



# ENCUESTA NACIONAL DE SALUD, SEXUALIDAD Y GÉNERO (ENSSEX)

## 2022-2023

### Manual Factores de expansión



**Ministerio de Salud**  
**Subsecretaría de Salud Pública**  
**División de Planificación Sanitaria**  
**Departamento de Epidemiología**  
**Mac Iver 541, Santiago**

**Cómo citar este documento**

Ministerio de Salud. Manual factores de expansión. Encuesta Nacional de Salud, Sexualidad y Género 2022-2023. Santiago de Chile; 2022. Disponible en <https://epi.minsal.cl/ensex-contenido-informativo-descargable/>

## Índice

<b>1</b>	<b>INTRODUCCIÓN .....</b>	<b>7</b>
<b>2</b>	<b>METODOLOGÍA .....</b>	<b>8</b>
2.1	Universo, muestra y error .....	8
2.1.1	Universo y población objetivo.....	8
2.1.2	Marco muestral.....	9
2.1.3	Tamaño de la muestra y errores de muestreo.....	11
2.1.4	No respuesta y muestra a seleccionar .....	24
2.2	Diseño muestral .....	25
2.2.1	Estratificación de la muestra.....	25
2.2.2	Etapas de selección .....	27
a)	Primera etapa de selección: comunas .....	27
b)	Segunda etapa de selección: manzanas.....	29
c)	Tercera etapa de selección: viviendas .....	32
d)	Cuarta etapa de selección: personas .....	33
<b>3</b>	<b>CÁLCULO DE FACTORES DE EXPANSIÓN .....</b>	<b>35</b>
3.1	Ajuste por probabilidades de selección .....	36
3.1.1	Ponderador de comuna .....	37
3.1.2	Ponderador de manzanas .....	39
3.1.2.1	Ajuste por omisión de manzanas .....	40
3.1.2.2	Ajuste por elegibilidad de manzanas .....	41
a)	Ajuste por elegibilidad desconocida de manzanas .....	42
b)	Ajuste por no elegibilidad de manzanas .....	43
3.1.2.3	Ponderador de manzanas ajustado .....	43
3.1.3	Ponderador de viviendas .....	43
3.1.3.1	Ajuste por elegibilidad de viviendas .....	46
a)	Ajuste por elegibilidad desconocida de viviendas .....	47
b)	Ajuste por no elegibilidad de viviendas .....	47
3.1.3.2	Ponderador de viviendas ajustado por elegibilidad.....	47
3.2	Ajuste explícitos por no respuesta.....	48
3.2.1	Modelo de no respuesta .....	49
3.2.1.1	Modelamiento y estimación de la probabilidad de responder.....	49
3.2.1.2	Creación de celdas de ajuste de no respuesta.....	51
3.2.2	Cálculo del factor de ajuste de no respuesta de vivienda.....	51
3.2.3	Ponderador de viviendas ajustado por no respuesta .....	52
3.3	Ajuste por probabilidad de selección de personas .....	53

3.3.1	Cálculo de probabilidad de selección de personas .....	53
3.3.2	Ponderador de selección de personas .....	55
3.4	Ajuste a totales poblacionales .....	56
3.4.1	Cálculo de totales poblacionales.....	56
3.4.2	Calibración mediante método raking.....	60
3.4.3	Ponderador de personas calibrado .....	61
3.5	Factor de expansión compensando por no respuesta módulo hogar .....	63
3.5.1	Modelo de no respuesta respecto módulo hogar .....	63
3.5.1.1	Modelamiento y estimación de la probabilidad de responder.....	63
3.5.1.2	Creación de celdas de ajuste de no respuesta.....	65
3.5.2	Cálculo del factor de ajuste de no respuesta de vivienda.....	65
3.5.3	Ponderador de personas ajustado por no respuesta a módulo hogar .....	66
3.5.4	Calibración de ponderador de personas ajustado por no respuesta a módulo hogar	66
3.6	Cálculo de varianzas.....	68
3.6.1	Conformación de variables de estratos y conglomerados.....	68
3.6.2	Ejemplo de cálculo de varianza bajo Diseño Muestral Complejo .....	70
<b>4</b>	<b>BIBLIOGRAFÍA.....</b>	<b>73</b>
<b>5</b>	<b>ANEXOS .....</b>	<b>74</b>
5.1	Estadísticas asociadas a Factores de Expansión.....	74
5.2	Cálculo de prevalencias, efecto diseño y coeficiente de variación.....	95

## Equipo de trabajo

La ENSSEX 2022-2023 fue desarrollada por el Ministerio de Salud, a través del Departamento de Epidemiología y en directa colaboración con la Universidad de Chile, la Universidad Alberto Hurtado y el *Institut National d'Etudes Démographiques* á Paris.

### **JEFE DE DEPARTAMENTO DE EPIDEMIOLOGÍA**

Dr. Christian García Calavaro

### **DIRECTORA DE PROYECTO**

Soc. Pía Álvarez Iturriaga

### **SUBDIRECTORES DE PROYECTO**

Est. Natalia Vergara Mallegas

Soc. José Alarcón Molina

### **INSTITUT NATIONAL D'ETUDES DÉMOGRAPHIQUES Á PARIS.**

Dr. Michel Bozón

### **UNIVERSIDAD DE CHILE**

Dra. Irma Palma Manríquez

### **UNIVERSIDAD ALBERTO HURTADO**

Dr. Jaime Barrientos Delgado

### **OFICINA DE VIGILANCIA DE ENFERMEDADES NO TRASMISIBLES Y ENCUESTAS POBLACIONALES**

Mat. Jessica Moya Cortes

Enf. Mat. Sylvina Alvarado Fick

Hist. Daniela Solís Díaz

Psi. Bert Kaempfe Vasquez

Soc. Rocío Martínez Gutiérrez

Dra. Clelia Vallebuona

**OFICINA DE ESTUDIOS Y ANÁLISIS ESTADÍSTICOS AVANZADOS**

Est. Nora Díaz Sanzana

Est. Tomas Bralic Muñoz

Est. Valentina Navarro Muñoz

Soc. Esteban Navarro Leal

**OFICINA DE VIGILANCIA DE ENFERMEDADES TRASMISIBLES**

Mat. Karen Cáceres Burton

El organismo ejecutor de esta encuesta (Licitación N°757-21-LR21), fue la Dirección de Estudios Sociales del Instituto de Sociología de la Pontificia Universidad Católica de Chile (DESUC), que contó con un equipo de profesionales interdisciplinarios a cargo de

**DIRECTOR DE PROYECTO**

Cristián Ayala Armanet

**COORDINADORAS DE PROYECTO**

Cristina Marchant Araya

Vicky Rojas Araya

**JEFE DE SUPERVISIÓN TERRENO**

Andrés Jaime

## 1 Introducción

El presente informe comprende la metodología utilizada para la construcción de los factores de expansión de Encuesta Nacional de Salud, Sexualidad y Género (ENSSEX 2022-2023) para el Ministerio de Salud.

El informe se compone de cinco apartados, incluyendo esta introducción. En el segundo apartado se abordan las definiciones metodológicas de la encuesta, con énfasis en el diseño muestral de ella. En el segundo apartado se desarrolla en detalle el proceso de construcción de los factores de expansión, dando cuenta del ajuste por probabilidades de selección, los ajustes explícitos de no respuesta, el ajuste por probabilidad de selección de la persona, ajustes a totales, el desarrollo del factor de expansión que compensa por la no respuesta al módulo hogar del cuestionario, y finalmente el cálculo de varianzas.

## 2 Metodología

A continuación, se presentan las principales características metodológicas de la implementación y desarrollo de la Encuesta Nacional de Salud, Sexualidad y Género 2022-2023.

### 2.1 Universo, muestra y error

En esta sección se detalla lo referente al universo, población objetivo, diseño muestral, tamaño de la muestra y errores de muestreo asociados al diseño metodológico de la encuesta.

Atendiendo a los requerimientos del estudio, se consideró un diseño muestral de áreas probabilístico, estratificado geográficamente y multietápico, para un tamaño **muestral a nivel nacional objetivo de 20.000 casos efectivos**, distribuidos apropiadamente en las dieciséis regiones del país.

#### 2.1.1 Universo y población objetivo

El **universo del estudio** se encuentra definido como: hombres y mujeres de 18 años y más, residentes en hogares particulares de áreas urbanas todas las regiones del país.

La población objetivo de la encuesta entonces corresponde a las **personas de 18 años y más**, chilenas o residentes en Chile, que habitan (residen en la vivienda por más de seis meses, no se encuentran en prisión o en casa de reposo, etc.) en viviendas particulares ocupadas localizadas en **zonas urbanas** de las **dieciséis regiones de Chile**.

Quedan excluidas de la población objetivo las personas que residen en áreas clasificadas como “Áreas de Difícil Acceso” (ADA)<sup>1</sup>, por el Instituto Nacional de Estadísticas<sup>2</sup> y DESUC<sup>3</sup>, ubicadas en 22 comunas.

---

<sup>1</sup> Las comunas de difícil acceso definidas por el INE durante el proceso de PreCenso 2011 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

<sup>2</sup> Para más información respecto a las Áreas de Difícil Acceso ver documento: “Definición de áreas especiales para el levantamiento censal”, septiembre 2015, Instituto Nacional de Estadísticas.

<sup>3</sup> A la lista de INE, DESUC agrega las comunas de Camarones, y Putre, debido a la verificación en levantamientos de datos previos de dificultades para acceder a ellas por sus condiciones físico-geográficas (por ejemplo, falta de infraestructura vial, o de medios de transporte públicos y privados constantes que permitan conectar con la zona, dificultades climáticas, entre otras).

## 2.1.2 Marco muestral

Para la ejecución de este diseño muestral se utilizó el marco de manzanas y entidades elaboradas para el Censo 2017<sup>4</sup>. INE puso a disposición en julio de 2018 la cartografía y microdatos a nivel de manzanas urbanas y entidades rurales del Censo 2017. Combinando ambas fuentes se obtiene la ubicación geográfica de manzanas –unidad geográfica básica que para fines estadísticos conforman zonas censales en el área urbana–, número de viviendas y personas que cada una de ellas posee (dato relevante para la selección de manzanas).

El marco fue creado a partir del análisis y actualización de las entidades de población realizadas durante el Precenso 2016 y levantamiento censal 2017, de acuerdo con la división político-administrativa y censal. La población objetivo de este estudio está completamente cubierta por la información disponible a partir del Censo.

Al momento de la selección de manzanas o secciones se excluyeron —por razones operativas—, aquellas con 7 o menos viviendas habitadas (Grupo 0 y Sin viviendas, en Tabla 1). Esta exclusión dejó fuera al 1% de las viviendas catastradas en el marco, pero removió del marco al 9% de sus manzanas. Se estima que esa pérdida de cobertura en el marco es compensada por una mayor eficiencia en los recursos de terreno.

**Tabla 1 Número de manzanas y entidades, viviendas y personas en el Censo 2017 según grupo de tamaño.**

Grupo de tamaño	Rango de total de viviendas	Total, de manzanas y entidades	Dist. de manzanas y entidades	Total, de Viviendas	Dist. de viviendas	Total, de Personas	Dist. de personas
	1	2	3	4	5	6	7
Total		180.499	100%	6.499.355	100%	17.574.003	100%
Sin viviendas	0	2	0%	0	0%	439	0%
Grupo 0	1 a 7	16.242	9%	69.492	1%	252.079	1%
Grupo 1 a 4	8 a 23	79.762	44%	1.249.628	19%	3.565.723	20%
Grupo 5 a 9	24 a 44	53.091	29%	1.665.145	26%	4.889.884	28%

<sup>4</sup> Ver Base Cartográfica Censal Junio 2018, INE

Grupo 10 a 19	45 a 81	18.060	10%	1.046.489	16%	2.907.868	17%
Grupo 20 a 28	82 a 154	8.003	4%	882.281	14%	2.320.225	13%
Grupo 29 y 30	155 y más	5.339	3%	1.586.320	24%	3.637.785	21%

Fuente: Elaboración propia a partir de Censo 2017

Una vez seleccionada la muestra de manzanas/secciones, se realizó una **actualización del total de viviendas seleccionadas en la muestra**. Este procedimiento, denominado “empadronamiento” (*listing*, en inglés), busca actualizar el total de viviendas registrado en el marco muestral a partir de las observaciones realizadas al momento del trabajo de campo. En concreto, el proceso de empadronamiento se realizó de la siguiente manera:

1. Ubicación del empadronador en la esquina nor–poniente de la manzana seleccionada.
2. Recorrido en sentido de las agujas del reloj de toda la manzana seleccionada (misma modalidad de recorrido utilizado en Censos).
3. Listado del total de edificaciones encontradas durante el recorrido de la manzana.
4. Asignación de Código de Disposición de Vivienda (que resume la naturaleza de la edificación en la manzana)
5. Identificación de viviendas elegibles (viviendas particulares ocupadas).

Además, durante este proceso de empadronamiento de manzanas se catastraron algunas características de estas que podrán utilizarse posteriormente para el ajuste explícito de no respuesta en la elaboración de factores de expansión (ver detalles en capítulo 3 del presente informe). Ejemplo de estas características son:

- Estado de edificios y casas de la manzana
- Presencia de basura en la manzana
- Presencia de vandalismo, grafitis o daño deliberado a la propiedad en el sector

De esta forma cambios derivados de nuevas construcciones, demoliciones, cambio de uso de las edificaciones y otros quedan incorporados en el estudio de dos formas:

- Las nuevas construcciones (en manzanas/secciones seleccionadas) tienen probabilidad mayor a cero de ser seleccionadas.
- El factor de expansión a través del ajuste de las probabilidades de selección de las viviendas incorpora el cambio en los pesos de las unidades seleccionadas para

incorporar los crecimientos/decrecimientos entre los registros desactualizados del marco y los registros actualizados del empadronamiento.

A su vez, en paralelo al proceso de empadronamiento, este fue supervisado y digitado en un **100%** en las oficinas de DESUC.

### 2.1.3 Tamaño de la muestra y errores de muestreo

La Tabla 2 presenta la distribución de la población objetivo por región, según las proyecciones poblacionales al año 2021<sup>5</sup> publicadas por INE con base 2017. En las primeras dos columnas se puede ver el total y la distribución de toda la población residente en áreas urbanas, que alcanzan los 17.430.714 personas. Centrándonos en la población objetivo, en la columna #3 se ve que las personas mayores de 18 años habitando en áreas urbanas son 13.439.667. Respecto a la distribución de dicha población, en la última columna se ve que se distribuye de manera muy similar a la de toda la población urbana. Así, el 46% se concentra en la región Metropolitana, y un 11% en la región de Valparaíso.

---

<sup>5</sup> Notar que se presentan totales poblacionales al año 2021 y no 2022, pues la selección de la muestra se realizó considerando los datos poblacionales del año 2021.

**Tabla 2 Distribución de la población objetivo del estudio:  
personas mayores de 18 años residentes de zonas urbanas de Chile, según región.**

Región	Personas habitando en áreas urbanas		Personas mayores de 18 años y más en áreas urbanas	
	Población	% col	Población	% col
	1	2	3	4
1 Tarapacá	371.788	2%	270.308	2%
2 Antofagasta	680.559	4%	508.680	4%
3 Atacama	289.760	2%	213.064	2%
4 Coquimbo	695.749	4%	523.772	4%
5 Valparaíso	1.814.427	10%	1.417.150	11%
6 O'Higgins	754.522	4%	572.789	4%
7 Maule	850.678	5%	646.246	5%
8 Biobío	1.485.892	9%	1.146.717	9%
9 La Araucanía	731.837	4%	556.150	4%
10 Los Lagos	668.895	4%	507.012	4%
11 Aysén	88.130	1%	64.290	0%
12 Magallanes	165.904	1%	129.261	1%
13 Metropolitana	7.937.816	46%	6.199.548	46%
14 Los Ríos	296.130	2%	228.753	2%
15 Arica y Parinacota	234.536	1%	176.290	1%
16 Ñuble	364.091	2%	279.637	2%
<b>Total</b>	<b>17.430.714</b>	<b>100%</b>	<b>13.439.667</b>	<b>100,0%</b>

Nota: Los totales poblacionales incluyen la población en las Áreas de Difícil Acceso.

Fuente: Proyecciones de población al año 2021 publicadas por INE con base 2017.

Para la distribución muestral se realizó una **distribución proporcional** de los casos, en tanto esto permite **optimizar los errores muestrales** a nivel regional.

Ello quiere decir que la distribución de casos objetivos de la muestra no responde exactamente a la proporción de la población en cada región. Así, en algunas regiones se obtendrá una muestra menor a la que correspondería según su peso poblacional, y en otras, por el contrario, se obtendrá una muestra mayor respecto a la que le correspondería si se sigue la distribución real de personas en dichas regiones.

En la Tabla 3 se muestran los tamaños muestrales objetivo de tres escenarios, junto con sus errores absolutos y relativos calculados bajo un supuesto de Muestreo Aleatorio Simple, a 95% nivel de confianza, con una prevalencia de 50% (varianza máxima). En concreto, para realizar estos cálculos se utilizaron las siguientes fórmulas según lo definido por Lohr (1999)<sup>6</sup>:

$$e_{abs} = Z \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$

$$e_{rel} = \frac{e_{abs}}{p}$$

Donde,

- Z : Nivel de confianza según distribución Z (95%)
- p : Prevalencia estimada
- n : Tamaño muestral
- N : Tamaño poblacional

Respecto a los tamaños muestrales, en todos los escenarios se consideró una muestra objetivo de 20.000 casos, pero la distribución de estos en regiones es diversa.

En el primer escenario, se tiene una distribución bastante proporcional de los casos, aunque con ajustes leves para lograr un mínimo de casos en regiones con baja proporción de población para lograr reducir los errores muestrales. En el segundo escenario, se tiene una distribución mucho menos proporcional, manteniendo tamaños muestrales por región relativamente similares, y con ello, manteniendo errores muestrales bastante análogos en todas las regiones. En el tercer escenario, se tiene una propuesta intermedia, con una distribución más proporcional de los casos, pero aumentando los mínimos regionales, y disminuyendo el máximo regional (en la región Metropolitana).

---

<sup>6</sup> Cabe señalar que la fórmula utilizada corresponde al cálculo de error muestral absoluto para poblaciones finitas, que corresponde a tamaños poblacionales menores a 100.000. Se utiliza esta fórmula porque la mayoría de las subpoblaciones se cumple con tener una población mayor a 100.000. Sin embargo, y aunque la región de Aysén tiene un tamaño poblacional menor a ese umbral, por parsimonia, se utiliza la misma fórmula para el cálculo de errores en todas las subpoblaciones.

En cuanto a los errores muestrales, al compartir el mismo tamaño muestral, los tres escenarios alcanzan un error absoluto de 0,7 y relativo de 1,4 puntos. Al analizar los errores absolutos por región, en el primer escenario el máximo error es de 4,9 puntos en las regiones con 400 casos. En el escenario 2 el máximo es de 3,4 puntos en la región de Magallanes. En el tercer escenario el mayor error absoluto se encuentra en la región de Aysén con 4,4 puntos.

En relación con los errores por sexo, todos los escenarios cuentan con un error absoluto de 1 punto y errores relativos cercanos a los 2 puntos. Respecto a los tramos etarios, también se comparten los mismos errores entre los tres escenarios, teniendo un máximo de error absoluto estimado de 4,2 puntos en el rango etario de 75 a 79 años.

Cabe destacar que, en ninguno de los escenarios, y en ninguna subpoblación se estimaron errores relativos superiores a los 20 puntos.

Teniendo estos elementos en consideración, es que en conjunto con la contraparte técnica **se acordó utilizar el escenario muestral 3 como abordaje para el levantamiento de los datos.** En específico, este acuerdo se llega en la reunión del día 5 de octubre de 2021, según consta en acta de la misma fecha.

Independiente de ello, se debe mencionar que - de todas formas -, la distribución apropiada de los casos del escenario se corrige mediante ajustes de postmuestreo. En concreto, se elaboró un factor de expansión que mediante el método de postestratificación ajusta las distribuciones y tamaños de la muestra a los estándares poblacionales para dar cuenta de la distribución real de la población en las regiones, tramos etarios y sexo.

