

ENUT

INE
Instituto Nacional de Estadísticas - Chile

Encuesta Nacional Sobre Uso del Tiempo



Documento Metodológico ENUT 2015

ENUT

**Encuesta Nacional
Sobre Uso del Tiempo**



**Documento Metodológico
ENUT 2015**

Departamento de Estudios Sociales Instituto Nacional de Estadísticas

“Documento Metodológico ENUT 2015”

Subdirectora Técnica: María Mercedes Jeria

Coordinador Estudios Sociales: Matías Pizarro

Jefa de Proyecto: Evelyn Larenas

Analistas investigadores(as): Diego Rodríguez

Rubén Ananías

Mabel Araya

Daniel Espinoza

Rigoberto Figueroa

Andrea Paola Marín

Karlla Muñoz

Verónica Valdivia

Jefa Departamento de

Investigación y Desarrollo: Denisse López

Analistas investigadores(as): Miguel Guerrero

Ximena Espinosa



Paseo Presidente Bulnes 418

Fono: 56-2/8944000 – Fax 6712169

Casilla de correo: 498 – Correo 3

Sitio Web: www.ine.cl

Correo electrónico: ine@ine.cl

Santiago de Chile

Departamento de Estudios Sociales
Departamento de Investigación y Desarrollo
Subdirección Técnica

Noviembre, 2016

ÍNDICE

1. INTRODUCCIÓN	8
2. CONTEXTO	9
2.1 Antecedentes del estudio	9
2.1.1 Tiempo, desigualdades socioeconómicas y de género	9
2.2 Necesidades de información	10
2.3 Objetivos del estudio o proceso estadístico	10
2.3.1 Objetivo general	10
2.3.2 Objetivos específicos	10
2.4 Referencias nacionales	11
2.5 Referencias internacionales	11
2.6 Nomenclaturas utilizadas	12
3. MARCO DE REFERENCIA	12
3.1 Marco conceptual	12
3.1.1 La centralidad del trabajo en la vida moderna	12
3.1.2 El Sistema de Cuentas Nacionales (SCN) como marco de análisis	13
3.1.3 La frontera general de la producción	13
3.1.4 La frontera de la producción del SCN: producción de los hogares	14
3.1.5 Hacia un concepto amplio del trabajo	15
3.2 Dimensiones de análisis para la ENUT 2015	17
3.2.1 Trabajo en la ocupación, producción de bienes para el autoconsumo (actividades dentro de la frontera de la producción del SCN) y actividades conexas	17
3.2.2 Trabajo no remunerado (fuera de la frontera de la producción del SCN)	18
3.2.3 Actividades personales (fuera de la frontera general de la producción del SCN)	19
3.3.4 Satisfacción con el uso del tiempo	19
4. DISEÑO ESTADÍSTICO	20
4.1 Universo	20
4.2 Población objetivo y cobertura	20
4.3 Unidades estadísticas	20
4.3.1 Unidad de observación	20

ÍNDICE

4.3.2.	Unidad de análisis	20
4.3.3.	Unidad de muestreo	20
4.4	Construcción de índices o metodologías compuestas	20
4.5	Definición de variables	21
4.5.1	Variable de estudio	21
4.5.2	Variables de clasificación	21
4.5.3	Variables de análisis	22
4.6	Metodología muestral	22
4.6.1	Marco muestral	22
4.6.2	Estratificación del marco muestral	23
4.6.3	Estratificación geográfica	23
4.6.4	Estratificación del MM 2015 por grupos de tamaño	24
4.6.5	Cobertura del marco muestral	26
4.7	Diseño muestral	29
4.7.1	Tipo de muestreo	29
4.7.2	Tamaño muestral teórico	30
4.7.3	Medidas de precisión	32
4.8	Selección muestral	34
4.8.1	Selección de unidades de primera etapa	34
4.8.2	Selección de unidades de segunda etapa	35
4.8.3	Consideraciones: distribución de la muestra	45
4.9	Tamaño muestral efectivo	45
4.10	Factores de expansión	46
4.10.1	Ponderador de selección de manzanas	47
4.10.2	Ajuste por omisión de manzanas	48
4.10.3	Omisión de comunas	50
4.10.4	Ponderador de selección de viviendas	51
4.10.5	Ajuste por elegibilidad de viviendas	52
4.10.5.1	Ajuste por elegibilidad desconocida	53
4.10.5.2	Ajuste por no elegibilidad	54
4.10.6	Ajuste por no respuesta	54
4.10.7	Suavizamiento del factor de expansión	57
4.10.8	Ponderador de personas ajustado a <i>stock</i> poblacional regional	59
4.10.9	Calibración a <i>stock</i> regional	60
4.10.10	Ponderador de personas calibrado, ajustado por ausencia de respuesta en el módulo Uso del Tiempo	61
4.10.11	Factor de hogar	63
4.10.12	Resumen factores	64
4.11	Estimación de varianzas	65

ÍNDICE

4.11.1	Métodos convencionales para estimación de la varianza	65
4.11.2	Varianza con muestras complejas	67
4.11.3	Algoritmo de cálculo en muestras complejas.	68
4.12	Algoritmo de cálculo de los estimadores	69
4.12.1	Algoritmo de cálculo de las varianzas en muestras complejas	71
4.12.2	Variables que identifican el diseño muestral complejo	73
4.13	Coefficiente de variación	74
5.	METODOLOGÍA DE RECOLECCIÓN Y PROCESAMIENTO DE DATOS	75
5.1	Sensibilización	75
5.2	Métodos de recolección de los datos	75
5.2.1	Periodicidad de la recolección de los datos	75
5.2.2	Período de recolección	76
5.2.3	Período de referencia de los datos	76
5.2.4	Tipo de entrevista	76
5.2.5	Criterios de logro	77
5.2.6	Medio de recolección de información	77
5.2.7	Conformación del equipo de trabajo	77
5.2.8	Métodos para incentivar la respuesta	78
5.2.9	Mecanismo de resolución de dudas	78
5.3	Diseño de formularios	78
5.3.1	Cuestionario del Hogar (CH)	78
5.3.2	Cuestionario de Uso del Tiempo (CUT)	79
5.4	Método de capacitación	82
5.5	Prueba piloto	83
5.6	Métodos de supervisión	85
5.7	Métodos de procesamiento de datos	85
5.7.1	Digitación	85
5.7.2	Análisis	86
6.	METODOLOGÍA DE ANÁLISIS	87
6.1	Análisis de información para la mejora de la base de datos	87
6.2	Análisis de suficiencia de información	87
6.3	Tratamiento de outliers de tiempo.	88
6.4	Definición sobre la imputación de datos de tiempo faltantes	89
6.5	Tratamiento de <i>outliers</i> de ingresos.	89
6.6	Imputación de ingresos del trabajo y jubilaciones	90
6.7	Análisis de comportamiento de resultados	91
7.	PRODUCTOS ESTADÍSTICOS	92
8.	BIBLIOGRAFÍA	93

1. INTRODUCCIÓN

El objetivo de este documento es informar a las y los usuarios sobre las principales decisiones metodológicas tomadas en las diferentes etapas de la producción de la Encuesta Nacional sobre Uso del Tiempo (ENUT) 2015.

El informe presenta los principales antecedentes del estudio, que incluye el contexto en que surge la encuesta y la necesidad de generar información sobre el tiempo que destinan las personas a los diferentes tipos de actividades que se consideran en la medición.

El segundo tema corresponde al marco de referencia utilizado para la implementación de la encuesta y profundiza tanto en los antecedentes conceptuales como en la descripción de la población estudiada.

El tercer punto aborda el diseño estadístico, en el que se explican las etapas asociadas al diseño muestral de la encuesta, así como el cálculo y ajustes de los factores de expansión.

El cuarto punto corresponde a la descripción de la metodología de recolección y procesamiento de datos. Aquí se detalla el procedimiento de generación de información en terreno considerando las estrategias metodológicas que se definieron para el resguardo de la calidad del dato producido.

En el quinto apartado del documento se mencionan las definiciones metodológicas asociadas al tratamiento de la base de datos. En este punto se destacan los controles de la calidad de la producción estadística desde la etapa de depuración y las definiciones asociadas al cálculo de los principales estimadores.

Finalmente, se detallan los medios utilizados para la difusión de los resultados de la encuesta.

2. CONTEXTO

2.1 ANTECEDENTES DEL ESTUDIO

La calidad de vida de las personas y su bienestar se ven afectados por el uso del tiempo, según qué actividades se priorizan y cómo se organizan. El tiempo, en tanto recurso, es lineal y limitado. Su uso está mediado por condicionantes sociales, como aspectos socioeconómicos, género, ciclo de vida, cultura, entre otros. Por tanto, al no ser una elección individual, este se distribuye de manera desigual en la población.

Las encuestas de Uso del Tiempo tienen como contenido básico las actividades que realizan las personas y el tiempo que dedican a estas, permitiendo conocer la forma en que una población organiza y distribuye el uso de este recurso.

2.1.1 Tiempo, desigualdades socioeconómicas y de género

A pesar que existe un ordenamiento estándar del tiempo a través de un año calendario en 365 días, semanas de siete días y las 24 horas de un día, el tiempo puede ser entendido como un recurso cuya disposición presenta un alto contenido social y de género. Siendo en sí mismo un medio que facilita o restringe las posibilidades de determinar las actividades a realizar, lo que repercute en el bienestar económico y subjetivo de las personas.

Los cambios en el mercado del trabajo han afectado las exigencias sociales respecto a la determinación del tiempo. El aumento de la participación femenina en el mercado laboral, sin que por ello se redistribuyan significativamente las tareas domésticas, conlleva que las mujeres deban asumir una doble o triple jornada de trabajo. Adicionando al trabajo en el mercado laboral los quehaceres domésticos y el cuidado de personas.

El fenómeno descrito presenta un importante sesgo socioeconómico, en tanto las mismas oportunidades de inserción al mercado laboral se ven afectadas por la posibilidad de acceso a redes de apoyo, familiar, público o de mercado, para la realización de las labores domésticas. Las conocidas brechas de género en el mercado laboral (en términos de participación y remuneraciones), junto a la debilidad de estas redes, repercuten en una reproducción de los roles de género y -como señala Juan Carlos Feres (2009)- en un aumento de la vulnerabilidad de las personas.

2.2 NECESIDADES DE INFORMACIÓN

Las encuestas de uso del tiempo son la mejor estrategia para medir el tiempo que la población dedica a las distintas actividades de interés en un período de referencia determinado. En ese sentido, con la medición del tiempo que las personas destinan al trabajo realizado en los hogares sin mediar remuneración, la ENUT 2015 avanza en el desafío planteado por la 19° Conferencia Internacional de Estadísticos del Trabajo (19° CIET) de generar información sobre las distintas formas de trabajo, independientemente de si su realización es con fines de mercado o para el consumo de los propios hogares.

Del mismo modo, permite generar información sobre el tiempo destinado a actividades como el cuidado personal, la vida social o los estudios. Contracara del tiempo destinado al trabajo, y un elemento central del bienestar en su dimensión de equilibrio vida-trabajo.

A partir de lo anterior, los datos de la ENUT 2015 permiten alinearse con las mediciones ya realizadas en la mayor parte de los países de la región y caracterizar las brechas existentes entre hombres y mujeres, así como entre otros grupos de la población en el tiempo destinado a las distintas actividades de trabajo o personales.

En relación específicamente a las desigualdades de género, la información proporcionada por la ENUT 2015 permite que las estadísticas oficiales de Chile cumplan con distintos compromisos internacionales asociados al adelanto de las mujeres, como los adquiridos en las Conferencias de la mujer, la Convención sobre la eliminación de todas las formas de discriminación contra la mujer (CEDAW), la Plataforma de Acción de Beijing, Beijing+5, entre otras.

2.3 OBJETIVOS DEL ESTUDIO O PROCESO ESTADÍSTICO

2.3.1 Objetivo general

Obtener información sobre el uso del tiempo de la población de 12 años y más, respecto a actividades de trabajo realizadas en el mercado, el trabajo no remunerado y las actividades personales.

2.3.2 Objetivos específicos

- Visibilizar la carga global de trabajo de la población de 12 años y más a través de la medición del tiempo que se dedica al trabajo en la ocupación y sus actividades conexas y al trabajo no remunerado.
- Evidenciar las desigualdades de género en el tiempo dedicado a trabajo doméstico y de cuidados no remunerado.
- Caracterizar la desigualdad en el uso del tiempo de acuerdo

a variables de clasificación como edad, sexo, vínculo con el mercado laboral y nivel socioeconómico.

- Obtener información sobre el uso del tiempo de las personas de 12 años y más en actividades personales.

2.4 REFERENCIAS NACIONALES

La ENUT 2015 es la segunda experiencia en medición de uso del tiempo que desarrolla el INE. La primera experiencia corresponde a la “Encuesta experimental de Uso del Tiempo en el Gran Santiago (EUT) 2007”, realizada el año 2007.

Si bien se revisó esta experiencia como antecedente, el uso de dos metodologías distintas de generación de información sobre el uso del tiempo, limita la comparabilidad entre ambas experiencias.

Entre las diferencias más relevantes destacan el tipo de cuestionario utilizado, ya que la EUT 2007 aplicó un diario de actividades (cuestionario que registra la respuesta espontánea sobre las actividades realizadas a lo largo de un día en orden cronológico), mientras la ENUT 2015 utilizó un listado de actividades (cuestionario estructurado en base a una selección de actividades de interés). Por otro lado, la EUT 2007 tuvo como cobertura el área urbana de la Región Metropolitana, mientras la ENUT 2015 tiene alcance nacional urbano.

2.5 REFERENCIAS INTERNACIONALES

Para el desarrollo de la encuesta se tomaron como referencia las principales recomendaciones internacionales para la realización de encuestas de uso del tiempo.

Para la toma de definiciones sobre el tipo de cuestionario a utilizar para el relevamiento del uso del tiempo, los contenidos a incluir para su caracterización y como base conceptual y metodológica para la medición, se usó la “Guía de elaboración de estadísticas sobre el empleo del tiempo para medir el trabajo remunerado y no remunerado” (UN, 2006), el documento para discusión de “Directrices y referentes conceptuales para armonizar las encuestas sobre uso del tiempo en América Latina y el Caribe”, (Gómez, 2010), la guía de UNECE “Guidelines for Harmonising Time Use Surveys”, (UNECE, 2013). Además de la revisión de la documentación metodológica de las principales encuestas de uso del tiempo a nivel regional.

Como sustento conceptual para la organización de las actividades humanas en base al trabajo y su vínculo con el mercado, se revisó el manual del “Sistema de Cuentas Nacionales 2008”, (SCN, 2009) y el Informe de Conferencia de la 19° Conferencia Internacional de Estadísticos del Trabajo (OIT, 2013) y el Manual de medición del trabajo voluntario (OIT, 2011).

2.6 NOMENCLATURAS UTILIZADAS

La selección y ordenamiento lógico y conceptual de las actividades consultadas tomó como base la Clasificación de Actividades de Uso del Tiempo para América Latina y el Caribe CAUTAL 2015, la que fue adaptada a las necesidades de información y características nacionales de uso del tiempo. Al utilizar un listado de actividades predefinidas, la clasificación se realizó con anterioridad a la aplicación.

Para la caracterización ocupacional de los informantes se utilizó la Clasificación Internacional Uniforme de Ocupaciones (CIUO 88) para la clasificación de grupos ocupacionales. Mientras que la situación en el empleo se clasificó a partir de la Clasificación Internacional de la Situación en el Empleo (CISE 93).

3. MARCO DE REFERENCIA

3.1 MARCO CONCEPTUAL

3.1.1 La centralidad del trabajo en la vida moderna

La forma en que se organiza el trabajo es uno de los elementos determinantes de la fisonomía que adquiere una sociedad. Es uno de los elementos centrales en el establecimiento de jerarquías sociales, diferencias de género y en la conformación de identidades de los sujetos (Gómez, 2008).

Los cambios en el mercado del trabajo han generado transformaciones en la forma en que las personas organizan su tiempo para las distintas actividades, alterándose así la organización de la vida doméstica (Aguirre, 2009). La flexibilización del trabajo ha devenido en una variedad de tipos de jornada laboral, adoptando distintos niveles de formalidad o informalidad; los avances tecnológicos permiten esta flexibilización, pero a la vez exigen mayores capacidades y nuevas formas de socialización e inserción social.

El trabajo doméstico y de cuidados no remunerado, así como las transferencias de trabajo no remunerado entre hogares, también presentan características particulares, y su actual fisonomía es reflejo de los cambios en la demanda de fuerza de trabajo. Ello se ve reflejado en la masificación del empleo femenino y las posibilidades diferenciadas de situar en el mercado los servicios asociados a la reproducción social, tanto en lo relacionado a la crianza de niños y niñas, la preparación de personas para su conversión en futuros trabajadores o la reproducción normativa del sistema social.

De esta forma, con el trabajo como elemento articulador de la vida social, el estudio de los tiempos destinados a las distintas formas que este adopta, como su relación con otros tipos de actividades, resulta fundamental en la caracterización de las condiciones de vida de la población.

3.1.2 El Sistema de Cuentas Nacionales (SCN) como marco de análisis

Uno de los elementos en deuda en el desarrollo de las estadísticas sobre uso del tiempo a nivel internacional, es la armonización metodológica y conceptual de estos estudios. Los intentos por superar este déficit han llevado a distintos organismos internacionales como Naciones Unidas, ONU Mujeres o la Comisión Económica para América Latina y el Caribe CEPAL a proponer estrategias que permitan la comparabilidad de la información entre países.

De esta manera, se ha propuesto al SCN como un marco de referencia para la conceptualización, medición y clasificación de las actividades en las encuestas de uso del tiempo, en tanto el SCN permite organizar todas las actividades humanas a partir del trabajo como eje, distinguiendo entre actividades personales y productivas, y entre las productivas, entre el trabajo realizado para el mercado y el efectuado para el consumo de los propios hogares.

En línea con lo anterior, los principales clasificadores internacionales de actividades para las encuestas sobre uso del tiempo (ICATUS Y CAUTAL), utilizan el marco conceptual del SCN, que a partir de una visión integral del trabajo permite un mayor alcance para el análisis económico, social y de género.

3.1.3 La frontera general de la producción

En el aparato conceptual del SCN, la definición de producción es un elemento central que permite ordenar las distintas actividades humanas. La producción se entiende como “(...) *un proceso físico, realizado bajo la responsabilidad control y gestión de una unidad institucional, en el que se utilizan mano de obra y activos para transformar insumos de bienes y servicios en productos de otros bienes y servicios*”, (SCN, 2009).

El trabajo representa un insumo de cualquier proceso productivo, y en relación a la estructura conceptual del SCN delimita los alcances de la frontera general de la producción (Gómez, 2008). La definición de la frontera general de la producción tiene relación con una conceptualización particular del trabajo, que es analizado desde una perspectiva integral, incorporando tanto el trabajo que se realiza con fines de mercado, como el trabajo no remunerado realizado para el consumo del propio hogar, dado el impacto que genera en el mercado laboral y en los hogares.

De esta manera, a partir de su actualización 2008, el SCN “(...) *incluye conceptos y definiciones que permiten un tratamiento del trabajo no remunerado acorde con el núcleo central de las cuentas nacionales*”, (Gómez, 2008).

Este giro conceptual en la concepción de la producción implica una valorización del trabajo más allá de su papel en el mercado, visibilizando

el aporte productivo del trabajo realizado en los hogares, y el papel que este tiene en el bienestar de las personas. A partir de lo anterior, el marco conceptual de la clasificación CAUTAL utilizada por la ENUT, *“(...) incorpora criterios económicos fundamentados en el SCN y parte de distinguir dos grandes conjuntos: las actividades productivas, vinculadas a la producción de bienes y servicios y las actividades no productivas o personales”*, (CAUTAL, 2015).

3.1.4 La frontera de la producción del SCN: producción de los hogares

Si bien el marco conceptual del SCN reconoce el carácter productivo de los distintos tipos de trabajo, al mismo tiempo establece dos distinciones para su contabilidad, la Frontera General de la Producción, que incorpora el concepto de trabajo en su sentido más amplio, y la Frontera de la Producción del Sistema de Cuentas Nacionales, que comprende los bienes y servicios producidos para el mercado, la producción de bienes para uso final propio (dada la posibilidad de decidir venderlos luego de su producción, por ejemplo en caso de una cosecha abundante), algunos tipos de trabajo voluntario y otras formas de producción no de mercado, como la realizada por el gobierno o las instituciones sin fines de lucro. Mientras la producción de servicios no remunerados para el hogar, para otros hogares y voluntariados no remunerados se identifica como parte de la frontera general de la producción, pero fuera de la frontera de la producción del SCN.

Las causas de la exclusión de este segundo tipo de actividades productivas de la contabilidad del SCN son variadas, pero se pueden resumir en que *“(...) la inclusión en las cuentas de estos grandes flujos no monetarios junto con los flujos monetarios podría oscurecer lo que está sucediendo en los mercados y reducir la utilidad analítica de los datos”*, (SCN, 2009), en tanto la decisión de producir servicios para el autoconsumo no está orientada por y no incide en la política económica, ni afectan la recaudación de impuestos, entre otros.

Este efecto distorsionador de los datos sobre la producción también se expresa en las estadísticas sobre mercado laboral, ya que de acuerdo con las recomendaciones de la OIT, la población económicamente activa o PEA se define como la dedicada a actividades productivas conforme a la definición del SCN. De esta forma, un concepto ampliado de producción desde el SCN generaría el absurdo de prácticamente eliminar el desempleo, distorsionando el sentido de su medición.

Como se puede apreciar, la producción en los hogares es un elemento particularmente relevante en las definiciones del SCN. En este ámbito una diferencia fundamental es la que se establece entre los bienes y los servicios generados, dado que los primeros sí son parte de la contabilidad mientras se excluye a los segundos.

