabilidades de selección desigual. Por eso es que los análisis se deben hacer con factores de expansión. Si no se usan factores de expansión, las viviendas con mayor probabilidad de selección pesan más en la muestra y distorsionan los resultados. Pero si se usan los factores, cada persona en la muestra pesa lo que corresponde para representar a su región (o comuna).

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

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

1. El Estimador

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

Supóngase que de una población U particionada en N unidades, $U = \{ 1, 2, 3, \dots, N \}$ se toma una muestra probabilística S de tamaño de tamaño n . La probabilidad de inclusión de la unidad k está dada por $\pi_k = P(k \in S) > 0$ (diseño

$$w_k = \frac{1}{\pi_k}$$

aleatorio), donde el término $\frac{1}{\pi_k}$ es el inverso de la probabilidad de inclusión o ponderador del diseño asociado al individuo k . La probabilidad que los elementos k y q pertenezcan a la muestra, está dada por

$\pi_{kq} = P(k \in S \text{ y } q \in S) > 0$, siendo $w_{kq} = \frac{1}{\pi_{kq}}$ el ponderador asociado a la inclusión conjunta de los elementos k y q en la muestra.

Algunos de los parámetros que se desea estimar son: $t = \sum_{k \in U} y_k$, el total poblacional de la variable de interés Y , y la variable $\bar{y}_U = \sum_{k \in U} y_k / N$, la media poblacional de la variable Y .

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

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

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

Donde:

- w_i = es el peso o ponderador del i -ésimo elemento según el diseño e indica a cuántos elementos representa en la población y
- $\hat{N} = \sum_{i \in S} w_i$ = es el tamaño de la población estimada.

En **muestreo aleatorio simple** para poblaciones finitas, donde se extrae una muestra aleatoria de n elementos de una población que contiene N de ellos, el

ponderador muestral es $w_k = \frac{N}{n}$, y el estimador de Horvitz-Thompson para el Total y es:

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

Siendo:

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

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

$$\hat{t}_{HT} = \sum_{h=1}^H \hat{t}_{HT_h} = \sum_{h=1}^H N_h \cdot \hat{\bar{y}}_h \quad (4)$$

Donde:

- $\hat{\bar{y}}_h = \sum_{k \in S_h} \frac{y_k}{n_h}$ = la media muestral en el estrato h , y
- n_h el número de elementos en la muestra dentro del estrato h .

Los **estimadores calibrados** son usados frecuentemente para obtener estimaciones más precisas, siendo necesario contar con información auxiliar correlacionada con las variables de interés y en donde los totales son conocidos o son fijados a nivel poblacional. La Encuesta Casen utiliza la calibración para ajustar las estimaciones de personas a los stocks poblacionales estimadas por medio de proyecciones que son fijadas como las poblacionales. Adicionalmente, utiliza otros ajustes como la corrección por no respuesta, de modo que el factor de expansión contiene varios componentes que son necesarios para la obtención de las estimaciones.

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

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

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

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

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

con $\hat{t}_{r_h} = \sum_{k \in S} w_k \cdot r_k$, siendo $r_k = y_k - \hat{R}_{\text{cal},y/z} \cdot z_k$

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

2. Varianza del Estimador

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

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

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

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

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

$$\text{son } w_k = \frac{N}{n} \text{ y } w_{kq} = \frac{N}{n} \cdot \frac{N-1}{n-1}.$$

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

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

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

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

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

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

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

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

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

Siendo $S_{S_h}^2 = \sum_{k \in S_h} (y_k - \bar{y}_h) / (n_h - 1)$ la varianza muestral en el estrato h, $f_h = n_h / N_h$ es la fracción de muestreo en el estrato h, y $(1-f_h) = fpc_h$ es la corrección por finitud o factor de ajuste para poblaciones finitas en el estrato h.

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

$$\widehat{Var}(\hat{t}_{cal}) = \sum_{h=1}^H \sum_{k \in S_h} \frac{(y_k \cdot w_k \cdot n_h - \hat{t}_h)^2}{n_h \cdot (n_h - 1)} \quad (12)$$

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

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

- Se asume que las unidades primarias de muestreo, en los diseños multi-etápicas, fueron seleccionadas con remplazo a partir de los estratos de muestreo. Se ignora cualquier corrección por población finita. Como resultado, las estimaciones de varianza estarán levemente sobre-estimadas.
- Se ignoran las etapas de selección al interior de la unidad primaria de muestreo. Es decir, se asume que las unidades fueron seleccionadas en una etapa al interior del conglomerado, por lo tanto la fórmula de estimación de la varianza sólo requiere el identificador de los las unidades primarias de muestreo y los estratos primarios. Toda la variabilidad al interior de la unidad primaria se captura en el estimador compuesto de varianza.

3. Variables que identifican el diseño muestral complejo

Un diseño de muestreo complejo se origina cuando interviene más de una etapa de selección, o se genera un muestreo polietápico o en varias etapas. Esta selección en más de una etapa suele ser generalmente además estratificado, 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 variables de interés.

La encuesta Casen 2011 presenta un diseño muestral que se puede caracterizar como probabilístico, estratificado y multietápico. En el muestreo estratificado la estimación de la varianza total se obtiene mediante la suma de las varianzas estimadas en cada estrato y la varianza en cada estrato se estima a partir de la variabilidad de la característica de interés entre los conglomerados.

En la base de datos de la Encuesta Casen, la variable que identifica los estratos "originales" utilizados en la selección de la muestra se denomina <Estrato> y la variable que identifica los conglomerados "originalmente" seleccionados en la muestra se denomina <Segmento>. En total considerando la muestra completa (submuestras 1 y 2), la muestra 2011 consideró 585 estratos y 10.479 conglomerados (9.236 manzanas y 1.243 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 2011 estableció, para la muestra completa, un mínimo de 2 conglomerados por estrato y un mínimo de 30 y 50 viviendas

por conglomerado para aquellos rurales (secciones) y urbanos (manzanas) respectivamente.

La implementación de la muestra 2011, sin embargo, se llevó a cabo en dos submuestras independientes con la finalidad de producir estimaciones oficiales a partir de la submuestra 2, la cual concentra un 67% de los conglomerados y viviendas de la muestra completa. Por otra parte, pérdidas asociadas a elegibilidad de las viviendas seleccionadas y diversas formas de no respuesta, implicó en algunas zonas una drástica reducción en la cantidad de viviendas de algunos conglomerados. Como consecuencia del menor tamaño muestral, se observan casos de estratos con 1 sólo conglomerado (en adelante, estratos unitarios) tanto en la submuestra 1 como en la submuestra 2. En esta última, hay 94 estratos unitarios. Ver Tabla VIII.2 en Anexo. El caso más crítico se observó en la comuna de El Quisco, donde inicialmente se seleccionaron 59 viviendas en 8 manzanas, sin embargo, sólo se logró entrevistar 18 viviendas en 6 manzanas (en dos de esas manzanas, sólo se logró entrevistar a 1 vivienda).

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 2011 y producir estimaciones de varianza más estables. Las estimaciones oficiales de la varianza compleja de Casen 2011 serán desarrolladas utilizando pseudo-estrato, pseudo-conglomerado y corrección por poblaciones finitas. A continuación se describen los procedimientos seguidos para la construcción de estas variables.

3.1. Creación de pseudo-estratos (*varstrat*)

La creación de pseudo-estratos tiene como objetivo corregir el problema de estimación de varianza que generan los estratos unitarios (1 conglomerado por estrato), el cual se traduce en una subestimación de la varianza de cualquier variable de interés. Los pseudo-estratos, en adelante *varstrat*, son una reagrupación de los 583 estratos de muestreo originales (agrupaciones de comuna-área) en donde 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, uniendo cada uno de los 94 estratos unitarios junto a otros estratos logran que en conjunto contengan al menos dos conglomerados. El procedimiento es como sigue:

1. Si el estrato unitario se encuentra en una provincia y área que posee dos o más estratos, se combina con otro estrato al interior de la provincia y área.

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

De un total de 585 estratos muestrales se crearon 170 varstrat, los que poseen al menos dos conglomerados. La Tabla VII.1 presenta el total de estratos y varstrat por región y submuestra.

Tabla VII.1: Total de estratos y varstrat según región y submuestra.

Región	Total de Estratos			Total de VarStrat		
	Replica 1	Replica 2	Total	Replica 1	Replica 2	Total
Arica y Parinacota	4	4	4	2	2	2
Tarapacá	8	9	9	5	5	5
Antofagasta	10	12	12	5	5	5
Atacama	14	15	15	6	6	6
Coquimbo	27	28	28	8	8	8
Valparaíso	61	67	67	15	15	15
Libertador General Bernardo OHiggins	60	64	64	16	16	16
Maule	57	60	60	15	15	15
Biobio	95	103	104	21	21	21
Araucanía	58	64	64	14	15	15
Los Lagos	43	48	48	14	14	14
Los Ríos	24	24	24	11	11	11
Aysén del General Carlos Ibañez del Campo	11	11	11	5	5	5
Magallanes y La Antártica Chilena	5	5	5	3	3	3
Metropolitana de Santiago	67	69	70	29	29	29
Total	544	583	585	169	170	170

3.2. Creación de pseudo-conglomerados (varunit)

La creación de pseudo-conglomerados tiene como objetivo estabilizar las estimaciones de varianza, la cual tiende a incrementarse cuando las estimaciones se basan en conglomerados con muy pocas unidades muestrales. Es importante recordar que tanto por diseño como por falta de respuesta existen conglomerados pequeños en la muestra. Por una parte en manzanas con 8 viviendas sólo 2 se seleccionan en la muestra, por otra parte por efecto de no respuesta un conglomerado con 15 viviendas seleccionadas puede terminar con solo 8 entrevistadas. Los pseudo-conglomerados, en adelante varunit, son una reagrupación de los 10.649 conglomerados originales, en donde cada uno contiene un número mínimo de viviendas.