De manera adicional, en la Tabla 4 y Tabla 5 se muestran los errores de muestreo absolutos y relativos bajo supuesto de Muestreo Aleatorio Simple, pero con prevalencias del 30% y 70% respectivamente.

**Tabla 3 Tamaños muestrales de los Escenarios 1,2 y 3 con errores absolutos y errores relativos al 95% nivel de confianza bajo Muestreo Aleatorio Simple y prevalencia de 50% según subpoblaciones de interés**

Sub poblaciones	Población	% Prevalencia	Prevalencia p=50%	Escenario 1			Escenario 2			Escenario 3		
				Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS
	1	2	3	4	5	6	7	8	9	10	11	12
País	13.439.667	100%	0,5	<b>20.000</b>	0,7	1,4	<b>20.000</b>	0,7	1,4	<b>20.000</b>	0,7	1,4
Tarapacá	270.308	2%	0,5	<b>450</b>	4,6	9,2	<b>930</b>	3,2	6,4	<b>580</b>	4,1	8,1
Antofagasta	508.680	4%	0,5	<b>720</b>	3,6	7,3	<b>1.080</b>	3,0	6,0	<b>880</b>	3,3	6,6
Atacama	213.064	2%	0,5	<b>400</b>	4,9	9,8	<b>1.040</b>	3,0	6,1	<b>520</b>	4,3	8,6
Coquimbo	523.772	4%	0,5	<b>910</b>	3,2	6,5	<b>1.310</b>	2,7	5,4	<b>1.150</b>	2,9	5,8
Valparaíso	1.417.150	11%	0,5	<b>2.130</b>	2,1	4,2	<b>1.330</b>	2,7	5,4	<b>1.640</b>	2,4	4,8
O'Higgins	572.789	4%	0,5	<b>1.070</b>	3,0	6,0	<b>1.020</b>	3,1	6,1	<b>1.410</b>	2,6	5,2
Maule	646.246	5%	0,5	<b>1.210</b>	2,8	5,6	<b>1.210</b>	2,8	5,6	<b>1.410</b>	2,6	5,2
Biobío	1.146.717	9%	0,5	<b>1.840</b>	2,3	4,6	<b>1.280</b>	2,7	5,5	<b>1.750</b>	2,3	4,7
La Araucanía	556.150	4%	0,5	<b>1.060</b>	3,0	6,0	<b>1.190</b>	2,8	5,7	<b>1.580</b>	2,5	4,9
Los Lagos	507.012	4%	0,5	<b>990</b>	3,1	6,2	<b>1.150</b>	2,9	5,8	<b>1.450</b>	2,6	5,1
Aysén	64.290	1%	0,5	<b>400</b>	4,9	9,8	<b>920</b>	3,2	6,4	<b>490</b>	4,4	8,8
Magallanes	129.261	1%	0,5	<b>400</b>	4,9	9,8	<b>840</b>	3,4	6,7	<b>510</b>	4,3	8,7
Metropolitana	6.199.548	45%	0,5	<b>6.790</b>	1,2	2,4	<b>3.870</b>	1,6	3,1	<b>4.380</b>	1,5	3,0
Los Ríos	228.753	2%	0,5	<b>560</b>	4,1	8,3	<b>980</b>	3,1	6,2	<b>650</b>	3,8	7,7

Sub poblaciones	Población	% Prevalencia	Prevalencia p=50%	Escenario 1			Escenario 2			Escenario 3		
				Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS
	1	2	3	4	5	6	7	8	9	10	11	12
Arica y Parinacota	176.290	1%	0,5	<b>400</b>	4,9	9,8	<b>870</b>	3,3	6,6	<b>550</b>	4,2	8,3
Ñuble	279.637	2%	0,5	<b>670</b>	3,8	7,6	<b>980</b>	3,1	6,2	<b>1.050</b>	3,0	6,0
Hombre	6.496.501	48%	0,5	<b>9.668</b>	1,0	2,0	<b>9.668</b>	1,0	2,0	<b>9.668</b>	1,0	2,0
Mujer	6.943.166	52%	0,5	<b>10.332</b>	1,0	1,9	<b>10.332</b>	1,0	1,9	<b>10.332</b>	1,0	1,9
18-19	460.507	4%	0,5	<b>685</b>	3,7	7,5	<b>685</b>	3,7	7,5	<b>685</b>	3,7	7,5
20-24	1.289.993	11%	0,5	<b>1.920</b>	2,2	4,5	<b>1.920</b>	2,2	4,5	<b>1.920</b>	2,2	4,5
25-29	1.489.525	11%	0,5	<b>2.217</b>	2,1	4,2	<b>2.217</b>	2,1	4,2	<b>2.217</b>	2,1	4,2
30-34	1.505.509	10%	0,5	<b>2.240</b>	2,1	4,1	<b>2.240</b>	2,1	4,1	<b>2.240</b>	2,1	4,1
35-39	1.310.784	9%	0,5	<b>1.951</b>	2,2	4,4	<b>1.951</b>	2,2	4,4	<b>1.951</b>	2,2	4,4
40-44	1.204.573	9%	0,5	<b>1.793</b>	2,3	4,6	<b>1.793</b>	2,3	4,6	<b>1.793</b>	2,3	4,6

Sub poblaciones	Población	% Prevalencia p=50%	Escenario 1			Escenario 2			Escenario 3			
			Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	
	1	2	3	4	5	6	7	8	9	10	11	12
45-49	1.130.166	9%	0,5	<b>1.682</b>	2,4	4,8	<b>1.682</b>	2,4	4,8	<b>1.682</b>	2,4	4,8
50-54	1.063.286	9%	0,5	<b>1.582</b>	2,5	4,9	<b>1.582</b>	2,5	4,9	<b>1.582</b>	2,5	4,9
55-59	1.009.793	8%	0,5	<b>1.503</b>	2,5	5,1	<b>1.503</b>	2,5	5,1	<b>1.503</b>	2,5	5,1
60-64	875.019	6%	0,5	<b>1.302</b>	2,7	5,4	<b>1.302</b>	2,7	5,4	<b>1.302</b>	2,7	5,4
65-69	723.764	5%	0,5	<b>1.077</b>	3,0	6,0	<b>1.077</b>	3,0	6,0	<b>1.077</b>	3,0	6,0
70-74	524.494	4%	0,5	<b>781</b>	3,5	7,0	<b>781</b>	3,5	7,0	<b>781</b>	3,5	7,0
75-79	365.956	3%	0,5	<b>545</b>	4,2	8,4	<b>545</b>	4,2	8,4	<b>545</b>	4,2	8,4
80 y más	486.298	3%	0,5	<b>724</b>	3,6	7,3	<b>724</b>	3,6	7,3	<b>724</b>	3,6	7,3

Nota: los tamaños muestrales para sexo y tramos etarios se estiman en función de la distribución de la muestra que se espera tomando como punto base la distribución de dichas variables en la población objetivo. Por lo tanto, dichas estimaciones pueden variar con la muestra alcanzada finalmente.

**Tabla 4 Tamaños muestrales de los Escenarios 1,2 y 3 con errores absolutos y errores relativos al 95% nivel de confianza bajo Muestreo Aleatorio Simple y prevalencia de 30% según subpoblaciones de interés**

Sub poblaciones	Población	% Prevalencia	Prevalencia p=30%	Escenario 1			Escenario 2			Escenario 3		
				Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS
	1	2	3	4	5	6	7	8	9	10	11	12
País	13.439.667	100%	0,3	<b>20.000</b>	0,6	2,1	<b>20.000</b>	0,6	2,1	<b>20.000</b>	0,6	2,1
Tarapacá	270.308	2%	0,3	<b>450</b>	4,2	14,1	<b>930</b>	2,9	9,8	<b>580</b>	3,7	12,4
Antofagasta	508.680	4%	0,3	<b>720</b>	3,3	11,1	<b>1.080</b>	2,7	9,1	<b>880</b>	3,0	10,1
Atacama	213.064	2%	0,3	<b>400</b>	4,5	15,0	<b>1.040</b>	2,8	9,3	<b>520</b>	3,9	13,1
Coquimbo	523.772	4%	0,3	<b>910</b>	3,0	9,9	<b>1.310</b>	2,5	8,3	<b>1.150</b>	2,6	8,8
Valparaíso	1.417.150	11%	0,3	<b>2.130</b>	1,9	6,5	<b>1.330</b>	2,5	8,2	<b>1.640</b>	2,2	7,4
O'Higgins	572.789	4%	0,3	<b>1.070</b>	2,7	9,1	<b>1.020</b>	2,8	9,4	<b>1.410</b>	2,4	8,0
Maule	646.246	5%	0,3	<b>1.210</b>	2,6	8,6	<b>1.210</b>	2,6	8,6	<b>1.410</b>	2,4	8,0
Biobío	1.146.717	9%	0,3	<b>1.840</b>	2,1	7,0	<b>1.280</b>	2,5	8,4	<b>1.750</b>	2,1	7,2
La Araucanía	556.150	4%	0,3	<b>1.060</b>	2,8	9,2	<b>1.190</b>	2,6	8,7	<b>1.580</b>	2,3	7,5
Los Lagos	507.012	4%	0,3	<b>990</b>	2,9	9,5	<b>1.150</b>	2,6	8,8	<b>1.450</b>	2,4	7,9
Aysén	64.290	1%	0,3	<b>400</b>	4,5	14,9	<b>920</b>	2,9	9,8	<b>490</b>	4,0	13,5
Magallanes	129.261	1%	0,3	<b>400</b>	4,5	14,9	<b>840</b>	3,1	10,3	<b>510</b>	4,0	13,2
Metropolitana	6.199.548	45%	0,3	<b>6.790</b>	1,1	3,6	<b>3.870</b>	1,4	4,8	<b>4.380</b>	1,4	4,5
Los Ríos	228.753	2%	0,3	<b>560</b>	3,8	12,6	<b>980</b>	2,9	9,5	<b>650</b>	3,5	11,7

Sub poblaciones	Población	% Prevalencia p=30%	Escenario 1			Escenario 2			Escenario 3			
			Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	
	1	2	3	4	5	6	7	8	9	10	11	12
Arica y Parinacota	176.290	1%	0,3	<b>400</b>	4,5	15,0	<b>870</b>	3,0	10,1	<b>550</b>	3,8	12,7
Ñuble	279.637	2%	0,3	<b>670</b>	3,5	11,6	<b>980</b>	2,9	9,5	<b>1.050</b>	2,8	9,2
Hombre	6.496.50											
	1	48%	0,3	<b>9.668</b>	0,9	3,0	<b>9.668</b>	0,9	3,0	<b>9.668</b>	0,9	3,0
Mujer	6.943.16											
	6	52%	0,3	<b>10.332</b>	0,9	2,9	<b>10.332</b>	0,9	2,9	<b>10.332</b>	0,9	2,9
18-19	460.507	4%	0,3	<b>685</b>	3,4	11,4	<b>685</b>	3,4	11,4	<b>685</b>	3,4	11,4
20-24	1.289.99											
	3	11%	0,3	<b>1.920</b>	2,0	6,8	<b>1.920</b>	2,0	6,8	<b>1.920</b>	2,0	6,8
25-29	1.489.52											
	5	11%	0,3	<b>2.217</b>	1,9	6,4	<b>2.217</b>	1,9	6,4	<b>2.217</b>	1,9	6,4
30-34	1.505.50											
	9	10%	0,3	<b>2.240</b>	1,9	6,3	<b>2.240</b>	1,9	6,3	<b>2.240</b>	1,9	6,3
35-39	1.310.78											
	4	9%	0,3	<b>1.951</b>	2,0	6,8	<b>1.951</b>	2,0	6,8	<b>1.951</b>	2,0	6,8
40-44	1.204.57											
	3	9%	0,3	<b>1.793</b>	2,1	7,1	<b>1.793</b>	2,1	7,1	<b>1.793</b>	2,1	7,1

Sub poblaciones	Población	% Prevalencia	Prevalencia p=30%	Escenario 1			Escenario 2			Escenario 3		
				Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS
	1	2	3	4	5	6	7	8	9	10	11	12
45-49	1.130.16											
	6	9%	0,3	<b>1.682</b>	2,2	7,3	<b>1.682</b>	2,2	7,3	<b>1.682</b>	2,2	7,3
50-54	1.063.28											
	6	9%	0,3	<b>1.582</b>	2,3	7,5	<b>1.582</b>	2,3	7,5	<b>1.582</b>	2,3	7,5
55-59	1.009.79											
	3	8%	0,3	<b>1.503</b>	2,3	7,7	<b>1.503</b>	2,3	7,7	<b>1.503</b>	2,3	7,7
60-64	875.019	6%	0,3	<b>1.302</b>	2,5	8,3	<b>1.302</b>	2,5	8,3	<b>1.302</b>	2,5	8,3
65-69	723.764	5%	0,3	<b>1.077</b>	2,7	9,1	<b>1.077</b>	2,7	9,1	<b>1.077</b>	2,7	9,1
70-74	524.494	4%	0,3	<b>781</b>	3,2	10,7	<b>781</b>	3,2	10,7	<b>781</b>	3,2	10,7
75-79	365.956	3%	0,3	<b>545</b>	3,8	12,8	<b>545</b>	3,8	12,8	<b>545</b>	3,8	12,8
80 y más	486.298	3%	0,3	<b>724</b>	3,3	11,1	<b>724</b>	3,3	11,1	<b>724</b>	3,3	11,1

Nota: los tamaños muestrales para sexo y tramos etarios se estiman en función de la distribución de la muestra que se espera tomando como punto base la distribución de dichas variables en la población objetivo. Por lo tanto, dichas estimaciones pueden variar con la muestra alcanzada finalmente.

**Tabla 5 Tamaños muestrales de los Escenarios 1,2 y 3 con errores absolutos y errores relativos al 95% nivel de confianza bajo Muestreo Aleatorio Simple y prevalencia de 70% según subpoblaciones de interés**

Sub poblaciones	Población	% Prevalencia	Prevalencia p=70%	Escenario 1			Escenario 2			Escenario 3		
				Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS
				4	5	6	7	8	9	10	11	12
País	13.439.667	100%	0,7	<b>20.000</b>	0,6	0,9	<b>20.000</b>	0,6	0,9	<b>20.000</b>	0,6	0,9
Tarapacá	270.308	2%	0,7	<b>450</b>	4,2	6,0	<b>930</b>	2,9	4,2	<b>580</b>	3,7	5,3
Antofagasta	508.680	4%	0,7	<b>720</b>	3,3	4,8	<b>1.080</b>	2,7	3,9	<b>880</b>	3,0	4,3
Atacama	213.064	2%	0,7	<b>400</b>	4,5	6,4	<b>1.040</b>	2,8	4,0	<b>520</b>	3,9	5,6
Coquimbo	523.772	4%	0,7	<b>910</b>	3,0	4,2	<b>1.310</b>	2,5	3,5	<b>1.150</b>	2,6	3,8
Valparaíso	1.417.150	11%	0,7	<b>2.130</b>	1,9	2,8	<b>1.330</b>	2,5	3,5	<b>1.640</b>	2,2	3,2
O'Higgins	572.789	4%	0,7	<b>1.070</b>	2,7	3,9	<b>1.020</b>	2,8	4,0	<b>1.410</b>	2,4	3,4
Maule	646.246	5%	0,7	<b>1.210</b>	2,6	3,7	<b>1.210</b>	2,6	3,7	<b>1.410</b>	2,4	3,4
Biobío	1.146.717	9%	0,7	<b>1.840</b>	2,1	3,0	<b>1.280</b>	2,5	3,6	<b>1.750</b>	2,1	3,1
La Araucanía	556.150	4%	0,7	<b>1.060</b>	2,8	3,9	<b>1.190</b>	2,6	3,7	<b>1.580</b>	2,3	3,2
Los Lagos	507.012	4%	0,7	<b>990</b>	2,9	4,1	<b>1.150</b>	2,6	3,8	<b>1.450</b>	2,4	3,4
Aysén	64.290	0%	0,7	<b>400</b>	4,5	6,4	<b>920</b>	2,9	4,2	<b>490</b>	4,0	5,8
Magallanes	129.261	1%	0,7	<b>400</b>	4,5	6,4	<b>840</b>	3,1	4,4	<b>510</b>	4,0	5,7
Metropolitana	6.199.548	46%	0,7	<b>6.790</b>	1,1	1,6	<b>3.870</b>	1,4	2,1	<b>4.380</b>	1,4	1,9
Los Ríos	228.753	2%	0,7	<b>560</b>	3,8	5,4	<b>980</b>	2,9	4,1	<b>650</b>	3,5	5,0
Arica y Parinacota	176.290	1%	0,7	<b>400</b>	4,5	6,4	<b>870</b>	3,0	4,3	<b>550</b>	3,8	5,5
Ñuble	279.637	2%	0,7	<b>670</b>	3,5	5,0	<b>980</b>	2,9	4,1	<b>1.050</b>	2,8	4,0

Sub poblaciones	Población	% Prevalencia	Prevalencia p=70%	Escenario 1			Escenario 2			Escenario 3		
				Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS
	1	2	3	4	5	6	7	8	9	10	11	12
Hombre	6.496.501	48%	0,7	<b>9.668</b>	0,9	1,3	<b>9.668</b>	0,9	1,3	<b>9.668</b>	0,9	1,3
Mujer	6.943.166	52%	0,7	<b>10.332</b>	0,9	1,3	<b>10.332</b>	0,9	1,3	<b>10.332</b>	0,9	1,3
18-19	460.507	3%	0,7	<b>685</b>	3,4	4,9	<b>685</b>	3,4	4,9	<b>685</b>	3,4	4,9
20-24	1.289.993	10%	0,7	<b>1.920</b>	2,0	2,9	<b>1.920</b>	2,0	2,9	<b>1.920</b>	2,0	2,9
25-29	1.489.525	11%	0,7	<b>2.217</b>	1,9	2,7	<b>2.217</b>	1,9	2,7	<b>2.217</b>	1,9	2,7
30-34	1.505.509	11%	0,7	<b>2.240</b>	1,9	2,7	<b>2.240</b>	1,9	2,7	<b>2.240</b>	1,9	2,7
35-39	1.310.784	10%	0,7	<b>1.951</b>	2,0	2,9	<b>1.951</b>	2,0	2,9	<b>1.951</b>	2,0	2,9
40-44	1.204.573	9%	0,7	<b>1.793</b>	2,1	3,0	<b>1.793</b>	2,1	3,0	<b>1.793</b>	2,1	3,0
45-49	1.130.166	8%	0,7	<b>1.682</b>	2,2	3,1	<b>1.682</b>	2,2	3,1	<b>1.682</b>	2,2	3,1

Sub poblaciones	Población	% Prevalencia p=70%	Escenario 1			Escenario 2			Escenario 3			
			Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	Tamaño muestra objetivo	Error absoluto 95% MAS	Error relativo 95% MAS	
	1	2	3	4	5	6	7	8	9	10	11	12
50-54	1.063.286	8%	0,7	<b>1.582</b>	2,3	3,2	<b>1.582</b>	2,3	3,2	<b>1.582</b>	2,3	3,2
55-59	1.009.793	8%	0,7	<b>1.503</b>	2,3	3,3	<b>1.503</b>	2,3	3,3	<b>1.503</b>	2,3	3,3
60-64	875.019	7%	0,7	<b>1.302</b>	2,5	3,6	<b>1.302</b>	2,5	3,6	<b>1.302</b>	2,5	3,6
65-69	723.764	5%	0,7	<b>1.077</b>	2,7	3,9	<b>1.077</b>	2,7	3,9	<b>1.077</b>	2,7	3,9
70-74	524.494	4%	0,7	<b>781</b>	3,2	4,6	<b>781</b>	3,2	4,6	<b>781</b>	3,2	4,6
75-79	365.956	3%	0,7	<b>545</b>	3,8	5,5	<b>545</b>	3,8	5,5	<b>545</b>	3,8	5,5
80 y más	486.298	4%	0,7	<b>724</b>	3,3	4,8	<b>724</b>	3,3	4,8	<b>724</b>	3,3	4,8

Nota: los tamaños muestrales para sexo y tramos etarios se estiman en función de la distribución de la muestra que se espera tomando como punto base la distribución de dichas variables en la población objetivo. Por lo tanto, dichas estimaciones pueden variar con la muestra alcanzada finalmente.

