



2023

EANNA

ENCUESTA DE ACTIVIDADES DE NIÑOS, NIÑAS Y ADOLESCENTES

INFORME DE DISEÑO MUESTRAL

Instituto Nacional de Estadísticas – Ministerio de Desarrollo Social y Familia

Junio 2023

Informe revisado y editado por el Ministerio de Desarrollo Social y Familia a partir de Informe Diseño Muestral de la Encuesta de Actividades de Niños, Niñas y Adolescentes 2023, entregado por el Instituto Nacional de Estadísticas (INE), en el marco del convenio *“Programa de trabajo conjunto para la elaboración del diseño para la Encuesta de Actividades de Niños, Niñas y Adolescentes (EANNA 2023) y la Encuesta de Bienestar Social (EBS 2023).”*

ÍNDICE

I.	ANTECEDENTES	4
II.	CARACTERÍSTICAS GENERALES DEL DISEÑO MUESTRAL DE EANNA	6
	II.1. Objetivo	6
	II.2. Población objetivo	6
	II.3. Unidad de información	6
	II.4. Marco muestral	6
	II.5. Dominios de estudio	7
	II.6. Objetivo de precisión y tamaño de muestra	7
	II.7. Estrategia muestral	8
III.	CÁLCULO Y DISTRIBUCIÓN DEL TAMAÑO MUESTRAL	9
	III.1. Etapas para la estimación del tamaño muestral para EANNA 2023	10
	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	21
	III.6. Tamaño muestral propuesto	22
	III.7. Errores esperados	24
IV.	SELECCIÓN DE UNIDADES MUESTRALES	25
V.	REFERENCIAS	27

ÍNDICE DE TABLAS

Tabla 1.	Resumen de parámetros utilizados y tamaños y errores resultantes según escenario, EANNA 2023	20
Tabla 2:	Estimación de tamaños muestrales según proyecciones de población al 15 de junio de 2023 con marco Casen 2022 actualizado	21
Tabla 3:	Tamaños y errores finales a nivel nacional, regional, urbano-rural, EANNA 2023	23
Tabla 4:	Tamaños y errores EANNA 2012 y propuesta EANNA 2023. Nacional Urbano - Rural	23
Tabla 5:	Errores esperados asociados al parámetro de interés según tamaño muestral objetivo ..	24
Tabla 6:	Distribución del número de NNA por región y área	26

ÍNDICE DE CUADROS

Cuadro 1.	Parámetros utilizados para el cálculo de tamaño muestral actual	14
Cuadro 2:	Parámetros comunes utilizados en escenarios EANNA 2023	19

I. ANTECEDENTES

La Encuesta Nacional sobre Actividades de Niños, Niñas y Adolescentes (en adelante EANNA) es un proyecto elaborado en conjunto por el Ministerio de Desarrollo Social y Familia (MDSF), el Ministerio del Trabajo y Previsión Social (Mintrab), la Organización Internacional del Trabajo (OIT) y el Fondo de Naciones Unidas para la Infancia (UNICEF).

El objetivo de la EANNA es recopilar información sobre la magnitud y principales características del trabajo infantil y adolescente en Chile (TI) así como actualizar el diagnóstico proporcionado por EANNA 2012 sobre trabajo infantil y las principales actividades que realizan los niños, niñas y adolescentes.

La encuesta constituye, además, un instrumento importante para el diseño, implementación, seguimiento y evaluación de políticas y programas sociales que se ejecutan en el país, especialmente en materias relacionadas con el trabajo infantil y las principales actividades que realizan los niños, niñas y adolescentes.

El trabajo de campo de la encuesta EANNA 2020 fue planificado para el mes de marzo del 2020, sin embargo, dadas las condiciones del país, relativas al Coronavirus y las restricciones impuestas de distanciamiento social, suspensión de clases y cuarentenas, fue necesario suspenderlo. La nueva versión de la encuesta será recolectada entre junio y agosto de 2023, para lo cual, el Ministerio de Desarrollo Social y Familia (MDSF) y el Instituto Nacional de Estadísticas (INE) han convenido un programa de trabajo conjunto para la elaboración del diseño muestral de la Encuesta de Actividades de Niños, Niñas y Adolescentes, EANNA 2023.

EANNA es una encuesta bifásica de la Encuesta de Caracterización Socioeconómica Nacional (Casen). Esta última se desarrolla generalmente, de manera bienal 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 individuos (niños, niñas y adolescentes) que serán encuestados en EANNA. Así, en el análisis y estudio de los datos levantados en la realización de esta encuesta, se tendrá la caracterización completa del hogar encuestado para conocer de mejor manera la magnitud del trabajo infantil y, a partir de la complementariedad con la encuesta Casen, profundizar en la caracterización de los hogares de los niños, niñas y adolescentes entrevistados.