Los varunit fueron construidos de acuerdo a criterios de orden jerárquico al interior de cada varstrat, según el total de viviendas entrevistadas en la submuestra 2. El procedimiento consistió en unir cada manzana o sección con menos de 30 viviendas entrevistadas, las cuales correspondieron al 99% de los conglomerados en la submuestra 2. El procedimiento utilizado es como sigue:

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 la cantidad de viviendas sigue siendo inferior a 30, o la cantidad de varunit es igual a 1 el varstart en el cual se realiza la agrupación de conglomerados, se une el varstrat con otro 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 135 conglomerados quedaban por debajo del criterio establecido (menos de 30 viviendas), lo que se debió a que el total de viviendas que respondieron en dichos conglomerados se encontraban muy cercano a 30 (superior a 24 e inferior a 30), por lo tanto al unir cualquiera de estos conglomerados a otro, el total de viviendas superaba el umbral establecido. Ante esto, se flexibilizaron los umbrales originales, obteniéndose con esto que el total de viviendas que concentran los varunit oscilen entre 28 y 62 viviendas.

A partir de un total de 10.649 conglomerados seleccionados en la muestra Casen 2011, se crearon 2.326 varunit, los cuales agrupan entre 28 y 62 viviendas entrevistadas en la submuestra 2. La Tabla VII.2 presenta el total de conglomerados y varunit por región y submuestra.

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

Región	Total Manzanas y secciones			Total de VarUnit		
	Replica 1	Replica 2	Total	Replica 1	Replica 2	Total
Arica y Parinacota	178	342	520	29	63	92
Tarapacá	243	517	760	39	104	143
Antofagasta	251	606	857	39	108	147
Atacama	331	639	970	35	83	118
Coquimbo	166	305	471	33	66	99
Valparaíso	313	577	890	56	124	180
Libertador General Bernardo OHiggins	194	360	554	49	93	142
Maule	214	422	636	61	124	185
Biobío	314	597	911	75	146	221
Araucanía	198	398	596	50	102	152
Los Lagos	224	417	641	58	114	172
Los Ríos	235	509	744	55	118	173
Aysén del General Carlos Ibañez del Campo	134	314	448	35	77	112
Magallanes y La Antártica Chilena	185	328	513	25	47	72
Metropolitana de Santiago	388	750	1.138	90	197	287
Total	3.568	7.081	10.649	729	1.566	2.295

4. Estimación de Varianzas en Muestra Compleja Casen 2011

Existen al menos tres estrategias para incorporar el diseño complejo en la estimación de la varianza, la primera es simplemente multiplicar el error estándar por un estimador (externo) de un efecto diseño, la segunda estrategia es que, al estimar un modelo se incluyan en él las variables del diseño muestral y la última es utilizar software o paquetes estadísticos que produzcan directamente estimaciones de las varianzas que den cuenta del diseño complejo de la muestra.

Tanto Stata como SPSS poseen rutinas que permiten desarrollar estimaciones de varianza bajo muestreo complejo. Ambos paquetes estadísticos entregan una variedad de opciones para obtener las estimaciones de los parámetros de interés (medias, proporciones, razones y totales) mediante análisis de frecuencias, descriptivos, Tablas de contingencia, razones, así como análisis, empleando el modelo lineal general como también los modelos de regresión logística, ordinal y de Cox. En cada análisis se pueden obtener estadígrafos como errores estándares, coeficientes de variación, efectos de diseño, intervalos de confianza, etc. En todos estos análisis, tanto Stata como SPSS asumen el diseño muestral definido en forma previa por el analista.

Como ya se ha mencionado, la encuesta Casen 2011 presenta un diseño muestral que se puede caracterizar como probabilístico, estratificado y multietápico. Las variables que permiten caracterizar este diseño muestral con el identificador de estratos, el identificador de conglomerados, la fracción de muestreo y el factor de expansión. Todas estas variables están disponibles en la base de datos Casen 2011-Submuestra Noviembre 2011-Enero 2012.

Existen varias opciones para estimar varianzas complejas a partir de las variables que identifican el diseño complejo de la encuesta Casen. Las diferencias entre las distintas opciones están relacionadas con los supuestos asociados a cada una. Distintos paquetes estadísticos tienen incorporadas más o menos opciones para realizar estimaciones de varianza bajo muestreo complejo.

Todas las opciones, en menor o mayor medida, permiten incorporar las características de un diseño complejo y por lo tanto permiten desarrollar estimaciones adecuadas - en contraste con las estimaciones bajo el supuesto de muestreo aleatorio simple (que es la opción implementada por defecto en todos los paquetes estadísticos). A continuación se presentan resultados de 6 distintas opciones de estimación de la varianza bajo muestreo complejo.

La Tabla VII.3 muestra las variantes de las 4 operacionalizaciones empleadas en Stata 11:

- Los métodos 1 y 2 utilizan los 583 estratos (estrato) y los 6.997 conglomerados (segmento) asociados al diseño muestral original. El método 2 incorpora la corrección por poblaciones finitas (fmr2).
- Los métodos 3 y 4 utilizan 170 pseudo-estratos (varstrat) y 1.566 pseudo-conglomerados (varunit) para aproximar las características del diseño muestral original. El método 2 incorpora la corrección por poblaciones finitas (fmr2).

La Tabla VII.4 muestra las 2 operacionalizaciones empleadas en SPSS 19:

- El método 5 utiliza los 583 estratos (estrato) y los 6.997 conglomerados (segmento) asociados al diseño muestral original.
- El método 6 utiliza 170 pseudo-estratos (varstrat) y 1566 pseudo-conglomerados (varunit) para aproximar las características del diseño muestral original.

Tabla VII.3: Opciones para estimación de varianza bajo muestreo complejo en Stata 11.

Método	Conglomerado		Estratos		Opciones Stata 11	
	Variable	No.	Variable	No.	singleunit	fpc
1.SE	segmento	6.997	estrato	583	certainty	-
2.SE, fpc	segmento	6.997	estrato	583	certainty	fmr2
3.VV	varunit	1.566	Varstrat	170	-	-
4.VV, fpc	varunit	1.566	Varstrat	170	-	fmr2

Tabla VII.4: Opciones para estimación de varianza bajo muestreo complejo en SPSS 19

Método	Conglomerado		Estratos		Opciones SPSS 19	
	Variable	No.	Variable	No.	singleunit	fpc
5.SE	segmento	6.997	estrato	583	certainty	-
6.VV	varunit	1.566	varstrat	170	-	n/a*

*n/a= no disponible en SPSS

Para efectos de la determinación de la estimación de la pobreza a nivel regional y nacional y su margen de error en la Encuesta Casen 2011, mostraremos las sintaxis o comandos que deben utilizarse para la creación del plan de muestreo y para la obtención de la Tabla con los resultados regionales de la pobreza, medida en porcentaje de pobres respecto al total de la población (personas pobres y no pobres). Se solicita adicionalmente el error típico o estándar.

En Stata 11, la definición del plan de muestreo se realiza a partir del comando `svyset`. La figura VII.1 presenta la sintaxis utilizada para definir los métodos 1, 2, 3 y 4. En SPSS 19, la definición del plan de muestreo se realiza mediante el comando `CSPLAN ANALYSIS`⁴⁷. La figura VII.2 presenta la sintaxis utilizada para definir los métodos 5 y 6. Los resultados de las estimaciones realizadas en base a los 6 métodos, para el nivel nacional, se presentan en la Tabla VII.5. Los resultados para las regiones Atacama y Aysén se presentan en las Tablas VII.6 y VII.7 respectivamente.

5. Resultados Nivel Nacional

La Tabla VII.5 muestra resultados para estimaciones asociadas al porcentaje de pobres (\hat{p}) a nivel nacional junto también con las estimaciones del error estándar $se(\hat{p})$ y el margen de error $moe(\hat{p})$ al 95% de confianza.

Las estimaciones fueron desarrolladas utilizando los softwares Stata 11 (los métodos 1, 2, 3 y 4) y SPSS 19 (los métodos 5 y 6). Para la estimación de la tasa de pobreza a nivel nacional, todos los métodos entregan el mismo resultado, independientemente del software que se utilice y de las unidades de estratificación y conglomeración.

⁴⁷ Con SPSS no es factible indicarle al programa que considere en el cálculo la corrección por finitud, a menos que las unidades de conglomeración o conglomerados hayan sido seleccionadas en forma aleatoria simple.

Para la estimación del error estándar de la tasa de pobreza, se observa que no existen diferencias usando STATA o SPSS para los métodos comparables. Tanto STATA como SPSS entregan exactamente las mismas estimaciones bajo la utilización de las mismas unidades de estratificación y conglomeración (SE ó VV), por lo que se puede hablar de equivalencia en uso de software entre los métodos 1 con 5 y 3 con 6, lo que el análisis se reduce a la comparación de los métodos del 1 al 4 en STATA.

Al considerar la misma estratificación y conglomeración SE (Segmento-Estrato) ó VV (Varunit-Varstrat), la diferencia en la estimación del error estándar es prácticamente nula en la comparación de los métodos 1 con 2 (SE sin y con fpc) y 3 con 4 (VV sin y con fpc). Uno esperaría que la incorporación de la corrección por finitud disminuyera el error estándar, sobre todo en aquellos estratos con pocas unidades de conglomeración en el Marco y donde la muestra representa una fracción de muestreo sobre el 5%. Sin embargo, a nivel nacional este efecto es absorbido por todos aquellos estratos con fracciones de muestreo muy por debajo del 5% y que en su conjunto, a nivel nacional, disminuyen esta fracción de muestreo.

