X ENCLA
Encuesta Laboral 2023

Décima Encuesta Laboral
Encla 2023

Metodología del Diseño Muestral

Encla 2023

Instituto Nacional de Estadísticas • Chile

Dirección del Trabajo
Gobierno de Chile

DEPARTAMENTO DE METODOLOGÍAS E INNOVACIÓN ESTADÍSTICA
SUBDEPARTAMENTO DE DISEÑO DE MARCOS Y MUESTRAS
SECCIÓN DE MUESTRAS ECONÓMICAS

Encuesta Laboral.
ENCLA 2023.

Informe de metodología del diseño muestral.
Entrega N°4 / Versión N°3, revisado septiembre 2024.

Instituto Nacional de Estadísticas, Chile.
28 de junio de 2024.

ÍNDICE

PRESENTACIÓN 6

I. OBJETIVO 8

I.1. Objetivo general 8

I.2. Objetivos específicos 8

II. CARACTERISTICAS DEL DISEÑO MUESTRAL 8

II.1. Población objetivo 8

II.2. Dominios de estudio 9

II.3. Período de referencia y periodicidad 9

II.4. Marco Muestral del producto 9

II.4.1. Cobertura temática 9

II.4.2. Cobertura geográfica 11

II.4.3. Construcción del marco muestral 12

II.4.4. Estratificación del marco muestral 15

II.5. Unidades estadísticas 17

II.5.1. Unidad de muestreo 17

II.5.2. Unidad de información y de análisis 17

III. METODOLOGÍA DE CÁLCULO DEL TAMAÑO MUESTRAL 18

III.1. Estrategia muestral 18

III.1.1. Consideraciones para el tamaño muestral 18

III.2. Cálculo del Tamaño muestral 19

III.3. Selección de unidades muestrales 23

IV. RESULTADOS DEL TAMAÑO MUESTRAL 24

V. DESARROLLO DEL FACTOR DE EXPANSIÓN 30

V.1. Factor de expansión para las empresas del tramo de inclusión aleatoria 30

V.1.1. Factor de expansión teórico 30

V.1.2. Ajuste por elegibilidad 31

V.1.3. Ajuste por no respuesta 37

V.1.4. Suavizamiento del factor de expansión 39

V.1.5. Calibración del factor de expansión 41

V.2. Factor de expansión para las empresas del tramo de inclusión forzosa 42

VI. ESTIMADORES Y MEDIDAS DE PRECISIÓN 43

VII. ESTIMACIONES DE LOS PARÁMETROS DE INTERÉS 44

VIII. ANEXOS 48

IX. BIBLIOGRAFIA 49

ÍNDICE DE CUADROS

Cuadro II.1 Secciones económicas en estudio para la Encla 2023. 10
Cuadro II.2 Secciones económicas excluidas del ámbito de estudio para la Encla 2023. 11
Cuadro II.3 Comunas excluidas del ámbito de estudio de la Encla 2023. 11
Cuadro II.4. Criterios de priorización para el tamaño de empresas en el marco muestral de la Encla 2023 15
Cuadro II.5. Agrupación regional a partir de las regiones, solicitadas por la DT 16
Cuadro II.6. Tamaño de empresas según número de trabajadores asalariados 16
Cuadro III.1. Elementos utilizados para el cálculo del tamaño muestral de Encla 2023 18
Cuadro VIII.1. Definición códigos de disposición final AAPOR 48

ÍNDICE DE TABLAS

Tabla IV.1. Tamaño muestral objetivo y efectivo según tamaños de empresas. 25
Tabla IV.2. Tamaño muestral objetivo y efectivo según región. 25
Tabla IV.3. Tamaño muestral objetivo y efectivo según agrupación de actividad económica. 26
Tabla IV.4. Tamaño muestral objetivo y efectivo según región y tamaño. 26
Tabla IV.5. Tamaño muestral objetivo y efectivo según agrupación de región y tamaño. 28
Tabla VII.1. Estimación y error relativo del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, según agrupación de actividad económica 45
Tabla VII.2. Estimación y error relativo del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, según tamaño de empresa 45
Tabla VII.3. Estimación y error relativo del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, según agrupación de región y tamaños 46
Tabla VII.4. Estimación y error relativo del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, según región. 47

ÍNDICE DE FIGURAS

Figura II.1. Diagrama del cruce entre el marco muestral F1887 y SUSESO 13
Figura II.2. Distribución de las empresas que no coinciden en ambas fuentes de información, según tamaño de la empresa 14
Figura II.3. Fuentes de información utilizadas para depurar el marco muestral 14
Figura II.4 Estratificación del marco muestral para la Encla 2023. 17

