



Observatorio Social



Informe Construcción Factores de Expansión

Encuesta Longitudinal de Primera Infancia Cuarta Ronda 2024

Noviembre 2025



**DIVISIÓN OBSERVATORIO SOCIAL
MINISTERIO DE DESARROLLO SOCIAL Y FAMILIA**

**INSTITUTO NACIONAL DE ESTADÍSTICAS
DEPARTAMENTO DE METODOLOGÍAS E INNOVACIÓN ESTADÍSTICA
SUBDEPARTAMENTO DE DISEÑO DE MARCOS Y MUESTRAS**

Encuesta longitudinal de Primera Infancia, ELPI 2024

Informe editado por el Ministerio de Desarrollo Social y Familia, a partir del “Informe metodológico sobre el diseño muestral. Incluye metodología de construcción de factores de expansión ELPI 2024” entregado por el Instituto Nacional de Estadísticas (INE), en el marco del convenio “Programa de trabajo conjunto para el desarrollo de diseños estadísticos de encuestas del Ministerio de Desarrollo Social y Familia”.

Versión revisada en noviembre de 2025.

ÍNDICE

ÍNDICE DE TABLAS	5
I. Introducción	6
II. Antecedentes de diseño muestral ELPI	9
II.1. Objetivo	9
II.2. Población objetivo.....	9
II.3. Unidad de información.....	9
II.4. Marco muestral	9
II.5. Dominios de estudio.....	10
II.6. Método de Selección de Unidades Primarias de Muestreo (comunas y estratos de comunas)10	
II.7. Método de Selección de Unidades Secundarias de Muestreo (niños y niñas)	10
III. Construcción de factores de expansión transversal.....	11
III.1. Ponderador de Selección/Diseño	11
III.2. Ajuste por no contacto	11
III.3. Ajuste de no respuesta.....	13
III.4. Suavizamiento	14
III.5. Ponderador de calibración	17
IV. Construcción del factor de expansión longitudinal del Panel	19
IV.1. Antecedentes de la propuesta del diagnóstico	19
IV.2. Ponderador de Selección/Diseño del año 2010	19
IV.3. Ajuste de no respuesta.....	21
IV.4. Suavizamiento	24
IV.5. Ponderador de calibración	24
V. Construcción de factores de expansión longitudinal del año 2010-2024	25
V.1. Antecedentes de la propuesta del diagnóstico	25
V.2. Ponderador de Selección/Diseño del año 2010	26
V.3. Ajuste de no respuesta.....	26
V.4. Suavizamiento	29
V.5. Ponderador de calibración	29
VI. Cálculo y evaluación de las estimaciones.....	30
VI.1. Estadísticos asociados a los parámetros de interés.....	31
VII. ANEXOS.....	35
VII.1. ANEXO N°1: Tablas marginales utilizadas para la calibración por Raking del factor de expansión transversal ELPI 2024	35
VII.2. ANEXO N°2: Tablas marginales utilizadas para la calibración por Raking del factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024.....	38
VIII. Bibliografía.....	40

ÍNDICE DE FIGURAS

Figura 1. Aplicación de método contracción a la media CM, en ambos conjuntos de factores; no atípicos y atípicos	16
Figura 2. Aplicación de método R-K en ambos conjuntos de factores atípicos y no atípicos suavizados por método CM atípicos	17

ÍNDICE DE TABLAS

Tabla 1: Característica de los factores de expansión de ELPI 2024.....	8
Tabla 2: Resumen de Métricas correspondiente a cada escenario.....	14
Tabla 3: Marco muestral por Región/Cohorte para cálculo de factor de expansión transversal	18
Tabla 4: Marco muestral por Región/Cohorte para cálculo de factor de expansión longitudinal del Panel	25
Tabla 5: Marco Muestral por Región/Cohorte para cálculo de factor de expansión longitudinal del Panel	30
Tabla 6. Estimación del Puntaje Estándar Hispanoamericano TVIP con sus respectivos errores muestrales a nivel nacional y regional, según factor de expansión transversal ELPI 2024	32
Tabla 7. Estimación del Puntaje Estándar Hispanoamericano TVIP con sus respectivos errores muestrales a nivel nacional y regional, según factor de expansión longitudinal panel, ELPI 2024.....	33
Tabla 8. Estimación del Puntaje Estándar Hispanoamericano TVIP con sus respectivos errores muestrales a nivel nacional y regional, según factor de expansión longitudinal 2010-2024.....	34
Tabla 8: Marco Muestral para cálculo de factor de expansión transversal	35
Tabla 9: Marco Muestral por Región/Sexo para cálculo de factor de expansión transversal	36
Tabla 10: Marco Muestral por Cohorte/Sexo para cálculo de factor de expansión transversal	37
Tabla 11: Marco Muestral para cálculo de factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024	37
Tabla 12: Marco Muestral por Región/Sexo para cálculo de factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024	38
Tabla 13: Marco Muestral por Cohorte/Sexo para cálculo de factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024	38
Tabla 14: Marco Muestral para cálculo de factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024	39

I. Introducción

La Encuesta longitudinal de Primera Infancia (en adelante, ELPI) es una encuesta realizada a niños, niñas y adolescentes (NNA), cuyo objetivo es hacer seguimiento periódico, a partir de una muestra, a lo largo de su nacimiento, infancia y adolescencia, con el fin de conocer sus trayectorias de vida, las características de su hogar y su entorno cercano, entre otras, contribuyendo así con la política pública, actual y futura, en la toma de decisiones con respecto a la infancia¹.

ELPI nace como una necesidad de obtener información del desarrollo infantil en Chile por ello, es una encuesta de tipo longitudinal. La primera ronda que se realizó de esta encuesta fue el año 2010, a cargo del Ministerio de Educación, seguida de la segunda ronda realizada el año 2012, esta vez a cargo del Ministerio del Trabajo. Finalmente, desde el año 2017, el desafío de realizar la tercera ronda queda a cargo del Ministerio de Desarrollo Social y Familia (MDSF) que, hasta el día de hoy, sigue cumpliendo estas funciones.

La encuesta, en sus diferentes versiones, se realiza en el hogar del menor seleccionado y es aplicada por un equipo de encuestadores y evaluadores especialmente capacitados. Sin embargo, en la última ronda, esta fue aplicada únicamente por encuestadores. El instrumento consta, en sus diversas rondas, de un cuestionario administrado al Cuidador Principal del NNA seleccionado/a, junto con evaluaciones de desarrollo aplicadas directamente al NNA. Además, el año 2017, se implementó un cuestionario autoaplicado dirigido al Segundo Cuidador Principal. Finalmente, se incorpora en los años 2017 y 2024, cuestionarios dirigidos a los propios NNA.

Actualmente, la ELPI es una encuesta longitudinal que se enfoca en hacer seguimiento a la muestra original del año 2010 correspondiente a 15.175 niños y niñas donde, para las siguientes rondas, se agregan muestras de refresco para capturar nuevas cohortes de niños y niñas.

Para dar continuidad al proyecto, durante el periodo 2021 - 2023 se levantaron diversos procesos de licitación pública, sin embargo, quedaron desiertos por falta de oferentes o por ofertas que excedían el presupuesto. Por lo que, para el año 2024 se decide realizar una cuarta ronda haciendo seguimiento a la muestra del año 2010, para mantener así una continuidad respecto de los procesos anteriores.

El presente documento es parte de los productos entregables del convenio INE-MDSF (Entrega N°4.1) y aborda los siguientes puntos:

1. Antecedentes y características generales del diseño muestral según la documentación compartida por MDSF, donde se revisa el objetivo de la encuesta, la población objetivo, la unidad de

¹ Encuesta Longitudinal de Primera Infancia: <https://observatorio.ministeriodesarrollosocial.gob.cl/elpi-primera-ronda>

información, los dominios de estudio, el marco muestral utilizado para la selección de unidades y los métodos de selección de estas.

2. Construcción del factor de expansión transversal 2024, en el cual se explica las mejoras metodológicas implementadas para el cálculo del factor de expansión transversal y aplicadas en la cuarta ronda. Estas mejoras se refieren principalmente, al cálculo del ponderador de selección/diseño, al ajuste por no contacto, ajuste por no respuesta, suavizamiento y a la construcción del ponderador de calibración.
3. Construcción del factor de expansión longitudinal del Panel, construido considerando únicamente a los niños, niñas y adolescentes que participaron efectivamente en las cuatro rondas de la ELPI. Su cálculo comienza con el factor transversal del año 2010, para luego, aplicar ajustes metodológicos que incluyen una corrección por elegibilidad, un ajuste por no contacto y un ajuste por no respuesta, condicional al contacto, mediante modelos de propensión de respuesta (Propensity Score), suavizamiento de valores extremos y calibración por Raking.
4. Construcción del factor de expansión longitudinal correspondiente a 2010 - 2024, que considera sólo aquellos casos que fueron contactados tanto en la primera como en la cuarta ronda de la ELPI, sin exigir continuidad en las mediciones intermedias. Al igual que en el caso del Panel, su cálculo comienza con el factor de expansión transversal de 2010, para luego, aplicar mejoras metodológicas que incluyen: ajuste por elegibilidad, un ajuste por no contacto y un ajuste por no respuesta, condicional al contacto, a partir de modelos logísticos, suavizamiento de valores atípicos y calibración.
5. Metodología de cálculo y evaluación de las estimaciones, donde se describe el detalle de las estimaciones obtenidas utilizando factores de expansión ad-hoc a cada estimación, según solicitud del MDSF.