Siguiendo la conceptualización de la 19° CIET, el trabajo doméstico y de cuidados no remunerado forman parte del trabajo de producción de servicios para el autoconsumo, definida en la 19° CIET como aquella producción “(...) cuyo destino previsto es principalmente para uso final del propio productor, como formación de capital fijo, o el consumo final de los miembros del hogar o de familiares que viven en otros hogares”, (OIT, 2013).

Ejemplos de bienes generados por el hogar, destinados al autoconsumo, que sí son registrados por el SCN son la producción y almacenamiento de productos agropecuarios, o la confección de utensilios como prendas de vestir, muebles u otros accesorios o el suministro de agua. Mientras que entre las actividades productivas que son excluidas de la contabilidad del SCN destacan la limpieza, decoración y mantenimiento de la vivienda ocupada por el hogar, incluyendo reparaciones menores, la elaboración y servicio de comida, el cuidado, formación e instrucción de los hijos o algunos servicios prestados para instituciones sin fines de lucro.

3.1.5 Hacia un concepto amplio del trabajo

A contar de la 19° Conferencia Internacional de Estadísticos del Trabajo (19° CIET), la OIT define trabajo como “(...) todas las actividades realizadas por personas de cualquier sexo y edad con el fin de producir bienes o prestar servicios para el consumo de terceros o para uso final propio”, (OIT, 2013), independientemente de la legalidad, formalidad o la mediación de pago por la actividad realizada. Definición que se encuentra en línea con las fronteras de la producción establecidas por el Sistema de Cuentas Nacionales del año 2008.

Una de las precisiones más relevantes al momento de hablar de trabajo o de su relación con otros tipos de actividades, es la delimitación del espectro de actividades posibles de catalogar como trabajo. Tradicionalmente “(...) se realiza un uso indistinto de las nociones de trabajo y empleo”, (Aguirre, 2009) restringiendo el alcance de este concepto a las actividades que tienen un reconocimiento social asociado al carácter de mercado del trabajo.

Al ampliar la noción del trabajo desde el empleo (o trabajo en la ocupación, según la 19° CIET) a una concepción que supere los criterios de mercado, se puede establecer que las distintas actividades realizadas por la población pueden distinguirse entre productivas, caracterizadas por cumplir con el llamado criterio de tercera persona, es decir “(...) si su realización puede encomendarse a otra persona con los mismo resultados deseados”; y no productivas o actividades personales, (UN, 2006).

Con esta nueva conceptualización, la 19° CIET identifica cinco formas de trabajo:

- a) El trabajo en la ocupación, realizado para terceros a cambio de una remuneración.
- b) El trabajo en formación no remunerado, realizado para terceros con la finalidad de adquirir experiencia.
- c) El trabajo voluntario, no remunerado y realizado para terceros.
- d) El trabajo de producción de bienes y servicios para el autoconsumo, no remunerado y realizado para el uso propio (de la persona o el hogar).
- e) Otras actividades productivas no definidas.

Desde esta lógica, es posible entender el trabajo como cualquier actividad que realiza una persona con la finalidad de obtener bienes y servicios, convirtiéndolo en un insumo de la producción, independientemente de que este trabajo sea o no remunerado. En ese sentido, Gómez planteará que *“(...) desde el punto de vista económico, el trabajo es la fuente del valor y de la riqueza económica”*, (Gómez, 2008).

Esta diferenciación es fundamental, ya que la contribución de los distintos tipos de actividades al bienestar de las personas y sociedades se sustenta tanto en el trabajo realizado en contextos de mercado como en las actividades productivas desarrolladas fuera de este, en las relaciones privadas y en las organizaciones sociales (Aguirre, 2009).

Desde la mirada de las estadísticas oficiales, el tránsito desde concepciones estrechas hacia un concepto integrador del trabajo se puede identificar desde la 18° Resolución sobre la medición del trabajo, donde se recomienda que

“(...) se mida todo el trabajo que se realiza en el país, independientemente de si el trabajo es remunerado o no remunerado, pues sólo (sic) un panorama amplio e integral podrá proporcionar mejores elementos para el diseño de las políticas laborales y de conciliación entre la actividad económica y vida familiar”, (OIT, 2008).

En esa lógica, la medición del tiempo destinado a la realización de los distintos tipos de trabajo, aparece como un elemento fundamental para dar cuenta de los objetivos establecidos por la OIT en torno a la medición de la participación en todas las formas de trabajo. Ya que estas actividades productivas en conjunto contribuyen al desarrollo económico, al sustento de los hogares y al bienestar de los individuos y la sociedad, así como al estudio del impacto económico y social que genera la participación de los distintos subgrupos poblacionales en las distintas formas de trabajo (OIT, 2013).

3.2 DIMENSIONES DE ANÁLISIS PARA LA ENUT 2015

A continuación se presentan las dimensiones de análisis abordadas en la medición de uso del tiempo, los tipos de actividades que las componen, siguiendo la estructura propuesta por CAUTAL; y la organización que estas tienen en los instrumentos de medición.

Como se especificará más adelante, la ENUT 2015 cuenta con dos instrumentos de medición: el Cuestionario del Hogar (CH), que registra las características sociodemográficas de los integrantes del hogar; y el Cuestionario de Uso del Tiempo (CUT), que contiene los módulos de uso del tiempo junto a la caracterización ocupacional de cada integrante del hogar de doce años o más y un módulo de percepción asociado a la satisfacción con el uso del tiempo.

Como el objetivo de la encuesta es medir el tiempo que destina la población a las distintas actividades y caracterizar las diferencias en el uso del tiempo entre las distintas subpoblaciones. Las principales dimensiones de análisis corresponden a los tipos de actividades sobre los cuales se organizó la medición de uso del tiempo.

3.2.1 Trabajo en la ocupación, producción de bienes para el autoconsumo (actividades dentro de la frontera de la producción del SCN) y actividades conexas

La primera dimensión de análisis y primera sección de la clasificación CAUTAL, incluye las actividades de trabajo relacionadas con la producción de bienes y servicios considerada dentro de la frontera del Sistema de Cuentas Nacionales. En ella se agrupa a todas las actividades que hacen posible, facilitan o fomentan la producción, incluyendo los traslados por motivos de trabajo.

Para la medición del tiempo destinado a actividades de trabajo dentro de la frontera general de la producción se incorporaron las categorías de:

- Trabajo en la ocupación y desplazamientos asociados a esta actividad: sección 1 “Tiempo de trabajo”, del módulo M del Cuestionario de Uso del Tiempo “Caracterización del tiempo de trabajo y traslado”.
- Búsqueda de empleo: sección 2 “Tiempo de búsqueda de trabajo”, del módulo M del Cuestionario de Uso del Tiempo “Caracterización del tiempo de trabajo y traslado”.
- Producción de bienes para el autoconsumo del hogar: preguntas O42 “Realizar actividades de construcción, ampliación o reparaciones mayores de la vivienda” y O63 “Recoger leña para calefaccionar o cocinar”, del módulo O del CUT “Trabajo doméstico”.

3.2.2 Trabajo no remunerado (fuera de la frontera de la producción del SCN)

La segunda dimensión de análisis y segunda sección de CAUTAL, es el trabajo no contabilizado por el SCN, parte de la frontera general de la producción. Correspondiente a la producción de servicios no remunerados para el hogar, para otros hogares y voluntariados no remunerados.

La medición del tiempo destinado a las actividades de trabajo fuera de la frontera de la producción del SCN se organiza en las siguientes subdimensiones:

- Cuidados no remunerados a miembros del hogar: Módulo N “Cuidados a integrantes del hogar” del CUT. Este módulo se estructura en función de las edades o características de salud de los integrantes del hogar, distinguiendo según la naturaleza del trabajo de cuidado, a sujetos de cuidado permanente de los otros integrantes del hogar. Comprende las secciones de:
 - ✓ Cuidados a personas que requieren cuidados de salud permanentes.
 - ✓ Cuidados a niñas y niños de 0 a 4 años.
 - ✓ Cuidados a niñas y niños de 5 a 14 años.
 - ✓ Cuidados a personas de 15 a 65 años.
 - ✓ Cuidados a personas de 66 años y más.
- Trabajo no remunerado en quehaceres del propio hogar: Módulo O “Trabajo doméstico” del CUT, excluyendo las actividades de autoproducción de bienes, que corresponden a la dimensión de Trabajo en la ocupación y producción de bienes para el autoconsumo. Comprende las secciones de:
 - ✓ Preparación y servicio de comida.
 - ✓ Limpieza de la vivienda.
 - ✓ Limpieza y cuidado de ropa y calzado.
 - ✓ Mantenimiento y reparaciones menores en el propio hogar.
 - ✓ Administración del hogar.
 - ✓ Abastecimiento del hogar.
 - ✓ Cuidado de mascotas y plantas.
- Trabajo no remunerado para otros hogares, para la comunidad y trabajo voluntario no remunerado: Módulo P “Ayuda no remunerada a otros hogares”. Comprende dos tipos de actividades diferenciadas según el destinatario de los trabajos:

- ✓ Quehaceres domésticos para otros hogares (incluye trabajo doméstico y de cuidados).
- ✓ Actividades comunitarias.

3.2.3 Actividades personales (fuera de la frontera general de la producción del SCN)

La tercera dimensión de análisis y sección de CAUTAL corresponde a las actividades no productivas. La ENUT 2015 tiene como foco el estudio del tiempo destinado al trabajo en toda su dimensión, siendo la realización de actividades personales el complemento del tiempo dedicado a las actividades productivas. Ello implica que el tiempo dedicado a estas actividades dependerá de cuánto tiempo se le dedica a las actividades de trabajo, lo que repercute directamente en la calidad de vida de las personas. Esta sección se compone de tres subdimensiones:

- Cuidados personales y necesidades fisiológicas: módulo Q “Cuidados personales”.
- Actividades de educación y aprendizaje: módulo R del CUT, que distingue según la formalidad de la instrucción entre asistencia al establecimiento educacional (sección 1) y Otras actividades de aprendizaje (sección 2).
- Ocio y vida social: correspondiente al módulo S del CUT y que incluye las secciones de Vida social (sección 1), Asistencia a eventos (sección 2), Juegos y aficiones (sección 3), Deportes (sección 4) y Uso de medios de comunicación (sección 5).

3.3.4 Satisfacción con el uso del tiempo

La cuarta dimensión de análisis corresponde a la evaluación de las y los informantes sobre su propio nivel de satisfacción con el tiempo destinado ellas.

A diferencia de los anteriores, el Módulo T, de “Satisfacción con el uso del tiempo” no indaga en el tiempo destinado a actividades, sino que en una serie de aspectos relacionados a la percepción de las personas respecto a su calidad de vida y la relación del tiempo destinado a los distintos tipos de actividades y su propia satisfacción con ello.

Este módulo cuenta con una sola sección.

4. DISEÑO ESTADÍSTICO

4.1 UNIVERSO

El universo de estudio lo constituye todas las personas de 12 años y más pertenecientes a las comunas en estudio.

4.2 POBLACIÓN OBJETIVO Y COBERTURA

La población objetivo la constituyen todas las personas de 12 años y más, que tengan condiciones físicas y mentales aptas para responder la encuesta, que residan habitualmente en viviendas particulares ocupadas (elegibles) y que se encuentren en estas durante el período de levantamiento. Esta población involucra al área urbana de todas las regiones del país, y específicamente pertenece a las principales comunas que acumulan el 85% de la población total nacional. Esto equivale a 118 comunas.

4.3 UNIDADES ESTADÍSTICAS

4.3.1. Unidad de observación

Las unidades de observación corresponden a los integrantes del hogar de 12 años o más, presentes en la vivienda durante el período de levantamiento de la encuesta, con capacidades físicas y mentales aptas para responder.

4.3.2. Unidad de análisis

Las unidades de análisis corresponden a personas, hogares y las actividades que estos declaran al momento de la entrevista.

4.3.3. Unidad de muestreo

Al ser una muestra bietápica, las unidades corresponden a manzanas como unidades de primera etapa, y viviendas como unidades de segunda etapa.

4.4 CONSTRUCCIÓN DE ÍNDICES O METODOLOGÍAS COMPUESTAS

Los indicadores de interés en ENUT, se refieren en general a la relación entre grupos poblacionales, la participación en la realización de determinadas actividades o categorías de actividades y el tiempo destinado a ellas.

Para la construcción del plan de indicadores centrales de la ENUT 2015 se definieron dos medidas sobre el uso del tiempo:

- La participación en la realización de una actividad

- El tiempo destinado a la realización de una actividad

Estas medidas generalmente se expresan en términos de promedios o proporciones, y corresponden a la relación entre la variable de análisis (la actividad declarada) y la población analizada.

La construcción de estos 2 indicadores se describe a continuación:

- Tasa de participación: corresponde al porcentaje de personas que realizaron una determinada actividad. Es la relación entre las personas que declararon realizar una actividad y el total de personas consultadas por dicha actividad.

$$\text{Participación en la actividad} = \frac{\text{Total de personas que realizaron la actividad}}{\text{Total de personas que se les consultó por la actividad}} \cdot 100$$

- Tiempo promedio por participante: corresponde a la media de tiempo destinado a una determinada actividad, entre las personas que efectivamente la realizaron y declaran tiempo para dicha actividad.

$$\text{Tiempo promedio por participante} = \frac{\text{Total de tiempo reportado en la actividad}}{\text{Total de personas que realizaron la actividad}}$$

4.5 DEFINICIÓN DE VARIABLES

4.5.1 Variable de estudio

Tal como se describe en el capítulo sobre definición del tamaño muestral de este documento, la variable de estudio corresponde a “Necesidades fisiológicas y cuidados personales”, sobre la cual se obtuvo la estimación y error asociado según EUT 2007, que sirvieron para el cálculo del tamaño muestral de ENUT 2015.

4.5.2 Variables de clasificación

Las variables para la construcción y análisis de indicadores de uso del tiempo referente a las subpoblaciones de interés corresponden a:

- Sexo: hombre y mujer
- Tramos de edad: 12 a 24, 25 a 45, 46 a 65 y 66 y más
- Vínculo con el mercado laboral (condición de actividad económica): ocupados, desocupados e inactivos
- Ámbito geográfico: nacional, cada una de las 15 regiones.

4.5.3 Variables de análisis

Si bien, existen múltiples estudios que se pueden realizar a partir de los resultados de esta encuesta, las variables de análisis o variables más relevantes consideradas corresponden a las definidas en el plan de tabulados 2016, las cuales se centran en los grupos de actividades de trabajo, es decir, actividades dentro de la frontera general de producción del Sistema de Cuentas Nacionales.

Según lo anterior, se definieron como prioritarias 14 dimensiones de actividades de uso del tiempo. Estas se detallan a continuación:

- Trabajo en la ocupación y producción de bienes para el autoconsumo (dentro de la frontera de la producción del SCN)
- Trabajo en la ocupación y traslados asociados
- Trabajo doméstico no remunerado para el propio hogar
- Trabajo de cuidados a integrantes del hogar que requieren cuidados permanentes de salud
- Trabajo de cuidados a integrantes del hogar de 0 a 4 años
- Trabajo de cuidados a integrantes del hogar de 5 a 14 años
- Trabajo de cuidados a integrantes del hogar de 15 a 65 años
- Trabajo de cuidados a integrantes del hogar de 66 años y más
- Trabajo de cuidados no remunerado a integrantes del hogar
- Trabajo no remunerado para otros hogares, para la comunidad y voluntario
- Trabajo no remunerado en quehaceres domésticos para otros hogares
- Trabajo no remunerado en actividades comunitarias
- Trabajo no remunerado (fuera de la frontera de la producción del SCN)
- Carga global de trabajo

4.6 METODOLOGÍA MUESTRAL

4.6.1 Marco muestral

Un marco muestral se define como un conjunto de unidades en que todos sus elementos son identificables uno a uno mediante algún procedimiento o listado, a los cuales se les puede asociar una probabilidad de selección distinta de cero de acuerdo a la estrategia de muestreo.

El marco muestral utilizado en este estudio es aquel generado a partir de cartografía digital proveniente del Censo 2002, actualizado al

año 2015¹. La información en el marco se actualiza con antecedentes de los registros administrativos asociados a nuevas construcciones otorgados por los municipios. Las unidades que componen este marco muestral en el área urbana se denominan “manzanas” y corresponden a delimitaciones geográficas fijas. En adelante, denominaremos a este marco como Marco de Manzanas (MM2015).

4.6.2 Estratificación del marco muestral

En el diseño de una muestra, la estratificación corresponde al proceso de agrupar a los elementos de una población según su homogeneidad, previo a la selección. Su propósito es mejorar la precisión estadística de los estimadores agrupando las unidades en clases homogéneas en su interior y que difieran de las características de unidades de otras clases. Los estratos deben ser mutuamente excluyentes, lo que significa que 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.

4.6.3 Estratificación geográfica

El INE cuenta con 2 listados de áreas geográficas conformadas a partir del Censo del año 2002, que en conjunto forman el marco muestral, el cual sirve de base para la selección de viviendas requeridas para las muestras de las encuestas de hogares, el cual cubre la totalidad del territorio con excepción de 22 áreas denominadas áreas de difícil acceso (ADA'S), según información del Censo de Población y Viviendas del año 2002².

Estos listados contienen las unidades geográficas del país organizadas de forma jerárquica.

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 urbano y rural. Estas áreas se encuentran definidas por la cantidad de población y en algunos casos, según la actividad económica preponderante en ellas, dando origen a las siguientes entidades:

- Ciudad (CD): centro urbano conformado por uno o un conjunto de centros urbanos adyacentes con 40.000 o más habitantes.
- Resto de Área Urbana (RAU): esta clasificación aparece cuando en una comuna existe una ciudad (CD) y uno o más centros urbanos con

1 El Marco muestral 2015 (capítulo 4.6), corresponde al marco muestral 2008, con actualización de 69 urbanos que fueron determinados para el levantamiento de la encuesta CASEN 2015. Esto equivale en términos porcentuales a una actualización de menos del 2% del total de manzanas del marco.

2 Corresponde a las comunas de: General Lagos, Colchane, Ollagüe, Juan Fernández, Isla de Pascua, Cochamó, Chaitén, Futaleufú, Hualaihué, Palena, Lago Verde, Guaitecas, O'Higgins, Tortel, Laguna Blanca, Río Verde, San Gregorio, Cabo de Hornos, Antártica, Primavera, Timaukel, Torres del Paine.

población entre 2.000 y 40.000 habitantes.

- Urbano (U): esta clasificación aparece cuando en una comuna no existe una ciudad (CD), solo centros urbanos, por lo que cada uno de estos 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 primarias³.

Cada listado está conformado por una o más de estas entidades. Así, el listado que contiene las áreas urbanas (ciudades y urbanos en su mayoría, y algunas de las áreas resto de área urbana) corresponde al MM2015, definido por subdivisiones denominadas manzanas. Por otro lado, el listado que contiene las entidades rurales y el resto de las áreas urbanas (ciudades, urbanos y resto de áreas urbanas) que no están en el MM2015, se denomina Marco de Secciones 2002 (MS2002), definido, como su nombre lo indica, por subdivisiones denominadas secciones.

Estas (manzanas y secciones) son las unidades geográficas más pequeñas y corresponden a las unidades de muestreo mayormente utilizadas en las encuestas de hogares diseñadas por el INE.

Cabe mencionar que para esta encuesta solo se consideró el área urbana de aquellas comunas que a nivel nacional concentran aproximadamente el 85% de la población urbana total, concentradas específicamente en 118 comunas cuyas unidades se encuentran en el MM2015 y, por lo tanto, las unidades seleccionadas corresponden a manzanas.

Si bien el marco del INE tiene cuatro estratos geográficos (CD, RAU, U y Rural), para la encuesta ENUT se trabajó considerando como variable de estratificación el área urbana (División Censal) como un todo.

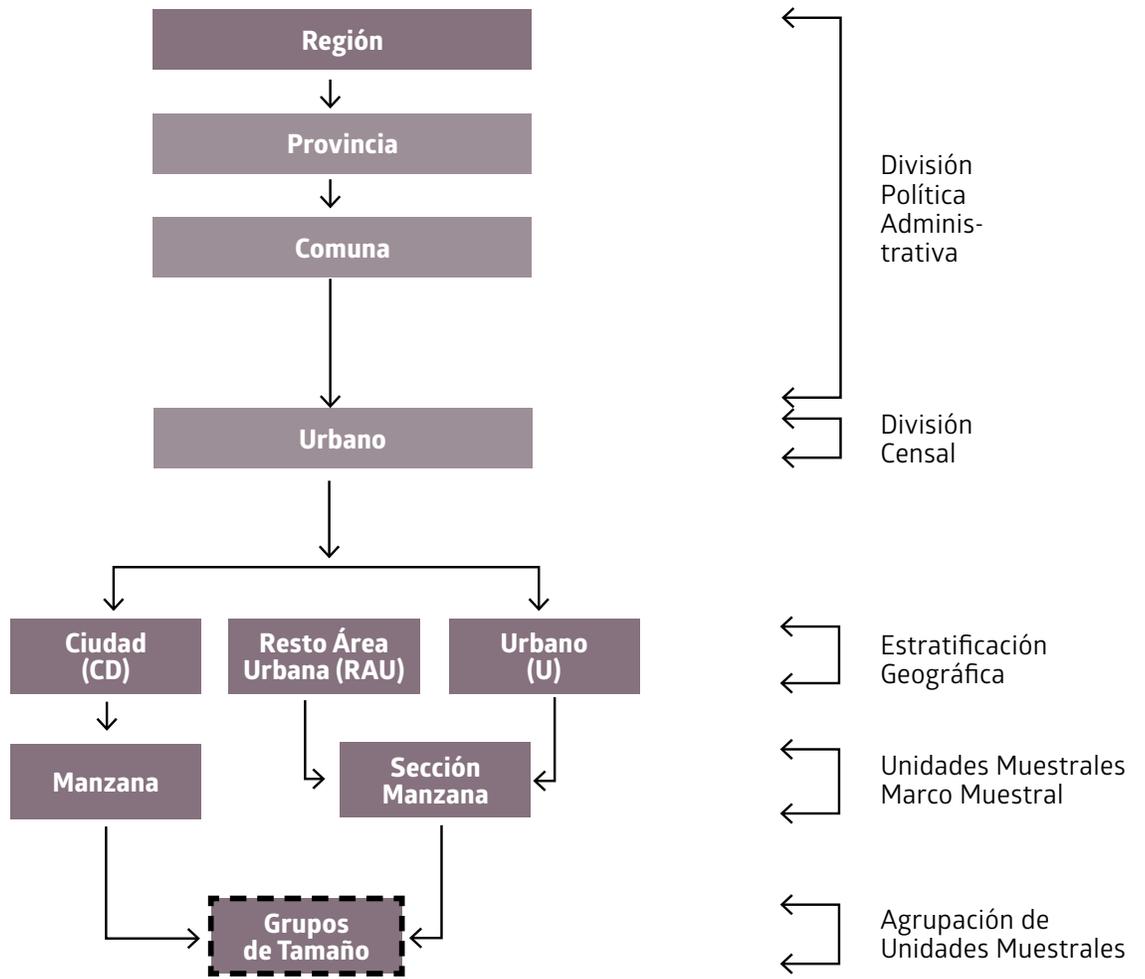
4.6.4 Estratificación del MM 2015 por grupos de tamaño

Para este estudio, el marco muestral de manzanas fue estratificado en 5 grupos de tamaño, según el número de viviendas que contienen. Esto significa que cada manzana del marco está estratificada en forma geográfica (división político-administrativa) y además por estos 5 grupos. Luego, los nuevos estratos de muestreo para el área urbana corresponden a la intersección del área urbana de las 118 comunas con los “Grupos de Tamaño”.