PRESENTACIÓN

Desde el año 1998 el Departamento de Estudios de la Dirección del Trabajo efectúa la Encuesta Laboral (Encla), cuyo objetivo es levantar información sobre las condiciones de trabajo y relaciones laborales en las empresas chilenas que se ubiquen en el territorio nacional, a partir de la cual se encarga de proveer información especializada y actualizada sobre relaciones laborales y transformaciones en el mundo del trabajo, permitiendo apoyar la elaboración de políticas públicas y entregar información confiable sobre los ámbitos vinculados al mundo del trabajo. A su vez, el Instituto Nacional de Estadísticas (INE), quién es el encargado de difundir información oficial, confiable y accesible garante de la calidad de sus estadísticas, ha participado en el diseño y ejecución de distintas actividades propias del desarrollo de la encuesta para versiones anteriores.

Así entonces, para la versión de la Encla 2023, el Departamento de Estudios encarga la realización de la encuesta al INE para que desarrolle diferentes actividades relativas al diseño metodológico, recopilación y procesamiento de la información para dar respuesta a los objetivos planteados, considerando las recomendaciones internacionales sobre las normas de calidad y las sugerencias propuestas realizadas por el Departamento de Estudios, en forma periódica y a través de las mesas de trabajo. Dentro de las actividades definidas para la ejecución del proyecto, se encuentra la elaboración del diseño muestral del cual se encarga el Subdepartamento de Diseños de Marcos y Muestras (SDMM) dependiente del INE, adoptando los requerimientos del estudio y definiendo los tamaños muestrales necesarios para entregar representatividad y confiabilidad de la estimación de interés dentro de los umbrales de errores muestrales aceptables por la institución. Las actividades consideradas para la elaboración del diseño corresponden a la conformación del marco muestral, definir la estrategia de muestreo, obtener los tamaños de muestra, desarrollar el cálculo de los factores de expansión, y estimar las principales variables y sus medidas de precisión.

En esta versión de la encuesta se incorporan cambios en el diseño del marco muestral, utilizando nuevas fuentes de información para la construcción del mismo como parte de las mejoras de completitud de la información. Así entonces, además de los insumos base, tales como el Marco Maestro de Empresas (MME) y el Formulario de trabajadores del Servicio de Impuestos Internos (SII) (F1887), se introduce el uso de registros administrativos provenientes de la información de la Superintendencia de Seguridad Social (SUSESO), registro que permite obtener la variable de diseño, la cual corresponde al número de trabajadores. En cuanto a los factores de expansión, se introduce la metodología de calibración que permite reconstruir los totales poblacionales a partir de un marco auxiliar elaborado de acuerdo a la información proveniente de SUSESO.

Respecto al cálculo del tamaño muestral objetivo y la potencial pérdida de unidades durante el trabajo de campo, al igual que en la Encla 2019, se establece como metodología la estrategia de sobremuestreo, que permite compensar las incidencias detectadas sobre las unidades en el proceso de recolección. Ahora bien, dicha estrategia considera como base la pérdida de unidades ocurrida en la versión inmediatamente anterior de la presente encuesta, lo que para este caso debiese ser Encla 2019. Sin embargo, dada las consecuencias de la contingencia de salud que vivió el país posterior a este periodo, se decide incorporar como complemento la información de logro proveniente de la XII Encuesta Nacional de Innovación (ENI), puesto que esta encuesta es la primera cuyo trabajo de campo se realiza en el año 2021, en un contexto nacional que difiere de lo observado en el año 2019.

Por último, para los dominios de estudio de la Encla 2023, particularmente para la agrupación regional, el Departamento de Estudios de la Dirección del Trabajo solicita un nuevo requerimiento para su conformación, estableciendo nueve agrupaciones regionales para esta versión en comparación con la Encla 2019, en la cual el muestreo se diseñó a nivel de seis agrupaciones regionales. Lo anterior se debe a la necesidad de contar con estimaciones confiables para el cruce de región y tamaño de empresa. Adicionalmente, para esta versión se incorpora la representatividad a nivel regional.