En síntesis, las principales características de los factores de expansión detallados en este documento se describen en la Tabla 1:

Tabla 1: Característica de los factores de expansión de ELPI 2024.

	Factor de Expansión Transversal 2024	Factor de Expansión Longitudinal Panel	Factor de Expansión Longitudinal 2010-2024
Descripción	Este factor se utiliza para estimar características de la población para el año 2024	Este factor se utiliza para estimar características de la población, a partir de una muestra de unidades que respondió las 4 rondas de ELPI.	Este factor se utiliza para estimar características de la población, a partir de una muestra de unidades que respondió en ELPI 2010 y 2024.
Población objetivo	Niños y niñas nacidos entre el 01 de enero del 2006 y el 31 de agosto del 2009		
Población total	876.944 ²	886.733 ³	886.733
Muestra total	10.003	7.012	10.003
Ajustes aplicados al ponderador de selección	Ajuste por no contacto con metodología Propensity Score, ajuste por no respuesta con ajuste simple por estrato, suavizamiento y calibración con metodología Raking.	Ajuste por no contacto y un ajuste por no respuesta condicional al contacto con metodología Propensity Score, suavizamiento y calibración con metodología de Raking.	Ajuste por no contacto y un ajuste por no respuesta condicional al contacto con metodología Propensity Score, suavizamiento y calibración con metodología de Raking.

² La población total del factor de expansión transversal presenta un valor menor, ya que este total debe representar a los niños y niñas vivos al hacer el cálculo del factor de expansión transversal ELPI 2024.

³ El total utilizado para la cuarta ronda presenta diferencias a las versiones anteriores de la ELPI, debido a desactualizaciones del Marco Muestral que son incorporadas en esta versión.

II. Antecedentes de diseño muestral ELPI

II.1. Objetivo

La ELPI tiene por objetivo generar información para caracterizar y analizar el desarrollo de niños, niñas y adolescentes a lo largo de las primeras etapas de su vida, contribuyendo con información valiosa respecto de sus trayectorias de vida, las características de su hogar y su entorno cercano, para formular políticas públicas actuales y futuras, en relación con la primera infancia.

II.2. Población objetivo

La población objetivo son niños y niñas residentes en todo Chile, tanto en zonas urbanas como rurales. Durante la primera ronda del año 2010, se consideran todos los niños y niñas que nacieron entre el 1° de enero de 2006 y el 31 de agosto de 2009, inscritos en el Servicio de Registro Civil e Identificación de Chile (SRCel). Para la segunda ronda, se agregó una muestra complementaria correspondiente a niños y niñas nacidos entre el 1° de septiembre de 2009 y el 31 de diciembre de 2011. Y, en el año 2017, para la tercera ronda, se agrega una nueva muestra complementaria constituida por niños y niñas que nacieron entre el 1° de enero de 2012 y el 31 de diciembre de 2016. Finalmente, para la cuarta ronda realizada el año 2024, debido a los problemas de licitación del proyecto, solo se consideraron a los niños y niñas nacidos entre el 1° de enero de 2006 y el 31 de agosto de 2009.

II.3. Unidad de información

La ELPI recoge información directamente de los niños y niñas seleccionados, del Cuidador Principal y únicamente en la tercera ronda del año 2017, al Segundo Cuidador Principal, en caso de que exista. Además, si el niño o la niña tiene más de siete años debe contestar un módulo de evaluación dentro del instrumento.

II.4. Marco muestral

En la primera ronda del año 2010, el Marco Muestral (MM) corresponde a los registros de niños y niñas, nacidos entre el 1° de enero de 2006 y el 31 de agosto de 2009, obtenidos desde el SRCel de Chile. Para la segunda ronda, se consideran los registros de niños y niñas nacidos entre el 1° de septiembre de 2009 y el 31 de diciembre de 2011. En el año 2017, para la tercera ronda, el MM corresponde a los registros de niños y niñas nacidos entre 1° de enero de 2012 y el 31 de diciembre de 2016. Para la cuarta ronda, solo se

consideran los registros de niños y niñas nacidos entre el 1° de enero de 2006 y el 31 de agosto de 2009. En todas las rondas, se excluyen, de los registros, a la población extranjera o que habita en localidades que no son comunas o bien, en comunas de difícil acceso.

II.5. Dominios de estudio

La ELPI se encuentra diseñada para obtener indicadores confiables con errores de muestreo aceptables⁴ para los niveles nacional, regional y según tramos de edad de los niños y niñas dependiendo de la ronda.

II.6. Método de Selección de Unidades Primarias de Muestreo (comunas y estratos de comunas)

El diseño muestral de la ELPI es un diseño probabilístico bietápico. La primera etapa corresponde a la selección de comunas donde, con probabilidad igual a 1 se incluyen 83 comunas que corresponden a todo el Gran Santiago más San Bernardo, Puente Alto, Coyhaique y, todas aquellas que tengan más de 60.000 habitantes, según el Censo 2002. Las restantes 263 comunas se agrupan en 33 estratos muestrales. Para ello, se ordenan según región, ingreso per cápita y población de niños y niñas según la información de Casen 2006 formando conjuntos de 100.000 habitantes aproximadamente en cada estrato. En cada uno de estos grupos se calcula la probabilidad proporcional al tamaño (PPT) para cada comuna, luego según esta probabilidad y el método de imagen inversa, se selecciona 1 comuna que represente al conjunto completo. Según lo anterior, la muestra contiene 116 comunas: 83 de inclusión automática y 33 seleccionadas desde los grupos de comunas creados.

II.7. Método de Selección de Unidades Secundarias de Muestreo (niños y niñas)

En la segunda etapa de selección, considerando la selección en la etapa anterior y la información recolectada en la encuesta Casen del año correspondiente, además de los datos proporcionados por el SRCel, dentro de cada comuna se ordenan los registros de los niños y niñas de acuerdo con su sexo y edad, luego se selecciona mediante un muestreo sistemático, a partir del salto calculado como $k = \frac{N_{ch}}{n_{ch}}$, donde N_{ch} es la población de niños y niñas en la comuna seleccionada y n_{ch} corresponde al número de niños y niñas a seleccionar en la muestra de la comuna seleccionada.

⁴ Errores de muestreo aceptables, según definición de tamaños y errores observadas en las Tablas 1, 2 y 3 del Capítulo 3 del Reporte Metodológico en el año 2017, a nivel nacional año-mes de nacimiento del NNA y regional para ELPI 2017.

III. Construcción de factores de expansión transversal

Durante el año 2025, en un trabajo conjunto entre el INE y el MDSF, se llegó al acuerdo de realizar una actualización en la metodología utilizada para el cálculo de los factores de expansión transversal de la ELPI, para ello se plantearon diferentes ajustes que podrían mejorar el factor de expansión transversal para su cuarta ronda del año 2024.

A continuación, se presenta el detalle de las mejoras implementadas sobre la construcción de los factores de expansión transversal para ELPI 2024 que fue consensado entre el INE y el Ministerio durante el año 2025.

III.1. Ponderador de Selección/Diseño

Para la recolección de la cuarta ronda de la ELPI, se decidió replicar la muestra lograda en el año 2010, correspondiente a 15.175 niños y niñas. Esto quiere decir, que no se incorporó una muestra de refresco en esta ronda, a diferencia de las rondas anteriores, por lo que, todos los niños y niñas que respondieron la encuesta en la primera ronda, se consideran elegibles para la encuesta. Esta decisión se realizó por los diferentes problemas de la licitación del proyecto para la cuarta ronda 2024.

Dado que no se realizó un nuevo proceso de selección muestral, no fue necesario calcular nuevos ponderadores de selección. En su lugar, se utilizó el ponderador de selección aplicado en la ronda de 2010 para cada uno de los casos originales que permanecen en la muestra. Este ponderador refleja las probabilidades iniciales de inclusión de cada unidad muestral y, en esta cuarta ronda, mantiene su vigencia al no haberse alterado la estructura de la muestra. En otras palabras, el ponderador de selección utilizado en el año 2024 corresponde al definido para la primera ronda de la ELPI, para asegurar la coherencia metodológica del proceso.

III.2. Ajuste por no contacto

El cálculo del factor de expansión considera un ajuste por contactabilidad, es decir, en lugar de asumir atricción de la muestra debido a cambios de estado en ciertas unidades, se opta por realizar un ajuste por no contacto del niño o la niña, con el cual se puede identificar en qué proporción existió o no, comunicación en esta nueva ronda 2024.⁵

⁵ Al tratarse de un seguimiento de unidades que ya se determinaron como elegibles previamente, se definió considerar que, para rondas posteriores a la ronda inicial, todas las unidades clasificadas como elegibles, mantienen su estado en las siguientes rondas. Por este motivo no corresponde ajustar por elegibilidad como suele realizarse en encuestas de corte transversal.

A partir de esta propuesta, se realizaron pruebas para determinar la mejor metodología de ajuste por no contacto, decidiendo que la aplicación de un modelo de no contacto resultaría adecuado, logrando agrupar las unidades de la muestra con similares características, asumiendo que sus probabilidades de contacto son también similares. Para ello se aplica un modelo logit utilizando variables que podrían estar relacionadas con el fenómeno del no contacto como predictores de la respuesta.

Para la generación del modelo se probaron varias combinaciones de variables agrupadas en diferentes dimensiones. Estas incluían:

- **Variables relacionadas con el adolescente seleccionado/a**, como el sexo y cohorte de nacimiento, que clasifica a los niños y niñas según su año de nacimiento.
- **Variables del responsable principal**, tales como el nivel educativo, edad y estado civil.
- **Variables del encuestador**, como su edad, nivel educativo y experiencia en la aplicación de encuestas.
- **Variables generales**, entre las que se encuentran la región geográfica y la condición de área urbana o rural.