Al comparar los métodos que utilizan distinta estratificación y unidades de conglomeración, pero bajo el mismo criterio de incorporación de la corrección por finitud, es decir la comparación de los métodos 1 con 3 y 2 con 4 respectivamente, observamos en ambos casos un aumento de un 4,2% en la estimación del error estándar. Este aumento se explica por los supuestos subyacentes en los procedimientos de estimación de los métodos 1, 2 (STATA) y 5 (SPSS) los cuales subestiman la verdadera varianza al considerar todos aquellos estratos con sólo una unidad de conglomeración como incluidos con certeza y por lo tanto con varianza cero (cero aporte a la varianza). En contraste, las nuevas variables de estratificación y conglomeración (VV) son más robustas y no asumen este supuesto, por lo tanto se espera que produzcan una estimación más cercana al verdadero error de muestreo (la cual puede ser mayor o menor que la varianza estimada con las unidades originales de muestreo SE).

Tabla VII.5. Estimaciones Tasa de Pobreza y Error estándar a nivel nacional.

Nivel	Software	Metodo	p	se(p)	moe(p) 95
Nacional	Stata 11	1. SE	14,45	0,40018	0,784
Nacional	Stata 11	2. SE, fpc	14,45	0,39850	0,781
Nacional	Stata 11	3. VV	14,45	0,41698	0,817
Nacional	Stata 11	4. VV, fpc	14,45	0,41520	0,814
Nacional	SPSS 19	5.SE	14,45	0,40018	0,784
Nacional	SPSS 19	6.VV	14,45	0,41698	0,817

6. Resultados Nivel Regional

Las Tablas VII.6 y VII.7 muestran a modo de ejemplo los resultados para estimaciones asociadas a la tasa de pobreza regional para las regiones de Atacama y Aysén. Al igual que a nivel nacional, los 6 métodos presentan los mismos resultados para las estimaciones de la tasa de pobreza (\hat{p}). Esto significa que independientemente del software que se utilice o de las unidades de estratificación y conglomeración definidas, las estimaciones de las tasas de pobreza regional resultan idénticas.

En cuanto a las estimaciones del error estándar, tanto STATA como SPSS entregan exactamente las mismas estimaciones bajo la utilización de las mismas unidades de estratificación y conglomeración (SE ó VV), por lo que se puede hablar de equivalencia en uso de software entre los métodos 1 con 5 y 3 con 6, lo que el análisis se reduce a la comparación de los métodos del 1 al 4 en STATA.

A diferencia del nivel nacional, las estimaciones con y sin corrección por poblaciones finitas presentan algunas diferencias. Este resultado resulta esperable ya que la muestra fue optimizada para alcanzar niveles de precisión determinados al nivel regional. En consecuencia, la adecuada especificación de estas características del diseño muestral permite obtener ganancias en términos de precisión a nivel regional cuando estas se especifican como parte del diseño de la muestra al momento de estimar la varianza compleja. Para la región de Atacama, la corrección por finitud disminuye el error estándar del método 1 de 1,021 a 0,999 (- 2,1%) y del método 3 de 1,255 a 1,229 (-2,1%), mientras que para la región de Aysén esta corrección disminuye el error estándar del método 1 de 0,989 a 0,938 (-5,2%) y del método 3 de 0,906 a 0,859 (-5,2%).

Las Tablas VII.6 y VII.7 muestran que la incorporación de pseudo estratos y pseudo conglomerados puede aumentar o disminuir el error estándar. Para la región de Atacama la incorporación de pseudo-estratos y pseudo-conglomerados aumentan la estimación del error estándar en un 23% (de 1,021 a 1,255 ó alternativamente de 0,999 a 1,229), en la región de Aysén lo disminuyen en un 8,4% (de 0,989 a 0,906 ó alternativamente, de 0,938 a 0,859). La pseudo-estratificación y pseudo-conglomeración simultánea corrige la conformación de los estratos de muestreo y de conglomerados, permitiendo que en cada uno de ellos haya tantas unidades de conglomeración como de viviendas de modo que la estimación de la varianza se pueda calcular en forma óptima, acercándose más precisamente al verdadero valor poblacional.

Tabla VII.6: Estimaciones Tasa de Pobreza y Error estándar, Región de Atacama.

Nivel	Software	Método	p	se(p)	moe(p)95
Región 3	Stata 11	1. SE	13,3	1,021	2,001
Región 3	Stata 11	2. SE, fpc	13,3	0,999	1,959
Región 3	Stata 11	3. VV	13,3	1,255	2,460
Región 3	Stata 11	4. VV, fpc	13,3	1,229	2,408
Región 3	SPSS 19	5. SE	13,3	1,021	2,001
Región 3	SPSS 19	6. VV	13,3	1,255	2,460

Tabla VII.7: Estimaciones Tasa de Pobreza y Error estándar, Región de Aysén.

Nivel	Software	Método	p	se(p)	moe(p)95
Región 11	Stata 11	1. SE	9,8	0,989	1,939
Región 11	Stata 11	2. SE, fpc	9,8	0,938	1,838
Región 11	Stata 11	3. VV	9,8	0,906	1,776
Región 11	Stata 11	4. VV, fpc	9,8	0,859	1,684
Región 11	SPSS 19	5. SE	9,8	0,989	1,939
Región 11	SPSS 19	6. VV	9,8	0,906	1,776

La Tabla VII.8 presenta los resultados de la estimación de la tasa de pobreza nacional y sus márgenes de error (al 95% de confianza) para las muestras Casen 1996 a 2011-S2 (submuestra 2). El margen de error de 2011 fue estimado en Stata 11 utilizando los parámetros del método 4 (varstrat, varunit, expr_r2, fmr2), considerando la utilización de pseudo-estratos y pseudo-conglomerados. Los márgenes de error de los años 1996 a 2009 fueron estimados en Stata 11 utilizando los parámetros del método 3 (estrato, segmento, expr_r2), ya que no se dispone para la serie histórica de pseudo-estratos (varstrat) y pseudo-conglomerados (varunit).

Tabla VII.8: Incidencia de la Pobreza y de la Pobreza Extrema, 1996-2011 (Estimación (%) y Margen de Error)

Año	Pobreza		Pobreza Extrema	
	Estimación (%)	Margen de error	Estimación (%)	Margen de error
1996	23,21	0,94	5,75	0,43
1998	21,66	0,75	5,62	0,39
2000	20,22	0,76	5,58	0,38
2003	18,66	0,62	4,67	0,30
2006	13,71	0,58	3,21	0,26
2009	15,13	0,64	3,74	0,30
2011	14,45	0,81	2,79	0,29

7. Márgenes de Error Proyectados, Efectivos e Históricos

La muestra de la Encuesta Casen 2011 fue optimizada al nivel regional, esto quiere decir que su diseño buscaba para lograr - en base a la muestra completa - márgenes de error entre 1 y 4 puntos porcentuales en cada región del país. La columna (2) en Tabla VII.8 muestra los márgenes de error proyectados para cada región en base a la muestra completa (n=90.122 hogares). Las regiones Metropolitana y de Los Ríos presentan los mínimos y máximos márgenes de error proyectados.

Los errores efectivos en base a la submuestra 2 (n=59.084 hogares) se presenta en la columna (4). A pesar de que esta muestra corresponde sólo al 67% de la muestra completa, los márgenes de error a nivel regional se mantuvieron dentro de los límites esperados de 1 a 4 puntos porcentuales. Para algunas regiones, sin embargo, los márgenes de error fueron incluso menores a los proyectados para la muestra completa. Destacan las regiones de El Maule (proyectado: 3,2 y efectivo: 1,65), Los Lagos (proyectado: 2,7 y efectivo: 1,87) y Los Ríos (proyectado: 4,0 y efectivo: 2,94).

A nivel nacional el margen de error efectivo (0,82 pp) fue mayor al proyectado (0,67 pp). Este es el resultado de contar, para la submuestra 2, sólo con el 65% del total de entrevistas proyectadas para la muestra completa (=59.084/90.122 hogares).

Tabla VII.8: Distribución Regional, muestra CASEN 2011.

Regiones	Muestra Completa		Submuestra 2	
	Entrevistas Proyectadas (1)	MOE Proyectado (2)	Entrevistas Efectivas (3)	MOE Efectivo (4)
1 Tarapacá *	6.241	2,4%	3.675	2,3%
2 Antofagasta	6.108	1,6%	3.905	2,0%
3 Atacama	4.382	2,7%	2.823	2,5%
4 Coquimbo	3.780	2,5%	2.418	2,6%
5 Valparaíso	8.064	1,9%	4.596	2,6%
6 O´Higgins	5.469	1,9%	3.656	2,2%
7 Maule	7.464	3,2%	4.987	1,7%
8 Bío Bío	8.652	2,1%	5.779	3,7%
9 La Araucanía	5.782	2,8%	3.995	2,4%
10 Los Lagos **	6.391	2,7%	4.392	1,9%
11 Aysén	3.941	2,3%	2.857	1,8%
12 Magallanes	2.315	1,8%	1.588	1,6%
13 Metropolitana	11.461	1,0%	7.790	1,5%
14 Los Rios	6.576	4,0%	4.364	2,9%
15 Arica y Parinacota	3.495	2,0%	2.259	2,8%
Nacional	90.122	0,7%	59.084	0,8%

Fuente: Elaborado por el INE y el Ministerio de Desarrollo Social.

Las bases de datos de la Encuesta Casen tienen disponibles, a partir de 1996, las variables que permiten identificar tres importantes aspectos del diseño muestral: estratos (estrato), conglomerados (segmento) y el factor de expansión (expr). Para efectos de comparación,

Las tablas Tabla VII.9 y VII.10 presentan los resultados de la estimación de la tasa de pobreza regional y sus márgenes de error (al 95% de confianza) para las muestras Casen 1996 a 2011-S2 (submuestra 2). El margen de error de 2011-S2 fue estimado en Stata 11 utilizando los parámetros del método 4 (varstrat, varunit, expr_r2, fmr2), considerando la utilización de pseudo-estratos y pseudo-conglomerados. Los márgenes de error de los años 1996 a 2009 fueron estimados en Stata 11 utilizando los parámetros del método 3 (estrato, segmento, expr_r2), ya que no se dispone para la serie histórica de pseudo-estratos (varstrat) y pseudo-conglomerados (varunit). Los tamaños muestrales asociados a las entrevistas efectivas de cada año se presentan en la última fila de la Tabla VII.8.

