

Encuesta de comportamiento lector 2014



www.plandelectura.cl



Gobierno
de Chile

www.gob.cl



Plan Nacional
de la Lectura

MANUAL DE DISEÑO MUESTRAL

ENCUESTA DE COMPORTAMIENTO LECTOR 2014

CONSEJO NACIONAL DE LA CULTURA Y LAS ARTES

**Consejo Nacional del Libro y la Lectura
Plan Nacional de la Lectura
Santiago, abril 2015**

INDICE

	Pág.
Presentación.....	01
1. ANTECEDENTES DEL DISEÑO MUESTRAL.....	02
2. POBLACIÓN OBJETIVO Y MARCO MUESTRAL.....	03
2.1. Población objetivo.....	03
2.2. Características del marco muestral.....	09
3. MUESTRA OBJETIVO: DISEÑO DE MUESTRA TEÓRICA.....	03
3.1. Tamaño de la muestra.....	06
3.2. Tipo de muestra.....	09
3.2.1. Estratificación de la muestra.....	09
3.2.2. Etapa de selección.....	10
a. Selección de unidades primarias de muestreo.....	10
b. Selección de unidades secundarias de muestreo.....	12
b. Selección de unidades terciarias de muestreo.....	14
b. Selección de las unidades últimas de muestreo.....	15
4. MUESTRA EFECTIVA: LOGRO DE TERRENO Y TASAS DE RESPUESTA.....	17
4.1. Logro de terreno.....	17
4.2. Códigos de disposición final de casos.....	20
5. FACTORES DE PONDERACIÓN Y EXPANSIÓN.....	22
6. COMPOSICIÓN DE LA MUESTRA EFECTIVA.....	24
6.1. Composición por comunas.....	24
7. CÁLCULO DE ERRORES EFECTIVOS SEGÚN MUESTRA LOGRADA.....	26
7.1. Estimación de varianza.....	26
7.2. Márgenes de error según variables de interés.....	29
7.3. Variables del diseño muestral.....	30
7.3.1. Variables dependiente asociadas a prácticas de lectura.....	31

PRESENTACIÓN

El siguiente manual es una guía para la correcta comprensión del diseño muestral.

Por ello, contiene:

- La población objetivo y el marco muestral utilizado.
- Explicación y descripción del diseño muestral.
- Información sobre la tasa de logro y la tasa de no respuesta en terreno.
- Muestra efectiva lograda, los mecanismos de ponderación y expansión, y los errores muestrales obtenidos.

1. ANTECEDENTES DEL DISEÑO MUESTRAL

La Encuesta de Comportamiento Lector 2014 se realizó por medio de entrevistas presenciales en hogares, sobre un universo que incluyó a personas de entre 9 y 65 años de edad, residentes en forma habitual en viviendas particulares ocupadas, localizadas en zonas urbanas de las 15 regiones de Chile. Tal como se señala en la tabla de síntesis, que se presenta a continuación, el trabajo de campo fue llevado a cabo en los meses de junio y julio de 2014 por el equipo de la Dirección de Estudios Sociales (Desuc) del Instituto de Sociología de la Pontificia Universidad Católica de Chile. El resultado fue una muestra nacional de 6990 personas, distribuidas en 116 comunas del país. Esto entrega un margen de error total de +/- 1,2%.¹ Se utilizó un sistema de sobredimensión de muestra y de registro de los resultados del trabajo de campo que permitieron determinar las tasas de respuesta (RR2: 84,1%) y cooperación (COOP2: 93,1%) a partir de los criterios establecidos por la American Association for Public Opinion Research (AAPOR).² Para variables específicas, se calculan errores muestrales que incluyen efecto de diseño.³

Debido a que el diseño muestral consideraba una muestra con afijación desproporcionada entre las distintas regiones del país, se aplicó un ponderador para asegurar la correcta representación de la población nacional. El ponderador fue calculado considerando las probabilidades de selección, el ajuste por no respuesta y el ajuste de posestratificación por sexo, edad, región y nivel educacional.

¹ Para este cálculo se asumió un muestreo aleatorio simple, por lo que no incluye efecto de diseño.

² Más información en www.aapor.org

³ El efecto de diseño refleja la eficiencia o cambio en la precisión de los estimadores al utilizar un diseño muestral complejo, es decir, estratificado, por conglomerados en múltiples etapas, y con distintas probabilidades de selección.

2. POBLACIÓN OBJETIVO Y MARCO MUESTRAL

2.1. Población objetivo

La población objetivo para este estudio corresponde a **personas de entre 9 y 65 años de edad, que residan habitualmente en viviendas particulares ocupadas localizadas en zonas urbanas de las 15 regiones de Chile.**

Respecto de otras mediciones realizadas a nivel nacional, esta Encuesta presenta⁴ un componente de interés en términos de la población objetivo que considera: constituye una encuesta dedicada especialmente a la lectura, con una muestra urbana nacional con representatividad estadística para diversos tramos etarios. Esto permite identificar aspectos de la trayectoria lectora de las personas en sus distintas etapas de vida.

Quedan, por su parte, excluidas⁵ de la población objetivo las personas que residen en áreas clasificadas como rurales⁶, así como también las que residan en zonas urbanas de las comunas clasificadas como “áreas de difícil acceso (ADA)”⁷ por el Instituto Nacional de Estadísticas.⁸

La

⁴ La ECL 2014 mantiene la población objetivo entre 9 y 65 años de edad considerada para la ECL 2011. Sin embargo, esta Encuesta se concentra solo en zonas urbanas.

⁵ En atención a la escasa cantidad de habitantes en la parte urbana de la comuna de Putre, se optó por excluir también esta comuna de la población objetivo del estudio.

⁶ Las áreas rurales definidas por el INE a partir del Censo 2002 corresponden al “conjunto de entidades clasificadas como rurales de acuerdo a un tamaño poblacional menor a 1.000 habitantes o entre 1.001 y 2.000 habitantes con predominio de población económicamente activa (PEA) dedicada a actividades primarias”. Se entiende por actividad primaria a toda aquella actividad relacionada con la extracción de recursos naturales (agricultura, caza, pesca, minería, etc.).

⁷ Las comunas de difícil acceso definidas por el INE a partir del Censo 2002 corresponden a: Ollagüe, Isla de Pascua y Juan Fernández, Chaitén, Futaleufú, Palena, Lago Verde, Guaitecas, Chile Chico, O’Higgins, Tortel, Torres del Paine, Río Verde, Laguna Blanca, San Gregorio, Primavera, Timaukel, Cabo de Hornos y Antártica.

⁸ Para mayor información respecto de las áreas de difícil acceso (ADA) ver documento *Actualización del diseño muestral. Nueva Encuesta Nacional de Empleo*, julio 2006, INE, disponible en la página web: www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/metodologia/metodologia.php

Tabla 1 presenta la distribución de la población objetivo de la ECL 2014, sobre la base de estimaciones obtenidas desde la encuesta Casen 2011.⁹

Tabla 1.

Distribución de la población objetivo del estudio: personas entre 9 y 65 años que residen en áreas urbanas*, según región

Región	Todas las personas en el país**		Personas de 9 a 65 años en áreas urbanas*	
	Total (1)	Porcentaje (2)	Total (3)	Porcentaje (4)
País	16.962.515	100,0	11.511.485	100,00
15 Arica y Parinacota	175.295	1,0	122.085	1,1
1 Tarapacá	302.659	1,8	225.870	2,0
2 Antofagasta	552.156	3,3	427.340	3,7
3 Atacama	273.870	1,6	192.304	1,7
4 Coquimbo	717.663	4,2	450.875	3,9
5 Valparaíso	1.753.449	10,3	1.222.822	10,6
6 O'Higgins	881.735	5,2	483.224	4,2
7 Del Maule	1.001.828	5,9	520.244	4,5
8 Del Biobío	2.018.886	11,9	1.323.946	11,5
9 La Araucanía	952.813	5,6	494.927	4,3
14 Los Ríos	366.727	2,2	196.213	1,7
10 Los Lagos	824.463	4,9	454.562	3,9
11 Aysén	99.466	0,6	64.772	0,6
12 Magallanes	150.355	0,9	115.387	1,0
13 Metropolitana	6.891.150	40,6	5.216.914	45,3

Notas: (*) Se entiende por áreas urbanas aquellas denominadas ciudad (CD), urbanos (U) y resto de área urbana (RAU). (**) Los totales poblacionales en la columna (3) excluyen la población en áreas rurales, en áreas de difícil acceso y Putre urbano (Región 15).

Fuente: Estimaciones obtenidas a partir de la encuesta Casen 2011.

