

Encuesta de Bienestar Social



## Metodología de Diseño Muestral





*Serie Metodología Encuesta de Bienestar Social – EBS 2023*  
**Metodología de Diseño Muestral**  
Noviembre de 2024

Subdepartamento de Diseño de Marcos y Muestras  
Departamento de Metodologías e Innovación Estadística  
Instituto Nacional de Estadística

División Observatorio Social  
Subsecretaría de Evaluación Social  
Ministerio de Desarrollo Social y Familia

Informe editado por el Ministerio de Desarrollo Social y Familia, a partir del Informe metodológico de la Encuesta de Bienestar Social, EBS 2023 entregado por el Instituto Nacional de Estadísticas (INE) en el marco del convenio de colaboración. Participaron en la elaboración de este documento: Marcela Pizarro, Jonathan Nuñez, Dafne Jaime y la jefatura de departamento de Investigación Social.

**ÍNDICE**

I.	ANTECEDENTES .....	6
II.	CARACTERÍSTICAS GENERALES DEL DISEÑO MUESTRAL .....	8
II.1.	Objetivo .....	8
II.2.	Población objetivo .....	8
II.3.	Unidad de información .....	8
II.4.	Marco muestral .....	8
II.5.	Dominios de estudio .....	9
II.6.	Objetivo de precisión y tamaño de muestra .....	9
II.7.	Estrategia muestral .....	9
III.	ESTIMACIÓN DEL TAMAÑO MUESTRAL .....	10
III.1.	Etapas para la estimación del tamaño muestral para EBS 2023 .....	11
III.2.	Parámetros utilizados para el cálculo del tamaño muestral .....	14
III.3.	Metodología de cálculo del tamaño muestral utilizando muestras complejas .....	15
III.4.	Resumen proceso de simulación .....	17
III.5.	Escenario escogido .....	20
III.5.1.	Escenario 2.2 .....	20
III.6.	Tamaño muestral propuesto .....	21
III.7.	Errores esperados .....	23
IV.	SELECCIÓN DE UNIDADES MUESTRALES .....	24
V.	CONSTRUCCIÓN FACTORES DE EXPANSIÓN .....	26
V.1.	Factor de expansión de primera fase .....	27
V.2.	Factor de expansión de segunda fase .....	28
V.2.1.	Ponderador de selección de personas como inverso de la probabilidad de selección .....	28
V.2.2.	Producto del factor de primera y segunda fase (Ponderador correspondiente al producto entre los ponderadores de primera y segunda fase) .....	29
V.3.	Ajustes por no contacto, elegibilidad y no respuesta .....	29
V.3.1.	Ajuste por no contacto .....	29
V.3.2.	Ajuste por no elegibilidad .....	32
V.3.3.	Ajuste por no respuesta .....	32
V.4.	Suavizamiento .....	34
V.4.1.	Método de contracción a la media (CM) .....	35
V.4.2.	Método R-K .....	35
V.4.3.	Método mixto .....	35
V.5.	Calibración de factores de expansión .....	36
V.6.	Redondeo Probabilístico .....	38
VI.	CÁLCULO Y EVALUACIÓN DE LAS ESTIMACIONES .....	39
VI.1.	Estimación de la varianza compleja .....	39
VI.2.	Variables que identifican el diseño muestral complejo .....	40



VI.3. Estadísticos asociados a los parámetros de interés.....	40
VII. ANEXOS.....	43
VII.1. Anexo N°1: Códigos de Disposición Final de casos EBS 2023.....	43
Anexo N°2: Sintaxis en R para la estimación de las variables de interés .....	44
REFERENCIAS.....	49

### ÍNDICE DE TABLAS

Tabla III.1: Resumen de parámetros utilizados y tamaños y errores resultantes según escenario, EBS 2023.....	19
Tabla III.2: Estimación de tamaños muestrales según proyecciones de población al 15 de septiembre de 2023 y total de personas de Casen 2022 según base actualizada al 10 de mayo de 2023.....	20
Tabla III.3: Tamaños y errores finales a nivel nacional, regional, urbano-rural, EBS 2023 .....	22
Tabla III.4: Tamaños y errores EBS 2021 y propuesta EBS 2023. Nacional Urbano - Rural.....	22
Tabla III.5: Errores esperados asociados al parámetro de interés según tamaño muestral objetivo .....	23
Tabla IV.1: Distribución del número de personas por región y área .....	25
Tabla V.1. Cantidad de personas según tipo de elegibilidad, EBS 2023.....	29
Tabla V.2. Variables utilizadas en el modelo de no contacto EBS 2023.....	30
Tabla V.3. Veintiles ajuste por no contacto EBS 2023.....	31
Tabla V.4. Variables utilizadas en el modelo de no respuesta EBS 2023 .....	33
Tabla V.5. Veintiles ajuste por no respuesta EBS 2023 .....	34
Tabla V.6. Proyecciones de población al 30 de octubre 2023 según región y área .....	37
Tabla V.7. Distribución de la muestra y tamaño poblacional expandido por región y área, EBS 2023 .....	38
Tabla VI.1 Errores muestrales por región y área para la variable Insatisfacción con la vida, EBS 2023 .....	41
Tabla VI.2. Errores muestrales por región y área para la variable balance afectivo negativo, EBS 2023 .....	42

### ÍNDICE DE CUADROS

Cuadro 1. Parámetros utilizados para el cálculo de tamaño muestral actual .....	14
Cuadro 2: Parámetros comunes utilizados en escenarios EBS 2023 .....	18

### ÍNDICE DE FIGURAS

Figura V.1. Diagrama de procesos de desarrollo de factor de expansión EBS 2023.....	27
Figura V.2. Ordenamiento de factores de expansión atípicos y no atípicos.....	35



## I. ANTECEDENTES

La Encuesta de Bienestar Social (en adelante EBS) es un proyecto del Ministerio de Desarrollo Social y Familia (MDSF) cuyo instrumento fue diseñado por la Subsecretaría de Evaluación Social y se basa en el modelo de bienestar planteado por la OCDE y en el enfoque de las Capacidades de Amartya Sen<sup>1</sup>.

El propósito de la EBS 2023 es generar información multidimensional, respecto al bienestar que experimentan las personas, la que complemente la información capturada en la Encuesta de Caracterización Socioeconómica Nacional, Casen, cuyo trabajo de campo ocurrió durante el año 2022, y así disponer de mayores insumos para el diseño y evaluación de la política social del país, lo cual es una función del Ministerio de Desarrollo Social y Familia.

La encuesta es un instrumento importante para el diseño, implementación, seguimiento y evaluación de políticas y programas sociales ejecutados en el país, especialmente en materias relacionadas con la percepción y oportunidades de la ciudadanía, y de los aspectos de calidad de vida que van más allá de lo material.

La EBS es una encuesta bifásica de Casen que se desarrolla generalmente de manera bianual con el objeto de dar continuidad a la serie de información sobre la situación socioeconómica de los hogares y la población del país. Lo anterior significa que, desde la base de los hogares encuestados en Casen, se realiza la selección de personas de 18 años o más que participarán en la encuesta. Así, en el análisis y estudio de los datos levantados en esta encuesta, se incluirá información recopilada en la encuesta Casen, lo que permitirá disponer de una medición multidimensional del bienestar de la población.

La muestra de la EBS es representativa a nivel nacional, nacional urbana, nacional rural y regional. El trabajo de campo se llevó a cabo durante el año 2023, para lo cual, el Ministerio de Desarrollo Social y Familia (MDSF) y el Instituto Nacional de Estadísticas (INE) establecieron un programa de trabajo conjunto para la elaboración del diseño muestral de la encuesta que tuvo inicio el año 2023 y se extiende hasta el año 2024.

El presente informe es parte de los productos entregables del convenio INE-MDSF y presenta el diseño metodológico del diseño muestral elaborado para EBS con el fin de responder a los objetivos propios de la encuesta. La metodología del diseño muestral de EBS 2023, describe el conjunto de opciones metodológicas y de procedimientos estadísticos adoptados a lo largo del proceso de planificación, que son considerados los más apropiados para seleccionar una muestra probabilística, estadísticamente representativa de las personas de 18 años o más de Chile y que permita estimar la tasa de insatisfacción con la vida, junto con otros indicadores de bienestar, a nivel nacional, nacional urbano, nacional rural y para cada una de las regiones que conforman el territorio nacional.

Este informe corresponde a la Entrega N°9 establecida en el convenio INE-MDSF que contempla el desarrollo de los siguientes capítulos:

---

<sup>1</sup> Encuesta de Bienestar Social: <http://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-bienestar-social>



- Características generales del diseño muestral
- Estimación del tamaño muestral
- Selección de unidades muestrales
- Construcción de factores de expansión
- Cálculo y evaluación de las estimaciones



## **II. CARACTERÍSTICAS GENERALES DEL DISEÑO MUESTRAL**

### **II.1. Objetivo**

---

La EBS tiene por objetivo generar información sobre cómo las personas evalúan su propia vida respecto a diversas dimensiones, generando una batería de indicadores de bienestar que complementen la información entregada por la encuesta Casen 2022, así como también, actualizar el diagnóstico proporcionado por EBS 2021 sobre el bienestar y las oportunidades que tienen las personas en el país y, además, obtener indicadores oficiales en diferentes dimensiones del bienestar.

### **II.2. Población objetivo**

---

La población objetivo es el universo de personas de 18 años o más, cumplidos a la fecha de realización de Casen 2022, que habitan el territorio nacional, excluyendo las áreas especiales definidas por el INE y representadas por Casen 2022.

### **II.3. Unidad de información**

---

La EBS recoge información directamente de las personas de 18 años o más lo que permite conocer su percepción y oportunidades que enfrentan, así como de los aspectos de calidad de vida que van más allá de lo meramente material, al mismo tiempo que se obtiene información sobre la composición de sus hogares y de sus variables sociodemográficas.

### **II.4. Marco muestral**

---

Las personas de 18 años o más entrevistadas en EBS 2023 son una muestra que proviene de integrantes de los hogares efectivamente entrevistados por la Encuesta Casen 2022. De esta manera, ambas encuestas comparten el mismo marco muestral basado en el Censo de Población y Viviendas de 2017 y actualizado al año 2020 (MMV 2020).

El marco muestral de viviendas utilizado en Casen 2022 fue elaborado a partir del MMV 2020 y está constituido por unidades primarias de muestreo (UPM) que corresponden a áreas geográficas homogéneas, en términos del número de viviendas particulares que las conforman, excluyendo viviendas de temporada y colectivas.

Las UPM presentan una estratificación geográfica, dada por la división político-administrativa, según la conformación región-provincia-comuna y desagregada, además según la división censal, clasificándolas en áreas urbanas y rurales. A esta estratificación se suma la clasificación



socioeconómica (con niveles bajo, medio y alto) construida a partir de un modelo que considera variables del Censo 2017<sup>2</sup>.

A través de las UPM se accede a las viviendas y hogares, cuyos integrantes son principalmente el objeto final de estudio. Para la selección de viviendas o unidades secundarias de muestreo (USM) de Casen 2022 se considera la actualización de las UPM seleccionadas en primera etapa donde, 60% fueron actualizadas a través de la verificación de direcciones en gabinete, a partir de los listados obtenidos de Precenso 2016 y utilizando imágenes satelitales, Certificados de Recepción Final de conjuntos habitacionales, servidores de mapas, entre otros. En el 40% restante la actualización corresponde a la enumeración en terreno donde, a partir de los listados de Precenso, se recorren las UPM verificando y/o añadiendo direcciones de nuevas edificaciones.

## **II.5. Dominios de estudio**

---

La encuesta EBS 2023 se encuentra diseñada para obtener indicadores confiables con errores de muestreo aceptables para los niveles nacional, nacional urbano, nacional rural y regional.

## **II.6. Objetivo de precisión y tamaño de muestra**

---

Considerando un levantamiento telefónico, con un tamaño objetivo ajustado de 12.369 personas, se espera obtener estimaciones del parámetro de interés con un error absoluto nacional de 1,0% y, un error relativo, de 8,0%. Con este tamaño los errores absolutos regionales no superan 3,8% y, los errores relativos, 33,6%. Pese a que existen errores relativos elevados, a partir del error estándar se aprecia que los valores a nivel regional cumplen con el estándar de calidad que utiliza el INE para la publicación de parámetros de interés de sus estudios.

## **II.7. Estrategia muestral**

---

El diseño muestral de EBS 2023 implica la obtención de una muestra probabilística bifásica. La primera fase está dada por el diseño muestral de Casen 2022, que corresponde a una muestra probabilística, estratificada y bietápica, siendo los estratos conformados por la combinación comuna-área-nivel socioeconómico<sup>3</sup>. La segunda fase se compone de una muestra probabilística estratificada, siendo los estratos muestrales conformados por la combinación región-área.

---

<sup>2</sup> Documentos de trabajo: Estratificación socioeconómica del marco muestral de viviendas 2017 MMV 2017. <https://www.ine.cl/inicio/documentos-de-trabajo/documento/estratificaci%C3%B3n-socioecon%C3%B3mica-del-marco-muestral-de-viviendas-2017>.

<sup>3</sup> El nivel socioeconómico se incorpora para algunas combinaciones de comuna-área.



### III. ESTIMACIÓN DEL TAMAÑO MUESTRAL

Durante el primer semestre de 2023, el INE tuvo la misión de desarrollar la estrategia de muestreo de EBS 2023 que consiste en proponer y elaborar, de acuerdo con los requerimientos planteados por el MDSF, simulaciones de tamaños muestrales que optimicen la selección de unidades muestrales, de acuerdo con los criterios de factibilidad y presupuesto.