#### 2.1.4 No respuesta y muestra a seleccionar

Para la selección del tamaño de la muestra objetivo se debió considerar el **fenómeno de la no respuesta** y la tasa de no respuesta asociada.

La tasa de no respuesta corresponde al porcentaje de direcciones en donde **no fue posible** concretar una entrevista, ya sea por dificultades para localizar a los residentes, rechazo de los residentes a participar, encontrarse con un local comercial en vez de una vivienda particular, encontrarse con un sitio eriazado en vez de una vivienda particular, etc.

La no respuesta es normal y esperable dentro del trabajo de encuestas. Siguiendo las mejores prácticas (Valliant, Dever, & Kreuter, 2013), el DESUC utiliza la **técnica de sobremuestreo** para administrar la no respuesta esperada de la encuesta. Esto quiere decir que se seleccionó ex ante, y de una sola vez, una muestra ampliada que considere una pérdida esperada.

Para este estudio se estimó que la tasa de no respuesta a nivel país será aproximadamente de **50%**, que corresponde a una tasa superior a la encontrada en estudios de DESUC con similares características en años anteriores. Se realizó esta corrección por la situación sanitaria en la que sería implementado el estudio y que ha afectado negativamente estudios en terreno que se han efectuado durante 2020 y 2021.

La estimación de participación es imprescindible para determinar el número final de casos que se seleccionarán desde el marco. De antemano se extrajeron una mayor cantidad de casos asumiendo que el 50% de estos no se traducirán en encuestas completas.

Esta técnica le pone, de facto, una cota a la posibilidad de reemplazos de la muestra ya que no se permite la selección de muestras adicionales de reemplazo, lo cual **asegura mantener la aleatoriedad y trazabilidad de la muestra**. Si la manzana no es residencial, si no se establece contacto con la vivienda o es rechazado su ingreso, o bien la persona elegida no responde, se registra como manzana inválida, no respuesta de la vivienda seleccionada, generando así mayor transparencia y rigor en el estado final de cada uno de los casos seleccionados.

En la Tabla 6 se puede ver la distribución muestral objetivo en el escenario 3 con 20.000 casos entre 16 estratos considerando las regiones del país. Para lograr la muestra objetivo, entonces, se debieron seleccionar 40.000 viviendas siguiendo la distribución de la población objetivo en el territorio. Esto implica que —dado que se seleccionan 10 viviendas en cada manzana— se seleccionaron de antemano aproximadamente 4.000 manzanas urbanas.

**Tabla 6 Distribución de muestra objetivo y sobredimensionada.**

Región	Escenario 3	
	Muestra objetivo	Casos sobredimensionados
	5	6
1 Tarapacá	580	1.160
2 Antofagasta	880	1.760
3 Atacama	520	1.040
4 Coquimbo	1.150	2.300
5 Valparaíso	1.640	3.280
6 O'Higgins	1.410	2.820
7 Maule	1.410	2.820
8 Biobío	1.750	3.500
9 La Araucanía	1.580	3.160
10 Los Lagos	1.450	2.900
11 Aysén	490	980
12 Magallanes	510	1.020
13 Metropolitana	4.380	8.760
14 Los Ríos	650	1.300
15 Arica y Parinacota	550	1.100
16 Ñuble	1.050	2.100
<b>Total</b>	<b>20.000</b>	<b>40.000</b>

## 2.2 Diseño muestral

El diseño muestral la encuesta 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 etapa, y posteriormente, se seleccionan **viviendas ocupadas** en una tercera etapa y **personas mayores de 18 años** en una cuarta etapa, asegurando así **aleatoriedad total en el proceso**.

A continuación, se describen en detalle la metodología implementada para la selección de la muestra.

### 2.2.1 Estratificación de la muestra

La estratificación de una muestra corresponde a la agrupación de los elementos de una población previo a la selección de ella. Uno de los motivos para utilizar un diseño estratificado (ver capítulo 4 en Lohr, 1999) es que éste mejora la eficiencia del muestreo para las

*Departamento de Epidemiología – Oficina de Vigilancia de Enfermedades No Transmisibles y Encuestas Poblacionales*

estimaciones de la población completa, al agrupar unidades que tienen propiedades similares de media y varianza. La estratificación por tamaño, con una distribución eficiente de casos entre estratos, es un ejemplo de lo señalado.

Este método utiliza una variable de tamaño que esté correlacionada con la variable de interés de la encuesta. Para una discusión en más detalle ver Valliant et al. (2013, cap 3) y Valliant et al. (2000, cap. 6). Así, se estratificó la muestra en dos instancias, con la finalidad de atender dos objetivos distintos.

Por una parte, se fijó como objetivo distribuir la muestra de manera aporportional entre ciertos dominios geográficos de regiones en Chile. Y, además, el diseño muestral también consideró la conformación de estratos de muestreo dentro de cada región que considerara comunas grandes (más de 50.000 habitantes), comunas medianas (entre 50.000 y 20.000 habitantes), y comunas pequeñas (menos de 20.000 habitantes).

Así estos estratos, también denominados “dominios de muestreo” (*sampling domains*), se conformaron entonces tomando como base las 16 regiones del país, además de los estratos por tamaño en su interior. Adicional a ello, se definieron ciertas comunas como relevantes o de Inclusión Forzosa (capitales regionales, comunas del Gran Santiago, o de gran tamaño), que se incluyeron con certeza en la muestra. Teniendo todo ello en consideración, se construyeron un total de **108 estratos de muestreo**, dentro de los cuales se incluyen comunas IF (de Inclusión Forzosa).

Por otra parte, se estratificaron implícitamente las manzanas a seleccionar ordenándolas – dentro de los estratos antes planteados— según la población que contenían<sup>7</sup> y seleccionarlas de acuerdo con un muestreo sistemático con probabilidad proporcional al tamaño. Esto con el objetivo de lograr homogeneizar la probabilidad de selección final de las personas participantes en el estudio.

---

<sup>7</sup> Este dato proviene desde el Marco Muestral utilizado para la selección de la muestra.  
Departamento de Epidemiología – Oficina de Vigilancia de Enfermedades No Transmisibles y Encuestas Poblacionales

### 2.2.2 Etapas de selección

A continuación, se describirán las distintas etapas de selección de la muestra.

#### a) Primera etapa de selección: comunas

Las Unidades Primarias de Muestreo (UPM) fueron las comunas. En concreto, dentro de los estratos de muestreo las comunas —que no fueran de inclusión forzosa— se escogieron con probabilidad de selección proporcional a su tamaño con método aleatorio (no muestreo sistemático), el cual fue medido por la cantidad de personas mayores de 18 años que habitaran en ellas.

La probabilidad de selección de las comunas quedó definida en:

$$P_h(i) = \begin{cases} 1 & \text{si } i \in h; \text{ es } IF \\ \frac{U_{hi}}{U_h} & \text{si } i \in h; \text{ es no } IF \end{cases}$$

Donde,

- $IF$  : Comuna de Inclusión Forzosa (capitales regionales o comunas de Gran Santiago)
- $h$  : Índice de los estratos de muestreo de primera etapa. En este caso, 16 regiones del país y otros estratos conformados según tamaño de comunas.
- $i$  : Índice de la UPM, correspondiente a la parte urbana de la comuna.
- $U_{hi}$  : Total, de personas mayores de 18 años que posee la UPM  $i$ , del estrato  $h$ , según el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.
- $U_h$  : Total, de personas mayores de 18 años que posee el estrato  $h$ , según el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.

De esta manera, la muestra de comunas seleccionadas permitió obtener una mayor presencia de comunas más grandes, aunque asegurando la inclusión de comunas más pequeñas en tanto se encontraron dentro de los estratos de muestreo.

En total, se seleccionaron 126 comunas. El listado de ellas se encuentra en la Tabla 7. Adicionalmente, se identifica a aquellas comunas que resultaron de Inclusión Forzosa, ya sea porque correspondían a capitales regionales (16), como por pertenecer al Gran Santiago (54), o por poseer más de 150.000 habitantes mayores de 18 años (2).

**Tabla 7 Listado de comunas seleccionadas**

<b>Tarapacá</b>	<b>Maule</b>	<b>Los Lagos</b>	<b>Metropolitana</b>
Alto Hospicio	Cauquenes	Ancud	Cerrillos IF
Iquique IF	Colbún	Frutillar	Cerro Navia IF
Pozo Almonte	Curicó	Llanquihue	Conchalí IF
<b>Antofagasta</b>	Linares	Los Muermos	El Bosque IF
Antofagasta IF	Maule	Osorno	Estación Central IF
Calama	Pencahue	Puerto Montt IF	Huechuraba IF
Taltal	Rauco	Puerto Varas	Independencia IF
Tocopilla	Talca IF	San Juan De La Costa	La Cisterna IF
<b>Atacama</b>	<b>Bíobío</b>	<b>Aysén</b>	La Florida IF
Caldera	Cabrero	Aysén	La Granja IF
Copiapó IF	Chiguayante	Cochrane	La Pintana IF
Diego De Almagro	Concepción IF	Coyhaique IF	La Reina IF
Vallenar	Curanilahue	<b>Magallanes</b>	Las Condes IF
<b>Coquimbo</b>	Florida	Natales	Lo Barnechea IF
Andacollo	Hualpén	Porvenir	Lo Espejo IF
Coquimbo IF	Hualqui	Punta Arenas IF	Lo Prado IF
La Serena	Los Ángeles	<b>Los Ríos</b>	Macul IF
Ovalle	Penco	Los Lagos	Maipú IF
Salamanca	Talcahuano	Paillaco	María Pinto
<b>Valparaíso</b>	Tucapel	Valdivia IF	Melipilla
Limache	Yumbel	<b>Arica y Parinacota</b>	Ñuñoa IF
Quilpué	<b>Araucanía</b>	Arica IF	Pedro Aguirre IF
San Antonio	Angol	<b>Ñuble</b>	Cerda
San Esteban	Cholchol	Chillán IF	Peñalolén IF
Santo Domingo	Curacautín	Chillán Viejo	Pirque IF
Valparaíso IF	Freire	Coelemu	Providencia
Viña Del Mar IF	Gorbea	Ñiquén	Pudahuel IF
<b>O'higgins</b>	Lonquimay	Pinto	Puente Alto IF
Chépica	Lumaco	San Carlos	Quilicura IF
Codegua	Nueva Imperial	Treguaco	Quinta Normal IF
Graneros	Padre Las Casas		Recoleta IF
Peralillo	Perquenco		Renca IF
Pichilemu	Pitrufquén		San Bernardo IF
Rancagua IF	Renaico		San Joaquín IF
			San Miguel IF

Rengo	Temuco	IF	San Ramón	IF
Santa Cruz			Santiago	IF
			Vitacura	IF

b) Segunda etapa de selección: manzanas

Las unidades secundarias de muestreo (USM) fueron las manzanas censales ubicadas en las áreas urbanas de las comunas seleccionadas del país.

Para este estudio se utilizó un método de selección sistemática con probabilidad proporcional al tamaño de las manzanas, medido éste como la cantidad de personas que habitan en la manzana según los registros del Marco muestral. Con esto se buscó uniformar la probabilidad de selección de las personas en esta muestra.

En concreto, la probabilidad de selección de las manzanas quedó definida en:

$$P_{hi}(j|i) = \frac{n_{hi} M_{hij}}{U_{hi}}$$

Donde,

- $j$  : Índice de la USM, correspondiente a manzana censal.
- $n_{hi}$  : Total, de USM a seleccionar en la UPM  $i$ , del estrato  $h$ .
- $U_{hi}$  : Total, de personas mayores de 18 años que posee la UPM  $i$ , del estrato  $h$ , según el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.
- $M_{hij}$  : Total, de personas mayores de 18 años que posee la USM  $j$ , en la UPM  $i$ , del estrato  $h$ , según el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.

En total, se seleccionaron 3.997 manzanas urbanas en distribuidas en las 16 regiones de Chile. En la Tabla 8 se encuentra el detalle del número de manzanas seleccionadas por comuna seleccionada.

**Tabla 8 Número de manzanas seleccionadas por comuna**

<b>Tarapacá</b>		<b>Maule</b>		<b>Los Lagos</b>		<b>Metropolitana</b>	
Alto Hospicio	36	Cauquenes	56	Ancud	30	Cerrillos	10
Iquique	74	Colbún	5	Frutillar	7	Cerro Navia	17
Pozo		Curicó	49	Llanquihue	14	Conchalí	16
Almonte	5	Linares	27	Los Muermos	16	El Bosque	20
<b>Antofagasta</b>		Maule	45	Osorno	71	Estación Central	20
Antofagasta	110	Pencahue	19	Puerto Montt	105	Huechuraba	12
Calama	48	Rauco	2	Puerto Varas	43	Independencia	13
Taltal	11	Talca	80	San Juan De La Costa	4	La Cisterna	12
Tocopilla	7	<b>Bíobío</b>		<b>Aysén</b>		La Florida	47
<b>Atacama</b>		Cabrero	18	Aysén	26	La Granja	15
Caldera	14	Chiguayante	44	Cochrane	9	La Pintana	21
Copiapó	60	Concepción	60	Coyhaique	63	La Reina	12
Diego De Almagro	11	Curanilahue	34	<b>Magallanes</b>		Las Condes	40
Vallenar	18	Florida	3	Natales	13	Lo Barnechea	12
<b>Coquimbo</b>		Hualpén	67	Porvenir	4	Lo Espejo	13
Andacollo	9	Hualqui	15	Punta Arenas	85	Lo Prado	13
Coquimbo	80	Los Ángeles	37	<b>Los Ríos</b>		Macul	16
La Serena	76	Penco	10	Los Lagos	16	Maipú	66
Ovalle	32	Talcahuano	39	Paillaco	39	María Pinto	4
Salamanca	32	Tucapel	5	Valdivia	75	Melipilla	73
<b>Valparaíso</b>		Yumbel	19	<b>Arica y Parinacota</b>		Ñuñoa	29
Limache	50	<b>Araucanía</b>		Arica	110	Pedro Aguirre	
Quilpué	66	Angol	37	<b>Ñuble</b>		Cerda	13
San Antonio	53	Cholchol	7	Chillán	107	Peñalolén	30
San Esteban	23	Curacautín	14	Chillán Viejo	32	Pirque	12
Santo Domingo	6	Freire	7	Coelemu	26	Providencia	20
Valparaíso	60	Gorbea	12	Ñiquén	2	Pudahuel	28
Viña Del Mar	69	Lonquimay	3	Pinto	22	Puente Alto	70
<b>O'higgins</b>		Lumaco	14	San Carlos	20	Quilicura	25
Chépica	21	Nueva Imperial	44	Treguaco	2	Quinta Normal	14
Codegua	7	Padre Las Casas	21			Recoleta	21
Graneros	33	Perquenco	5			Renca	18
Peralillo	12					San Bernardo	36

Pichilemu	16	Pitrufquén	20	San Joaquín	12
Rancagua	98	Renaico	6	San Miguel	14
Rengo	65	Temuco	126	San Ramón	11
Santa Cruz	30			Santiago	58
				Vitacura	11

### c) Tercera etapa de selección: viviendas

Las unidades terciarias de muestreo (UTM) corresponden a las viviendas en las manzanas ubicadas en las zonas urbanas del país. Se definió un número fijo de **10 viviendas seleccionadas cada manzana**.

A diferencia de las USM, las viviendas no son directamente identificables en el marco muestral que se dispone. Esto quiere decir que, previo a implementar la selección de las viviendas se debió “construir” un marco muestral, en terreno, para las USM seleccionadas en la muestra. El proceso de identificación y listado de las viviendas en las manzanas seleccionadas en la muestra se denomina “empadronamiento” (*listing*). Para realizar este procedimiento, el empadronador o encuestador contó con una Hoja de Ruta 1 que guía el proceso, y un mapa de la manzana seleccionada (el cual es elaborado por DESUC).

Durante este proceso de empadronamiento (el cual fue supervisado en su 100%) el encuestador hace un recorrido completo por la manzana seleccionada, lista las direcciones de todos los sitios y asigna un código único de acuerdo con el tipo de edificación que se encuentra en cada sitio. Las manzanas con edificaciones de uso habitacional particular son consideradas “elegibles” para el estudio. Las manzanas que no son de uso residencial (ej. sitios eriazos, plazas, establecimientos comerciales, educativos, etc.) se consideran “no elegibles”.

En las **manzanas empadronadas** (y con viviendas elegibles) el mecanismo de selección consistió en los siguientes pasos:

- identificar en el listado las viviendas “elegibles” en la manzana empadronada;
- calcular el total de viviendas elegibles y dividir por 10 para obtener el periodo del salto sistemático;
- seleccionar al azar una vivienda en cada manzana; y
- partiendo de la vivienda seleccionada, identificar las siguientes 9 viviendas a entrevistar aplicando el salto sistemático.

Este método, denominado **muestreo sistemático**, asignó igual probabilidad de selección a cada vivienda en la manzana seleccionada, la que queda definida según:

$$P_{hij}(k|j) = \frac{m_{hij}}{M_{hij}^*}$$

Donde,

- $k$  : Índice de la UTM, corresponde a la vivienda.
- $m_{hij}$  : Total de viviendas a seleccionar en la USM  $j$ , en la UPM  $i$ , del estrato  $h$ . Esto se encuentra definido por diseño en 10.
- $M_{hij}^*$  : Total de viviendas particulares elegibles (residenciales ocupadas) que posee la USM  $j$ , en la UPM  $i$ , del estrato  $h$ , según lo registrado en el proceso de empadronamiento

Este proceso de selección se realizó en paralelo a la ejecución de entrevistas dentro de las viviendas seleccionadas. Finalmente, al terminar del proceso de empadronamiento, y habiéndose realizado las supervisiones al 100% de estos, se procedió a la digitación de una base de empadronamiento, la cual tendrá tantas filas como manzanas seleccionadas en la muestra.

#### d) Cuarta etapa de selección: personas

Las unidades últimas de muestreo (UUM) en este estudio corresponden a las personas mayores de 18 años o más que son residentes habituales de las viviendas seleccionadas. La persona seleccionada a este nivel será el único informante por entrevistar.

Tampoco se dispone de un “listado” de las unidades (las personas) que cumplen con estos requerimientos, por lo tanto, se debió implementar un procedimiento que permitiera identificar a las personas elegibles en cada vivienda y luego seleccionar al informante.

Para la encuesta se utilizó un método totalmente probabilístico, donde se seleccionó a una persona elegible dentro de la vivienda del siguiente modo:

1. El encuestador enlista en el software de aplicación de encuestas en su dispositivo electrónico a todas las personas de la vivienda, registrando datos como nombre de pila, sexo, edad.
2. La función de la edad de las personas, el software definirá la característica de elegible o no elegible para cada una de las personas. Cabe señalar que, en consonancia con lo estipulado en las Bases de Licitación, se excluyeron del listado y, por tanto, de la categoría de elegibles a aquellas personas que mostraran actitudes violentas con los encuestadores. Esto queda debidamente registrado en los códigos de disposición final de casos.
3. El software automáticamente escoge entre las personas elegibles, a una persona que es la seleccionada para participar del estudio.

Adicionalmente, y atendiendo a lo solicitado por la contraparte técnica, se decidió sobremuestrear a la población de menor de 30 años, en tanto constituye una subpoblación de interés en elementos relativos a diversidad sexual e Infecciones de Transmisión Sexual (ITS). Dado ello, dicha población tuvo el doble de probabilidad de ser escogidos.