Entre 1996 y 2009, los márgenes de error de la tasa de pobreza a nivel regional promediaban un 3,2% con un mínimo 1% en 2006 en la Región de Metropolitana y un máximo de 6,7% en 2009 en la Región de los Ríos. Estos resultados eran el reflejo de los tamaños muestrales en cada región y del diseño muestral empleado cada año. En 2011, los errores muestrales de la tasa de pobreza a nivel regional promedian un 2,3%. Como se puede apreciar en la Tabla VII.9, para 10 de las 15 regiones del país, los márgenes de error de la tasa de pobreza regional corresponden a los mínimos históricos, por ejemplo la región de Tarapacá con un 2,3% y Magallanes y Aysén con un 1,6% y 1,8% respectivamente. Este hecho, tan sólo a partir de la submuestra 2, implica que se logró el objetivo planteado en la etapa de diseño de la muestra 2011 – mejorar las precisión de las estimaciones de las tasas de pobreza regional.

Tabla VII.9: Tasa de Pobreza (p) en la muestra Casen por Año, según Región

Región	1996	1998	2000	2003	2006	2009	2011
1 Tarapacá *	21,4	15,6	20,8	18,5	11,8	15,8	13,1
2 Antofagasta	16,4	12,7	13,4	11,2	7,3	8,0	7,5
3 Atacama	26,4	28,7	23,9	24,9	10,5	17,5	13,3
4 Coquimbo	30,1	24,8	24,8	21,5	15,9	16,6	15,3
5 Valparaíso	22,3	18,8	18,8	19,4	15,3	15,1	16,9
6 O´Higgins	26,6	22,9	20,9	19,2	11,4	12,8	10,1
7 Maule	32,6	29,3	25,3	23,1	17,7	20,7	16,2
8 Bío Bío	34,1	32,2	27,1	28,0	20,7	21,0	21,5
9 La Araucanía	35,7	34,8	32,7	29,7	20,1	27,1	22,9
10 Los Lagos **	32,3	29,2	25,6	21,8	11,8	14,2	15,0
11 Aysén	22,4	15,1	14,3	14,2	9,2	15,1	9,8
12 Magallanes	13,7	12,0	11,1	12,3	6,3	9,1	5,8
13 Metropolitana	14,7	15,4	15,1	13,1	10,6	11,5	11,5
14 Los Rios					18,8	20,4	17,5
15 Arica y Parinacota					18,6	12,8	15,7
Nacional	23,2	21,7	20,2	18,7	13,7	15,1	14,4
n (hogares)	33.636	48.107	65.036	71.321	73.720	71.460	59.084

Tabla VII.10: Margen de Error (MOE(p)) al 95% de Confianza de la Tasa de Pobreza (p) en la muestra Casen por Año, según Región

Región	1996	1998	2000	2003	2006	2009	2011
1 Tarapacá *	5,14	4,69	4,96	4,53	3,28	5,61	2,27
2 Antofagasta	3,96	3,56	4,16	3,01	2,67	3,03	2,00
3 Atacama	5,38	7,02	3,94	4,87	3,45	4,74	2,41
4 Coquimbo	3,16	3,10	3,23	2,93	2,92	2,75	2,61
5 Valparaíso	2,65	2,39	2,12	1,88	1,84	1,94	2,62
6 O´Higgins	2,77	2,53	2,32	2,33	1,81	1,77	2,13
7 Maule	2,72	2,86	2,35	2,72	2,30	3,48	1,64
8 Bío Bío	3,41	2,24	2,27	1,92	1,66	1,81	3,63
9 La Araucanía	3,08	2,77	2,50	2,37	1,95	2,71	2,31
10 Los Lagos **	3,03	3,94	3,05	2,17	1,92	2,97	1,85
11 Aysén	4,77	3,96	4,54	3,70	5,55	4,74	1,68
12 Magallanes	5,37	4,89	4,76	4,81	2,97	3,22	1,52
13 Metropolitana	1,48	1,10	1,24	0,96	1,01	0,93	1,45
14 Los Rios					3,55	6,73	2,88
15 Arica y Parinacota					5,64	4,61	2,72
Nacional	0,94	0,75	0,76	0,62	0,58	0,64	0,81
n (hogares)	33.636	48.107	65.036	71.321	73.720	71.460	59.084

Figura VII.11: Sintaxis de estimación en Stata 11.

En primer lugar, es necesario crear variable "cortep" en base de datos principal de Encuesta CASEN 2011:

```
recode corte (1/2=1 "Total Pobres")(3=2 "No Pobres"),gen(cortep)
```

```
* -----
* metodo 1: estrato, segmento
* -----
svyset segmento [pw=expr_r2], strata(estrato) singleunit(certainty)
svy: prop cortep, level(95)
svy: prop cortep, level(95) over(region)

* -----
* método 2: estrato, segmento, fmr2
* -----
svyset segmento [pw=expr_r2], strata(estrato) singleunit(certainty) fpc(fmr2)
svy: prop cortep, level(95)
svy: prop cortep, level(95) over(region)

* -----
* método 3: varstrat, varunit
* -----
svyset varunit [pw=expr_r2], strata(varstrat)
svy: prop cortep, level(95)
svy: prop cortep, level(95) over(region)

* -----
* método 4: varstrat, varunit, fmr2
* -----
svyset varunit [pw=expr_r2], strata(varstrat) fpc(fmr2)
svy: prop cortep, level(95)
svy: prop cortep, level(95) over(region)
```

Figura VII.12: Sintaxis de estimación en SPSS 19.

En primer lugar, es necesario crear variable "cortep" en base de datos principal de Encuesta CASEN 2011:

```
RECODE corte (1 thru 2=1) (3 thru Highest=2) INTO cortep.
VARIABLE LABELS cortep 'Situación de pobreza'.
EXECUTE.
VALUE labels cortep
  1 'Total pobre'
  2 'No pobre'.
```

```
* -----
* método 5: estrato, conglomerado
* -----
CSPLAN ANALYSIS
/PLAN FILE='C:\Directorio\PLAN_1_Congl_Estrat_Personas.csaplan'
/PLANVARS ANALYSISWEIGHT=expr_r2
/SRSESTIMATOR TYPE=WR
/PRINT PLAN
/DESIGN STRATA=area comuna CLUSTER=segmento
```

```

/ESTIMATOR TYPE=WR.
CSTABULATE
/PLAN FILE='C:\Directorio\PLAN_1_Congl_Estrat_Personas.csaplan'
/TABLES VARIABLES=region BY cortep
/CELLS ROWPCT
/STATISTICS SE

* -----
* método 6: varstrat, variunit
* -----

CSPLAN ANALYSIS
/PLAN FILE='C:\Directorio\PLAN_2_VarUnit_VarStrat_Personas.csaplan'
/PLANVARS ANALYSISWEIGHT=expr_r2
/SRSESTIMATOR TYPE=WR
/PRINT PLAN
/DESIGN STRATA=varstrat CLUSTER=variunit
/ESTIMATOR TYPE=WR.
CSTABULATE
/PLAN FILE='C:\Directorio\PLAN_2_VarUnit_VarStrat_Personas.csaplan'
/TABLES VARIABLES=region BY cortep
/CELLS ROWPCT
/STATISTICS SE
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.

```

VIII. Anexos

1. Anexo A: Referencias

- Czajka, J.L., Hirabayashi, S.M., Little, R.J.A. and Rubin, D.B. (1992). "Projecting from Advance Data Using Propensity Modeling: An Application to Income and Tax Statistics." *Journal of Business and Economic Statistics*, 10(2), 117-132.
- Ehling, M., and T. Körner (2007), *Handbook on Data Quality Assessment Methods and Tools*, Wiesbaden: European Commission, Eurostat.
- Sitter RR (1992) Comparing Three Bootstrap Methods for Survey Data. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 20:135-154
- Valliant R, Brick JM, Dever JA (2008) Weight Adjustments for the Grouped Jackknife Variance Estimator. *Journal of Social Statistics* 24(3):469-488.
- Valliant, R. Dever, J.A., Kreuter, F. (2013, chapter 14).
- Valliant, R. Dever, J.A., Kreuter, F. (2013, section 6.5.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- Valliant R, Rust KF (2010) Degrees of Freedom Approximations and Rules-of-Thumb. *Journal of Social Statistics* 26:585-602.
- Wolter KM (2007) *Introduction to Variance Estimation*, 2nd edn. Springer, New York
- Valliant, R. Dever, J.A., Kreuter, F. (2013, chapter 15) *Westat (2007) WesVar 4.3 Users Guide*. Westat, Rockville MD, URL www.westat.com

2. Anexo B: Tablas

Tabla VIII.1 : Estimaciones de la muestra Casen 2009

Regiones	Tasa de Pobreza P	Error Estandar Tasa Pobreza SE_p^2	Efecto Diseño Tasa pobreza $Deff_p$	Tamaño Muestral (viviendas) $n \cdot \bar{m}$	Cuasi-Varianza S_p^2
I Región	,158	,029	4,764	1.252	,215
II Región	,080	,015	4,678	1.793	,092
III Región	,175	,024	3,061	1.518	,290
IV Región	,166	,014	3,168	3.174	,198
V Región	,151	,010	4,358	7.835	,177
VI Región	,128	,009	2,002	6.473	,263
VII Región	,207	,018	6,354	6.468	,321
VIII Región	,210	,009	3,381	11.782	,298
IX Región	,271	,014	3,028	6.286	,398
X Región	,142	,015	4,846	5.427	,257
XI Región	,151	,024	1,404	1.080	,449
XII Región	,091	,016	1,647	761	,125
RM Región Metropolitana	,115	,005	4,519	13.202	,066
XIV Región	,204	,034	8,764	2.476	,333
XV Región	,128	,024	2,633	714	,150
País	,151	,003	4,289	70.241	,173