A continuación, se muestra la estructura del marco urbano descrita anteriormente:

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

Figura N° 1



En el cuadro a continuación, se puede observar la composición de cada uno de los grupos de tamaño.

Cuadro N° 1: Composición de grupos de tamaño en cuanto a número de manzanas y viviendas

Grupo de tamaño	Total viviendas por grupo	Número de manzanas	Número de viviendas	% del total de manzanas	% del total de viviendas
Total		119.466	3.947.184	100,00%	100,00%
1	8-23	63.319	994.825	53,00%	25,20%
2	24 - 44	39.267	1.216.764	32,90%	30,80%
3	45 - 81	10.334	591.573	8,70%	15,00%
4	82 - 154	3.990	439.327	3,30%	11,10%
5	155 - 1.341	2.556	704.695	2,10%	17,90%

Por ejemplo, el grupo N°1, que es el grupo que contiene las manzanas más pequeñas, está formado por manzanas cuyo tamaño oscila entre 8 y 23 viviendas. Dentro de este grupo, el número de manzanas en el marco alcanza a 63.319, que equivale a 53%; y el número de viviendas que concentra es de 994.825, que equivale a 25,2%. Por otro lado, el grupo n° 5, es el grupo que contiene las manzanas con mayor número de viviendas, este número oscila entre 155 y 1.341 viviendas. Dentro de este grupo el número de manzanas en el marco, alcanza a 2.556, que equivale a 2,1%; y el número de viviendas que concentra es de 704.695, que equivale a 17,9%.

4.6.5 Cobertura del marco muestral

La cobertura es una propiedad estadística asociada al marco muestral que se utiliza para la selección de la muestra. En el caso de ENUT 2015, las 118 comunas consideradas abarcan 86,9% de las manzanas del marco urbano; y 89,2% de las viviendas, tal como se muestra a continuación en el cuadro N°2.

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. 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. Para la selección de las muestras realizadas por el INE, se efectúan exclusiones intencionadas de aquellas manzanas con menos de 8 viviendas debido a potenciales problemas operativos y de costos. A continuación se detalla la incidencia de esto en el MM2015.

Cuadro N° 2: Cobertura total área urbana de comunas seleccionadas en relación al MM2015

Región	MM2015		Marco 118 comunas seleccionadas		% Cobertura urbana del MM2015	
	Número de manzanas	Número de viviendas	Número de manzanas	Número de viviendas	Número de manzanas	Número de viviendas
Total País	119.466	3.947.184	103.855	3.519.034	86,90%	89,20%
Arica y Parinacota	2.123	51.773	2.123	51.773	100,0%	100,0%
Tarapacá	2.004	69.510	1.900	67.328	94,8%	96,9%
Antofagasta	4.904	129.422	4.561	121.617	93,0%	94,0%
Atacama	3.237	71.817	2.352	55.521	72,7%	77,3%
Coquimbo	5.918	149.351	5.162	132.528	87,2%	88,7%
Valparaíso	16.255	495.190	14.187	435.748	87,3%	88,0%
Metropolitana	43.369	1.774.906	39.982	1.659.903	92,2%	93,5%
O'Higgins	5.072	156.319	3.500	113.201	69,0%	72,4%
Maule	6.089	182.483	5.211	157.924	85,6%	86,5%
Biobío	14.599	438.329	11.630	364.841	79,7%	83,2%
La Araucanía	6.309	167.122	5.092	134.599	80,7%	80,5%
Los Ríos	2.261	65.746	1.781	53.290	78,8%	81,1%
Los Lagos	4.617	134.446	3.968	116.039	85,9%	86,3%
Aysén	858	19.295	858	19.295	100,0%	100,0%
Magallanes	1.851	41.475	1.548	35.427	83,6%	85,4%

Cuadro N° 3: Total de manzanas y viviendas MM2015, según intervalo de clasificación de viviendas

Región	Intervalo de clasificación	Número de manzanas	Número de viviendas	% del total de manzanas	% del total de viviendas
Total	Menor igual 7 viviendas	13.894	53.578	10,4%	1,3%
	8 o más viviendas	119.466	3.947.184	89,6%	98,7%
	Total	133.360	4.000.762	100,0%	100,0%
ARICA Y PARINACOTA	Menor igual 7 viviendas	292	1.092	12,1%	2,1%
	8 o más viviendas	2.123	51.773	87,9%	97,9%
	Total	2.415	52.865	100,0%	100,0%
TARAPACÁ	Menor igual 7 viviendas	194	652	8,8%	0,9%
	8 o más viviendas	2.004	69.510	91,2%	99,1%
	Total	2.198	70.162	100,0%	100,0%
ANTOFAGASTA	Menor igual 7 viviendas	710	2.524	12,6%	1,9%
	8 o más viviendas	4.904	129.422	87,4%	98,1%
	Total	5.614	131.946	100,0%	100,0%
ATACAMA	Menor igual 7 viviendas	763	2.869	19,1%	3,8%
	8 o más viviendas	3.237	71.817	80,9%	96,2%
	Total	4.000	74.686	100,0%	100,0%
COQUIMBO	Menor igual 7 viviendas	899	3.605	13,2%	2,4%
	8 o más viviendas	5.918	149.351	86,8%	97,6%
	Total	6.817	152.956	100,0%	100,0%
VALPARAÍSO	Menor igual 7 viviendas	3004	12.588	15,6%	2,5%
	8 o más viviendas	16.255	495.190	84,4%	97,5%
	Total	19.259	507.778	100,0%	100,0%
METROPOLITANA	Menor igual 7 viviendas	2658	9.116	5,8%	0,5%
	8 o más viviendas	43.369	1.774.906	94,2%	99,5%
	Total	46.027	1.784.022	100,0%	100,0%
O'HIGGINS	Menor igual 7 viviendas	809	3.071	13,8%	1,9%
	8 o más viviendas	5.072	156.319	86,2%	98,1%
	Total	5.881	159.390	100,0%	100,0%
MAULE	Menor igual 7 viviendas	614	2.572	9,2%	1,4%
	8 o más viviendas	6.089	182.483	90,8%	98,6%
	Total	6.703	185.055	100,0%	100,0%
BIOBÍO	Menor igual 7 viviendas	1893	7.481	11,5%	1,7%
	8 o más viviendas	14.599	438.329	88,5%	98,3%
	Total	16.492	445.810	100,0%	100,0%

Cuadro N° 3: Total de manzanas y viviendas MM2015, según intervalo de clasificación de viviendas

Región	Intervalo de clasificación	Número de manzanas	Número de viviendas	% del total de manzanas	% del total de viviendas
LA ARAUCANÍA	Menor igual 7 viviendas	849	3.261	11,9%	1,9%
	8 o más viviendas	6.309	167.122	88,1%	98,1%
	Total	7.158	170.383	100,0%	100,0%
LOS RÍOS	Menor igual 7 viviendas	286	1.044	11,2%	1,6%
	8 o más viviendas	2.261	65.746	88,8%	98,4%
	Total	2.547	66.790	100,0%	100,0%
LOS LAGOS	Menor igual 7 viviendas	604	2.317	11,6%	1,7%
	8 o más viviendas	4.617	134.446	88,4%	98,3%
	Total	5.221	136.763	100,0%	100,0%
AYSÉN	Menor igual 7 viviendas	132	487	13,3%	2,5%
	8 o más viviendas	858	19.295	86,7%	97,5%
	Total	990	19.782	100,0%	100,0%
MAGALLANES	Menor igual 7 viviendas	187	899	9,2%	2,1%
	8 o más viviendas	1.851	41.475	90,8%	97,9%
	Total	2.038	42.374	100,0%	100,0%

Si los fenómenos de interés no están relacionados con el tamaño de las manzanas o no están relacionados con la aislación geográfica, entonces es probable que la falta de cobertura del marco muestral no afecte las inferencias hechas a partir de los datos de la encuesta en cuestión. En el caso de existir evidencia de una relación entre los fenómenos de estudio y las variables mencionadas, entonces los investigadores deberán ser cautelosos al momento de hacer inferencia a la población bajo estudio⁴.

4.7 DISEÑO MUESTRAL

4.7.1 Tipo de muestreo

El diseño muestral de la ENUT 2015 responde a las características de un diseño probabilístico, bietápico y estratificado geográficamente y por tamaño, en cuanto al número de viviendas en el área urbana.

Las etapas del diseño se describen a continuación:

- Unidad primaria de muestreo: Corresponde a manzanas
- Unidad secundaria de muestreo: Corresponde a viviendas particulares ocupadas.

⁴ Para mayor detalle de incidencia a nivel de comunas seleccionadas ver anexo I.

En cuanto a los niveles de estimación, el diseño muestral fue elaborado con el objetivo de obtener estimaciones con validez estadística de aquellas variables más relevantes asociadas al uso del tiempo. Generalmente, si se cuenta con información de una encuesta anterior, se utiliza una variable referencial para fijar los criterios de precisión de los niveles de estimación. Para esta encuesta, se utilizó la variable “Necesidades fisiológicas y cuidados personales”, proveniente de los resultados de EUT 2007.

Esta encuesta, solo preveía información del Gran Santiago, por lo que hubo que extrapolar los errores observados en todas las demás regiones del país, de acuerdo al cambio proporcional no lineal del tamaño de viviendas del MM2015.

De acuerdo a la fijación de los criterios de precisión y la observación de estos en la encuesta anterior, se determinó que las estimaciones tendrán validez estadística solo para los niveles de estimación formados por el área Nacional Urbana y Regional Urbana.

4.7.2 Tamaño muestral teórico

Para el cálculo del tamaño muestral se estudiaron distintos escenarios, los cuales en un principio consideraban el área urbana y rural de las áreas en estudio. Sin embargo, por razones presupuestarias, finalmente se optó por una muestra que representara solo el área urbana de las áreas en cuestión.

El tamaño calculado, se basó en los resultados obtenidos en EUT 2007, levantada solo en la Región Metropolitana, en las comunas del Gran Santiago. La variable de interés correspondió a “necesidades fisiológicas y cuidados personales”, cuyos datos fueron extrapolados al resto de las regiones, a fin de contar con información para todo el país.

Para el cálculo del tamaño se consideró representatividad a nivel regional urbano. El tamaño en las regiones fue asignado según la distribución del marco muestral; al igual que la asignación de los tamaños a nivel comunal.

Así, se obtuvo un tamaño total nacional de 11.623 viviendas lo que asegura un error relativo para la variable de interés correspondiente a 0,55% a nivel nacional.

La fórmula utilizada para el cálculo de los tamaños a nivel regional corresponde a:

$$m = \frac{k^2 S^2}{e^2} Def f = 11.623 \text{ viviendas}$$

Donde:

k : Percentil de la distribución Normal asociado a un intervalo de confianza bilateral del 95% (este valor corresponde a 1,96).

S^2 : Varianza de la variable de interés.

e : Error absoluto requerido para la precisión deseada.

$Deff$: Factor de ajuste del muestreo bietápico llamado Efecto del Diseño.

Los tamaños a nivel regional se muestran en el siguiente cuadro:

Cuadro N°4: Tamaño objetivo de viviendas a nivel regional

Región	N° viviendas objetivo
Total país	11.623
Arica y Parinacota	186
Tarapacá	249
Antofagasta	453
Atacama	271
Coquimbo	665
Valparaíso	1.150
Metropolitana	3.450
O'Higgins	797
Maule	900
Biobío	1.150
La Araucanía	890
Los Ríos	344
Los Lagos	686
Aysén	277
Magallanes	155

A estos tamaños objetivos, se les incorporó un porcentaje adicional para cubrir los efectos de la no respuesta. Este porcentaje fue calculado a nivel comunal, considerando en la mayoría de las comunas 30%; y en comunas que históricamente han presentado bajos niveles de respuesta, 50%. A nivel nacional, el número de viviendas adicionales correspondió a 3.689, por lo que en total de gestionaron 15.312 viviendas.

Los tamaños con sobremuestreo a nivel regional se muestran en el siguiente cuadro:

Cuadro N°5: Tamaño de viviendas con sobremuestreo a nivel regional

Región	N° viviendas objetivo	N° viviendas con sobremuestreo
Total país	11.623	15.312
Arica y Parinacota	186	242
Tarapacá	249	324
Antofagasta	453	589
Atacama	271	353
Coquimbo	665	864
Valparaíso	1.150	1.506
Metropolitana	3.450	4.670
O'Higgins	797	1.036
Maule	900	1.168
Biobío	1.150	1.497
La Araucanía	890	1.157
Los Ríos	344	448
Los Lagos	686	893
Aysén	277	361
Magallanes	155	204

4.7.3 Medidas de precisión

A partir de los resultados de EUT 2007, se desarrolló un set de simulaciones, con el objetivo de estudiar distintos escenarios. Estos fueron considerados según el cruce de 10 actividades en estudio con las siguientes variables:

- Sexo
- Tramo de edad (12 a 18, 19 a 59, 60 o más)
- Día (lunes-viernes y sábado-domingo).

Los cruces que se realizaron fueron:

- Sexo / tramo edad
- Día / tramo edad
- Sexo / día.

Esto permite tener antecedentes en los niveles de representatividad de los distintos escenarios según variables:

- Sociodemográfica (sexo - edad)

- Temporal (días de la semana)
- Geográfica (nacional / regional).

Sin embargo, la encuesta proporcionará información más precisa para la variable de interés anteriormente mencionada: “Necesidades fisiológicas y cuidados personales”.

Con el tamaño muestral objetivo, es posible obtener estimaciones confiables a nivel nacional urbano y regional urbano, esperándose un error relativo de 0,55% para el total nacional. A nivel de regiones, se espera obtener errores relativos que oscilen entre 1,01% y 4,75% para la Región Metropolitana y Magallanes, respectivamente. Los errores relativos esperados por región se presentan en el cuadro N°6.

Cuadro N°6: Tamaños y errores relativos

Región	N° viviendas objetivo	N° viviendas sobre-muestreo	Error relativo %
Total país	11.623	15.312	0,55
Arica y Parinacota	186	242	4,34
Tarapacá	249	324	3,75
Antofagasta	453	589	2,78
Atacama	271	353	3,59
Coquimbo	665	864	2,29
Valparaíso	1.150	1.506	1,74
Metropolitana	3.450	4.670	1,01
O'Higgins	797	1.036	2,1
Maule	900	1.168	1,97
Biobío	1.150	1.497	1,74
La Araucanía	890	1.157	1,98
Los Ríos	344	448	3,19
Los Lagos	686	893	2,26
Aysén	277	361	3,55
Magallanes	155	204	4,75

4.8 SELECCIÓN MUESTRAL

La selección de las manzanas, se realizó de forma independiente y sistemática con igual probabilidad, al interior de cada comuna y grupo de tamaño. Luego, al interior de cada manzana se seleccionaron las viviendas de forma sistemática y con igual probabilidad.

4.8.1 Selección de unidades de primera etapa

Como se mencionó en el capítulo referente al marco muestral, el MM2015 está particionado en manzanas y estratificado en 5 grupos de tamaño que reflejan el tamaño de las manzanas en el marco en cuanto al número de viviendas que contienen. La selección de las manzanas se realiza en cada comuna de la muestra y estrato de tamaño de forma independiente, sistemática y con igual probabilidad. El procedimiento descrito permite seleccionar en la muestra manzanas grandes y pequeñas, de acuerdo a la distribución de éstas en la comuna. La selección se implementa en base a un algoritmo de selección que considera las dos etapas que se describen a continuación.

Etapa 1: distribución del total de viviendas según estrato de tamaño.

Las manzanas se agrupan en t grupos de tamaño. Esta información se cruza con las áreas urbanas de cada comuna, de modo que en cada comuna-área pueden existir a lo más 5 grupos ($i = 1, 2, \dots, z$; donde $z \leq 5$). Posteriormente, se cuenta el total de viviendas que contiene cada grupo de tamaño según información del marco y se determina la proporción de viviendas del total que contiene cada uno de estos grupos. De acuerdo a lo anterior, se asigna el total de viviendas a seleccionar en cada grupo de tamaño. Como cada grupo de tamaño tiene un número de viviendas a seleccionar fijo, se calcula el total de manzanas a seleccionar.

Etapa 2: selección de una manzana.

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

Cuadro N° 7: Grupos de tamaño de manzanas y total de viviendas a encuestar según grupo

Grupo tamaño	Total viviendas en la manzana	Número promedio de viviendas a encuestar
1	8 - 23	4
2	24 - 44	8
3	45 - 81	14
4	82 - 154	26
5	155 - 1341	30

Probabilidad de selección de las unidades de primera etapa

Dadas las características de selección de las manzanas, la probabilidad de inclusión del conglomerado i , del estrato h y grupo de tamaño g es,

$$p_{hg}(i) = \frac{n_{hg}}{N_{hg}}$$

Donde,

h : representa el índice del estrato (comuna).

g : representa el índice del grupo de tamaño.

n_{hg} : es el número de manzanas seleccionadas en el estrato h y grupo de tamaño g .

N_{hg} : es el total de manzanas en el marco en el estrato h , y grupo de tamaño g .

4.8.2 Selección de unidades de segunda etapa

Uno de los métodos aplicados para reducir los errores de cobertura de marco debido a desactualizaciones en el tiempo, es el proceso de enumeración o empadronamiento.

El proceso de enumeración se entiende como la actualización de las direcciones de viviendas y de su estado (particulares, otro uso, etc.) realizado previo a la selección de viviendas, en aquellas manzanas seleccionadas en la muestra.

Proceso de enumeración de las unidades de segunda etapa

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.

Material de trabajo del enumerador.

A cada enumerador se entrega el siguiente set de materiales,

- Plano comunal: plano de la comuna donde se encuentran identificadas todas las manzanas seleccionadas en la muestra.
- Croquis de la manzana: plano a menor escala que identifica solo la manzana seleccionada.
- Formulario de enumeración: en él se registran todas las construcciones encontradas en la manzana según el recorrido (orden en que fueron encontradas en terreno).