El desarrollo del diseño metodológico de la muestra para Encla 2023, se presenta en el siguiente informe, que se estructura de la siguiente forma: en primer lugar, se describen los objetivos de la encuesta; en el segundo capítulo se detallan las características del diseño muestral, tales como: población objetivo, dominios de estudios, cobertura temática y geográfica, marco muestral y unidades estadísticas, entre otros. En el tercer capítulo se describe la estrategia de muestreo y la metodología de cálculo del tamaño muestral. Posteriormente, en el cuarto capítulo se presentan los resultados del tamaño muestral. En el quinto capítulo se expone el desarrollo del factor de expansión. Posteriormente, en los capítulos seis y siete, se presenta el estimador y sus medidas de precisión, así como también sus resultados finales. Por último, se incorporan las conclusiones, anexo y bibliografía respectiva.

I. OBJETIVO

I.1. Objetivo general

El objetivo de la Encla 2023 es diagnosticar el estado y evolución de las condiciones de empleo y trabajo, de las relaciones laborales y la igualdad de género en las empresas que cuentan con 5 o más trabajadores, regidas por el Código del Trabajo en Chile.

I.2. Objetivos específicos

1. Caracterizar las empresas en relación con su entorno económico.
2. Caracterizar las condiciones de empleo y trabajo en las empresas.
3. Analizar las relaciones laborales que se dan al interior de las empresas.
4. Analizar la relación entre condiciones de empleo y trabajo, y las relaciones laborales.
5. Identificar la situación de las empresas en materia de igualdad de género.
6. Detectar y describir fenómenos laborales emergentes.

II. CARACTERISTICAS DEL DISEÑO MUESTRAL

II.1. Población objetivo

Empresas formales[1] y vigentes que desarrollan su actividad económica principal en algunos de los sectores en estudio, según el Clasificador Internacional Industrial Uniforme de todas las actividades económicas, revisión 4 adaptación nacional (CIIU4.CL 2012), dentro de los límites geográficos del país, excluyendo algunas comunas definidas como difícil acceso[2]; que declararon en los registros administrativos del Formulario 1887[3] del SII y SUSESO en el mes de noviembre del año comercial 2020-2021; que cuentan con cinco o más trabajadores contratados directamente (excluyendo honorarios); y cuyas relaciones laborales se rigen por el Código del Trabajo.

[1] Empresa formal es aquella con iniciación de actividades en el Servicio de Impuestos Internos (SII) vigente al año 2021, que posee RUT y contabilidad propia, y declara IVA.
[2] Para mayor detalle de las comunas excluidas ver Cuadro III.3.
[3] Contiene declaraciones sobre rentas del Art. 42 N°1 (sueldos) otros componentes de la remuneración y retención del impuesto único de segunda categoría de la ley de Rentas.

II.2. Dominios de estudio

Los dominios de estudio en los cuales la muestra tiene representatividad para la estimación del total de trabajadores, asegurando un error muestral esperado que no supere el 15% en estos niveles establecidos para el estudio, corresponden a:

- Nacional,
- Región (16 regiones),
- Agrupación Actividad Económica (13 agrupaciones),
- Tamaño de empresa (Grande, Mediana, Pequeña, Micro), y
- Agrupación regional y tamaño de empresas (9 agrupaciones regionales)

No obstante, los niveles de estudio para publicación quedarán sujetos a la precisión y normas de calidad de las estimaciones a partir de los resultados obtenidos en el trabajo de campo.

II.3. Período de referencia y periodicidad

El período de referencia de la información para la Encla 2023 corresponderá al período 2021-2022. Respecto a la periodicidad de la encuesta, habitualmente se realiza cada dos o tres años. Sin embargo, producto de la contingencia de salud a nivel mundial (COVID-19), el periodo se extiende para esta versión.

II.4. Marco Muestral del producto

II.4.1. Cobertura temática

La Encla 2023 cubre 18 sectores económicos desde el ámbito silvoagropecuario hasta servicios, que se encuentran clasificados según el Clasificador Industrial Internacional Uniforme de todas las Actividades Económicas, adaptación nacional (CII4.CL 2012). Dicho clasificador permite la comparabilidad de la información a nivel de actividad económica, de acuerdo con estándares internacionales.

Tras la definición de los sectores económicos, algunos fueron agrupados, conformando 13 categorías de actividades económicas, las cuales se mantienen desde la Encla 2019 y se detallan en el siguiente cuadro.

Cuadro II.1 Secciones económicas en estudio para la Encla 2023.