Independiente de ello, este mecanismo de selección se trata de un método probabilístico, ya que asigna una probabilidad de selección conocida a cada persona en la vivienda, preservando el carácter probabilístico del diseño muestral aplicado en las etapas de selección previas. En específico, la probabilidad de selección de la persona queda definido por:

$$P_{hijk}(l|k) = \frac{q_l}{\sum q_l}$$

Donde,

$q_l$  : Es 2 cuando persona tiene <30 años, y 1 cuando persona tiene >= 30 años.

Dado que este proceso se encuentra incorporado en el software de aplicación, esto asegura que no existan errores en el proceso de selección de la persona.

### 3 Cálculo de Factores de Expansión

En la presente encuesta —al igual que en otros levantamientos de datos— se sobre-representó a grupos de la población con el fin de obtener indicadores fiables para ellos, aunque, por su peso poblacional, podrían ser poco precisos. Una variable de diseño para este estudio es la geografía, donde es importante obtener indicadores regionales para algunas con baja población. Por ejemplo, las regiones de Aysén y Magallanes. Además, también se sobre-muestreó a la población menor de 30 años.

Considerando esa característica del diseño muestral, es necesario que la muestra obtenida se pondere, esto es, se ajuste por las distintas tasas de sub o sobre representación de los estratos de la muestra. Para ello se construyó un **factor de expansión** que permitiera corregir esos desajustes respecto a las distribuciones poblacionales.

Los **factores de expansión** se utilizan comúnmente para asignarle a los individuos de una muestra un “peso” de tal manera que las estimaciones ponderadas de la encuesta representen de la mejor manera posible a la población de objetivo.

En la práctica, la construcción de dichos factores puede variar. En general contempla tres etapas: ponderación por probabilidades de selección, ajustes de no respuesta y ajuste de calibración a totales poblacionales externos.

En la primera etapa, se genera un ponderador base que refleja de cerca las probabilidades de selección de los individuos de la muestra. En la segunda, se efectúan ajustes para compensar por la no respuesta de los individuos inicialmente seleccionados. Y en la tercera, se realiza una corrección por fallas de cobertura, para que:

1. la muestra se asemeje a la población en variables auxiliares relevantes, y
2. mejorar la precisión en estimaciones de subpoblaciones de interés (Kalton & Flores-Cervantes, 2003)

De esta manera, se construyeron dos factores de expansión:

- **Factor de expansión general:** que tiene como principal objetivo que las personas que contestaron la encuesta puedan representar a la población objetivo del estudio, es decir, personas de 18 años y más de las zonas urbanas de las 16 regiones de Chile. Este factor de expansión considera ajuste por probabilidades de selección, ajuste explícito de no respuesta, y calibración a totales poblacionales. Dicho factor se encuentra disponible para todos los respondentes de la encuesta, y se encuentra en la base de datos con el nombre ***w\_personas\_cal***.

- **Factor de expansión por cuestionario incompleto:** que tiene como principal objetivo que las personas que contestaron el cuestionario completo (y en particular, el módulo de hogar), puedan representar apropiadamente a aquellos que no lo contestaron. Este factor incluye los ajustes por probabilidades basales, ajuste explícito por no respuesta al módulo de hogar, y calibración a totales poblacionales. Este factor se encuentra disponible solo para aquellos que respondieron el módulo hogar, y su nombre en la base de datos es ***w\_personas\_cal\_nr\_mod\_hogar***.

### 3.1 Ajuste por probabilidades de selección

En primera instancia se considera un **ponderador base** que, como ya fue mencionado previamente, reflejara las probabilidades de selección desiguales en el contexto del estudio. El ponderador base se construye como el inverso de la probabilidad de selección de las viviendas, quedando expresado de la siguiente manera.

$$wb_{viviendas} = \frac{1}{P(i)}$$

Donde  $P(i)$  corresponde a la probabilidad de selección de cada vivienda seleccionada. A su vez, esto se construye considerando la probabilidad de selección de la vivienda en una manzana, cuya probabilidad es condicional a las probabilidades de que la manzana fuera seleccionada, y esta probabilidad está a su vez condicionada por la probabilidad de que la comuna fuera seleccionada para incluirse en la encuesta. Esta expresión también se puede entender como el total de unidades (en este caso viviendas) a las cual representa una determinada vivienda  $i$  que es parte de la muestra.

A continuación, se describirá la construcción del ponderador base.

### 3.1.1 Ponderador de comuna

Como fue mencionado previamente, la ENSSEX 2022-2023 posee un diseño muestral probabilístico, multietápico y estratificado por área geográfica. La muestra fue seleccionada en cuatro etapas: 1) comunas, 2) manzanas, 3) viviendas y 4) personas, donde las comunas corresponden a las Unidades Primarias de Muestreo (UPM).

Sin embargo, previo a la selección de comunas se excluyeron del Marco Muestral aquellas consideradas “Áreas de Dificil Acceso” (ADA)<sup>8</sup>, por el Instituto Nacional de Estadísticas<sup>9</sup> y DESUC<sup>10</sup>, correspondiendo a 25 comunas excluidas.

Luego de ello, se definieron comunas como relevantes o de Inclusión Forzosa (capitales regionales, comunas del Gran Santiago, o de gran tamaño), que se incluyeron con certeza en la muestra, es decir, con probabilidad igual a 1.

La selección de comunas siguió el siguiente procedimiento. Dentro de los estratos de muestreo las comunas —que no fueran de inclusión forzosa— se escogieron con una probabilidad de selección proporcional a su tamaño, el cual fue medido por la cantidad de personas mayores de 18 años que habiten en ellas.

La probabilidad de selección de las comunas quedó definida en:

$$P_h(i) = \begin{cases} 1 & \text{si } i \in h; \text{ es IF} \\ \frac{U_{hi}}{U_h} & \text{si } i \in h; \text{ es no IF} \end{cases}$$

---

<sup>8</sup> Las comunas de difícil acceso definidas por el INE durante el proceso de PreCenso 2011 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

<sup>9</sup> Para más información respecto a las Áreas de Dificil Acceso ver documento: “Definición de áreas especiales para el levantamiento censal”, septiembre 2015, Instituto Nacional de Estadísticas.

<sup>10</sup> A la lista de INE, DESUC agrega las comunas de Camarones, Chile Chico y Putre, debido a la verificación en levantamientos de datos previos de dificultades para acceder a ellas por sus condiciones físico-geográficas (por ejemplo, falta de infraestructura vial, o de medios de transporte públicos y privados constantes que permitan conectar con la zona, dificultades climáticas, entre otras).

Donde,

- $IF$  : Comuna de Inclusión Forzosa (capitales regionales o comunas de Gran Santiago)
- $h$  : Índice de los estratos de muestreo de primera etapa. En este caso, 16 regiones del país y otros estratos conformados según tamaño de comunas.
- $i$  : Índice de la UPM, correspondiente a la parte urbana de la comuna.
- $U_{hi}$  : Total, de personas mayores de 18 años que posee la UPM  $i$ , del estrato  $h$ , según el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.
- $U_h$  : Total, de personas mayores de 18 años que posee el estrato  $h$ , según el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.

Teniendo ello en consideración, el ponderador de selección de comunas se define como el inverso de probabilidad de selección de las comunas:

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

Este ponderador constituye la base para el desarrollo de los factores de expansión, aunque usualmente no es utilizado para desarrollar estimaciones de manera directa.

En la Anexos

### 3.2 Estadísticas asociadas a Factores de Expansión

Tabla 17 se pueden ver las estadísticas descriptivas asociadas al ponderador de comunas de las 126 comunas seleccionadas en la muestra de ENSSEX 2022-2023. La columna #2 muestra el número de comunas seleccionadas por región. La columna #3 muestra el valor mínimo del ponderador, donde los valores 1 corresponden a comunas de Inclusión Forzosa, y las mayores a 1 corresponden a ponderadores de comunas sin inclusión forzosa. Dentro de ellas, el máximo se encuentra en la región del Maule con un 14,6.

En el resto de las columnas se presentan resultados asociados a percentiles de interés del ponderador, junto con la media y suma de este.

### 3.2.1 Ponderador de manzanas

Al momento de la selección de manzanas se excluyeron —por razones operativas—, aquellas con 7 o menos viviendas habitadas (Grupo 0 y Sin viviendas). Esta exclusión dejó fuera al 1% de las viviendas catastradas en el marco, pero removi6 del marco al 9% de sus manzanas.

Luego de ello, se procedi6 a la selecci6n de manzanas censales dentro de las comunas previamente seleccionadas, correspondiendo a las Unidades Secundarias de Muestreo (USM). Para dicha selecci6n se utiliz6 un m6todo de selecci6n sistem6tica con probabilidad proporcional al tama6o de las manzanas, medido 6ste como la cantidad de personas que habitan en la manzana seg6n los registros del Marco muestral. Con esto se busc6 uniformar la probabilidad de selecci6n de las personas en esta muestra.

En concreto, la probabilidad de selecci6n de las manzanas qued6 definida en:

$$P_{hi}(j|i) = \frac{n_{hi} M_{hij}}{U_{hi}}$$

Donde,

- $j$  : 6ndice de la USM, correspondiente a manzana censal.
- $n_{hi}$  : Total de USM a seleccionar en la UPM  $i$ , del estrato  $h$ .
- $U_{hi}$  : Total de personas mayores de 18 a6os que posee la UPM  $i$ , del estrato  $h$ , seg6n el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.
- $M_{hij}$  : Total de personas mayores de 18 a6os que posee la USM  $j$ , en la UPM  $i$ , del estrato  $h$ , seg6n el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.

Ahora, para poder definir el n6mero de manzanas que cada USM representan en la poblaci6n, se hace necesario definir la probabilidad incondicional de selecci6n de cada manzana. Esto se define como el producto de la probabilidad de selecci6n de las comunas y la probabilidad condicional de las USM, seg6n lo se6alado a continuaci6n:

$$P_{hi}(j) = P_h(i) * P_{hi}(j|i)$$

A su vez, y considerando la existencia de comunas con y sin inclusi6n forzosa, se tiene que:

$$P_{hi}(j) = \begin{cases} \frac{n_{hi} M_{hij}}{U_{hi}} & \text{si } i \in h; \text{ es IF} \\ \frac{U_{hi}}{U_h} * \frac{n_{hi} M_{hij}}{U_{hi}} & \text{si } i \in h; \text{ es no IF} \end{cases}$$

De esta manera, el ponderador de manzanas queda definido como el inverso de la probabilidad de selección de cada manzana, según lo siguiente:

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

Se establece la diferencia de ponderadores para manzanas en comunas sin y con inclusión forzosa de la siguiente forma:

$$w_{hij} = \begin{cases} \frac{U_{hi}}{n_{hi} M_{hij}} & \text{si } i \in h; \text{ es IF} \\ \frac{U_h}{U_{hi}} * \frac{U_{hi}}{n_{hi} M_{hij}} & \text{si } i \in h; \text{ es no IF} \end{cases}$$

Así, el ponderador  $w_{hij}$  puede ser interpretado como la cantidad o número de manzanas en la población que representa cada manzana seleccionada en la muestra de ENSSEX 2022-2023.

### 3.2.1.1 Ajuste por omisión de manzanas

Como fue mencionado anteriormente, de manera previa a la selección de las manzanas se excluyeron del marco aquellas que tuvieran menos de 7 viviendas según la información contenida en el Marco Muestral. Dado esta omisión de manzanas pequeñas, el ponderador calculado previamente no considera a dichas manzanas, y por lo tanto, la suma de dicho ponderador subestimaría el total de viviendas que se encuentra contenida en el Marco Muestral.

Debido a ello, es que se desarrolló un ajuste para corregir esta omisión. Este ajuste corresponde a la razón en cada estrato de muestreo entre la totalidad de viviendas que posee el Marco Muestral, y la estimación total de viviendas que se obtiene a partir del ponderador de manzanas recién construido. Ello queda expresado de la siguiente forma:

$$R_{h,omit}^{MZ} = \frac{M_h}{\sum_{i \in \Omega_h} \sum_{j \in \Omega_{hi}} w_{hij} * M_{hij}}$$

Donde:

- $M_h$  : Total de viviendas en el estrato  $h$  que posee el Marco de muestreo.  
 $\Omega_h$  : Conjunto de comunas o pseudocomunas seleccionadas en el estrato  $h$ .  
 $\Omega_{hi}$  : Conjunto de manzanas seleccionadas en las comunas o pseudocomunas  $i$ , del estrato  $h$ .  
 $M_{hij}$  : Total de viviendas que posee la USM  $j$ , en la UPM  $i$ , del estrato  $h$ , según el Marco de muestreo, sin considerar manzanas de 7 viviendas o menos.

De esta forma, el ponderador de selección de manzanas ajustado por omisión se define en:

$$w'_{hij} = w_{hij} * R_{h,omit}^{MZ}$$

En la Tabla 18 se pueden ver las estadísticas descriptivas asociadas al ajuste por omisión de manzanas. Como se puede ver, el mínimo se encuentra en 1<sup>11</sup> y el máximo en 1,88 en la región de Valparaíso.

### 3.2.1.2 Ajuste por elegibilidad de manzanas

Los ponderadores o factores de expansión a nivel de manzanas tienen por objeto representar a las unidades elegibles respecto a la población objetivo de la encuesta. Para ENSSEX, esto corresponde a manzanas donde existe al menos una vivienda particular ocupada en el momento de levantamiento de datos.

Por otro lado, existen manzanas que son “no elegibles”, que son aquellas donde se verifica que no poseen viviendas particulares en el proceso de empadronamiento. Un ejemplo de ello es una manzana que en función de la información del Marco Muestral poseía viviendas particulares, pero en el momento del levantamiento de datos ninguna de sus edificaciones es una vivienda particular. Además de ello, también existen manzanas de “elegibilidad desconocida”, en las cuales por – por distintos motivos – no se pudo determinar si eran manzanas elegibles o no. Un ejemplo de ello son manzanas ubicadas en lugares extremos, por lo cual no fueron enviadas a terreno para ser empadronadas.

Respecto a la distribución de esta condición, de las 3.997 manzanas seleccionadas en la muestra de ENSSEX 2022-2023, no existieron manzanas no elegibles, pero sí se presentan 335

---

<sup>11</sup> Los valores menores a 1 se imputaron en 1 por parsimonia.

manzanas con elegibilidad desconocida, en tanto no fueron empadronadas en terreno. Con ello, se tuvieron 3.662 manzanas elegibles.

a) Ajuste por elegibilidad desconocida de manzanas

Para realizar el ajuste por elegibilidad desconocida a nivel de manzanas, en primera instancia se identificaron aquellas manzanas de elegibilidad desconocida durante el proceso de empadronamiento. En la Tabla 19 se muestra la distribución de dichas 335 manzanas con elegibilidad desconocida por región.

Luego de ello, se procede a redistribuir el peso de dichas manzanas entre aquellas que tengan una elegibilidad conocida. Esto se realiza al interior de cada área  $a$ . Los pesos se redistribuyeron siguiendo la siguiente fórmula:

$$R_{a, know}^{MZ} = \begin{cases} \frac{\sum_{h \in a} \sum_{i \in \Omega_h} \sum_{j \in \theta_i} w'_{hij}}{\sum_{h \in a} \sum_{i \in \Omega_h} \sum_{j \in \theta_{i, know}} w'_{hij}} & \text{si } j \in \text{eleg. conocida} \\ 0 & \text{si } j \in \text{eleg. desconocida} \end{cases}$$

Donde:

- $a$  : Indicador de área, que toma valores de 1 a 16, identificando a cada una de las regiones.
- $w'_{hij}$  : Ponderador de selección corregido por omisión de manzanas asociadas a la USM  $j$ , en la UPM  $i$ , del estrato  $h$
- $\Omega_h$  : Conjunto de comunas o pseudocomunas seleccionadas en el estrato  $h$ .
- $\theta_i$  : Conjunto de manzanas con elegibilidad conocida y desconocida seleccionadas en las comunas o pseudocomunas  $i$ .
- $\theta_{i, know}$  : Conjunto de manzanas con elegibilidad conocida seleccionadas en las comunas o pseudocomunas  $i$ .

De esta forma, este ajuste corresponde a la razón entre la estimación total de USM en la población, y la estimación total de USM con elegibilidad conocida, al interior de cada región. En la Tabla 20 se muestra este ajuste para cada una de las regiones.

Cabe señalar que este ajuste solo se realiza para manzanas con elegibilidad conocida, mientras que las de elegibilidad desconocida son descartadas a partir de ahora del cálculo de factores de expansión. Sin embargo, y debido al ajuste, el peso poblacional asociadas a dichas

manzanas se conserva gracias a la distribución del mismo entre las manzanas con elegibilidad conocida.

#### b) Ajuste por no elegibilidad de manzanas

Luego del cálculo de ajuste por elegibilidad desconocida a nivel de manzanas, se puede proceder a descartar a las manzanas no elegibles, puesto que en principio no debieron ser consideradas al no pertenecer a la población objetivo del estudio. Cabe destacar que esta eliminación debe realizarse con posterioridad al ajuste por elegibilidad desconocida, puesto que para el cálculo del ajuste se consideran manzanas tanto elegibles como no elegibles. El ajuste por no elegibilidad de manzanas se define en:

$$R_{a,elec}^{MZ} = \begin{cases} 1 & \text{si } j \in \text{elegible} \\ 0 & \text{si } j \in \text{no elegible} \end{cases}$$

#### 3.2.1.3 Ponderador de manzanas ajustado

Así, el ponderador de selección de manzanas ajustado tanto por “no elegibilidad” como por “elegibilidad desconocida” se define en:

$$w''_{hij} = w'_{hij} * R_{a,known}^{MZ} * R_{a,elec}^{MZ}$$

En la Tabla 21 se pueden observar las estadísticas descriptivas del ponderador de manzanas ajustado, a nivel de región. El mayor ponderador se encuentra en la región Metropolitana, mientras que la región de Aysén presenta los menores ponderadores. La suma de este factor de expansión corresponde a 144.441 manzanas urbanas. Como referencia, se menciona que el total de manzanas en el marco era de 151.545 manzanas.

#### 3.2.2 Ponderador de viviendas

Las Unidades Terciarias de Muestreo (UTM) en ENSSEX 2022-2023 correspondieron a viviendas particulares, las cuales fueron escogidas en cada manzana (USM) seleccionada.

El procedimiento de selección se realiza con posterioridad al proceso de empadronamiento, en el cual todas las viviendas particulares son identificadas y listadas. Luego de ello, se procedió a seleccionar mediante un método de salto sistemático a 10 viviendas al interior de cada manzana. Cabe señalar que si en el proceso de empadronamiento la manzana poseía un

número menor a 10<sup>12</sup> de viviendas particulares, todas las viviendas fueron seleccionadas con certeza.

De este modo, la probabilidad condicional de incluir a cada k-ésima vivienda particular en la muestra, dada la selección de la USM en la UPM, queda definida en la fórmula:

$$P_{hij}(k|j) = \frac{m_{hij}}{M_{hij}^*}$$

Donde:

- $k$  : Indicador de la UTM, correspondiente a una vivienda.
- $m_{hij}$  : Total, de viviendas seleccionadas en la USM  $j$ , de la UPM  $i$ , del estrato  $h$ .
- $M_{hij}^*$  : Total, de viviendas particulares ocupadas que posee la USM  $j$ , de la UPM  $i$ , del estrato  $h$ , de acuerdo a la información obtenida a partir del proceso de empadronamiento.

Adicionalmente a ello, se debe considerar las siguientes probabilidades de selección:

$$P_{hij}(k) = P_{hi}(j) * P_{hij}(k|j)$$

$$\approx \left[ \frac{1}{w''_{hij}} \right] * P_{hij}(k|j)$$

Donde:

- $P_{hij}(k)$  : Probabilidad incondicional de selección de la k-ésima vivienda, perteneciente a la USM  $j$ , de la UPM  $i$ , del estrato  $h$ .
- $P_{hij}(k|j)$  : Probabilidad condicional de selección de la k-ésima vivienda, perteneciente a la USM  $j$ , de la UPM  $i$ , del estrato  $h$ .
- $P_{hi}(j)$  : Probabilidad incondicional de selección de una manzana en la muestra, en la UPM  $i$ , del estrato  $h$ .