El presente informe es parte de los productos entregables del convenio INE-MDSF y presenta el diseño metodológico elaborado para EANNA con el fin de responder a los objetivos propios de la encuesta, requerimientos que se deben considerar al momento de determinar un tamaño muestral

adecuado. Además, el tamaño muestral debe balancear en su diseño el presupuesto disponible y los errores de muestreo considerados como admisibles para fines del estudio. La conjugación de estos tres elementos (requerimientos, errores muestrales admisibles y presupuesto disponible) permite obtener tamaños muestrales óptimos¹ para posteriormente realizar la selección de las unidades muestrales.

La metodología del diseño muestral de la EANNA 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 los niños, niñas y adolescentes (NNA) de Chile y que permita estimar la tasa de trabajo infantil en cada región del territorio nacional, así como también, en las áreas geográficas urbanas y rurales a nivel nacional y para el país en su conjunto.

Este informe corresponde a la Entrega N°6 establecida en el convenio INE-MDSF que contempla el desarrollo de los siguientes capítulos:

- Características generales del diseño muestral de EANNA
- Cálculo y distribución del tamaño muestral
- Selección de unidades muestrales

¹ Hablar de tamaños muestrales óptimos no denota un proceso matemático de maximización/minimización de una función objetivo, sino el proceso de las múltiples iteraciones para las distintas simulaciones, distintos parámetros y requerimientos. Así como también los acuerdos llegados entre el INE y la contraparte del Ministerio de Desarrollo Social y Familia en las reuniones de trabajo sostenidas.

II. CARACTERÍSTICAS GENERALES DEL DISEÑO MUESTRAL DE EANNA

II.1. Objetivo

Recopilar información sobre la magnitud y principales características del trabajo infantil y adolescente en Chile, así como actualizar el diagnóstico (proporcionado por EANNA 2012) sobre trabajo infantil y las principales actividades que realizan los niños, niñas y adolescentes y que favorecen el ejercicio de sus derechos.

II.2. Población objetivo

La población objetivo es el universo de niños, niñas y adolescentes (NNA) de entre 5 y 17 años cumplidos a la fecha de realización de la encuesta, que habitan el territorio nacional y son representados por la Encuesta de Caracterización Socioeconómica Nacional (Casen) 2022.

II.3. Unidad de información

La EANNA recoge información directamente de los niños entre 5 y 17 años, que permite identificar quienes están en situación de trabajo infantil, conocer sus características y analizar las condiciones en que desarrollan la actividad laboral. Al mismo tiempo, mediante entrevistas a los padres o adultos responsables se obtiene información sobre sus hogares y sobre su percepción respecto del trabajo por parte de los niños.

II.4. Marco muestral

Los NNA entrevistados en EANNA 2023 son una muestra que proviene de los individuos pertenecientes a 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.

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 2017².

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 consideró 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 EANNA 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 presencial, con un tamaño objetivo ajustado de 18.069 NNA, se espera obtener estimaciones del parámetro de interés con un error absoluto nacional de 0,6% y, un error relativo, de 8,5%. Con este tamaño los errores absolutos regionales no superan 3,4% y, los errores relativos, 51,1%. 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.

Dados los tamaños muestrales objetivos a nivel nacional y regional, para la distribución de los tamaños a nivel de estratos (región-área-tramo etario) se establece un mínimo de 10 NNA, especialmente, en áreas rurales.

² 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>.

II.7. Estrategia muestral

El diseño muestral de EANNA 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 y nivel socioeconómico³. La segunda fase se compone de una muestra probabilística estratificada, siendo los estratos muestrales conformados por región-área-tramo etario⁴.

³ El nivel socioeconómico se incorpora para algunas combinaciones de comuna-área.

⁴ Se consideran 4 tramos etarios: 5 a 8 años, 9 a 11 años, 12 a 14 años y 15 a 17 años.

III. CÁLCULO Y DISTRIBUCIÓN DEL TAMAÑO MUESTRAL

Entre enero y marzo de 2023, el INE tuvo la misión de desarrollar la estrategia de muestreo de EANNA 2023 que consistió en proponer y elaborar, de acuerdo con los requerimientos planteados por el MDSF, simulaciones de tamaños muestrales con los términos específicos de errores muestrales aceptables y distribución muestral, optimizando la selección de las unidades muestrales, de acuerdo con criterios de factibilidad y de presupuesto.

Durante esta etapa, previa al desarrollo de la estrategia de trabajo de campo, se estableció 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 EANNA 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 EANNA 2012, 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 estableció 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 los 33.989 NNA que fueron resultado del levantamiento de Casen 2022⁵. El error absoluto propuesto para cumplir con el estándar de calidad y el orden de la distribución regional de la población es de 0,6% a nivel nacional. Con errores absolutos entre 1,5% y 3,4% a nivel regional se obtiene un tamaño bajo muestreo aleatorio simple de 8.072 NNA a nivel nacional.