⁹ Para este estudio se utilizaron estimaciones de población desarrolladas a partir de la encuesta Casen 2011, y no las proyecciones publicadas por el Instituto Nacional de Estadísticas, ya que para los análisis que se desarrollan más adelante se requería de desagregaciones por grupos etarios a nivel comunal, que no están disponibles en las proyecciones del INE. Para la encuesta Casen 2011, el INE desarrolló estimaciones de población total a noviembre de 2011, las cuales fueron ajustadas en forma posterior para representar exclusivamente a la parte de la población que reside en viviendas particulares ocupadas.

2.2. Características del marco muestral

Para la selección de la muestra de la Encuesta de Comportamiento Lector 2014 se utilizaron dos marcos muestrales que mantiene vigente el INE para el 2013, y que son mutuamente excluyentes: el **marco urbano de manzanas (MM2008-B)** y el **marco de secciones (MS2002)**.

El **marco MM2008-B** es generado a partir de cartografía digital de alta validez métrica, actualizado al segundo semestre del año 2008. Las unidades primarias de muestreo en el área urbana se denominan **manzanas** y corresponden a delimitaciones geográficas fijas.

Por su parte, el marco de secciones (**MS2002**) fue producido a partir del Censo de Población y Vivienda 2002. Las unidades primarias de muestreo se denominan **secciones** y corresponden al área que pudo recorrer un censista el día del censo.

Si bien es cierto el **MM2008-B** cubre a nivel nacional aproximadamente el 97% de las viviendas en áreas urbanas, su cobertura llega al 90% aproximadamente en regiones como O'Higgins, el Maule y La Araucanía. Por otra parte, hay 87 comunas donde la totalidad de las viviendas urbanas están registradas en el marco de secciones. Por lo tanto, fue importante para la representatividad de este tipo de comunas el utilizar también dicho marco.¹⁰

¹⁰ Una descripción más detallada del marco muestral del INE en el documento *Metodología del diseño muestral y factores de expansión Casen 2011* (páginas 15 a 20). Las últimas versiones de encuestas de hogares de gran relevancia en Chile, como la Encuesta Nacional de Empleo (2010, 2011, 2012, 2013), la Encuesta de Caracterización Socioeconómica Nacional (2011, 2013) y la Encuesta de Seguridad Ciudadana (2010, 2011, 2012, 2013), han utilizado los marcos muestrales de secciones (MS2002) y manzanas (MM2008-B) para la selección de sus muestras.

3. MUESTRA OBJETIVO: DISEÑO DE MUESTRA TEÓRICA

3.1. Tamaño de la muestra

Para establecer el tamaño de la muestra, se consideró:

- Asumir un nivel mínimo de precisión (representatividad) estadística a nivel general, y de los grupos y variables de interés.
- Identificar la o las variables de interés principal de la encuesta, y eventuales *proxy*, a fin de poder hacer estimaciones de varianza, error y efecto diseño.
- Contar con una aproximación del efecto diseño asociado al diseño muestral a implementar (Valliant *et al.*, 2013).
- Estimar la tasa de no respuesta para anticipar pérdidas durante el trabajo de campo y seleccionar, por lo tanto, muestra adicional que permita compensar por estas pérdidas.

Los detalles y pasos referentes a estos cuatro elementos críticos para la propuesta de tamaño de muestra a nivel país y regional se presentan en el anexo II.A. Como fruto de este análisis, se propuso una muestra objetivo de 6900 casos.

Para efectos de determinar el tamaño de muestra “por seleccionar”, es imprescindible tener una idea de la pérdida de casos que se va a enfrentar por razones de no logro de entrevistas debido a rechazo o problemas con las direcciones¹¹. A partir de la construcción de estimadores, se proyectó una tasa de no respuesta de referencia para la ECL 2014 de entre 10% y 25% a nivel regional y en torno al 18% a nivel nacional.

¹¹ Se deben considerar las pérdidas de casos muestrales que ocurren en forma natural durante el proceso de recolección de datos, ya sea por problemas con las direcciones seleccionadas (por ejemplo, no existe dirección, no se puede acceder al lugar, etc.) o con los hogares seleccionados (no se logra contactar a los residentes, los residentes rechazan participar en la entrevista, etc.).

Esto significa que en total se propuso seleccionar aproximadamente 8.455 viviendas para lograr una muestra objetivo de aproximadamente 6.900 viviendas por encuestar ($=6.900/0.82=8.455$).

A su vez, para determinar el tamaño muestral objetivo en cada región se siguió la metodología utilizada por la encuesta Casen 2011, la cual incorpora procedimientos que siguen las recomendaciones de Valliant *et al.* (2013), respecto de la determinación del tamaño muestral para un parámetro poblacional de una variable de interés.

La columna 1 de la Tabla 2 muestra los resultados del ejercicio de determinación del tamaño muestral propuesto, a nivel regional y nacional, para la ECL 2014. En total, se propuso tamaños regionales que variaban entre 320 y 1.100 casos, a partir los parámetros estimados de Casen (variable \bar{Y}_2 , tasa de respuesta), el efecto diseño de referencia (Def_{fc}), niveles de precisión (error absoluto y error relativo) y los tamaños muestrales en cada etapa del proceso (bajo muestreo aleatorio simple, bajo diseño complejo y sobredimensionado por no respuesta) (Ver pasos seguidos en el anexo II. A).

La columna 4 muestra la tasa esperada de no respuesta para cada una de las regiones, la cual varía entre 15% y 25%. La columna 5, por su parte, refleja el tamaño de la muestra que se debe seleccionar para compensar las pérdidas de direcciones, lo que se denomina **muestra con sobredimensión**. El tamaño final de la muestra por seleccionar son aproximadamente 8.630 viviendas, con tamaños regionales entre 376 (Arica y Parinacota) y 1.467 (Metropolitana).

Una vez determinado el tamaño muestral para cada región, se procedió a distribuir la muestra en los estratos de muestreo que se conformaron en cada una de ellas, utilizando un mecanismo de asignación proporcional que estableció mayor parte de la muestra definida a nivel regional a las áreas con mayor cantidad de viviendas (según los totales en los marcos muestrales 2002 y 2008B).

Dados los tamaños muestrales en la columna 1, se espera obtener márgenes de error regional, considerando el diseño muestral del estudio, que fluctúen entre los 4,9 y 8,3 puntos porcentuales, para la variable de interés.¹² La misma información a nivel de macrozona se presenta en la Tabla 3.

Es importante recordar que los márgenes de error efectivos (obtenidos luego de recolectados los datos) para cada una de las variables de interés de la encuesta son presentados en la sección 6.2 de este manual. Estos son utilizados como referencia al momento de realizar el cálculo del tamaño requerido de muestra (estimados antes de recolectar los datos).

Tabla 2.
Distribución del tamaño muestral por región

Regiones	Tamaño muestral objetivo (teórico)	Parámetros referenciales para ECL 2014			Tamaño muestral por seleccionar (sobredimensión)
		Variable objetivo Y_2	Deff regional conservador	Tasa de no respuesta	
	(1)	(2)	(3)	(4)	(5)
Total	6.900	0,27	4,5	0,20	8.630
15	320	0,28	1,2	0,15	376
1	360	0,29	1,3	0,20	450
2	380	0,27	1,8	0,20	475
3	320	0,25	1,1	0,20	400
4	360	0,28	3,2	0,20	450
5	750	0,29	2,7	0,20	938
6	397	0,28	2,3	0,20	496
7	453	0,26	1,6	0,15	533
8	810	0,26	4,5	0,20	1.013
9	320	0,29	2,6	0,18	390
14	360	0,28	2,1	0,18	439
10	320	0,27	1,8	0,18	390
11	330	0,27	1,2	0,20	413
12	320	0,26	1,2	0,20	400
13	1100	0,27	5,0	0,25	1.467

¹² Se ha realizado un ejercicio de simulación de errores muestrales proyectados para la variable Y_2 "proporción p=50%". Este ejercicio toma como base el tamaño muestral, el efecto diseño y la tasa de respuesta utilizados en la Tabla 2. Es de interés revisar esta proporción del 50%, ya que a ella se atribuye la variabilidad máxima entre todas las proporciones, y por lo tanto sirve para ilustrar los errores proyectados más grandes que se podrían llegar a observar en la ECL 2014, de lograrse la muestra objetivo.