---

<sup>12</sup> Menor a 10, pero superior a 6, dado que manzanas con menos de 7 viviendas quedaron excluidas del marco.

Entonces, la probabilidad incondicional de selección de cada vivienda seleccionada en ENSSEX 2022-2023 se puede definir como la primera expresión. Ahora, se debe considerar que la probabilidad de selección de la manzana ( $P_{hi}(j)$ ) fue corregida por omisión y elegibilidad, por lo que se debe considerar éstas últimas correcciones para la definición de la probabilidad de selección de la vivienda.

Con ello, la probabilidad incondicional de selección de cada vivienda seleccionada en la encuesta, ajustada por omisión y elegibilidad de manzanas, queda definida como la segunda expresión.

Dado todo lo anterior, el ponderador de selección de viviendas se calcula como el inverso de la probabilidad incondicional de las viviendas, lo cual se puede expresar en:

$$w_{hijk} = \frac{1}{P_{hij}(k)}$$

O, reemplazando el valor de la probabilidad incondicional de selección de la vivienda, se expresa en:

$$w_{hijk} = w''_{hij} * \frac{1}{P_{hij}(k|j)}$$

Una vez calculado el ponderador, se puede estimar el total de viviendas que existen en la población durante el trabajo de campo de ENSSEX 2022-2023, realizando la suma de dicho ponderador.

Luego de ello, se revisó la existencia de valores extremos, observándose que 15 ponderadores (asociados a 15 manzanas), presentaron valores superiores a 5 veces el valor de la media de los ponderadores dentro de cada estrato de muestreo. Dichos ponderadores fueron truncados al valor máximo admisible (5 veces la media del ponderador en el estrato) y posteriormente sus pesos fueron distribuidos al interior de los estratos con el siguiente ajuste:

$$R_{ajust} = \frac{\sum_h w_{hijk}}{\sum_h w_{hijk_{trunc}}}$$

$\sum_h w_{hijk}$  : Suma de ponderadores de selección de viviendas al interior de cada estrato  $h$ .

$\sum_h w_{hijk}^{trunc}$  : Suma de ponderadores de selección de viviendas truncados al interior de cada estrato  $h$ .

Con ello, el valor máximo de dicho ponderador fue de 1.031.

### 3.2.2.1 Ajuste por elegibilidad de viviendas

Los ponderadores o factores de expansión a nivel de vivienda tienen por objetivo representar a la población elegible de ENSSEX 2022-2023. En concreto, ello se traduce en representar a las viviendas particulares ocupadas por al menos una persona de 18 años y más. Sin embargo, y al igual que en el caso de manzanas, existen viviendas que son “no elegibles” o con “elegibilidad desconocida”.

Un ejemplo de lo primero es una vivienda donde solo residan menores de edad, lo que la calificaría como no elegible para el caso de ENSSEX 2022-2023. Por otra parte, un ejemplo de lo segundo son aquellas viviendas donde – por diversos motivos – no fue posible determinar si en dicha vivienda habitaban personas de 18 años y más, ya sea porque no se envió a terreno, o porque no se pudo establecer contacto con los residentes.

En concreto, para el caso de ENSSEX, de las 36.620 viviendas seleccionadas (provenientes de 3.662 manzanas elegibles), 2 resultaron de elegibilidad desconocida y 66.618 con elegibilidad conocida. De estas con elegibilidad conocida, 18 resultaron ser no elegibles, y 66.600 elegibles.

Al igual que en el caso de manzanas, el ponderador de viviendas deberá ser ajustado por estas condiciones de elegibilidad.

### a) Ajuste por elegibilidad desconocida de viviendas

Para realizar el ajuste por elegibilidad desconocida de vivienda se siguieron dos pasos. En primera instancia, se identificó aquellas viviendas con elegibilidad desconocida, que correspondieron a solo 2 viviendas en total. La distribución de dichas viviendas por región se puede ver en la Tabla 22. En segundo lugar, se debe distribuir el peso de dichas viviendas entre aquellas que tienen una elegibilidad conocida. Dado que la cantidad de viviendas con elegibilidad desconocida es muy baja en algunas regiones, se optó por omitir este ajuste, y de este modo, evitar inducir variabilidad indeseada al ponderador.

Esto se puede expresar en:

$$R_{a,known}^{VIV} = \begin{cases} 1 & \text{si } k \in \text{eleg. conocida} \\ 0 & \text{si } k \in \text{eleg. desconocida} \end{cases}$$

Con ello, de ahora en adelante, las viviendas de elegibilidad desconocida se descartan del proceso de construcción de factores de expansión.

### b) Ajuste por no elegibilidad de viviendas

En ENSSEX busca representar a personas que son parte de la población objetivo, es decir, personas mayores de 18 años. Dado ello, en esta etapa se debe proceder a descartar los factores de expansión asociados a unidades que sean “no elegibles”. Este ajuste se realiza mediante:

$$R_{a,eleg}^{VIV} = \begin{cases} 1 & \text{si } k \in \text{elegible} \\ 0 & \text{si } k \in \text{no elegible} \end{cases}$$

En este caso, se detectaron 18 casos de viviendas no elegibles. La distribución de dichas viviendas según región se muestra en la Tabla 23.

### 3.2.2.2 Ponderador de viviendas ajustado por elegibilidad

Luego, el ponderador de selección de vivienda ajustado por elegibilidad (tanto de no elegibilidad como de elegibilidad desconocida), se define en:

$$w'_{hijk} = w_{hijk} * R_{a,known}^{VIV} * R_{a,eleg}^{VIV}$$

En la Tabla 24 se exponen las estadísticas descriptivas del ponderador de viviendas ajustado por elegibilidad, a nivel de región. Como se puede ver, el valor mínimo es de 11,6 en la región de Aysén, y el máximo de 1.032 en la región Metropolitana. La suma de este ponderador es la representación de 3.586.570 viviendas particulares.

### 3.3 Ajuste explícitos por no respuesta

La segunda etapa en la construcción del factor corresponde a la realización de ajustes explícitos por no respuesta a la unidad. Esto se realiza debido a que de las 36.600 viviendas elegibles para ENSSEX 2022-2023, se pudo concretar entrevistas solo en 20.392 casos. El resto de los casos (16.208) fueron rechazos o no contactos. En la Tabla 25 se puede ver la distribución de los resultados de las visitas de las 36.600 viviendas elegibles, por región.

En un escenario donde las viviendas que participaron del estudio tuvieran características similares a las que no respondieron la encuesta, la no respuesta no constituiría un problema para realizar estimaciones a partir del estudio. Sin embargo, este escenario no es comprobable, por lo cual nos podríamos enfrentar a un posible “sesgo de no respuesta”. Una manera de enfrentar el problema es realizar un ajuste explícito a la no respuesta.

El ajuste de no respuesta a través de ponderadores busca redistribuir los ponderadores de los elementos seleccionados que no respondieron entre elementos seleccionados que sí respondieron de características similares. Dicho de otra forma, a los que sí responden se les aumenta el “peso” para que representen a los individuos que no respondieron (Groves & Peytcheva, 2008; Kalton & Flores-Cervantes, 2003).

Para realizar el ajuste explícito de no respuesta se utilizará la técnica de “*propensity stratification*”, método introducido por Rosenbaum y Rubin (1985). En ese método, se crean celdas de ajustes a partir de la predicción de responder o no en la encuesta en base a una serie de variables auxiliares, que en el caso ideal debiesen estar correlacionadas con la probabilidad de responder y una variable de interés del estudio.

A continuación, se describirán las etapas de la realización de este ajuste.

### 3.3.1 Modelo de no respuesta

Para construir el modelo de no respuesta se realizan dos procesos; primero, se estima una probabilidad de responder, y segundo, se conforman celdas de ajuste en función de dicha probabilidad.

#### 3.3.1.1 Modelamiento y estimación de la probabilidad de responder

Para realizar el ajuste de no respuesta, en primera instancia es necesario estimar, para cada una de las viviendas elegibles en la muestra, una probabilidad de responder la encuesta.

En concreto, esta estimación se realiza a partir de modelos. En ENSSEX se realizó un modelo de regresión lineal múltiple para modelar la probabilidad de responder a nivel de manzana. Cabe destacar que se realiza este modelamiento a nivel de manzana pues las variables utilizadas para el modelamiento no se encuentran a nivel de manzana.

De esta forma, la variable dependiente del modelo corresponde a la tasa de respuesta de la manzana  $j$ . En cada manzana se calculó la razón no ponderada entre el total de las viviendas elegibles que responden y el total de viviendas elegibles. Esto se puede expresar en:

$$tr_j = \frac{n_{j,eleg,resp}}{n_{j,eleg}}$$

Donde:

$n_{j,eleg,resp}$  : Total de viviendas de la USM  $j$  que son elegibles seleccionadas y que responden

$n_{j,eleg}$  : Total de viviendas de la USM  $j$  que son elegibles y seleccionadas

Ahora, las variables independientes que se incluyeron en el modelo son tanto a nivel de manzanas, como a nivel de comuna. El detalle de las variables utilizadas se expone en la Tabla 9.

**Tabla 9 Variables independientes utilizadas para el modelo de no respuesta**

<b>Variable</b>	<b>Descripción</b>	<b>Nivel</b>
tasa_mujeres	Tasa de mujeres en la manzana, según lo reportado en Marco Muestral	Manzana
tasa_ninos	Tasa de personas entre 0 a 5 años en la manzana, según lo reportado en Marco Muestral	Manzana
tasa_adultos_mayores	Tasa de personas mayores a 65 años en la manzana, según lo reportado en Marco Muestral	Manzana
tasa_migrantes	Tasa de personas migrantes en la manzana, según lo reportado en Marco Muestral	Manzana
tasa_mat_irrec	Tasa de viviendas con materialidad irrecuperable, según lo reportado en Marco Muestral	Manzana
viv_marco	Número de viviendas particulares, según lo reportado en Marco Muestral	Manzana
ingreso_total	Ingreso total promedio de la comuna, según resultados Casen 2020	Comuna
ingreso_trabajo	Ingreso proveniente del trabajo promedio de la comuna, según resultados Casen 2020	Comuna
pobreza	Porcentaje de hogares bajo la línea de la pobreza en la comuna, según resultados Casen 2020	Comuna
decil	Decil de ingresos promedio de los hogares de la comuna, según resultados Casen 2020	Comuna
educ_sup_comp_jh	Porcentaje de hogares de la comuna con jefaturas de hogar con educación superior completa, según resultados Casen 2020	Comuna
50onasa_jh	Porcentaje de hogares de la comuna con jefaturas de hogar afiliados a FONASA, según resultados Casen 2020	Comuna
hacinamiento	Porcentaje de hogares de la comuna con hacinamiento, según resultados Casen 2020	Comuna
ind_san_def	Porcentaje de hogares de la comuna con índice de saneamiento deficitario, según resultados Casen 2020	Comuna
alleg_int	Porcentaje de hogares de la comuna con allegamiento interno, según resultados Casen 2020	Comuna
alleg_ext	Porcentaje de hogares de la comuna con allegamiento externo, según resultados Casen 2020	Comuna

Las variables que resultaron predecir de mejor manera la respuesta fue la tasa de migrantes en la manzana, el porcentaje de hogares bajo la línea de pobreza de la comuna, el porcentaje de hogares con jefes de hogar con educación superior completa, y el porcentaje de hogares con allegamiento externo en la comuna.

Luego de ello, y a partir de un modelo optimizado (que solo consideró las variables que eran significativas para predecir la respuesta), se calculó la tasa de respuesta predicha a nivel de manzana, para cada una de las viviendas elegibles en la muestra. Esto, en tanto la tasa de respuesta predicha de la manzana puede ser considerado como un proxy de respuesta de la vivienda.

### 3.3.1.2 Creación de celdas de ajuste de no respuesta

Para ENSSEX 2022-2023 se optó por utilizar la tasa de respuesta predicha de manera agrupada, y en específico, el método de clases llamado “*propensity stratification*” (Rosenbaum y Rubin (1985)). Este método consiste en el siguiente procedimiento:

1. Calcular la probabilidad de responder para cada una de las unidades de la muestra en función de modelo de regresión utilizado.
2. Ordenar los casos de mayor a menor en función de la probabilidad predicha.
3. Formar celdas que agrupen a un mismo número de elementos de la muestra, que incluirán a respondentes y no respondentes.

Así, se agrupan viviendas en clases donde se comparten predictores, o se tiene una probabilidad de responder cercana. De esta manera, se construyeron 40 celdas ( $g$ ) de ajuste de igual tamaño a partir de la predicción de responder. Con ello, la celda #1 agrupa viviendas con la tasa predicha de responder más baja, y la celda #40 a aquellas con la tasa predicha de responder más alta. En la Tabla 26 se puede observar la tasa de respuesta efectiva y predicha de las viviendas de cada una de las celdas de ajuste.

### 3.3.2 Cálculo del factor de ajuste de no respuesta de vivienda

Tal como fue mencionado anteriormente, el objeto de realizar un ajuste explícito de no respuesta consiste en la reducción de un posible sesgo de no respuesta. Esto se realiza redistribuyendo los pesos de los casos que no respondieron o participaron de la encuesta, entre aquellos que sí lo hicieron. Esto se realiza al interior de cada una de las celdas de ajuste  $g = 40$ .

Para ello, se calcula una razón de ajuste en cada celda usando el ponderador de viviendas corregido por elegibilidad y suavizado. De esta manera, el numerador queda conformado por

la estimación total de viviendas elegibles, y en el denominador la estimación del total de viviendas elegibles que responde. Esto puede ser expresado con la siguiente fórmula:

$$\hat{R}_{g,Resp} = \begin{cases} \frac{\sum_{k \in \varphi_{g,eleg}} w'_{hijk}}{\sum_{k \in \varphi_{g,eleg,Resp}} w'_{hijk}} & \text{si } k \in \text{responde} \\ 0 & \text{si } k \in \text{no responde} \end{cases}$$

Donde,

- $w'_{hijk}$  : Ponderador de viviendas ajustado por elegibilidad y suavizado, asociado a la vivienda  $k$ , de la manzana  $j$ , seleccionada al interior de la comuna  $i$  en el estrato  $h$ .
- $g$  : Índice de las 40 clases (celdas) creadas para el ajuste de no respuesta.
- $\varphi_{g,eleg}$  : Conjunto de viviendas elegibles pertenecientes a la celda  $g$ .
- $\varphi_{g,eleg,Resp}$  : Conjunto de viviendas elegibles que respondieron la encuesta, pertenecientes a la celda  $g$ .

En la Tabla 27 se puede ver los valores de  $\hat{R}_{g,Resp}$  para cada una de las 40 celdas creadas. Destaca que el factor de ajuste mayor se encuentra en la celda #2, con un ajuste de 2,29, y la celda con menor ajuste es la #23, con una razón de 1,43.

### 3.3.3 Ponderador de viviendas ajustado por no respuesta

De esta manera, se puede obtener el ponderador de viviendas ajustado por no respuesta, que queda expresado en:

$$w''_{hijk} = w'_{hijk} * \hat{R}_{g,Resp}$$

Considerando que el  $\hat{R}_{g,Resp}$  toma valor 0 para casos que no responden, a partir de acá se descartan los casos que no respondieron la encuesta, reteniéndose solo los factores de expansión asociados a casos que participaron de ENSSEX 2022-2023.

En la Tabla 28 se presentan las estadísticas descriptivas del ponderador de viviendas ajustado por no respuesta. Como se puede ver, el valor mínimo se encuentra en la región de Aysén, y el máximo en la región Metropolitana. La región con menor variabilidad (100 unidades) fue la región de Aysén, con un mínimo de 18 unidades y un máximo de 118 unidades. En cambio, la

región con mayor variabilidad fue la Metropolitana, con una diferencia entre mínimo y máximo de 1.734 unidades (1.775 – 41 unidades).

Este ponderador permite estimar el total de viviendas elegibles. La suma del factor de expansión corresponde a 3.586.571 viviendas. Como referencia, en el marco muestral se tienen 5.473.126 viviendas particulares<sup>13</sup>.

Finalmente, se encontraron 7 viviendas con valores extremos en su ponderador de vivienda (con valores 5 veces o más la media del ponderador dentro del estrato). Estos valores fueron truncados y sus pesos distribuidos al interior de cada estrato entre aquellos ponderadores que se encontraron bajo el umbral, con la misma metodología explicada previamente.

### 3.4 Ajuste por probabilidad de selección de personas

Ahora, teniendo el ponderador de viviendas ajustado por no respuesta, es necesario pasar a representar a la población objetivo a nivel de individuos. Para eso, debe incorporar el ponderador de selección de personas, considerando la probabilidad de selección del encuestado en la vivienda (la cual se encuentra disponible solo para aquellas viviendas que decidieron participar de la encuesta).

#### 3.4.1 Cálculo de probabilidad de selección de personas

Teniendo definida la población objetivo de ENSSEX 2022-2023, se utilizó una estrategia de selección del encuestado totalmente aleatorio y probabilístico, donde a cada persona elegible de la vivienda (personas de 18 años y más) se le otorga una probabilidad de selección que dependió de su edad. Lo que se definió en:

$$q_l = \begin{cases} 1 & \text{si } l \in h; \text{ es } \geq 30 \text{ años} \\ 2 & \text{si } l \in h; \text{ es } < 30 \text{ años} \end{cases}$$

Así, la probabilidad de selección queda definida en definida en:

$$P_{hijk}(l|k) = \frac{q_l}{\sum q_l}$$

Donde,

---

<sup>13</sup> Esta discrepancia se explica en función de desactualizaciones del marco versus lo encontrado en terreno.

$q_l$  : Es 2 cuando persona tiene <30 años, y 1 cuando persona tiene  $\geq$  30 años.

Entonces, la probabilidad de selección de la  $l$ -ésima persona que pertenece a la muestra de ENSSEX 2022-2023 se define en:

$$P_{hijk}(l) = [P_{hij}(k)] * P_{hijk}(l|k) \quad (1)$$

$$= \left[ P_h(i) * P_{hi}(j|i) * \frac{1}{R_{h,omit}^{MZ}} * \frac{1}{R_{a,know}^{MZ}} * \frac{1}{R_{a,eleg}^{MZ}} * P_{hij}(k|j) * \frac{1}{R_{a,know}^{VIV}} \right. \\ \left. * \frac{1}{R_{a,eleg}^{VIV}} * \frac{1}{R_{g,resp}^{VIV}} \right] * P_{hijk}(l|k) \quad (2)$$

$$= \left[ \frac{1}{w''_{hijk}} \right] * P_{hijk}(l|k) \quad (3)$$

Donde,

$P_h(i)$  : Probabilidad incondicional de selección de la comuna/pseudocomuna en el estrato  $h$

$P_{hij}(k)$  : Probabilidad incondicional de la  $k$ -ésima vivienda de la manzana  $j$  de la comuna/pseudocomuna  $i$  dentro del estrato  $h$  de pertenecer a la muestra

$P_{hijk}(l)$  : Probabilidad incondicional de la  $l$ -ésima persona de la vivienda  $k$  de la manzana  $j$  de la comuna  $i$  del estrato  $h$  pertenezca a la muestra.