Para obtener el efecto de diseño se contempla el efecto de diseño de EANNA 2012, el efecto de diseño utilizado para Casen 2022 y el efecto de diseño resultante del cálculo de la tasa de pobreza en hogares con NNA 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 Coquimbo, Metropolitana, Ñuble, Los Ríos, Aysén y Magallanes. A nivel regional se establece que los efectos de diseño utilizados fluctúen entre 1,5 y 3,0.

⁵ Corresponde a la última actualización realizada por el MDSF con fecha 11 de abril de 2023 y puesto a disposición del INE el 20 de abril de 2023.

Luego, el tamaño objetivo ajustado por efecto de diseño es de 18.188 NNA que, al aplicar el ajuste por finitud, considerando como universo las proyecciones de población al 15 de junio de 2023⁶, resulta un tamaño de 18.066 NNA a nivel nacional. La distribución de los tamaños por estrato de muestreo (región área y tramo etario), se realiza en forma proporcional a las proyecciones de población a junio de 2023 y se establece un mínimo de 10 NNA por cada uno, aumentando la muestra a 18.069 NNA con un error absoluto es de 0,6% y un error relativo de 8,5%. Si bien a nivel regional se observan errores relativos elevados, a partir del error estándar se aprecia que estos valores cumplen con el estándar de calidad.

Finalmente, para obtener el tamaño con sobremuestreo, se utiliza una tasa de no logro a nivel regional donde se selecciona el máximo entre: la tasa de no logro de Casen 2022, ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y Casen 2022, que se pondera con EANNA 2012, que equivale a 60% y, Casen a 40%. Este procedimiento dio lugar a un tamaño con sobremuestreo de 23.489 NNA.

III.1. Etapas para la estimación del tamaño muestral para EANNA 2023

El cálculo del tamaño muestral para EANNA 2023 inicia con la obtención de parámetros asociados a la última aplicación de la encuesta (EANNA 2012) en los niveles de estimación definidos para el nuevo estudio (EANNA 2023). Así, a nivel regional se obtiene la tasa de trabajo infantil junto con sus estadísticos asociados (error estándar y efecto del diseño). Luego, se prevé una serie de procedimientos que corresponden 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 definir el tamaño muestral con el total de NNA a encuestar, en principio, a nivel de región, luego a nivel de estrato de muestreo (región-área-tramo etario). Obtenidos los tamaños muestrales para el escenario elegido se determina un tamaño de muestreo por estrato de manera tal que existan, al menos, 10 NNA a encuestar por estrato.

El cálculo del tamaño para una encuesta que presenta un cambio en su diseño muestral, debe considerar la obtención de un tamaño bajo m.a.s. para luego, aplicar los correspondientes ajustes según el efecto del nuevo diseño. A continuación, se detallan los pasos a seguir:

⁶ Estas proyecciones consideran el 15 de junio de 2023 como fecha de referencia; fecha que se encuentra dentro del periodo estipulado para el levantamiento.

Paso 1: Estimación de parámetros

Se obtienen estimaciones, a nivel regional, de la tasa de trabajo infantil (TI) junto con el error estándar y el efecto del diseño asociado a este indicador, usando los resultados del trabajo de campo de EANNA 2012. A partir de estos parámetros se obtiene la cuasivarianza poblacional $S(r)_r^2$ según la ecuación (1):

$$S(ti)_r^2 = \frac{n \cdot \bar{m}}{Def f(ti)_r} \cdot SE(ti)_r^2 \quad (1)$$

Donde:

- n : Número de conglomerados logrados en la región r .
- m : Número promedio de NNA por conglomerado en la región r .
- $n \cdot \bar{m} = m_r$: Número de NNA logrados en la región r .
- $Def f(ti)_r$: Efecto del diseño asociado a la tasa de trabajo infantil (ti) en la región r .
- $SE(ti)$: Error estándar de la estimación de la tasa de trabajo infantil (ti) en la región r .

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

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

Donde:

- $\sigma^2(ti)_{MC}$: Cálculo de la varianza bajo un diseño complejo.
- $\sigma^2(ti)_{MAS}$: 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 encuestas que hacen uso de diseños muestrales complejos y que la estimación del efecto del diseño se encuentra 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 $Def f$ 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 el investigador 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.

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

Paso 3: Ajuste por efecto del diseño

Dado que el diseño muestral definido para EANNA 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):

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

Donde:

m_{0r} : Es el número de NNA a encuestar en la región r bajo un muestreo aleatorio simple.

m_{1r} : Es el número de NNA a encuestar en la región r considerando un diseño complejo.

$Deff(ti)_r$: Es el efecto de diseño en la región r obtenido desde la estimación de alguna variable vinculada a la tasa de trabajo infantil.

Generalmente, los valores de $Deff(ti)_r$ 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 $Deff$ 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):

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

Donde:

m_{2r} : Es el total de NNA a encuestar en la región r considerando un diseño complejo y asumiendo finitud de la población.