Agrupación de Sección Económica	Glosa de la Sección Económica
A	Agricultura, ganadería, silvicultura y pesca
B	Explotación de minas y canteras
C	Industrias manufactureras
D-E	Suministro de electricidad, gas, vapor y aire acondicionado y Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación
F	Construcción
G	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas
H-J	Transporte y almacenamiento e Información y comunicaciones
I	Actividades de alojamiento y de servicio de comidas
K-L	Actividades financieras y de seguros y Actividades inmobiliarias
M-N	Actividades profesionales, científicas y técnicas y Actividades de servicios administrativos y de apoyo
P	Enseñanza
Q	Actividades de atención de la salud humana y de asistencia social
R-S	Actividades artísticas, de entretenimiento y recreativas y Otras actividades de servicios

Elaborado por el Instituto Nacional de Estadísticas.

Como casos particulares, se excluyen del estudio las cooperativas de agua, así como también, las instituciones públicas que, clasificadas en otra actividad económica, pertenecen a municipalidades y corporaciones municipales u otro organismo estatal que se consideran fuera del alcance del estudio, según el catastro de instituciones públicas que dispone el INE.

Para las actividades de Enseñanza y de atención de la salud humana y asistencia social, P y Q, respectivamente, se aplican otras exclusiones complementarias al catastro que dispone el INE. En primer lugar, las subclases 85311, 86101, 86201 se excluyen directamente, mientras que, para las subclases 85001 y 85420 la exclusión consiste en identificar, a nivel de subclase[4], unidades que declaren pertenecer a sociedades públicas asociadas a las municipales, identificando las corporaciones municipales. Por último, también se excluyen las secciones que O, T y U, tal como se presenta en el cuadro II.2.

[4] Clasificación económica a nivel de cinco dígitos de las unidades en el Marco Maestro de Empresas, correspondiente a la adaptación nacional del clasificador CIIU. Rev.4.

Cuadro II.2 Secciones económicas excluidas del ámbito de estudio para la Encla 2023.

Sección Económica	Glosa de la Sección Económica	Subclase CIIU Rev.4.CL	Glosa de la Subclase
P	Enseñanza	85001	Enseñanza preescolar, primaria, secundaria científico humanista y técnico profesional pública
		85311	Enseñanza superior en universidades públicas
		85420	Enseñanza cultural
Q	Actividades de atención de la salud humana y de asistencia social	86101	Actividades de hospitales públicos
		86201	Actividades de centros de salud municipalizados (servicios de salud pública)
O	Administración pública y defensa; planes de seguridad social de afiliación obligatoria		
T	Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio		
U	Actividades de organizaciones y órganos extraterritoriales		

Elaborado por el Instituto Nacional de Estadísticas.

II.4.2. Cobertura geográfica

La cobertura geográfica de la Encla 2023 comprende a todas las empresas formales que realicen su actividad económica dentro del territorio nacional, es decir, en algunas de las 16 regiones del país, excluyendo de este estudio a 22 comunas que son definidas como de difícil acceso, o que presentan un alto costo de levantamiento. Las comunas excluidas del estudio de la Encla 2023 se describen en el Cuadro II.3.

Cuadro II.3 Comunas excluidas del ámbito de estudio de la Encla 2023.

Región	Provincia	Comunas
Arica y Parinacota	Parinacota	General Lagos
Tarapacá	Tamarugal	Colchane
Antofagasta	El Loa	Ollagüe
Valparaíso	Isla De Pascua	Isla de Pascua
	Valparaíso	Juan Fernández
Los Lagos	Llanquihue	Cochamó

Región	Provincia	Comunas
	Palena	Chaitén - Futaleufú - Hualaihué - Palena
Aysén del General Carlos Ibáñez del Campo	Aisén	Guaitecas
	Coihayque	Lago Verde
	Capitán Prat	Villa O'Higgins - Tortel
Magallanes y de la Antártica Chilena	Antártica Chilena	Antártica - Cabo de hornos
	Magallanes	Laguna Blanca - Río Verde - San Gregorio
	Tierra del Fuego	Timaukel - Primavera
	Última Esperanza	Torres del Paine

Elaborado por el Instituto Nacional de Estadísticas.

II.4.3. Construcción del marco muestral

A diferencia de las versiones anteriores, el marco muestral se conforma a partir de tres fuentes de información: el Marco Maestro de Empresas INE (MME), año contable 2021; el directorio de trabajadores asalariados obtenido desde Formulario N°1887 (F1887), año comercial 2020; y los registros administrativos de la Superintendencia de Seguridad Social (SUSESO), año 2021. El detalle de las fuentes de información utilizadas y la depuración que se realiza para la construcción del marco muestral se detalla a continuación:

II.4.3.1. Marco Maestro de Empresas (MME) año 2021

Se utiliza el insumo más actualizado del MME, que corresponde al año contable 2021. A partir de éste, se seleccionan aquellas empresas que ejercen su actividad económica principal en alguna de las mencionadas en el Cuadro II.1.