Tras múltiples pruebas, y considerando tanto la significancia estadística de las variables asociadas a la probabilidad de contacto con el niño o niña, como indicadores de desempeño del modelo (precisión, sensibilidad y especificidad), se determinó que el mejor modelo queda conformado por:

- **Región Metropolitana**: una variable dicotómica que indica si una observación pertenece (1) o no pertenece (0) a la Región Metropolitana.
- **Quintil de Ingreso Familiar**: Se creó una variable que categoriza por quintiles el ingreso per cápita familiar de cada observación, dada la información de ELPI 2010.
- **Sexo**: una variable dicotómica que indica si el responsable principal es Hombre (1) o si es Mujer (2).
- **Tramo Educativo**: Se refiere al nivel educacional del responsable principal, puede tomar el valor de 13 niveles educacionales a considerar para el año 2024, complementada con información de rondas anteriores. Estos son: Sin educación formal, Básica incompleta, Básica completa, Media humanista incompleta, Media técnica profesional incompleta, Media humanista completa, Media técnica profesional completa, Técnico nivel superior incompleta, Técnico nivel superior completo, Profesional incompleto, Posgrado incompleto, Profesional completo, Posgrado completo y no especificado.

- Situación ocupacional: Variable categórica que describe la situación de ocupación del responsable principal. Las categorías son: ocupado, desocupado e Inactivo.
- Estado Civil: Se genera una variable según el estado civil del responsable principal. Las categorías son: separado (que incluye si es separado, divorciado o anulado), casado(a), soltero, pareja (que incluye conviviente o pareja sin acuerdo de unión civil), AUC (que corresponde a los convivientes con acuerdo de unión civil) y otros (que incluye si es viudo o no se señala).
- Tramo Etario del Encuestador, definiendo las siguientes categorías: joven, entre 18 y 30 años; joven adulto, entre 31 y 40 años; adulto, entre 41 y 59 años; adulto mayor, entre 60 y 75 años y; no especificado, si la edad del encuestador no está disponible.

Luego de ajustar el modelo logit, se procede a ordenar las probabilidades resultantes de menor a mayor, dividiéndolas en diez grupos o deciles. Esta estratificación permite agrupar las observaciones según niveles similares de propensión a ser contactados, lo que facilita aplicar un ajuste por no contacto dentro de cada uno de estos grupos. Con este ajuste se busca corregir el sesgo que podría introducirse si ciertas características influyen de manera sistemática en la probabilidad de ser contactado.

III.3. Ajuste de no respuesta

El INE inicialmente propone la aplicación de un ajuste por no respuesta, basado en un modelo de propensión de respuesta (Propensity Score), como alternativa al enfoque utilizado en las últimas tres rondas, basado en un ajuste simple por estrato. Lo anterior, debido a que el ajuste simple no permite capturar adecuadamente la heterogeneidad en las probabilidades de respuesta entre las unidades dentro de un mismo estrato, y suele ser insuficiente cuando la no respuesta está asociada a variables individuales relevantes, ya que este asume probabilidades homogéneas de respuesta dentro de cada estrato. En esa línea, el modelo de Propensity Score permite realizar un ajuste más preciso al incorporar múltiples variables de manera simultánea que permiten explicar el comportamiento de la no respuesta dentro de deciles donde, las covariables (o variables explicativas del modelo) se comportan de manera similar en su interior, resultando ser una mejor alternativa.

Sin embargo, por condiciones de la encuesta, como no resultaba aplicable un ajuste previo por elegibilidad, y considerando que ya se había aplicado un modelo para explicar el no contacto utilizando una metodología basada también en Propensity Score, resultaba complejo ajustar dos modelos destinados a corregir fenómenos similares (la no respuesta y el no contacto), especialmente dado el número limitado de variables disponibles. No obstante, dado que el ajuste por no respuesta es sumamente relevante, en este tipo de encuestas, pues podría impactar significativamente en los valores del factor de expansión, se

optó finalmente por asumir la atrición de la muestra aplicando un ajuste simple por estrato para su tratamiento. Esta decisión se tomó luego de evaluar los siguientes escenarios:

- **Escenario 1:** Realizar un ajuste simple para la contactabilidad, y un modelo logit para tratar la no respuesta de aquellos que lograron ser contactados.
- **Escenario 2:** Realizar solo un ajuste por no respuesta a partir de un modelo logit.
- **Escenario 3:** Realizar un modelo logit para tratar el no contacto y un ajuste simple para la no respuesta.

Para determinar la mejor opción según los datos disponibles, se ajustó cada uno de ellos y se obtuvieron las correspondientes métricas con los siguientes resultados, presentados en la Tabla 2:

Tabla 2: Resumen de Métricas correspondiente a cada escenario

Métrica / Escenarios	Escenario 1	Escenario 2	Escenario 3
Exactitud (Accuracy)	65,89%	65,66%	64,78%
Precisión	76,44%	78,13%	82,72%
Sensibilidad	69,73%	66,53%	65,55%
Especificidad	58,44%	63,98%	62,69%

Al comparar los tres escenarios, el escenario 3 destaca como el más adecuado, a pesar de no contar con la mayor **especificidad** (62,69%), logra un equilibrio atractivo entre una **precisión alta** (82,72%) y una **sensibilidad** (65,55%) muy cercana a la de los otros escenarios. Este balance lo convierte en una opción sólida cuando se busca mantener un buen nivel de acierto en las categorías de la matriz de confusión, sin sacrificar significativamente la capacidad de detección de casos reales.

Según lo anterior, se concluye por aplicar un modelo de Propensity Score para ajustar el no contacto y un ajuste simple por estrato para abordar la no respuesta, justificándolo adicionalmente, en gran parte porque el no contacto resulta ser considerablemente más prevalente que el rechazo. De hecho, la encuesta presenta una alta tasa de cooperación (COOP1) de 90,1%, pero una menor tasa de contacto (CON1), que alcanza 73,2%.

III.4. Suavizamiento

En distintas encuestas realizadas por el INE, se ha observado en ocasiones la presencia de valores atípicos en los factores de expansión obtenidos. Estos, pueden distorsionar los resultados y afectar la calidad de los análisis estadísticos, motivo por el cual, se hace necesario evaluar la aplicación de algún tipo de tratamiento.

Desde el año 2019, el INE ha implementado un método de suavizamiento para tratar datos atípicos en la Encuesta longitudinal de Empresas (ELE₅) procedimiento que fue posteriormente adoptado por encuestas

de carácter social como: la Encuesta Social Covid-19, EANNA y EBS, y continúa utilizándose hasta la actualidad. En este contexto, se propuso aplicar dicha metodología en el cálculo del factor de expansión transversal de la ELPI.

La metodología de suavizamiento empleada por el INE consiste en la aplicación de un método mixto (MM), denominado así por combinar dos enfoques comúnmente utilizados: el método de contracción a la media (CM) y el método R-K.

El método de contracción de la media (CM) consiste básicamente en contraer o acercar aquellos factores inusualmente altos (o bajos) a la media de ellos, lo que resulta en una distribución más homogénea de los mismos. Por otro lado, el método R-K, se caracteriza por dar un tratamiento diferenciado a los datos del conjunto, aumentando proporcionalmente los factores no atípicos, mientras que disminuye proporcionalmente los factores atípicos. Esto se logra a través de un sistema de ecuaciones para distribuir los pesos, donde se asegura que el factor más grande de los valores no atípicos sea menor que el factor más pequeño de los valores atípicos suavizados.

Ahora bien, el método mixto (MM)⁶ consiste en separar los factores de expansión en dos subconjuntos: uno, conformado por factores no atípicos y otro, por factores atípicos. (Figura 1). Posteriormente se procede de la siguiente manera:

1. Para el conjunto Φ_1 , de factores no atípicos, se aplica el método de contracción a la media con un valor de beta relativamente cercano a la unidad, de modo que la distribución se mantenga casi intacta, es decir, que no varíe mucho respecto a la original, por ejemplo, aplicar el método CM con $\beta = 0.8$.
2. Para el conjunto Φ_2 de factores atípicos, se aplica el mismo método CM, pero con un valor de beta cercano a cero, de modo de potenciar la homogeneidad de ellos en este conjunto. Por ejemplo, al aplicar el método CM con $\beta = 0.2$. En el caso extremo que se encuentre tan solo un valor atípico, el método CM no tiene efecto sobre este valor, es decir, el factor suavizado coincide con el valor atípico.

Cabe resaltar que, en la aplicación diferenciada del método CM a ambos subconjuntos (factores atípicos y no-atípicos) se deben contraer levemente los factores no atípicos y fuertemente los atípicos, por lo que se recomienda usar intensidades de contracciones complementarias a la unidad, es decir, si para los no atípicos se aplica β , para los atípicos se aplique $(1 - \beta)$.

3. Habiendo aplicado en ambos conjuntos por separado el método CM, se aplica entonces el método R-K, juntando ambas distribuciones de factores (no atípicos con atípicos) en un continuo que se

⁶ Esta explicación fue tomada del documento de trabajo Análisis del efecto del COVID-19 sobre el diseño muestral y las estimaciones de la ENE en EFM 2020.

puede lograr con valores de λ cercanos a la unidad (Figura 2). Específicamente, se recomienda un valor de λ entre 0.95 a 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.