Durante esta etapa, previa al desarrollo de la estrategia de trabajo de campo, se establece como hito inicial del convenio INE-MDSF obtener un tamaño muestral óptimo para la estimación del principal parámetro de interés, con márgenes de precisión que permitan que EBS 2023 tenga representatividad a nivel nacional, nacional urbano, nacional rural, y regional. Para ello, se simularon distintos escenarios o alternativas de tamaño muestral y distribución, basados en la unificación de criterios estadísticos solicitados por MDSF y discutidos en reuniones de trabajo.

Las simulaciones se realizaron considerando los parámetros del levantamiento de EBS 2021, Casen en Pandemia 2020, Piloto Casen Cambio de Marco y Casen 2022. Se formularon diferentes alternativas de distribución de la muestra objetivo que fueron presentadas en reuniones de trabajo conformadas por los equipos de la División de Observatorio Social del Ministerio y el Subdepartamento de Diseño de Marcos y Muestras del INE. Finalmente, se establece un escenario que se ajusta a los requerimientos establecidos por el MDSF y considera el presupuesto acordado.

El escenario elegido utiliza como marco de referencia las 145.051 personas de 18 años y más que fueron resultado del levantamiento de Casen 2022<sup>4</sup>. El error absoluto propuesto para cumplir con el estándar de calidad y el orden de la distribución regional de la población es de 1,0% a nivel nacional. Con errores absolutos entre 2,8% y 3,8% a nivel regional se obtiene un tamaño bajo muestreo aleatorio simple de 6.812 personas a nivel nacional.

Para obtener el efecto de diseño se contempla el efecto de diseño de EBS 2021, el efecto de diseño utilizado para Casen 2022 y el efecto de diseño resultante del cálculo de la tasa de pobreza del Piloto Cambio de Marco Casen 2022 donde, el efecto de diseño propuesto va a corresponder al máximo entre estos tres valores, con excepción de las regiones de Arica y Parinacota, Atacama, Ñuble, Los Ríos y Magallanes donde se ajustan para que el mínimo del efecto diseño sea 1,5 y en la Región Metropolitana se reduce a 3,0 que es definido como valor máximo.

Luego, el tamaño objetivo ajustado por efecto de diseño es de 12.379 personas que, al aplicar el ajuste por finitud, considerando como universo las proyecciones de población al 15 de septiembre de 2023<sup>5</sup>, resulta un tamaño de 12.369 personas a nivel nacional. La distribución de los tamaños por estrato de muestreo (región – área) se realiza en forma proporcional a las proyecciones de población a septiembre de 2023 y la muestra se mantiene en 12.369 personas con un error absoluto es de 1,0% y un error relativo de 8,0%. Si bien a nivel regional se observan errores relativos superiores a 30% en las regiones de Aysén y Magallanes, a partir del error estándar se observa que estos valores cumplen con el estándar de calidad.

<sup>4</sup> Corresponde a la última actualización realizada por el MDSF con fecha 10 de mayo de 2023.

<sup>5</sup> Estas proyecciones consideran el 15 de septiembre de 2023 como fecha de referencia; fecha que se encuentra dentro del periodo estipulado para el levantamiento.



Finalmente, para obtener el tamaño con sobremuestreo, se utiliza una tasa de no logro a nivel regional que corresponde al promedio entre las tasas de no logro a nivel regional de Casen en Pandemia 2020, Piloto Cambio de Marco Casen 2022, ENDIDE precontacto y EBS 2021. Este procedimiento da lugar a un tamaño con sobremuestreo de 26.568 personas.

### III.1. Etapas para la estimación del tamaño muestral para EBS 2023

El cálculo del tamaño muestral comienza con la obtención de parámetros asociados a la última aplicación de la encuesta (EBS 2021) en los niveles de estimación definidos para el nuevo estudio (EBS 2023). Dado que no está definido un parámetro de interés, para efectos del cálculo del tamaño muestral se obtiene la tasa de insatisfacción con la vida a nivel regional junto con sus estadísticos asociados (error estándar y efecto del diseño).

El cálculo del tamaño muestral para EBS 2023 prevé una serie de procedimientos que corresponden, en principio, al paso desde un diseño bajo muestreo aleatorio simple a un diseño complejo, seguido de otros dos ajustes, determinados por la finitud de la población y la tasa de no logro, obteniendo así el tamaño con sobremuestreo.

En los párrafos siguientes se describe, en detalle, cada uno de los nueve pasos realizados para obtener el tamaño muestral con el total de personas a encuestar, en principio, a nivel de región, luego a nivel de estrato de muestreo (región-área):

#### Paso 1: Estimación de parámetros

Se obtienen estimaciones, a nivel regional, de la tasa de insatisfacción con la vida junto con el error estándar y el efecto del diseño asociados a este indicador, usando los resultados del trabajo de campo de EBS 2021. A partir de estos parámetros se obtiene la cuasivarianza poblacional  $S(r)_r^2$  según la ecuación (1):

$$S(r)_r^2 = \frac{m_r}{Def f(r)_r} \cdot SE(r)_r^2 \quad (1)$$

Donde:

- $m_r$  : Número de personas logradas en la región  $r$ .
- $Def f(r)_r$  : Efecto del diseño asociado a la tasa de insatisfacción con la vida ( $r$ ) en la región  $r$ .
- $SE(r)$  : Error estándar de la estimación de la tasa de insatisfacción con la vida ( $r$ ) en la región  $r$ .

El efecto del diseño asociado al parámetro de interés se define como:

$$Def f(r) = \sigma^2(r)_{MC} / \sigma^2(r)_{MAS} \quad (2)$$

Donde:

- $\sigma^2(r)_{MC}$  : Cálculo de la varianza bajo un diseño complejo.
- $\sigma^2(r)_{MAS}$  : Cálculo de la varianza bajo un muestreo aleatorio simple.

En el proceso de obtención del efecto de diseño de una encuesta, es posible que se presenten valores menores que 1 o mayores que lo esperado desde el diseño muestral o respecto de lo



observado en diferentes niveles de estimación. En los casos de las encuestas que hacen uso de diseños muestrales complejos y que la estimación del efecto del diseño se encuentre por debajo del valor 1, puede deberse a que los algoritmos de cálculo de la varianza se basan en aproximaciones que, en general, utilizan el método de conglomerado último, cuyo efecto puede traducirse en una subestimación de la varianza. Luego, al comparar esta estimación subestimada con la varianza bajo diseño aleatorio simple, esta última resulta mayor, obteniéndose un *Def* menor que 1. Ante esta situación, y bajo el supuesto que una muestra polietápica tiene un mayor error de muestreo que bajo muestreo aleatorio simple, se justifica que se considere un valor mínimo de 1 para el efecto de diseño utilizado en el cálculo de tamaños muestrales.

Por otra parte, en los casos en que la persona investigadora considere que los efectos de diseño estimados exceden de manera importante lo esperado, puede tomar la decisión de establecer cotas superiores para su uso en el cálculo de tamaños, bajo el supuesto que los altos valores observados se deben a factores particulares del diseño y de la muestra recolectada y, considerando además, que la inferencia se realiza sobre una submuestra particular de un gran número de muestras posibles obtenidas desde el levantamiento de la encuesta.

### **Paso 2: Cálculo de Tamaño bajo M.A.S.**

Se obtienen los errores absolutos “objetivo” por región ( $d_{0r}$ ) y luego, los errores relativos ( $e_{0r}$ ) según una aproximación del cuantil de una distribución t de Student con nivel de confianza de 95% que, para efectos de la simulación, tiene un valor correspondiente a  $t = 2$ . Luego, junto con la cuasivarianza calculada en la ecuación (1) se obtiene un tamaño muestral inicial a nivel regional ( $m_{0r}$ ) considerando un muestreo aleatorio simple monoetápico.

$$m_{0r} = \frac{t^2_{1-\alpha/2} \cdot S(r)_r^2}{d_{0r}^2} \quad (3)$$

### **Paso 3: Ajuste por efecto del diseño**

Dado que el diseño muestral definido para EBS 2023 es un diseño complejo, el tamaño calculado bajo m.a.s. se debe ponderar por un valor que representa el efecto provocado en la varianza al muestrear bajo un diseño estratificado en lugar de un m.a.s.

Así, el tamaño asociado al diseño complejo se obtiene desde la ecuación (4):

$$m_{1r} = m_{0r} \cdot Def(r)_r \quad (4)$$

Donde:

- $m_{0r}$  : Es el número de personas a encuestar en la región  $r$  bajo un muestreo aleatorio simple.
- $m_{1r}$  : Es el número de personas a encuestar en la región  $r$  considerando un diseño complejo.
- $Def(r)_r$  : Es el efecto de diseño en la región  $r$  obtenido desde la estimación de alguna variable vinculada a la tasa de insatisfacción con la vida.

Generalmente, los valores de  $Def(r)_r$  son acotados al rango (1 – 5) basado en el supuesto que, por un lado, para un mismo tamaño muestral, una muestra polietápica tiene un mayor error de muestreo que bajo un muestreo aleatorio simple, lo que implica que *Def* no puede ser menor que



1 y, por otro, dado que la evidencia empírica en estudios de este tipo apunta a que la ganancia en precisión, producto de corregir el tamaño muestral por un efecto del diseño mayor a 5, es marginal respecto al incremento de los costos de levantamiento del nuevo tamaño muestral.

#### **Paso 4: Corrección por población finita**

El tamaño  $m_{1r}$  a su vez, es ajustado por un ponderador que da cuenta de la finitud de la población. Su forma de cálculo se observa en la ecuación (5):

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

Donde:

- $m_{2r}$  : Es el total de personas a encuestar en la región  $r$  considerando un diseño complejo y asumiendo finitud de la población.
- $M_r$  : Es el total de personas en la región  $r$ , según marco de selección.

Finalmente, este tamaño es el que determina la muestra objetivo regional, que se debe distribuir en cada uno de los estratos de muestreo definidos ( $ra$ ).

#### **Paso 5: Distribución por estrato de muestreo: región – área**

Una vez determinado el tamaño muestral objetivo  $m_{2r}$  a nivel regional, se distribuye este valor, en forma proporcional al marco de selección, en cada uno de los estratos que las conforman.

Es importante mencionar que, el tamaño inicial calculado en cada estrato de muestreo puede ajustarse en la medida que estos no cumplan con tamaños mínimos definidos, por lo que, los tamaños regionales calculados en el paso 4 eventualmente podrían aumentar, caso en el cual, se deben recalcular los errores muestrales propuestos inicialmente.

#### **Paso 6: Ajuste por no logro**

Una vez definido el total de unidades a encuestar para estimar la tasa de insatisfacción con la vida, se debe tener en consideración que, en el trabajo de campo es posible no lograr algunas unidades por diversas razones, tales como: rechazos, moradores ausentes, entre otras. Por este motivo se hace necesario ajustar el tamaño muestral definido en el paso 5 por un ponderador que da cuenta de las posibles pérdidas de unidades muestrales.

Para calcular este ponderador, se utiliza información relacionada con el logro de la encuesta en levantamientos anteriores o de operaciones estadísticas de similar temática o método de recolección. El nivel de desagregación idóneo para el ponderador de ajuste es a nivel de estrato de muestreo. Obtenidos los tamaños con sobremuestreo a este nivel, los tamaños a nivel regional son calculados como la agregación de los tamaños de los estratos que las conforman.



Luego, el tamaño con sobremuestreo a nivel de estrato  $m_{3ra}$  es calculado en la ecuación (6):

$$m_{3ra} = \frac{m_{2ra}}{1 - tnl_{ra}} \quad (6)$$

Donde:

$m_{3ra}$  : Número total de personas a encuestar en el estrato de muestreo  $ra$ .

$tnl_{ra}$  : Tasa de no logro aplicada en el estrato de muestreo  $ra$ .

### Paso 7: Estimación de la tasa de no respuesta a niveles agregados

Dado que el tamaño de muestra regional con sobremuestreo  $m_{3r}$ , es calculado como la suma o agregación de estratos de muestreo  $ra$ , del total de personas de cada una de las regiones, se puede despejar la tasa de no logro y estimar la tasa regional a partir de la ecuación (7):

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

De forma análoga se estima la tasa de no logro a nivel nacional.

## III.2. Parámetros utilizados para el cálculo del tamaño muestral

Para el cálculo del tamaño muestral se consideran los elementos presentados en el Cuadro 1:

### Cuadro 1. Parámetros utilizados para el cálculo de tamaño muestral actual

Parámetro	Descripción
Parámetro asociado	Tasa de insatisfacción con la vida (porcentaje de personas que respondieron “totalmente insatisfecho” e “insatisfecho” en la pregunta ¿Cuán satisfecho(a) o insatisfecho(a) está usted con su vida en este momento?)
Estimador asociado	Estimador de razón de insatisfacción con la vida según resultados obtenidos de EBS 2021: $r = \frac{\text{Número de personas insatisfechas con su vida}}{\text{Número total de personas}}$
Tamaño de la población objetivo	1. Se disponen de proyecciones de población oficiales para el periodo de recolección de la encuesta. 2. Se dispone además de la Población estimada a partir de Casen 2020 de personas de 18 años o más por región y área.
Variable de diseño	$X_j = \begin{cases} 0 & , \text{si la persona no está insatisfecha con su vida} \\ 1 & , \text{si la persona está insatisfecha con su vida} \end{cases}$
Niveles de estimación	Nacional Nacional urbano, Nacional rural Regional



Parámetro	Descripción
<b>Errores de muestreo</b>	<p>Errores absolutos de acuerdo con el estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares (INE, 2020), cuya definición de aceptación se establece mediante el siguiente umbral:</p> $\text{Máximo ee tolerable} = \begin{cases} \sqrt[3]{r^2/9} & ; r < 0,5 \\ \sqrt[3]{(1-r)^2/9} & ; r \geq 0,5 \end{cases}$
<b>Cuasivarianza</b>	EBS 2021
<b>Efecto del diseño</b>	<p>Se utiliza el máximo entre:</p> <ol style="list-style-type: none"> <li>1. El efecto obtenido en EBS 2021.</li> <li>2. El efecto obtenido para la tasa de pobreza en Casen Piloto cambio de marco.</li> <li>3. El efecto utilizado para el cálculo del tamaño muestral de Casen 2022.</li> </ol> <p>Luego, se define un mínimo de 1,5 y un máximo de 3,0.</p>
<b>Tasa de sobremuestreo</b>	Promedio entre las tasas de no logro a nivel regional de Casen en Pandemia 2020, Piloto Cambio de Marco Casen 2022, ENDIDE precontacto y EBS 2021.