Figura n°1: plano comunal



Figura n°2: croquis de manzana

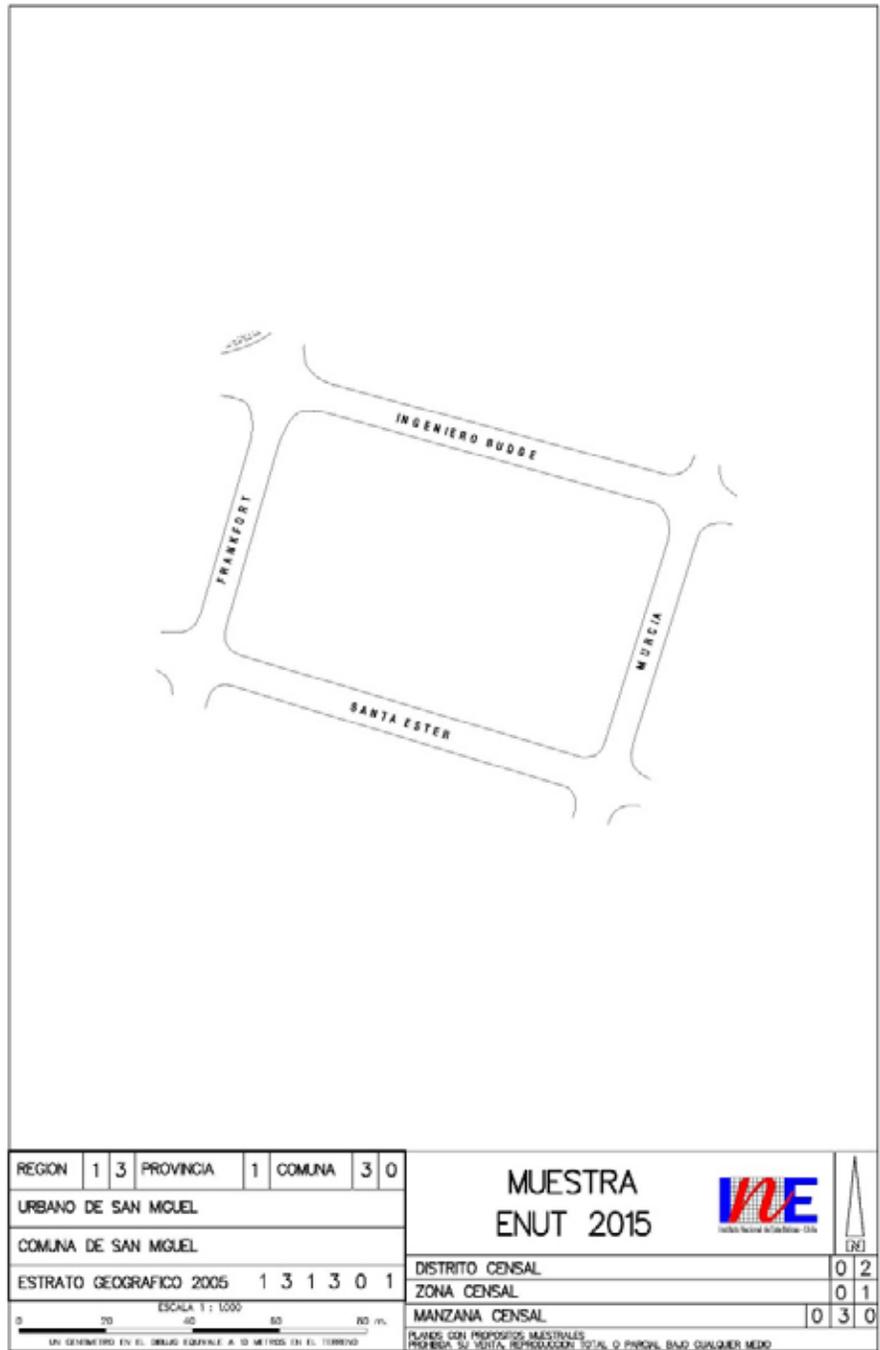


Figura n°3: Formulario de Enumeración en áreas geográficas urbanas

ENUMERACIÓN DE VIVIENDAS EN ÁREAS GEOGRÁFICAS URBANAS

CON DATOS QUE EL INE SOLICITA SIN COMPENSACIONES, EL ARTÍCULO DE LEY 17.374 ESTABLECE EL SECRETO ESTADÍSTICO



INSTITUTO NACIONAL DE ESTADÍSTICAS
SUBDIRECCIÓN DE OPERACIONES
Departamento de Estadísticas de Hogares
Subdepartamento de Proyecto de Hogares
Avenida Bulnes 206, Oficina 31 Santiago
Fono: 02-7962400 02-7962447 02-7962448

IDENTIFICACIÓN GEOGRÁFICA

Hoja N° ____ de ____

USO ESTADÍSTICO DE SECCIONES

Estrato Sección

Comuna Distrito Zona Manzana Nombre de Comuna

N° de Línea	N° de Orden de la Edificación	Dirección					Código del Registro	N° de Orden de la Vivienda	N° de Orden del Hogar	Número Total de Personas en el Hogar <small>(Incluir menores de 15 años)</small>	Anotaciones complementarias.	
		Tipo de Dirección	Nombre de Dirección <small>(Otro: Indique nombre si corresponde)</small>	Número de Domicilio	Block o Torre	N° de Piso					N° o letra del Depto. o Pieza	Anote el número del código en la columna correspondiente y detalle en la columna descripción y/u observación.
		1.- Avanzada								1.- Vivienda sin numeración (Anotar descripción detallada para identificar y diferenciar la vivienda).	2.- Cualquier anotación importante para identificar la vivienda.	
		2.- Calle								3.- Fachado	3.- Observaciones sin vivienda	
		3.- Pasaje								4.- Describir tipo de vivienda colectiva	6.- Describir actividad de otro uso de la edificación.	
		4.- Camino								5.- Observaciones sin vivienda		
		5.- Paseo								6.- Describir actividad de otro uso de la edificación.		
		6.- Otro								Código	Descripción y/u Observación	
1												
2												
3												
4												
5												
6												
7												
8												
9												

Recorrido y enumeración urbana

El enumerador debe confirmar que se encuentra en la manzana seleccionada, verificándolo con el plano comunal y el croquis respectivo. Una vez realizado esto, debe situarse en la esquina norponiente de la manzana 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.

Durante el proceso de enumeración, algunas manzanas todavía pueden estar sujetas a cambios (reemplazos) por diversos motivos,

1. La manzana puede estar temporalmente destruida porque se van a levantar nuevas edificaciones, ya sean casas nuevas o edificios.
2. La mayoría de las viviendas cambiaron de uso o destino a otro que no es de vivienda.
3. Es imposible ingresar a las viviendas de la manzana porque están dentro de un edificio o condominio que tiene un conserje que les impide el ingreso.

4. La manzana no tiene el número de viviendas mínimo requerido para la selección de las viviendas.
5. La estructura en terreno de la manzana difiere de la enviada en el croquis, debido a que esta 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 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.

Selección de las viviendas

Una vez determinadas las manzanas definitivas de la muestra, se procede a la selección de las viviendas al interior de cada una de ellas.

El MM2015 tiene en promedio un mínimo de 4 viviendas a encuestar y para no concentrar la muestra en manzanas grandes (en cuanto a tamaño de viviendas), se establece en promedio una cota superior de 30 viviendas. 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. Tal como se mencionó anteriormente, al interior de cada manzana la selección de viviendas se realiza de manera sistemática con igual probabilidad.

La selección de viviendas fue implementada en SPSS, bajo el módulo de análisis *Muestras Complejas*, específicamente en el procedimiento *seleccionar una muestra*. Bajo este procedimiento se realizó la selección sistemática de viviendas donde se utiliza una semilla fija, a fin de poder replicar la selección.

Sean M_i el número de viviendas que posee la manzana, obtenido de la enumeración, y m_i el número de viviendas a seleccionar. Para la selección de m_i viviendas el software sigue los siguientes pasos:

Paso 1

En primera instancia se ordenan geográficamente todos los elementos según la división político administrativa. Para las unidades procedentes del MM2015, se ordenan de forma ascendente según las variables: región, comuna⁵, distrito, zona, manzana y número de orden de vivienda.

Paso 2

Al interior de cada manzana se calcula el período $k = M_i/m_i$. Notar que k puede ser un número real, no entero (puede tener decimales).

5 Comuna corresponde al código único territorial.

Paso 3

Luego calcula el arranque A o primera selección como un número aleatorio entre 1 y el período k . Para la selección de este número se define una semilla fija que en esta selección fue 999999999.

Paso 4

Posteriormente se va sumando sucesivamente el período k al arranque A para obtener distintos valores, los que van generando las sucesivas selecciones: $A, A+K, A+2K, A+3K, \dots, A+(m_i-1)K$.

La primera vivienda seleccionada es A y es un número 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 $A+(m_i-1)K$.

Probabilidad condicional de selección de las viviendas

La probabilidad de inclusión condicional de la j -ésima unidad de segunda etapa (vivienda) dentro de la i -ésima unidad primaria (manzana), perteneciente al estrato h , $p_{hi}(j|i)$ es igual a:

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

siendo,

m_{hi} : el número de viviendas seleccionadas en la manzana i del estrato h .

M'_{hi} : el número actualizado⁶ de viviendas en la manzana i del estrato h . El total de unidades muestrales seleccionadas en cada estrato se presenta en el siguiente cuadro:

⁶ Puede ocurrir que al visitar alguna de las manzanas seleccionadas, el encuestador se encuentre con que el número de viviendas es mayor (o menor) al reportado por el marco de muestreo, en este caso, se anota el nuevo número de viviendas y se registra bajo la denominación M'_{hi} . En caso de no haber cambios, $M'_{hi} = M_{hi}$.

Cuadro N°8: Total de unidades muestrales seleccionadas ENUT 2015

Región	RPC	Comuna	Nro. manzanas seleccionadas	Nro. viviendas a encuestar
Total			2.038	15.312
1	1101	IQUIQUE	24	217
1	1107	ALTO HOSPICIO	16	107
2	2101	ANTOFAGASTA	50	358
2	2201	CALAMA	27	161
2	2301	TOCOPILLA	11	70
3	3101	COPIAPÓ	31	199
3	3201	CHAÑARAL	15	69
3	3301	VALLENAR	16	85
4	4101	LA SERENA	41	295
4	4102	COQUIMBO	47	303
4	4201	ILLAPEL	15	75
4	4203	LOS VILOS	12	66
4	4301	OVALLE	21	125
5	5101	VALPARAÍSO	31	248
5	5103	CONCÓN	8	71
5	5107	QUINTERO	10	59
5	5109	VIÑA DEL MAR	33	321
5	5301	LOS ANDES	11	66
5	5501	QUILLOTA	11	69
5	5502	CALERA	9	64
5	5601	SAN ANTONIO	13	81
5	5602	ALGARROBO	8	59
5	5603	CARTAGENA	11	60
5	5604	EL QUISCO	11	62
5	5701	SAN FELIPE	10	68
5	5801	QUILPUÉ	18	127
5	5802	LIMACHE	8	59
5	5804	VILLA ALEMANA	14	92
6	6101	RANCAGUA	63	516
6	6105	DOÑIHUE	10	61
6	6106	GRANEROS	10	87
6	6108	MACHALÍ	11	91
6	6301	SAN FERNANDO	15	131

CONTINÚA ▶

Cuadro N°8: Total de unidades muestrales seleccionadas ENUT 2015

Región	RPC	Comuna	Nro. manzanas seleccionadas	Nro. viviendas a encuestar
6	6303	CHIMBARONGO	10	65
6	6310	SANTA CRUZ	13	85
7	7101	TALCA	59	415
7	7102	CONSTITUCIÓN	9	83
7	7109	SAN CLEMENTE	8	66
7	7201	CAUQUENES	10	88
7	7301	CURICÓ	30	222
7	7401	LINARES	20	155
7	7404	PARRAL	11	74
7	7407	VILLA ALEGRE	10	65
8	8101	CONCEPCIÓN	20	221
8	8102	CORONEL	15	96
8	8103	CHIGUAYANTE	8	77
8	8106	LOTA	13	74
8	8107	PENCO	9	69
8	8108	SAN P. DE LA PAZ	12	111
8	8110	TALCAHUANO	23	148
8	8111	TOMÉ	12	72
8	8112	HUALPÉN	14	83
8	8205	CURANILAHUE	11	65
8	8301	LOS ÁNGELES	16	125
8	8305	MULCHÉN	11	65
8	8306	NACIMIENTO	10	65
8	8401	CHILLÁN	19	161
8	8406	CHILLÁN VIEJO	11	65
9	9101	TEMUCO	73	498
9	9108	LAUTARO	14	70
9	9111	NUEVA IMPERIAL	10	62
9	9112	PADRE LAS CASAS	13	92
9	9115	PUCÓN	10	68
9	9201	ANGOL	20	108
9	9202	COLLIPULLI	12	61
9	9203	CURACAUTÍN	11	57
9	9210	TRAIQUÉN	13	60

CONTINÚA ▶

Cuadro N°8: Total de unidades muestrales seleccionadas ENUT 2015

Región	RPC	Comuna	Nro. manzanas seleccionadas	Nro. viviendas a encuestar
9	9211	VICTORIA	16	81
10	10101	PUERTO MONTT	38	286
10	10105	FRUTILLAR	9	59
10	10107	LLANQUIHUE	10	59
10	10201	CASTRO	12	87
10	10202	ANCUD	12	85
10	10208	QUELLÓN	8	65
10	10301	OSORNO	38	252
11	11101	COYHAIQUE	40	241
11	11201	AYSÉN	23	120
12	12101	PUNTA ARENAS	36	204
13	13101	SANTIAGO	20	329
13	13102	CERRILLOS	10	66
13	13103	CERRO NAVIA	14	95
13	13104	CONCHALÍ	15	94
13	13105	EL BOSQUE	14	112
13	13106	ESTACIÓN CENTRAL	11	94
13	13107	HUECHURABA	8	66
13	13108	INDEPENDENCIA	6	66
13	13109	LA CISTERNA	9	75
13	13110	LA FLORIDA	30	260
13	13111	LA GRANJA	12	91
13	13112	LA PINTANA	15	127
13	13113	LA REINA	8	92
13	13114	LAS CONDES	21	296
13	13115	LO BARNECHEA	5	77
13	13116	LO ESPEJO	9	75
13	13117	LO PRADO	9	79
13	13118	MACUL	10	91
13	13119	MAIPÚ	43	358
13	13120	ÑUÑO A	13	198
13	13121	P. AGUIRRE CERDA	11	85
13	13122	PEÑALOLÉN	16	144
13	13123	PROVIDENCIA	9	186

CONTINÚA ▶

Cuadro N°8: Total de unidades muestrales seleccionadas ENUT 2015

Región	RPC	Comuna	Nro. manzanas seleccionadas	Nro. viviendas a encuestar
13	13124	PUDAHUEL	18	142
13	13125	QUILICURA	14	133
13	13126	QUINTA NORMAL	9	86
13	13127	RECOLETA	11	100
13	13128	RENCA	11	96
13	13129	SAN JOAQUÍN	10	75
13	13130	SAN MIGUEL	7	98
13	13131	SAN RAMON	10	66
13	13132	VITACURA	9	92
13	13201	PUENTE ALTO	44	385
13	13401	SAN BERNARDO	21	176
13	13605	PEÑAFLORES	7	65
14	14101	VALDIVIA	33	247
14	14107	PAILLACO	11	60
14	14201	LA UNIÓN	11	73
14	14204	RÍO BUENO	14	68
15	15101	ARICA	37	242

4.8.3 Consideraciones: distribución de la muestra

Dentro de los requerimientos solicitados, estaba considerada la distribución de la muestra según nivel socioeconómico, período de referencia y quincena de levantamiento. Si bien la muestra no fue distribuida de acuerdo al nivel socioeconómico de la manzana, es posible identificar cada unidad seleccionada en algún nivel, esto es: alto, medio, bajo, de acuerdo a la estratificación realizada en base al Censo de Población de Vivienda 2002⁷. Los períodos de referencia (par de días por los cuales será consultado el entrevistado) fueron construidos según la siguiente distribución:

7 Para mayores antecedentes consultar "Metodología de clasificación socioeconómica de los hogares chilenos", http://www.ine.cl/canales/chile_estadistico/censos_poblacion_vivienda/censo2002/clasificacion/metodoclasificacionhogares.pdf

Cuadro N°9: Correspondencia entre período de referencia y par de días asignado

Período de referencia	Par de días asignados	
1	Lunes	Sábado
2	Martes	Sábado
3	Miércoles	Sábado
4	Jueves	Sábado
5	Viernes	Sábado
6	Lunes	Domingo
7	Martes	Domingo
8	Miércoles	Domingo
9	Jueves	Domingo
10	Viernes	Domingo

Así, a cada unidad de primera etapa (manzana), se le asignó un nivel socioeconómico, y según esto, cada una de las comunas fue ordenada por este nivel. Luego en cada manzana de las comunas se distribuyeron las viviendas a encuestar de modo que a nivel comunal se obtuviera un número similar de viviendas por estrato alto, medio y bajo, en los distintos períodos de referencia, y en las distintas quincenas de levantamiento (en total 6 quincenas). Finalmente, se cuadró la muestra a nivel regional.

4.9 TAMAÑO MUESTRAL EFECTIVO

Un elemento central en la evaluación de la calidad de las estimaciones es el nivel de logro alcanzado en relación a la muestra objetivo.

Si se observa la tabla, a nivel nacional se logró entrevistar el 90% de las viviendas esperadas, con un desempeño dispar entre regiones.

Cuadro N°10: Tasas de logro

Región	Muestra objetivo	Viviendas en muestra (sobre-muestra)	Viviendas elegibles	Viviendas no elegibles	Viviendas de elegibilidad desconocida	Viviendas logradas	Tasa de logro bruta (1)	Tasa de logro ajustada (2)
Total país	11623	15312	14532	687	93	10502	69%	90%
Arica y Parinacota	186	242	234	7	1	164	68%	88%
Tarapacá	249	324	290	29	5	257	79%	103%
Antofagasta	453	589	543	43	3	388	66%	86%
Atacama	271	353	344	9	0	252	71%	93%
Coquimbo	665	864	806	56	2	580	67%	87%
Valparaíso	1150	1506	1386	107	13	1051	70%	91%
Metropolitana	3450	4670	4518	135	17	2915	62%	84%
O'Higgins	797	1036	986	46	4	848	82%	106%
Maule	900	1168	1089	48	31	678	58%	75%
Biobío	1150	1497	1433	59	5	1097	73%	95%
Araucanía	890	1157	1103	46	8	776	67%	87%
Los Ríos	344	448	426	22	0	334	75%	97%
Los Lagos	686	893	840	49	4	695	78%	101%
Aysén	277	361	338	23	0	291	81%	105%
Magallanes	155	204	196	8	0	176	86%	114%

(1) Relación entre las viviendas logradas y el total de viviendas en muestra.

(2) Relación entre viviendas logradas y muestra objetivo.

4.10 FACTORES DE EXPANSIÓN

El factor de expansión se interpreta como la cantidad de unidades en la población que representa una unidad de la muestra. Este es distinto para cada manzana, y esta diferencia proviene del hecho que la muestra no es autoponderada.

El factor de expansión, según el diseño muestral, corresponde al aplicado a una muestra en dos etapas, y depende del número de viviendas que tiene la manzana seleccionada y del número de manzanas que tiene cada comuna por grupo de tamaño.

El factor base se calcula como el inverso de las probabilidades de selección de las unidades de primera y segunda etapa.

En el factor se incluye un ajuste a la población de personas por comuna, dada por las proyecciones de población, por sexo y tramo de edad, correspondientes al período de levantamiento de la encuesta.

Por otro lado, se incluyen además ajustes por parámetros referentes

a manzanas y viviendas, con el objetivo de aumentar la precisión y el alcance de las estimaciones.

Los ajustes correspondientes a manzanas, dan cuenta de:

- La omisión de comunas,
- La existencia de manzanas que quedan fuera de la muestra, debido a ciertas características inherentes que se mencionan más adelante.

Los ajustes correspondientes a viviendas, corresponden a:

- Ajuste por elegibilidad, que da cuenta de la condición de elegibilidad de las viviendas al momento de la entrevista,
- Ajuste por no respuesta, que da cuenta de la falla de no poder levantar absolutamente todas las viviendas de la muestra objetivo.

Finalmente, se calcula un factor que corresponde al ajuste por personas que contestan el módulo sobre uso del tiempo.

Por otro lado, se considera también la posibilidad de realizar un suavizamiento del factor de expansión, en la medida de que exista un aumento considerable en la variabilidad de la estimación de variables de interés.

4.10.1 Ponderador de selección de manzanas

El ponderador de selección de manzanas se calcula como el inverso de las probabilidades de selección de cada manzana. Según lo descrito en el punto 4.7.2.1 Selección de unidades de primera etapa, en el ítem “Probabilidad de selección de las unidades de primera etapa”, de este informe, el factor de selección de manzanas entonces, está dado por:

$$w_{hgi} = \frac{1}{P_{hg}(i)}$$

$$w_{hgi} = \frac{N_{hg}}{n_{hg}}$$

Siendo,

w_{hgi} : El ponderador de selección del conglomerado i , perteneciente al grupo de tamaño g de la comuna h .

$P_{hg}(i)$: La probabilidad de selección de la manzana i , perteneciente al grupo de tamaño g de la comuna h .

El ponderador de selección de manzanas puede ser interpretado como el número de unidades de la población que representa cada manzana. Por ejemplo, cada manzana presente en la muestra representa a W_{hgi} manzanas de la población. De igual forma, al realizar la suma de los ponderadores de selección de todas las manzanas en la muestra se puede obtener 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.

Cabe señalar que, debido a la forma de seleccionar, todas las manzanas de un mismo grupo de tamaño, al interior de una comuna, tienen el mismo ponderador de selección.

Tal como se mencionó anteriormente, las manzanas con menos de 8 viviendas, no son consideradas en la selección, considerándoseles como grupo de tamaño "0".

La distribución de las manzanas por grupo de tamaño en el marco de muestreo, incluyendo las manzanas del grupo de tamaño "0", se muestra en el cuadro N° 11:

Cuadro N°11: Total de unidades según grupo de tamaño en el MM2008

Grupo	N° de viviendas	N° viviendas	N° Manzanas
0	1 - 7	53.578	13.894
1	8 - 23	994.825	63.319
2	24 - 44	1.216.764	39.267
3	45 - 81	591.573	10.334
4	82 - 154	439.327	3.990
5	155 y más	704.695	2.556
Total	Total	4.000.762	133.360

Luego, si se quiere estimar desde la base de datos el total de manzanas a nivel nacional, este valor va a diferir de las 133.360 manzanas que contiene el marco y que está reflejado en el cuadro N° 11. Esto se debe a la exclusión de unidades muestrales, que puede ser corregido a partir de un ajuste de sus ponderadores, tal como se expone en el siguiente punto.

4.10.2 Ajuste por omisión de manzanas

Con el objetivo de reducir costos operativos y de tiempo, previo a la selección, se distribuyeron las unidades muestrales en el área urbana de cada región entre todas las comunas que acumularan aproximadamente el 80% de la población. Así, por ejemplo, en el área urbana de la Región de Arica y Parinacota solo se seleccionaron unidades de la comuna de Arica, ya que concentra más del 90% de la población total.

Definidas las comunas que participan en la muestra, se excluyen 13.894 manzanas que pertenecen al grupo de tamaño “0”, es decir, aquellas que tienen menos de 8 viviendas. Estas concentran aproximadamente 53.500 viviendas, lo que equivale a 1% del MM2015.

Junto a lo anterior, dentro de algunas comunas se excluyeron algunos estratos de tamaño, debido a que al momento de distribuir la muestra de forma proporcional en cada uno de ellos, el número de unidades asignado, era demasiado pequeño (insuficiente para asignar una manzana en el estrato).