Tabla 3.
Distribución del tamaño por macrozona

Macrozona/Región	Tamaño muestral objetivo (teórico) (1)	Parámetros referenciales para ECL 2014			Tamaño muestral por seleccionar (9)
		Variable objetivo Y_2 (2)	Deff macrozona conservador (3)	Tasa de no respuesta (4)	
Total	6.900	0,27	4,5	0,2	8.630
Norte 1 15, 1	680	0,28	1,2	0,2	826
Norte 2 2,3,4	1.060	0,27	2,2	0,2	1.325
Centro 1 5	750	0,29	2,7	0,2	938
Centro 2 6,7	850	0,27	2,1	0,2	1.029
Centro 3 8	810	0,26	4,5	0,2	1.013
Sur 1 9,10,14	1.000	0,28	2,1	0,2	1.219
Sur 2 11,12	650	0,27	1,2	0,2	813
Metro 13	1.100	0,27	5,0	0,3	1.467

3.2. Tipo de muestra

El diseño muestral ejecutado se puede caracterizar como un muestreo de áreas probabilístico, estratificado geográficamente y multietápico, donde se seleccionaron "comunas" en una primera etapa, "manzanas censales" en una segunda, "viviendas ocupadas" en la tercera y "personas entre 9 y 65 años de edad" en una cuarta etapa. A continuación se describe en detalle la metodología para implementar la selección de la muestra objetivo propuesta de n=6.900 personas.

3.2.1. Estratificación de la muestra

En el diseño de una muestra, la estratificación corresponde al proceso de agrupar los elementos de una población en forma previa a la selección de la muestra. Los diseños estratificados se utilizan por distintos motivos (ver capítulo 4 en Lohr, 1999) y –en particular para ECL 2014– principalmente como una forma de mejorar la eficiencia del muestreo para las estimaciones de la población completa, al agrupar unidades que tienen propiedades similares de media y varianza.

Para ello, primero se fijó como objetivo producir estimaciones en cada una de las 15 regiones del país con cierto nivel de precisión estadística. El diseño muestral propuesto consideró la conformación de estratos de muestreo, también denominados “dominios de muestreo” (*sampling domains*), a partir de las 15 regiones del país.

En segundo lugar, se estratificó la muestra en forma previa a la selección de una submuestra de las áreas geográficas o comunas por incluir en la muestra. La siguiente sección –en la etapa de selección a nivel de comunas– describe los criterios utilizados para la identificación de las comunas que fueron incluidas en forma forzosa en la muestra, así como el método de submuestreo aplicado al resto de las comunas del país.

3.2.2. Etapas de selección

a. Selección de las unidades primarias de muestreo (comunas)

Las unidades primarias de muestreo (UPM) en la ECL 2014 corresponden a parte de las comunas ubicadas en las zonas urbanas del país, las que fueron estratificadas y seleccionadas según los criterios que se detallan a continuación.

Para esta Encuesta, las áreas geográficas que conforman las UPM corresponden a las 309 comunas con población urbana que pertenece a la población objetivo del estudio.¹³ Como es la práctica en muestreo de áreas geográficas, se seleccionaron las comunas en la muestra con probabilidad proporcional al tamaño, medido este último sobre la base del total de viviendas en la comuna respectiva, según el marco muestral de manzanas (MM2008).

¹³ De las 346 comunas que hay en el país, se excluyen las 18 comunas identificadas como áreas de difícil acceso (ADA), las 18 comunas que tienen solo población rural (y no están entre las ADA) y una comuna con población urbana muy pequeña (Putre). De esta forma, la población objetivo del estudio se concentra en 309 comunas con población urbana a lo largo del país.

Antes de la selección de comunas, sin embargo, se determinó que aquellas que cumplieran con alguno de los siguientes tres requisitos fueran incluidas con certeza en la muestra, y por esta razón se les denominó “comunas con inclusión forzosa (IF)”:

1. Tuvieran más de 100.000 habitantes entre 9 y 65 años, según las proyecciones de población del Instituto Nacional de Estadísticas a 2013.
2. Correspondieran a alguna de las 34 comunas de la Región Metropolitana pertenecientes al área conocida como Gran Santiago.
3. Correspondieran a la capital regional de alguna de las 15 regiones del país.

La

Tabla 4 presenta una clasificación de las 310 comunas en la población objetivo de la encuesta según su cantidad de habitantes, de las cuales **56 fueron incluidas en la muestra ECL 2014 por inclusión forzosa**.

Tabla 4.
Universo: número de comunas según cantidad de habitantes urbanos entre 9 y 65 años

Región	Todas las comunas Total	Comunas no-IF			Comunas IF		
		Menos de 25 mil hab.	Entre 25 mil y 100 mil hab.	Total	Entre 25 mil y 100 mil hab.	Más de 100 mil hab.	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
País	310	218	36	254	18	38	56
15 Arica y Parinacota	1					1	1
1 Tarapacá	4	2	1	3		1	1
2 Antofagasta	7	5		5		2	2
3 Atacama	8	6	1	7		1	1
4 Coquimbo	13	10	1	11		2	2
5 Valparaíso	36	25	8	33		3	3
6 O'Higgins	32	28	3	31		1	1
7 Del Maule	30	26	3	29		1	1
8 Del Biobío	54	43	7	50		4	4
9 La Araucanía	32	29	2	31		1	1
14 Los Ríos	12	11		11		1	1
10 Los Lagos	23	19	2	21		2	2
11 Aysén	4	3		3	1		1
12 Magallanes	3	2		2	1		1
13 Metropolitana	51	9	8	17	16	18	34

Nota: El total de habitantes utilizado en la clasificación de las comunas según cantidad de habitantes excluye a la población en áreas rurales, en áreas de difícil acceso y Putre urbano (región 15).

Fuente: La población entre 9 y 65 años, estimada a partir de Casen 2011.

Por su parte, a nivel de universo, **las comunas sin inclusión forzosa** (no-IF) congregan alrededor de un tercio de la población objetivo de la encuesta ECL 2014, es decir, aproximadamente 3,4 de las 11,5 millones de personas entre 9 y 65 años de edad. Las comunas no-IF se distribuyen a lo largo de todo el territorio nacional y presentan gran variabilidad en cuanto al total de personas que las componen.

Para la selección de las comunas no-IF se llevó a cabo un proceso de construcción de 49 estratos de muestreo con sus respectivas 169 UPM, de acuerdo a los procedimientos y fundamentos desarrollados en el anexo II.A, utilizando un método de selección proporcional al tamaño de las UPM.

De esta forma, se procedió a la selección de las 49 UPM sin inclusión forzosa, que se sumaron a las 56 UPM con inclusión forzosa, lo que dio un **total de 105 UPM** que conformaron la muestra ECL 2014. El resumen de la cantidad de UPM seleccionadas a nivel regional se encuentra en la Tabla 5.

Tabla 5
Unidades primarias de muestreo (UPM) seleccionadas por región

Región	UPM	% UPM
1	2	1,9
2	3	2,9
3	2	1,9
4	4	3,8
5	12	11,4
6	5	4,8
7	6	5,7
8	14	13,3
9	5	4,8
10	4	3,8
11	2	1,9
12	2	1,9
13	41	39,0
14	2	1,9
15	1	1,0
Total	105	100

b. Selección de las unidades secundarias de muestreo (manzanas)

Las unidades secundarias de muestreo (USM) en la ECL 2014 corresponden a las manzanas censales o secciones ubicadas en las zonas urbanas del país. Con el objetivo de controlar los costos del proceso de encuesta en terreno, se propuso definir un número fijo de 5 viviendas por seleccionar en cada manzana. Las manzanas en el MM2008B presentan tamaños muy disímiles, con las más pequeñas con solo 2 viviendas. Para evitar que manzanas con tamaños pequeños sean seleccionadas en la muestra, se excluyeron del proceso de selección aquellas manzanas con menos de 8 viviendas. Esta estrategia es actualmente implementada por el INE en la selección de manzanas de las encuestas NENE, Casen y ENUSC.¹⁴ La columna 1 en la

Tabla 6 presenta el total de manzanas/secciones seleccionadas en la muestra en cada región del país.

Para la selección de manzanas se pueden emplear distintos métodos. Como se indicó en las secciones previas, los métodos que consideran estratificación resultan más eficientes que el método aleatorio simple. El marco muestral fue ordenado geográficamente para tener una estratificación implícita geográfica. Posteriormente, la selección de las manzanas dentro de las comunas y pseudocomunas se hizo mediante un método de selección sistemático y con probabilidad de selección proporcional al tamaño (medido en términos del total de viviendas en la manzana, según los registros en el marco muestral utilizado para la selección).

El total de manzanas por seleccionar en cada comuna (UPM) viene determinado por la distribución proporcional de la muestra regional entre los 105 estratos de muestreo definidos para la ECL 2014 (ver detalles en el anexo II.A).