M_r : Es el total de NNA 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.

Paso 5: Distribución por estrato de muestreo: región – área – tramo etario

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, primero a nivel de área y luego a nivel de tramo etario.

Es importante mencionar que, el tamaño inicial calculado en cada estrato de muestreo puede ajustarse en la medida que los estratos 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 TI, 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 de la misma encuesta o de operaciones estadísticas de similar temática o método de recolección. El nivel de desagregación idóneo 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.

Para calcular el ponderador de ajuste (tasa de no logro) se utiliza el máximo entre la tasa de Casen 2022 ponderada por la proporción de la tasa de no logro (TNL) de EANNA 2011 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022 ponderando, un 60% EANNA y, un 40% Casen. Esta tasa se encuentra a nivel regional y se aplica a cada estrato de muestreo de EANNA.

Luego, el tamaño con sobremuestreo a nivel de estrato m_{3em} es calculado como se observa en la ecuación (6):

$$m_{3em} = \frac{m_{2em}}{1 - tnr_{em}} \quad (6)$$

Donde:

m_{3em} : Número total de NNA a encuestar en el estrato de muestreo em .

tnr_{em} : Tasa de no respuesta aplicada en el estrato de muestreo em .

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 em , del total de NNA de cada una de las regiones, se puede despear la tasa de no logro y estimar la tasa regional a partir de la ecuación (7):

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

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 trabajo infantil (TI) en la frontera de producción (porcentaje de NNA en situación de TI)
Estimador asociado	Estimador de razón de trabajo infantil (TI) en la frontera de producción (porcentaje de NNA en situación de TI) según resultados obtenidos de EANNA 2012: $ti = \frac{\text{Número de NNA en TI}}{\text{Número total de NNA}}$
Tamaño de la población objetivo	Población estimada de niños, niñas y adolescentes (NNA) y de los hogares que los contienen por región, área y tramos etario (5 a 8 años, 9 a 11 años, 12 a 14 años y 15 a 17 años) a partir de las proyecciones de población oficiales para el periodo de recolección de la encuesta (15 de junio de 2023).
Variable de diseño	$X_j = \begin{cases} 0, & \text{si el NNA no está en situación de TI} \\ 1, & \text{si el NNA está en situación de TI} \end{cases}$

Parámetro	Descripción
Niveles de estimación	Nacional Nacional urbano, Nacional rural Regional
Errores de muestreo	<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> $\text{Máximo ee tolerable} = \begin{cases} \sqrt[3]{r^2/9} & ; r < 0,5 \\ \sqrt[3]{(1-r)^2/9} & ; r \geq 0,5 \end{cases}$
Cuasivarianza	EANNA 2012
Efecto del diseño	<p>Se utiliza el máximo entre:</p> <ol style="list-style-type: none"> 1. El efecto obtenido en EANNA 2012 2. El efecto obtenido para la tasa de pobreza en viviendas con NNA en Casen Piloto cambio de marco (CM). 3. El efecto utilizado para el cálculo del tamaño muestral de Casen 2022. <p>Luego, se define un mínimo de 1,5 y un máximo de 3,0</p>
Tasa de sobremuestreo	Obtenida como el máximo entre la tasa de Casen 2022, ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022, ponderando 60% con la TNL de EANNA 2012 y 40% para Casen.

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 EANNA 2023, hacen necesaria la confección de un diseño eficiente para la selección de una muestra de NNA, en términos de ejecución presupuestaria e implementación operativa.

El muestreo aleatorio simple de viviendas necesitaría, en principio, de un marco de NNA 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, los NNA seleccionados de un marco de este tipo, se encontrarían tan dispersos 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 NNA) se determinan después de varias etapas; en consecuencia, no es necesario mantener un marco de NNA actualizado, lo cual reduce significativamente los costos.

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). 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 de la pobreza en las unidades secundarias de muestreo o como la correlación de las unidades secundarias al interior de los conglomerados, respecto 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 aplicáramos 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 EANNA 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.

La obtención de tamaños muestrales para EANNA 2023 comienza con la determinación del tamaño bajo muestreo aleatorio simple y, a partir de este tamaño se obtiene un error de estimación. Luego, 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 EANNA 2012.

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.

De este modo, para EANNA 2023 se obtuvieron tamaños de muestra regionales, asegurando representatividad a nivel nacional, nacional urbano, nacional rural y regional.

Posteriormente, se realiza un procedimiento mediante el cual se distribuye el tamaño de muestra de cada región por el estrato de muestreo que incluye el área y tramo etario 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 junio de 2023.

Bajo esta distribución, los tamaños con sobremuestreo son calculados a partir de una tasa de no logro que considera el máximo entre la tasa de no logro de Casen 2022, ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022, ponderando 60% para EANNA y 40% para Casen. Finalmente, para que la muestra sea eficiente, en términos de la cantidad de NNA 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, es decir, desagregación por área (urbana – rural) y por estratos de muestreo (región - área - tramo etario).