Tabla VIII.2: Unidades Primarias de Muestreo, por región y área

REGIÓN	AREA	Manzanas	Secciones	Total PSUs	Estratos Urbanos	Estratos Rurales	Total Strata
1	Urbana	737	6	743	3	1	4
	Rural	0	25	25	0	5	5
	Total 1	737	31	768	3	6	9
2	Urbana	855	4	859	6	1	7
	Rural	0	14	14	0	5	5
	Total 2	855	18	873	6	6	12
3	Urbana	916	0	916	8	0	8
	Rural	0	31	31	0	7	7
	Total 3	916	31	947	8	7	15
4	Urbana	406	9	415	10	3	13
	Rural	0	53	53	0	15	15
	Total 4	406	62	468	10	18	28
5	Urbana	774	15	789	31	5	36
	Rural	0	71	71	0	31	31
	Total 5	774	86	860	31	36	67

... continuación

REGIÓN	AREA	Manzanas	Secciones	Total PSUs	Estratos Urbanos	Estratos Rurales	Total Strata
6	Urbana	398	29	427	22	10	32
	Rural	0	95	95	0	32	32
	Total 6	398	124	522	22	42	64
7	Urbana	462	51	513	14	16	30
	Rural	0	123	123	0	30	30
	Total 7	462	174	636	14	46	60
8	Urbana	730	58	788	34	20	54
	Rural	0	125	125	0	50	50
	Total 8	730	183	913	34	70	104
9	Urbana	450	41	491	19	13	32
	Rural	0	107	107	0	32	32
	Total 9	450	148	598	19	45	64
10	Urbana	477	36	513	13	10	23
	Rural	0	106	106	0	25	25
	Total 10	477	142	619	13	35	48
11	Urbana	375	18	393	2	3	5
	Rural	0	40	40	0	6	6
	Total 11	375	58	433	2	9	11
12	Urbana	494	8	502	2	1	3
	Rural	0	9	9	0	2	2
	Total 12	494	17	511	2	3	5
13	Urbana	1.096	4	1.100	49	2	51
	Rural	0	44	44	0	19	19
	Total 13	1.096	48	1.144	49	21	70
14	Urbana	575	15	590	9	3	12
	Rural	0	92	92	0	12	12
	Total 14	575	107	682	9	15	24
15	Urbana	491	3	494	1	1	2
	Rural	0	11	11	0	1	1
	Total 15	491	14	505	1	2	3
Total General	Urbana	9.236	297	9.533	223	89	312
	Rural	0	946	946	0	272	272
	Total	9.236	1.243	10.479	223	361	584

Tabla VIII.3: Grupos de Tamaño de marco muestral INE en Áreas Urbanas

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

Tabla VIII.4: Total de UPM y USM seleccionadas, por región

Nombre Región	Unidades Primarias	Total	Total Viviendas
TOTAL PAÍS	Manzanas	9.236	84.410
	Secciones	1.243	28.461
	Total	10.479	112.871
Región de Tarapacá	Manzanas	737	7.003
	Secciones	31	760
	Total	768	7.763
Región de Antofagasta	Manzanas	855	7.118
	Secciones	18	397
	Total	873	7.515
Región de Atacama	Manzanas	916	5.907
	Secciones	31	677
	Total	947	6.584
Región Coquimbo	Manzanas	406	3.142
	Secciones	62	1.376
	Total	468	4.518
Región de Valparaíso	Manzanas	774	7.867
	Secciones	86	1.828
	Total	860	9.695
Región de O'Higgins	Manzanas	398	3.830
	Secciones	124	2.780
	Total	522	6.610
Región del Maule	Manzanas	462	4.365
	Secciones	174	3.863
	Total	636	8.228
Región del BíoBío	Manzanas	730	6.588
	Secciones	183	3.926
	Total	913	10.514
Región de la Araucanía	Manzanas	450	3.630
	Secciones	148	3.312
	Total	598	6.942
Región de los Lagos	Manzanas	477	4.465
	Secciones	142	3.382
	Total	619	7.847
Región de Aisén	Manzanas	375	4.017
	Secciones	58	1.566
	Total	433	5.583
Región de Magallanes	Manzanas	494	3.351
	Secciones	17	422
	Total	511	3.773
Región Metropolitana	Manzanas	1.096	14.496
	Secciones	48	963
	Total	1.144	15.459
Región de los Ríos	Manzanas	575	5.081
	Secciones	107	2.817
	Total	682	7.898
Región de Arica	Manzanas	491	3.550
	Secciones	14	392
	Total	505	3.942

Tabla VIII.5: Estratos donde fueron seleccionados 1

Región	Estrato		Viviendas			Unidades Muestrales		
	Comuna	Área	Replica 1	Replica 2	Total	Replica 1	Replica 2	Total
		Total País	1.665	2.432	4.097	95	94	189
Arica y Parinacota	Camarones	RURAL	20	30	50	1	1	2
Taracá	Pica	RURAL	21	62	83	1	1	2
Antofagasta	Taltal	RURAL	18	27	45	1	1	2
Atacama	Copiapó	RURAL	36	52	88	2	1	3
	Diego de Almagro	RURAL	16	22	38	1	1	2
Coquimbo	Andacollo	RURAL	14	20	34	1	1	2
	La Higuera	RURAL	21	30	51	1	1	2
	Los Vilos	RURAL	20	29	49	1	1	2
Valparaíso	Algarrobo	RURAL	14	21	35	1	1	2
	Cabildo	RURAL	21	31	52	1	1	2
	Calera	RURAL	14	20	34	1	1	2
	Calle Larga	RURAL	19	27	46	1	1	2
	Cartagena	RURAL	14	21	35	1	1	2
	Casablanca	RURAL	18	26	44	1	1	2
	Hijuelas	RURAL	21	30	51	1	1	2
	Los Andes	RURAL	18	26	44	1	1	2
	Nogales	RURAL	15	22	37	1	1	2
	Olmué	RURAL	18	26	44	1	1	2
	Panquehue	RURAL	16	24	40	1	1	2
	Puchuncaví	RURAL	14	20	34	1	1	2
	Putendo	RURAL	20	30	50	1	1	2
	Quilpué	RURAL	14	20	34	1	1	2
	Quintero	RURAL	15	22	37	1	1	2
	Rinconada	RURAL	14	21	35	1	1	2
	San Esteban	RURAL	17	24	41	1	1	2
	San Felipe	RURAL	20	29	49	1	1	2
	Santa María	RURAL	19	27	46	1	1	2
	Santo Domingo	RURAL	14	21	35	1	1	2
Valparaíso	RURAL	14	19	33	1	1	2	
Villa Alemana	RURAL	14	19	33	1	1	2	
Zapallar	RURAL	14	20	34	1	1	2	
O'Higgins	Doñihue	RURAL	16	24	40	1	1	2
	Litueche	RURAL	20	29	49	1	1	2
	Machalí	RURAL	18	26	44	1	1	2
	Mostazal	RURAL	19	28	47	1	1	2
	Nancagua	RURAL	21	31	52	1	1	2
	Navidad	URBANA	22	32	54	1	1	2
	Olivar	RURAL	22	32	54	1	1	2
	Paredones	RURAL	21	30	51	1	1	2
	Peralillo	RURAL	20	28	48	1	1	2
	Peumo	RURAL	22	32	54	1	1	2
Pichilemu	RURAL	18	27	45	1	1	2	
Pumanque	RURAL	19	27	46	1	1	2	
Maule	Empedrado	RURAL	15	22	37	1	1	2
	Licantén	RURAL	20	29	49	1	1	2
	Pelluhue	RURAL	20	29	49	1	1	2
	Vichuquén	RURAL	22	31	53	1	1	2

Continúa

Tabla VIII.5: Estratos donde fueron seleccionados 1

Región	Estrato		Viviendas			Unidades Muestrales		
	Comuna	Área	Replica 1	Replica 2	Total	Replica 1	Replica 2	Total
Biobío	Alto Biobío	RURAL	18	26	44	1	1	2
	Alto Biobío	URBANA	22	31	53	1	1	2
	Antuco	RURAL	16	22	38	1	1	2
	Bulnes	RURAL	20	28	48	1	1	2
	Cabrero	RURAL	18	27	45	1	1	2
	Coelemu	RURAL	22	32	54	1	1	2
	Contulmo	RURAL	14	21	35	1	1	2
	Curanilahue	RURAL	18	26	44	1	1	2
	Hualqui	RURAL	16	23	39	1	1	2
	Los Alamos	RURAL	18	27	45	1	1	2
	Mulchén	RURAL	17	24	41	1	1	2
	Negrete	RURAL	16	22	38	1	1	2
	Ninhue	RURAL	17	24	41	1	1	2
	Ñiquén	URBANA	22	32	54	1	1	2
	Penco	RURAL	13	19	32	1	1	2
	Pinto	RURAL	17	25	42	1	1	2
	Portezuelo	RURAL	15	21	36	1	1	2
	Quilaco	RURAL	16	22	38	1	1	2
	Quilleco	RURAL	16	24	40	1	1	2
	Quirihue	RURAL	16	22	38	1	1	2
	San Fabián	RURAL	16	23	39	1	1	2
	San Nicolás	RURAL	20	28	48	1	1	2
	San Rosendo	RURAL	12	18	30	1	1	2
	Santa Bárbara	RURAL	16	24	40	1	1	2
Santa Juana	RURAL	20	28	48	1	1	2	
Tirúa	RURAL	22	32	54	1	1	2	
Tomé	RURAL	18	25	43	1	1	2	
Tucapel	RURAL	15	22	37	1	1	2	
Yungay	RURAL	18	25	43	1	1	2	
Araucanía	Curacautín	RURAL	20	28	48	1	1	2
	Los Sauces	RURAL	17	24	41	1	1	2
	Melipeuco	RURAL	20	28	48	1	1	2
	Renaico	RURAL	16	24	40	1	1	2
	Traiguén	RURAL	21	31	52	1	1	2
Los Lagos	Puqueldón	RURAL	22	32	54	1	1	2
Los Ríos	Corral	RURAL	18	26	44	1	1	2
Metropolitana de Santiago	Alhué	RURAL	14	19	33	1	1	2
	Buín	RURAL	18	25	43	1	1	2
	El Monte	RURAL	14	21	35	1	1	2
	Isla de Maipo	RURAL	18	27	45	1	1	2
	María Pinto	RURAL	17	25	42	1	1	2
	Padre Hurtado	RURAL	15	22	37	1	1	2
	Peñaflor	RURAL	14	19	33	1	1	2
	San Bernardo	RURAL	15	21	36	1	1	2
	San José de Maipo	RURAL	14	20	34	1	1	2
	San Pedro	RURAL	16	24	40	1	1	2
	Talagante	RURAL	18	25	43	1	1	2
Tiltil	RURAL	16	23	39	1	1	2	