Como se verá a continuación para el proceso de selección de viviendas, se debió efectuar un proceso de empadronamiento¹⁵ (*listing*) de las manzanas y viviendas

¹⁴ Ver metodologías en Ministerio de Desarrollo Social (2012), e Instituto Nacional de Estadísticas (2011, 2013).

¹⁵ Durante el proceso de empadronamiento, el encuestador hace un recorrido completo por la manzana/sección seleccionada, lista las direcciones de todos los sitios y asigna un código único de acuerdo al tipo de edificación que se encuentra en cada sitio. Las manzanas/secciones con edificaciones de uso habitacional particular son consideradas "elegibles" para el estudio; las manzanas/secciones que no son de uso residencial (ej. sitios eriazos, plazas, establecimientos

seleccionadas. Siguiendo el protocolo establecido, en cada manzana empadronada se deben seleccionar 5 viviendas (5 casos). Fruto de este proceso de empadronamiento, de las 1727 manzanas (equivalentes a 8635 casos por encuestar) que consideraba la sobredimensión de la muestra, se confirmaron 1629 manzanas residenciales, correspondientes a 8145 casos.¹⁶

Tabla 6
Relación entre manzanas seleccionadas y manzanas válidas para la elección de viviendas

Región	Manzanas seleccionadas	Manzanas válidas	% válidas	Casos por encuestar Muestra sobredimensionada
	(1)	(2)	(3)	(4)
				335
15	75	67	89	
1	90	84	93	420
2	95	91	96	455
3	80	77	96	385
4	90	85	94	425
5	188	183	97	915
6	99	92	93	460
7	107	100	93	500
8	203	195	96	975
9	78	74	95	370
14	88	77	88	385
10	78	73	94	365
11	83	72	87	360
12	80	72	90	360
13	293	287	98	1435
Total	1727	1629	94	8145

c. Selección de las unidades terciarias de muestreo (viviendas)

Las unidades terciarias de muestreo (UTM) en la ECL 2014 corresponden a las viviendas en las manzanas/secciones ubicadas en las zonas urbanas del país. Como se indicó en la sección precedente, con el objetivo de reducir los costos del proceso de encuesta en terreno se propuso seleccionar un número fijo de **5 viviendas en cada manzana seleccionada**.

comerciales, educacionales, etc.) se consideran "no elegibles" para el estudio. Pero, además, hay otras manzanas que se considera "fuera de rango".

¹⁶ La diferencia de manzanas correspondió a manzanas no residenciales (ej., comercio, colegios) o manzanas que no fue posible localizar en terreno.

A diferencia de las comunas y las manzanas censales, las viviendas no son directamente identificables en el marco muestral de áreas urbanas del que se dispone (MM2008). Esto quiere decir que, antes de implementar la selección de las viviendas se debió construir un marco muestral en terreno, por medio del proceso de empadronamiento ya mencionado.

Finalmente, en la ECL 2014 para la selección de las viviendas por encuestar se utilizó el siguiente procedimiento: 1) identificar en el listado las viviendas “elegibles” en la manzana empadronada; 2) calcular el total de viviendas elegibles y dividir por 5 para obtener el período del salto sistemático; 3) seleccionar al azar una vivienda en cada manzana, y 4) a partir de la vivienda seleccionada, identificar las siguientes 4 viviendas por entrevistar aplicando el salto sistemático. Este método, denominado muestreo sistemático, asigna igual probabilidad de selección a cada vivienda en la manzana seleccionada.

d. Selección de las unidades últimas de muestreo (personas)

Las unidades últimas de muestreo (UUM) en este estudio corresponden a las personas de entre 9 y 65 años que son residentes habituales de las viviendas seleccionadas en la muestra. Al igual que sucede con la etapa anterior de selección, no se dispone de un listado de las unidades (las personas) que cumplen con estos requerimientos antes de iniciar el proceso de encuesta y, por lo tanto, se debe implementar un procedimiento que permita identificar a las personas elegibles en cada vivienda y luego seleccionar a una persona para entrevistar.

El método de Kish, considerado el método probabilístico por excelencia en la metodología de encuestas, asigna a cada persona elegible en la vivienda la misma probabilidad de selección en la muestra. A modo de resumen, el método consiste en:

1. Asignar a cada vivienda seleccionada una letra (letra Kish, asociada a la tabla Kish), que indica qué persona deberá ser entrevistada en la vivienda seleccionada.

2. Listar a todos los residentes de la vivienda seleccionada.
3. Enumerar de 1 a 6 a todas las personas que son parte de la población objetivo del estudio ("elegibles"), partiendo primero con los hombres, de mayor a menor edad, y luego con las mujeres, de mayor a menor edad.
4. Identificar en la tabla Kish, según el cruce de la letra Kish y la cantidad de personas elegibles en el hogar, el número de orden que define a qué persona de la vivienda se deberá entrevistar.

La ECL 2014 utilizó una adaptación del método de Kish que tiene las siguientes características:¹⁷

1. No requiere de la asignación previa de una letra/tabla Kish, sino que utiliza el último dígito de la dirección de la vivienda seleccionada para identificar a la persona que deberá ser entrevistada.
2. Lista a todos los residentes de la vivienda seleccionada.
3. Enumera de 1 a 10 a todos aquellos que son parte de la población objetivo del estudio ("elegibles", personas de 9 a 65 años), partiendo primero con el jefe(a) de hogar y siguiendo con el resto de los miembros de mayor a menor edad.
4. Identifica en la tabla Kish, según el cruce del último dígito de la dirección y el número de personas "elegibles" en el hogar, el número de orden que define a qué persona de la vivienda se deberá entrevistar.

Para este estudio se utilizó una adaptación del método, usando como mecanismo de asignación probabilístico el último dígito de la dirección de las viviendas seleccionadas en la muestra. Bajo el supuesto de que la distribución de los últimos dígitos es uniforme, entonces se puede asumir que cada dígito de 0 a 9 tiene una probabilidad de 1/10 de aparecer en las direcciones de la muestra.

¹⁷ El método original, desarrollado durante los años cuarenta en EE.UU., buscaba establecer una metodología simple que permitiera (1) la selección de personas al azar, y (2) que fuera de fácil aplicación y supervisión. El diseño original contempló 8 configuraciones que dan aproximadamente igual probabilidad de selección a cada adulto en las viviendas seleccionadas. Las tablas de Kish (1949) asignan igual probabilidad de selección a los adultos que residen en viviendas con 1, 2, 3, 4 y 6 personas. En viviendas con 5 adultos, las probabilidades de selección de cada adulto difieren ligeramente.

Para mayores detalles en relación con la implementación del método de selección, revisar los documentos Anexo II. B, “Manual del Trabajo de Campo ECL 2014” y “Hoja de Ruta 3 ECL 2014”.

4. MUESTRA EFECTIVA: LOGRO DE TERRENO Y TASAS DE RESPUESTA

4.1. Logro de terreno

Con el fin de obtener tasas de logro más conservadoras, para el cálculo de los siguientes indicadores se consideran la totalidad de los casos sorteados inicialmente. Es decir, se toma como base una muestra sobredimensionada de 8635 encuestas (5 casos en cada una de las 1727 manzanas) y se toma como muestra objetivo definitiva un total de 6906¹⁸ casos.

A continuación se presentan las tasas de encuestas completas (logro), rechazadas, no elegibles y con elegibilidad desconocida (ver Tabla 7).

En las manzanas elegibles utilizadas se obtuvieron **6990 encuestas completas**, lo que corresponde a una tasa de logro agregada de un **101% respecto de la muestra objetivo** y de un 80,9% respecto de la muestra sobredimensionada.

Esta tasa de logro es variable, dependiendo de las regiones de aplicación. Por ejemplo, considerando la muestra objetivo, en la Región Metropolitana se obtuvo un total de 1214 encuestas, lo que representa un logro de 110%, y en la VIII Región del Biobío y X Región de Los Lagos se alcanza un 97% de la muestra objetivo.

Bajo la columna de **rechazos** se consideran tanto los de los encuestados a nivel de hogar o informantes como los de la persona seleccionada mediante la tabla Kish (incluyendo los casos de Kish mal aplicadas y rechazadas por supervisión). En total

18 Si bien inicialmente se establecieron 6900 casos como muestra objetivo, producto de diferentes asignaciones de casos (a nivel regional y asignación de 5 encuestas por manzana), se obtiene una muestra objetivo de 6906 casos.

hubo 516 casos rechazados, los que corresponden a un 6% de la muestra sobredimensionada. La región que presenta una mayor tasa de rechazo es la Metropolitana, con un 12%, mientras que la XIV Región de Los Ríos y la XV Región de Arica y Parinacota no presentan casos de rechazos. Bajo la columna “**otro rechazo**” se consideran los casos en que se produce un rechazo indirecto, dados por problemas de idioma o de discapacidad física o mental de la persona seleccionada por tabla de Kish. También en esta categoría se incluyen los casos que fueron rechazados por el equipo de supervisión. Esta tasa alcanza un total de 4% a nivel país.