II.4.3.2. Formulario 1887 referencia noviembre 2020

Desde el directorio proveniente del F1887 se obtiene el total de trabajadores asalariados en el mes de noviembre del año 2020, siendo este el año más actualizado que posee el INE al momento de la construcción del marco muestral para el estudio. De este insumo, se incluyen las empresas que reportan 5 o más trabajadores en el mes de noviembre de 2020.

II.4.3.3. Registros administrativos de la SUSESO referencia noviembre 2021

Es un registro administrativo que proviene de SUSESO y que el INE recibe periódicamente. Este proporciona mensualmente datos nominados, entre los cuales se puede obtener el número de trabajadores que reporta cada empresa. Al igual que para el F1887, se incluyen las empresas que reportan 5 trabajadores o más en el mes de noviembre de 2021.

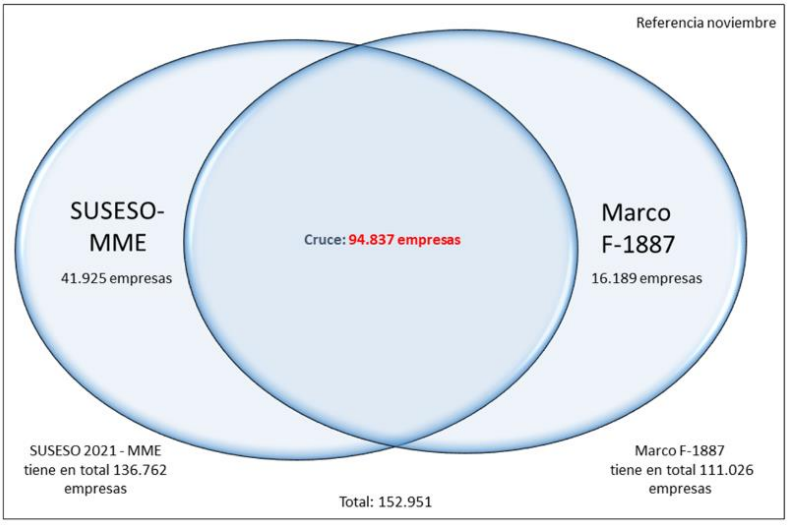
II.4.3.4. II.4.3.4 Depuración del marco muestral

Paso 1: Se cruza el MME con F1887, obteniendo la actividad económica y región desde el MME y el número de trabajadores del F1887. A partir de la intersección de ambos directorios, y al excluir las empresas con 4 o menos trabajadores y las actividades que se encuentran fuera de estudio, quedan 111.026 empresas.

Por otra parte, se cruza el MME con SUSESO y se realizan las mismas exclusiones mencionadas anteriormente, es decir, se excluyen las actividades fuera del estudio y las empresas con 4 o menos trabajadores. Como resultado de este cruce se obtiene un total 136.762 empresas.

Paso 2: Se realiza el cruce con las dos bases de datos obtenidas en el paso 1. Para continuar con la construcción del marco muestral, se analiza tanto el resultado de la intercepción como del complemento del cruce. La figura II.1 ilustra el resultado obtenido.