Figura 1. Aplicación de método contracción a la media CM, en ambos conjuntos de factores; no atípicos y atípicos

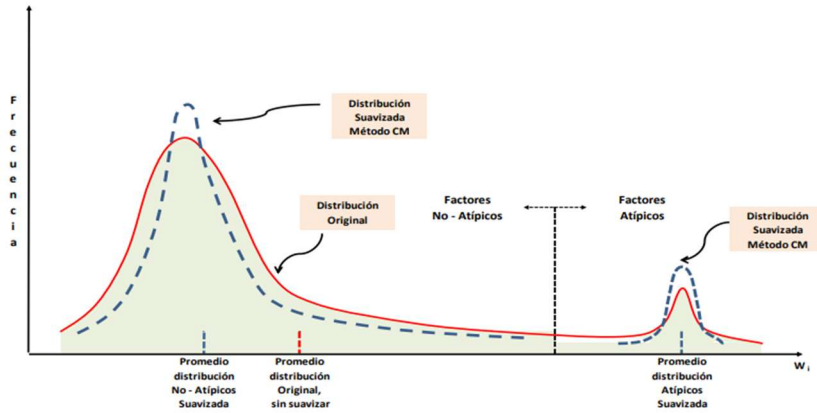
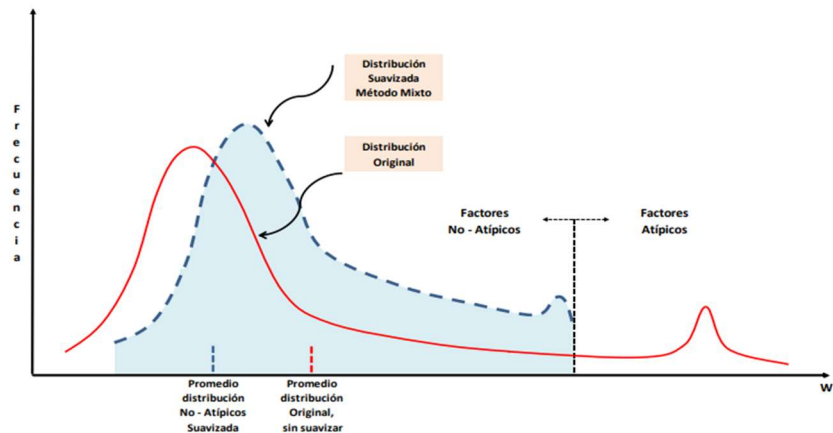


Figura 2. Aplicación de método R-K en ambos conjuntos de factores atípicos y no atípicos suavizados por método CM atípicos



En la propuesta presentada al MDSF, se indicó que la aplicación de la metodología de suavizamiento debía ser evaluada en función de la presencia de datos atípicos antes de la calibración. Según lo anterior, una vez finalizado el ajuste por no respuesta, en la cuarta ronda de ELPI, se identificaron 1.129 factores de expansión considerados atípicos, lo que justificó la necesidad de aplicar el procedimiento de suavizamiento. Como resultado de la aplicación, el rango de los factores de expansión se redujo, pasando de valores entre 44,08 y 147,42 a un rango más acotado de 46,14 a 129,6. Esta disminución en la dispersión de los factores de expansión refleja una mejora considerable en la estabilidad del ponderador, contribuyendo a una mejor precisión en las estimaciones derivadas.

III.5. Ponderador de calibración

El propósito de este ajuste, al igual que en el cálculo del factor de expansión transversal en las versiones anteriores de la ELPI, es obtener estimaciones coherentes con totales poblacionales conocidos, definidos según características geográficas, demográficas, entre otras. En términos prácticos, esto significa incorporar al factor información auxiliar específica, con el objetivo de generar estimaciones más precisas y representativas de la población en los diferentes dominios de estudio, concretamente: a nivel nacional, regional y según tramos de edad definidos de la ronda.

Históricamente, la ELPI ha aplicado un ajuste por post-estratificación en los factores de expansión a los totales poblacionales conocidos. Si bien este método resulta útil para reducir sesgos mediante el uso de información auxiliar, también introduce una mayor variabilidad en los ponderadores, lo que puede afectar la precisión de las estimaciones. Esto se debe a que, al incorporar variables con múltiples categorías, se aumenta la complejidad del ajuste y, con ello, la dispersión de los factores post-estratificados.

Con el fin de disminuir este efecto y mejorar la eficiencia del ajuste, se propone como alternativa metodológica la implementación del método de *Raking Generalizado*. Este método permite disminuir la

distancia entre los ponderadores originales y los calibrados, logrando estimaciones igualmente válidas, pero con menor variabilidad, lo que redonda en una mayor estabilidad de los resultados.

Para aplicar esta metodología de calibración, se requiere contar con los totales poblacionales de referencia correspondientes a los niños y niñas nacidos entre el 1 de enero de 2006 y el 31 de agosto de 2009 que se encuentren vivos al momento del cálculo del factor de expansión. Para ello, a partir de la información del registro civil e Identificación (SRCel) y del Ministerio de Salud (MINSAL), el Subdepartamento de Demografía del INE preparó tablas a nivel nacional y regional, por cohorte y sexo con total de niños y niñas nacidos. A estas tablas se les agregó la información sobre defunciones utilizando la tasa de mortalidad proporcionada por la División de Información Social (DIS) de la Subsecretaría de Evaluación Social (SES).

Con esta información se construyó el marco muestral de calibración, cuyo detalle desagregado por región y cohorte de nacimiento se presenta en la Tabla 3:

Tabla 3: Marco muestral por Región/Cohorte para cálculo de factor de expansión transversal

Región	Población de niños y niñas por Región y Cohorte			
	2006	2007	2008	2009 ⁷
Nacional⁸	228.713	237.912	243.910	166.409
Arica y Parinacota	3.063	3.123	3.328	2.375
Tarapacá	4.706	4.961	5.230	3.725
Antofagasta	8.871	9.315	9.601	6.584
Atacama	4.594	4.668	4.751	3.429
Coquimbo	9.264	9.686	10.334	7.083
Valparaíso	21.422	22.409	22.866	15.721
Metropolitana	92.513	96.518	98.791	67.840
O'Higgins	11.384	12.064	12.456	8.348
Maule	12.948	13.416	13.893	9.340
Ñuble	5.874	6.056	6.111	4.014
Biobío	20.842	21.320	21.367	14.535
La Araucanía	12.866	13.270	13.348	8.789
Los Ríos	5.095	5.097	5.249	3.406
Los Lagos	11.622	12.199	12.634	8.647
Aysén	1.546	1.661	1.712	1.092
Magallanes	2.103	2.149	2.239	1.481

Fuente: Instituto Nacional de Estadísticas

Finalmente, la metodología aplicada en el factor de expansión transversal de la ELPI para su cuarta ronda 2024, corresponde al producto entre el ponderador de selección, el ajuste por no contacto, el ajuste por

⁷ Corresponde a los niños y niñas nacidas entre el 1 de enero y el 31 de agosto del 2009

⁸ El total utilizado para el cálculo del factor de expansión transversal 2024 presenta diferencias a las versiones anteriores de la ELPI, debido a desactualizaciones del Marco Muestral que son incorporadas en esta versión.

no respuesta, suavizamiento y el ponderador de calibración usando la metodología de Raking Generalizado.

IV. Construcción del factor de expansión longitudinal del Panel

IV.1. Antecedentes de la propuesta del diagnóstico

Uno de los principales objetivos de la ELPI, es caracterizar y analizar el desarrollo de diferentes cohortes desde la primera infancia hasta la adolescencia. En otras palabras, busca realizar un seguimiento continuo de los niños y niñas que conformaron la muestra original capturada en el año 2010. En este contexto, se pone especial atención en aquellos casos que lograron ser contactados y entregar respuesta en cada una de las cuatro rondas del estudio. Estas unidades conforman el denominado Panel, cuya situación difiere significativamente de quienes no respondieron en alguna ronda o abandonaron el estudio.

Por esta razón, la metodología de cálculo del factor de expansión longitudinal asociados al Panel completo incorpora ajustes orientados a reflejar de mejor forma la trayectoria de estas unidades a lo largo del tiempo. Estos ajustes buscan, principalmente, optimizar el comportamiento de los ponderadores y aumentar la precisión de las estimaciones derivadas, asegurando así una representación más exacta sobre el comportamiento de la población objetivo.

A continuación, se presenta el detalle de las propuestas metodológicas implementadas en la construcción de los factores de expansión longitudinal para ELPI 2024. Estas fueron consensuadas entre el INE y el MDSF durante el año 2025 e incluyen, entre otros aspectos, una actualización del factor de expansión transversal correspondiente al año 2010, cuyo detalle también se aborda en el documento.

IV.2. Ponderador de Selección/Diseño del año 2010

Como se mencionó en el capítulo precedente, el cálculo del factor transversal de ELPI 2024 introdujo etapas adicionales con el objetivo de representar de mejor forma, por un lado, la estructura actual de la población objeto de estudio y, por otro, el resultado del trabajo de campo. Según lo anterior, para el cálculo del factor longitudinal, que utiliza como base el factor transversal inicial correspondiente al año 2010 se hace necesario que este último incorpore las mejoras aplicadas al factor del año 2024.

Para implementar estos cambios era necesario contar con la información detallada de los Códigos de Disposición Final (CDF) resultantes del trabajo de campo en el año 2010, así como con el listado completo de la muestra sobredimensionada que fue encuestada en la primera ronda. Sin embargo, debido al tiempo transcurrido y los diversos equipos que han gestionado la ELPI en estos 14 años, solo fue posible recuperar

información parcial; en concreto, se recuperó la muestra sobredimensionada de 25.570 niños y niñas, diferente a la publicada en el informe metodológico del año 2010, cuya muestra sobredimensionada es de 22.728 niños y niñas.