La columna de “no contacto” presenta la tasa de casos elegibles en que, tras tres visitas al hogar, no fue posible establecer contacto con la persona seleccionada. Es decir, se estableció contacto a nivel de vivienda, pero no con la persona seleccionada mediante la tabla Kish. En esta categoría hay 122 casos, los que representan un 1,4% respecto de la muestra sobredimensionada. La IV Región de Coquimbo presenta la mayor tasa de no contacto, con un total del 10,2% de los casos asignados.

Se obtuvo, por otra parte, una tasa de casos no elegibles (viviendas no residenciales) de un 3,7%. Este número contempla aquellos casos en manzanas no elegibles (sin viviendas residenciales) o viviendas en que no hubiesen habitantes que cumplieran con los criterios de elegibilidad establecidos. La XIV Región de Los Ríos presenta la tasa de no elegibilidad más alta, con un 10,2%.

Finalmente, un 3,9% de los casos a nivel agregado presenta una elegibilidad desconocida, donde no se estableció contacto con el hogar. Esta situación se da principalmente por complicaciones al momento de localizar las viviendas. La IX Región de La Araucanía tiene la tasa de no elegibilidad más alta, con un 9,7%.

Tabla 7
Tasas de logro, rechazo y elegibilidad de encuestas, por región

Región	MUESTRA		CASOS ELEGIBLES								OTRA ELEGIBILIDAD				
	Objetivo (1)	Sobredimensión (2)	Logro	% Logro / (1)*	% Logro / (2)*	Rechazo	% Rechazo / (2)*	Otro rechazo	% Otro / (2)*	No contacto	% No contacto / (2)*	No elegible (NE)	% NE / (2)*	ED**	% ED / (2)*
1	360	450	354	98,3	78,7	27	6,0	26	5,8	13	2,9	15	3,3	15	3,3
2	380	475	408	107,4	85,9	20	4,2	21	4,4	2	0,4	9	1,9	15	3,2
3	320	400	318	99,4	79,5	22	5,5	35	8,8	10	2,5	5	1,3	10	2,5
4	360	450	357	99,2	79,3	9	2,0	12	2,7	46	10,2	16	3,6	10	2,2
5	752	940	746	99,2	79,4	67	7,1	63	6,7	35	3,7	22	2,3	7	0,7
6	396	495	396	100,0	80,0	53	10,7	11	2,2	0	0,0	20	4,0	15	3,0
7	455	535	449	98,7	83,9	16	3,0	13	2,4	0	0,0	25	4,7	32	6,0
8	812	1015	790	97,3	77,8	86	8,5	39	3,8	4	0,4	29	2,9	67	6,6
9	320	390	317	99,1	81,3	1	0,3	9	2,3	0	0,0	25	6,4	38	9,7
10	320	390	311	97,2	79,7	18	4,6	12	3,1	1	0,3	30	7,7	18	4,6
11	332	415	332	100,0	80,0	5	1,2	18	4,3	5	1,2	25	6,0	30	7,2
12	320	400	319	99,7	79,8	17	4,3	13	3,3	1	0,3	38	9,5	12	3,0
13	1099	1465	1214	110,5	82,9	175	11,9	39	2,7	2	0,1	12	0,8	23	1,6
14	361	440	362	100,3	82,3	0	0,0	22	5,0	0	0,0	45	10,2	11	2,5
15	319	375	317	99,4	84,5	0	0,0	15	4,0	3	0,8	5	1,3	35	9,3
País	6906	8635	6990	101,2	80,9	516	6,0	348	4,0	122	1,4	321	3,7	338	3,9

Notas: (*) El valor "(1)" corresponde a muestra objetivo y el valor "(2)" a muestra sobredimensionada.
(**) Elegibilidad desconocida (ED).

4.2. Códigos de disposición final de casos (AAPOR)

A continuación se presentan diferentes tasas de logro utilizando los códigos de disposición final de casos (CDF), según las categorizaciones sugeridas por AAPOR.

Como se explicó anteriormente, se seleccionaron más casos de los requeridos para contrarrestar las pérdidas esperadas por no respuesta. Cada caso seleccionado tiene un CDF, el cual señala el estado en que quedó al cerrarse el proceso de levantamiento de información. Por ejemplo, existen códigos para encuestas completas, encuestas rechazadas, viviendas no elegibles o individuos no contactados. Esta metodología permite calcular indicadores como tasa de respuesta, tasa de rechazo, tasa de cooperación y tasa de contacto, los cuales se presentan a continuación.

Es necesario recordar que para los cálculos de estos indicadores se toman en cuenta **todos** los casos que fueron sorteados inicialmente, es decir, aquellos casos seleccionados antes de comenzar con la fase del empadronamiento. Al utilizar la muestra inicial se obtienen tasas más conservadoras y que describen el logro completo de terreno. El resumen de los CDF para el total de las encuestas se presenta en la Tabla 8.

Tabla 8.
Resumen de disposición final de casos

Código general	Casos	%	
Completas	6990	80,9	
Contacto Rechazos	516	6,0	
Casos elegibles	Total contacto	7506	86,9
	No contacto	122	1,4
	Otros	348	4,0
	Total casos elegibles	7976	92,4
No elegible	321	3,7	
Elegibilidad desconocida	338	3,9	
Total	8635	100	

Utilizando esta información, se procede a calcular la tasa de respuesta, la tasa de rechazo, la tasa de contacto y la tasa de cooperación, según los estándares de AAPOR¹⁹. Es necesario considerar que en las tasas calculadas a continuación los denominadores del cálculo no siempre son los mismos.

- **Tasa de respuesta (RR2):** Corresponde al número de entrevistas completas y parciales (aquí se consideraron válidas aquellas con más del 80% de las preguntas con respuesta, y con respuesta en las preguntas más importantes del cuestionario), partido por el total de casos elegibles seleccionados (se excluyen viviendas deshabitadas, casas de veraneo, locales comerciales, es decir, los casos no elegibles). Para una aproximación más conservadora, esta tasa asume que aquellos casos con elegibilidad desconocida son elegibles. En esta encuesta la tasa de respuesta fue de **84,1%**²⁰.
- **Tasa de rechazo (REF2):** Corresponde a la proporción de rechazos (se considera rechazo hogar, persona y encuestas incompletas) del total de casos elegibles (se excluyen viviendas deshabitadas, casas de veraneo, locales comerciales, es decir, los casos no elegibles). Para una aproximación más conservadora, esta tasa asume que aquellos casos con elegibilidad desconocida son elegibles. En esta encuesta la tasa de rechazo fue de **6,2%**.²¹
- **Tasa de cooperación (COOP2):** Corresponde a la proporción de casos en que se realizó una entrevista completa, respecto del total de casos contactados de la muestra. En esta encuesta la tasa de cooperación fue de **93,1%**²²
- **Tasa de contacto (CON2):** Corresponde a la proporción de casos en que se llegó a algún tipo de contacto (ya sea encuesta completa o rechazo), sobre el total de casos elegibles de la muestra. Para una aproximación más conservadora, esta tasa asume que aquellos casos

¹⁹ Para mayor detalle revisar http://www.aapor.org/Standard_Definitions2.htm

²⁰ $RR2 = \frac{\text{Completas}}{\text{Casos Elegibles+Elegibilidad Desconocida}} = \frac{6990}{7976+338} = \frac{6990}{8314} = 84,1\%$

²¹ $REF2 = \frac{\text{Rechazo}}{\text{Casos Elegibles+Elegibilidad Desconocida}} = \frac{516}{7976+338} = \frac{516}{8314} = 6,2\%$

²² $COOP2 = \frac{\text{Completas}}{\text{Completa+Rechazos}} = \frac{6990}{6990+516} = \frac{6990}{7506} = 93,1\%$

con elegibilidad desconocida son elegibles. En esta encuesta la tasa de contacto fue de **90,3%**²³

5. FACTORES DE PONDERACIÓN Y EXPANSIÓN

La Encuesta de Comportamiento Lector 2014, al igual que muchas otras encuestas, selecciona a ciertos grupos de la población a una mayor tasa en forma deliberada, con el objetivo de sobrerrepresentar o “fijar” a los grupos pequeños en la muestra. Esto permite a los analistas la posibilidad de realizar análisis tanto para la población completa como para esos pequeños, pero importantes, grupos de la población.