Tabla VIII.6: Total de conglomerados con 30 o más viviendas que responde según submuestra

Región	Replica 1			Replica 2			Muestra completa		
	>=30	<30	Total	>=30	<30	Total	>=30	<30	Total
Arica y Parinacota	3	175	178	0	342	342	3	517	520
Tarapacá	0	243	243	0	517	517	0	760	760
Antofagasta	0	251	251	0	606	606	0	857	857
Atacama	0	331	331	1	638	639	1	969	970
Coquimbo	1	165	166	1	304	305	2	469	471
Valparaíso	0	313	313	0	577	577	0	890	890
Libertador General Bernardo OHiggins	2	192	194	5	355	360	7	547	554
Maule	0	214	214	2	420	422	2	634	636
BioBio	0	314	314	3	594	597	3	908	911
Araucanía	6	192	198	3	395	398	9	587	596
Los Lagos	4	220	224	13	404	417	17	624	641
Los Ríos	0	235	235	2	507	509	2	742	744
Aysén del General Carlos Ibañez del Campo	0	134	134	5	309	314	5	443	448
Magallanes y La Antártica Chilena	0	185	185	0	328	328	0	513	513
Metropolitana de Santiago	0	388	388	6	744	750	6	1.132	1.138
Total	16	3.552	3.568	41	7.040	7.081	57	10.592	10.649

Tabla VIII.7: Total de USM seleccionadas, por comuna.

REGION	COMUNA	Población Viviendas	Porcentaje Comunal	Total de viviendas muestral	Porcentaje Comunal
PAIS	TOTAL	4.594.668	-	112.938	-
1 Tarapacá	Total	71.684	100,0	7.734	100,0
	1101 Iquique	45.532	63,5	4.996	64,6
	1107 Alto Hospicio	20.214	28,2	1.978	25,6
	1401 Pozo Almonte	2.378	3,3	282	3,6
	1402 Camiña	421	0,6	67	0,9
	1404 Huara	1.827	2,5	207	2,7
	1405 Pica	1.312	1,8	204	2,6
2 Antofagasta	Total	123.340	100,0	7.486	100,0
	2101 Antofagasta	80.105	64,9	4.610	61,6
	2102 Mejillones	1.995	1,6	161	2,2
	2103 Sierra Gorda	271	0,2	46	0,6
	2104 Taltal	2.895	2,3	246	3,3
	2201 Calama	29.206	23,7	1.664	22,2
	2203 San Pedro de Atacama	1.133	0,9	183	2,4
	2301 Tocopilla	6.090	4,9	421	5,6
2302 María Elena	1.645	1,3	155	2,1	
3 Atacama	Total	68.332	100,0	6.610	100,0
	3101 Copiapó	31.984	46,8	2.918	44,1
	3102 Caldera	3.714	5,4	431	6,5
	3103 Tierra Amarilla	3.425	5,0	341	5,2
	3201 Chañaral	3.281	4,8	324	4,9
	3202 Diego de Almagro	5.997	8,8	598	9,0
	3301 Vallenar	13.469	19,7	1.225	18,5
	3302 Alto del Carmen	1.608	2,4	166	2,5
	3303 Freirina	1.848	2,7	243	3,7
3304 Huasco	3.006	4,4	364	5,5	
4 Coquimbo	Total	198.469	100,0	4.511	100,0
	4101 La Serena	60.179	30,3	1.048	23,2
	4102 Coquimbo	47.928	24,1	890	19,7
	4103 Andacollo	3.411	1,7	151	3,3
	4104 La Higuera	1.426	0,7	108	2,4
	4105 Paiguano	1.509	0,8	54	1,2
	4106 Vicuña	7.608	3,8	208	4,6
	4201 Illapel	10.533	5,3	245	5,4
	4202 Canela	2.947	1,5	147	3,3
	4203 Los Vilos	10.903	5,5	257	5,7
	4204 Salamanca	6.899	3,5	190	4,2
	4301 Ovalle	26.854	13,5	578	12,8
	4302 Combarbalá	4.670	2,4	169	3,7
	4303 Monte Patria	9.633	4,9	254	5,6
	4304 Punitaqui	2.697	1,4	157	3,5
4305 Río Hurtado	1.272	0,6	55	1,2	

... Continuación Tabla VIII.7

REGION	COMUNA	Población Viviendas	Porcentaje Comunal	Total de viviendas muestral	Porcentaje Comunal
5 Valparaíso	Total	491.287	100,0	9.692	100,0
	5101 Valparaíso	83.131	16,9	1.208	12,5
	5102 Casablanca	7.549	1,5	166	1,7
	5103 Concón	8.950	1,8	161	1,7
	5105 Puchuncaví	4.621	0,9	134	1,4
	5106 Quilpué	46.851	9,5	756	7,8
	5107 Quintero	5.714	1,2	148	1,5
	5108 Villa Alemana	35.111	7,1	578	6,0
	5109 Viña del Mar	95.726	19,5	1.288	13,3
	5301 Los Andes	24.190	4,9	412	4,3
	5302 Calle Larga	2.593	0,5	116	1,2
	5303 Rinconada	2.079	0,4	113	1,2
	5304 San Esteban	2.025	0,4	112	1,2
	5401 La Ligua	9.084	1,8	221	2,3
	5402 Cabildo	5.367	1,1	162	1,7
	5403 Papudo	1.278	0,3	70	0,7
	5404 Petorca	2.678	0,5	123	1,3
	5405 Zapallar	1.833	0,4	109	1,1
	5501 Quillota	23.195	4,7	436	4,5
	5502 Calera	12.714	2,6	279	2,9
	5503 Hijuelas	2.950	0,6	137	1,4
	5504 La Cruz	3.395	0,7	150	1,5
	5505 Limache	24.770	5,0	467	4,8
	5506 Nogales	6.166	1,3	164	1,7
	5507 Olmué	4.612	0,9	164	1,7
	5601 San Antonio	24.282	4,9	434	4,5
	5602 Algarrobo	2.858	0,6	132	1,4
	5603 Cartagena	4.634	0,9	160	1,7
	5604 El Quisco	3.258	0,7	92	0,9
	5605 El Tabo	2.556	0,5	86	0,9
	5606 Santo Domingo	1.974	0,4	110	1,1
	5701 San Felipe	17.879	3,6	336	3,5
5702 Catemu	2.798	0,6	125	1,3	
5703 Llay Llay	5.440	1,1	162	1,7	
5704 Panquehue	1.598	0,3	104	1,1	
5705 Putaendo	3.743	0,8	135	1,4	
5706 Santa María	3.685	0,8	142	1,5	
6 Libertador Bernardo O'Higgins	Total	244.622	100,0	6.630	100,0
	6101 Rancagua	61.299	25,1	1.180	17,8
	6102 Codegua	3.386	1,4	160	2,4
	6103 Coinco	1.872	0,8	120	1,8
	6104 Coltauco	4.741	1,9	157	2,4
	6105 Doñihue	5.551	2,3	189	2,9

... continuación Tabla VIII.7

REGION	COMUNA	Población Viviendas	Porcentaje Comunal	Total de viviendas muestral	Porcentaje Comunal	
6 Libertador Bernardo O´Higgins	6106 Graneros	14.640	6,0	311	4,7	
	6107 Las Cabras	6.403	2,6	164	2,5	
	6108 Machalí	9.043	3,7	249	3,8	
	6109 Malloa	3.671	1,5	135	2,0	
	6110 Mostazal	6.433	2,6	208	3,1	
	6111 Olivar	3.466	1,4	147	2,2	
	6112 Peumo	4.625	1,9	173	2,6	
	6113 Pichidegua	4.071	1,7	140	2,1	
	6114 Quinta de Tilcoco	3.440	1,4	138	2,1	
	6115 Rengo	17.119	7,0	324	4,9	
	6116 Requínoa	6.319	2,6	180	2,7	
	6117 San Vicente	15.231	6,2	285	4,3	
	6201 Pichilemu	3.933	1,6	137	2,1	
	6202 La Estrella	999	0,4	95	1,4	
	6203 Litueche	2.045	0,8	111	1,7	
	6204 Marchihue	2.638	1,1	118	1,8	
	6205 Navidad	1.994	0,8	117	1,8	
	6206 Paredones	1.873	0,8	115	1,7	
	6301 San Fernando	20.104	8,2	379	5,7	
	6302 Chépica	3.910	1,6	150	2,3	
	6303 Chimbarongo	9.045	3,7	200	3,0	
	6304 Lolol	1.817	0,7	108	1,6	
	6305 Nancagua	4.869	2,0	150	2,3	
	6306 Palmilla	3.467	1,4	134	2,0	
	6307 Peralillo	2.735	1,1	127	1,9	
	6308 Placilla	2.254	0,9	123	1,9	
	6309 Pumanque	1.000	0,4	45	0,7	
	6310 Santa Cruz	10.629	4,3	261	3,9	
		Total	288.879	100,0	8.211	100,0
	7 Maule	7101 Talca	52.845	18,3	1.126	13,7
7102 Constitución		13.353	4,6	328	4,0	
7103 Curepto		2.616	0,9	134	1,6	
7104 Empedrado		1.165	0,4	103	1,3	
7105 Maule		5.741	2,0	202	2,5	
7106 Pelarco		2.131	0,7	124	1,5	
7107 Pencahue		2.559	0,9	138	1,7	
7108 Río Claro		4.214	1,5	165	2,0	
7109 San Clemente		10.213	3,5	267	3,3	
7110 San Rafael		2.481	0,9	126	1,5	
7201 Cauquenes		10.571	3,7	299	3,6	
7202 Chanco		2.170	0,8	136	1,7	
7203 Pelluhue		1.836	0,6	119	1,4	
7301 Curicó		68.213	23,6	1.534	18,7	