$P_{hi}(j|i)$  : Probabilidad de selección de la manzana  $j$  condicionada por la selección de la comuna  $i$  en el estrato  $h$

$P_{hij}(k|j)$  : Probabilidad de selección de la vivienda  $k$  condicionada por la selección de la manzana  $j$  de la comuna  $i$  del estrato  $h$

$P_{hijk}(l|k)$  : Probabilidad de selección de la persona  $l$  condicionada por la selección de la vivienda  $k$  de la manzana  $j$  de la comuna  $i$  del estrato  $h$

$R_{h,omit}^{MZ}$  : Ajuste por omisión de manzanas en el estrato  $h$

$R_{a,know}^{MZ}$  : Ajuste por elegibilidad desconocida de manzanas en las regiones  $a$

$R_{a,eleg}^{MZ}$  : Ajuste por no elegibilidad de manzanas en las regiones  $a$

$R_{a,know}^{VIV}$  : Ajuste por elegibilidad desconocida de viviendas en las regiones  $a$

- $R_{a,elec}^{VIV}$  : Ajuste por no elegibilidad de viviendas en las regiones  $a$
- $R_{g,resp}^{VIV}$  : Ajuste por no respuesta de viviendas en la celda de ajuste  $g$
- $w''_{hijk}$  : Ponderador de viviendas ajustado

La primera expresión lo que señala es que, dado que la probabilidad incondicional de las viviendas  $P_{hij}(k)$  recibió ajustes por diversas fuentes durante la selección de manzanas (omisión  $R_{h,omit}^{MZ}$ , elegibilidad desconocida  $R_{a,know}^{MZ}$  y no elegibilidad  $R_{a,elec}^{MZ}$ ) y viviendas (elegibilidad desconocida  $R_{a,know}^{VIV}$ , no elegibilidad  $R_{a,elec}^{VIV}$ , y no respuesta  $R_{g,resp}^{VIV}$ ), entonces se puede definir que la probabilidad no condicional de que la  $l$ -ésima persona haya sido seleccionada en la muestra de ENSSEX 2022-2023 se puede definir en la segunda expresión, donde se observan con claridad todos los ajustes incorporados. Con eso, la tercera expresión expone entre paréntesis cuadrados el inverso del ponderador de viviendas ajustado.

### 3.4.2 Ponderador de selección de personas

Considerando que al interior de cada vivienda se seleccionó a una persona aleatoriamente entre todas aquellas elegibles, con probabilidad de selección variable dependiendo de la edad, entonces se puede definir el ponderador de selección de persona como el producto de la multiplicación del ponderador de selección de vivienda ajustado  $w''_{hijk}$ , y el inverso de la probabilidad condicional de haber escogido a una de las personas elegibles de la vivienda. Esto puede definirse en:

$$w_{hijkl} = \frac{1}{P_{hijk}(l)}$$

$$= w''_{hijk} * \left[ \frac{1}{P_{hijk}(l|k)} \right]$$

Donde,

- $P_{hijk}(l|k)$  : Probabilidad de selección de la personas  $l$  condicionada a la selección de la vivienda  $k$  de la manzana  $j$  de la comuna  $i$  del estrato  $h$
- $P_{hijk}(l)$  : Probabilidad incondicional de selección de la persona  $l$  de la vivienda  $k$  de la manzana  $j$  de la comuna  $i$  dentro del estrato  $h$
- $w''_{hijk}$  : Ponderador de selección de viviendas ajustado

En la Tabla 29 se puede ver las estadísticas descriptivas asociadas al ponderador de selección de personas, según región. En dicha tabla se observa que la suma de dicho factor de expansión es de 7.684.381 personas, que serían las representadas hasta esta fase. El valor mínimo se encuentra en 18,12 personas en la región de Aysén, y el máximo en 9.765 personas en la región Metropolitana.

### 3.5 Ajuste a totales poblacionales

A continuación, se describe el proceso asociado al ajuste a totales poblacionales mediante calibración.

#### 3.5.1 Cálculo de totales poblacionales

Hasta este momento, los factores de expansión calculados debieran permitir que las unidades muestrales representen los stocks poblacionales. Sin embargo, debido a fallas en el Marco Muestral, elementos de la población objetivo de ENSSEX 2022-2023 podrían haber quedado excluidas por el estudio.

Debido a estos desbalances, es importante realizar un ajuste que permita corregir por ellos para que, de esta manera, la muestra resultante del estudio sea representativa de la población objetivo (mayores de 18 años de todas las regiones de Chile en áreas urbanas).

Para este propósito, se consideró que las mejores fuentes para obtener parámetros poblacionales externos son las Estimaciones y Proyecciones de Población al año 2022, con base al Censo 2017, publicadas por el Instituto Nacional de Estadísticas en diciembre del año 2019. Estas proyecciones de población permiten disponer de los totales poblacionales por sexo y edad, a nivel de país, regiones y área geográfica.

Usando este documento, para el caso de ENSSEX 2022-2023 se utilizaron como referencia dos sets de totales poblacionales:

- Celda sexo-edad: totales poblacionales definidos por 26 celdas conformadas por la combinación de 2 categorías de sexo (hombre, mujer) y 13 categorías de edad (18-24,25-29,30-34,35-39,40-44,45-49,50-54,55-59,60-64,65-69,70-74,75-79,80 y más).
- Celda región: totales poblacionales definidos por 16 celdas correspondientes a las 16 regiones del país.

Estas estimaciones entregar una población objetivo total de 13.584.611 personas, correspondiendo a mayores de 18 años de las zonas urbanas de las 16 regiones de Chile. En la Tabla 10 y Tabla 11 se exponen dichos totales poblacionales por celdas.

**Tabla 10 Totales poblacionales de celda sexo y edad a partir de Proyecciones Poblacionales al año 2022, según INE**

<b>Tramo edad</b>	<b>Hombre</b>	<b>Mujer</b>
18-24 años	863.714	842.682
25-29 años	738.064	720.256
30-34 años	779.039	761.035
35-39 años	671.742	661.556
40-44 años	608.795	612.914
45-49 años	565.064	584.133
50-54 años	515.480	549.142
55-59 años	487.897	536.971
60-64 años	418.954	477.757
65-69 años	340.782	406.829
70-74 años	243.778	309.665
75-79 años	159.352	223.604
80 años y más	175.268	330.138
<b>Total</b>	<b>6.567.929</b>	<b>7.016.682</b>

**Tabla 11 Totales poblacionales de celda región a partir de Proyecciones Poblacionales al año 2022, según INE**

<b>Región</b>	<b>Total</b>
Tarapacá	274.858
Antofagasta	514.635
Atacama	214.469
Coquimbo	532.706
Valparaíso	1.433.636
O'Higgins	580.867
Maule	655.575
Biobío	1.154.963
La Araucanía	561.315
Los Lagos	513.296
Aysén	64.871
Magallanes	130.605
Metropolitana	6.260.620
Los Ríos	230.811

Arica y	178.619
Parinacota	
Ñuble	282.765
<b>Total</b>	<b>13.584.611</b>

Adicionalmente, y considerando la importancia del nivel educacional en la descripción de fenómenos sociales y de salud, se decidió incorporar esta variable en el proceso de ajuste a totales externos. Para este propósito se estimó la población en cuatro grupos educacionales a partir de la Encuesta Nacional de Empleo para el trimestre Septiembre-Octubre-Noviembre del año 2022<sup>14</sup>, lo cual coincide en gran parte con las fechas de levantamiento de terreno de ENSSEX 2022-2023. Con ello, se estableció el siguiente set de totales poblacionales:

- Celda educación: totales poblacionales definidos por 4 celdas conformadas por 4 categorías de nivel educacional (Básica incompleta o completa, Media incompleta, Media completa, Superior incompleta o completa).

Esta estimación dio un total poblacional de 13.547.793 personas. Este total difiere del estimado según Proyecciones Poblacionales debido a que en caso de ENE se excluyeron de la estimación aquellas personas donde no se pudo conocer su nivel educacional, y por tanto, el total poblacional se encontrará marginalmente subestimado. Dado eso, los totales poblacionales de ENE se recalcularon proporcionalmente para que convergieran con el total estimado a partir de Proyecciones Poblacionales.

**Tabla 12 Totales poblacionales de celda nivel educacional a partir de Encuesta Nacional de Empleo SON 2022, ajustado por Proyecciones Poblacionales al año 2022, según INE**

Nivel educacional	Total
Básica incompleta o completa	2.221.285
Media incompleta	1.318.305
Media completa	4.504.008
Superior incompleta o completa	5.541.012
<b>Total</b>	<b>13.584.610</b>

<sup>14</sup> Se utiliza esta fuente de información por motivos de coherencia, en tanto en el proceso de calibración de ENCAVI 2015 se utilizó este mismo parámetro para el cálculo de tamaños poblacionales respecto a educación de los entrevistados. Adicionalmente, para este cálculo se excluyeron de la base de datos de ENE personas menores de 18 años, y pertenecientes a áreas rurales.

### 3.5.2 Calibración mediante método raking

Para el caso de ENSSEX 2022-2023 se optó por utilizar el método de *raking* para realizar la calibración a totales poblacionales. El método *raking* es un procedimiento de ajuste proporcional e iterativo en el cual se ajustan los valores de las variables a ponderar mediante la coincidencia de los totales de fila de la variable obtenida, con los datos de la variable poblacional elegida para posteriormente ajustar los totales de columna. Este procedimiento se repite hasta lograr la convergencia entre ambas (Kalton & Flores-Cervantes, 2003).

En concreto, se utilizó el comando *ipkraking* (Kolenikov, 2014) en Stata 17 para el desarrollo de los ponderadores calibrados utilizando el método de *raking*.

En términos generales, el algoritmo de *raking* consiste en un ciclo externo que chequea criterios de convergencia y un ciclo interno que itera sobre las variables de control. La notación multi-índice de los ponderadores intermedios,  $w_f^{k,v}$ , indica el factor de expansión de la unidad  $f$  calculado en el ciclo externo  $k$  luego de postestratificar con respecto a la variable  $v$ -ésima. De esta manera,  $k$  va desde 1 a un número predeterminado de iteraciones  $K$ ;  $v$  va desde "0" (que indica el ponderador de inicio para una determinada iteración) pasando por "1" (que indica ajuste con respecto a la primera variable de control) y llegando a "p" (que indica ajuste con respecto a la última variable de control).

Con ello, el proceso básico del *raking* consiste en 7 pasos:

1. Iniciar el contador de iteraciones  $k \leftarrow 0$  y los factores como  $w_f^{0,p}$ . (Es decir, use los factores o ponderadores base para iniciar el *raking*; el superíndice 0,p se usa solo por consistencia con la notación de los siguientes pasos).
2. Incremente el contador de la iteración  $k \leftarrow k + 1$ , y actualice los factores  $w_f^{k,0} \leftarrow w_f^{k-1,p}$ . (Es decir, use el resultado de la iteración del ciclo externo previo para inicializar los factores para el ciclo externo de iteración actual).
3. Ciclo interno: vaya desde la variable de control  $v = 1, \dots, p$ , y actualice los factores según:

$$w_f^{k,v} = \begin{cases} w_f^{k,v-1} \frac{T(X_v)}{\sum_{l \in S} w_l^{k,v-1} x_{vl}} & , x_{vf} \neq 0 \\ w_f^{k,v-1} & x_{vf} = 0 \end{cases}$$

(Es decir, postestratificar con respecto a la  $v$ -ésima variable de control.)

4. Si las discrepancias entre los totales ponderados  $\sum_{f \in S} w_f^{k,p} x_v$  y los totales de control  $T(X_v)$  están dentro de los márgenes de tolerancia especificados para todas las variables  $v = 1, \dots, p$ , se declara la convergencia y procede al paso 7.
5. Si el número de iteraciones  $k$  alcanza el límite pre-especificado  $K$ , declara no convergencia y vaya al paso 7.
6. De otra forma, regresar al paso 2. (Es decir, la precisión alcanzada de los totales objetivo es insuficiente, y se requiere más trabajo.)
7. Entregar los factores  $w_f^{k,p}$  de la etapa final como los ponderadores calibrados.

El algoritmo recién descrito en los pasos 1 a 7 corresponde a un ajuste de postestratificación que considera a cada variable de control como la variable de postestratificación y comienza el ciclo sobre cada una de estas en cada iteración.

Adicionalmente, utilizando opciones de *ipfraking* los ponderadores se truncaron al 99% para controlar valores extremos. Para detalles acerca de este proceso ver Kolenikov (2014, pág. 29-31).

Cabe mencionar que para implementar el proceso de raking se requiere que todas las variables sobre las cuales se va a calibrar tengan valores válidos y no existan datos perdidos (*missing*). Debido a ello, en procesos de supervisión y validación de los datos se estableció como parámetro mínimo para considerar una encuesta válida no tener valores perdidos en las variables de sexo, edad, región y preguntas asociadas al nivel educacional del entrevistado. Como resultado de ello, no existen valores perdidos en dichas variables en la base de datos de ENSSEX 2022-2023.

### 3.5.3 Ponderador de personas calibrado

Entonces, el factor de expansión de personas ajustado por no respuesta y calibrado quería expresado en:

$$w'_{hijkl} = w_{hijkl} * R_l^{CAL}$$

Donde;

$R_l^{CAL}$  : ajuste de calibración población total para cada persona  $l$ .

En la Tabla 30 se pueden observar las estadísticas descriptivas según región del factor de expansión de personas calibrado. Como se ve, la suma de este corresponde a 13.584.610 personas (posee una diferencia de -1 con el total poblacional externo). Su mínimo se encuentra en la región de Aysén con un valor de 10, y el máximo es de 4.650 (correspondiente al valor truncado al 99%).

### 3.6 Factor de expansión compensando por no respuesta módulo hogar

Adicionalmente, también se construyó un factor de expansión específico que dé cuenta de escenarios donde una persona no haya respondido de manera completa el instrumento, y en específico, donde no haya respondido el módulo hogar del cuestionario. Esto se piensa, en particular, para aquellas personas que no desean responder algunos módulos particulares del cuestionario, y por lo tanto, se tendrían diferentes poblaciones de inferencia (una población que contesta todo el cuestionario, y otra que no lo contesta por completo).

Para hacer frente a ello, es que se construyó un factor de expansión específico mediante el cual aquellas personas que sí respondieron completo el instrumento puedan pasar a representar a aquellas que lo respondieron parcialmente. En concreto, se realizó un ajuste explícito de no respuesta para este fenómeno.

#### 3.6.1 Modelo de no respuesta respecto módulo hogar

Para construir el modelo de no respuesta se realizan dos procesos; primero, se estima una probabilidad de responder, y segundo, se conforman celdas de ajuste en función de dicha probabilidad.

##### 3.6.1.1 Modelamiento y estimación de la probabilidad de responder

Para realizar este ajuste de no respuesta, en primera instancia es necesario estimar, para cada una de las personas que respondieron el cuestionario, una probabilidad de responder el módulo de ingreso.

Para ello, se realizó un modelo de regresión logística para modelar la probabilidad de responder a nivel de persona. Cabe destacar que se realiza este modelamiento a nivel de persona pues – a diferencia del caso de viviendas – las variables utilizadas para el modelamiento se encuentran a nivel de persona, y también de manzana.

De esta forma, la variable dependiente del modelo corresponde a si el cuestionario tiene respuestas en el módulo de hogar (valor = 1), o el módulo se encuentra sin responder (valor = 0).

Ahora, las variables independientes que se incluyeron en el modelo son tanto a nivel de manzanas, como a nivel de persona. El detalle de las variables utilizadas se expone en la siguiente tabla.

**Tabla 13 Variables independientes utilizadas para el modelo de no respuesta**

<b>Variable</b>	<b>Descripción</b>	<b>Nivel</b>
tasa_mujeres	Tasa de mujeres en la manzana, según lo reportado en Marco Muestral	Manzana
tasa_ninos	Tasa de personas entre 0 a 5 años en la manzana, según lo reportado en Marco Muestral	Manzana
tasa_adultos_mayores	Tasa de personas mayores a 65 años en la manzana, según lo reportado en Marco Muestral	Manzana
tasa_migrantes	Tasa de personas migrantes en la manzana, según lo reportado en Marco Muestral	Manzana
tasa_mat_irrec	Tasa de viviendas con materialidad irrecuperable, según lo reportado en Marco Muestral	Manzana
mujer	Indicador de que respondente es mujer (valor = 1)	Persona
edad_18_29	Indicador de que respondente tiene entre 18 y 29 años (valor = 1)	Persona
edad_30_44	Indicador de que respondente tiene entre 30 y 44 años (valor = 1)	Persona
edad_45_59	Indicador de que respondente tiene entre 45 y 59 años (valor = 1)	Persona
basica	Indicador de que respondente tiene educación básica completa o incompleta (valor = 1)	Persona
media_incom	Indicador de que respondente tiene educación media incompleta (valor = 1)	Persona
media_com	Indicador de que respondente tiene educación media completa (valor = 1)	Persona
alguna_ocupacion	Indicador de que respondente realizó alguna labor remunerada la semana pasada (valor = 1)	Persona

Las variables que resultaron predecir de mejor manera la respuesta fue la tasa de adultos mayores, de migrantes, y de viviendas con materialidad irrecuperable a nivel de manzana, y todas las variables asociadas a la persona.

Luego de ello, y a partir de un modelo optimizado (que solo consideró las variables que eran significativas para predecir la respuesta), se calculó la probabilidad de responder para cada una de las personas en el estudio.

### 3.6.1.2 Creación de celdas de ajuste de no respuesta

Al igual que en el ajuste a nivel de viviendas, se optó por utilizar la probabilidad de responder predicha de manera agrupada, y en específico, el método de clases llamado “*propensity stratification*” (Rosenbaum y Rubin (1985)).

Así, se agrupan personas en clases donde se comparten predictores, o se tiene una probabilidad de responder cercana. De esta manera, se construyeron 20 celdas ( $g$ ) de ajuste a partir de la predicción de responder. Con ello, la celda #1 agrupa personas con la probabilidad predicha de responder más baja, y la celda #20 a aquellas con la probabilidad predicha de responder más alta. En la Tabla 31 se puede observar la probabilidad de respuesta efectiva y predicha de las personas de cada una de las celdas de ajuste.

### 3.6.2 Cálculo del factor de ajuste de no respuesta de vivienda

Tal como fue mencionado anteriormente, el objeto de realizar un ajuste explícito de no respuesta consiste en la reducción de un posible sesgo de no respuesta. Esto se realiza redistribuyendo los pesos de los casos que no respondieron el módulo de ingresos, entre aquellos que sí lo hicieron. Esto se realiza al interior de cada una de las celdas de ajuste  $g = 20$ .

Para ello, se calcula una razón de ajuste en cada celda usando el ponderador de personas calibrado (truncado y suavizado). De esta manera, el numerador queda conformado por la estimación total de personas, y en el denominador la estimación del total de personas que respondió el módulo de hogar. Esto puede ser expresado con la siguiente fórmula:

$$\hat{R}_{g,Resp\_hogar} = \begin{cases} \frac{\sum_{k \in \varphi_{g,pers}} w'_{hijkl}}{\sum_{k \in \varphi_{g,pers,Resp}} w'_{hijkl}} & \text{si } l \in \text{responde módulo} \\ 0 & \text{si } l \in \text{no responde módulo} \end{cases}$$

Donde,

- $w'_{hijkl}$  : Ponderador de personas calibrado a totales poblacionales (truncado y suavizado).
- $g$  : Índice de las 20 clases (celdas) creadas para el ajuste de no respuesta.
- $\varphi_{g,pers}$  : Conjunto de personas pertenecientes a la celda  $g$ .
- $\varphi_{g,pers,Resp}$  : Conjunto de personas que respondieron el módulo, pertenecientes a la celda  $g$ .



- Celda sexo-edad: totales poblacionales definidos por 8 celdas conformadas por la combinación de 2 categorías de sexo (hombre, mujer) y 4 categorías de edad (18-29,30-44,45-59,60 y más). Estos totales se pueden ver en la Tabla 14.
- Celda región: totales poblacionales definidos por 16 celdas correspondientes a las 16 regiones del país (ver Tabla 11).
- Celda educación: totales poblacionales definidos por 4 celdas de nivel educacional (ver Tabla 12).

**Tabla 14 Totales poblacionales de celda sexo y edad para ponderador de no respuesta a módulo de hogar a partir de Proyecciones Poblacionales al año 2022, según INE**

Tramo edad	Hombre	Mujer
18-29 años	1.601.778	1.562.938
30-44 años	2.059.576	2.035.505
45-59 años	1.568.441	1.670.246
60 años y más	1.338.134	1.747.993
<b>Total</b>	<b>6.567.929</b>	<b>7.016.682</b>

Este proceso de calibración también se realizó mediante el comando de ipfraking de Stata 17. Adicionalmente, utilizando opciones de ipfraking los ponderadores se truncaron al 99% para controlar valores extremos.