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 EANNA 2012, desde donde se obtiene la estimación de la tasa de trabajo infantil 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 que se describen a continuación:

- Escenario MDSF: Este escenario fue formulado por el MDSF y utiliza como parámetros base los resultados de EANNA 2012, sin embargo, no considera el efecto que provoca el nuevo diseño muestral de la encuesta, así como tampoco la corrección por población finita. Además, en la obtención de los errores de muestreo no se consideran las orientaciones del estándar de calidad que utiliza el INE.
- Escenario n°0: Utiliza como marco muestral los NNA de Casen en Pandemia al no contar, al momento de la simulación, con los resultados de Casen 2022 y, a su vez, presenta información relativa a los parámetros de EANNA 2012 desde donde se obtuvo una estimación de la tasa de trabajo infantil nacional de 6,6%, asociada a un error absoluto de 0,8% y a un error relativo de 11,8%.
- Escenario n°1: Esta simulación considera como marco muestral la estimación de los NNA de Casen en Pandemia 2020 y se utilizan los totales de NNA logrados en Casen en Pandemia 2020 para el cálculo de la fracción de muestreo.
- Escenario n°1.1: Mantiene los errores y efectos del diseño utilizados en el escenario n°1 y 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 enero de 2023.
- Escenario n°2: Utiliza como insumo para la corrección por población finita las proyecciones de población al 15 de junio 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. Además, este escenario presenta un ajuste en la distribución urbano-rural debido a que no todos los estratos poseen suficientes NNA para seleccionar.
- Escenario n°2.1: Este escenario presenta una diferencia con respecto al escenario n° 2 y se refiere a una actualización del marco muestral a utilizar para el cálculo de las fracciones de muestreo. La fecha de actualización por parte de MDSF corresponde al 16 de marzo de 2023, que fue recibida por el INE el 22 de marzo de 2023.
- Escenario n°2.2: Este escenario es el seleccionado. Utiliza los mismos parámetros que el escenario n°2, sin embargo, las fracciones de muestreo son obtenidas a partir del marco actualizado por MDSF al 11 de abril de 2023 y recibido por el INE, el 20 de abril de 2023.

Obtenidos los tamaños muestrales objetivos del escenario elegido (escenario n°2.2) 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 EANNA 2023

Parámetro	Descripción
Parámetro a estimar	Tasa de trabajo infantil (TI)
Variable de diseño	$X_j = \begin{cases} 0, & \text{si el NNA no está en situación de TI} \\ 1, & \text{si el NNA está en situación de TI} \end{cases}$
Estimador asociado	Estimador de razón: $t_i = \frac{\text{Número de NNA que trabajan}}{\text{Número total de NNA}}$

Elaborado por el Instituto Nacional de Estadísticas.

La Tabla 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 EANNA 2012 (escenario n°0) y luego, presentar el resto de los escenarios.

Tabla 1. Resumen de parámetros utilizados y tamaños y errores resultantes según escenario, EANNA 2023

Nro.	N° NNA según marco utilizado	Niveles de estimación	Error Absoluto	Error Relativo	Efectos de Diseño	Tamaño Muestral Objetivo	Tamaño Muestral Objetivo (ajustado) ⁷	Tasa de no logro	Tamaño con sobremuestreo
MDSF		Nacional		8,5%			18.470	TNR EANNA 2012 corregida por la variación entre la TNR de Casen 2011 y el máximo entre la TNR Casen 2017 y la TNR Cambio Marco	21.782
		Regional		18,2% - 45,5%					
0	32.499	Nacional	0,8%	11,8%	2,5	6.830			
		Regional	1,4% - 6,3%	24,2% - 61,2%	0,7 - 3,4				
1	3.435.717	Nacional	0,6%	8,5%	2,1	15.938	15.940	TNR EANNA 2012 corregida por la variación entre la TNR de Casen 2011 y el máximo entre la TNR Casen 2017 y la TNR Cambio Marco	19.661
		Regional	1,4% - 3,5%	21,0% - 51,0%	1,0 - 3,0				
1.1	3.290.782	Nacional	0,6%	8,5%	2,1	15.921	15.922	TNR EANNA 2012 corregida por la variación entre la TNR de Casen 2011 y el máximo entre la TNR Casen 2017 y la TNR Cambio Marco	19.625
		Regional	1,4% - 3,5%	21,0% - 51,0%	1,0 - 3,0				
2 - 2.1 - 2.2	3.287.283	Nacional	0,6%	8,5%	2,1	18.066	18.069	Máximo entre la TNL de Casen 2022 ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y 2022 ponderando un 60% EANNA y un 40% Casen.	23.489
		Regional	1,4% - 3,4%	21,0% - 51,1%	1,0 - 3,0				

Elaborado por el Instituto Nacional de Estadísticas

⁷ Los errores absolutos y relativos son calculados con el tamaño objetivo ajustado.