Con esta información recolectada, fue necesario recalculer la fracción de muestreo utilizada en el año 2010. Sabiendo que la muestra sobredimensionada total tuvo un tamaño de 25.570 niños y niñas, y considerando la población de 877.084, la fracción de muestreo (f) se calcula como:

$$f = \frac{n \text{ muestral}}{N \text{ poblacional}} = \frac{25.570}{877.084} = 0.02915342 \approx 0.029 \quad (1)$$

Es importante señalar que el diseño muestral del 2010 corresponde a un muestreo bietápico autoponderado, es decir, que la probabilidad de selección de un niño o niña es la misma independiente de la comuna y estrato en que resulte seleccionado. Por lo tanto, el ponderador de selección se define como el inverso de la fracción de muestreo total, dada por:

$$f = \frac{n}{N} = \frac{n_h}{N_h} \quad (2)$$

Además, por construcción se tiene que el inverso de la fracción de muestreo, corresponde al producto entre el inverso de la probabilidad de selección de la primera etapa (f_1) y el inverso de la probabilidad de selección del niño y niña de cada comuna seleccionada (f_2):

$$F_{ci} = \frac{1}{f} = \frac{1}{f_1 \times f_2} = \left(\frac{N_{hc}}{N_h} \times \frac{n_{hc}}{N_{hc}} \right)^{-1} = \left(\frac{n_{hc}}{N_h} \right)^{-1} = \left(\frac{n_c}{N_h} \right)^{-1} = \left(\frac{n}{N} \right)^{-1} = \frac{N}{n} \quad (3)$$

Este diseño es equivalente, en términos prácticos, a un muestreo aleatorio simple (MAS) aplicado a lo largo de todos los estratos, ya que solo se seleccionaba una comuna por estrato, simplificando así la estructura del diseño complejo. En resumen, la probabilidad de selección de un niño o niña es igual a $\frac{n}{N}$ independientemente de en qué comuna o estrato se encuentre, vale decir, como si se hubiera realizado un MAS de tamaño $n = n_1 + n_2 + \dots + n_L$, a través de todos los estratos con tamaño $N = N_1 + N_2 + \dots + N_L$.

Con el ponderador de selección corregido, se procedió a aplicar un ajuste por elegibilidad, utilizando los datos del CDF de 2010 recuperados por el MDSF. No obstante, debido a que no se dispone de variables observadas para los niños y niñas que no respondieron la encuesta, no fue posible ajustar por no respuesta mediante la metodología de Propensity Score, en su lugar, se aplicó un ajuste simple por estrato.

En síntesis, se incorporaron las posibles mejoras metodológicas al factor de expansión transversal del año 2010, que es el elemento base para el cálculo del factor de expansión longitudinal del panel para ELPI 2024. Esto permite contar con una estimación más coherente con las metodologías utilizadas actualmente en encuestas longitudinales de este tipo.

IV.3. Ajuste de no contacto y ajuste de no respuesta condicional al contacto

De acuerdo con lo señalado anteriormente en [IV.2 Ponderador de Selección/Diseño del año 2010](#), el factor de expansión longitudinal panel del año 2024 se construye a partir del factor de expansión transversal del año 2010, el cual fue corregido incorporando un ajuste por elegibilidad y un ajuste simple por no respuesta, debido a limitaciones en la información disponible.

No obstante, en el caso del panel actual, sí se cuenta con las variables necesarias para implementar una metodología más robusta de ajuste por no respuesta, mediante el uso de modelos de propensión de respuesta (Propensity Score). Por lo tanto, dado que la no respuesta constituye una de las principales fuentes de sesgo en estudios longitudinales, especialmente cuando está asociada a características individuales observables, se hace esencial aplicar un procedimiento que permita modelar la probabilidad de contacto y la respuesta condicionada al contacto.

Según lo anterior, por un lado, se construye un modelo logístico de no contacto, lo que permite obtener una estimación precisa de la propensión total de cada unidad a ser contactada a lo largo del tiempo. A continuación, se describe en detalle la construcción de la variable y la metodología utilizada para el cálculo de los respectivos pesos de ajuste.

- Variable contacto:

$$c = \begin{cases} 1, & \text{si se logró contacto en la ronda 2024 y hubo respuesta en las rondas previas} \\ 0, & \text{en caso contrario} \end{cases}$$

Esta variable representa a los casos en que fue posible establecer contacto durante la ronda 2024, siempre que también se haya obtenido respuesta en las rondas anteriores. Se identifican 7.418 NNA con valor 1, lo que corresponde al universo sobre el cual se puede estimar la probabilidad de contacto en la última ronda.

De manera similar que en la sección [III.2 Ajuste por no contacto](#), se probaron varias combinaciones de variables agrupadas en diferentes dimensiones, priorizando aquellas más relevantes para el análisis. Las variables finalmente incluidas en el modelo de no contacto se describen a continuación:

- Región Metropolitana: una variable dicotómica que indica si una observación pertenece (1) o no pertenece (0) a la Región Metropolitana.
- Quintil de Ingreso Familiar: Se creó una variable que categoriza por quintiles el ingreso per cápita familiar a cada observación, dada la información de ELPI 2010.
- Sexo: una variable dicotómica que indica si el responsable principal es Hombre (1) o si es Mujer (2)
- Tramo Educativo: Se definen 13 niveles educativos definidos para el año 2024, complementada con información de rondas anteriores. Las categorías son: Sin educación formal, Básica incompleta, Básica completa, Media humanista incompleta, Media técnica profesional incompleta, Media humanista completa, Media técnica profesional completa, Técnico nivel superior incompleta, Técnico nivel superior completo, Profesional incompleto, Posgrado incompleto, Profesional completo, Posgrado completo y No especificado.
- Estado Civil: Se genera una variable según el estado civil del responsable principal. Las categorías son: Separado (que incluye si es separado, divorciado o anulado), Casado(a), Soltero, Pareja (que incluye Conviviente o pareja sin acuerdo de unión civil), AUC (que corresponde a los convivientes con acuerdo de unión civil) y Otros (que incluye si es viudo o no se señala).
- Actividad: Variable categórica construida según la actividad que realiza el responsable principal. Las categorías son Desocupados, Ocupados e Inactivos.
- Tramo Etario del Encuestador: Las categorías son joven, entre 18 y 30 años; joven adulto, entre 31 y 40 años; adulto, entre 41 y 59 años; adulto mayor, entre 60 y 75 años, no especificado, si la edad del encuestador no está disponible.

A partir de este modelo, se obtiene la probabilidad de que un individuo haya sido contactado, es decir:

$$P(c) = \hat{p}(c = 1|x)$$

Luego, las probabilidades estimadas se ordenan de menor a mayor y se estratifican en 10 grupos o deciles, con ellos, se distribuye el peso del total de los NNA en los que efectivamente contestaron en las primeras 3 rondas y además fueron contactados en 2024.

Por otro lado, se construye un modelo logístico de no respuesta condicional al contacto, lo que permite obtener una estimación precisa de la propensión total de cada unidad a permanecer en el panel a lo largo del tiempo, dado que ya fue contactado. A continuación, se describe en detalle la construcción de la variable y la metodología utilizada para el cálculo de los respectivos pesos de ajuste.

- Variable respuesta condicional al contacto:

$$r = \begin{cases} 1, & \text{si hubo respuesta en todas las rondas} \\ 0, & \text{en caso contrario} \end{cases}$$

Esta variable identifica a los niños, niñas y adolescentes (NNA) cuya participación fue completa a lo largo de todas las rondas de la encuesta. En total, se registran 7.012 casos con valor, lo que indica que estos NNA respondieron en todas las mediciones realizadas.

Para la construcción del modelo de no respuesta condicional al contacto se utilizaron las mismas variables que para el modelo de no contacto, mencionadas anteriormente. Con dicho modelo, se obtiene la probabilidad de que un individuo haya respondido dado que fue contactado, es decir:

$$\hat{p}(r = 1 | x, c = 1)$$

Luego, las probabilidades estimadas se ordenan de menor a mayor y se estratifican en 10 grupos o deciles, distribuyendo el peso del total de los NNA entre los que efectivamente contestaron en las todas las rondas.

Finalmente, este procedimiento permite agrupar las observaciones según niveles similares de propensión a permanecer en el panel, facilitando la aplicación de ajustes posteriores, para mitigar los sesgos asociados a la no respuesta condicionada al contacto y mejorar la precisión de las estimaciones longitudinales.

IV.4. Suavizamiento

La metodología de suavizamiento aplicada en el factor de expansión longitudinal del Panel, sigue los mismos lineamientos descritos en el apartado [III.4 Suavizamiento](#), y corresponde al método mixto que, como se explicó previamente, combina el enfoque de contracción de la media y el método R-K.

Al igual que en el caso del factor de expansión transversal, la propuesta presentada al MDSF señalaba que la aplicación de este procedimiento debía evaluarse solo en caso de detectarse valores atípicos antes del proceso de calibración. En esta instancia, una vez completado el ajuste por no respuesta, se identificaron 443 factores de expansión considerados atípicos, lo que justificó la aplicación del suavizamiento.

Como resultado de este tratamiento, el rango de los factores de expansión se redujo, de un intervalo entre 65,2 y 1.216,09 a un rango más acotado entre 69,06 y 1.188,03. Esta disminución en la dispersión indica una mejora en la estabilidad de los ponderadores, lo que, a su vez, contribuye a generar estimaciones más consistentes y representativas para el análisis longitudinal de la encuesta.