Figura II.1. Diagrama del cruce entre el marco muestral F1887 y SUSESO

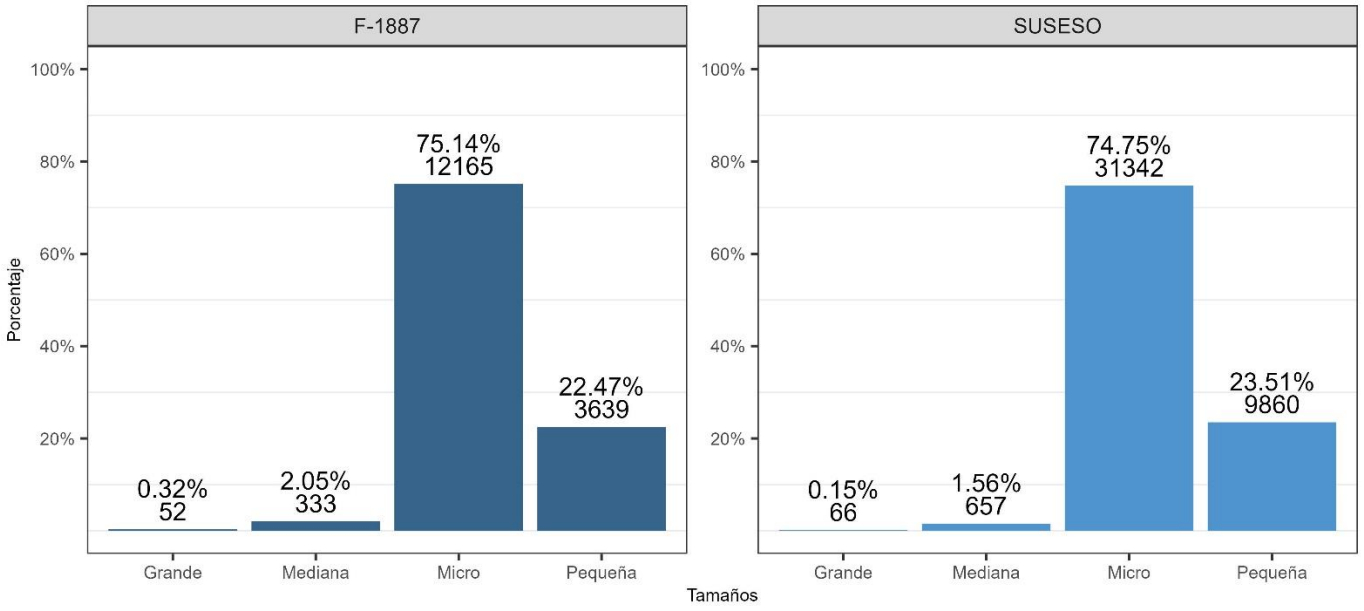


Elaborado por el Instituto Nacional de Estadísticas.

Tal como se observa, existen 94.837 empresas que están en ambas fuentes de información (F1887 y SUSESO), 41.925 empresas que solo están en SUSESO y 16.189 que solo están en el F1887. La suma de todas las partes da un total de 152.951 empresas.

Además, con el fin de profundizar en el análisis se verifica la distribución del tamaño de las empresas que están solo en F1887 o solo en SUSESO. Tal como se muestra en la Figura II.2, la mayoría de las empresas que están en una u otra fuente son de tamaño micro, seguidas por las de tamaño pequeña.

Figura II.2. Distribución de las empresas que no coinciden en ambas fuentes de información, según tamaño de la empresa



Elaborado por el Instituto Nacional de Estadísticas.

Considerando que el F1887 tiene como referencia el 2020, es decir, cuenta con un año de desfase con respecto de las demás fuentes utilizadas, siendo además un año atípico debido la pandemia de COVID-19, es muy probable que estas empresas ya no existan en el país. Por lo tanto, junto a la DT, se decide excluir del marco muestral a las empresas pequeñas y micro que solo se encuentran en este insumo.

Paso 3: Finalmente, se realiza una depuración de empresas a través de la información más reciente del levantamiento de encuestas realizadas por el INE, presentadas en la figura II.3. Para esta depuración se excluyen del marco muestral todas las unidades que presentan estados de levantamiento *sin movimiento y término de giro* en dichas encuestas.

Figura II.3. Fuentes de información utilizadas para depurar el marco muestral



Elaborado por el Instituto Nacional de Estadísticas.

Después de la depuración, el marco muestral queda conformado por 134.035 empresas.

Paso 4: Finalmente, una vez obtenido el listado final de unidades que pertenecen al marco muestral, se asigna el tamaño de empresa a cada unidad, según los siguientes criterios:

Cuadro II.4. Criterios de priorización para el tamaño de empresas en el marco muestral de la Encla 2023

Fuente de información en la que está la empresa	Información que se mantiene en el marco muestral
Sólo F1887 2020	F1887
Solo SUSESO 2021	SUSESO 2021
En ambas (F1887 2020 – SUSESO 2021)	SUSESO 2021

Elaborado por el Instituto Nacional de Estadísticas.

Este criterio se define con la finalidad de priorizar la información más actualizada disponible de la variable número de trabajadores.

II.4.4. Estratificación del marco muestral

Una vez construido el marco muestral, se realiza la estratificación. Esta permite reunir elementos de una población en grupos homogéneos previo a la selección, con el propósito de mejorar las medidas de precisión de las estimaciones agrupando las unidades del marco muestral en clases homogéneas en su interior, y que difieran de las características del resto. Los estratos deben ser mutuamente excluyentes y exhaustivos colectivamente, es decir, cada elemento en la población debe ser asignado a un solo estrato y ninguno puede quedar excluido.