III.5. Escenario escogido

Tabla 2: Estimación de tamaños muestrales según proyecciones de población al 15 de junio de 2023 con marco Casen 2022 actualizado⁸

Región	NNA Casen 2022	Proyección NNA 15/06/23	Tasa TI	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
Nivel País	33.989	3.287.283	6,6%	2,1	18.188	0,6%	8,5%	18.066	18.069	0,6%	8,5%	23.489
Urbano	27.622	2.926.900	6,1%	2,1	16.187	0,5%	8,2%	16.079	15.743	0,5%	8,3%	20.430
Rural	6.367	360.383	9,7%	2,1	2.001	1,9%	19,4%	1.987	2.326	1,7%	18,0%	3.059
Arica y Parinacota	1.490	46.845	4,9%	1,8	652	2,3%	46,1%	644	644	2,3%	46,3%	839
Tarapacá	1.795	76.333	4,5%	1,7	717	2,0%	44,7%	711	711	2,0%	44,9%	887
Antofagasta	1.667	125.710	5,5%	1,7	774	2,1%	37,2%	770	770	2,1%	37,3%	1.033
Atacama	1.701	61.560	5,2%	1,7	672	2,2%	42,3%	665	665	2,2%	42,5%	912
Coquimbo	1.420	154.893	5,3%	1,9	908	2,1%	38,9%	903	903	2,1%	39,0%	1.310
Valparaíso	3.303	323.958	5,1%	2,1	2.098	1,4%	27,5%	2.085	2.085	1,4%	27,6%	2.589
Metropolitana	6.336	1.306.947	6,7%	3,0	3.798	1,4%	21,0%	3.787	3.787	1,4%	21,0%	5.027
O'Higgins	2.213	173.978	8,7%	1,8	999	2,4%	27,6%	994	995	2,4%	27,6%	1.196
Maule	2.186	197.757	4,8%	1,7	1.221	1,5%	31,5%	1.214	1.215	1,5%	31,6%	1.701
Ñuble	1.312	86.134	10,3%	2,4	775	3,4%	33,1%	769	769	3,4%	33,2%	927
Biobío	3.234	282.035	4,4%	2,6	1.916	1,5%	34,3%	1.904	1.905	1,5%	34,4%	2.302
La Araucanía	2.244	179.839	11,6%	2,1	1.029	2,9%	25,1%	1.024	1.024	2,9%	25,2%	1.464
Los Ríos	1.852	68.634	11,9%	1,5	693	3,0%	25,1%	687	687	3,0%	25,2%	867
Los Lagos	1.716	153.757	8,0%	2,6	887	2,9%	36,3%	882	882	2,9%	36,4%	1.148
Aysén	732	20.263	7,7%	1,5	504	2,9%	37,6%	492	492	2,9%	38,0%	613
Magallanes	788	28.640	3,0%	1,5	545	1,5%	50,6%	535	535	1,5%	51,1%	674

Elaborado por el Instituto Nacional de Estadísticas.

⁸ Marco actualizado con fecha 11 de abril de 2023 y puesto a disposición del INE el día 20 de abril de 2023.

Este escenario se elabora con base en los parámetros obtenidos del trabajo de campo de EANNA 2012 (escenario n°0). La diferencia con los escenarios anteriores se encuentra en que el marco de referencia corresponde a 33.989 NNA que fueron resultado del levantamiento de Casen 2022. El error absoluto propuesto es ajustado respecto al escenario n°1 para cumplir con el estándar de calidad y el orden de la distribución regional de la población. Con errores absolutos entre 1,4% y 3,4% a nivel regional se obtiene un tamaño bajo m.a.s. de 8.072 NNA a nivel nacional.

Para obtener el efecto de diseño se contempla el efecto de diseño de EANNA 2012, el efecto de diseño utilizado para Casen 2022 y el efecto de diseño resultante del cálculo de la tasa de pobreza en hogares con NNA 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 Coquimbo, Metropolitana, Ñuble, Los Ríos, Aysén y Magallanes. A nivel regional se establece que los efectos de diseño utilizados fluctúen entre 1,5 y 3,0.

Luego, el tamaño objetivo ajustado por efecto de diseño es de 18.188 NNA que, al aplicar el ajuste por finitud, considerando como universo las proyecciones de población al 15 de junio de 2023, resulta un tamaño de 18.066 NNA a nivel nacional.

La distribución de los tamaños por estrato de muestreo (región-área-tramo etario) se realiza en forma proporcional a las proyecciones de población a junio de 2023 y se establece un mínimo de 10 NNA por cada uno, aumentando la muestra a 18.069 NNA con un error absoluto es de 0,6% y un error relativo de 8,5%. Si bien a nivel regional se observan errores relativos elevados, a partir del error estándar se aprecia que estos valores cumplen con el estándar de calidad.