IV.5. Ponderador de calibración

Como se expuso en la sección [III.5 Ponderador de calibración](#) del documento, la implementación del método de *Raking Generalizado* permite mejorar la eficiencia de los factores de expansión al reducir su variabilidad sin sacrificar la coherencia de los totales poblacionales conocidos. Por esta razón, es que se mantiene esta metodología como parte de la construcción del factor de expansión longitudinal.

En este contexto, se utilizó nuevamente como marco de calibración para los totales poblacionales, la información recuperada por el Departamento de Demografía del INE, correspondiente a los niños y niñas nacidos entre el 1 de enero de 2006 y el 31 de agosto de 2009. Sin embargo, existe una leve diferencia de lo realizado en el factor de expansión transversal previamente calculado, ya que, para este caso no se aplicó la corrección por defunciones. Debido a que, el interés principal del análisis longitudinal radica en observar los cambios y trayectorias de la población originalmente encuestada en la primera ronda de la ELPI, independientemente de su situación actual. En otras palabras, se busca mantener constante la base de análisis para permitir comparaciones válidas a lo largo del tiempo.

Con ello, el marco muestral que se utilizó para calcular el ponderador de calibración con la metodología de *Raking Generalizado* se observa en la Tabla 4:

Tabla 4: Marco muestral por Región/Cohorte para cálculo de factor de expansión longitudinal del Panel

Región	Población de niños y niñas por Región y Cohorte			
	2006	2007	2008	2009 ⁹
Nacional¹⁰	231.383	240.569	246.581	168.200
Arica y Parinacota	3.099	3.158	3.364	2.401
Tarapacá	4.761	5.017	5.287	3.765
Antofagasta	8.975	9.419	9.706	6.655
Atacama	4.648	4.720	4.803	3.466
Coquimbo	9.372	9.794	10.447	7.159
Valparaíso	21.672	22.659	23.116	15.890
Metropolitana	93.593	97.596	99.873	68.570
O'Higgins	11.517	12.199	12.592	8.438
Maule	13.099	13.566	14.045	9.440
Ñuble	5.943	6.124	6.178	4.057
Biobío	21.085	21.558	21.601	14.691
La Araucanía	13.016	13.418	13.494	8.884
Los Ríos	5.155	5.154	5.306	3.443
Los Lagos	11.758	12.335	12.772	8.740
Aysén	1.563	1.679	1.733	1.104
Magallanes	2.127	2.173	2.264	1.497

Fuente: Instituto Nacional de Estadísticas

Finalmente, la metodología aplicada en el factor de expansión longitudinal del Panel de la ELPI para su cuarta ronda 2024, corresponde al producto entre el factor de expansión transversal de la primera ronda mejorado (antes de calibrar), el ajuste por no contacto y no respuesta condicional al contacto suavizado y el ponderador de calibración usando la metodología de Raking Generalizado.

V. Construcción de factores de expansión longitudinal del año 2010-2024

V.1. Antecedentes de la propuesta del diagnóstico

En la sección anterior, dedicada a la construcción del factor de expansión longitudinal del Panel, se destacó la importancia de analizar el comportamiento de la población que participó en todas las rondas de la ELPI, ya que esto permite comprender la evolución de los niños y niñas a lo largo del tiempo. Sin embargo, el cálculo del factor de expansión longitudinal 2010-2024 se plantea con un objetivo complementario: entender cómo se comportó la población que solo logró ser contactada en la primera y en la última ronda de la ELPI.

⁹ Corresponde a los niños y niñas nacidas entre el 1 de enero y el 31 de agosto del 2009

¹⁰ El total utilizado para el cálculo del factor de expansión transversal 2024 presenta diferencias a las versiones anteriores de la ELPI, debido a desactualizaciones del Marco Muestral que son incorporadas en esta versión.

De manera similar al factor del Panel, este factor incorpora las mismas mejoras metodológicas que buscan reflejar con mayor precisión la trayectoria de estas unidades, optimizando el comportamiento de los ponderadores y mejorando la exactitud de las estimaciones derivadas. De esta manera, se asegura una representación más exacta de la población objetivo.

A continuación, se detalla el conjunto de metodologías implementadas en la construcción del factor de expansión longitudinal 2010-2024, las que fueron consensuadas entre el INE y el Ministerio durante el año 2025, siguiendo la misma línea de trabajo que para el factor del Panel.

V.2. Ponderador de Selección/Diseño del año 2010

En el cálculo del factor longitudinal 2010-2024, el ponderador de selección mantiene la misma lógica de construcción que se aplicó para el factor de expansión longitudinal del Panel de la ELPI 2024. Esto implica replicar la metodología de mejora sobre el factor de expansión transversal del año 2010, ya que como el objetivo principal de este factor es comprender el comportamiento de la población originalmente encuestada en 2010 y que, tras varios años, logró ser nuevamente contactada en la cuarta ronda, independientemente de si respondió o no en todas las mediciones intermedias.

Siguiendo la misma línea metodológica que en el factor del panel, se corrigió el ponderador de selección y se aplicó un ajuste por elegibilidad, utilizando los CDF recuperados. Nuevamente, no fue posible aplicar un modelo de propensión (Propensity Score); por lo que se mantiene el ajuste simple a nivel de estrato.

Con este procedimiento se logró actualizar el factor de expansión longitudinal 2010-2024 a las prácticas metodológicas actuales, asegurando así una estimación más coherente y alineada con los objetivos del estudio longitudinal.

V.3. Ajuste de no contacto y ajuste de no respuesta condicional al contacto

Para la construcción del factor de expansión longitudinal correspondiente al período 2010–2024, se utilizó el mismo enfoque metodológico aplicado en el Panel completo, mencionado en [IV.3. Ajuste de no contacto y ajuste de no respuesta condicional al contacto](#), basado en modelos de propensión de respuesta (*Propensity Score*). Este enfoque permite modelar de manera más precisa la probabilidad de que un individuo permanezca en el panel entre ambos años, y así corregir los sesgos asociados a la no respuesta condicionada al contacto a lo largo del tiempo.

Si bien se mantienen las mismas variables explicativas utilizadas previamente en [IV.3. Ajuste de no respuesta](#)—agrupadas en dimensiones como características del niño o niña, del responsable principal, del encuestador y del contexto territorial—, en este caso particular la construcción de las variables

dependientes difiere, ya que el modelo considera únicamente la continuidad entre las mediciones de los años 2010 y 2024, sin requerir respuesta en las rondas intermedias del panel.

Dado que todos los niños y niñas incluidos en la muestra respondieron la encuesta en 2010, el análisis se centra exclusivamente en el comportamiento observado en la ronda de 2024. Para la estimación de los modelos de contacto y respuesta condicional al contacto en este período, se definieron las siguientes variables binarias:

Variable contacto:

$$c = \begin{cases} 1, & \text{si se logró contacto en la ronda 2024} \\ 0, & \text{en caso contrario} \end{cases}$$

Esta variable distingue entre aquellos casos en que fue posible establecer contacto con el hogar en 2024 y aquellos en que no se logró dicha conexión. Se identifican 11.101 NNA con valor 1, lo que representa el conjunto de casos en los cuales se tuvo la oportunidad de aplicar efectivamente el instrumento, independientemente de si finalmente se obtuvo respuesta. Este conjunto constituye la base para estimar la probabilidad de respuesta condicional al contacto.

- Variable respuesta condicional al contacto (r):

$$r = \begin{cases} 1, & \text{si hubo respuesta en la ronda 2024} \\ 0, & \text{en caso contrario} \end{cases}$$

Esta variable permite identificar a los niños, niñas y adolescentes (NNA) que continuaron participando activamente en el estudio durante la medición del año 2024. En total, se registran 10.003 casos con valor $r = 1$, lo que representa a las unidades que respondieron efectivamente la encuesta en ambas rondas — 2010 y 2024—, aunque, como se señaló previamente, la participación en 2010 es condición de inclusión en el análisis. Por lo tanto, estos casos constituyen el subgrupo longitudinal completo sobre el cual se estima la probabilidad de respuesta condicional al contacto.

A partir del conjunto de variables iniciales x , mencionadas en el apartado [III.3 Ajuste de no respuesta](#) se realizó una selección de variables para construir ambos modelos, priorizando aquellas más relevantes para el análisis. Las variables finalmente incluidas se describen a continuación.

- Quintil de Ingreso Familiar: Se creó una variable que categoriza por quintiles el ingreso per cápita familiar a cada observación, dada la información de ELPI 2010.
- Sexo: una variable dicotómica que indica si el responsable principal es Hombre (1) o si es Mujer (2)

- Tramo Educativo: Se definen 13 niveles educativos definidos para el año 2024, complementada con información de rondas anteriores. Las categorías son: Sin educación formal, Básica incompleta, Básica completa, Media humanista incompleta, Media técnica profesional incompleta, Media humanista completa, Media técnica profesional completa, Técnico nivel superior incompleta, Técnico nivel superior completo, Profesional incompleto, Posgrado incompleto, Profesional completo, Posgrado completo y No especificado.
- Estado Civil: Se genera una variable según el estado civil del responsable principal. Las categorías son: Separado (que incluye si es separado, divorciado o anulado), Casado(a), Soltero, Pareja (que incluye Conviviente o pareja sin acuerdo de unión civil), AUC (que corresponde a los convivientes con acuerdo de unión civil) y Otros (que incluye si es viudo o no se señala).
- Actividad: Variable categórica construida según la actividad que realiza el responsable principal. Las categorías son Desocupados, Ocupados e Inactivos.
- Tramo Etario del Encuestador: Las categorías son joven, entre 18 y 30 años; joven adulto, entre 31 y 40 años; adulto, entre 41 y 59 años; adulto mayor, entre 60 y 75 años y, no especificado, si la edad del encuestador no está disponible.