Además de la omisión de ciertos grupos de tamaño por comuna, por razones presupuestarias, se excluyó el área urbana de algunas comunas, siendo las más pequeñas de cada región.

Como medida de mitigación a esta falta de cobertura, el factor de expansión de los conglomerados fue ajustado en dos etapas, la cuales se detallan a continuación:

Exclusión de manzanas de menos de 8 viviendas

El ajuste aplicado para corregir esta falta de cobertura, denominado razón de ajuste por omisión de manzanas, consiste en ponderar el inverso de la probabilidad de selección de las manzanas con la razón obtenida a partir del cociente entre:

- El total de viviendas⁸ contenida en las manzanas del MM2015 en la comuna h ;
- La suma del total de viviendas observadas en las manzanas, ponderada por el factor de selección de conglomerado, al interior de cada comuna h .

Este procedimiento se resume a partir de la siguiente fórmula:

$$w'_{hgi} = w_{hgi} \cdot \ddot{R}_h$$

Donde

w_{hgi} : Ponderador de selección del conglomerado i , perteneciente al grupo de tamaño g de la comuna h .

w'_{hgi} : Ponderador de selección del conglomerado i , perteneciente al grupo de tamaño g de la comuna h ajustado por omisión de manzanas pertenecientes a algún grupo de tamaño.

\ddot{R}_h : Razón de ajuste por omisión de manzanas en la comuna h , con la cual se pondera el factor de selección de conglomerado. Se calcula a partir de la siguiente expresión:

⁸ Este total de viviendas considera las viviendas de todos los grupos de tamaño, inclusive el grupo cero, de manzanas de 7 viviendas o menos.

$$\hat{R}_h = \frac{MM_h}{\sum_{g=1}^5 \sum_{i \in O_{hg}} MM_{hgi} \cdot w_{hgi}}$$

Siendo

h : Es el índice de comuna.

g : Es el índice de grupo de tamaño.

i : Es el índice de manzana.

M_h : Es el total de viviendas en las manzanas de la comuna h registradas en el MM2015.

M_{hgi} : Es el total de viviendas en la manzana i , del grupo de tamaño g en la comuna h registradas en el MM2015.

O_{hg} : Es el listado de manzanas seleccionadas para ENUT en el grupo de tamaño g en la comuna h .

Nota: se analizó la posibilidad de realizar el ajuste a nivel de manzanas lo que llevaría a estimar exactamente el total de estas en el marco con el ponderador corregido, sin embargo, esto conlleva a una sobrestimación del número de viviendas, ya que el grupo de tamaño que se excluye, en relación a las manzanas, representa un mayor porcentaje de lo que representan las viviendas que lo componen.

4.10.3 Omisión de comunas

El ajuste aplicado para corregir la exclusión del área urbana de ciertas comunas, es un procedimiento similar al descrito en el apartado anterior, y consiste en ponderar el inverso de la probabilidad de selección de las manzanas, con la razón obtenida a partir del cociente entre:

- El total de viviendas⁹ observadas en MM2015 en la región r .
- La suma del total de viviendas observadas en los conglomerados, ponderada por el factor de selección de conglomerado, al interior de cada región r . Este procedimiento se resume en la siguiente fórmula:

$$w'_{hgi} = w_{hgi} \cdot \ddot{R}_r$$

Siendo,

w'_{hgi} : Ponderador de selección del conglomerado i , perteneciente al grupo de tamaño g de la comuna h ajustado por omisión de áreas urbanas de ciertas comunas.

⁹ Nuevamente se consideran todas las viviendas de todos los grupos de tamaño, inclusive el grupo cero o de manzanas de 7 viviendas o menos. El total considera a todas las comunas de la región, estén o no representadas por la muestra.

w'_{hgi} : Ponderador de selección del conglomerado i , perteneciente al grupo de tamaño g de la comuna h ajustado por omisión de manzanas pertenecientes a algún grupo de tamaño.

\hat{R}_r : Razón de ajuste por omisión de comunas en la región r , con la cual se pondera el factor de selección de conglomerado. Se calcula a partir de la siguiente expresión:

$$\hat{R}_r = \frac{M_r}{\sum_{h \in \varphi_r} \sum_{g=1}^5 \sum_{i \in \Theta_{hg}} M_{hgi} \cdot w'_{hgi}}$$

Donde

M_r : Es el total de viviendas en las manzanas de la región r registradas en MM2015.

M_{hgi} : Es el total de viviendas en la manzana i , del grupo de tamaño g en la comuna h registradas en el MM2015.

Θ_{hg} : Es el listado de manzanas seleccionadas para ENUT en el grupo de tamaño g en la comuna h .

φ_r : Es el listado de comunas seleccionadas para ENUT en la región r .

4.10.4 Ponderador de selección de viviendas

Según la probabilidad de selección de las viviendas, descrita en el punto 4.7.2.2 Selección de unidades de segunda etapa, en el ítem “Probabilidad condicional de selección de las viviendas”, al ser las viviendas seleccionadas en un proceso multi-etápico, la probabilidad de selección de las viviendas debe dar cuenta de la probabilidad de selección de la unidad a la cual pertenece. Luego, la probabilidad de selección de la vivienda j , perteneciente a la manzana i del grupo de tamaño g en la comuna h , puede ser calculadas como el producto de las dos probabilidades definidas anteriormente, tal como se indica a continuación:

$$P_{hgi}(j) = P_{hg}(i) \cdot P_{hg}(j|i), \text{ con } P_{hg}(j|i) = \frac{m_{hgi}}{M'_{hgi}}$$

Donde

$P_{hgi}(j)$: Es la probabilidad de seleccionar en la muestra la vivienda j , perteneciente a la manzana i del grupo de tamaño g en la comuna h .

$P_{hg}(i)$: Es la probabilidad de seleccionar en la muestra a la manzana i del grupo de tamaño g en la comuna h .

$P_{hg}(j|i)$: Es la probabilidad condicional de seleccionar la vivienda j perteneciente a la manzana i del grupo de tamaño g en la comuna h , dado que en la etapa de selección anterior la manzana i fue seleccionada.

m_{hgi} : Es el número de viviendas logradas en la sección i del grupo de tamaño g en la comuna h .

M'_{hgi} : Es el número de viviendas total actualizadas en terreno mediante la enumeración, en la sección i del grupo de tamaño f en la comuna h .

El ponderador de selección de vivienda se calcula como el inverso de la probabilidad de selección de la vivienda j , $P_{hgi}(j)$, en el grupo de tamaño g de la comuna h , tal como se enseña en la siguiente fórmula:

$$w_{hgi} = \frac{1}{P_{hgi}(j)}$$

Sin embargo, debe considerarse el ponderador de selección de conglomerados ajustado, en lugar del inverso de la probabilidad de selección, ya que este ponderador da cuenta de las correcciones a causa de la omisión intencionada de ciertas áreas geográficas y unidades muestrales.

Luego el factor de selección de viviendas está dado por:

$$w_{hgi} = w''_{hgi} \cdot \frac{M'_{hgi}}{m_{hgi}}$$

Donde

w''_{hgi} : corresponde al ajuste por omisión de comunas descrito anteriormente.

Este ponderador puede ser interpretado como el número de unidades en la población que representa cada una de las viviendas seleccionadas para ENUT 2015. La suma de este ponderador, a nivel de vivienda, proporciona una estimación del total de viviendas en el área urbana del país.

4.10.5 Ajuste por elegibilidad de viviendas

La población objetivo de la ENUT 2015, corresponde a todas aquellas personas de 12 y más años, que poseen sus capacidades físicas y mentales aptas para responder, que residen en viviendas particulares del área urbana del país y además deben estar presentes en la vivienda en la quincena de levantamiento. Por ello, previo al levantamiento de la información, se seleccionan las viviendas identificadas como viviendas particulares al momento de la enumeración. Sin embargo, pese a todos los esfuerzos realizados previamente, al momento de realizar el trabajo de campo es probable que ciertas unidades hayan cambiado de estado, por ejemplo, una vivienda clasificada como particular ocupada en el proceso de enumeración, al momento de realizar el trabajo de campo podría encontrarse en condiciones de arrendamiento o venta. También puede ocurrir que fallas en el proceso de enumeración

deriven en la selección de unidades colectivas o de otro uso (empresas, viviendas de uso temporal, etc.).

Por otro lado, existe la posibilidad que, en el trabajo de campo, por diversas razones, no se logre ubicar la vivienda, razón por la cual se desconoce la condición de ésta, en términos de si podría ser elegible o no.

Lo anterior significa que una porción de las unidades muestrales fueron seleccionadas cuando sus características no corresponden a unidades elegibles, o que por diversas razones se desconoce la real condición de la vivienda, razón por la cual las probabilidades de selección de las viviendas deben ser ajustadas.

Una vez finalizado el trabajo de campo, todas las viviendas seleccionadas son clasificadas en tres grandes grupos:

- Elegibles: son todas aquellas direcciones en las que el encuestador logró determinar que correspondían a viviendas particulares ocupadas, con moradores presentes o ausentes, que hayan respondido o no la encuesta.
- No elegibles: edificaciones identificadas como viviendas colectivas, deshabitadas, de uso temporal, destruidas, negocios, etc.
- Elegibilidad desconocida: corresponden a direcciones que no pueden verificar la condición y estado de la vivienda, por ejemplo, direcciones no enviadas a terreno, dirección a la cual no se pudo llegar por razones climáticas o de peligro para el encuestador, etc.

En el siguiente apartado se presenta la metodología de ajuste.

4.10.5.1 Ajuste por elegibilidad desconocida

Todas las viviendas seleccionadas para ENUT 2015, fueron clasificadas como viviendas elegibles, ya que correspondían a viviendas particulares ocupadas, sin embargo, algunas de ellas cambiaron de estatus en el tiempo transcurrido entre la enumeración y el período de levantamiento de información, pasando a ser unidades no elegibles. Junto a lo anterior, existen unidades que no fue posible identificarlas como unidad elegible o no elegible, estas unidades corresponden a elegibilidad desconocida.

El ponderador asociado a estas últimas unidades es redistribuido entre aquellas unidades de elegibilidad conocida (elegibles y no elegibles) al interior de cada comuna. La distribución de los pesos de las viviendas de elegibilidad desconocida se realiza a partir de la razón entre dos estimaciones obtenidas con el ponderador de selección de las viviendas, estas son:

- En el numerador se encuentra la estimación del total de viviendas de la población;
- El denominador corresponde a la estimación del total de viviendas de elegibilidad conocida (elegibles y no elegibles). La razón de ajuste se presenta en la siguiente expresión:

$$\hat{R}_{h,known} = \frac{\sum_{i \in \Theta_h} \sum_{j \in \Phi_{hi}} w_{hgi j}}{\sum_{i \in \Theta_h} \sum_{j \in \Phi_{hi,known}} w_{hgi j}}$$

Siendo

Θ_h : Es el listado de manzanas seleccionadas para ENUT en la comuna h .

Φ_{hi} : Conjunto de viviendas seleccionadas en la manzana i de la comuna h .

$\Phi_{hi,known}$: Conjunto de viviendas seleccionadas y de elegibilidad conocida en la manzana i de la comuna h .

Esta razón es igual para todas las viviendas pertenecientes a la misma comuna.

Luego, el ponderador de selección de vivienda ajustado por elegibilidad desconocida fue obtenido a partir del producto entre el ponderador de selección de vivienda y la razón explicitada anteriormente, tal como se señala a continuación:

$$w'_{hgi j} = w_{hgi j} \cdot \hat{R}_{h,known}$$

Cabe señalar que, en esta etapa, solo las unidades elegibles y no elegibles tienen un factor asociado. Finalmente, una vez realizado el ajuste por elegibilidad desconocida, las unidades No elegibles y de elegibilidad desconocida son descartadas de la base de datos.

4.10.5.2 Ajuste por no elegibilidad

De las viviendas seleccionadas en la muestra, luego del trabajo de campo se identificaron viviendas de elegibilidad conocida y de elegibilidad desconocida. De las viviendas que fueron identificadas como de elegibilidad conocida, se pueden observar viviendas elegibles y viviendas no elegibles.

Dado que el objetivo de la encuesta se centra en las personas de 12 o más años residentes en viviendas particulares, analíticamente no es necesario contar con la información de las viviendas no elegibles, por ello a partir de esta etapa y para fines de cálculo del factor de expansión se excluyen las unidades clasificadas en este tipo de elegibilidad.

Luego, el total de viviendas particulares de la población puede ser estimado a partir de la suma del factor de expansión ajustado por elegibilidad.

Por otro lado, del total de viviendas elegibles, en algunos casos, no se logró concretar una entrevista u obtener la información mínima requerida, por lo que es necesario realizar un ajuste para compensar la pérdida de unidades que, siendo elegibles, no se logró obtener la información esperada. En el siguiente punto, se aborda este tema.

4.10.6 Ajuste por no respuesta

A partir de esta etapa solo se consideran las viviendas elegibles, ya que en ellas se encuentra la población objetivo. Sin embargo, es posible que los residentes de algunas de estas viviendas hayan decidido no participar en la encuesta.

Utilizar un factor de expansión sin considerar la existencia de no respuesta, implica tener estimaciones que solo representan a las unidades que responden y, por tanto, la población que no responde, no sería cubierta por la encuesta. Si no existe relación entre el patrón de la no respuesta y el fenómeno que se desea medir, entonces, posiblemente la no respuesta afectaría la precisión de las estimaciones. De lo contrario se estarían obteniendo estimaciones sesgadas a partir de la muestra recogida.

Para minimizar los problemas generados por la ausencia de respuesta de la unidad de muestreo, se procede a realizar un ajuste para compensar dichas pérdidas a partir de un método denominado propensity score.

En términos generales, el método de ajuste consiste en estimar la probabilidad de que las unidades muestrales respondan la encuesta. Sin embargo, al no existir información relacionada a las viviendas, se ajusta un modelo de regresión lineal para predecir la tasa de respuesta de las manzanas, donde la variable dependiente es la tasa de respuesta y las variables independientes o predictoras son principalmente a nivel de comuna, provenientes de registros administrativos u otras encuestas; y a nivel, de manzana, se incorpora el puntaje PRINCALS, proveniente del marco 2002, que ha sido identificado en el marco 2015.

Las variables incorporadas en el modelo son:

1. A nivel de manzanas:

- Puntaje de clasificación socioeconómica, según Censo 2002.

2. A nivel de comunas:

- Porcentaje de personas en situación de pobreza, 2011. Metodología de Imputación de Medias por Conglomerados. Fuente: Ministerio de Desarrollo Social.
- Porcentaje de afiliados a seguro de cesantía del 40% nacional de menores ingresos, abril 2013. Fuente: Administradora de Fondos de Cesantía de Chile.

- Porcentaje de matrícula en establecimientos particulares subvencionados, 2012. Fuente: Ministerio de Educación.
- Promedio de Puntaje Simce 2012.
- Promedio de Puntaje Prueba de Selección Universitaria, 2012. Fuente: Consejo de Rectores.
- Tasa de mortalidad general 2005 - 2010 (ajustada a 1.000 habitantes). Fuente: Epidemiología, año 2003, Ministerio de Salud.
- Esperanza de vida al nacer de mujeres, 2005 - 2010. Epidemiología, Ministerio de Salud.
- Esperanza de vida al nacer de hombres, 2005 - 2010. Epidemiología, Ministerio de Salud.
- Porcentaje de población de 0 a 6 años en control, que está con sobrepeso. Fuente: DEIS, Ministerio de Salud, año 2011.

Luego, se ordenan las manzanas según la tasa de respuesta predicha con el modelo, creando un número de grupos de tamaño, al interior del cual se realizará el ajuste por falta de respuesta, distribuyendo el peso de aquellas unidades que no responden entre aquellas que responden al interior del grupo creado, denominado celdas de ajuste.

En resumen, la determinación de las celdas de ajuste, se realiza en tres pasos:

- Selección del modelo de predicción de la tasa de respuesta de manzanas.
- Predicción de la tasa de respuesta para cada manzana.
- Ordenamiento de las tasas predichas, de mayor a menor, y creación de celdas de ajuste. Si el modelo de predicción es correcto, al interior de las 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 siguiente fórmula:

$$\hat{R}_{d,R} = \frac{\sum_{i \in \Theta_d} \sum_{j \in \Phi_{di,elec}} w' h_{gij}}{\sum_{i \in \Theta_d} \sum_{j \in \Phi_{di,elec,R}} w' h_{gij}}$$

y se realiza al interior de cada celda, 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, la estimación del total de viviendas elegibles que responde,

Los parámetros de la fórmula se describen a continuación:

d : Índice de la celda al interior de la cual se realiza el ajuste por falta de respuesta

Θ_d : Listado de manzanas seleccionadas para ENUT asignadas a la celda de ajuste d .

$\Phi_{d,i,elig}$: Conjunto de viviendas elegibles pertenecientes a la manzana i clasificada en la celda de ajuste d .

$\Phi_{d,i,elig,R}$: Conjunto de viviendas elegibles y que responden en la manzana i clasificada en la celda de ajuste d .

W'_{ngij} : Ponderador de selección de vivienda ajustado por elegibilidad, asociado a la vivienda j de la manzana i , en el grupo de tamaño g de la comuna h .

Luego, el ponderador de selección ajustado por falta de respuesta puede ser expresado como:

$$W''_{ngij} = W'_{ngij} \cdot \ddot{R}_{d,R}$$

En este proceso, las unidades que no respondieron no se les asigna un valor, por lo tanto, tienen un valor perdido o missing. En las siguientes etapas se trabaja solo con las unidades que responden, mientras que las otras son excluidas de la base de datos.

4.10.7 Suavizamiento del factor de expansión

Debido a las etapas de selección, los ponderadores iniciales que componen el Factor de Expansión, a medida que se van ajustando, aumentan significativamente su variabilidad. Como medida de mitigación se propone realizar un suavizamiento en los factores. Para ello se requiere como insumo la principal variable de interés a medir, que permitirá identificar el umbral de suavizamiento a través de la minimización de su error cuadrático medio.

En este contexto, la variable de interés que se midió es "*Necesidades fisiológicas y cuidados personales*" obtenida a nivel de comuna. Para aplicar la metodología de suavizamiento se utilizó la base datos a nivel de personas, donde se calculó la prevalencia de la variable antes mencionada.

Una vez obtenida la estimación, se procedió a realizar el suavizamiento a partir de la implementación de las siguientes etapas:

- i. Se inspeccionó la existencia de valores extremos en la distribución del ponderador al interior de cada una de las comunas,
- ii. Se determinaron puntos de corte a partir de los cuales realizar el suavizamiento,
- iii. Se suavizaron los valores extremos identificados,

- iv. Se estimó el error cuadrático medio (ECM) para los distintos puntos de corte,
- v. Se eligió la opción de corte que minimizaba el ECM,

Considerando lo anterior, se consideraron 7 puntos de cortes distintos definidos como sigue:

Sea

$$M_{ha} = a * \bar{F}_h, \quad a = 4, 5, 6, 7, 8, 9, 10, \quad \text{con } \bar{F}_h = \frac{\sum_{i \in \Theta_h} \sum_{j \in \Phi_i} w''_{hgijk}}{TV_h}$$

M_{ha} : Corresponde a la cantidad o número asociado al punto de corte, correspondiente a “a” veces el promedio del factor de expansión en la comuna h .

\bar{F}_h : Corresponde al promedio del factor de expansión en la comuna h , es decir, a la suma de todos los factores de viviendas en la comuna h , dividida por todas las viviendas que responden la entrevista en la comuna h .

a : Corresponde a todos los puntos de corte considerados para el suavizamiento.

Φ_i : Es el conjunto de viviendas seleccionadas en la manzana i .

Θ_h : Es el conjunto de manzanas seleccionadas en la comuna h .

TV_h : Es el total de viviendas entrevistadas en la comuna h .

Luego, para realizar el suavizamiento se procede a truncar aquellos ponderadores identificados como valores extremos de la siguiente forma,

$$w''_{hgij}{}^t = \begin{cases} w''_{hgijk} & \text{si } w''_{hgij} \leq M_{ha} \\ M_{ha} & \text{si } w''_{hgij} > M_{ha} \end{cases}$$

Si se suman todos los valores $w''_{hgij}{}^t$, se obtiene un total de unidades estimadas inferior que al sumar los ponderadores base, por lo tanto se debe distribuir la diferencia faltante en el resto de los ponderadores que no fueron truncados. Los pesos fueron distribuidos al interior de cada comuna de la siguiente forma:

$$w_{hgij}{}^{t*} = \begin{cases} w''_{hgij} \cdot \frac{(\sum_{g \in \Psi_h} \sum_{i \in \Theta_g} \sum_{j \in \Omega_i} w''_{hgij} - \sum_{g \in \Psi_h} \sum_{i \in \Theta_g} \sum_{w''_{hgij} > M_{ha}} w''_{hgij})}{\sum_{g \in \Psi_h} \sum_{i \in \Theta_g} \sum_{w''_{hgij} \leq M_{ha}} w''_{hgij}} & , w''_{hgij} \leq M_{ha} \\ M_{ha} & , w''_{hgij} > M_{ha} \end{cases}$$

Donde $w_{hgij}{}^{t*}$ es el factor suavizado.

La idea es que, aquellos ponderadores identificados como valores extremos, según el suavizamiento utilizado, son truncados al valor máximo establecido $M_{ha} = a * \bar{F}_h$, mientras que el peso “sobrante” de estos ponderadores truncados es distribuido sobre el resto de los ponderadores.

Luego, para determinar el punto de corte donde se realiza finalmente el suavizamiento, se calcula un estadígrafo que da cuenta del sesgo y de la variabilidad.