Finalmente, para obtener el tamaño con sobremuestreo, se utiliza una tasa de no logro a nivel regional donde se selecciona el máximo entre: la tasa de no logro de Casen 2022, ponderada por la proporción de la TNL de EANNA 2012 respecto de Casen 2011 y el máximo de la TNL entre Casen 2017 y Casen 2022, que se pondera con EANNA 2012 que equivale a 60% y Casen a 40%. Este procedimiento dio lugar a un tamaño con sobremuestreo de 23.489 NNA.

III.6. Tamaño muestral propuesto

El tamaño calculado a nivel regional se debe distribuir por área (urbana – rural) y, al interior de estas, por los tramos etarios definidos según una distribución conocida. Para esto, se cuenta con la distribución resultante de Casen 2022 o bien, con las proyecciones de población estimadas al 15 de junio de 2023 (fecha que coincide con el periodo de levantamiento de la encuesta).

El tamaño objetivo ajustado del escenario elegido (n°2.2) definido para obtener estimaciones de la tasa de trabajo infantil con suficiente grado de precisión a nivel regional, alcanza 18.069 NNA a nivel

nacional, cuyos errores absolutos y relativos asociados son de 0,6% y 8,5%, 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 23.489 NNA.

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

Tabla 3: Tamaños y errores finales a nivel nacional, regional, urbano-rural, EANNA 2023

Región	Error Absoluto	Error Relativo	Tamaño objetivo			Tamaño con sobremuestreo		
			Total	Urbano	Rural	Total	Urbano	Rural
Nivel País	0,6%	8,5%	18.069	15.743	2.326	23.489	20.430	3.059
Arica y Parinacota	2,3%	46,3%	644	595	49	839	775	64
Tarapacá	2,0%	44,9%	711	671	40	887	839	48
Antofagasta	2,1%	37,3%	770	730	40	1.033	981	52
Atacama	2,2%	42,5%	665	617	48	912	847	65
Coquimbo	2,1%	39,0%	903	760	143	1.310	1.077	233
Valparaíso	1,4%	27,6%	2.085	1.906	179	2.589	2.367	222
Metropolitana	1,4%	21,0%	3.787	3.627	160	5.027	4.815	212
O'Higgins	2,4%	27,6%	995	779	216	1.196	936	260
Maule	1,5%	31,6%	1.215	940	275	1.701	1.317	384
Ñuble	3,4%	33,2%	769	567	202	927	683	244
Biobío	1,5%	34,4%	1.905	1.703	202	2.302	2.058	244
La Araucanía	2,9%	25,2%	1.024	743	281	1.464	1.062	402
Los Ríos	3,0%	25,2%	687	504	183	867	635	232
Los Lagos	2,9%	36,4%	882	685	197	1.148	891	257
Aysén	2,9%	38,0%	492	425	67	613	529	84
Magallanes	1,5%	51,1%	535	491	44	674	618	56

Elaborado por el Instituto Nacional de Estadísticas

Luego, a partir de las desagregaciones de tamaños por estrato geográfico (comuna-área) y de los resultados del trabajo de campo de EANNA 2012 se obtienen, por agregación, tamaños y errores asociados para los niveles Nacional Urbano y Nacional Rural que se muestra en la Tabla 4.

Tabla 4: Tamaños y errores EANNA 2012 y propuesta EANNA 2023. Nacional Urbano - Rural

Región	Resultados EANNA 2012				Propuesta EANNA 2023			
	NNA logrados	Tasa TI	Error Absoluto	Error Relativo	Tamaño Objetivo	Error Absoluto	Error Relativo	Tamaño con Sobremuestreo
Nivel País	9.978	6,6%	0,8%	11,8%	18.069	0,6%	8,5%	23.489
Urbano	7.859	6,1%	0,8%	13,1%	15.743	0,5%	8,2%	20.430
Rural	2.119	9,7%	1,6%	16,5%	2.326	1,9%	19,4%	3.059

Elaborado por el Instituto Nacional de Estadísticas.

A nivel nacional urbano, con un tamaño objetivo ajustado de 15.743 NNA se obtiene una estimación para la tasa de TI con un error absoluto de 0,5% y un error relativo de 8,2%. Por otra parte, a nivel nacional rural, con un tamaño objetivo de 2.326 NNA se obtiene una estimación para la tasa de TI con un error absoluto de 1,9% y un error relativo de 19,4%.

III.7. Errores esperados

La Tabla 5 presenta los errores esperados, asociados a la estimación del parámetro de interés para un tamaño muestral objetivo de 18.069 NNA⁹.