Además, se incorpora una interacción entre dos variables geográficas, cuyo objetivo es capturar diferencias en los patrones de contacto y respuesta según la combinación de región y área de residencia. Las variables utilizadas para construir esta interacción se describen a continuación:

- Región Metropolitana: una variable dicotómica que indica si una observación pertenece (1) o no pertenece (0) a la Región Metropolitana.
- Área: Variable que identifica el tipo de área geográfica, tomando el valor 1 para zonas urbanas y 2 para zonas rurales.

La interacción entre ambas variables permite identificar subgrupos específicos, que podrían presentar patrones particulares de contacto o respuesta. En el modelo, la combinación **Área rural dentro de la Región Metropolitana** se utiliza como categoría de referencia. Por lo tanto, todos los efectos estimados para las demás combinaciones (urbano en RM, urbano fuera de RM, rural fuera de RM) se interpretan con relación a este grupo base.

Esta especificación busca mejorar la capacidad explicativa del modelo, permitiendo detectar diferencias que podrían quedar ocultas si ambas variables se incluyeran de manera independiente.

V.4. Suavizamiento

Nuevamente la metodología de suavizamiento aplicada en el factor de expansión longitudinal del año 2010-2024, sigue los mismos lineamientos descritos en el apartado [III.4 Suavizamiento](#), que corresponde al método mixto, que mezcla el enfoque de contracción de la media y el método R-K.

Al igual que en los casos anteriores, la propuesta presentada al MDSF señala que la aplicación de este procedimiento debe aplicarse solo en el caso de detectarse valores atípicos antes del proceso de calibración. En esta instancia, una vez completado el ajuste por no respuesta, se identificaron 615 factores de expansión considerados atípicos, lo que justificó la aplicación del suavizamiento.

Como resultado de este tratamiento, el rango de los factores de expansión se redujo, de un intervalo entre 45,68 y 735,54 a un rango más acotado entre 48,77 y 717,80. Esta disminución en la dispersión de los factores de expansión refleja una mejora en la estabilidad del ponderador, contribuyendo a una mejora en las estimaciones derivadas.

V.5. Ponderador de calibración

Para finalizar, en el ajuste del factor de expansión longitudinal 2010-2024, se implementó nuevamente el método de Raking Generalizado, dado que, como se ha mencionado, este permite mejorar la eficiencia de los factores de expansión al reducir su variabilidad sin sacrificar la coherencia con los totales poblacionales conocidos.

En este contexto, se utilizó como marco de calibración, al igual que en el factor anterior, la información recuperada por el Departamento de Demografía del INE, correspondiente a los niños y niñas nacidos entre el 1 de enero de 2006 y el 31 de agosto de 2009. En esta ocasión, se siguió el mismo criterio aplicado para el factor de expansión longitudinal del Panel, es decir, no se incorporó la corrección por defunciones, ya que el interés principal de este análisis es observar la evolución y trayectoria de la población originalmente encuestada en la primera ronda de la ELPI, independientemente de su situación actual.

En síntesis, el marco muestral que se empleó para calcular el ponderador de calibración mediante el método de *Raking Generalizado* corresponde al que se detalla en la Tabla 5:

Tabla 5: Marco Muestral por Región/Cohorte para cálculo de factor de expansión longitudinal del Panel

Región	Población de niños y niñas por Región y Cohorte			
	2006	2007	2008	2009 ¹¹
Nacional¹²	231.383	240.569	246.581	168.200
Arica y Parinacota	3.099	3.158	3.364	2.401
Tarapacá	4.761	5.017	5.287	3.765
Antofagasta	8.975	9.419	9.706	6.655
Atacama	4.648	4.720	4.803	3.466
Coquimbo	9.372	9.794	10.447	7.159
Valparaíso	21.672	22.659	23.116	15.890
Metropolitana	93.593	97.596	99.873	68.570
O'Higgins	11.517	12.199	12.592	8.438
Maule	13.099	13.566	14.045	9.440
Ñuble	5.943	6.124	6.178	4.057
Biobío	21.085	21.558	21.601	14.691
La Araucanía	13.016	13.418	13.494	8.884
Los Ríos	5.155	5.154	5.306	3.443
Los Lagos	11.758	12.335	12.772	8.740
Aysén	1.563	1.679	1.733	1.104
Magallanes	2.127	2.173	2.264	1.497

Fuente: Instituto Nacional de Estadísticas

En síntesis, la metodología aplicada en el factor de expansión longitudinal del año 2010-2024 de la ELPI 2024, incluye el producto entre el factor de expansión transversal de la primera ronda (actualizado y antes de calibrar), el ajuste por no contacto, el ajuste por no respuesta condicional al contacto, suavizamiento y la calibración usando la metodología de Raking Generalizado.

VI. Cálculo y evaluación de las estimaciones

El objetivo de este capítulo es presentar las estimaciones de los parámetros de interés definidos para este estudio, considerando la variable Puntaje Estándar Hispanoamericano TVIP, junto con sus respectivos errores muestrales, para los dominios previamente establecidos.

¹¹ Corresponde a los niños y niñas nacidas entre el 1 de enero y el 31 de agosto del 2009

¹² El total utilizado para el cálculo del factor de expansión transversal 2024 presenta diferencias a las versiones anteriores de la ELPI, debido a desactualizaciones del Marco Muestral que son incorporadas en esta versión.

VI.1. Estadísticos asociados a los parámetros de interés

A partir de la muestra obtenida en la Encuesta Longitudinal de Primera Infancia (ELPI) 2024, se calculan los estadísticos asociados al parámetro de interés definido por el Ministerio, correspondiente a la variable Puntaje Estándar Hispanoamericano TVIP.

Esta variable se construye a partir de los resultados del Test de Vocabulario en Imágenes Peabody (TVIP), instrumento que evalúa el vocabulario receptivo o auditivo de los niños y niñas, es decir, su capacidad para comprender palabras al escucharlas. El test consiste en que el niño escucha una palabra y debe identificar la imagen correcta entre cuatro opciones, siendo ampliamente utilizado en contextos educativos, clínicos y de investigación.

Las Tablas 6, 7 y 8 presentan las estimaciones del Puntaje Estándar Hispanoamericano TVIP, junto con sus respectivos errores muestrales, expresados a través del error estándar, error absoluto, error relativo y los límites del coeficiente de variación (CV), tanto a nivel nacional como regional. Cada tabla corresponde a uno de los tres factores de expansión considerados en este documento.

Tabla 6. Estimación del Puntaje Estándar Hispanoamericano TVIP con sus respectivos errores muestrales a nivel nacional y regional, según factor de expansión transversal ELPI 2024

Región	TVIP					
	Estimación	Error estándar	Error absoluto	Error relativo	IC límite inferior	IC límite superior
Nacional	107.806	0,154	0,302	0,003	107,504	108,108
Arica y Parinacota	109.482	1,002	1,980	0,018	107,519	111,446
Tarapacá	106.837	1,057	2,085	0,020	104,765	108,909
Antofagasta	111.587	0,615	1,208	0,011	110,382	112,791
Atacama	110.722	1,259	2,486	0,022	108,254	113,189
Coquimbo	109.809	0,636	1,251	0,011	108,562	111,057
Valparaíso	107.509	0,518	1,016	0,009	106,494	108,523
Metropolitana	107.720	0,253	0,497	0,005	107,223	108,216
O'Higgins	107.949	0,565	1,110	0,010	106,841	109,056
Maule	110.953	0,549	1,078	0,010	109,877	112,030
Ñuble	107.078	0,895	1,763	0,016	105,324	108,831
Biobío	106.994	0,498	0,977	0,009	106,018	107,969
La Araucanía	104.918	0,720	1,414	0,013	103,507	106,328
Los Ríos	108.580	1,223	2,414	0,022	106,184	110,977
Los Lagos	104.667	0,699	1,372	0,013	103,298	106,036
Aysén	110.026	1,555	3,098	0,028	106,978	113,075
Magallanes	104.000	2,023	4,022	0,039	100,036	107,965

Elaborado por el Instituto Nacional de Estadísticas.

Tabla 7. Estimación del Puntaje Estándar Hispanoamericano TVIP con sus respectivos errores muestrales a nivel nacional y regional, según factor de expansión longitudinal panel, ELPI 2024.

Región	TVIP					
	Estimación	Error estándar	Error absoluto	Error relativo	IC límite inferior	IC límite superior
Nacional	108,030	0,188	0,368	0,003	107,661	108,398
Arica y Parinacota	109,320	1,255	2,486	0,023	106,861	111,779
Tarapacá	107,286	1,175	2,324	0,022	104,982	109,589
Antofagasta	111,258	0,723	1,423	0,013	109,840	112,675
Atacama	110,530	1,300	2,571	0,023	107,982	113,078
Coquimbo	109,931	0,731	1,438	0,013	108,498	111,364
Valparaíso	108,342	0,587	1,152	0,011	107,192	109,491
Metropolitana	107,961	0,328	0,642	0,006	107,320	108,603
O'Higgins	108,316	0,635	1,248	0,012	107,071	109,560
Maule	110,960	0,668	1,313	0,012	109,650	112,269
Ñuble	107,096	1,045	2,063	0,019	105,047	109,145
Biobío	106,268	0,606	1,189	0,011	105,081	107,456
La Araucanía	104,947	0,941	1,850	0,018	103,103	106,791
Los Ríos	109,972	1,537	3,048	0,028	106,960	112,985
Los Lagos	105,495	0,786	1,546	0,015	103,954	107,036
Aysén	110,786	1,655	3,310	0,030	107,543	114,030
Magallanes	105,864	2,257	4,506	0,043	101,440	110,288

Elaborado por el Instituto Nacional de Estadísticas,

Tabla 8. Estimación del Puntaje Estándar Hispanoamericano TVIP con sus respectivos errores muestrales a nivel nacional y regional, según factor de expansión longitudinal 2010-2024, ELPI 2024.