En este caso, el estadígrafo corresponde al Error Cuadrático Medio (ECM) asociado a la variable de interés. Como los objetivos de esta encuesta pretenden proveer información sobre el uso del tiempo de la población de 12 años y más; y al observar la alta prevalencia en la población de la variable utilizada para el cálculo del tamaño muestral, se utilizó esta como variable de interés, es decir, "*Necesidades fisiológicas y cuidados personales*".

Luego, se calculó dicho indicador por región y su desviación típica. De esa forma el sesgo y el ECM, respectivamente, se calculan según las siguientes fórmulas:

$$sesgo(\hat{P}_{ra}) = P_r - \hat{P}_{ra}$$

$$ECM(\hat{P}_{ra}) = Sesgo^2(\hat{P}_{ra}) + Var(\hat{P}_{ra})$$

Siendo,

r : Subíndice de la región. Toma valores entre 1 y 15 según numeración de las regiones del país.

P_r : La proporción de personas que realizan actividades según la variable de interés en la región r , obtenida con el factor de expansión sin truncar $w''_{hgi j}$

\hat{P}_{ra} : La proporción de personas que realizan actividades según la variable de interés, en la región r con el factor suavizado en el punto a .

Del ejercicio realizado, se estableció que el valor mínimo de la mediana del ECM para cada región se alcanza cuando el ponderador es truncado a 7 veces la media, por lo que finalmente ese fue el criterio a utilizar.

Posteriormente, utilizando como insumo el ponderador base suavizado, se realiza la calibración al *stock* poblacional, el cual se detalla en el siguiente punto.

4.10.8 Ponderador de personas ajustado a *stock* poblacional regional

Los factores de expansión de selección de unidades muestrales, proporcionan una estimación del total de estas. En base a ellos, podemos realizar una estimación del total de manzanas y también de viviendas.

Así, al aplicar al factor de expansión de viviendas, un ajuste por falta de respuesta, y asignarlo a cada persona dentro de la base de datos, permite realizar una estimación del total de personas. Sin embargo, este valor difiere de las estimaciones provenientes de otras fuentes,

como, por ejemplo, las proyecciones de población generadas a partir de modelos socio-demográficos. Por este motivo, se procede a ajustar los factores de expansión de selección de unidades, al *stock* según la desagregación de interés, que genera la Unidad de Demografía del INE.

Para la ENUT 2015, los factores de expansión se ajustan a las proyecciones de población generadas con fecha 30 de octubre de 2015¹⁰, para el área urbana de cada región, según sexo y tramo de edad, tal como se describe a continuación:

El cuadro a continuación muestra las proyecciones de población a nivel regional en el área urbana del país, según sexo y tramo etario:

Cuadro N°12: Proyecciones de población al 30 de octubre por sexo y tramo etario para cada región

Región	Hombres			Mujeres			Total
	0 - 11	12 o +	Total	0 - 11	12 o +	Total	
Total general	1.332.453	6.304.968	7.637.421	1.286.185	6.675.658	7.961.843	15.599.264
Arica y Parinacota	14.181	57.738	71.919	13.405	69.946	83.351	155.270
Tarapacá	31.248	132.722	163.970	30.584	131.831	162.415	326.385
Antofagasta	57.474	247.440	304.914	55.713	232.669	288.382	593.296
Atacama	24.938	107.620	132.558	24.108	106.571	130.679	263.237
Coquimbo	55.767	252.078	307.845	53.271	268.455	321.726	629.571
Valparaíso	135.771	695.702	831.473	130.832	736.113	866.945	1.698.418
Metropolitana	589.730	2.799.416	3.389.146	568.699	3.004.536	3.573.235	6.962.381
O'Higgins	54.297	274.963	329.260	52.396	278.876	331.272	660.532
Maule	59.562	286.649	346.211	57.219	305.706	362.925	709.136
Biobío	147.376	713.673	861.049	141.999	766.433	908.432	1.769.481
La Araucanía	59.785	270.218	330.003	58.540	300.021	358.561	688.564
Los Ríos	22.294	105.448	127.742	21.311	115.095	136.406	264.148
Los Lagos	58.107	258.966	317.073	56.637	260.085	316.722	633.795
Aysén	9.357	38.933	48.290	9.180	37.256	46.436	94.726
Magallanes y Antártica	12.566	63.402	75.968	12.291	62.065	74.356	150.324

4.10.9 Calibración a *stock* regional

El método de calibración a stocks poblacionales, utilizado en ENUT 2015, fue el método de *raking*, el cual permite ajustar en forma simultánea múltiples subpoblaciones de interés con un nivel de precisión adecuado. Esta es una de las ventajas de este método por sobre otros, como, por ejemplo, el ajuste de razón (o postestratificación) tradicionalmente utilizado para calibrar factores de expansión en encuestas desarrolladas en Chile.

¹⁰ Correspondiente al período central del levantamiento de información realizado.

En palabras simples, el algoritmo básico consiste en el desarrollo de un ciclo externo que chequea criterios de convergencia, y un ciclo interno que itera sobre ciertas variables de control.

El algoritmo puede dar origen a ponderadores calibrados de alta variabilidad y/o con algunos valores extremos (*outliers*). Estos valores en los ponderadores pueden generar problemas al momento de realizar análisis, especialmente en subpoblaciones. Por este motivo, usualmente se truncan los ponderadores extremos.

El método también dispone de un algoritmo de *raking* modificado, que permite implementar distintas opciones para truncar los ponderadores de manera simultánea a la implementación del procedimiento de calibración.

Luego, el factor calibrado¹¹ corresponde a:

$$w_{hgi}^{Cal1} = w_{hgi}^{t*} \cdot Calib_{hgi}^{1,rse}$$

Siendo,

w_{hgi}^{Cal1} : Factor calibrado, correspondiente al factor de expansión suavizado asociado a la vivienda j de la manzana i en el grupo de tamaño g de la comuna h , ponderado por la razón de ajuste de calibración $Calib_{hgi}^{1,rse}$

$Calib_{hgi}^{1,rse}$: Corresponde al ponderador de ajuste de calibración obtenido por el método de calibración raking asociado a cada persona de la vivienda j en la manzana i del grupo de tamaño g en la comuna h , realizado a nivel regional r , por sexo s (hombre, mujer) y tramo etario e (menores de 12 y de 12 o más años)

w_{hgi}^{t*} : Es el factor de expansión suavizado obtenido de la etapa anterior.

4.10.10 Ponderador de personas calibrado, ajustado por ausencia de respuesta en el cuestionario de Uso del Tiempo

Por diversos motivos, algunas de las personas elegibles, esto es, de 12 o más años con capacidades físicas y mentales, no respondieron el cuestionario de Uso del Tiempo. Por lo tanto, al realizar un análisis estadístico de las variables del cuestionario de Uso del Tiempo con el ponderador de personas, existirá una discrepancia entre el total de personas estimado a partir de las variables socio-demográficas y el universo de personas estimado que responde las preguntas del cues-

¹¹ El factor calibrado es el resultado final de la calibración aplicada a los factores suavizados en la etapa previa. Esta calibración es un método iterativo que minimiza las distancias entre los factores iniciales y los obtenidos en las distintas iteraciones hasta alcanzar cierta convergencia.

tionario de Uso de Tiempo. Para subsanar esta diferencia, entre el número de personas que responde la información socio demográfica y el CUT, bajo la misma metodología antes expuesta se realizó la calibración a nivel regional, por sexo y tramo de edad, considerando solo las personas que respondieron el CUT.

Cabe señalar además que, a diferencia del proceso descrito en el punto sobre calibración por stocks poblacionales, se calibró considerando que la muestra asignada a cada uno de los pares de días debía alcanzar los mismos stocks. A modo de ejemplo se presenta en el siguiente cuadro los stocks a los cuales debiera llegar cada uno de los 10 pares de días en los que se dividió la información:

Cuadro N°13: *Stocks* poblacionales para cada región y para cada par de días

Región	Tamaño poblacional para cada par de días
Total	1.559.928
Arica y Parinacota	15.527
Tarapacá	32.639
Antofagasta	59.330
Atacama	26.324
Coquimbo	62.957
Valparaíso	169.842
Metropolitana	696.238
O'Higgins	66.053
Maule	70.914
Biobío	176.948
La Araucanía	68.856
Los Lagos	63.380
Los Ríos	26.415
Aysén	9.473
Magallanes	15.032

Luego, el factor calibrado¹² corresponde a:

$$w_{hgi j}^{Cal2} = w_{hgi j}^{t*} \cdot Calib_{hgi j}^{2, rse, t}$$

¹² Para este factor se aplica el mismo algoritmo Raking para cuadrar a los stocks poblacionales de cada región, según sexo y tramo etario, pero además se agrega un nuevo stock a cuadrar, considerando que la muestra asignada a cada uno de los pares de días de la semana, debía alcanzar los mismos tamaños. Además, esta calibración se aplica solo a las personas que contestaron los módulos del cuestionario de uso del tiempo CUT.

Siendo,

w_{hgi}^{Cal2} : Factor calibrado, correspondiente al factor de expansión suavizado asociado a la vivienda j de la manzana i en el grupo de tamaño g de la comuna h , ponderado por la razón de ajuste de calibración $Calib_{hgi}^{2,rse,t}$

$Calib_{hgi}^{2,rse,t}$: Corresponde al ponderador de ajuste de calibración obtenido por el método de calibración *raking* asociado a cada persona de la vivienda j en la manzana i del grupo de tamaño g en la comuna h , realizado a nivel regional r , por sexo s (hombre, mujer) y tramo etario e (menores de 12 y de 12 o más años), además calibrado a los *stocks* de pares de días de la semana t .

w_{hgi}^{t*} : Es el factor de expansión suavizado obtenido de la etapa anterior.

4.10.11 Factor de hogar

En ENUT 2015 la última unidad de muestreo son las viviendas, mientras que las unidades de información y análisis son las personas y los hogares inherentes a estas.

En este contexto, se requieren dos factores de expansión, uno que permita hacer los análisis de las personas y otro que permita analizar los hogares.

En ENUT 2015 se entrevistan a todas las personas de 12 o más años que residen en la vivienda seleccionadas, y que cuenten con condiciones físicas y mentales aptas para responder la encuesta. Por lo tanto, la probabilidad de selección de cada hogar no es más que la probabilidad de selección de la vivienda w_{hgi}^{t*} , el cual se asigna a cada hogar y a través del hogar se asigna a cada persona, sin diferenciar por sexo o tramo etario.

Sin embargo, la utilización del ponderador de selección de vivienda conllevaría a una estimación del total de hogares que no considera los crecimientos poblacionales, sino solo los crecimientos de edificación de viviendas particulares. Por esta razón es que el ponderador del hogar, que es el asignado desde la vivienda a la cual pertenece, es calibrado nuevamente mediante el método *raking*, al stock poblacional de la región, sin diferenciar por características socio-demográficas, ya que estas son propias de las personas, no así de los hogares.

Luego, el factor del hogar, calibrado por el método *raking*, corresponde a:

$$w_{hgijk}^{Cal3} = w_{hgijk}^{t*} \cdot Calib_{hgijk}^{3,r}$$

Siendo,

W_{hgiijk}^{Cal3} : Factor calibrado, correspondiente al factor de expansión suavizado asociado al hogar k de la vivienda j en la manzana i del grupo de tamaño g de la comuna h , ponderado por la razón de ajuste de calibración $Calib_{hgiijk}^3$

$Calib_{hgiij}^{3,r}$: Corresponde al ponderador de ajuste de calibración obtenido por el método de calibración raking asociado a cada persona de la vivienda j en la manzana i del grupo de tamaño g en la comuna h , realizado a nivel regional r .

W_{hgiijk}^{t*} : Es el factor de expansión suavizado obtenido de la etapa anterior, asociado al hogar k de la vivienda f , obtenido de la etapa anterior.

4.10.12 Resumen factores

Uso de los factores de expansión

Los resultados que se obtienen a partir de la muestra son representativos a nivel nacional urbano y regional urbano, por sexo y tramos de edad. Para la ENUT se tienen diferentes unidades de observación, ya que, si bien la última unidad de selección es la vivienda, dentro de ella se levanta información relativa al hogar y a sus integrantes. Y en el caso de los integrantes del hogar la información del cuestionario de uso del tiempo solamente se levanta para las personas de 12 años y más con condiciones físicas y mentales para responder y que se encuentren en la vivienda durante la quincena de levantamiento. Por tanto, para obtener estimaciones insesgadas se deben utilizar los factores de expansión apropiados teniendo en cuenta la unidad de análisis y tipo de estimación que se desea obtener:

Cuadro N°14: Resumen factores de expansión

Unidad de análisis	Factor de expansión	Nombre
Hogar	Factor de expansión del hogar.	W_{hgiijk}^{Cal3}
Personas integrantes del hogar.	Factor de expansión de las personas, calibrado por sexo-región y tramo de edad-región	W_{hgiij}^{Cal1}
Personas integrantes del hogar que responden el cuestionario de uso del tiempo.	Factor de expansión de las personas, calibrado por sexo-región, tramo de edad-región, par de día-región.	W_{hgiij}^{Cal2}

4.11 ESTIMACIÓN DE VARIANZAS

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

4.11.1 Métodos convencionales para estimación de la varianza

Existen principalmente cuatro métodos convencionales para estimar las varianzas o errores muestrales para estimaciones basadas en datos de una encuesta¹³.

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

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

Métodos exactos

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

Método del conglomerado último o final

El método del conglomerado último para estimar la varianza (véase Hansen, Hurwitz y Madow, 1953, págs. 257-259) se emplea para estimar las estimaciones basadas en una muestra obtenida a partir de un diseño muestral complejo, generalmente polietápico o polifásico. Según este método, el conglomerado último consiste en considerar a la unidad primaria de muestreo UPM como la unidad de información última que conglera la información de todas las unidades de

¹³ Para más información ver Kish y Frankel (1974), Wolter (1985) y Lehtonen y Pahkinen (1995).

muestreo de etapas posteriores dentro de ella, como por ejemplo, si la unidad primaria de muestreo corresponde a un conglomerado que contiene unidades secundarias de muestreo como las viviendas, y a su vez estas unidades secundarias de muestreo contienen unidades terciarias como las personas, y se está interesado en alguna variable de interés como la ocupación, entonces cada UPM tendrá asociado el total de ocupados, como la sumatoria sobre todas las personas de todas las viviendas contenidas en la UPM. Las estimaciones de la varianza se calculan utilizando solo totales entre UPM, sin tener que calcular los componentes de la varianza en cada etapa de selección.

Método de linealización

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

Se aplica la linealización generalmente para la estimación de razones o cocientes de dos variables las que se desarrollan en una serie de Taylor infinita centrada en el valor esperado (estimado) del numerador y el valor esperado (estimado) del denominador. Seguidamente, el estimador no lineal se somete a aproximación algebraica conservando solo los primeros términos de la serie infinita de Taylor, lo que produce una expresión algebraica que ya no es una función no lineal, sino una función lineal de los datos muestrales; esto es, el estimador no lineal cociente se ha “linealizado”. Ahora es posible obtener directamente la varianza estimada de la función linealizada (incluidos los términos de covarianza pertinentes). Una vez linealizada, la varianza de la estimación no lineal puede aproximarse mediante los métodos exactos descritos anteriormente (véanse Cochran, 1997, y Lohr, 1999, para consultar información técnica detallada sobre el proceso de linealización ilustrada con ejemplos).

Método de replicación

Este método consiste en tomar submuestras o réplicas de la muestra total. Se estima con la muestra total la variable de interés Y , de acuerdo al diseño probabilístico, sea cual fuere. Enseguida se generan réplicas o se seleccionan submuestras de la muestra total, de manera que cada réplica refleje el plan de muestreo y los ajustes y procedimientos de ponderación de la muestra completa, de modo que con cada una de ellas se llegue a la misma estimación que con la muestra completa. Existen varias técnicas para hacerlo y en algunos software ya vienen incluidos los módulos de selección de réplicas, entre los cuales los más

utilizadas son los grupos aleatorios; la replicación repetida equilibrada (BRR); la replicación jackknife (JK1, JK2, y JKn); el bootstrap.

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

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

4.11.2 Varianza con muestras complejas

Una aproximación común (WR: With Replacement) para la descripción de muchos planes de muestreo complejo, normalmente es utilizada en las encuestas complejas de muestreo polietápico por conglomerados. Además, es habitual la estratificación de las UPM de la población antes del muestreo en la primera etapa. Por otra parte, la estratificación de las unidades de la segunda y posteriores etapas (en una UPM seleccionada) puede tener lugar antes del muestreo en estas etapas. Sin embargo, con los métodos aproximados de estimación de varianzas comúnmente utilizados para estos diseños complejos, no es necesario tomar en consideración todas las etapas de muestreo y estratificación. El muestreo complejo en las etapas posteriores queda automáticamente cubierto de manera adecuada con la aproximación “con reemplazo” para la primera etapa de muestreo que hemos visto más arriba. De hecho, pocos paquetes de software tienen la capacidad de incluir por separado todas las etapas de muestreo en la estimación de varianzas, en los casos en los que no se practica la aproximación con reemplazo en la primera etapa.

La estimación de varianzas por conglomerados finales (EVCF) puede efectuarse indistintamente por linealización mediante el desarrollo en serie de Taylor o con una técnica de replicación. Con este enfoque, las UPM se tratan como si se hubieran seleccionado con reemplazo entre los estratos de la primera etapa. En tal caso, cada elemento o individuo solo tiene que identificarse por el estrato de la primera etapa y por la UPM (dentro del estrato) de la que se haya seleccionado. A los efectos de la estimación de varianzas no se necesita información sobre las etapas de muestreo por debajo del nivel de las UPM. De este

modo la descripción del plan de muestreo se simplifica de forma que se asemeja a un muestreo por conglomerados estratificados en una etapa, esto es, una muestra estratificada de conglomerados finales completamente enumerados. Este enfoque por conglomerados finales produce una buena aproximación para la estimación de la varianza, siempre que sea razonable partir de la hipótesis de una primera etapa con reemplazo.

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

- La variable de ponderación de la muestra WT o factor de expansión (necesaria también para las estimaciones): en ENUT 2015 designada como w_{hgi}^{Cal1} , w_{hgi}^{Cal2} , w_{hgi}^{Cal3}
- La variable de estratificación (o variable de pseudo-estratificación): en ENUT 2015 designado VarStrat.
- La variable UPM (o pseudo-conglomerado): en ENUT 2015 designada como VarUnit.

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

4.11.3 Algoritmo de cálculo en muestras complejas

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

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

Si una vivienda se incluye en la muestra con una probabilidad p_{hij} , su ponderación básica, viene dada por $w_{hij} = 1/p_{hij}$, es decir, el inverso de la probabilidad de selección y se denomina factor de expansión teórico, de acuerdo al diseño probabilístico. Este ponderador es igual para todas las viviendas dentro del conglomerado (manzana o sección) y, por lo tanto, para todas las personas dentro de ellas. En la ENUT 2015, posteriormente, este factor se ajustó por omisión de conglomerados, elegibilidad desconocida y también por no-respuesta de las viviendas y además se calibró para alcanzar los stocks poblacionales de personas según las proyecciones de población.

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

$z_{hij} = w_{hij} \cdot y_{hij}$ es el valor de la variable observada en la vivienda j a través de todos los hogares que componen la vivienda, expandida a la población.

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

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

n_h es el número de UPM (manzanas) seleccionadas en el estrato h en la muestra ENUT 2015.

N_h es el número de UPM (manzanas) en el estrato h , contenidas en el marco muestral utilizado en ENUT 2015.

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

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

4.12 ALGORITMO DE CÁLCULO DE LOS ESTIMADORES

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

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

Es el total estimado en el estrato h de la variable de interés Y

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

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

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

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

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

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

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

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

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

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

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

Es el estimador de razón entre las variables Y y X, en el estrato h, es decir,

$$R_h = \frac{Y_h}{X_h} = \frac{\bar{Y}_h}{\bar{X}_h} \quad (8)$$

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

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

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

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

Son las estimaciones del total y la media de la variable Y en el estrato h , obtenidas mediante el estimador de razón

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

Son las estimaciones del total y la media de la variable Y a nivel agregado, obtenidas mediante el estimador de razón.

Para las fórmulas (7), (8), (9) y (10), se asume que el total X o la media \bar{X} son conocidas o dadas como fijas¹⁴. Pero, generalmente se desconoce el valor poblacional y entonces son estimados por \hat{X} y $\bar{\hat{X}}$ en la expresión (7), o son estimadas por \hat{X}_h y $\bar{\hat{X}}_h$ en la expresión (8).

4.12.1 Algoritmo de cálculo de las varianzas en muestras complejas

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

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

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

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

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

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

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

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

Finalmente, se muestran las estimaciones de las varianzas de los estimadores de totales, medias y razones.

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

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

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

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

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

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

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

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

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

Es la varianza del estimador de razón entre las variables Y y X , en el estrato h , con: $\hat{X}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \cdot x_{hij}$

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

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

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

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

Es la estimación de la varianza del total y la media de la variable Y , en el estrato h , mediante el estimador de razón.

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

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

4.12.2 Variables que identifican el diseño muestral complejo

Un diseño de muestreo complejo se origina cuando interviene más de una etapa de selección. Esta selección en más de una etapa suele ser generalmente además estratificada, lo que añade otro grado de complejidad. Bajo la estratificación, en un muestreo bietápico, las unidades muestrales se suelen llamar conglomerados o unidades de conglomeración (que en la ENUT 2015 corresponden a manzanas), que son grupos de unidades de selección de las últimas unidades de muestreo sobre las cuales se obtienen las estimaciones de las variables de interés.

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

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 ENUT estableció, en lo posible, un mínimo de 2 conglomerados por estrato y un mínimo de 50 viviendas por conglomerado (manzanas).