Elaborado por el Instituto Nacional de Estadísticas.

### III.3. Metodología de cálculo del tamaño muestral utilizando muestras complejas

La población objetivo, los niveles de estimación deseados y la cobertura geográfica de la encuesta EBS 2023, hacen necesaria la confección de un diseño eficiente para la selección de una muestra de personas de 18 años o más, en términos de ejecución presupuestaria e implementación operativa.

El muestreo aleatorio simple de personas necesitaría, en principio, de un marco de personas de 18 años o más con sus respectivas direcciones de domicilios, generado por un censo de población y viviendas que, por un lado, resultaría muy costoso de mantener actualizado y, por otro, las personas seleccionadas de un marco de este tipo, se encontrarían tan dispersas geográficamente que, los costos asociados al levantamiento excederían la asignación presupuestaria, por tanto, el muestreo aleatorio simple iría en desmedro de la eficiencia deseada.

Por otra parte, en los diseños de muestra complejos, las unidades que finalmente componen la muestra (en este caso personas de 18 años o más) se determinan después de varias etapas; en consecuencia, no es necesario mantener un marco de personas actualizado, lo cual reduce significativamente los costos.

Se debe considerar que la conformación de estratos de muestreo, la conglomeración de unidades muestrales y la definición de las probabilidades de inclusión, caracterizan los diseños de muestra complejos (Heeringa, West, & Berglund, 2010, pág. 13) y el diseño de Casen presenta elementos que lo complejizan, específicamente, por el hecho que sea probabilístico, estratificado y bietápico.

Mientras, algunos de estos elementos tienden a reducir la varianza, como la estratificación, otros tienden a aumentarla, como la conglomeración. Las distintas etapas de muestreo tienden a aumentar la varianza mientras que, la disminución del tamaño de los conglomerados tiende a disminuirla. También es necesario considerar otros efectos tales como la distribución de la



prevalencia del parámetro de interés en las unidades secundarias de muestreo o como la correlación de las unidades secundarias al interior de los conglomerados, respecto también, a la característica de interés, llamada correlación intraclásica, que afecta la varianza y es el principal factor en la medición del efecto del diseño o efecto de no considerar un muestreo aleatorio simple (Kish, 1965).

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

Luego, para determinar el tamaño de la muestra compleja se incorpora este factor de ajuste a partir del tamaño de una muestra aleatoria simple, lo que permite aproximarse al número de viviendas necesarias para que el diseño complejo proporcione la misma varianza (Kish, 1965).

En este contexto, la metodología de cálculo del tamaño muestral utilizada en EBS 2023 se inicia con la determinación de un tamaño muestral bajo un muestreo aleatorio simple, fijando previamente el error absoluto a alcanzar con base en los resultados de encuestas anteriores. Posteriormente, se realizan tres ajustes secuenciales: el primero, que da cuenta del efecto del diseño; el segundo, que corresponde a un ajuste por población finita y; el tercero, referido al ajuste por la tasa de no logro.

Utilizando efectos que dan cuenta del diseño complejo, se calculan tamaños necesarios que permitan obtener estimaciones con el grado de precisión calculado en el paso anterior. Los efectos de diseño utilizados están basados en: los resultados del levantamiento del Piloto Cambio de Marco de Casen, los utilizados en el diseño de Casen 2022 y los efectos obtenidos en el levantamiento de EBS 2021.

Además, la corrección por finitud de la población es estimada a partir de la población completa sobre la que se desea realizar inferencias (el universo de estudio) y por esta razón se utilizan las proyecciones de población o las estimaciones de población proveniente de la encuesta Casen en Pandemia 2020.

Posteriormente, se realiza un procedimiento mediante el cual se distribuye el tamaño de muestra de cada región en los estratos de muestreo que las conforman (región-área) donde, se busca, por un lado, resguardar coherencia entre los tamaños muestrales y aquellos de la población (según el marco muestral utilizado) en términos de proporcionalidad y, por otro, que dicha distribución no difiriera mayormente de la distribución de tamaños de las proyecciones de población al 15 de septiembre de 2023.

Bajo esta distribución, los tamaños con sobremuestreo son calculados a partir de una tasa de no logro que considera el promedio entre las tasas de no logro a nivel regional de diferentes levantamientos realizados por el MDSF que corresponden a Casen en Pandemia 2020, Piloto Casen Cambio de Marco, ENDIDE precontacto y EBS 2021.

Finalmente, para que la muestra sea eficiente, en términos de la cantidad de personas a seleccionar, la fracción de muestreo respecto del marco debe considerar el establecimiento de cotas (especialmente superiores) que, idealmente no debieran superar 70% a nivel regional. Sin embargo, se hace necesario también la revisión de las fracciones de muestreo a niveles más desagregados, en este caso, desagregación por área (urbana-rural) y por estratos de muestreo (región-área).



### III.4. Resumen proceso de simulación

---

En este apartado se presentan las simulaciones de tamaños muestrales derivadas de distintos escenarios que fueron trabajados inicialmente, hasta llegar a la propuesta elegida. La estructura del capítulo mantiene la cronología de las simulaciones realizadas, así como también, la metodología empleada en ellas.

Como base para el desarrollo de todos los escenarios trabajados se presentan los resultados del trabajo de campo de EBS 2021, desde donde se obtiene la estimación de la tasa de insatisfacción con la vida junto con sus estadísticos asociados, a saber, el error estándar, el coeficiente de variación y el efecto de diseño, tanto a nivel nacional como regional.

A partir de estos resultados, fueron analizados seis escenarios, el primero de ellos elaborado por el MDSF (Escenario MDSF) y, los siguientes, por el INE. En relación con los escenarios trabajados por el INE, el primero (escenario n°0) corresponde a un escenario base que toma los parámetros obtenidos de EBS 2021. Luego, a partir de él, se desarrolla el escenario n°1 que es una simulación inicial considerando como marco muestral la estimación de las personas de 18 años o más de Casen en Pandemia 2020 y se utilizan los totales de personas de 18 años o más logradas en Casen en Pandemia 2020 para calcular la fracción de muestreo.

El siguiente escenario (n°1.1) mantiene los errores y efectos del diseño utilizados en el escenario n°1, la modificación se encuentra en el insumo utilizado para calcular la corrección por finitud ya que, en este caso, se utilizan las proyecciones de población al 15 de febrero de 2023.

El escenario n°2 utiliza como insumo para la corrección por población finita las proyecciones de población al 15 de septiembre de 2023 y, los resultados del levantamiento de Casen 2022, para calcular las fracciones de muestreo. Además, se realizan ajustes a los efectos del diseño definidos en los escenarios anteriores para establecer un mínimo de 1,5 y un máximo de 3,0 y se modifica el ponderador de ajuste para el cálculo del tamaño con sobremuestreo.

Finalmente, los escenarios n°2.1 y n°2.2 (escenario escogido) utilizan como base el escenario n°2, ajustando los errores y considerando una actualización de la Base de levantamiento de Casen 2022, con fecha 2 de mayo de 2023, en el escenario n°2.1 y; 10 de mayo de 2023, en el escenario n°2.2, para el cálculo de las fracciones de muestreo.

Obtenidos los tamaños muestrales objetivos del escenario elegido se realiza una comparación entre su distribución y la distribución de las proyecciones de población a nivel regional, con la finalidad de mostrar que no existen mayores diferencias con dichos referentes.



El Cuadro 2 presenta los parámetros comunes utilizados en los escenarios considerados.

**Cuadro 2: Parámetros comunes utilizados en escenarios EBS 2023**

Parámetro	Descripción
Parámetro a estimar	Tasa de insatisfacción con la vida
Variable de diseño	$X_j = \begin{cases} 0, & \text{si la persona no está insatisfecha con su vida} \\ 1, & \text{si la persona está insatisfecha con su vida} \end{cases}$
Estimador asociado	Estimador de razón: $r = \frac{\text{Número de personas insatisfechas con su vida}}{\text{Número total de personas}}$

Elaborado por el Instituto Nacional de Estadísticas.

La Tabla III.1 presenta un resumen de los parámetros utilizados en cada una de las simulaciones, así como también, los tamaños y errores muestrales resultantes, por orden cronológico, comenzando con el escenario propuesto por el MDSF, para seguir con los resultados del trabajo de campo de EBS 2021 (escenario n°0) y luego, presentar el resto de los escenarios.



**Tabla III.1: Resumen de parámetros utilizados y tamaños y errores resultantes según escenario, EBS 2023**

Nro.	Niveles de estimación	Error Absoluto	Error Relativo	Efectos de Diseño	Tamaño Muestral Objetivo	Tamaño Muestral Objetivo (ajustado) <sup>6</sup>	Tasa de no logro	Tamaño con sobremuestreo
MDSF	Nacional	0,9%	1,8%			11.300	Máximo entre las tasas de no respuesta regional entre Casen 2017, ENDIDE precontacto y Cambio de Marco Casen 2022	14.551
	Regional	3,0% - 4,0%	2,0% - 8,0%					
0	Nacional	1,0%	8,0%	2,5	9.439			
	Regional	2,1% - 3,2%	15,8% - 29,1%	1,0 - 1,5				
1	Nacional	1,0%	8,0%	2,5	16.490	16.490	Máximo entre las tasas de no respuesta regional entre Casen 2017, ENDIDE precontacto y Cambio de Marco Casen 2022	20.378
	Regional	2,3% - 3,2%	17,3% - 28,1%	1,2 - 3,2				
1.1	Nacional	1,0%	8,0%	1,8	16.490	16.490	Máximo entre las tasas de no respuesta regional entre Casen 2017, ENDIDE precontacto y Cambio de Marco Casen 2022	20.378
	Regional	2,3% - 3,2%	17,3% - 28,1%	1,2 - 3,2				
2	Nacional	1,0%	8,0%	1,8	14.326	14.326	Promedio entre las tasas de no logro regional de Casen en Pandemia 2020, Piloto Cambio de Marco Casen 2022, ENDIDE precontacto y EBS 2021.	22.428
	Regional	2,6% - 3,5%	19,1% - 32,6%	1,5 - 3,0				
2.1- 2.2	Nacional	1,0%	8,0%	1,8	12.369	12.369	Promedio entre las tasas de no logro regional de Casen en Pandemia 2020, Piloto Cambio de Marco Casen 2022, ENDIDE precontacto y EBS 2021	26.568
	Regional	2,8% - 3,8%	20,6% - 33,6%	1,5 - 3,0				

Elaborado por el Instituto Nacional de Estadísticas.

<sup>6</sup> Los errores absolutos y relativos son calculados con el tamaño objetivo ajustado.