Con ello, se tiene el ponderador de personas ajustado por no respuesta al módulo de hogar calibrado a totales poblacionales:

$$\begin{cases} w'''_{ijkl\_hogar} = w''_{ijkl\_hogar} * R_l^{CAL}; & \text{si persona contesta módulo hogar} \\ . & ; \text{si persona no contesta módulo hogar} \end{cases}$$

Donde;

$w''_{ijkl\_hogar}$  : Ponderador de personas ajustado por no respuesta a módulo hogar

$R_l^{CAL}$  : ajuste de calibración población total para cada persona  $l$

En la Tabla 34 se pueden ver las estadísticas descriptivas asociadas a este factor de expansión. Como se ve, su suma corresponde al total poblacional estimado de la población objetivo (13.584.610 personas). Su valor mínimo se encuentra en la región de Aysén (25), y su máximo en la región Metropolitana (15.259).

### 3.7 Cálculo de varianzas

En la base de datos de ENSSEX 2022-2023 se pueden encontrar cuatro variables que permiten identificar en forma adecuada el diseño muestral complejo de la encuesta – el factor de expansión de personas (“*w\_personas\_cal*”), el factor de expansión de personas ajustando por no respuesta al módulo de ingresos (“*w\_personas\_cal\_nr\_mod\_hogar*”), el indicador de estratos (“*varstrat*”) y el indicador de conglomerados (“*varunit*”). A continuación se describen los pasos empleados para la conformación de las unidades de cálculo de los errores estándar y se presenta un ejemplo de estimación que toma en cuenta el diseño muestral complejo de la encuesta.

#### 3.7.1 Conformación de variables de estratos y conglomerados

La varianza muestral es una medida del nivel de precisión (variabilidad) de un estadístico bajo el paradigma del muestreo repetido. Con ello, la varianza muestral viene determinada por el diseño de la muestra y el estadístico que se quiere estimar.

Siguiendo las recomendaciones en la literatura especializada (Valliant *et al.* 2013, Heeringa *et al.* 2010) y la práctica de encuestas de hogares con diseños similares en encuestas previas de MINSAL (ej: ENCAVI 2015), se optó por la creación de unidades de cómputo de errores estándares, denominadas *varstrat* y *varunit* que se incorporan en la base de datos de ENSSEX. El objetivo del desarrollo de estas unidades es doble:

1. Eliminar estratos con solo un conglomerado (a través de la combinación de estratos)
2. Eliminar conglomerados (manzanas censales) que tengan pocos entrevistados (a través de la combinación de conglomerados).

El primer objetivo permitirá que se puedan realizar estimaciones de varianza para la ENSSEX 2022-2023 a través de paquetes estadísticos como R, Stata, SPSS y SAS. El segundo objetivo permite estabilizar las estimaciones de varianza. La combinación de conglomerados con pocos entrevistados permite reducir la variabilidad de las estimaciones de varianza comparado con el uso de conglomerados que varíen demasiado en cuanto a su tamaño.

Para ello, en primera instancia, se estableció como un piso mínimo que cada conglomerado tuviera un mínimo de 50 casos entrevistados<sup>15</sup>. Luego de ello, se revisó que cada estrato tuviera al menos 2 conglomerados en su interior.

---

<sup>15</sup> Sin embargo, en algunos casos, debido a tamaños poblacionales y de muestra efectiva, se establecieron conglomerados de tamaño menor a 50, pero mayor a 30 casos.  
Departamento de Epidemiología – Oficina de Vigilancia de Enfermedades No Transmisibles y Encuestas Poblacionales

En la Tabla 15 se puede ver la distribución de varstrat y varunit por región. Como se ve, se tiene un total de 84 pseudo-estratos y 370 pseudo-conglomerados.

**Tabla 15 Distribución de varstrat y varunit, según región**

<b>Región</b>	<b>Varstrat</b>	<b>Varunit</b>
<b>Tarapacá</b>	2	11
<b>Antofagasta</b>	3	17
<b>Atacama</b>	3	9
<b>Coquimbo</b>	4	21
<b>Valparaíso</b>	6	31
<b>O'Higgins</b>	5	26
<b>Maule</b>	4	25
<b>Biobío</b>	8	34
<b>La Araucanía</b>	6	27
<b>Los Lagos</b>	6	27
<b>Aysén</b>	2	9
<b>Magallanes</b>	2	9
<b>Metropolitana</b>	24	83
<b>Los Ríos</b>	3	12
<b>Arica y Parinacota</b>	1	9
<b>Ñuble</b>	5	20
<b>Total</b>	<b>84</b>	<b>370</b>

### 3.7.2 Ejemplo de cálculo de varianza bajo Diseño Muestral Complejo

Para estimar la varianza en encuestas con un diseño muestral complejo se requiere del uso de procedimientos que reflejen de forma apropiada los principales efectos de las distintas características del diseño muestral empleado en la selección de la muestra. En la ENSSEX 2022-2023 se deben considerar tres elementos principales que dan cuenta de estas características: (1) la estratificación de las unidades primarias de muestreo; (2) la selección de unidades primarias de muestreo (conglomerados); y (3) factores de expansión que reflejan distintas probabilidades de selección, no respuesta y calibración a totales externos.

A continuación se presentan resultados sobre ello, los cuales fueron desarrollados utilizando el comando para el análisis de encuestas complejas en el software Stata17. En general, siempre que se trabaje con encuestas de diseño muestral probabilístico, se recomienda dar cuenta de todos los elementos del diseño complejo de la encuesta al momento de realizar análisis con los datos.

Para desarrollar estimaciones a partir de los datos de la ENSSEX 2022-2023 se deben seguir tres pasos:

1. **Paso 1: Definir la variable de análisis.** Para efectos de ilustración, se definieron dos variables de interés: “porcentaje de personas que declaran tener un buen o muy buen estado de salud” (variable *buena\_salud*) y “porcentaje de personas que declara haber tenido relaciones sexuales de manera voluntaria” (variable *relaciones sexuales*).
2. **Paso 2: Especificar las variables que definen el diseño complejo.** Se debe especificar la variable de estratificación o Strata (*varstrat*), la definición del conglomerado o cluster (*varunit*) y la identificación del factor de expansión (*w\_personas\_cal*). Esto, en Stata 17 se declara:

```
svyset varunit [w= w_personas_cal], strata(varstrat) singleunit(certainty)
```

3. **Paso 3: Realizar el análisis de datos.** Una vez definido el plan de muestreo, es posible desarrollar diversos tipos de análisis estadísticos que tengan en consideración el diseño complejo de la muestra de la encuesta al momento de estimar los errores de muestreo.

La Tabla 16 presenta los resultados de la estimación del porcentaje de personas que declaran tener una “buena o muy buena salud” y porcentaje de personas que declara “haber mantenido relaciones sexuales voluntarias” para el nivel país, región, sexo y tramos etarios tomando en cuenta el diseño complejo de la ENSSEX 2022-2023. A nivel nacional el porcentaje de personas que tener una buena o muy buena salud es de 66%. Respecto a variabilidad en efectos diseños, el mínimo se encuentra en la región de Tarapacá y el máximo en la región de O’Higgins.

En cuanto a la prevalencia de mantener relaciones sexuales voluntarias, la proporción a nivel país es de 93%. El mínimo de efecto diseño estimado se encuentra en la región de Atacama, y el máximo en la región de Aysén.

En la sección 0 se puede ver el código utilizado para obtener estas estimaciones.

**Tabla 16 Estimación de Efecto Diseño y Coeficiente de variación bajo Diseño Muestral Complejo para variables de interés**

Sub poblaciones	N muestral	Percepción buena o muy buena salud			Ha tenido relaciones sexuales voluntarias		
		P	DEFF	CV	P	DEFF	CV
		1	2	3	4	5	6
País	<b>20.392</b>	0,66	4,19	1,03	0,93	3,46	0,35
Tarapacá	616	0,65	0,51	2,11	0,84	4,37	3,72
Antofagasta	892	0,72	1,81	2,79	0,95	1,32	0,89
Atacama	530	0,68	1,33	3,41	0,97	0,36	0,48
Coquimbo	1.176	0,72	2,59	2,93	0,95	2,94	1,15
Valparaíso	1.713	0,70	2,29	2,37	0,95	1,40	0,68
O'Higgins	1.422	0,64	6,80	5,22	0,95	2,28	0,91
Maule	1.451	0,57	3,99	4,53	0,90	4,66	1,94
Biobío	1.869	0,76	1,64	1,67	0,89	1,49	0,99
La Araucanía	1.495	0,62	3,04	3,53	0,96	0,78	0,48
Los Lagos	1.526	0,68	3,52	3,26	0,88	7,45	2,60
Aysén	502	0,74	3,32	4,77	0,91	9,76	4,46
Magallanes	451	0,75	3,14	4,80	0,98	1,25	0,74
Metropolitana	4.392	0,63	2,94	1,98	0,94	2,52	0,60
Los Ríos	663	0,69	2,59	4,15	0,94	1,42	1,18
Arica y Parinacota	524	0,65	2,10	4,62	0,90	5,41	3,32
Ñuble	1.170	0,68	4,16	4,06	0,91	3,84	1,77
Hombre	6.838	0,72	2,49	1,78	0,94	2,60	0,51
Mujer	13.554	0,60	4,41	1,47	0,93	3,09	0,42
18-24	3.028	0,80	2,19	1,35	0,71	3,29	2,11
25-29	2.189	0,77	2,25	1,75	0,95	1,53	0,62
30-34	2.129	0,76	2,84	2,05	0,96	2,27	0,64
35-39	1.695	0,76	1,86	1,87	0,97	2,52	0,72
40-44	1.502	0,69	2,52	2,76	0,97	1,56	0,59
45-49	1.599	0,70	2,48	2,55	0,98	2,78	0,64
50-54	1.625	0,64	2,70	3,05	0,97	1,44	0,49
55-59	1.550	0,57	3,59	4,19	0,97	1,86	0,63
60-64	1.453	0,50	2,51	4,19	0,95	3,29	1,05
65-69	1.317	0,43	3,43	5,83	0,98	1,05	0,42
70-74	1.000	0,45	2,76	5,81	0,97	1,21	0,57
75-79	735	0,43	1,86	5,83	0,94	2,92	1,62
80 y más	570	0,41	2,82	8,52	0,94	1,62	1,31

#### 4 Bibliografía

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer.

Lohr, S. L. (1999). *Sampling: design and analysis*. Brooks.

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American statistical Association*, 44(247), 380-387.

Kalton, G., & Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 81-97.

Rosenbaum, P. R., & Rubin, D. B. (1985). The bias due to incomplete matching. *Biometrics*, 103-116.

Deville, J. C., & Särndal, C. E. (1994). Variance estimation for the regression imputed Horvitz-Thompson estimator. *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM*, 381.

Kreuter, F., & Valliant, R. (2007). A survey on survey statistics: What is done and can be done in Stata. *The Stata Journal*, 7(1), 1-21.

## 5 Anexos

### 5.1 Estadísticas asociadas a Factores de Expansión

**Tabla 17 Estadísticas descriptivas Ponderador de comunas, según región**

Región	N	Mínimo	Percentil 5	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Mediana	Suma
Tarapacá	3,0	1,0	1,0	1,0	1,0	1,0	1,3	1,4	1,5	1,5	1,2	4
Antofagasta	4,0	1,0	1,0	1,0	1,0	1,0	1,6	2,6	2,9	3,3	1,6	6
Atacama	4,0	1,0	1,0	1,0	1,0	1,5	2,1	2,2	2,3	2,3	1,6	6
Coquimbo	5,0	1,0	1,0	1,0	1,0	1,0	2,5	4,1	4,6	5,1	2,1	11
Valparaíso	7,0	1,0	1,0	1,0	1,6	3,2	4,9	8,3	9,7	11,0	4,0	28
O'Higgins	8,0	1,0	1,6	2,2	2,7	3,2	3,4	4,1	4,9	5,6	3,1	25
Maule	8,0	1,0	1,0	1,0	1,0	1,0	3,7	7,6	11,1	14,6	3,5	28
Biobío	12,0	1,0	1,0	1,0	1,0	1,9	3,2	4,3	4,4	4,5	2,2	27
La Araucanía	13,0	1,0	1,0	1,1	1,6	2,5	2,5	5,0	6,0	7,5	2,9	37
Los Lagos	8,0	1,0	1,0	1,0	1,0	1,6	2,3	3,2	3,6	4,1	1,9	15
Aysén	3,0	1,0	1,0	1,0	1,0	1,0	1,7	2,2	2,3	2,4	1,5	4
Magallanes	3,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	3
Metropolitana	37,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	5,5	8,6	1,5	55
Los Ríos	3,0	1,0	1,3	1,5	2,3	3,6	5,2	6,2	6,5	6,8	3,8	11

Arica y Parinacota	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1
Ñuble	7,0	1,0	1,0	1,0	1,0	1,9	3,4	4,5	5,4	6,2	2,6	2,6	18
<b>Total</b>	<b>126,0</b>	<b>1,0</b>	<b>1,0</b>	<b>1,0</b>	<b>1,0</b>	<b>1,0</b>	<b>2,7</b>	<b>4,8</b>	<b>6,5</b>	<b>14,6</b>	<b>2,2</b>	<b>2,2</b>	<b>280</b>

**Tabla 18 Estadísticas descriptivas de ajuste por omisión de manzanas, según región**

Región	Mínimo	Mediana	Máximo	Media
Tarapacá	1,03	1,03	1,22	1,06
Antofagasta	1,00	1,00	1,26	1,01
Atacama	1,00	1,16	1,19	1,13
Coquimbo	1,08	1,13	1,26	1,13
Valparaíso	1,00	1,12	1,88	1,19
O'Higgins	1,00	1,11	1,14	1,08
Maule	1,00	1,03	1,41	1,05
Biobío	1,00	1,07	1,21	1,10
La Araucanía	1,01	1,04	1,20	1,07
Los Lagos	1,00	1,00	1,13	1,04
Aysén	1,03	1,03	1,17	1,05
Magallanes	1,00	1,00	1,15	1,01
Metropolitana	1,00	1,08	1,46	1,09
Los Ríos	1,08	1,08	1,31	1,14
Arica y Parinacota	1,11	1,11	1,11	1,11
Ñuble	1,00	1,03	1,16	1,04
<b>Total</b>	<b>1,00</b>	<b>1,06</b>	<b>1,88</b>	<b>1,09</b>

**Tabla 19 Distribución de manzanas con elegibilidad desconocida, según región**

Región	Elegibilidad desconocida
Tarapacá	21
Antofagasta	11
Atacama	15
Coquimbo	14
Valparaíso	12
O'Higgins	21
Maule	20
Biobío	19
La Araucanía	40
Los Lagos	24
Aysén	6

Magallanes	13
Metropolitana	81
Los Ríos	16
Arica y Parinacota	13
Ñuble	9
<b>Total</b>	<b>335</b>

**Tabla 20 Ajuste por elegibilidad desconocida de manzanas, según región**

<b>Región</b>	<b>Ajuste elegibilidad desconocida</b>
Tarapacá	1,24
Antofagasta	1,08
Atacama	1,18
Coquimbo	1,07
Valparaíso	1,05
O'Higgins	1,10
Maule	1,08
Biobío	1,06
La Araucanía	1,14
Los Lagos	1,09
Aysén	1,08
Magallanes	1,15
Metropolitana	1,09
Los Ríos	1,19
Arica y Parinacota	1,13
Ñuble	1,04

**Tabla 21 Estadísticas descriptivas de ponderador de manzanas ajustado por elegibilidad, según región**

Región	Mínimo	Percentil 5	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Mediana	Suma
Tarapacá	20,32	20,32	20,32	20,32	20,32	31,05	31,05	31,05	61,31	25,17	2.365,70
Antofagasta	27,01	27,01	27,01	27,01	27,01	34,55	38,63	41,35	54,42	30,7	5.065,67
Atacama	35,03	35,03	35,03	35,03	35,03	44,26	56,56	60,71	60,71	41,97	3.693,25
Coquimbo	29,88	29,88	29,88	29,88	37,61	37,61	38,14	38,14	58,18	35,56	7.645,10
Valparaíso	51,15	51,15	51,15	51,54	52,58	61,37	65,12	101,56	101,56	58,34	18.378,65
O'Higgins	21,68	21,68	22,27	25,98	26,43	29,99	29,99	29,99	56,51	27,55	7.191,15
Maule	31,24	31,24	31,24	32,14	33,07	33,55	35,87	35,87	35,87	33,22	8.737,08
Biobío	26,89	26,89	26,89	34,76	42,33	49,88	53,59	53,59	60,49	41,9	13.912,37
La Araucanía	24,64	24,64	24,64	24,64	29,94	30,86	40,8	42,35	47	30,98	8.551,76
Los Lagos	20,3	20,3	22,78	22,78	22,82	26,86	26,86	29,74	29,74	24,41	6.491,75
Aysén	11,63	11,63	11,63	11,63	11,63	11,82	11,82	19,17	19,17	12,25	1.127,27
Magallanes	21,29	21,29	21,29	21,29	21,29	21,29	27,29	27,29	37,34	22,64	2.015,08
Metropolitana	24,81	24,81	37,38	55,72	67,61	74,73	75,67	78,1	122,06	62,86	49.845,45
Los Ríos	25,11	25,11	25,11	25,11	25,11	42,14	42,14	42,91	42,91	31,01	3.535,04
Arica y Parinacota	23,79	23,79	23,79	23,79	23,79	23,79	23,79	23,79	23,79	23,79	2.307,91
Ñuble	16,05	16,05	16,05	16,05	16,05	20,34	21,45	21,56	21,56	17,72	3.578,63
<b>Total</b>	<b>11,63</b>	<b>16,05</b>	<b>21,29</b>	<b>25,11</b>	<b>33,55</b>	<b>52,58</b>	<b>71,2</b>	<b>74,73</b>	<b>122,06</b>	<b>39,44</b>	<b>144.441,86</b>

**Tabla 22 Distribución de viviendas con elegibilidad desconocida, según región**

<b>Región</b>	<b>Elegibilidad desconocida</b>
Tarapacá	0
Antofagasta	0
Atacama	0
Coquimbo	0
Valparaíso	0
O'Higgins	1
Maule	0
Biobío	0
La Araucanía	1
Los Lagos	0
Aysén	0
Magallanes	0
Metropolitana	0
Los Ríos	0
Arica y Parinacota	0
Ñuble	0
<b>Total</b>	<b>2</b>

**Tabla 23 Distribución de viviendas no elegibles, según región**

<b>Región</b>	<b>No elegibles</b>
Tarapacá	0
Antofagasta	0
Atacama	9
Coquimbo	0
Valparaíso	0
O'Higgins	2
Maule	0
Biobío	0
La Araucanía	0
Los Lagos	2
Aysén	0
Magallanes	0
Metropolitana	0
Los Ríos	4
Arica y Parinacota	1
Ñuble	0
<b>Total</b>	<b>18</b>