... continuación Tabla VIII.7

REGION	COMUNA	Población Viviendas	Porcentaje Comunal	Total de viviendas muestral	Porcentaje Comunal
7 Maule	7302 Hualañé	3.672	1,3	157	1,9
	7303 Licantén	1.848	0,6	128	1,6
	7304 Molina	9.467	3,3	273	3,3
	7305 Rauco	2.198	0,8	132	1,6
	7306 Romeral	5.134	1,8	176	2,1
	7307 Sagrada Familia	4.844	1,7	182	2,2
	7308 Teno	8.226	2,8	251	3,1
	7309 Vichuquén	1.216	0,4	118	1,4
	7401 Linares	22.837	7,9	507	6,2
	7402 Colbún	4.908	1,7	175	2,1
	7403 Longaví	8.301	2,9	232	2,8
	7404 Parral	10.095	3,5	278	3,4
	7405 Retiro	6.237	2,2	194	2,4
	7406 San Javier	10.594	3,7	273	3,3
	7407 Villa Alegre	3.873	1,3	154	1,9
	7408 Yerbas Buenas	5.321	1,8	180	2,2
	Total	553.948	100,0	10.487	100,0
8 Bío Bío	8101 Concepción	64.219	11,6	820	7,8
	8102 Coronel	28.325	5,1	374	3,6
	8103 Chiguayante	29.846	5,4	364	3,5
	8104 Florida	2.967	0,5	128	1,2
	8105 Hualqui	4.931	0,9	134	1,3
	8106 Lota	13.972	2,5	193	1,8
	8107 Penco	12.297	2,2	225	2,1
	8108 San Pedro de la Paz	25.172	4,5	322	3,1
	8109 Santa Juana	3.114	0,6	127	1,2
	8110 Talcahuano	41.520	7,5	516	4,9
	8111 Tomé	16.273	2,9	240	2,3
	8112 Hualpén	23.016	4,2	299	2,9
	8201 Lebu	6.949	1,3	162	1,5
	8202 Arauco	12.418	2,2	198	1,9
	8203 Cañete	8.447	1,5	214	2,0
	8204 Contulmo	1.151	0,2	91	0,9
	8205 Curanilahue	12.473	2,3	220	2,1
	8206 Los Alamos	6.926	1,3	161	1,5
	8207 Tirúa	2.369	0,4	118	1,1
	8301 Los Ángeles	54.255	9,8	633	6,0
8302 Antuco	1.197	0,2	105	1,0	
8303 Cabrero	9.901	1,8	182	1,7	
8304 Laja	5.519	1,0	151	1,4	
8305 Mulchén	8.609	1,6	157	1,5	
8306 Nacimiento	8.383	1,5	166	1,6	
8307 Negrete	2.902	0,5	114	1,1	

... continuación Tabla VIII.7

REGION	COMUNA	Población Viviendas	Porcentaje Comunal	Total de viviendas muestral	Porcentaje Comunal
8 Bío Bío	8308 Quilaco	1.169	0,2	95	0,9
	8309 Quilleco	2.856	0,5	106	1,0
	8310 San Rosendo	1.094	0,2	102	1,0
	8311 Santa Bárbara	2.921	0,5	106	1,0
	8312 Tucapel	3.503	0,6	111	1,1
	8313 Yumbel	5.893	1,1	142	1,4
	8314 Alto Biobío	1.388	0,3	100	1,0
	8401 Chillán	45.573	8,2	587	5,6
	8402 Bulnes	6.176	1,1	148	1,4
	8403 Cobquecura	1.328	0,2	111	1,1
	8404 Coelemu	5.671	1,0	153	1,5
	8405 Coihueco	6.580	1,2	172	1,6
	8406 Chillán Viejo	9.942	1,8	206	2,0
	8407 El Carmen	3.176	0,6	131	1,2
	8408 Ninhue	1.320	0,2	104	1,0
	8409 Ñiquén	2.800	0,5	120	1,1
	8410 Pemuco	2.396	0,4	114	1,1
	8411 Pinto	2.569	0,5	113	1,1
	8412 Portezuelo	1.143	0,2	93	0,9
	8413 Quillón	4.838	0,9	155	1,5
	8414 Quirihue	2.797	0,5	130	1,2
	8415 Ránquil	1.902	0,3	116	1,1
8416 San Carlos	13.034	2,4	223	2,1	
8417 San Fabián	1.137	0,2	105	1,0	
8418 San Ignacio	5.812	1,0	161	1,5	
8419 San Nicolás	2.921	0,5	119	1,1	
8420 Treguaco	1.495	0,3	111	1,1	
8421 Yungay	5.363	1,0	139	1,3	
	Total	267.546	100,0	6.924	100,0
9 La Araucanía	9101 Temuco	74.106	27,7	1.231	17,8
	9102 Carahue	6.465	2,4	181	2,6
	9103 Cunco	5.072	1,9	155	2,2
	9104 Curarrehue	2.133	0,8	144	2,1
	9105 Freire	8.264	3,1	231	3,3
	9106 Galvarino	3.104	1,2	128	1,8
	9107 Gorbea	4.561	1,7	145	2,1
	9108 Lautaro	7.997	3,0	206	3,0
	9109 Loncoche	7.609	2,8	226	3,3
	9110 Melipeuco	1.657	0,6	113	1,6
	9111 Nueva Imperial	8.464	3,2	201	2,9
	9112 Padre Las Casas	22.078	8,3	435	6,3
	9113 Perquenco	1.916	0,7	109	1,6
	9114 Pitrufuquén	8.370	3,1	185	2,7
	9115 Pucón	6.340	2,4	186	2,7
	9116 Saavedra	3.815	1,4	137	2,0
	9117 Teodoro Schmidt	5.527	2,1	161	2,3

... continuación Tabla VIII.7

REGION	COMUNA	Población Viviendas	Porcentaje Comunal	Total de viviendas muestral	Porcentaje Comunal
9 La Araucanía	9118 Toltén	3.354	1,3	135	1,9
	9119 Vilcún	10.172	3,8	229	3,3
	9120 Villarrica	15.653	5,9	294	4,2
	9121 Cholchol	2.651	1,0	123	1,8
	9201 Angol	15.708	5,9	359	5,2
	9202 Collipulli	5.810	2,2	181	2,6
	9203 Curacautín	5.438	2,0	158	2,3
	9204 Ercilla	2.351	0,9	143	2,1
	9205 Lonquimay	2.960	1,1	134	1,9
	9206 Los Sauces	1.499	0,6	108	1,6
	9207 Lumaco	2.292	0,9	143	2,1
	9208 Purén	3.165	1,2	158	2,3
	9209 Renaico	2.392	0,9	127	1,8
9210 Traiguén	6.716	2,5	189	2,7	
9211 Victoria	9.907	3,7	269	3,9	
	Total	228.440	100,0	7.832	100,0
10 Los Lagos	10101 Puerto Montt	52.450	23,0	1.389	17,7
	10102 Calbuco	11.122	4,9	365	4,7
	10104 Fresia	3.584	1,6	213	2,7
	10105 Frutillar	4.632	2,0	236	3,0
	10106 Los Muermos	5.760	2,5	232	3,0
	10107 Llanquihue	4.630	2,0	230	2,9
	10108 Maullín	5.695	2,5	288	3,7
	10109 Puerto Varas	8.183	3,6	518	6,6
	10201 Castro	13.371	5,9	425	5,4
	10202 Ancud	15.193	6,7	432	5,5
	10203 Chonchi	4.208	1,8	195	2,5
	10204 Curaco de Vélez	1.118	0,5	53	0,7
	10205 Dalcahue	3.913	1,7	167	2,1
	10206 Puqueldón	1.103	0,5	55	0,7
	10207 Queilén	1.648	0,7	125	1,6
	10208 Quellón	9.331	4,1	307	3,9
	10209 Quemchi	2.692	1,2	136	1,7
	10210 Quinchao	3.011	1,3	143	1,8
10301 Osorno	50.642	22,2	1.299	16,6	
10302 Puerto Octay	2.964	1,3	141	1,8	
10303 Purranque	8.734	3,8	258	3,3	
10304 Puyehue	2.921	1,3	141	1,8	
10305 Río Negro	5.793	2,5	199	2,5	
10306 San Juan de la Costa	2.286	1,0	128	1,6	
10307 San Pablo	3.456	1,5	157	2,0	