**III.5. Escenario escogido****III.5.1. Escenario 2.2****Tabla III.2: Estimación de tamaños muestrales según proyecciones de población al 15 de septiembre de 2023 y total de personas de Casen 2022 según base actualizada al 10 de mayo de 2023**

Región	Personas ≥18 años Casen 2022	Proyección Personas ≥18 años 15/09/23	Tasa insatisfacción con la vida	Efecto Diseño Ajustado	Tamaño ajustado efecto diseño	Error absoluto	Error Relativo	Tamaño ajustado por finitud	Tamaño Objetivo Ajustado	Error Absoluto ajustado	Error Relativo tamaño ajustado	Tamaño con Sobre-muestreo
<b>Nivel País</b>	<b>145.051</b>	<b>15.520.055</b>	<b>12,6%</b>	<b>1,8</b>	<b>12.379</b>	<b>1,0%</b>	<b>8,0%</b>	<b>12.369</b>	<b>12.369</b>	<b>1,0%</b>	<b>8,0%</b>	<b>26.568</b>
<b>Urbano</b>	<b>115.455</b>	<b>13.737.414</b>	<b>12,4%</b>	<b>1,8</b>	<b>10.957</b>	<b>1,3%</b>	<b>8,7%</b>	<b>10.948</b>	<b>10.490</b>	<b>1,3%</b>	<b>10,5%</b>	<b>22.635</b>
<b>Rural</b>	<b>29.596</b>	<b>1.782.641</b>	<b>14,1%</b>	<b>1,8</b>	<b>1.422</b>	<b>3,2%</b>	<b>19,4%</b>	<b>1.421</b>	<b>1.879</b>	<b>2,8%</b>	<b>20,0%</b>	<b>3.933</b>
Arica y Parinacota	5.320	197.544	15,5%	1,5	548	3,8%	20,7%	547	547	3,8%	24,6%	1.240
Tarapacá	6.021	296.544	13,6%	1,6	602	3,5%	21,0%	601	601	3,5%	25,8%	1.394
Antofagasta	5.822	541.235	12,0%	1,7	760	3,1%	23,2%	759	759	3,1%	26,1%	1.625
Atacama	6.351	237.255	10,6%	1,5	590	3,2%	24,8%	589	589	3,2%	29,7%	1.264
Coquimbo	5.605	665.742	12,2%	2,2	763	3,5%	23,3%	763	763	3,5%	28,5%	1.802
Valparaíso	15.299	1.581.992	12,0%	1,9	960	2,9%	20,2%	960	960	2,9%	24,0%	1.926
Metropolitana	28.005	6.554.944	13,3%	3,0	1.809	2,8%	15,8%	1.809	1.809	2,8%	20,6%	4.094
O'Higgins	9.786	788.017	10,9%	1,8	769	3,1%	24,2%	769	769	3,1%	28,3%	1.539
Maule	10.220	900.081	12,7%	1,7	821	3,0%	22,7%	821	821	3,0%	23,5%	1.583
Ñuble	6.450	406.884	10,6%	1,5	663	3,0%	23,3%	662	662	3,0%	28,2%	1.358
Biobío	14.924	1.308.986	12,3%	1,8	884	2,9%	20,9%	884	884	2,9%	23,7%	1.968
La Araucanía	9.769	790.070	12,0%	1,7	775	3,1%	22,2%	775	775	3,1%	25,5%	1.591
Los Ríos	7.067	320.854	12,4%	1,5	656	3,1%	22,1%	655	655	3,1%	25,4%	1.355
Los Lagos	8.093	704.201	13,1%	2,1	764	3,5%	21,4%	764	764	3,5%	27,1%	1.662
Aysén	2.670	81.582	10,0%	1,5	467	3,4%	29,1%	465	465	3,4%	33,6%	964
Magallanes	3.649	144.124	9,0%	1,5	548	3,0%	27,9%	546	546	3,0%	33,0%	1.203

Elaborado por el Instituto Nacional de Estadísticas.

El escenario n°2.2 corresponde al escenario elegido y se basa en la versión más reciente de los resultados del levantamiento de Casen 2022. La diferencia respecto al escenario n°2.1 se encuentra en las fracciones de muestreo que se obtienen entre el tamaño con sobremuestreo a nivel regional y las personas elegibles obtenidas de la actualización del marco de selección.

Este escenario, al igual que los anteriores, se elabora con base en los parámetros obtenidos del trabajo de campo de EBS 2021 (escenario n°0); y con un marco de referencia correspondiente a 145.051 personas resultantes del levantamiento de Casen 2022. Respecto al efecto de diseño, éste se basa en el efecto de diseño de la EBS 2021, el efecto de diseño utilizado en Casen 2022 y efecto de diseño resultante de la tasa de pobreza en hogares del Piloto Cambio de Marco Casen 2022, y corresponde al máximo entre estos tres valores. Además, a nivel regional se establece que los efectos de diseño fluctúen entre 1,5 y 3,0.

Luego, el tamaño objetivo ajustado por efecto de diseño es de 12.379 que, al aplicar el ajuste por finitud, considerando como universo las proyecciones de población al 15 de septiembre de 2023, resulta un tamaño de 12.369 personas a nivel nacional, tamaño que se mantiene al ajustar la distribución de los tamaños por estrato de muestreo (región-área) en forma proporcional a las proyecciones de población al 15 de septiembre de 2023.

Finalmente, para obtener el tamaño con sobremuestreo, se utiliza una tasa de no logro a nivel regional a partir de la media entre las tasas de no logro a nivel regional de Casen en Pandemia 2020, Piloto Cambio de Marco Casen 2022, ENDIDE precontacto y EBS 2021. Este procedimiento dio lugar a un tamaño con sobremuestreo de 26.568 personas.

### **III.6. Tamaño muestral propuesto**

---

El tamaño calculado a nivel regional se debe distribuir por área (urbana – rural). Para esto, dependiendo del escenario, se cuenta con la distribución resultante de Casen en pandemia 2020 o bien, con las proyecciones de población estimadas al 15 de septiembre de 2023 (fecha que coincide con el periodo de levantamiento de EBS 2023) al mismo nivel.

El tamaño objetivo ajustado del escenario elegido (n°2.2) definido para obtener estimaciones de la tasa de insatisfacción con la vida con suficiente grado de precisión a nivel regional, alcanza 12.369 personas a nivel nacional, cuyos errores absolutos y relativos asociados son de 1,0% y 8,0%, respectivamente. Aplicando las tasas de no logro a nivel regional con ciertos ajustes (con el objetivo de que el marco muestral cuente con suficientes unidades para seleccionar la muestra) se obtiene el tamaño con sobremuestreo que alcanza 26.568 personas.

La Tabla III.3 muestra los tamaños objetivos ajustados junto con sus errores muestrales asociados y el tamaño con sobremuestreo a nivel nacional y regional, desagregados por área urbano-rural.



**Tabla III.3: Tamaños y errores finales a nivel nacional, regional, urbano-rural, EBS 2023**

Región	Error Absoluto	Error Relativo	Tamaño objetivo			Tamaño con sobremuestreo		
			Total	Urbano	Rural	Total	Urbano	Rural
<b>Nivel País</b>	<b>1,0%</b>	<b>8,0%</b>	<b>12.369</b>	<b>10.948</b>	<b>1.421</b>	<b>26.568</b>	<b>22.635</b>	<b>3.933</b>
Arica y Parinacota	3,8%	24,6%	547	502	45	1.240	1.142	98
Tarapacá	3,5%	25,8%	601	567	34	1.394	1.324	70
Antofagasta	3,1%	26,1%	759	730	29	1.625	1.538	87
Atacama	3,2%	29,7%	589	536	53	1.264	1.148	116
Coquimbo	3,5%	28,5%	763	622	141	1.802	1.508	294
Valparaíso	2,9%	24,0%	960	881	79	1.926	1.752	174
Metropolitana	2,8%	20,6%	1.809	1.744	65	4.094	3.959	135
O'Higgins	3,1%	28,3%	769	576	193	1.539	1.152	387
Maule	3,0%	23,5%	821	608	213	1.583	1.164	419
Ñuble	3,0%	28,2%	662	466	196	1.358	982	376
Biobío	2,9%	23,7%	884	786	98	1.968	1.764	204
La Araucanía	3,1%	25,5%	775	556	219	1.591	1.095	496
Los Ríos	3,1%	25,4%	655	476	179	1.355	1.006	349
Los Lagos	3,5%	27,1%	764	565	199	1.662	1.253	409
Aysén	3,4%	33,6%	465	374	91	964	751	213
Magallanes	3,0%	33,0%	546	501	45	1.203	1.097	106

Elaborado por el Instituto Nacional de Estadísticas

Luego, a partir de las desagregaciones de tamaños por estrato geográfico (región-área) y de los resultados del trabajo de campo de EBS 2021 se obtienen, por agregación, tamaños asociados para los niveles Nacional Urbano y Nacional Rural y, a partir de estos, se calculan los errores absolutos y relativos resultantes para dichos niveles que se muestran en la Tabla III.4.

**Tabla III.4: Tamaños y errores EBS 2021 y propuesta EBS 2023. Nacional Urbano - Rural**

	Resultados EBS 2021				Propuesta EBS 2023			
	Muestra efectiva	Tasa Insatisfacción con la vida	Error Absoluto	Error Relativo	Tamaño Objetivo	Error Absoluto	Error Relativo	Tamaño con Sobremuestreo
<b>Nivel País</b>	10.921	12,6%	1,0%	8,0%	12.369	1,0%	8,0%	26.568
Urbano	9.307	12,4%	1,1%	8,7%	10.490	1,3%	10,5%	22.635
Rural	1.614	14,1%	2,7%	19,4%	1.879	2,8%	20,0%	3.933

Elaborado por el Instituto Nacional de Estadísticas.

A nivel nacional urbano, con un tamaño objetivo ajustado de 10.490 personas se obtiene una estimación para la tasa de insatisfacción con la vida con un error absoluto de 1,3% y un error relativo de 10,5%.



A nivel nacional rural, con un tamaño objetivo de 1.879 personas se obtiene una estimación para la tasa de insatisfacción con la vida con un error absoluto de 2,8% y un error relativo de 20,0%.

### III.7. Errores esperados

La Tabla III.5 presenta los errores esperados, asociados a la estimación del parámetro de interés para un tamaño muestral objetivo de 12.369 personas<sup>7</sup>.

**Tabla III.5: Errores esperados asociados al parámetro de interés según tamaño muestral objetivo**

Región	Tasa insatisfacción con la vida EBS 2021	Error Absoluto Propuesto EBS 2023	Error Relativo Propuesto EBS 2023	Tamaño ajustado por finitud EBS 2023	Tamaño objetivo ajustado EBS 2023	Tamaño con sobremuestreo EBS 2023
<b>Nacional</b>	<b>12,6%</b>	<b>1,0%</b>	<b>8,0%</b>	<b>12.369</b>	<b>12.369</b>	<b>26.568</b>
Arica y Parinacota	15,5%	3,8	20,7	547	547	1.240
Tarapacá	13,6%	3,5	21,0	601	601	1.394
Antofagasta	12,0%	3,1	23,2	759	759	1.625
Atacama	10,6%	3,2	24,8	589	589	1.264
Coquimbo	12,2%	3,5	23,3	763	763	1.802
Valparaíso	12,0%	2,9	20,2	960	960	1.926
Metropolitana	13,3%	2,8	15,8	1.809	1.809	4.094
O'Higgins	10,9%	3,1	24,2	769	769	1.539
Maule	12,7%	3,0	22,7	821	821	1.583
Ñuble	10,6%	3,0	23,3	662	662	1.358
Biobío	12,3%	2,9	20,9	884	884	1.968
La Araucanía	12,0%	3,1	22,2	775	775	1.591
Los Ríos	12,4%	3,1	22,1	655	655	1.355
Los Lagos	13,1%	3,5	21,4	764	764	1.662
Aysén	10,0%	3,4	29,1	465	465	964
Magallanes	9,0%	3,0	27,9	546	546	1.203

Elaborado por el Instituto Nacional de Estadísticas.

<sup>7</sup> La estimación de los errores se obtiene a partir del tamaño objetivo y no del tamaño con sobremuestreo, pues dadas las pérdidas naturales de unidades muestrales, se espera obtener un tamaño de 12.369 personas.



## IV. SELECCIÓN DE UNIDADES MUESTRALES

La selección de unidades muestrales se realiza desde el marco muestral actualizado al 10 de mayo de 2023, conformado por personas logradas en Casen 2022, de 18 años o más, en forma sistemática y con igual probabilidad al interior de cada estrato de muestreo (región-área) y considerando sólo aquellas personas que no pertenecen al servicio doméstico puertas adentro.

La selección de personas es implementada en el *software* R, utilizando el paquete “Sampling”, fijando una semilla aleatoria, a fin de que pueda ser replicable en cualquier momento.

Sean  $M_{ra}$  el número de personas que conforman el estrato  $ra$  y  $m_{ra}^{sel}$  el número de personas a seleccionar. Para la selección de  $m_{ra}^{sel}$  personas, el *software* ejecuta los siguientes pasos:

### **Paso 1**

En primera instancia se ordenan geográficamente y en forma ascendente todas las personas según región, comuna, área, identificador de la UPM, sexo y edad<sup>8</sup>.

### **Paso 2**

Al interior de cada estrato de muestreo se calcula el período ( $k$ ) que corresponde a:

$$k = M_{ra} / m_{ra}^{sel}$$

Notar que “ $k$ ” puede ser un número real, no entero (puede tener decimales).

### **Paso 3**

Luego se determina el arranque “ $A$ ” o primera selección, que corresponde a una semilla aleatoria propia para la encuesta.

### **Paso 4**

Posteriormente se suma sucesivamente el período “ $k$ ” al arranque “ $A$ ” para obtener distintos valores, los que dan origen a la selección de unidades de la siguiente forma: “ $A$ ”, “ $A + k$ ”, “ $A + 2k$ ”, “ $A + 3k$ ”, ..., “ $A + (m_{ra}^{sel} - 1)k$ ”.

La primera persona seleccionada es “ $A$ ” y es un número entero, el segundo es el redondeo de “ $A + k$ ”, el tercero es el redondeo de “ $A + 2k$ ” y así sucesivamente, hasta la  $m_{ra}^{sel}$  selección, dada por el redondeo de “ $A + (m_{ra}^{sel} - 1)k$ ”.

Luego, la probabilidad de inclusión de la  $k$  – ésima persona dentro del estrato  $ra$  está dada por:

$$P_{ra,k}(k) = \frac{m_{ra}^{sel}}{M_{ra}} \quad (8)$$

<sup>8</sup> En la sintaxis enviada a MDSF la variable “correlativo” replica este ordenamiento.



Donde:

$m_{ra}^{sel}$  : Número de personas seleccionadas en el estrato  $ra$

$M_{ra}$  : Número actualizado de personas en el marco de selección en el estrato  $ra$ .

La Tabla IV.1 presenta la distribución del marco de personas de 18 años o más logradas en Casen 2022 y el total de personas seleccionadas, según región y área (Urbano - Rural).