Los estratos que se conforman en el marco muestral corresponden al cruce entre región, agrupación de actividades económicas de acuerdo con la CIIU4.CL 2012, tamaño de empresa según número de trabajadores y tramos de inclusión forzosa (IF) y aleatoria (IA).

En la siguiente subsección se describen las variables consideradas para conformar los estratos en el marco muestral de la Encla 2023.

II.4.4.1. Región y agrupación regional

En el marco muestral se consideran las 16 regiones del país, las cuales corresponden a un nivel de estimación. Por otro lado, se agrupa las regiones de acuerdo con sus características geográficas, conformando 9 agrupaciones regionales solicitadas por la Dirección del Trabajo (DT) [5]. En el Cuadro III.5 se muestra la nueva agrupación a partir de las 16 regiones.

[5] La solicitud de la Dirección del Trabajo de pasar de 6 a 9 agrupaciones regionales permite asegurar representatividad para una desagregación mayor a nivel regional considerando el cruce con tamaño de empresas.

Cuadro II.5. Agrupación regional a partir de las regiones, solicitadas por la DT.

Agrupación Regional	Nombre de región
Norte Grande	Arica y Parinacota
	Tarapacá
Antofagasta	Antofagasta
Norte Chico	Atacama
	Coquimbo
Valparaíso	Valparaíso
Metropolitana	Metropolitana de Santiago
Centro Sur	Libertador General Bernardo O´Higgins
	Maule
	Ñuble
Bio-Bío	Bio-Bío
Sur	Araucanía
	Los Ríos
	Los Lagos
Zona Austral	Aysén del General Carlos Ibáñez del Campo
	Magallanes y Antártica Chilena

Elaborado por el Instituto Nacional de Estadísticas

II.4.4.2. Agrupación de actividades económicas

Las actividades económicas especificadas en la cobertura temática (ver Cuadro II.1), se agrupan conformando las 13 agrupaciones de actividad económicas en estudios, que forman una variable de estratificación en el marco muestral.

II.4.4.3. Tamaño de empresas

La clasificación de las empresas según tamaños para la Encla 2023 establece cuatro categorías: grande, mediana, pequeña y micro, definidas según el total del número de trabajadores asalariados contratados directamente por la empresa en el periodo noviembre 2021. En el Cuadro II.6 se observan los límites que definen los tramos de número de trabajadores.

Cuadro II.6. Tamaño de empresas según número de trabajadores asalariados

Tamaño Empresa	Número de Trabajadores Asalariados	
	Límite inferior	Límite Superior
Grande	200	Más
Mediana	50	199
Pequeña	10	49
Micro	5	9

Elaborado por el Instituto Nacional de Estadísticas

II.4.4.4. Tramo de inclusión

El tramo de inclusión permite diferenciar las unidades que se incluyen forzosamente (IF) de aquellas cuya inclusión es aleatoria (IA). Para la Encla 2023, a diferencia de las versiones anteriores de la encuesta, las unidades IF son escogidas por reportar un total de trabajadores asalariados mayor o igual a 2.000 en el período noviembre 2021; siendo consideradas como unidades autorepresentadas, ya que presentan un alto aporte en el sector económico al cual pertenecen. Esto motiva la conveniencia de crear estratos de inclusión forzosa (IF) dentro de estas actividades, con la intención de mejorar la precisión de las estimaciones de interés.

Respecto al tramo de inclusión aleatoria (IA), este se conforma con el resto de las unidades económicas en la población, con las cuales se estima un tamaño muestral a través de un diseño probabilístico.

Figura II.4 Estratificación del marco muestral para la Encla 2023.



Elaborado por el Instituto Nacional de Estadísticas.

II.5. Unidades estadísticas

II.5.1. Unidad de muestreo

La unidad de muestreo es la empresa formal, que cuenta con iniciación de actividades en el SII; que desarrolla su actividad económica en el ámbito de estudio y dentro de los límites territoriales del país; y que, durante el mes de noviembre 2021, haya declarado tener 5 o más trabajadores asalariados contratados directamente.

II.5.2. Unidad de información y de análisis

La unidad de información y de análisis es la empresa formal que cumple los requisitos para formar parte de la población objetivo.