Región	TVIP					
	Estimación	Error estándar	Error absoluto	Error relativo	IC límite inferior	IC límite superior
Nacional	108,023	0,155	0,304	0,003	107,718	108,327
Arica y Parinacota	109,613	1,029	2,035	0,019	107,595	111,63
Tarapacá	106,988	1,04	2,05	0,019	104,95	109,025
Antofagasta	111,471	0,641	1,259	0,011	110,215	112,727
Atacama	110,999	1,2	2,37	0,021	108,646	113,352
Coquimbo	109,998	0,635	1,249	0,011	108,753	111,244
Valparaíso	107,877	0,509	0,999	0,009	106,879	108,874
Metropolitana	108,047	0,259	0,507	0,005	107,54	108,554
O'Higgins	108,224	0,59	1,158	0,011	107,068	109,38
Maule	110,954	0,559	1,098	0,01	109,858	112,049
Ñuble	107,246	0,903	1,78	0,017	105,476	109,016
Biobío	106,852	0,504	0,989	0,009	105,864	107,839
La Araucanía	105,028	0,749	1,471	0,014	103,56	106,495
Los Ríos	108,982	1,258	2,483	0,023	106,517	111,447
Los Lagos	104,765	0,715	1,404	0,013	103,364	106,166
Aysén	110,605	1,384	2,758	0,025	107,892	113,319
Magallanes	104,639	2,014	4,005	0,038	100,69	108,587

Elaborado por el Instituto Nacional de Estadísticas,

VII. ANEXOS

VII.1. ANEXO N°1: Tablas marginales utilizadas para la calibración por Raking del factor de expansión transversal ELPI 2024

Tabla 9: Marco Muestral para cálculo de factor de expansión transversal

Región	Población de niños y niñas por Región y Cohorte							
	Masculino				Femenino			
Sexo								
Cohorte	2006	2007	2008	2009	2006	2007	2008	2009
Nacional	118,203	122,952	125,858	85,750	113180	117,617	120,723	82,450
Arica y Parinacota	1,544	1,558	1,707	1,231	1,555	1,600	1,657	1,170
Tarapacá	2,496	2,597	2,744	1,951	2,265	2,420	2,543	1,814
Antofagasta	4,556	4,844	4,935	3,421	4,419	4,575	4,771	3,234
Atacama	2,363	2,470	2,415	1,738	2,285	2,250	2,388	1,728
Coquimbo	4,856	5,061	5,261	3,569	4,516	4,733	5,186	3,590
Valparaíso	10,988	11,452	11,850	8,049	10,684	11,207	11,266	7,841
Metropolitana	47,817	49,862	51,171	35,057	45,776	47,734	48,702	33,513
O'Higgins	5,834	6,300	6,435	4,285	5,683	5,899	6,157	4,153
Maule	6,789	6,934	7,065	4,870	6,310	6,632	6,980	4,570
Ñuble	3,038	3,124	3,131	2,073	2,905	3,000	3,047	1,984
Biobío	10,898	11,073	11,001	7,533	10,187	10,485	10,600	7,158
La Araucanía	6,592	6,815	6,881	4,447	6,424	6,603	6,613	4,437
Los Ríos	2,526	2,636	2,683	1,765	2,629	2,518	2,623	1,678
Los Lagos	5,973	6,211	6,568	4,444	5,785	6,124	6,204	4,296
Aysén	803	883	894	572	760	796	839	532
Magallanes	1,130	1,132	1,117	745	997	1,041	1,147	752

Fuente: Instituto Nacional de Estadísticas

Tabla 10: Marco Muestral por Región/Sexo para cálculo de factor de expansión transversal

Región	Población de niños y niñas por región	
	Masculino	Femenino
Nacional	447,765	429,179
Arica y Parinacota	5,973	5,916
Tarapacá	9,680	8,942
Antofagasta	17,560	16,811
Atacama	8,887	8,555
Coquimbo	18,541	17,826
Valparaíso	41,872	40,546
Metropolitana	181,877	173,785
O'Higgins	22,602	21,650
Maule	25,375	24,222
Ñuble	11,240	10,815
Biobío	40,058	38,006
La Araucanía	24,462	23,811
Los Ríos	9,504	9,343
Los Lagos	22,940	22,162
Aysén	3,116	2,895
Magallanes	4,078	3,894

Fuente: Instituto Nacional de Estadísticas

Tabla 11: Marco Muestral por Cohorte/Sexo para cálculo de factor de expansión transversal

Región	Población de niños y niñas por Cohorte	
	Masculino	Femenino
Nacional	447,765	429,179
2006	116,839	111,874
2007	121,594	116,318
2008	124,495	119,415
2009	84,837	81,572

Fuente: Instituto Nacional de Estadísticas

Tabla 12: Marco Muestral para cálculo de factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024

Región	Población de niños y niñas por Región
Nacional	886,733
Arica y Parinacota	11,889
Tarapacá	18,622
Antofagasta	34,371
Atacama	17,442
Coquimbo	36,367
Valparaíso	82,418
Metropolitana	355,662
O'Higgins	44,252
Maule	49,597
Ñuble	22,055
Biobío	78,064
La Araucanía	48,273
Los Ríos	18,847
Los Lagos	45,102
Aysén	6,011
Magallanes	7,972

Fuente: Instituto Nacional de Estadísticas

VII.2. ANEXO N°2: Tablas marginales utilizadas para la calibración por Raking del factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024

Tabla 13: Marco Muestral por Región/Sexo para cálculo de factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024

Región	Población de niños y niñas por Región y Sexo	
	Masculino	Femenino
Nacional	452,763	433,970
Arica y Parinacota	3,099	3,158
Tarapacá	4,761	5,017
Antofagasta	8,975	9,419
Atacama	4,648	4,720
Coquimbo	9,372	9,794
Valparaíso	21,672	22,659
Metropolitana	93,593	97,596
O'Higgins	11,517	12,199
Maule	13,099	13,566
Ñuble	5,943	6,124
Biobío	21,085	21,558
La Araucanía	13,016	13,418
Los Ríos	5,155	5,154
Los Lagos	11,758	12,335
Aysén	1,563	1,679
Magallanes	2,127	2,173

Fuente: Instituto Nacional de Estadísticas

Tabla 14: Marco Muestral por Cohorte/Sexo para cálculo de factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024

Región	Población de niños y niñas por Cohorte y Sexo	
	Masculino	Femenino
Nacional	452,763	433,970
2006	118,203	113,180
2007	122,952	117,617
2008	125,858	120,723
2009	85,750	82,450

Fuente: Instituto Nacional de Estadísticas

Tabla 15: Marco Muestral para cálculo de factor de expansión longitudinal Panel y factor de expansión longitudinal 2010-2024

Región	Población de niños y niñas por Región
Nacional	886,733
Arica y Parinacota	12,022
Tarapacá	18,830
Antofagasta	34,755
Atacama	17,637
Coquimbo	36,772
Valparaíso	83,337
Metropolitana	359,632
O'Higgins	44,746
Maule	50,150
Ñuble	22,302
Biobío	78,935
La Araucanía	48,812
Los Ríos	19,058
Los Lagos	45,605
Aysén	6,079
Magallanes	8,061

Fuente: Instituto Nacional de Estadísticas

VIII. Bibliografía

- CEEL UC, C, d, (2018), *Encuesta Longitudinal de Primera Infancia (III Ronda)*, Santiago,
- CMD, C, d, (2010), *Diseño Muestral y Factores de Expansión Primera Ronda ELPI*, Santiago,
- CMD, C, d, (2012), *Levantamiento Final de la Encuesta Longitudinal de la Primera Infancia, Segunda Ronda*, Santiago,
- Cochran, W, (1998), *Técnicas de Muestreo*, México D,F,: Compañía Editorial Continental, S,A, de C,V,
- INE, (2019), *Quinta Encuesta Longitudinal de Empresas*, Santiago,
- INE, (Diciembre de 2019), *www,ine,gob,cl*, Obtenido de https://www,ine,gob,cl/docs/default-source/documentos-de-trabajo/documento-de-trabajo-fundamentos-de-la-nueva-calibraci%C3%B3n-de-los-factores-de-expansi%C3%B3n-en-la-ene,pdf?sfvrsn=3de3a0e1_4
- INE, (2020), *Análisis del efecto del COVID-19 sobre el diseño muestral y las estimaciones de la ENE en EFM 2020*, Santiago,
- INE, (2020), *La nueva metodología de calibración de la Encuesta Nacional de Empleo: método de calibración Raking*, Santiago,
- INE, (2022 de Febrero de 2022), *Especificación de la Nueva Metodología de Diseño de Factores de Expansión de la encuesta Casen*, Santiago,
- Social, D, O, (Febrero de 2022), <https://observatorio,ministeriodesarrollosocial,gob,cl>, Obtenido de https://observatorio,ministeriodesarrollosocial,gob,cl/storage/docs/casen/2020/Nota_tecnica8_Nueva_metodologia_Calibracion,pdf
- Social, D, O, (2024), *Proceso de Diseño de Cuestionario ELPI Cuarta Ronda*, Santiago,
- Social, O,, & Social, S, d, (2021), *Manual de Usuario Cuarta Ronda*, Santiago,