**Tabla IV.1: Distribución del número de personas por región y área**

Región	Área	Total personas	
		Marco Casen 2022	Muestra objetivo EBS 2023
<b>Total País</b>		<b>145.051</b>	<b>12.369</b>
Arica y Parinacota	Urbano	4.844	502
	Rural	476	45
Tarapacá	Urbano	5.546	567
	Rural	475	34
Antofagasta	Urbano	5.253	730
	Rural	569	29
Atacama	Urbano	5.527	536
	Rural	824	53
Coquimbo	Urbano	4.289	622
	Rural	1.316	141
Valparaíso	Urbano	12.362	881
	Rural	2.937	79
Metropolitana	Urbano	25.785	1.744
	Rural	2.220	65
O'Higgins	Urbano	6.704	576
	Rural	3.082	193
Maule	Urbano	7.082	608
	Rural	3.138	213
Ñuble	Urbano	4.054	466
	Rural	2.396	196
Biobío	Urbano	11.867	786
	Rural	3.057	98
La Araucanía	Urbano	6.240	556
	Rural	3.529	219
Los Ríos	Urbano	5.095	476
	Rural	1.972	179
Los Lagos	Urbano	5.499	199
	Rural	2.594	764
Aysén	Urbano	2.002	374
	Rural	668	91
Magallanes	Urbano	3.306	501
	Rural	343	45

Elaborado por el Instituto Nacional de Estadísticas.



## V. CONSTRUCCIÓN FACTORES DE EXPANSIÓN

Los análisis y estudios basados en encuestas de hogares con diseño muestral probabilístico utilizan un ponderador en la estimación de parámetros de interés para que estos tengan validez sobre la población objetivo. Este ponderador guarda relación con las probabilidades de selección de las distintas unidades de muestreo y da cuenta del número de personas de la población que representa cada individuo que participa en la encuesta. Este ponderador es conocido como factor de expansión.

Para EBS 2023 se desarrollan factores que expanden a la población tanto nacional como regional y se calculan para cada persona presente en la muestra. El cálculo de estos factores responde al diseño muestral bifásico cuya primera fase se compone del factor que se encuentra asociado a las viviendas logradas en Casen 2022 y, en la segunda fase se calcula un factor de selección de personas de 18 años o más en forma sistemática y con igual probabilidad al interior de cada estrato de muestreo correspondiente al diseño de la encuesta.

La metodología de cálculo de los factores de expansión consiste en la obtención secuencial de una serie de ponderadores que, como se menciona en párrafo precedente, se inicia con el factor de expansión de primera fase que corresponde al ponderador de no respuesta de las viviendas logradas en Casen 2022 y que da cuenta de su diseño muestral. Luego, se obtiene el factor de expansión de segunda fase el que refleja cuántas personas en la población representa la persona seleccionada en la muestra de EBS.

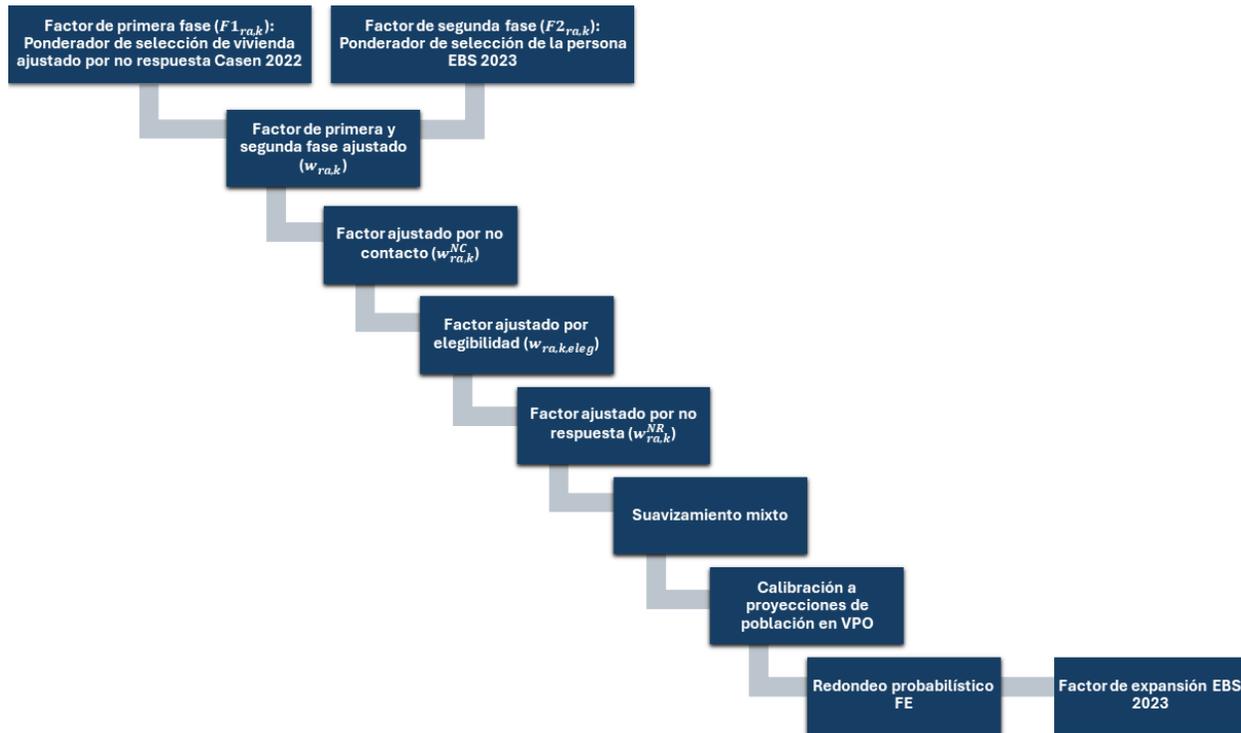
A continuación, se obtiene el producto entre el factor de primera y segunda fase, al que se le aplican una serie de ajustes según el resultado del trabajo de campo y la definición del estado de logro de las personas (no contacto, no elegibilidad y no respuesta) con el objetivo de reducir el sesgo asociado a la falla en conseguir la cooperación de las personas seleccionadas en la muestra. A este ponderador ajustado por no respuesta, se le realiza un suavizamiento para aminorar el efecto de valores inusualmente atípicos y finalmente, la calibración con el fin de cuadrar las estimaciones poblacionales obtenidas de la encuesta con las estimaciones poblacionales o proyecciones de población desarrolladas y utilizadas por el INE en todas las encuestas de hogares.

Para la calibración, se utiliza la técnica por *Raking* que permite controlar por diversas variables simultáneamente y mantener el diseño muestral disminuyendo el sesgo que se produce por errores no muestrales, provocados por la no respuesta y/o por la imperfección de los marcos muestrales, además de mejorar la precisión de las estimaciones. En el proceso de calibración, se utiliza la información auxiliar proveniente de las Proyecciones de Población de 18 años y más según sexo y edad para distintos niveles de desagregación.

La Figura V.1 muestra el flujo del proceso del cálculo de los factores de expansión para EBS 2023 y a lo largo del capítulo se detalla cada una de las etapas del proceso de construcción de los factores de expansión.



Figura V.1. Diagrama de procesos de desarrollo de factor de expansión EBS 2023



Elaborado por Instituto Nacional de Estadísticas.

## V.1. Factor de expansión de primera fase

En Casen 2022 las unidades muestrales se obtienen desde el MMV 2020 a partir de un diseño probabilístico, estratificado y bietápico, donde los estratos muestrales están conformados por la combinación Comuna-Área-NSE, donde aplique, y los conglomerados (Unidades Primarias de Muestreo o UPM), al interior de los estratos, son seleccionados en forma sistemática con probabilidad proporcional al tamaño, según el número de viviendas particulares que los conforman. Por su parte, las viviendas o unidades secundarias de muestreo son seleccionadas en forma aleatoria y con igual probabilidad al interior de cada UPM elegida.

El cálculo del factor de expansión de primera fase de EBS considera el diseño muestral de Casen 2022 donde se realiza un ajuste por elegibilidad y un ajuste por no respuesta. De la encuesta Casen 2022 se obtiene información sobre los conglomerados y viviendas entrevistadas, luego se realiza la corrección de elegibilidad y la corrección por no respuesta, de donde se obtiene la información de todas las personas que pertenecen a las viviendas que respondieron la encuesta Casen 2022.

Para EBS 2023 el factor de expansión de primera fase ( $F1_{ra,k}$ ) corresponde al ponderador ajustado por no respuesta de Casen 2022 ( $w_{hij}^{NR}$ ) donde se calcula una razón de ajuste para que los ponderadores asociados



a las viviendas elegibles que no respondieron se redistribuyan entre las viviendas elegibles que sí respondieron.

Por lo tanto, este factor de primera fase corresponde al factor de selección de viviendas logradas en Casen 2022 y se expresa según la ecuación (9):

$$F1_{ra,k} = w_{hij}^{NR} = \hat{R}_{g,R} \cdot w'_{hij} \quad (9)$$

Donde,

$ra, k$  : Índice del estrato de muestreo región-área y de la persona seleccionada, respectivamente.

$w_{hij}^{NR}$  : Ponderador de selección de viviendas de Casen 2022 ajustado por no respuesta.

## V.2. Factor de expansión de segunda fase

---

Dado que EBS presenta un diseño bifásico, luego del cálculo de los ponderadores asociados a las viviendas que respondieron Casen 2022 es necesario obtener los ponderadores de las personas seleccionadas en la muestra.

### V.2.1. Ponderador de selección de personas como inverso de la probabilidad de selección

Este ponderador corresponde al inverso de la probabilidad de selección de la persona para la muestra de EBS según el estrato de muestreo (región – área). Se interpreta como el número de personas de 18 años o más en la población, que representa la persona seleccionada en la muestra de EBS.

El ponderador de selección de la persona  $k$  es calculado como el inverso de su probabilidad de selección al interior del estrato  $ra$ . Esto se observa en la ecuación (10):

$$F2_{ra,k} = \frac{M_{ra}}{m_{ra\ sel}} \quad (10)$$

Donde,

$ra, k$  : Índice del estrato de muestreo región-área y de la persona  $k$  seleccionada, respectivamente.

$F2_{ra,k}$  : Ponderador de selección de la persona  $k$  en el estrato  $ra$ .

$M_{ra}$  : Total de personas disponibles en el estrato, según el levantamiento de Casen 2022.

$m_{ra\ sel}$  : Total de personas seleccionadas en EBS 2023 en el estrato  $ra$ .



### V.2.2. Producto del factor de primera y segunda fase (Ponderador correspondiente al producto entre los ponderadores de primera y segunda fase)

Este ajuste da cuenta de la selección de personas considerando ambas fases de muestreo, es decir, contempla la selección de la k-ésima persona para EBS, dado que fue entrevistada y lograda en Casen 2022. Este factor corresponde al factor de selección y se expresa según la ecuación (11):

$$w_{ra,k} = F1_{ra,k} * F2_{ra,k} \quad (11)$$

Donde,

$w_{ra,k}$  : Ponderador de selección de personas considerando la primera y segunda fase.

### V.3. Ajustes por no contacto, elegibilidad y no respuesta

En principio la muestra incluye personas elegibles ya que estas pertenecen a hogares encuestados en Casen 2022. Sin embargo, durante el levantamiento hay situaciones en las que no es posible establecer contacto con la persona seleccionada o bien, hay personas que dejan de ser consideradas elegibles por diferentes motivos como, por ejemplo, dejar de pertenecer al hogar encuestado en Casen.

Estas situaciones se identifican en el cuestionario hoja de ruta a través de diferentes códigos de disposición final (CDF)<sup>9</sup> donde, en el caso de EBS 2023, al tratarse de un levantamiento telefónico, las personas seleccionadas son clasificadas en elegibles y no elegibles. La tabla V.1 muestra la distribución de la elegibilidad en EBS 2023.

**Tabla V.1. Cantidad de personas según tipo de elegibilidad, EBS 2023**

Elegibilidad	Total de personas	Porcentaje del total
<b>Total</b>	<b>26.568</b>	
Elegibles entrevistadas	11.234	42,3%
Elegibles no entrevistadas <sup>10</sup>	5.140	19,3%
Elegibles no contactadas	7.960	30,0%
No elegibles	2.234	8,4%

Elaborado por el Instituto Nacional de Estadísticas.

#### V.3.1. Ajuste por no contacto

Luego del trabajo de campo de EBS 2023 el no contacto corresponde a 30% de la muestra gestionada. Entre las principales razones del no contacto se encuentran: el número telefónico no existe; el teléfono marca, pero no hay respuesta o responde el buzón de voz. Este ajuste se realiza porque, ante la imposibilidad de

<sup>9</sup> Ver Anexo n°1

<sup>10</sup> Esta categoría incluye a las personas con entrevista parcial, rechazo y otros motivos de no entrevista.



establecer contacto, no es posible conocer con certeza la elegibilidad de la persona, por lo tanto, el objetivo de este ajuste es determinar la probabilidad de que la persona sea contactada y aplicarlo para todas las personas que sí lo fueron.

Esta probabilidad se obtiene para las 26.568 personas seleccionadas a través del modelamiento utilizando variables de Casen 2022 y la hoja de ruta de EBS 2023, a partir de una regresión logística donde las probabilidades de respuesta son ordenadas de menor a mayor y las personas, con características similares, son agrupadas en veintiles de ajuste a nivel nacional.

En la tabla V.2 se presentan las variables utilizadas en el modelo de no contacto.