En la ECL 2014, por ejemplo, la Región de Arica y Parinacota representa el 4% de la muestra; sin embargo, en la población esta región corresponde a solo un 1,1% de las personas entre 9 y 65 años de edad que residen en las áreas urbanas del país. La Región Metropolitana, por otra parte, se encuentra subrepresentada, ya que corresponde a un 45% de la población pero solo a un 16% de la muestra ECL 2014.

Los “factores de expansión” (o “ponderadores”) permiten ajustar por las distintas tasas de subrepresentación o sobrerrepresentación de los casos en la muestra y son indispensables para realizar análisis con los datos de la ECL 2014.

El factor de expansión (asociado a cada uno de los entrevistados) se puede interpretar como el número de personas en la población que representa el entrevistado. Por ejemplo, si el factor de expansión asociado a una mujer es 2.800, entonces ella representa a 2.800 mujeres en la población. En la ECL 2014 se asignó un factor de expansión a cada persona entrevistada. El factor de expansión asignado a cada entrevistado contiene tres componentes.

²³ $CON2 = \frac{Completas+Rechazos}{Casos Elegibles+Elegibilidad Desconocida} = \frac{7506}{8314} = 90,3\%$

El **primer componente corresponde al inverso de las probabilidades de selección** de cada entrevistado. Si la probabilidad de selección es 1 en 2.000, entonces el factor de expansión inicial es de 2.000.

El **segundo factor corresponde a un ajuste de no respuesta**, con base en la probabilidad de que alguien en el hogar complete el registro de residentes y la probabilidad de completar la entrevista con el adulto seleccionado. Si la tasa de respuesta es 80%, entonces el factor de expansión ajustado por no respuesta es 2.500 ($=2.000/0.8$).

El **tercer factor** corresponde al ajuste a poblaciones totales, conocido también como "**calibración**" o **posestratificación**, y permite hacer calzar los totales poblacionales de sexo, edad, región y nivel educacional²⁴. Las tres primeras variables son estimadas a partir de las proyecciones de población desarrolladas por el INE en septiembre de 2014. Por su parte, el nivel educacional es estimado utilizando la distribución del nivel educacional del entrevistado según la Nueva Encuesta Nacional de Empleo (NENE) 2014.

Al término de cada uno de los pasos descritos se procedió a evaluar la presencia de valores extremos (*outliers*) del factor de expansión creado. Siguiendo la misma metodología empleada en la encuesta EANNA 2012, se evaluaron distintos puntos de corte para los valores extremos del factor de expansión.

²⁴ Para la ECL 2014, se posestratificó adicionalmente por nivel educacional, debido a que se trata de un determinante para las variable dependiente de interés.

6. COMPOSICIÓN DE LA MUESTRA EFECTIVA

6.1. Composición por comunas

A continuación se expone la composición de la muestra a nivel regional y comunal. Cabe recordar que las comunas fueron seleccionadas en un proceso de estratificación, señalado en el punto anterior. La distribución de casos por comuna en cada región se encuentra en la Tabla 9.

Tabla 9
Distribución de la muestra
en comunas seleccionadas

Región	Comuna	Casos por comuna	% Comuna / casos por Región	Casos por Región
1	ALTO HOSPICIO	112	32	354
	IQUIQUE	242	68	
2	ANTOFAGASTA	259	63	408
	CALAMA	114	28	
	TOCOPILLA	35	9	
3	CHAÑARAL	166	52	318
	COPIAPO	152	50	
4	CANELA	5	1	357
	COQUIMBO	131	37	
	ILLAPEL	62	17	
	LA SERENA	128	36	
	LOS VILOS	31	9	
5	ALGARROBO	26	3	746
	LA CALERA	25	3	
	EL QUISCO	45	6	
	LIMACHE	41	5	
	LOS ANDES	27	4	
	NOGALES	32	4	
	PAPUDO	5	1	
	QUILLOTA	32	4	
	QUILPUÉ	71	10	
	QUINTERO	53	7	
	SAN ANTONIO	34	5	
	VALPARAISO	145	19	
	VILLA ALEMANA	50	7	
VIÑA DEL MAR	156	21		
ZAPALLAR	4	1		
6	CODEGUA	10	3	396
	DOÑIHUE	67	17	
	GRANEROS	39	10	
	MACHALÍ	64	16	
	RANCAGUA	163	41	
7	SANTA CRUZ	53	13	449
	CAUQUENES	67	15	
	CURICÓ	90	20	
	LINARES	46	10	

	MAULE	35	8	
	PENCAHUE	9	2	
	SAN JAVIER	60	13	
	TALCA	142	32	
	ARAUCO	60	8	
	CHIGUAYANTE	49	6	
	CHILLÁN	49	6	
	CONCEPCIÓN	131	17	
	CORONEL	56	7	
	HUALPÉN	46	6	
	LAJA	52	7	
8	LOS ÁNGELES	74	9	790
	LOTA	43	5	
	MULCHÉN	24	3	
	NEGRETE	8	1	
	SAN PEDRO DE LA PAZ	41	5	
	TALCAHUANO	82	10	
	TOMÉ	34	4	
	YUMBEL	41	5	
	FREIRE	25	8	
	NUEVA IMPERIAL	57	18	
	PITRUFQUÉN	30	9	
9	PUERTO SAAVEDRA	1	0,3	317
	TEMUCO	145	46	
	TEODORO SCHMIDT	15	5	
	VILLARRICA	44	14	
	ANCUD	59	19	
10	FRUTILLAR	54	17	311
	OSORNO	85	27	
	PUERTO MONTT	113	36	
11	COIHAYQUE	244	73	332
	PUERTO AYSÉN	88	27	
12	PUERTO NATALES	52	16	319
	PUNTA ARENAS	267	84	
	BUIN	22	2	
	CALERA DE TANGO	5	0	
	CERRILLOS	19	2	
	CERRO NAVIA	30	2	
	COLINA	15	1	
	CONCHALÍ	25	2	
	EL BOSQUE	30	2	
	EL MONTE	15	1	
	ESTACIÓN CENTRAL	15	1	
	HUECHURABA	12	1	
	INDEPENDENCIA	15	1	
13	LA CISTERNA	10	1	1214
	LA FLORIDA	78	6	
	LA GRANJA	20	2	
	LA PINTANA	39	3	
	LA REINA	19	2	
	LAS CONDES	58	5	
	LO BARNECHEA	5	0,4	
	LO ESPEJO	20	2	
	LO PRADO	15	1	
	MACUL	25	2	
	MAIPÚ	92	8	
	MELIPILLA	13	1	
	ÑUÑO A	55	5	
	PADRE HURTADO	12	1	

	PEDRO AGUIRRE CERDA	25	2	
	PEÑAFLORES	14	1	
	PEÑALOLÉN	39	3	
	PROVIDENCIA	18	1	
	PUDAHUEL	45	4	
	PUENTE ALTO	113	9	
	QUILICURA	40	3	
	QUINTA NORMAL	24	2	
	RECOLETA	33	3	
	RENCA	24	2	
	SAN BERNARDO	39	3	
	SAN JOAQUÍN	15	1	
	SAN MIGUEL	20	2	
	SAN RAMÓN	17	1	
	SANTIAGO	57	5	
	TALAGANTE	10	1	
	VITACURA	17	1	
14	LOS LAGOS	178	49	362
	VALDIVIA	184	51	
15	ARICA	317	100	317
	TOTAL	6990		6990

6.2. Composición por variables sociodemográficas y variables de análisis

La Tabla 1 presenta variables sociodemográficas (sexo, edad, región, macrozona y educación del entrevistado) utilizadas en el proceso de posestratificación. Comparando las columnas con las proporciones de dichas variables, se puede observar la diferencia entre la muestra no ponderada y la ponderada, lo que expresa cómo la muestra es corregida según los parámetros considerados.

La importancia del ponderador se observa con mayor claridad en las variables geográficas (región o macrozona), debido a las características del tipo de muestra (con afijación).