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 ENUT 2015 y producir estimaciones de varianza más estables. Las estimaciones oficiales de la varianza compleja de ENUT 2015 han sido desarrolladas utilizando pseudo-estrato (VarStrat), pseudo-conglomerado (VarUnit) y variable de ponderación de la muestra

$$w_{hgi}^{Cal1}, w_{hgi}^{Cal2}, w_{hgi}^{Cal3}$$

4.13 COEFICIENTE DE VARIACIÓN

El coeficiente de variación es una medida de variabilidad relativa que corresponde al cociente entre el error estándar $\hat{\sigma}_{\hat{\theta}}$ y la estimación $\hat{\theta}$, en el contexto de una estimación intervalar para la estimación de la variable de interés. En fórmula, podemos escribir:

$$Cv_{\hat{\theta}} = Cv(\hat{\theta}) = \hat{\sigma}_{\hat{\theta}}/\hat{\theta} = \sqrt{\hat{V}(\hat{\theta})}/\hat{\theta} \quad (22)$$

Siendo $\hat{\theta}$ la estimación de la variable de interés θ , $\hat{\sigma}_{\hat{\theta}}$ el error estándar de la estimación, que corresponde a la raíz cuadrada de la varianza, dada por $\hat{V}(\hat{\theta})$.

En el contexto de una estimación intervalar de muestreo repetido, con una confianza dada de $(1-\alpha)\%$, el intervalo de confianza correspondiente para la variable de interés θ , está dado por:

$$\theta \in [\hat{\theta} \mp t_{1-\alpha/2}^v * \hat{\sigma}_{\hat{\theta}}]_{1-\alpha} \Leftrightarrow [\hat{\theta} - t_{1-\alpha/2}^v * \hat{\sigma}_{\hat{\theta}} ; \hat{\theta} + t_{1-\alpha/2}^v * \hat{\sigma}_{\hat{\theta}}]_{1-\alpha} \quad (23)$$

Siendo $t_{1-\alpha/2}^v$ el percentil $(1-\alpha/2)$ de la distribución t-Student con v grados de libertad.

El error absoluto de la estimación intervalar corresponde exactamente a la parte que se suma y resta de la estimación, esto es, $E_{abs}(\hat{\theta}) = t_{1-\alpha/2}^v * \hat{\sigma}_{\hat{\theta}}$, que al dividirla por la misma estimación, se obtiene el error relativo, dado por:

$$E_{rel}(\hat{\theta}) = t_{1-\alpha/2}^v * \hat{\sigma}_{\hat{\theta}}/\hat{\theta} = t_{1-\alpha/2}^v * Cv_{\hat{\theta}} = E_{abs}(\hat{\theta})/\hat{\theta} \quad (24)$$

Para muestras grandes, el percentil de la distribución t-Student converge al percentil de la distribución Normal, es decir, $t_{1-\alpha/2}^v \approx Z_{1-\alpha/2}$, que para una confianza de 95%, corresponde al valor 1,96, esto es: $Z_{0,975} = 1,96$ cuando la confianza es $(1 - \alpha) = 95\%$

Si solo se considera el coeficiente de variación como medida del error relativo, en el contexto de estimación intervalar, es semejante a construir intervalos de confianza de aproximadamente $(1 - \alpha) = 68\%$ de confianza, ya que $Z_{1-\alpha/2} = Z_{0,84} \approx 1$ entonces, aproximadamente $E_{rel}(\hat{\theta}) = Cv_{\hat{\theta}}$

Se considera que el resultado de una estimación debe leerse con precaución¹⁵ si su coeficiente de variación es mayor al 20%, es decir, en el contexto de una estimación intervalar de 95% de confianza, si su error relativo es mayor a 40%.

5. Metodología de recolección y procesamiento de datos

5.1 SENSIBILIZACIÓN

El trabajo de sensibilización de las y los informantes tuvo como objetivo fomentar la participación de los hogares seleccionados en la encuesta, y se realizó a través la comunicación de dos elementos centrales: confianza en la institución y relevancia de la encuesta.

Este trabajo fue realizado por los equipos de levantamiento (encuestadores y supervisores) de cada región. Por lo que su capacitación incorporó aspectos como la Ley 17.374 que resguarda el secreto estadístico, el enfoque de género de la encuesta y fundamentalmente la necesidad de medir y visibilizar el tiempo destinado al trabajo doméstico y de cuidados no remunerado.

Además de la capacitación al personal de terreno, con la finalidad de facilitar el contacto y convencimiento de los informantes, se generó material de apoyo como cartas para las viviendas seleccionadas y dípticos informativos sobre las características de la encuesta, como sus objetivos y alcance.

5.2 MÉTODOS DE RECOLECCIÓN DE LOS DATOS

5.2.1 Periodicidad de la recolección de los datos: la ENUT 2015 es la primera encuesta sobre Uso del Tiempo de carácter nacional realizada en Chile. Se espera que dada la relevancia de la medición del trabajo realizado por fuera de la frontera de la producción del SCN, esta medición pueda repetirse de manera periódica.

¹⁵ Comúnmente si el error relativo de una estimación es superior a 30%, ya es motivo de alerta, en el sentido que la estimación no está siendo tan precisa, lo que es aproximadamente equivalente a tolerar un coeficiente de variación no superior a 15%.

5.2.2 Período de recolección: el período regular de recolección de la encuesta comenzó el 21 de septiembre y terminó el 13 de diciembre de 2015.

De manera especial se generó un levantamiento complementario durante el mes de febrero del 2016, con la finalidad de recuperar información faltante sobre las características ocupacionales de los informantes no contactados en las viviendas logradas del período regular.

5.2.3 Período de referencia de los datos: la encuesta contó con distintos períodos de referencia dependiendo de las preguntas realizadas. Para las preguntas de vínculo con el mercado laboral e ingresos del trabajo se siguieron los estándares de la OIT: semana y mes anterior, respectivamente; y para las preguntas de otros ingresos también se usó como período de referencia el mes anterior.

Para las preguntas del Módulo F Ayudas que recibe el hogar, se usó como período la semana anterior a la entrevista.

Para las preguntas de uso del tiempo se asignó muestralmente a cada vivienda dos períodos de referencia: un día de semana y uno de fin de semana específico.

Esta asignación permitió que los distintos días de la semana tuviesen la misma representación en la muestra.

En relación a la presentación de resultados de uso del tiempo, asignar muestralmente a cada vivienda dos períodos de referencia permite obtener estimaciones para días de semana y para días de fin de semana. Por otro lado, para poder presentar resultados que en una sola estimación representen la realidad de los distintos tipos de días de forma simple y explicativa, se construye el día tipo, que busca ser el reflejo de las actividades que se realizan los distintos períodos de referencia.

Se considera que el informante participa de la actividad en un día tipo si realiza la actividad en al menos uno de los dos períodos de referencia, esto es, si la ejecuta en el día de semana o en el día de fin de semana. En ese mismo sentido, se considera que no realiza la actividad, cuando esta no es desarrollada en ninguno de los dos períodos consultados.

Para las personas que participan en el día tipo, el tiempo destinado se calcula haciendo la suma ponderada del tiempo, con proporciones de 5/7 para el tiempo destinado en día de semana y 2/7 para el tiempo destinado en el fin de semana.

5.2.4 Tipo de entrevista: los dos cuestionarios que comprende la ENUT se aplicaron en base a entrevistas cara a cara.

El Cuestionario del Hogar, que contiene la caracterización demográfi-

ca de los integrantes del hogar, se aplicó a través de la entrevista a un informante idóneo (integrante del hogar de 18 años o más).

El Cuestionario de Uso del Tiempo, que contiene la caracterización ocupacional, ingresos del trabajo, jubilaciones y aportes previsionales y las preguntas sobre uso del tiempo, se aplicó de manera directa a cada integrante del hogar de 12 años o más.

Para los casos en que no se logró el contacto con el informante directo, el módulo de caracterización ocupacional e ingresos del CUT fue aplicado por las y los supervisores de terreno a través de la entrevista cara a cara a un informante idóneo o “sustituto”.

De manera excepcional, se realizaron entrevistas telefónicas para los casos que no se lograron durante el período de recolección oficial y fue necesario pasar a un levantamiento complementario.

5.2.5 Criterios de logro: como la encuesta considera distintas unidades de información, se definieron distintos criterios de logro, diferenciando entre el logro de las unidades muestrales, en este caso las viviendas; y las unidades de información, correspondientes a los integrantes del hogar de 12 años y más.

El logro de la vivienda se definió como la realización del Cuestionario del Hogar y de al menos un Cuestionario de Uso del tiempo.

El logro del Cuestionario de Uso del Tiempo se definió como la aplicación de al menos los módulos de caracterización ocupacional y de ingresos y de tiempo destinado a los distintos tipos de trabajo. Los casos que no cumplieran con este criterio fueron identificados como *break off*, o no logro por insuficiencia de información.

Dada la dificultad de definir a priori un criterio de logro más específico, se definió implementar este criterio amplio para el trabajo de terreno y realizar un segundo análisis de suficiencia de información a nivel de base de datos, el que se detalla en el apartado de Análisis de información para la mejora de la base de datos.

5.2.6 Medio de recolección de información: todas las encuestas se realizaron a través de formularios en papel, los que posteriormente fueron digitados de manera centralizada.

5.2.7 Conformación del equipo de trabajo: cada equipo de levantamiento se conformó de una persona responsable de la supervisión y coordinación del grupo y de tres a cinco encuestadores y encuestadoras, contratados específicamente para el proyecto. La cantidad de equipos por región dependió del tamaño de la muestra asignada.

5.2.8 Métodos para incentivar la respuesta: como se señaló en el apartado de sensibilización, el elemento central en la motivación de las personas seleccionadas fue el relevamiento de la importancia de la medición del tiempo destinado al trabajo no remunerado, junto a la explicación de los mecanismos de resguardo de la privacidad de los datos que tiene el INE.

5.2.9 Mecanismo de resolución de dudas: las dudas surgidas en el trabajo de campo se respondieron de manera centralizada y distinguiendo según el tipo de consulta. Así, para las preguntas de carácter operativo, las respuestas fueron de responsabilidad de la coordinación del levantamiento; mientras las preguntas sobre conceptos y criterios para la aplicación de los cuestionarios fueron de responsabilidad del equipo técnico de la ENUT.

5.3 DISEÑO DE FORMULARIOS

La ENUT contó con dos formularios diferenciados por la unidad de observación y el tipo de información que recaban.

5.3.1 Cuestionario del Hogar (CH): tiene como objetivo caracterizar al hogar entrevistado y a las distintas personas que forman parte de él. En este instrumento se delimita quiénes son integrantes del hogar, se pregunta por las características sociodemográficas de estos integrantes (sexo, edad, relaciones de parentesco, nivel educacional, entre otras), además de algunas características del hogar como la presencia de servicio doméstico u otras personas que ayuden a las labores domésticas.

La información del Cuestionario del Hogar la entrega una persona que sea informante idónea, que corresponde a cualquier integrante del hogar, de 18 años o más, que tenga los conocimientos necesarios para poder responder por el resto de los y las integrantes de su hogar.

Si en una vivienda seleccionada se identificó más de un hogar, a cada uno se asignó un Cuestionario del Hogar independiente.

El Cuestionario del Hogar se estructura en base a nueve módulos temáticos:

Módulo A. Hoja de Ruta: la Hoja de Ruta registra información propia del levantamiento en sus distintas etapas como una bitácora para la encuesta, además de ser un instrumento que ordena y simplifica la labor de los y las entrevistadoras. La HR está anexada tanto al Cuestionario del Hogar como al Cuestionario de Uso del Tiempo y registra todo el proceso de levantamiento de los mismos, como la elegibilidad

de la vivienda, los registros de visitas de apertura y los códigos de disposición final, según cuestionario.

Módulo B. Identificación del Hogar: tiene por objetivo identificar el número de hogares en una vivienda. Este módulo considera 4 preguntas explícitas que ayudarán a reconocer el número de hogares que existe por cada vivienda.

Módulo C. Características sociodemográficas: busca establecer las personas que componen el hogar y registrar la información socio demográfica, como sexo, edad, etnia, estado conyugal, entre otras.

Módulo D. Educación: el objetivo del módulo es identificar y caracterizar la situación educacional de todos y todas los integrantes del hogar.

Módulo E. Cuidados permanentes: este módulo tiene dos objetivos asociados entre sí. Por un lado identificar a las personas que requieran apoyo de otros para realizar sus actividades cotidianas a causa de dificultades permanentes de salud. Y por otro identificar si corresponde aplicar la Sección 1 en el Módulo N del Cuestionario de Uso del Tiempo.

Módulo F. Ayudas que recibe el hogar: su objetivo es identificar si el hogar recibe apoyo de una o más personas externas al hogar en el trabajo doméstico y de cuidados, distinguiendo si éste es remunerado o no remunerado.

Módulo G. Ingresos del hogar: el objetivo de medir los ingresos del hogar se relaciona directamente con la necesidad de construir información que permita identificar como el ingreso mensual influye o afecta las diversas actividades que realizan las personas del hogar.

Con la idea de capturar esta información, el cuestionario del hogar considera las fuentes de ingreso que, en promedio, son las más recibidas por los hogares: pensión de alimentos, dinero aportado por personas ajenas al hogar, dinero por arriendo de bienes muebles e inmuebles, asignación familiar y subsidio familiar. Las cuatro primeras fuentes de ingreso se asocian a transferencias de dinero privadas y la última -subsidio familiar- es una transferencia que corresponden a subsidios del estado.

Módulo H. Observaciones: tiene por objetivo que el encuestador o encuestadora pueda registrar información relevante del hogar que dé cuenta de alguna situación anormal.

Módulo I. Vinculación con el Cuestionario de Uso del Tiempo: en este módulo el encuestador o encuestadora sistematiza información clave del Cuestionario del Hogar, para una correcta aplicación de los Cuestionarios de Uso del Tiempo.

5.3.2 Cuestionario de Uso del Tiempo (CUT): busca identificar las actividades que los distintos integrantes del hogar realizaron en un período

do de referencia determinado. Este período de referencia corresponde a dos días asignados a la vivienda desde el diseño muestral: un día de semana y uno de fin de semana. Además, en el Cuestionario de Uso del Tiempo se pregunta al informante sobre su condición ocupacional y sobre los ingresos del trabajo, jubilaciones y aportes previsionales.

La sección de Uso del Tiempo del CUT corresponde a un listado de actividades, ordenado en base a siete módulos temáticos (módulos M al S) en base al marco de CAUTAL y el SCN: trabajo dentro y fuera de las fronteras del SCN, y actividades personales.

Como se señaló anteriormente, existen dos metodologías fundamentales para la captura del tiempo en encuestas: la medición a través de diarios y la medición a través de listados de actividades.

El uso del listado de actividades permite al personal de terreno estructurar la entrevista relevando información sobre el conjunto de actividades definidas desde el diseño de la encuesta. De este modo se evita la subdeclaración de actividades como el trabajo doméstico y de cuidados, que por ser socialmente poco valoradas o no conceptualizadas como trabajo en el caso de los cuidados, corren el riesgo de ser invisibilizadas en una entrevista de respuesta espontánea (como los diarios de actividades)¹⁶.

El objetivo del listado de actividades es relevar las actividades de interés y determinar la realización de las personas en estas y el tiempo que se destina. Es un instrumento que simplifica errores de terreno, de procesamiento en cuanto a codificación y clasificación de actividades, e incurre en un costo menor. Sin embargo, tiene la desventaja de no recoger cronológicamente la información, lo que impide reconstruir el día de la persona informante, en tanto solo se recoge información de las actividades consultadas y no de todas las actividades posibles.

La información de los Cuestionarios de Uso del Tiempo se obtiene a través de entrevistas directas a todos y todas las integrantes de cada hogar desde los 12 años en adelante. Las únicas excepciones que pueden darse a este criterio es que la persona no vaya a estar presente en la quincena definida para la entrevista o que no tenga capacidades físicas o mentales para responder (esto se detallará más adelante). En este caso se habla de un informante “no elegible”.

A continuación se describen los módulos del CUT:

Módulo J. Hoja de Ruta Cuestionario de Uso del Tiempo: su objetivo es registrar los procesos de levantamiento y en este caso específico, para cada integrante del hogar elegible.

Módulo K. Trabajo en la ocupación: este módulo tiene por objetivo caracterizar la relación de la población con el mercado laboral y registrar sus

¹⁶ Siguiendo a Naciones Unidas (UN, 2006), la lista de actividades al no contextualizar cronológicamente la información, puede presentar errores donde los encuestados subestiman o sobrestiman el tiempo que dedican a distintas actividades.

ingresos producto del trabajo, jubilaciones y otros aportes previsionales, con la finalidad de comprender sus patrones de uso del tiempo.

Es importante destacar que la medición de ingresos en la ENUT (tanto los ingresos del hogar como los del trabajo) tiene como finalidad caracterizar el uso del tiempo de las personas en función de su posición socioeconómica, mostrando las brechas existentes en ese sentido. Y no generar estimaciones sobre el ingreso de los hogares, por lo que no es recomendable comparar los datos obtenidos desde la ENUT con otras encuestas que miden en detalle los ingresos.

Módulo L. Contextualización al uso del tiempo: desde una perspectiva operativa, el módulo busca permitir a las y los informantes recordar las actividades realizadas los días asignados para el Cuestionario de Uso del Tiempo a través de una breve lista de actividades de distintos tipo. Desde una perspectiva analítica, la información generada es un insumo para el proceso de validación de datos anómalos, al identificar si estos días fueron habituales (o normales) o pasó algo distinto a un día normal que alterara el uso del tiempo de la persona.

Módulo M. Caracterización del tiempo de trabajo y traslado: su objetivo es medir las horas efectivas de Trabajo en la Ocupación y de traslado al lugar de trabajo para las personas ocupadas y el tiempo de búsqueda de trabajo de las personas desocupadas.

Módulo N. Cuidados a integrantes del hogar: el propósito de este módulo es identificar la participación y el tiempo dedicado por las y los integrantes del hogar a actividades de trabajo de cuidados para los otros integrantes del hogar en función de sus edades o condición de salud.

Módulo O. Trabajo doméstico: su objetivo es capturar información sobre el tiempo que destinan las personas integrantes del hogar a las distintas actividades de trabajo doméstico, como preparación de comidas, aseo, mantenimiento del hogar, entre otras.

Módulo P. Ayuda a otros hogares: el objetivo general del módulo es visibilizar los distintos tipos de transferencias de trabajo no remuneradas de los integrantes del hogar entrevistado para otros hogares o la comunidad.

Módulo Q. Cuidados personales: siguiendo el marco de análisis de la encuesta, desde el módulo Q en adelante, las actividades preguntadas no corresponden a la categoría de trabajo, sino a actividades personales, que se comprenden como aquellas actividades vinculadas al tiempo de “no trabajo” o el tiempo disponible.

El objetivo de este primer módulo de actividades personales es identificar la participación y el tiempo dedicado por las y los integrantes del hogar de 12 años y más a actividades de cuidados personales.

Módulo R. Actividades de educación y aprendizaje: busca identificar la participación y el tiempo dedicado por las y los integrantes del ho-

gar de 12 años y más a actividades de aprendizaje y educación. Distinguiendo el tiempo destinado a la educación formal de la no formal o informal.

Módulo S. Ocio y vida social: el objetivo es identificar la participación y el tiempo dedicado por las y los integrantes del hogar de 12 años y más a actividades de ocio y vida social.

Módulo T. Satisfacción con el uso del tiempo: el módulo consta de cuatro baterías de preguntas. Las tres primeras indagan en el nivel de satisfacción de cada persona con el tiempo destinado y la distribución de distintas actividades, mientras que la última batería busca conocer el nivel y las fuentes de estrés de la persona.

Los objetivos del módulo son: conocer el nivel de satisfacción de las personas integrantes del hogar de 12 años y más con el tiempo que destinan a sus distintas actividades; y conocer la satisfacción de las personas integrantes del hogar de 12 años y más con la forma en que se distribuyen los trabajos al interior del hogar.

Módulo U. Observaciones: al igual que en el módulo de observaciones del CH, el del CUT busca registrar información relevante del hogar, que dé cuenta de alguna situación anormal de las respuestas. Este complementa al módulo de contextualización (Módulo L).

5.4 MÉTODO DE CAPACITACIÓN

Con la finalidad de resguardar la calidad en la producción estadística se desarrolló un plan de capacitación que tuvo como orientación el entendimiento de la capacitación como un proceso continuo de formación y retroalimentación entre los equipos técnicos y de levantamiento.

La aplicación del plan de capacitación fue responsabilidad directa del equipo técnico de la ENUT, que realizó capacitaciones presenciales en todas las capitales regionales del país.

Junto a las capacitaciones realizadas por las y los analistas ENUT, se diseñó un Manual de Trabajo de Campo, que sirvió de base para la consulta permanente del personal de terreno; además de un manual específico con la especificación de las tareas de supervisión.

El plan de capacitación continua tuvo como principales hitos las siguientes etapas:

Autocapacitación: con la finalidad de estandarizar los contenidos y la forma de transmitirlos a los distintos equipos regionales, se capacitó de manera centralizada al personal responsable de realizar los distintos cursos en regiones.

Curso de inducción a encargados operativos de regiones: presentación de la encuesta a los responsables regionales del trabajo de campo con

la finalidad de explicar los objetivos y alcances operativos del proyecto.

Curso de inducción a encuestadores en regiones: corresponde al primer acercamiento de las y los encuestadores al trabajo de la ENUT. En esta etapa se impartieron talleres teóricos y prácticos que incluyeron tres grandes tipos de contenidos: conceptos básicos de las encuestas de hogares y protocolos de trabajo en terreno; marco de análisis de la ENUT, relevancia de la medición del trabajo no remunerado y enfoque de género; y reglas de aplicación de los cuestionarios.

Capacitación de supervisores: como complemento al curso de inducción de encuestadores, a las personas responsables de tareas de supervisión se les realizó una capacitación adicional enfocada en los criterios de revisión de cuestionarios, supervisión de labores de terreno y elementos administrativos de la encuesta.