**Tabla 24 Estadísticas descriptivas de ponderador de viviendas ajustado por elegibilidad, según región**

Región	Mínimo	Percentil 5	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Media	Suma
Tarapacá	20,44	20,44	20,44	31,8	44,96	69,96	89,04	114,49	414,13	57,37	53.925,07
Antofagasta	27,01	32,42	38	51,32	64,83	89,14	114,01	136,47	216,1	73,27	120.890,85
Atacama	35,03	42,04	49,05	63,06	87,59	113,12	154,15	158,37	218,56	92,91	80.923,43
Coquimbo	29,88	37,61	41,37	52,66	67,78	95,36	124,12	160,53	218,15	76,64	164.784,82
Valparaíso	51,15	52,15	61,38	76,72	117,22	168,26	257,65	347,81	735,91	146,33	460.931,75
O'Higgins	21,68	27,48	29,99	38,99	54,32	77,99	122,13	178,65	386,93	70,89	184.805,05
Maule	31,68	32,14	33,96	45,62	63,74	88,7	123,54	161,02	432,59	75,75	199.219,62
Biobío	27,03	34,76	43,25	62,18	88,54	118,2	165,08	203,63	594,96	99,67	330.892,63
La Araucanía	24,64	30,86	36,95	49,27	65,81	89,75	122,39	157,39	446,68	77,09	212.699,91
Los Lagos	22,78	24,44	29,32	41,63	56,96	77,47	102,63	115,49	239,23	63,18	167.921,09
Aysén	11,63	11,63	13,96	19,78	29,09	41,5	49,65	52,36	87,26	30,73	28.275,85
Magallanes	21,29	21,29	23,42	29,81	38,32	48,97	63,87	76,65	171,95	43,05	38.317,92
Metropolitana	24,81	56,38	71,82	93,45	131,53	179,34	302,67	432,71	1.031,58	168,61	1.337.092,04
Los Ríos	25,77	25,77	30,93	46,35	61,85	81,53	120,14	133,02	320,72	70,79	80.418,93
Arica y Parinacota	23,79	23,79	23,79	28,55	42,83	59,48	88,03	95,17	166,55	49,76	48.213,88
Ñuble	16,05	19,26	20,34	24,52	35,31	44,95	57,22	64,21	162,12	38,25	77.257,63
<b>Total</b>	<b>11,63</b>	<b>25,68</b>	<b>32,23</b>	<b>47,59</b>	<b>73,6</b>	<b>117,21</b>	<b>176,2</b>	<b>245,97</b>	<b>1.031,58</b>	<b>97,99</b>	<b>3.586.570,49</b>

**Tabla 25 Distribución de resultados de visitas a viviendas, según región**

<b>Región</b>	<b>Aplicadas</b>	<b>Rechazos</b>	<b>No contacto</b>	<b>Total</b>
Tarapacá	616	263	61	940
Antofagasta	892	685	73	1.650
Atacama	530	269	72	871
Coquimbo	1.176	903	71	2.150
Valparaíso	1.713	1.224	213	3.150
O'Higgins	1.422	838	347	2.607
Maule	1.451	1.070	109	2.630
Biobío	1.869	1.230	221	3.320
La Araucanía	1.495	980	284	2.759
Los Lagos	1.526	1.018	114	2.658
Aysén	502	329	89	920
Magallanes	451	334	105	890
Metropolitana	4.392	2.965	573	7.930
Los Ríos	663	380	93	1.136
Arica y Parinacota	524	363	82	969
Ñuble	1.170	704	146	2.020
<b>Total</b>	<b>20.392</b>	<b>13.555</b>	<b>2.653</b>	<b>36.600</b>

**Tabla 26 Tasa de respuesta efectiva y predicha en celdas de ajuste de no respuesta**

<b>Celdas</b>	<b>Tasa de respuesta efectiva</b>	<b>Tasa de respuesta predicha</b>
1	0,47	0,48
2	0,51	0,51
3	0,58	0,52
4	0,50	0,53
5	0,62	0,53
6	0,53	0,53
7	0,49	0,54
8	0,47	0,54
9	0,44	0,55
10	0,55	0,55
11	0,58	0,55
12	0,55	0,55
13	0,57	0,55
14	0,49	0,55
15	0,58	0,55
16	0,50	0,56
17	0,60	0,56
18	0,51	0,56
19	0,68	0,56
20	0,64	0,56
21	0,53	0,56
22	0,60	0,56
23	0,65	0,56
24	0,51	0,56
25	0,68	0,56
26	0,44	0,57
27	0,61	0,57
28	0,55	0,57
29	0,65	0,57
30	0,49	0,57
31	0,54	0,57
32	0,66	0,57
33	0,55	0,57

34	0,50	0,57
35	0,55	0,58
36	0,56	0,58
37	0,55	0,58
38	0,59	0,59
39	0,59	0,59
40	0,62	0,60

**Tabla 27 Valor de la razón de ajuste de no respuesta por celda de ajuste**

<b>Celdas</b>	<b>N</b>	<b>Razón de ajuste</b>
1	920	2,07
2	919	2,29
3	920	1,56
4	920	1,83
5	919	1,59
6	920	1,88
7	920	1,93
8	920	1,72
9	919	2,14
10	920	1,69
11	920	1,72
12	920	2,08
13	920	1,71
14	919	1,78
15	920	1,64
16	918	1,87
17	920	1,82
18	920	2,06
19	920	1,46
20	920	1,45
21	919	1,86
22	918	1,58
23	908	1,43
24	910	1,73

25	909	1,59
26	909	2,03
27	910	1,64
28	910	1,66
29	910	1,58
30	909	1,91
31	910	1,79
32	906	1,56
33	910	1,72
34	908	1,82
35	910	1,50
36	910	1,79
37	910	1,70
38	910	1,63
39	910	1,58
40	910	1,47

**Tabla 28 Estadísticas descriptivas de ponderador de viviendas ajustado por no respuesta, por región**

Región	Mínimo	Percentil 5	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Media	Suma
Tarapacá	30,61	35,18	42,39	62,78	93,37	126,66	186,41	215,53	705,82	109,38	67.379,13
Antofagasta	51,47	55,8	65,09	97,38	119,86	166,6	207,6	269,95	459,39	136,57	121.820,87
Atacama	55,52	83,88	90,09	107,85	155	200,06	245,53	300,09	353,49	163,04	86.408,90
Coquimbo	49,04	63,75	71,17	90,24	122,19	166,74	253,88	308,15	377,34	141,09	165.918,91
Valparaíso	74,75	97,44	109,28	148,85	209,45	317,47	509,31	702,91	1.498,57	281,52	482.248,52
O'Higgins	37,24	48,8	57,17	73,21	102,9	155,26	251,53	371,29	666,04	137,37	195.345,46
Maule	47,45	65,88	70,05	85,52	114,62	152,99	217,54	273,92	647,95	134,23	194.769,86
Biobío	42,48	66,7	85,87	121,79	160,08	201,07	278,27	344,82	1.003,37	177,55	331.848,46
La Araucanía	35,82	54,65	69,19	87,66	117,35	161,45	203,74	268,25	688,36	131,88	197.165,69
Los Lagos	32,55	41,72	49,31	70,42	96,82	128,95	183,07	221,75	398,23	109,42	166.971,86
Aysén	18,12	18,49	21,74	35,06	48,91	72,46	82,13	98,21	117,7	53,41	26.813,05
Magallanes	38,92	42,82	49,24	59,67	73,96	101,2	115,29	152,25	356,66	87,76	39.581,77

Metropolitana	40,65	110,08	134,24	180,37	242,8	329,6	505,2	605,44	1.775,63	294,23	1.292.254,63
Los Ríos	48,01	57,61	62,68	86,93	117,97	148,82	205,1	360,05	508,22	135,98	90.153,14
Arica y Parinacota	33,99	41,21	46,58	57,78	82,49	123,73	144,47	229,17	311,12	94,53	49.531,64
Ñuble	25,3	33,38	33,97	46	61,27	78,41	101,47	124,1	328,56	66,97	78.358,61
<b>Total</b>	<b>18,12</b>	<b>48,12</b>	<b>60,68</b>	<b>87,5</b>	<b>133,54</b>	<b>210,4</b>	<b>321,66</b>	<b>442,44</b>	<b>1.775,63</b>	<b>175,88</b>	<b>3.586.570,49</b>

**Tabla 29 Estadísticas descriptivas Ponderador de selección de personas, según región**

Región	Mínimo	Percentil 5	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Media	Suma
Tarapacá	30,61	52,54	61,22	84,78	141,32	271,30	430,05	764,67	3.529,08	236,34	145.583,17
Antofagasta	51,47	67,65	83,15	109,98	167,15	259,96	413,69	532,52	1.655,13	216,24	192.886,19
Atacama	71,99	90,09	105,48	153,03	210,96	335,52	600,63	769,46	2.474,45	299,41	158.686,70
Coquimbo	49,04	72,19	86,28	122,19	189,24	356,46	598,20	756,95	1.827,74	279,04	328.145,76
Valparaíso	74,75	134,61	171,94	265,74	444,15	784,91	1.392,95	1.860,60	8.991,41	668,64	1.145.388,16
O'Higgins	38,58	64,18	84,87	117,13	201,09	353,07	605,32	917,02	3.330,22	307,02	436.576,12
Maule	47,45	73,54	89,59	122,57	197,40	336,84	555,13	744,43	2.455,52	280,16	406.518,72
Biobío	42,48	87,66	117,64	171,73	281,27	406,87	643,21	809,50	5.016,84	347,39	649.270,93
La Araucanía	46,48	84,16	99,99	154,90	246,02	410,26	610,18	801,03	4.130,14	326,77	488.527,48
Los Lagos	32,55	55,33	68,73	94,82	141,44	220,76	375,89	496,58	1.586,58	189,64	289.384,96
Aysén	18,12	20,34	29,46	44,37	66,11	105,07	162,55	226,86	408,49	85,36	42.851,12

Magallanes	38,92	49,24	56,61	66,30	99,45	152,25	245,22	353,99	1.426,65	140,10	63.186,94
Metropolitana	40,65	160,50	196,63	283,09	482,54	796,16	1.290,90	1.775,63	9.765,00	655,66	2.879.643,62
Los Ríos	48,01	72,01	86,93	111,76	188,03	301,63	537,68	835,32	2.184,81	272,90	180.933,42
Arica y Parinacota	33,99	54,38	73,53	108,76	185,59	288,95	466,83	664,17	1.655,21	247,05	129.455,95
Ñuble	25,30	34,46	42,82	61,17	96,13	151,78	244,67	314,50	1.363,14	125,93	147.341,36
<b>Total</b>	<b>18,12</b>	<b>61,84</b>	<b>80,12</b>	<b>129,79</b>	<b>235,33</b>	<b>445,32</b>	<b>798,47</b>	<b>1.147,52</b>	<b>9.765,00</b>	<b>376,83</b>	<b>7.684.380,61</b>

**Tabla 30 Estadísticas descriptivas Ponderador de selección de personas calibrado, según región**

Región	Mínimo	Percentil 5	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Mediana	Suma
Tarapacá	26	69	87	138	262	502	1.019	1.450	4.650	446	274.858
Antofagasta	53	135	165	251	401	685	1.119	1.508	4.650	577	514.635
Atacama	45	86	109	175	277	464	828	1.164	3.761	405	214.469
Coquimbo	39	81	101	165	286	565	1.028	1.307	4.650	453	532.706
Valparaíso	35	127	173	301	530	1.005	1.839	2.707	4.650	837	1.433.636
O'Higgins	22	63	85	143	247	478	874	1.302	4.650	408	580.867

Maule	41	94	121	192	309	533	853	1.232	4.650	452	655.575
Biobío	42	127	167	266	447	747	1.218	1.678	4.650	618	1.154.963
La Araucanía	28	78	98	155	257	457	765	1.000	4.162	375	561.315
Los Lagos	21	68	91	139	237	399	666	972	3.059	336	513.296
Aysén	10	26	35	58	98	164	260	331	1.153	129	64.871
Magallanes	39	65	82	118	181	336	545	873	2.647	290	130.605
Metropolitana	68	272	359	565	1.036	1.890	3.216	4.480	4.650	1.425	6.260.619
Los Ríos	33	66	78	121	237	418	734	1.041	4.285	348	230.811
Arica y Parinacota	20	62	90	139	237	405	657	1.004	3.239	341	178.619
Ñuble	24	47	63	95	167	307	483	627	2.164	242	282.765
<b>Total</b>	<b>10</b>	<b>77</b>	<b>106</b>	<b>187</b>	<b>372</b>	<b>776</b>	<b>1.572</b>	<b>2.371</b>	<b>4.650</b>	<b>666</b>	<b>13.584.610</b>

**Tabla 31 Probabilidad de respuesta efectiva y predicha en celdas de ajuste de no respuesta para módulo de hogar**

<b>Celdas</b>	<b>Tasa efectiva</b>	<b>Probabilidad predicha</b>
1	0,21	0,22
2	0,23	0,25
3	0,27	0,26
4	0,29	0,27
5	0,30	0,28
6	0,30	0,29
7	0,28	0,30
8	0,31	0,31
9	0,30	0,32
10	0,32	0,32
11	0,34	0,33
12	0,32	0,34
13	0,36	0,35
14	0,35	0,36
15	0,36	0,37
16	0,41	0,38
17	0,40	0,39
18	0,38	0,40
19	0,43	0,42
20	0,44	0,46

**Tabla 32 Valor de la razón de ajuste de no respuesta a módulo de hogar por celda de ajuste**

<b>Celdas</b>	<b>N</b>	<b>Razón de ajuste</b>
1	1020	4,58
2	1020	5,03
3	1020	4,30
4	1020	3,76
5	1020	3,66
6	1020	3,39
7	1020	4,26
8	1020	3,73
9	1020	3,97
10	1020	3,08
11	1020	3,47
12	1020	3,84
13	1019	3,25
14	1019	2,89
15	1019	3,50
16	1019	2,57
17	1019	2,53
18	1019	3,21
19	1019	2,43
20	1019	2,51

**Tabla 33 Estadísticas descriptivas de ponderador de personas ajustado por no respuesta a módulo de hogar, por región**

Región	Mínimo	Percentil 15	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Mediana	Suma
Tarapacá	279	383	448	715	1.360	2.173	4.466	5.779	15.781	1.986	299.846
Antofagasta	180	363	533	833	1.530	2.448	4.403	5.247	18.839	2.138	453.162
Atacama	130	319	385	580	962	1.583	2.345	3.664	8.974	1.287	400.185
Coquimbo	130	313	375	707	1.298	2.361	3.711	4.945	8.609	1.780	272.359
Valparaíso	137	370	502	861	1.456	2.855	5.157	7.337	16.115	2.315	1.870.350
O'Higgins	75	209	281	467	809	1.424	2.767	4.193	16.254	1.300	694.196
Maule	131	309	372	598	996	1.591	2.483	3.189	9.972	1.287	472.189
Biobío	160	412	554	856	1.399	2.449	4.406	6.130	19.823	2.096	1.056.267
La Araucanía	78	223	304	475	834	1.478	2.578	3.639	15.511	1.222	1.032.181
Los Lagos	53	183	261	432	757	1.320	2.179	2.931	6.738	1.043	585.894
Aysén	32	87	111	168	323	509	818	1.186	4.222	422	80.566
Magallanes	147	207	274	398	657	1.027	1.637	2.269	7.804	918	185.420
Metropolitana	380	955	1.273	2.114	3.603	6.639	10.853	13.470	23.386	5.014	5.339.691
Los Ríos	101	246	318	494	846	1.483	3.035	4.143	12.232	1.329	216.564
Arica y Parinacota	92	188	238	429	700	1.142	1.839	2.338	9.228	963	214.843
Ñuble	96	175	215	393	666	1.192	1.765	2.476	7.593	919	410.896
<b>Total</b>	<b>32</b>	<b>237</b>	<b>337</b>	<b>573</b>	<b>1.117</b>	<b>2.279</b>	<b>4.716</b>	<b>7.105</b>	<b>23.386</b>	<b>2.016</b>	<b>13.584.610</b>

**Tabla 34 Estadísticas descriptivas de ponderador de personas ajustado por no respuesta a módulo de hogar y calibrado, por región**

Región	Mínimo	Percentil 15	Percentil 10	Percentil 25	Mediana	Percentil 75	Percentil 90	Percentil 95	Máximo	Mediana	Suma
Tarapacá	250	328	403	653	1.227	1.930	4.096	5.408	15.094	1.820	274.858
Antofagasta	201	411	622	969	1.797	2.814	4.931	6.261	15.123	2.428	514.635
Atacama	68	171	207	306	498	833	1.279	1.894	4.892	690	214.469
Coquimbo	252	604	731	1.343	2.521	4.823	7.329	9.257	14.493	3.482	532.706
Valparaíso	101	280	386	653	1.102	2.237	4.026	5.505	12.733	1.774	1.433.636
O'Higgins	62	168	236	389	680	1.167	2.269	3.458	13.150	1.088	580.867
Maule	176	421	497	845	1.354	2.180	3.525	4.443	13.348	1.786	655.575
Biobío	184	452	602	934	1.568	2.733	4.562	6.928	15.123	2.292	1.154.963
La Araucanía	46	123	165	257	452	816	1.421	1.992	8.360	664	561.315
Los Lagos	45	162	228	372	657	1.165	1.919	2.546	6.006	913	513.296
Aysén	25	68	86	140	253	399	638	958	3.554	340	64.871
Magallanes	102	151	187	290	450	730	1.173	1.649	5.900	647	130.605
Metropolitana	465	1.188	1.573	2.559	4.414	8.363	13.319	15.122	15.123	5.879	6.260.619
Los Ríos	114	262	313	542	901	1.597	3.205	4.092	13.823	1.416	230.811
Arica y Parinacota	74	165	192	342	583	1.000	1.564	2.028	7.680	801	178.619
Ñuble	67	119	153	263	437	824	1.244	1.717	5.687	633	282.765
<b>Total</b>	<b>25</b>	<b>173</b>	<b>241</b>	<b>443</b>	<b>951</b>	<b>2.233</b>	<b>5.153</b>	<b>8.111</b>	<b>15.123</b>	<b>2.016</b>	<b>13.584.610</b>

## 5.2 Cálculo de prevalencias, efecto diseño y coeficiente de variación

```

use "base.dta"

* Declaración diseño complejo

svyset varunit [w= w_personas_cal], strata(varstrat) singleunit(certainty)

*** Buena o muy buena salud

* Total

svy: prop buena_salud
estat effects, deff
estat cv

* Por región

svy: prop buena_salud if region==1
estat effects, deff
estat cv

svy: prop buena_salud if region==2
estat effects, deff
estat cv

svy: prop buena_salud if region==3
estat effects, deff
estat cv

svy: prop buena_salud if region==4
estat effects, deff
estat cv

svy: prop buena_salud if region==5
estat effects, deff
estat cv

```

```

svy: prop buena_salud if region==6
estat effects, deff
estat cv

svy: prop buena_salud if region==7
estat effects, deff
estat cv

svy: prop buena_salud if region==8
estat effects, deff
estat cv

svy: prop buena_salud if region==9
estat effects, deff
estat cv

svy: prop buena_salud if region==10
estat effects, deff
estat cv

svy: prop buena_salud if region==11
estat effects, deff
estat cv

svy: prop buena_salud if region==12
estat effects, deff
estat cv

svy: prop buena_salud if region==13
estat effects, deff
estat cv

svy: prop buena_salud if region==14
estat effects, deff
estat cv

svy: prop buena_salud if region==15
estat effects, deff

```

```

estat cv

svy: prop buena_salud if region==16
estat effects, deff
estat cv

* Por sexo

svy: prop buena_salud if sexo==1
estat effects, deff
estat cv

svy: prop buena_salud if sexo==2
estat effects, deff
estat cv

* Por tramo de edad

svy: prop buena_salud if tramo_edad==1
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==2
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==3
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==4
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==5
estat effects, deff
estat cv

```

```

svy: prop buena_salud if tramo_edad==6
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==7
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==8
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==9
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==10
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==11
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==12
estat effects, deff
estat cv

svy: prop buena_salud if tramo_edad==13
estat effects, deff
estat cv

*** Relaciones sexuales

* Total

svy: prop relaciones_sexuales
estat effects, deff

```

```

estat cv

* Por región

svy: prop relaciones_sexuales if region==1
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==2
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==3
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==4
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==5
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==6
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==7
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==8
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==9
estat effects, deff

```

```

estat cv

svy: prop relaciones_sexuales if region==10
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==11
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==12
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==13
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==14
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==15
estat effects, deff
estat cv

svy: prop relaciones_sexuales if region==16
estat effects, deff
estat cv

* Por sexo

svy: prop relaciones_sexuales if sexo==1
estat effects, deff
estat cv

svy: prop relaciones_sexuales if sexo==2
estat effects, deff

```

```

estat cv

* Por tramo de edad

svy: prop relaciones_sexuales if tramo_edad==1
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==2
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==3
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==4
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==5
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==6
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==7
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==8
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==9
estat effects, deff

```

```
estat cv

svy: prop relaciones_sexuales if tramo_edad==10
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==11
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==12
estat effects, deff
estat cv

svy: prop relaciones_sexuales if tramo_edad==13
estat effects, deff
estat cv
```