Tabla 10
Composición de la muestra según variables sociodemográficas

Variables	Valores	Población representada	Casos de la muestra	Distribución muestral no ponderada	Distribución muestral ponderada
Total		12.180.471	6990	100%	100%
Sexo	Hombre	6.036.611	3081	44%	50%
	Mujer	6.143.860	3909	56%	50%
Edad	9 a 14 años	1.279.583	621	9%	11%
	15 a 24 años	2.446.829	1049	15%	20%
	25 a 44 años	4.629.483	2576	37%	38%
	45 a 65 años	3.824.576	2744	39%	31%
Región	Tarapacá	241.995	354	5%	2%
	Antofagasta	473.872	408	6%	4%
	Atacama	217.661	318	5%	2%
	Coquimbo	480.787	357	5%	4%
	Valparaíso	1.278.733	746	11%	10%
	O'Higgins	509.220	396	6%	4%
	Maule	549.491	449	6%	5%
	Biobío	1.390.586	790	11%	11%
	La Araucanía	524.397	317	5%	4%
	Los Lagos	469.617	311	4%	4%
	Aysén	72.490	332	5%	1%
	Magallanes	120.176	319	5%	1%
	Metropolitana	5.478.466	1214	17%	45%
	Los Ríos	215.239	362	5%	2%
Arica y Parinacota	157.741	317	5%	1%	
Macrozona	Norte	1.572.056	1754	25%	13%
	Centro	2.337.444	1591	23%	19%
	Sur	2.792.505	2431	35%	23%
	Región Metropolitana	5.478.466	1214	17%	45%
Educación entrevistado	Básica completa	2.958.813	1500	21%	24%
	Media comp. o inc.	5.401.987	3441	49%	44%
	Superior inc. o más	3.819.671	2049	29%	31%

En la Tabla **Tabla 1111** se presenta la composición de la muestra por región según las variables que se utilizarán en el análisis de los resultados, como la edad (dos categorías, adultos y menores), la educación del jefe de hogar y la actividad del encuestado.

Tabla 11.
Composición de la muestra por variables de análisis

Variables	Valores	Población representada	Casos de la muestra	Distribución muestral no ponderada	Distribución muestral ponderada
Edad	9 a 17 años	1.849.402	829	12%	15%
	18 a 65 años	10.331.069	6161	88%	85%
Educación del jefe de hogar	Básica completa	2.958.813	1500	21%	24%
	Media completa o incompleta.	5.401.987	3441	49%	44%
	Superior incompleta o más	3.819.671	2049	29%	31%
Actividad del encuestado	Trabajando	6.045.623	3633	52%	50%
	Estudiando	3.075.567	1318	19%	25%
	Realizando labores del hogar	1.811.820	1128	16%	15%

7. CÁLCULO DE ERRORES EFECTIVOS SEGÚN MUESTRA LOGRADA

7.1. Estimación de varianza

La varianza muestral es una medida del nivel de precisión de un valor estadístico bajo el paradigma del muestreo repetido, y está determinada por el diseño de la muestra y el valor estadístico que se quiere estimar. En el caso de la ECL 2014, por ejemplo, la varianza muestral refleja el hecho de que se entrevistó a 6.990 personas y no a los 12,8 millones de personas entre 9 y 65 años de edad en Chile.

La estimación de la varianza para la ECL 2014 requirió del uso de procedimientos que reflejaran en forma apropiada los principales efectos de las distintas características del diseño muestral empleado²⁵. La Base de datos ECL 2014 contiene tres variables que permiten identificar en forma adecuada el diseño muestral complejo de la encuesta: el factor de

²⁵ Los resultados que se presentan a continuación fueron desarrollados utilizando los módulos/comandos para el análisis de encuestas complejas en los paquetes estadísticos SPSS 19 y Stata 12, respectivamente. El uso de paquetes estadísticos que no tengan la capacidad de considerar el diseño complejo de la encuesta generará varianzas más pequeñas (subestimadas), y como consecuencia, intervalos de confianza y test estadísticos más pequeños, que pueden llevar a rechazar una hipótesis nula en forma incorrecta. Por esta razón, se recomienda explicitar todos los elementos del diseño complejo de la encuesta al momento de realizar análisis con los datos ECL 2014.

expansión ("FEXP"), el indicador de estratos ("VARSTRAT") y el indicador de conglomerados ("VARUNIT").

7.2. Márgenes de error según variables de interés

A continuación se presentan los márgenes de error para la ECL 2014, utilizando diferentes supuestos del tipo de muestra: aleatorio simple (MAS) y complejo, que incluye *efecto de diseño*. Si bien el error calculado suponiendo MAS es el indicador convencional que se suele presentar en diversos estudios, el más apropiado y riguroso es el segundo, considerando que esta encuesta contempla un diseño multietápico y estratificado.

Los diferentes errores se estiman para distintas variables de interés y sus respectivas varianzas. En primera instancia, se presenta el indicador utilizado en los cálculos del diseño muestral, el cual corresponde al porcentaje de personas que leen diariamente 15-20 minutos. Luego se presentan tres variables dependientes relevantes, que están asociadas a la prevalencia del nivel de lectura: personas que leen de vez en cuando de 15 a 20 minutos, lectura por motivos de ocio/entretención y lectura por motivos de estudio/trabajo²⁶. Finalmente, también se exhibe errores que consideran varianza máxima, asumiendo el escenario más conservador para el resto de variables no presentadas aquí.

Los errores de las variables dependientes se especifican para distintos segmentos de interés, asociados a variables relevantes para el análisis de resultados (sexo, tramo etario, región, educación del jefe de hogar y actividad del encuestado).

²⁶ Las preguntas a las que se hace referencia provienen de una recodificación de: A2 y A3 "¿Ud. lee, aunque sea de vez en cuando y por más de 15 minutos seguidos, libros, revistas, diarios, textos digitales o algún otro tipo de material?" y "En los últimos 12 meses, ¿con qué frecuencia ha leído durante 15 a 20 minutos de forma continua diarios, revistas, cómics, libros u otro material de lectura (en cualquier formato, es decir, impreso y/o digital?"; B4A1 "¿Cuántos libros leyó en cualquier formato, sea impreso o digital, por motivos de estudio/trabajo en los últimos 12 meses?"; y B4B1 "¿Cuántos libros leyó en cualquier formato, sea impreso o digital, por motivos de entretenimiento/ocio en los últimos 12 meses?".

7.3. Variable del diseño muestral

En primera instancia se revisarán los errores efectivos para la variable utilizada para el cálculo del diseño muestral de la encuesta (*porcentaje de personas que leen diariamente 15-20 minutos*).

La Tabla 12 presenta el **error absoluto o margen de error al 95% de confianza (suponiendo MAS)**, el que varía entre regiones, marcando el menor valor en la Región Metropolitana, con un 2,8%, y el mayor en Los Lagos, con un 5,5%. En la Tabla 12 también se presenta el **error absoluto considerando efecto de diseño**, que resulta menor en la Región de Coquimbo (1,5%) y el más alto en La Araucanía (12,3%).

Tabla 12.
Porcentaje de personas que leen diariamente 15-20 minutos, por región.
Estimación bajo muestreo aleatorio simple y diseño muestral complejo, ECL 2014

Región	Frecuencia de lectura (Dummie)						
	CÁLCULO SUPONIENDO MUESTREO ALEATORIO SIMPLE			CÁLCULO CON VARIABLES DEL DISEÑO			
	Estimación	Margen de error absoluto	Margen de error relativo	Estimación	Margen de error absoluto	Margen de error relativo	Efecto del diseño
1 Tarapacá	60,7	±5,1	±8,4	61,3	±10,1	±16,5	1,71
2 Antofagasta	38,0	±4,7	±12,4	36,1	±7,1	±19,7	1,4
3 Atacama	47,5	±5,5	±11,6	43,3	±10,7	±24,7	1,42
4 Coquimbo	36,4	±5,0	±13,7	39,1	±1,5	±3,9	0,07
5 Valparaíso	51,7	±3,6	±6,9	50,5	±4,7	±9,3	1,62
6 O'Higgins	43,7	±4,9	±11,2	41,4	±2,8	±6,7	0,23
7 Maule	32,7	±4,3	±13,3	32	±4,6	±14,4	0,72
8 Biobío	44,1	±3,5	±7,9	44,8	±3,3	±7,4	0,88
9 La Araucanía	60,3	±5,4	±8,9	60,2	±12,3	±20,5	5,59
10 Los Lagos	45,7	±5,5	±12,1	48,7	±2,1	±4,3	0,12
11 Aysén	48,2	±5,4	±11,2	48,4	±5,1	±10,5	0,11
12 Magallanes	51,1	±5,5	±10,7	50	±8,6	±17,2	0,52
13 Metropolitana	52,8	±2,8	±5,3	51,9	±3,0	±5,7	2,82
14 Los Ríos	44,8	±5,1	±11,4	42,3	±11,5	±27,1	1,61
15 Arica y Parinacota	41,0	±5,4	±13,2	38,1	±2,2	±5,7	0,05
Total	47,1	±1,2	±2,5	48,4	±1,7	±3,5	1,97

7.3.1. Variables dependientes asociadas a prácticas de lectura

Se presentan tres variables dependientes que están asociadas a la prevalencia del nivel de lectura: personas que leen de vez en cuando de 15 a 20 minutos, lectura de al menos un libro en los últimos 12 meses por motivos de ocio o entretención, y lectura de al menos un libro en los últimos 12 meses por motivos de estudio o trabajo.