Visitas técnicas: durante el desarrollo del trabajo de campo se realizaron dos rondas de visitas técnicas a regiones, en las que se acompañó a las y los encuestadores en sus labores de campo y revisión en oficina; identificando dificultades, necesidades de reforzamiento técnico u operativo y retroalimentando a los equipos en función de lo observado.

Los aprendizajes generados a partir de las visitas técnicas demostraron que estas son un elemento central en el aseguramiento de la calidad de la generación de información estadística, en tanto permite la evaluación y retroalimentación de las prácticas realizadas por los equipos de terreno.

Capacitación de personal nuevo: para enfrentar la rotación de personal en los distintos equipos de levantamiento se diseñó un sistema mixto que consideró una inducción abreviada bajo responsabilidad de los analistas técnicos, realizada a través de videoconferencias y un protocolo de monitoreo a cargo de los mismos equipos en regiones.

Capacitación de analistas: el diseño organizacional bajo el que se levantó la ENUT 2015 consideró un equipo de analistas de operaciones que tuvo como finalidad revisar de manera centralizada las encuestas recibidas de terreno de modo de identificar errores y corregir o devolver a terreno en caso de detectar errores que requiriesen contactar a los informantes.

Para este equipo se diseñó una capacitación alternativa, que relevó los elementos más críticos de cada formulario y los vínculos entre ellos, como guía para la revisión.

5.5 PRUEBA PILOTO

Con la finalidad de evaluar el diseño de la encuesta, se realizó una prueba piloto que fue ejecutada en tres regiones diferentes: Arica y Parinacota (100 casos urbano); Metropolitana (100 casos urbano) y Aysén (50 casos urbanos y 50 rurales).

El ejercicio tuvo como objetivos evaluar en terreno el diseño del operativo de campo, cuestionarios, proceso de capacitación y el comportamiento de los principales indicadores, en función de las siguientes dimensiones:

a) Operativo de campo

1. Evaluar el diseño de la metodología de levantamiento propuesta.
2. Conocer la cantidad de visitas necesarias para el logro completo de los hogares en las viviendas muestreadas.
3. Medir y caracterizar las razones de no logro para ajustar la planificación del terreno.

b) Diseño de cuestionario

1. Evaluar la comprensión de las preguntas por parte de los informantes y encuestadores.
2. Conocer la capacidad de los distintos informantes de entregar la información solicitada a partir del nivel de completitud de las respuestas.
3. Estudiar el flujo y ritmo de la aplicación de los distintos cuestionarios, así como el tiempo que ésta demora.
4. Evaluar la aplicabilidad del set de preguntas para la identificación de hogares.

c) Proceso de capacitación

1. Evaluar las dificultades y nudos críticos en el proceso de transmisión de conocimientos.

d) Principales indicadores

1. Analizar el comportamiento de los principales indicadores de la encuesta.
2. Estudiar la consistencia de las relaciones entre variables teóricamente relacionadas.

El ejercicio del piloto contó con distintas fuentes de información sobre el proceso de producción de datos sobre uso del tiempo, que permitieron incorporar e integrar aprendizajes de distinto tipo.

A continuación se presentan las principales etapas del piloto y las distintas fuentes de información utilizadas para generar información sobre cada una de ellas.

Capacitación

- Entrevistas: información cualitativa en base a una entrevista grupal (o grupo de discusión) de los capacitadores del equipo ENUT, y entrevistas a encuestadores y supervisores.

Levantamiento de datos

- Entrevistas: información cualitativa sobre la experiencia de encues-

taje del personal de campo (encuestadores y supervisores) y de los y las analistas técnicos del equipo ENUT.

- Registro de levantamiento: información cuantitativa sobre el proceso de encuestaje generada a partir de la hoja de ruta de las encuestas.

Diseño de cuestionarios

- Entrevistas: información cualitativa sobre la experiencia de encuestaje de encuestadores y supervisores/as, así como sobre la observación y experiencia de encuestaje de parte de los y las analistas técnicos del equipo ENUT.
- Pautas de observación: información cuantitativa sobre las dificultades en la aplicación de encuestas por el personal de campo.
- Base de datos Piloto ENUT: información cuantitativa obtenida a partir de la aplicación de los distintos cuestionarios.

A partir de la sistematización y análisis de la información generada en el piloto se realizaron distintos ajustes al diseño original, mejorando el desempeño de la metodología, capacitaciones e instrumentos de generación de datos.

5.6 MÉTODOS DE SUPERVISIÓN

Cada equipo de levantamiento fue dirigido por un supervisor o supervisora, quien realizó dos formas de supervisión:

1. Gabinete: revisión de un 100% de la carga diligenciada por el personal de campo.
2. Terreno: considera un 10% de la carga regional mediante la aplicación de una re-entrevista a un informante idóneo de la vivienda seleccionada en terreno.

Además de este 10%, se supervisó en terreno:

- La totalidad de viviendas identificadas como no elegibles o de elegibilidad dudosa.
- Los hogares donde se haya rechazado participar de la encuesta.
- Los hogares donde se haya identificado un integrante del hogar no elegible por condición física o mental.

5.7 MÉTODOS DE PROCESAMIENTO DE DATOS

5.7.1 Digitación

Las encuestas visadas por las y los supervisores en terreno fueron digitadas de manera centralizada en las oficinas del INE en Santiago. En esta etapa se ingresó toda la información registrada por el encuestador, a través del Sistema Operativo de Gestión de los Hogares SGOH.

El diseño de la interfaz gráfica del formulario de digitación se realizó en base al formulario en papel, para así evitar distorsiones al momento del ingreso de la información.

A medida que el personal de digitación ingresa los datos, el sistema informático verifica, mediante la aplicación de una serie de reglas de validación, si la información fue ingresada de manera correcta, detectando los errores e informando explícitamente de cada uno de ellos al momento en que se producen.

Este procedimiento, propio de la validación de primer nivel, tiene por objetivo comprobar que durante el proceso de digitación se ingresen únicamente valores válidos. Un ejemplo de validación de primer nivel es la limitación de rango de respuesta, es decir, si en formulario la respuesta válida es valor 1 o 2, esta validación arroja error si es que existe un valor distinto a estos (incluso si es vacío cuando debe ir contestada).

La tarea del personal de digitación se remite exclusivamente a traspassar la información registrada en el papel al sistema informático. De esta forma, si se encuentra con errores desde su origen, no realiza correcciones al formulario y solo puede resolver errores originados en su propia digitación en el momento del ingreso de la información, controlados mediante el sistema de validación en línea. Cualquier problema que no tenga relación con el ingreso, debe ser reportado para ser solucionado en la fase de análisis.

5.7.2 Análisis

El análisis constituye un nivel especializado en el control de calidad de la información, enfocado en las características lógicas y procedimentales de la encuesta. En ese sentido, los y las analistas requieren un conocimiento profundo y acabado del instrumento en cuanto a las relaciones lógicas entre las distintas preguntas y variables.

Para la realización de su labor, cada analista además de la revisión visual de cada cuestionario, se debe apoyar en el sistema informático y en los registros realizados por encuestadores y supervisores, con especial atención a las anotaciones realizadas en la página de observaciones incluida al final del cuestionario.

En la etapa de análisis se da tratamiento a las validaciones de primer nivel no corregidas en la digitación, y a las validaciones de segundo nivel, que se ejecutan sobre el total del formulario ya ingresado en el sistema (digitado), y se enfocan en la coherencia de información entre secciones, saltos de preguntas y preguntas dependientes entre sí.

En caso de que no se detecten errores, la encuesta se valida y pasa a ser almacenada, completando de esta forma su flujo. En caso contrario, si se detectan errores, el sistema informático arroja un informe de ellos, el que debe ser recibido por el analista.

Además del análisis y validación realizada desde el equipo de opera-

ciones, en el equipo técnico de la encuesta se realizó un proceso de revisión de los distintos datos de ingresos en función de las características particulares de cada partida.

Validación de montos por disposiciones legales: Uno de los elementos revisados fue los ingresos con disposiciones legales que establecen restricciones y límites en los montos de cada una de las variables. Las variables revisadas fueron la K55: Aporte Previsional Solidario de Invalidez (APSI), K56a: Jubilaciones y K56b: Aporte Previsional Solidario de Vejez (APSV). Se revisaron las distintas combinatorias entre estos tres montos, junto a las restricciones de edad y topes máximos de montos, estableciendo soluciones para cada combinación. En general, cuando los montos excedían el límite se hicieron modificaciones a las variables para así conservar la mayor cantidad de datos provenientes de terreno.

Para la validación de la pregunta G11d: Asignación Familiar se utilizó una variable *proxy* de la cantidad de cargas familiares en el hogar, estableciendo el máximo que debería reportar un hogar, reduciendo los montos que sobrepasaran el máximo. Por su parte, la G11e: Subsidio Familiar se validó estableciendo los múltiplos del monto especificado por ley, revisando aquellos que no estuvieran dentro de las posibilidades y corrigiendo, según *proxy* de cargas.

Ingresos cero: se revisaron en papel todos los casos con ingresos cero en las variables K51/K52: Ingresos de la ocupación principal, K54: Ingresos de segundas ocupaciones y K56a: Jubilaciones con el objetivo de corregir errores de digitación.

En la mayoría de los casos se observó que había una justificación de que la persona declaró no saber o no querer entregar el monto, por lo que se asignó el código de no respuesta. En los casos sin justificación, también se asignó el código de no respuesta, siguiendo el criterio de ingreso devengado. Por otro lado, cuando la persona declaró no haber tenido ganancia se mantuvo el cero.

6. Metodología de análisis

6.1 ANÁLISIS DE INFORMACIÓN PARA LA MEJORA DE LA BASE DE

Los distintos procedimientos de mejora de la información contenida en la base de datos se pueden agrupar en tres grandes actividades que se detallan a continuación: análisis de suficiencia de información, identificación y tratamiento de *outliers* e imputaciones.

6.2 ANÁLISIS DE SUFICIENCIA DE INFORMACIÓN

El análisis de suficiencia de información tuvo por objetivo responder a los problemas de no respuesta al ítem en las variables de uso del tiempo. Para ello, se identificaron las actividades con mayor impacto

en la explicación de la varianza del estimador del tiempo de trabajo no remunerado, y cuya ausencia generaría una subestimación en los principales indicadores de la encuesta.

Para la definición de las actividades consideradas como excluyentes, se consideró tanto las medias de tiempo como la prevalencia de cada actividad. En función de este análisis se definió un set de variables cuya presencia fue definida como criterio excluyente de logro del CUT. De este modo, los cuestionarios sin respuesta en alguna de estas actividades prioritarias pasaron a formar parte de un segundo nivel de *break off* y fueron eliminados de la base de datos.

6.3 TRATAMIENTO DE OUTLIERS DE TIEMPO.

El listado de actividades con que se midió el uso del tiempo de la población, permitió que cada persona declarase el tiempo destinado a cada actividad de manera abierta, considerando como rango válido de respuesta desde un minuto a las 24 horas del día.

Esta determinación se tomó en base a que la inexistencia de antecedentes nacionales empíricos sobre el tiempo destinado impedía definir validaciones con rangos de tiempo por actividad más acotados. Por otro lado, el uso de cuestionarios en papel en vez de dispositivos digitales, dificulta la aplicación de estos rangos.

En este contexto, luego de generada la información en terreno fue necesario evaluar el comportamiento de los datos observados, donde la revisión de las medidas de tendencia central y dispersión confirmaron la existencia de datos anómalos para las distintas actividades consultadas.

A partir de este análisis se procedió a definir un modelo estadístico para la identificación de *outliers*. El modelo definido fue el de Hadi, en tanto permite incorporar distintas variables de contextualización y evaluar así el comportamiento de los datos en función de las características de la persona informante.

Para la generación de los perfiles de informantes se probó con distintas variables de clasificación en función de las actividades evaluadas, pero al no observarse resultados consistentes con las variables probadas, y dada la cantidad de actividades medidas, se optó por utilizar las variables de sexo y rangos de edad para todas las actividades del cuestionario.

Definidas las variables de agrupamiento de los casos, el modelo identificó los casos atípicos en función de la distancia de las observaciones al centroide de cada perfil de informantes, identificando los datos estadísticamente anómalos para cada una de las variables. Sin embargo, la falta de antecedentes sobre la variabilidad del tiempo destinado a las distintas actividades, llevó a tomar la definición de generar un segundo corte dentro de los *outliers* identificados en una primera instancia, es-

tandarizando las distancias calculadas por el modelo en un puntaje de 0 a 1, y definiendo como outliers los casos donde la distancia estandarizada fuese mayor a 0,5. Identificando así como *outliers* los casos más lejanos entre los estadísticamente definidos como atípicos.

Todos los *outliers* fueron revisados en papel de modo de identificar errores de digitación y elementos de contexto que justificaran un comportamiento poco habitual. Para ello se utilizó la información del módulo L, de contextualización de la encuesta, donde se consultó sobre características generales y la tipicidad de los días de referencia; y el módulo de observaciones de la encuesta, donde el personal de campo registró cualquier información complementaria que permitiera explicar informaciones atípicas. Con estos antecedentes se evaluó los datos anómalos de cada variable, corrigiendo la digitación, validando el dato a partir de la información de contexto, o recodificando como valor perdido en los casos donde no se logró una justificación del comportamiento atípico.

6.4 DEFINICIÓN SOBRE LA IMPUTACIÓN DE DATOS DE TIEMPO

Al analizar la falta de respuesta a las variables de tiempo se observó que la prevalencia de no respuesta era bastante baja tanto a nivel de personas (pocas actividades sin respuesta por informante), como a nivel de actividades (pocos informantes sin respuesta por variable). Por otro lado, se observó que la no respuesta distribuía aleatoria por sexo y edad para parte importante de las variables.

En función de lo anterior, de las depuraciones previas donde se eliminaron los casos sin respuesta en las variables excluyentes, la falta de experiencias internacionales de imputación de datos de tiempo, las recomendaciones de expertos del mismo INE y de CEPAL; y considerando los riesgos de generar sesgos en las estimaciones, se definió no imputar datos de tiempo para la ENUT 2015.

6.5 TRATAMIENTO DE *OUTLIERS* DE INGRESOS.

Para validar los datos de ingresos se utilizaron diferentes tratamientos dependiendo de la partida de ingresos revisada.

En las variables G11a: Pensión de alimentos, G11b: Dinero aportado por personas ajenas al hogar, G11c: Dinero recibido por arriendo de casas, departamentos, oficinas, locales comerciales, estacionamientos, tierras u otros, K51/K52: Ingreso de la ocupación principal, K54: Ingreso de segundas ocupaciones, K56a: Jubilaciones; se aplicó el método estadístico *Adjusted Outlyingness*, tomando en consideración algunas variables de clasificación según la partida de ingresos analizada. Para cada variable se seleccionó un porcentaje de los casos con mayor dis-

tancia. Todos estos casos se corrigieron o mantuvieron según lo indicado en la encuesta física.

6.6 IMPUTACIÓN DE INGRESOS DEL TRABAJO Y JUBILACIONES

Debido a la presencia de no respuesta parcial en los datos de ingresos reportados, fue necesario realizar un proceso de imputación en los ingresos más relevantes, a saber, ingresos del trabajo y jubilaciones.

A nivel muestral, de un total 14.287 personas clasificadas como ocupadas, 14.014 señalan que el ingreso disponible proviene de la ocupación descrita previamente en el módulo K, lo que representa el 98,09% de los casos que pueden ser donantes o estar sujetos a imputación en caso de tener el dato faltante.

De este conjunto, 932 son personas ocupadas dependientes y 339 son personas ocupadas independientes que no poseen respuesta en sus ingresos, representando una tasa de 8,7% y 10,27% de no respuesta respectivamente. Esto significa que las variables son posibles de imputar de acuerdo a los criterios planteados por Galván & Medina (2007).

En el caso de las personas que solo perciben ingresos de la jubilación (2.960 observaciones válidas), 150 casos presentan problemas de no respuesta en el ítem, lo que equivale al 5,07% del total. Tal como en el caso de los ingresos del trabajo, el porcentaje de casos sin respuesta se encuentra entre los rangos recomendados para la imputación.

A partir del análisis de aleatoriedad (test de Little), de correlaciones y la revisión de las mejores prácticas desarrolladas por la oficina estadística, se optó por una variante del método de *Hot Deck*, con niveles definidos de acuerdo a las variables disponibles en la encuesta.

Para la imputación de las variables de ingresos del trabajo se utilizaron las variables de grupo ocupacional, clasificación de la situación en el empleo, sexo, edad, nivel educacional y horas trabajadas a la semana.

Para la imputación de las jubilaciones se utilizó el sexo, la edad y el nivel educacional de la persona.

En ambos casos las imputaciones se realizaron según área geográfica. Además de la no respuesta al ítem en los Cuestionarios de Uso del Tiempo logrados, se da la situación de falta de respuesta en las variables de situación en el mercado laboral e ingresos cuando el cuestionario individual no fue logrado. Esto cobra relevancia al momento de construir la variable agregada de ingreso monetario, junto con quintil de ingresos del hogar, ya que no se cuenta con la información de algunos integrantes del hogar de 15 años y más que podrían ser perceptores de ingresos.

Respecto a la magnitud, en el 18,8% de los hogares hay al menos una persona de 15 años o más sin datos en el CUT. En términos de individuos, de las 26.601 personas elegibles, 2.985 tienen 15 años o más y

no respondieron el CUT, lo que representa el 11.2%.

Debido a que en estos casos la cantidad de información disponible es menor respecto a los cuestionarios logrados, se realizó un proceso de imputación diferenciado.

Para la imputación de ingresos del trabajo, se efectuó previamente la imputación de la condición de actividad económica (ocupado/no ocupado) con una variante del método *Hot Deck* usando variables de área geográfica, sexo, edad, nivel educacional, condición de proveedor principal del hogar, estado civil, razón de no estudio por motivos de trabajo, cantidad de personas ocupadas en el hogar y cantidad de menores de 15 años en el hogar.

Luego, sobre los casos donde se imputó la condición de ocupado, se imputó la mediana del ingreso del trabajo también con una variante de *Hot deck*, en este caso con las variables de área geográfica, sexo, edad, nivel educacional, condición de proveedor principal del hogar, estado civil y razón de no estudio por motivos de trabajo.

Además del ingreso del trabajo, se imputó el ingreso por jubilaciones para el subconjunto de personas de 60 años y más con cuestionario no logrado, los que representan el 1,8% del total personas elegibles. Nuevamente se utilizó una variante del método *Hot Deck*, tomando la mediana del ingreso de cada uno de los niveles definidos con las variables de área geográfica, estado civil, sexo, edad, nivel educacional y condición de proveedor principal del hogar.

6.7 ANÁLISIS DE COMPORTAMIENTO DE RESULTADOS

Las estimaciones obtenidas en las variables de tiempo y participación fueron validadas a partir de dos procesos de crítica: la comparación con otras encuestas de uso del tiempo de la región y la crítica de expertos.

La comparación con otras encuestas de uso del tiempo de la región se realizó fundamentalmente con las estimaciones de variables similares de México 2014, Uruguay 2013 y Colombia 2012. Si bien lógicamente existen diferencias en los valores de las estimaciones, los patrones de comportamiento de los datos resultan concordantes con la revisión efectuada.

A nivel nacional, se compararon los resultados con la Encuesta Experimental de Uso del Tiempo del año 2007, que si bien presenta importantes diferencias tanto en la metodología del cuestionario como en su cobertura geográfica (la ENUT 2015 tiene cobertura nacional urbana, mientras la EUT 2007 abarcó solamente la Región Metropolitana), corresponde al único antecedente oficial en Chile de una encuesta de uso del tiempo. También se observan patrones de tiempo similares.

En cuanto a la crítica de expertos, los datos fueron presentados en mesas técnicas al interior de la oficina de estadísticas y a la División de Asuntos de Género de la CEPAL.

7. Productos estadísticos

La base de datos de la ENUT 2015 se encuentra disponible para todo usuario en el sitio web de la encuesta. Para facilitar su uso, se generó un manual de usuario con las especificaciones necesarias para el manejo de la base en SPSS y STATA.

Junto a la base de datos y su manual de usuario, en la página www.ine.cl se puede descargar la documentación metodológica y el documento de resultados.

8. Bibliografía

- **Aguirre, Rosario** (2009). La necesaria redefinición de la noción de trabajo
- **CEPAL** (2015). Clasificación de Actividades de Uso del Tiempo para América Latina y el Caribe (CAUTAL).
- **Feres, Juan Carlos** (2009). "Pobreza y Uso del Tiempo". Curso Internacional: "Redistribución del Tiempo, un indicador de igualdad". CEPAL, 2008.
- **Galván y Medina** (2007). Imputación de datos: teoría y práctica.
- **Gómez Luna, María Eugenia** (2010). Directrices y referentes conceptuales para armonizar las encuestas sobre uso del tiempo en América Latina y el Caribe, Documento para discusión. México.
- **INE** (2006). Encuesta Nacional De Empleo Diseño Muestral. http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/pdf/ene_espanol.pdf
- **Naciones Unidas** (2006). Guía de elaboración de estadísticas sobre el empleo del tiempo para medir el trabajo remunerado y no remunerado.
- **OIT** (2008). 18° Conferencia Internacional de Estadísticos del Trabajo, CIET
- **OIT** (2011). Manual de medición del trabajo voluntario http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_167833.pdf
- **OIT** (2013). 19° Conferencia Internacional de Estadísticos del Trabajo, CIET
- **SCN** (2009). Sistema de Cuentas nacionales 2008.
- **UNECE** (2013). Guidelines for Harmonising Time Use Surveys. Descargado en noviembre de 2014 de http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/bur/2013/12-Add.1_TimeUseSurvey_Guidelines_UNECE.pdf



Instituto Nacional de Estadísticas • Chile

Departamento de Estudios Sociales

www.ine.cl

 /ChileINE  /INE_Chile