**Tabla V.2. Variables utilizadas en el modelo de no contacto EBS 2023**

<b>Variable</b>	<b>Fuente</b>
Horario de contacto (mañana, tarde o noche)	Hoja de ruta EBS 2023
Cantidad de encuestas realizadas por la persona encuestadora en tramos	Hoja de ruta EBS 2023
Hogar seleccionado en EANNA 2023	Hoja de ruta EBS 2023
Región	Casen 2022
Número de personas en el hogar (excluyendo servicio doméstico puertas adentro)	Casen 2022
Nivel educacional más alto al que asiste o alcanzó	Casen 2022
Hogar unipersonal	Casen 2022
Jefe de hogar o parentesco con este (pareja, otro familiar, no familiar)	Casen 2022
Estado civil	Casen 2022
Edad en tramos ([18-29], [30-45], [46-59], [60+])	Casen 2022
La persona trabajó al menos una hora la semana pasada, sin considerar quehaceres del hogar	Casen 2022
Frecuencia en que ha presenciado rayados o daños a casas o vehículos en el área de residencia	Casen 2022
Frecuencia en que ha presenciado consumo de drogas o alcohol en el área de residencia	Casen 2022
Dificultad para ver incluso si lleva lentes	Casen 2022
Condición de tenencia de la vivienda	Casen 2022
Proporción de personas desocupadas en el hogar	Casen 2022
Acceso a internet	Casen 2022
Sexo de la persona	Casen 2022
La persona tiene dificultad para recordar o concentrarse	Casen 2022
Participación en alguna organización social, política o corporativa	Casen 2022
Hogar cuenta con un vehículo de uso laboral o particular	Casen 2022

Elaborado por el Instituto Nacional de Estadísticas.

Los veintiles de ajuste tienen como característica que agrupan personas de similares características por lo que, las personas contactadas podrán representar a aquellas que no lo fueron y, por lo tanto, el peso de las 7.960 personas no contactadas será distribuido entre aquellas que sí lo fueron, a través de una razón de ajuste realizada al interior de cada veintil, utilizando el ponderador de selección de primera y segunda fase como se observa en la ecuación (12).



En el numerador se incluye la estimación del total de personas en la muestra, y en el denominador la estimación del total de personas que fueron contactadas.

$$\hat{R}_{g,C} = \frac{\sum_{i \in \Phi_g} \sum_{j \in \theta_g} w_{ra,k}}{\sum_{i \in \Phi_g} \sum_{j \in \theta_{g,cont}} w_{ra,k}} \quad (12)$$

Donde,

- $g$ : : Índice del grupo para la corrección de no contacto (veintil).
- $\Phi_g$ : : Conjunto de personas asignadas al veintil  $g$ .
- $\theta_g$ : : Conjunto de personas pertenecientes al veintil  $g$  en la muestra.
- $\theta_{g,cont}$ : : Conjunto de personas pertenecientes al veintil  $g$  y que fueron contactadas.
- $w_{ra,k}$ : : Ponderador de selección de la persona de primera y segunda fase

Una vez estimados los factores de ajuste, el ponderador de selección corregido por no contacto  $w_{ra,k}^{NC}$  se expresa según la ecuación (13):

$$w_{ra,k}^{NC} = \hat{R}_{g,C} \cdot w_{ra,k} \quad (13)$$

Este ponderador se asigna a las 16.374 personas contactadas. A las personas no contactadas se asigna un valor blanco (“missing”).

La Tabla V.3 presenta la distribución de los veintiles y el promedio de la probabilidad de contacto de cada uno de ellos.

**Tabla V.3. Veintiles ajuste por no contacto EBS 2023**

Veintil	Personas seleccionadas	Personas contactadas	Personas no contactadas	Promedio Probabilidad de contacto
Total país	26.568	18.608	7.960	
1	1.329	696	633	0,321
2	1.329	760	569	0,373
3	1.329	813	516	0,400
4	1.329	841	488	0,420
5	1.329	846	483	0,437
6	1.329	872	457	0,451
7	1.329	881	448	0,465
8	1.329	938	391	0,478
9	1.328	912	416	0,490
10	1.328	960	368	0,503
11	1.328	953	375	0,515
12	1.328	963	365	0,526
13	1.328	978	350	0,538
14	1.328	993	335	0,551
15	1.328	990	338	0,564



Veintil	Personas seleccionadas	Personas contactadas	Personas no contactadas	Promedio Probabilidad de contacto
16	1.328	1027	301	0,579
17	1.328	1030	298	0,594
18	1.328	1035	293	0,614
19	1.328	1042	286	0,639
20	1.328	1078	250	0,687

Elaborado por el Instituto Nacional de Estadísticas.

### V.3.2. Ajuste por no elegibilidad

A partir de esta etapa no se consideran, para fines analíticos, aquellas personas que no conforman la población objetivo de la encuesta, es decir, personas consideradas como no elegibles por no residir en la dirección seleccionada al momento de la entrevista o nunca haber residido en dicha dirección, entre otras razones de no elegibilidad.

Este ajuste consiste en asignar un valor blanco (“missing”) en el ponderador ajustado por no contacto a aquellas personas clasificadas como no elegibles.

### V.3.3. Ajuste por no respuesta

En esta etapa solo son consideradas las personas contactadas que resultaron elegibles, pues son estas las que cumplen con las características necesarias para participar en la encuesta. Sin embargo, es posible que la persona seleccionada no desee participar o no cuente con las condiciones para responder la encuesta, como es el caso, por ejemplo, de aquellos que se encuentran impedidos física o mentalmente al momento de la entrevista. En EBS 2023, de las 24.334 personas que resultaron elegibles, 11.234 fueron entrevistadas lo que significa que la tasa de respuesta es de 46.2%.

Si en la construcción del factor de expansión no se considera el fenómeno de la no respuesta a la unidad, las estimaciones serán representativas solamente de aquellas unidades que participan en la encuesta y, por lo tanto, una parte de la población no sería cubierta. Además, existe la posibilidad de que las estimaciones sean sesgadas en caso de que exista alguna relación entre la no respuesta y la variable de interés.

El método utilizado para minimizar los problemas ocasionados por la falta de respuesta es el propensity score donde se modela la probabilidad de respuesta a partir de variables predictoras obtenidas desde Casen 2022 y de la hoja de ruta de EBS 2023. Una vez obtenida la probabilidad de respuesta, las personas se ordenan de menor a mayor probabilidad y son agrupadas en veintiles (20 grupos de igual tamaño).

En la tabla V.4 se presentan las variables utilizadas en el modelo.



**Tabla V.4. Variables utilizadas en el modelo de no respuesta EBS 2023**

<b>Variable</b>	<b>Fuente</b>
Sexo persona encuestadora	Hoja de ruta EBS 2023
Horario de contacto (mañana, tarde o noche)	Hoja de ruta EBS 2023
Cantidad de encuestas realizadas por la persona encuestadora en tramos	Hoja de ruta EBS 2023
Hogar seleccionado en EANNA 2023	Hoja de ruta EBS 2023
Región	Casen 2022
Nivel educacional más alto al que asiste o alcanzó	Casen 2022
Proporción de personas fuera de la fuerza de trabajo en el hogar	Casen 2022
Jefe/a de hogar y relación con la jefatura de hogar	Casen 2022
Número de personas ocupadas en el hogar	Casen 2022
Edad en tramos ([18-29], [30-45], [46-59], [60+])	Casen 2022
Sistema previsional de salud	Casen 2022
Condición de tenencia de la vivienda	Casen 2022
Sexo de la persona	Casen 2022
Dificultad para recordar o concentrarse	Casen 2022
Dificultad para comunicarse, comprender o ser comprendido por otros	Casen 2022
Participación en alguna organización social, política o corporativa	Casen 2022
Hogar cuenta con un vehículo de uso laboral o particular	Casen 2022

Elaborado por el Instituto Nacional de Estadísticas.

Con el modelo de predicción se espera que al interior de las 20 celdas de ajuste las personas sean similares en sus características y así, aquellas entrevistadas puedan representar a las no entrevistadas. Bajo este supuesto, los ponderadores asociados a las personas contactadas elegibles que no respondieron deben ser redistribuidos entre las personas que sí respondieron. Para estos efectos, se debe calcular una razón de ajuste.

La determinación de la razón de ajuste viene dada por la ecuación (14) y se realiza al interior de cada veintil utilizando el ponderador ajustado por elegibilidad.

En el numerador se incluye la estimación del total de personas elegibles, y en el denominador la estimación del total de personas que responden.

$$\hat{R}_{v,R} = \frac{\sum_{i \in \Phi_v} \sum_{j \in \theta_v} w_{ra,k,elec}}{\sum_{i \in \Phi_v} \sum_{j \in \theta_{v,R}} w_{ra,k,resp}} \quad (14)$$

Donde,

- $v$ : : Índice del grupo para la corrección de no respuesta (veintil).
- $\Phi_v$ : : Conjunto de personas asignadas al veintil  $v$ .
- $\theta_v$ : : Conjunto de personas pertenecientes al veintil  $v$  elegibles y contactadas.
- $\theta_{v,R}$ : : Conjunto de personas pertenecientes al veintil  $v$  que respondieron.
- $w_{ra,k,elec}$ : : Estimación del total de personas elegibles
- $w_{ra,k,resp}$ : : Estimación del total de personas que responde



Una vez estimados los factores de ajuste, el ponderador de selección corregido por no respuesta  $w_{ra,k}^{NR}$  se expresa según la ecuación (15):

$$w_{ra,k}^{NR} = \hat{R}_{v,R} \cdot w_{ra,k,elec} \quad (15)$$

Este ponderador se asigna a las 11.234 personas que respondieron y a las personas elegibles y no entrevistadas se les asigna un valor blanco (“missing”).

La Tabla V.5 presenta la distribución de los veintiles y la tasa de respuesta de cada uno de ellos.

**Tabla V.5. Veintiles ajuste por no respuesta EBS 2023**

Veintil	Personas elegibles y contactadas	Personas contactadas y no entrevistadas	Personas entrevistadas	Promedio probabilidad de respuesta
<b>Total país</b>	<b>16.374</b>	<b>5.140</b>	<b>11.234</b>	
1	819	539	280	0,194
2	819	385	434	0,340
3	819	302	517	0,388
4	819	315	504	0,417
5	819	310	509	0,439
6	819	303	516	0,458
7	819	260	559	0,474
8	819	272	547	0,489
9	819	255	564	0,504
10	819	253	566	0,519
11	819	224	595	0,532
12	819	223	596	0,546
13	819	232	587	0,559
14	819	226	593	0,573
15	818	183	635	0,587
16	818	192	626	0,603
17	818	201	617	0,621
18	818	162	656	0,642
19	818	167	651	0,668
20	818	136	682	0,726

Elaborado por el Instituto Nacional de Estadísticas.

#### V.4. Suavizamiento

El suavizamiento de los factores de expansión tiene por objetivo reducir los valores atípicos y para ello se aplica una metodología mixta que corresponde al método de contracción a la media (CM) y el método R-K.



**V.4.1. Método de contracción a la media (CM)**

Este método tiene por objetivo que la distribución de los factores sea más homogénea y para ello contrae los factores de expansión, acercándolos al valor promedio de éstos. El factor suavizado o contraído a la media (o promedio) viene dado por:

$$w_i^s = \beta \cdot w_i + (1 - \beta) \cdot \bar{w}; \quad 0 \leq \beta \leq 1 \tag{16}$$

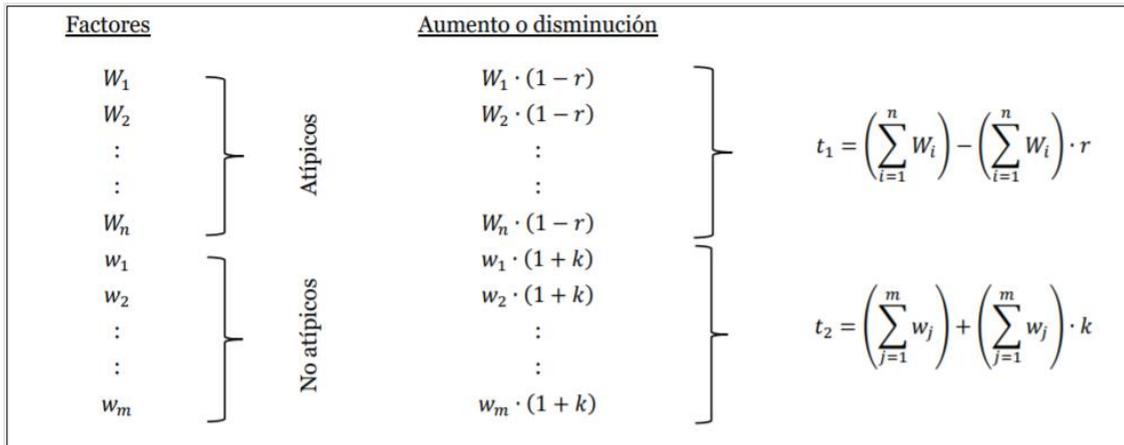
Donde,

- $\beta$  : Es un parámetro real que permite controlar la intensidad de la contracción
- $w_i$ : Factor ajustado por no respuesta
- $\bar{w}$ : Media de los factores a nivel de estrato región-área y de valores atípicos y no atípicos

**V.4.2. Método R-K**

Este método da un tratamiento diferenciado a los conjuntos de datos atípicos o no atípicos. Por un lado, genera un aumento proporcional en los factores no atípicos y, por otro lado, produce una disminución proporcional en los factores atípicos. Esto significa que el método R-K disminuye en una proporción R los valores atípicos y aumenta en una proporción K los valores no atípicos como se indica en la figura V.2.