... continuación Tabla VIII.7

REGION	COMUNA	Población Viviendas	Porcentaje Comunal	Total de viviendas muestral	Porcentaje Comunal
11 Aysén	Total	28.211	100,0	5.566	100,0
	11101 Coyhaique	17.546	62,2	3.421	61,5
	11201 Aysén	7.098	25,2	1.248	22,4
	11202 Cisnes	950	3,4	225	4,0
	11301 Cochrane	690	2,4	240	4,3
	11401 Chile Chico	1.244	4,4	307	5,5
	11402 Río Ibáñez	683	2,4	125	2,2
12 Magallanes Y La Antártica Chilena	Total	45.118	100,0	3.826	100,0
	12101 Punta Arenas	37.836	83,9	3.048	79,7
	12301 Porvenir	1.589	3,5	277	7,2
	12401 Natales	5.693	12,6	501	13,1
13 Región Metropolitana	Total	1.833.728	100,0	15.488	100,0
	13101 Santiago	90.805	5,0	796	5,1
	13102 Cerrillos	22.147	1,2	168	1,1
	13103 Cerro Navia	37.258	2,0	243	1,6
	13104 Conchalí	35.915	2,0	218	1,4
	13105 El Bosque	46.840	2,6	301	1,9
	13106 Estación Central	36.930	2,0	248	1,6
	13107 Huechuraba	20.413	1,1	163	1,1
	13108 Independencia	21.285	1,2	166	1,1
	13109 La Cisterna	24.840	1,4	225	1,5
	13110 La Florida	101.439	5,5	755	4,9
	13111 La Granja	33.195	1,8	258	1,7
	13112 La Pintana	47.667	2,6	280	1,8
	13113 La Reina	28.550	1,6	266	1,7
	13114 Las Condes	98.041	5,3	1.336	8,6
	13115 Lo Barnechea	19.980	1,1	339	2,2
	13116 Lo Espejo	24.287	1,3	182	1,2
	13117 Lo Prado	29.826	1,6	191	1,2
	13118 Macul	32.931	1,8	262	1,7
	13119 Maipú	141.904	7,7	777	5,0
	13120 Ñuñoa	62.925	3,4	654	4,2
	13121 Pedro Aguirre Cerda	29.360	1,6	212	1,4
	13122 Peñalolén	60.811	3,3	421	2,7
	13123 Providencia	60.277	3,3	736	4,8
	13124 Pudahuel	55.195	3,0	331	2,1
	13125 Quilicura	47.043	2,6	271	1,7
	13126 Quinta Normal	28.541	1,6	194	1,3
13127 Recoleta	39.928	2,2	261	1,7	

... continuación Tabla VIII.7

REGION	COMUNA	Población Viviendas	Porcentaje Comunal	Total de viviendas muestral	Porcentaje Comunal
13 Región Metropolitana	13128 Renca	37.943	2,1	239	1,5
	13129 San Joaquín	27.099	1,5	232	1,5
	13130 San Miguel	27.337	1,5	219	1,4
	13131 San Ramón	25.012	1,4	177	1,1
	13132 Vitacura	28.148	1,5	508	3,3
	13201 Puente Alto	141.442	7,7	754	4,9
	13202 Pirque	5.012	0,3	116	0,7
	13203 San José de Maipo	3.841	0,2	99	0,6
	13301 Colina	20.293	1,1	267	1,7
	13302 Lampa	24.591	1,3	206	1,3
	13303 Tiltil	4.325	0,2	103	0,7
	13401 San Bernardo	65.821	3,6	419	2,7
	13402 Buin	19.125	1,0	196	1,3
	13403 Calera de Tango	4.851	0,3	119	0,8
	13404 Paine	18.260	1,0	188	1,2
	13501 Melipilla	26.185	1,4	228	1,5
	13502 Alhué	1.604	0,1	89	0,6
	13503 Curacaví	7.581	0,4	123	0,8
	13504 María Pinto	2.764	0,2	98	0,6
	13505 San Pedro	1.913	0,1	42	0,3
13601 Talagante	16.863	0,9	187	1,2	
13602 El Monte	7.226	0,4	121	0,8	
13603 Isla de Maipo	8.899	0,5	143	0,9	
13604 Padre Hurtado	10.309	0,6	150	1,0	
13605 Peñaflores	18.951	1,0	211	1,4	
	Total	107.635	100,0	7.895	100,0
14 Los Ríos	14101 Valdivia	49.666	46,1	3.283	41,6
	14102 Corral	978	0,9	157	2,0
	14103 Lanco	4.201	3,9	329	4,2
	14104 Los Lagos	5.117	4,8	379	4,8
	14105 Máfil	1.665	1,5	199	2,5
	14106 Mariquina	6.119	5,7	514	6,5
	14107 Paillaco	5.810	5,4	457	5,8
	14108 Panguipulli	8.481	7,9	595	7,5
	14201 La Unión	10.191	9,5	770	9,8
	14202 Futrono	3.191	3,0	289	3,7
	14203 Lago Ranco	1.355	1,3	182	2,3
	14204 Río Bueno	10.861	10,1	741	9,4
	Total	43.429	100,0	4.046	100,0
15 Arica y Parinacota	15101 Arica	42.742	98,4	3.875	95,8
	15102 Camarones	255	0,6	50	1,2
	15201 Putre	432	1,0	121	3,0

TABLA VIII.8: Distribución de Viviendas, por Comuna y Zona.

REGIÓN	COMUNA	ZONA	Población Viviendas	Distribución Viviendas (%)	Total de viviendas Muestra	Urbanos Rurales (%)
XIV Los Ríos	14105 Máfil	1 Urbano	928	55,7	104	58,4
		2 Rural	737	44,3	74	41,6
		Total	1.665	100,0	178	100,0
	14106 Mariquina	1 Urbano	3.105	50,7	271	52,5
		2 Rural	3.014	49,3	245	47,5
		Total	6.119	100,0	516	100,0
	14107 Paillaco	1 Urbano	3.204	55,1	231	56,6
		2 Rural	2.606	44,9	177	43,4
		Total	5.810	100,0	408	100,0
	14108 Panguipulli	1 Urbano	3.014	35,5	219	39,3
		2 Rural	5.467	64,5	338	60,7
		Total	8.481	100,0	557	100,0
	14201 La Unión	1 Urbano	6.610	64,9	511	64,6
		2 Rural	3.581	35,1	280	35,4
		Total	10.191	100,0	791	100,0
	14202 Futrono	1 Urbano	1.842	57,7	155	58,9
2 Rural		1.349	42,3	108	41,1	
Total		3.191	100,0	263	100,0	
14203 Lago Ranco	1 Urbano	494	36,5	92	47,2	
	2 Rural	861	63,5	103	52,8	
	Total	1.355	100,0	195	100,0	
14204 Río Bueno	1 Urbano	5.980	55,1	376	55,9	
	2 Rural	4.881	44,9	297	44,1	
	Total	10.861	100,0	673	100,0	
		Urbano	73.389	68,2	5.402	68,4
		Rural	34.246	31,8	2.493	31,6
		TOTAL	107.635	100,0	7.895	100,0
XV Arica y Parinacota	15101 Arica	1 Urbano	39.360	92,1	3.550	91,5
		2 Rural	3.382	7,9	330	8,5
		Total	42.742	100,0	3.880	100,0
	15102 Camarones	2 Rural	255	100,0	50	100,0
	15201 Putre	1 Urbano	160	37,0	63	54,3
		2 Rural	272	63,0	53	45,7
Total		432	100,0	116	100,0	
		Urbano	39.520	91,0	3.613	89,3
		Rural	3.909	9,0	433	10,7
		TOTAL	43.429	100,0	4.046	100,0

3. Anexo C: Plan de Muestreo en SPSS

Pasos para Definir Plan de Muestreo en SPSS

En SPSS, previamente a la realización de cualquier análisis o estimación que considere un diseño complejo, se requiere generar un plan de muestreo.

Generación de un PLAN de muestreo en SPSS

La generación de un plan de muestreo se realiza en 4 pasos:

1. En el módulo de muestras complejas, hay que generar un Plan para el análisis posterior, mediante los comandos:

< Analizar > / < Muestras complejas > / < Preparar para el análisis >

2. En la ventana que se abre se marca la opción:

- Crear un archivo de plan
En Examinar se debe indicar el directorio donde guardará el archivo con el plan de muestreo, al cual deberá asignarle un nombre, por ejemplo: "PLAN_Ensayo". Una vez asignado el nombre, marque <Siguiente> para pasar al siguiente nivel de definición del plan de muestreo.

3. En este nivel deberá introducir tres variables para que quede el plan definido. Estas son:

Estratos → Corresponden a las agrupaciones derivadas de la estratificación definida en el diseño muestral. En el caso específico de la Casen, corresponden originalmente al cruce o interposición de la división política administrativa a nivel comunal (322 comunas del país), con la división geográfica del territorio nacional en las áreas urbanas y rurales (2 áreas)⁴⁸. En Casen, también se creó una Post-Estratificación con el objetivo de corregir estimaciones de la varianza, lo que dio origen a 170 pseudo-estratos denominados VarStrat.

Conglomerados → Son las unidades de selección definidas en el diseño muestral, que generalmente corresponden a unidades muestrales previas a la última etapa. En Casen son denominadas manzanas (5.600) y secciones (736) en las partes urbana y rural de las comunas, respectivamente. En forma simultánea a la creación de pseudo-estratos o VarStrat, en Casen se generaron agrupaciones de manzanas y de secciones, denominadas VarUnit, también con el objetivo de mejorar las estimaciones de la varianza, agrupando los conglomerados de manera de contener una cantidad aproximada mínima de 30 viviendas.

⁴⁸ Cabe hacer notar que no todas las comunas contienen ambas áreas geográfica, por lo que el número de estratos generados de la interposición corresponde a 585 estratos de diseño para la Casen.

Ponderación muestral → Se refiere al factor de expansión que posee cada unidad de selección. En Casen corresponde al factor de expansión regional Expr_r2.

Una vez ingresadas estas tres variables, marque <Siguiete> para pasar al siguiente nivel de definición del plan de muestreo.

4. Llegado a este paso, o incluso en el paso previo, se puede dar por finalizado el proceso de crear un plan de análisis en muestras complejas. Si se escoge la opción de muestreo con reemplazo (CR), no es necesario definir alguna etapa de muestreo posterior y el proceso puede darse por finalizado. La opción de incluir el factor de corrección para poblaciones finitas o factor de corrección por finitud (FPC) aplica sólo si las unidades de conglomeración son seleccionadas bajo el supuesto de muestreo aleatorio simple. Si se escoge cualquiera de las otras dos opciones, que implican un muestreo con reemplazo, se requiere que sea definida una segunda etapa y también que se tengan calculadas las probabilidades de inclusión conjunta de las unidades de conglomeración. En Casen se recomienda la opción CR ya que la estimación de la varianza no difiere prácticamente de las otras opciones simplificando enormemente los cálculos y la expresión de ésta es sencilla.