Las siguientes tablas presentan los márgenes de error (a un 95% de confianza) bajo supuesto de muestreo aleatorio simple (asumiendo varianza máxima y varianza real para variables de interés), y otro utilizando el diseño complejo de la muestra. En la primera tabla (Tabla 13) se presenta la apertura de las tres variables dependientes, segmentadas por las variables de posestratificación (edad, sexo, región, educación del encuestado), mientras que la segunda tabla (Tabla 14) presenta los mismos indicadores pero segmentados por variables analíticas (sexo, educación del jefe de hogar y actividad del encuestado).

Tal como se observa en la Tabla 13, a nivel total el error MAS con varianza máxima es de $\pm 1,2\%$. Este error aumenta para las submuestras regionales y macrozonas, dado el menor número de casos. A nivel de variable, se observa que "lectura de vez en cuando por 15 minutos seguidos" alcanza la mayor precisión, con un margen de error de $\pm 0,9$, mientras que las otras dos se mantienen al nivel del estimador bajo varianza máxima.

Ahora bien, los márgenes de error que consideran las características del diseño muestral, como es de esperar, aumentan en la mayoría de los casos, dado que el efecto de diseño suele ser superior a 1. A nivel de totales, como se observa en la Tabla 8, en las tres variables dependientes el error aumenta al compararlo con el MAS.

La Tabla 8 presenta un ejercicio similar, pero considerando los grupos o segmentos resultantes de las variables analíticas. La comparación por medio de los métodos de cálculo de la varianza se mantiene similar a lo

revisado en la tabla anterior. Dado a que en todas las variables el efecto de diseño es superior a 1, los márgenes de error siempre son mayores si se considera el efecto del diseño muestral al compararlo con el supuesto MAS con varianza real.

Tabla 13.
Márgenes de error bajo diferentes supuestos, por variables sociodemográficas (al 95% de confianza)

Variables	Categorías	Margen de error bajo supuestos MAS y varianza máxima	Margen de error bajo supuestos de muestreo aleatorio simple			Margen de error considerando las características del diseño de la muestra (diseño complejo)			Efecto de diseño (deff)		
			Lee, por más de 15 minutos seguidos	Entretención y ocio	Estudio y trabajo	Lee por más de 15 minutos seguidos	Entretención y ocio	Estudio y trabajo	Lee, por más de 15 minutos seguidos	Entretención y ocio	Estudio y trabajo
Total		±1,2	±0,9	±1,2	±1,2	±1,2	±2,2	±1,6	1,82	3,29	1,84
Edad	9 a 14 años	±3,9	±2,6	±3,9	±1,9	±3,7	±4,0	±1,9	2,16	1,25	1,11
	15 a 24 años	±3,0	±2,0	±3,0	±2,9	±2,1	±5,0	±3,8	1,49	3,57	2,25
	25 a 44 años	±1,9	±1,4	±1,9	±1,9	±2,1	±3,3	±3,4	2,15	2,95	3,24
	45 a 65 años	±1,9	±1,6	±1,9	±1,5	±2,7	±3,3	±2,6	2,27	2,42	2,35
Sexo	Hombre	±1,8	±1,3	±1,8	±1,8	±1,8	±3,0	±2,1	1,84	3,07	1,53
	Mujer	±1,6	±1,2	±1,6	±1,6	±2,0	±2,6	±2,8	2,52	2,43	2,89
Educación del entrevistado	Básica completa o menos	±2,5	±2,3	±2,4	±2,5	±3,3	±3,4	±3,9	2,29	2,21	2,67
	Media completa o incompleta	±1,7	±1,3	±1,7	±1,6	±1,8	±2,4	±3,1	1,64	1,83	3,35
	Superior incompleta o más	±2,2	±1,1	±2,1	±2,1	±1,4	±3,6	±3,1	1,70	3,20	2,21
Región	Tarapacá	±5,2	±3,8	±4,8	±5,2	±7,2	±9,4	±4,7	1,30	1,49	0,31
	Antofagasta	±4,8	±3,6	±4,6	±4,8	±3,3	±4,5	±6,9	0,52	0,61	1,34
	Atacama	±5,5	±4,9	±5,5	±5,4	±8,6	±14,6	±15,7	1,20	2,83	3,39
	Coquimbo	±5,2	±4,4	±5,0	±4,9	±5,9	±11,5	±6,5	1,34	3,99	1,32
	Valparaíso	±3,6	±2,5	±3,6	±3,6	±2,3	±5,5	±5,0	0,80	2,22	1,88
	O'Higgins	±4,9	±4,0	±4,9	±4,5	±2,9	±4,9	±5,0	0,38	0,71	0,88
	Maule	±4,6	±3,4	±4,6	±4,5	±4,5	±12,5	±7,9	1,15	5,19	2,11
	Biobío	±3,5	±2,7	±3,5	±3,5	±4,9	±4,5	±4,7	3,22	1,61	1,78
	La Araucanía	±5,5	±3,8	±5,4	±5,5	±5,9	±5,9	±2,3	2,13	1,07	0,16
	Los Lagos	±5,6	±4,9	±5,2	±5,3	±6,0	±3,1	±2,0	1,21	0,30	0,11
	Aysén	±5,4	±4,4	±5,3	±5,4	±6,4	±4,1	±6,8	0,25	0,07	0,19
	Magallanes	±5,5	±3,2	±5,4	±5,3	±5,7	±9,7	±6,5	0,62	0,68	0,32
	Metropolitana	±2,8	±2,1	±2,8	±2,8	±1,9	±3,8	±2,7	2,22	4,68	2,30
	Los Ríos	±5,1	±3,8	±5,1	±5,1	±10,3	±9,9	±1,6	2,30	1,24	0,03
	Arica y Parinacota	±5,5	±5,0	±5,0	±5,3	±12,1	±9,2	±7,6	1,69	0,95	0,57
Macrozona	Norte	±2,3	±1,9	±2,3	±2,3	±3,1	±5,8	±3,8	1,31	3,29	1,35
	Centro	±2,5	±1,8	±2,5	±2,4	±1,7	±4,2	±3,8	0,70	2,37	1,98
	Sur	±2,0	±1,5	±2,0	±2,0	±2,9	±2,8	±2,5	2,23	1,24	1,02
	Región Metropolitana	±2,8	±2,1	±2,8	±2,8	±1,9	±3,8	±2,7	2,22	4,68	2,30

Tabla 8.
Márgenes de error bajo diferentes supuestos, por variables de análisis (al 95% de confianza)

Variables	Categorías	Margen de error bajo supuestos MAS y varianza máxima	Margen de error bajo supuestos de muestreo aleatorio simple			Margen de error considerando las características del diseño de la muestra (diseño complejo)			Efecto de diseño (deff)		
			Lee por más de 15 minutos seguidos.	Entretención y ocio	Estudio y trabajo	Lee por más de 15 minutos seguidos.	Entretención y ocio	Estudio y trabajo	Lee por más de 15 minutos seguidos.	Entretención y ocio	Estudio y trabajo
Edad Análisis	9 a 17 años	±3,4	±2,3	±3,4	±2,2	±3,7	±4,4	±2,7	3,22	2,06	1,89
	18 a 65 años	±1,2	±1,0	±1,2	±1,2	±1,4	±2,4	±2,0	2,00	3,47	2,61
Educación del jefe de hogar	Básica completa	±2,3	±1,1	±2,2	±2,2	±1,4	±3,3	±2,7	1,49	2,36	1,59
	Media com. o inc.	±1,8	±1,3	±1,8	±1,8	±2,2	±3,2	±2,7	2,53	2,73	1,98
	Superior inc. o más	±2,1	±1,9	±2,1	±1,9	±2,5	±3,6	±2,8	1,71	3,11	2,25
Actividad del encuestado	Trabajando	±1,6	±1,2	±1,6	±1,5	±2,2	±2,2	±1,8	2,96	1,74	1,28
	Estudiando	±2,7	±1,6	±2,7	±1,9	±2,1	±4,2	±2,1	2,20	3,17	1,60
	Realizando labores del hogar	±2,9	±2,5	±2,9	±2,4	±4,0	±3,9	±4,0	2,14	1,62	2,40