**Figura V.2. Ordenamiento de factores de expansión atípicos y no atípicos**



Fuente: Instituto Nacional de Estadísticas.

**V.4.3. Método mixto**

El método mixto combina los métodos CM y R-K, este consiste en separar los factores de expansión en un subconjunto de factores no atípicos y otro de factores atípicos<sup>11</sup>. Para los factores no atípicos se aplica el

<sup>11</sup> Al igual que en EBS 2021, para identificar los factores atípicos se utiliza como criterio el percentil de distribución del factor de expansión, para ello se define el valor  $h = 1,5(p_{75} - p_{25})$  donde los valores con un valor inferior a  $(p_{25} - h)$  o superior a  $(p_{75} + h)$  se consideran atípicos (MDSF, 2021).



método de contracción a la media considerando un  $\beta$  de 0,8 mientras que, para los factores atípicos el  $\beta$  es de 0,2 con el fin de que, en caso de tener un único valor atípico al interior del estrato, el método CM no tenga efecto sobre ese valor y se conserve el valor del factor atípico.

Luego, se aplica el método R-K donde se unen las distribuciones tanto de factores atípicos como en un continuo que se puede lograr con valores de  $\lambda$  cercanos a la unidad. En este caso se utiliza un  $\lambda=0,99$  para conservar una pequeña distancia entre el mayor de los factores no atípicos y el menor de los factores atípicos, después del suavizamiento.

## **V.5. Calibración de factores de expansión**

---

La calibración de los factores de expansión sirve para ajustar los factores a los totales poblacionales dados por las proyecciones de población. En el caso de EBS 2023 las proyecciones de población utilizadas son al 30 de octubre de 2023, fecha que coincide con la mitad del periodo de levantamiento de la encuesta.

Para la calibración se utiliza el método *Raking* que permite incorporar información auxiliar y tiene tres propósitos, el primero de ellos es que la encuesta, a partir de los pesos calibrados, sea capaz de reproducir totales poblacionales conocidos, el segundo objetivo es alcanzar mejoras en la precisión de los parámetros de interés por medio de la obtención de estimaciones con un menor error de muestreo y finalmente, se pretende obtener una mayor precisión en las estimaciones (Alvarado, Pizarro & Guarda, 2020).

Las variables marginales de calibración utilizadas corresponden a las siguientes:

- Población total por región y área.
- Población total por región y sexo.
- Población total por tramo etario ([18-29], [30-44], [45-59], [60-79], [80+] y sexo.

Dado que la población objetivo del estudio es la que reside en viviendas particulares es necesario realizar un ajuste a las proyecciones de población. Para esto se considera la proporción de población residente en estas viviendas sobre el total de población en el país según los resultados del Censo de Población y Vivienda 2017. El ajuste se realiza al nivel de las marginales de calibración, lo que da como resultado diferentes stocks poblacionales para cada marginal por lo que es necesario realizar un ajuste adicional que toma como referencia la marginal de región-sexo y los stocks poblacionales de la marginal región-área y tramo etario-sexo son ajustados para que las tres marginales cuadren y sea posible realizar la calibración con el método *Raking*.

La Tabla V.6 muestra las proyecciones de población de 18 años y más, según región, y aquellas ajustadas a población residente en viviendas particulares utilizadas en EBS 2023.



**Tabla V.6. Proyecciones de población al 30 de octubre 2023 según región y área**

Región	Estimación de población total			Estimación de población residente en viviendas particulares		
	Urbano	Rural	Total	Urbano	Rural	Total
<b>Total</b>	<b>13.752.795</b>	<b>1.783.754</b>	<b>15.536.549</b>	<b>13.485.237</b>	<b>1.712.890</b>	<b>15.198.127</b>
Arica y Parinacota	181.407	16.404	197.811	173.635	12.136	185.771
Tarapacá	280.240	16.794	297.034	271.294	7.684	278.978
Antofagasta	521.003	20.822	541.825	489.481	4.877	494.358
Atacama	216.224	21.206	237.430	210.985	18.881	229.866
Coquimbo	544.170	122.812	666.982	534.728	116.966	651.694
Valparaíso	1.454.261	129.688	1.583.949	1.425.519	125.779	1.551.298
Metropolitana	6.325.444	235.593	6.561.037	6.229.046	228.466	6.457.512
O'Higgins	590.966	198.115	789.081	584.706	194.672	779.378
Maule	667.253	234.016	901.269	656.560	232.093	888.653
Ñuble	286.692	120.604	407.296	282.352	120.129	402.481
Biobío	1.164.864	145.060	1.309.924	1.147.900	143.129	1.291.029
La Araucanía	567.661	223.086	790.747	556.249	221.550	777.799
Los Ríos	233.513	87.643	321.156	227.601	85.977	313.578
Los Lagos	521.183	183.891	705.074	509.297	180.051	689.348
Aysén	65.644	16.010	81.654	62.105	13.080	75.185
Magallanes	132.270	12.010	144.280	123.779	7.420	131.199

Elaborado por el Instituto Nacional de Estadísticas.

El resultado del proceso de calibración se presenta en la tabla V.7 donde se observa que, luego del ajuste, los factores de expansión alcanzan un total de 15.198.127 personas.



**Tabla V.7. Distribución de la muestra y tamaño poblacional expandido por región y área, EBS 2023**

Región	Tamaño de muestra lograda	Tamaño expandido a población de 18 años y más
<b>Total</b>	<b>11.234</b>	<b>15.198.127</b>
Urbano	9.360	13.485.237
Rural	1.874	1.712.890
Arica y Parinacota	497	185.771
Tarapacá	535	278.978
Antofagasta	637	494.358
Atacama	534	229.866
Coquimbo	690	651.694
Valparaíso	859	1.551.298
Metropolitana	1.633	6.457.512
O'Higgins	693	779.378
Maule	755	888.653
Ñuble	616	402.481
Biobío	807	1.291.029
La Araucanía	726	777.799
Los Ríos	599	313.578
Los Lagos	724	689.348
Aysén	431	75.185
Magallanes	498	131.199

Elaborado por el Instituto Nacional de Estadísticas.

## V.6. Redondeo Probabilístico

Finalmente, como los resultados son referidos a personas, para obtener un número entero en cada uno de los factores de expansión, se aplica un redondeo de forma probabilística con el objetivo de que los nuevos pesos calibrados no produzcan un sesgo en las estimaciones de interés.

El redondeo probabilístico distribuye los excesos de la parte decimal en forma aleatoria con base en una distribución de probabilidades entre los factores, a partir de semillas de aleatorización diferentes para cada factor que permiten replicar el procedimiento y alcanzar el stock poblacional a nivel nacional; a diferencia del redondeo determinístico que aproxima el valor de los factores al entero más cercano.

Por último, cabe señalar que, en el caso del factor regional, para el redondeo son utilizadas las marginales de calibración región-área, región-sexo y grupo edad-sexo, sin embargo, esto no garantiza que se alcance el stock poblacional para cada una de las categorías de las marginales y solo es posible aproximarse al total nacional como fue señalado previamente.



## VI. CÁLCULO Y EVALUACIÓN DE LAS ESTIMACIONES

Este capítulo tiene como objetivo presentar los métodos de estimación de varianza en diseños muestrales complejos, en este caso un diseño bifásico elaborado a partir de Casen 2022. Además, se presenta la estimación de la variable insatisfacción con la vida y balance afectivo negativo, junto con sus errores muestrales, para los dominios de estudio definidos.

### VI.1. Estimación de la varianza compleja

---

Los métodos exactos de estimación de la varianza en diseños de muestras estándares, cuando son aplicables, son la mejor forma de estimarla. Sin embargo, los diseños muestrales empleados en encuestas sociales, como es el caso de EBS, son más complejos que el muestreo aleatorio simple o estratificado.

Un diseño de muestreo complejo se origina cuando existe más de una etapa de selección. En el caso de EBS la selección fue realizada a partir de la muestra lograda en Casen 2022<sup>12</sup>, encuesta que ya posee un diseño complejo, lo que añade otro grado de complejidad al cálculo de las estimaciones debido a la varianza adicional que significa el submuestreo de segunda fase (Gutiérrez, 2015).

En el muestreo estratificado la estimación de la varianza total se obtiene mediante la suma de las varianzas estimadas en cada estrato y la varianza en cada estrato se estima a partir de la variabilidad de la característica de interés entre los conglomerados. En este punto cabe señalar que EBS cuenta con un diseño estratificado donde los conglomerados no corresponden a una etapa en la selección, pero sí se encuentran presentes en el diseño de EBS ya que estos se obtienen desde Casen 2022.

Una aproximación para la estimación de varianza comúnmente utilizada en encuestas complejas, de muestreo polietápico por conglomerados, corresponde al método de estimación de varianza por conglomerado último (EVCU) o por reemplazo (WR por sus siglas en inglés). Este método permite capturar cierto grado de covarianza por medio de la correlación intraclásica de las unidades al interior de los conglomerados, aun cuando las estimaciones se basen en muestras sin traslape. También es posible la captura de covarianza a través de la correlación intraclásica de las unidades al interior de los estratos o dominios comunes (INE, 2022).

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

---

<sup>12</sup> Casen 2022 presenta un diseño muestral probabilístico, estratificado y bietápico.



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

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

1. La variable de ponderación de la muestra o factor de expansión designado “fexp” donde el nivel de expansión territorial es regional.
2. La variable de estratificación correspondiente a la combinación de región-área en EBS 2023.
3. La variable UPM (o pseudo-conglomerado) designada como “varunit” obtenida desde Casen 2022.

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

## **VI.2. Variables que identifican el diseño muestral complejo**

---

En la muestra de EBS 2023, para identificar los estratos definidos en el diseño se utilizan las variables “glosa\_region” y “glosa\_area”. En relación con los conglomerados, la selección de personas se realizó directamente al interior de cada estrato, sin embargo, el diseño de la primera fase (Casen 2022), sí contempla conglomerados por lo que, al momento de realizar las estimaciones, es necesario considerarlos en la definición del diseño complejo a través de la variable “varunit” obtenida desde Casen 2022. En total, en la muestra lograda las personas se distribuyen en 32 estratos y 7.576 conglomerados.

## **VI.3. Estadísticos asociados a los parámetros de interés**

---

A partir de la muestra lograda en EBS 2023, se calculan los estadísticos asociados a parámetros de interés definidos por el Ministerio y que corresponden a las siguientes variables:

- Insatisfacción con la vida (a1) correspondiente al porcentaje de personas con la categoría “Totalmente insatisfecho” o “Insatisfecho” con su vida.



- Balance afectivo negativo (affective\_sw) que corresponde al porcentaje de personas que presentan un balance negativo, es decir, las emociones negativas (enojado(a), triste) superan las positivas (tranquilo(a), feliz)<sup>13</sup>.

Las tablas VI.1 y VI.2 presentan la estimación y errores de la variable insatisfacción con la vida y balance afectivo negativo, respectivamente, a nivel nacional, por área y a nivel regional.

**Tabla VI.1 Errores muestrales por región y área para la variable Insatisfacción con la vida, EBS 2023**

Región	Insatisfacción con la vida		
	Estimación	Error absoluto	Error relativo
<b>Nacional</b>	11,0%	0,8%	7,4%
Urbano	11,2%	0,9%	7,9%
Rural	10,1%	1,9%	19,1%
Arica y Parinacota	11,3%	2,9%	25,5%
Tarapacá	10,2%	2,7%	26,2%
Antofagasta	10,9%	2,7%	24,8%
Atacama	8,8%	2,6%	29,5%
Coquimbo	7,8%	2,1%	26,7%
Valparaíso	10,5%	2,1%	20,1%
Metropolitana	12,2%	1,7%	13,6%
O'Higgins	10,2%	2,3%	22,3%
Maule	10,9%	2,2%	20,5%
Ñuble	11,4%	2,7%	23,5%
Biobío	10,8%	2,2%	20,8%
La Araucanía	10,4%	2,3%	21,8%
Los Ríos	10,7%	2,5%	23,7%
Los Lagos	8,8%	2,4%	27,4%
Aysén	8,5%	2,8%	32,6%
Magallanes	6,8%	2,3%	34,2%

Elaborado por el Instituto Nacional de Estadísticas.

<sup>13</sup> El balance negativo se construye considerando las personas que presentan valores negativos en la variable affective\_sw.



**Tabla VI.2. Errores muestrales por región y área para la variable balance afectivo negativo, EBS 2023**

Región	Balance afectivo negativo		
	Estimación	Error absoluto	Error relativo
<b>Total</b>	9,0%	0,7%	8,0%
Urbano	9,2%	0,8%	8,6%
Rural	7,9%	1,5%	18,7%
Arica y Parinacota	11,1%	2,8%	25,0%
Tarapacá	10,8%	2,7%	24,8%
Antofagasta	8,2%	2,3%	27,7%
Atacama	7,8%	2,4%	31,1%
Coquimbo	7,8%	2,1%	26,2%
Valparaíso	10,4%	2,1%	20,4%
Metropolitana	8,9%	1,4%	16,1%
O'Higgins	10,8%	2,3%	20,9%
Maule	10,6%	2,3%	21,2%
Ñuble	8,9%	2,4%	26,5%
Biobío	8,6%	2,0%	22,7%
La Araucanía	7,3%	2,0%	27,0%
Los Ríos	8,5%	2,2%	26,3%
Los Lagos	7,4%	2,1%	28,1%
Aysén	6,5%	2,3%	35,6%
Magallanes	6,9%	2,1%	30,8%

Elaborado por el Instituto Nacional de Estadísticas.