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

Región	Tasa TI EANNA 2012	Error Absoluto Propuesto EANNA 2023	Error Relativo Propuesto EANNA 2023	Tamaño ajustado por finitud EANNA 2023	Tamaño objetivo ajustado EANNA 2023	Tamaño con sobremuestreo EANNA 2023
Nacional	6,6%	0,6%	8,5%	18.066	18.069	23.489
Arica y Parinacota	4,9%	2,3	46,1	644	644	839
Tarapacá	4,5%	2,0	44,7	711	711	887
Antofagasta	5,5%	2,1	37,2	770	770	1.033
Atacama	5,2%	2,2	42,3	665	665	912
Coquimbo	5,3%	2,1	38,9	903	903	1.310
Valparaíso	5,1%	1,4	27,5	2.085	2.085	2.589
Metropolitana	6,7%	1,4	21,0	3.787	3.787	5.027
O'Higgins	8,7%	2,4	27,6	994	995	1.196
Maule	4,8%	1,5	31,5	1.214	1.215	1.701
Ñuble	10,3%	3,4	33,1	769	769	927
Biobío	4,4%	1,5	34,3	1.904	1.905	2.302
La Araucanía	11,6%	2,9	25,1	1.024	1.024	1.464
Los Ríos	11,9%	3,0	25,1	687	687	867
Los Lagos	8,0%	2,9	36,3	882	882	1.148
Aysén	7,7%	2,9	37,6	492	492	613
Magallanes	3,0%	1,5	50,6	535	535	674

Elaborado por el Instituto Nacional de Estadísticas.

⁹ 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 18.069 NNA.

IV. SELECCIÓN DE UNIDADES MUESTRALES

La selección de unidades muestrales se realiza desde el marco muestral actualizado al 11 de abril de 2023, conformado por NNA logrados en Casen 2022, en forma sistemática y con igual probabilidad al interior de cada estrato de muestreo (región - área - tramo etario) contemplando aquellos registros de NNA entre 5 y 17 años cumplidos a junio de 2023.

La selección de NNA 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_{em} el número de NNA encontrado en el estrato y m_{em}^{sel} el número de NNA a seleccionar. Para la selección de m_{em}^{sel} NNA, el *software* ejecuta los siguientes pasos:

Paso 1

En primera instancia se ordenan geográficamente y en forma ascendente todos los NNA según estrato de muestreo, comuna, sexo e identificador de la UPM¹⁰.

Paso 2

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

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

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 ", " $A + k$ ", " $A + 2k$ ", " $A + 3k$ ", ..., " $A + (m_{em}^{sel} - 1)k$ ".

El primer NNA seleccionado 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_{em}^{sel} selección, dada por el redondeo de " $A + (m_{em}^{sel} - 1)k$ ".

¹⁰ En la sintaxis enviada a MDSF la variable correlativo replica este ordenamiento.

Luego, la probabilidad de inclusión del j – ésimo NNA dentro del estrato em está dada por:

$$P_{re}(j) = \frac{m_{em}^{sel}}{M_{em}} \quad (1)$$

Donde:

m_{em}^{sel} : Número de NNA seleccionados en el estrato em

M_{em} : Número actualizado de NNA en el marco de selección en el estrato em .

La Tabla 6 presenta la distribución del total de NNA seleccionados, según región y área (Urbano - Rural) obtenida a partir de la distribución del marco de NNA logrados en Casen 2022.

Tabla 6: Distribución del número de NNA por región y área

Región	Área	Total NNA	
		Marco Casen 2022	Selección EANNA 2023
Total País		33.989	23.489
Arica y Parinacota	Urbano	1.294	775
	Rural	196	64
Tarapacá	Urbano	1.659	839
	Rural	136	48
Antofagasta	Urbano	1.519	981
	Rural	148	52
Atacama	Urbano	1.531	847
	Rural	170	65
Coquimbo	Urbano	1.104	1.077
	Rural	316	233
Valparaíso	Urbano	2.696	2.367
	Rural	607	222
Metropolitana	Urbano	5.812	4.815
	Rural	524	212
O'Higgins	Urbano	1.626	936
	Rural	587	260
Maule	Urbano	1.605	1.317
	Rural	581	384
Ñuble	Urbano	874	683
	Rural	438	244
Biobío	Urbano	2.584	2.058
	Rural	650	244
La Araucanía	Urbano	1.431	1.062
	Rural	813	402
Los Ríos	Urbano	1.308	635
	Rural	544	232
Los Lagos	Urbano	1.268	891
	Rural	448	257
Aysén	Urbano	606	529
	Rural	126	84
Magallanes	Urbano	705	618
	Rural	83	56

Elaborado por el Instituto Nacional de Estadísticas.

V. REFERENCIAS

Cochran, W. (1998). *Técnicas de Muestreo*. México D.F.: Compañía Editorial Continental, S.A. de C.V.

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*. Obtenido de <https://www.ine.gob.cl/inicio/documentos-de-trabajo/documento/fundamentos-del-estandar-para-la-evaluacion-de-la-calidad-de-las-estimaciones-en-encuestas-de-hogares>

Kish, L. (1965). *Survey Sampling*. Nueva York: John Wiley & Sons.

MDSF. (2022). *Minuta de EANNA 2023*.