## VII.ANEXOS

## VII.1. Anexo N°1: Códigos de Disposición Final de casos EBS 2023

Glosa código de disposición final	Total	Porcentaje
<b>Total</b>	<b>26.568</b>	<b>100%</b>
<b>Personas Entrevistadas</b>	<b>11.489</b>	
11. entrevista completa	11.234	42,3%
12. entrevista parcial	255	1,0%
<b>Personas elegibles, no entrevistados</b>	<b>4.885</b>	
21. rechazo nivel hogar	1.156	4,4%
22. rechazo informante	2.459	9,3%
41. persona informante impedido(a) física y/o mentalmente para contestar	660	2,5%
42. persona informante no habla idioma español	15	0,1%
43. muerte de la persona informante o de algún(a) familiar	144	0,5%
44. concertó cita con la persona informante	57	0,2%
45. información falseada	68	0,3%
46. pérdida o errores en la transmisión de datos	17	0,1%
49. otra razón de no entrevista	114	0,4%
81. calidad de audio mala o intermitente	37	0,1%
82. lugar o actividad de la persona informante no permite una entrevista	158	0,6%
<b>Personas elegibles, no contactadas</b>	<b>7.960</b>	
33. persona informante no ubicable, se contacta a otra persona	1.114	4,2%
71. marca tono, pero sin respuesta	1.625	6,1%
72. responde buzón de voz	1.232	4,6%
73. suena ocupado	423	1,6%
74. se encuentra apagado o fuera del área de cobertura	1.061	4,0%
75. número telefónico no existe	1.726	6,5%
76. número telefónico equivocado	673	2,5%
78. no se gestionó la llamada	2	0,0%
79. otra razón de no contacto	104	0,4%
<b>Personas no elegibles</b>	<b>2.234</b>	
61. establecimiento, oficina de gobierno u otra organización	5	0,0%
62. institución (por ejemplo: hospital, cárcel, residencias para personas mayores, etc.)	2	0,0%
63. dormitorio colectivo (por ejemplo: militar, de trabajo, internado, hostales, residenciales, etc.)	5	0,0%
64. vivienda en demolición, incendiada, destruida o erradicada	5	0,0%
65. vivienda desocupada (por ejemplo: en venta, arriendo, etc.)	3	0,0%
66. vivienda de veraneo o de uso temporal	6	0,0%
67. edificación no calificable (por ejemplo: en construcción, en reparaciones, etc.)	1	0,0%
69. otra razón para ser considerada no elegible	28	0,1%
91. persona informante/hogar ya no reside en la dirección seleccionada	1.588	6,0%
92. persona informante/hogar nunca residió en la dirección seleccionada	591	2,2%

Elaborado por el Instituto Nacional de Estadísticas.



**Anexo N°2: Sintaxis en R para la estimación de las variables de interés****1. Cargar base**

```
library(openxlsx)
library(haven)
library(tidyverse)
library(survey)

rm(list = ls())
options(scipen = 999)
setwd("C:/Users/ifmedinas/OneDrive - Instituto Nacional de Estadísticas/Documentos/EBS 2023")
load("Estimación/BD Estimación EBS 2023/Base de datos EBS 2023.RData")
EBS2023 <- EBS_2023
```

**2. Crear variables para estimación**

```
EBS2023 <- EBS2023 %>% mutate(estrato = paste0(glosa_region,"-",glosa_area),
  insatis = as.factor(iffelse(a1 == 1 | a1 == 2,1,0)),
  a2 = as.factor(case_when(affective_sw < 0 ~ 1,
    affective_sw >= 0 & affective_sw <= 8 ~ 0,
    .default = NA)),
  cod_region = case_when(glosa_region == "Tarapaca" ~1,
  glosa_region == "Antofagasta" ~2,
  glosa_region == "Atacama" ~3,
  glosa_region == "Coquimbo" ~4,
  glosa_region == "Valparaiso" ~5,
  glosa_region == "Libertador General Bernardo O'higgins" ~6,
  glosa_region == "Maule" ~7,
  glosa_region == "Biobio" ~8,
  glosa_region == "La Araucania" ~9,
  glosa_region == "Los Lagos" ~10,
  glosa_region == "Aysen Del General Carlos Ibañez Del Campo" ~11,
  glosa_region == "Magallanes Y De La Antartica Chilena" ~12,
  glosa_region == "Metropolitana De Santiago" ~13,
  glosa_region == "Los Rios" ~14,
  glosa_region == "Arica Y Parinacota" ~15,
  glosa_region == "Ñuble" ~16,
  .default = NA),
  cod_area = iffelse(glosa_area == "Urbano",1,2),
  aux = 1)
```

**3. Diseño para estimación con diseño original EBS (1 fase) y conglomerado Casen**

```
d_1f_upm <- svydesign(ids=~varunit,
  strata = ~estrato, #region-área
  weights = ~fexp,
  data= EBS2023)
```



#### 4. Estimación variable insatisfacción con la vida

##### 4.1 Estimación nacional

```
###Tasa
nac_prop <- svyby(~insatis, ~aux, d_1f_upm, na.rm = TRUE, level = 0.95,
  vartype = c("se", "cv"), svymean) %>%
  mutate(region = ifelse(aux==1, "Nacional", 0)) %>%
  select(-c("aux"))

###GL
GL_nac <- data.frame(region = "Nacional", GL = degf(d_1f_upm))
```

##### 4.2 Estimación regional

```
###Tasa
reg_prop <- svyby(~insatis, ~cod_region, d_1f_upm, na.rm = TRUE, level = 0.95,
  vartype = c("se", "cv"), svymean) %>%
  rename(region = cod_region) %>%
  mutate(region = as.factor(region))

###GL por región
#n varunit - 2 (U-R)
gl_reg <- 1:16 %>% lapply(function(r) {
  design <- subset(d_1f_upm, cod_region == r)
  data.frame(region = as.factor(r), GL = degf(design))}) %>%
  bind_rows()
```

##### 4.3 Estimación por área

```
###Tasa
area_prop <- svyby(~insatis, ~cod_area, d_1f_upm, na.rm = TRUE, level = 0.95,
  vartype = c("se", "cv"), svymean) %>%
  rename(region = cod_area) %>%
  mutate(region = ifelse(region == 1, "Urbano", "Rural"))

###GL por área
#n varunit - 16 (regiones)
gl_area <- data.frame(region = c("Urbano", "Rural"),
  GL = sapply(1:2, function(a) degf(subset(d_1f_upm, cod_area == a))))
```

##### 4.4 Totales muestra

```
n_reg <- EBS2023 %>% group_by(cod_region) %>%
  summarise(n = n()) %>%
  rename(region = cod_region)

n_area <- EBS2023 %>% group_by(cod_area) %>%
  summarise(n = n()) %>%
  rename(region = cod_area) %>%
  mutate(region = ifelse(region == 1, "Urbano", "Rural"))

n_nac <- data.frame(region = "Nacional", n = nrow(EBS2023))

n_ebs <- rbind(n_nac, n_area, n_reg)
```



## 4.5 Unir resultados

```
prop_a1 <- bind_rows(nac_prop, area_prop, reg_prop)
gl <- bind_rows(GL_nac, gl_area, gl_reg)

a1 <- left_join(prop_a1, gl, by = "region") %>%
  left_join(n_ebs, by = "region")
```

## 4.6 Calcular error absoluto y relativo y aplicar estándar de calidad INE

```
a1 <- a1 %>% select(7,2,4,6,8,9) %>%
  rename(SE = se.insatis1) %>%
  mutate(Ts = qt(0.975, GL),
         error_abs = Ts * SE,
         error_rel = error_abs/insatis1,
         SE_max = ifelse(insatis1 < 0.5,
                        (insatis1^(2/3))/9,
                        ((1-insatis1)^(2/3))/9)),
         est_calidad = case_when(n >= 60 & GL >= 9 & SE < SE_max ~ "fiable",
                                n >= 60 & GL >= 9 & SE > SE_max ~ "poco fiable",
                                n < 60 | GL < 9 ~ "no fiable",
                                TRUE ~ NA))
```

**5. Estimación variable balance afectivo negativo**

## 5.1 Estimación nacional

```
###Tasa
nac_prop <- svyby(~a2, ~aux, d_1f_upm, na.rm = TRUE, level = 0.95,
                vartype = c("se", "cv"), svymean) %>%
  mutate(region = ifelse(aux==1, "Nacional", 0)) %>%
  select(-c("aux"))
```

```
###GL
GL_nac <- data.frame(region = "Nacional", GL = degf(d_1f_upm))
```

## 5.2 Estimación Regional

```
###Tasa
reg_prop <- svyby(~a2, ~cod_region, d_1f_upm, na.rm = TRUE, level = 0.95,
                vartype = c("se", "cv"), svymean) %>%
  rename(region = cod_region) %>%
  mutate(region = as.factor(region))
```

```
###GL por región
#n varunit - 2 (U-R)
gl_reg <- 1:16 %>% lapply(function(r) {
  design <- subset(d_1f_upm, cod_region == r)
  data.frame(region = as.factor(r), GL = degf(design))}) %>%
  bind_rows()
```



## 5.3 Estimación por área

```

###Tasa
area_prop <- svyby(~a2, ~cod_area, d_1f_upm, na.rm = TRUE, level = 0.95,
  vartype = c("se", "cv"), svymean) %>%
  rename(region = cod_area) %>%
  mutate(region = ifelse(region == 1, "Urbano", "Rural"))

###GL por área
#n varunit - 16 (regiones)
gl_area <- data.frame(region = c("Urbano", "Rural"),
  GL = sapply(1:2, function(a) degf(subset(d_1f_upm, cod_area == a))))

```

## 5.4 Totales muestra

```

n_reg <- EBS2023 %>% group_by(cod_region) %>%
  summarise(n = sum(!is.na(a2))) %>%
  rename(region = cod_region)

n_area <- EBS2023 %>% group_by(cod_area) %>%
  summarise(n = sum(!is.na(a2))) %>%
  rename(region = cod_area) %>%
  mutate(region = ifelse(region == 1, "Urbano", "Rural"))

n_nac <- data.frame(region = "Nacional", n = sum(!is.na(EBS2023$a2)))

n_ebs <- rbind(n_nac, n_area, n_reg)

```

## 5.5 Unir resultados

```

prop_a2 <- bind_rows(nac_prop, area_prop, reg_prop)
gl <- bind_rows(GL_nac, gl_area, gl_reg)

a2 <- left_join(prop_a2, gl, by = "region") %>%
  left_join(n_ebs, by = "region")

```

## 5.6 Calcular error absoluto y relativo y aplicar estándar de calidad INE

```

a2 <- a2 %>% select(7,2,4,6,8,9) %>%
  rename(SE = se.a21,
  a2 = a21) %>%
  mutate(Ts = qt(0.975, GL),
  error_abs = Ts * SE,
  error_rel = error_abs/a2,
  SE_max = ifelse(a2 < 0.5,
    (a2^(2/3))/9,
    ((1-a2)^(2/3))/9),
  est_calidad = case_when(n >= 60 & GL >= 9 & SE < SE_max ~ "fiable",
    n >= 60 & GL >= 9 & SE > SE_max ~ "poco fiable",
    n < 60 | GL < 9 ~ "no fiable",
    TRUE ~ NA))

```



**6. Guardar resultados**

```
{wb <- createWorkbook("Estimaciones EBS")

  addWorksheet(wb, sheetName = "a1")
  addWorksheet(wb, sheetName = "a2")

  writeData(wb,1,a1)
  writeData(wb,2,a2)}

xlsx <- paste0(paste0("Estimación/BD Estimación EBS 2023/",gsub("-", "", Sys.Date()), "_", "Estimación EBS 2023_informe.xlsx"))
saveWorkbook(wb, xlsx, overwrite = TRUE)
```



**REFERENCIAS**

- Alvarado, M., Pizarro, M. & Guarda, N. (2020). Documentos de trabajo. La nueva metodología de calibración de la Encuesta Nacional de Empleo: método de calibración Raking. Obtenido de: [https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/documentos-de-trabajo/documentos-de-trabajo/calibracion-factores-ene-metodologia.pdf?sfvrsn=cc1f6a74\\_8](https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/documentos-de-trabajo/documentos-de-trabajo/calibracion-factores-ene-metodologia.pdf?sfvrsn=cc1f6a74_8)
- Cochran, W. (1998). *Técnicas de Muestreo*. México D.F.: Compañía Editorial Continental, S.A. de C.V.
- Gutiérrez, A. (2015). *Estrategias de Muestreo, Diseño de Encuestas y Estimación de Parámetros*. Bogotá.
- Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied Survey Data Analysis*. USA: Chapman & Hall / CRC Press.
- INE. (marzo de 2020). [www.ine.gob.cl](http://www.ine.gob.cl). Obtenido de <https://www.ine.gob.cl/inicio/documentos-de-trabajo/documento/fundamentos-del-est%C3%A1ndar-para-la-evaluaci%C3%B3n-de-la-calidad-de-las-estimaciones-en-encuestas-de-hogares>
- INE. (2022). *Separata técnica. Metodología de estimación anual Encuesta Nacional de Empleo (ENE)*. Obtenido de <https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/publicaciones-y-anuarios/separatas/tematicas/estimacion-anual-2022.pdf>
- Kish, L. (1965). *Survey Sampling*. Nueva York: John Wiley & Sons.
- MDSF. (2021). *Metodología de Diseño Muestral EBS 2021*.



Encuesta de Bienestar Social

**EBS**

2 0 2 3