III. METODOLOGÍA DE CÁLCULO DEL TAMAÑO MUESTRAL

III.1. Estrategia muestral

El diseño aplicado a la Encla 2023 corresponde a un muestreo probabilístico y estratificado, donde los estratos se encuentran definidos según el cruce de las variables región, agrupación de actividades económicas de acuerdo con la CIIU4.CL 2012, tamaño de empresa según número de trabajadores y tramos de inclusión forzosa (IF) y aleatoria (IA).

III.1.1. Consideraciones para el tamaño muestral

En el siguiente cuadro se detallan los elementos considerados para calcular el tamaño de la muestra correspondiente a la Encuesta Laboral, en su versión 2023:

Cuadro III.1. Elementos utilizados para el cálculo del tamaño muestral de Encla 2023

Parámetro	Descripción
Variable de diseño	Número de trabajadores: y
Estimador asociado	Estimador del total del número de trabajadores: $\hat{Y}$
Parámetro asociado	Total del número de trabajadores: Y
Niveles de estimación para los cuales se establece el error de muestreo	- Nacional: 5%. - Tamaño de empresas: 6% - Región: 6% - Agrupación de actividad económica: 11% - Cruce de agrupación regional y tamaño de empresas: 15%.
Nivel de confianza	95%

Elaborado por el Instituto Nacional de Estadísticas

III.2. Cálculo del Tamaño muestral

El diseño muestral de la Encla 2023 es probabilístico y estratificado. Los estratos están conformados, entre otras variables, por el tramo de inclusión forzosa e inclusión aleatoria, donde la primera inclusión corresponde al grupo unidades autorepresentadas, es decir, que sólo se representan a sí mismas, y su método de selección se detalla en el apartado sobre la estratificación del marco muestral. En cambio, para las unidades que pertenecen al tramo inclusión aleatoria, se calcula un tamaño muestral por medio del método que se describe a continuación.

Paso 1: Desde la Encla 2019 se obtuvieron los tamaños de la muestra iniciales calculados por estrato. Este diseño estratificado considera como variable de interés el número de trabajadores, fijando errores relativos del 15% a nivel del estrato, con un coeficiente de confianza del 95%.

$$n_{rhv} = \frac{\left(\sum_{i=1}^{I} N_{rhv_i} \cdot \sigma_{rhv_i}\right)^2 \cdot \left(Z_{1-\frac{\alpha}{2}}\right)^2}{e_r \cdot \sum_{i=1}^{I} Y_i + \left(\sum_{r=1}^{R} N_{rhv_i} \cdot \sigma_{rhv_i}^2\right) \cdot \left(Z_{1-\frac{\alpha}{2}}\right)^2} \quad (1)$$

Donde:

n_{rhv} : Tamaño de la muestra de empresas a nivel de cada estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

N_{rhv_i} : Número total de empresas i en el marco muestral para el estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

σ_{rhv_i} : Desviación estándar del número de trabajadores de las empresas i en el estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

e_r : Error relativo máximo admisible de 15%.

Y_i : Total del número de trabajadores en el marco muestral para las empresas i en el estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

$Z_{1-\frac{\alpha}{2}}$: Estadístico normal estándar correspondiente a un nivel de confianza del 95% ($Z_{0,975}$=1,96≈2).

Paso 2: El tamaño de muestra objetivo obtenido en cada estrato de región o agrupación regional, agrupación de actividad y tamaño de empresas, se ajusta siguiendo las siguientes reglas:

- Si el tamaño de muestra objetivo (n_{rhv}) calculado en el paso 1 es mayor al número de empresas disponibles en el marco muestral (N_{rhv}) para el estrato, se incluyen todas las empresas del marco muestral en la muestra.

- Si el tamaño de muestra objetivo (n_{hvm}) calculado en el paso 1 es menor al número de empresas disponible en el marco muestral (N_{rhv}) para el estrato, se mantiene el tamaño de la muestral calculada.

Paso 3: Luego para obtener el tamaño de la muestra objetivo a nivel de región o agrupación regional r, agrupación de actividad h y tamaño de empresas v, respectivamente, se suman los tamaños muestrales obtenidos para los estratos a nivel más desagregado.

- Tamaño muestral a nivel de tamaño de empresas:

$$n_v = \sum_{r=1}^{R} \sum_{h=1}^{H} n_{rhv} \quad (2)$$

Donde:

n_v : Tamaño de la muestra de empresas a nivel de tamaño de empresas v.

n_{rhv} : Tamaño de la muestra a nivel de cada estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

- Tamaño muestral a nivel de agrupación de actividad económica:

$$n_h = \sum_{r=1}^{R} \sum_{v=1}^{V} n_{rhv} \quad (3)$$

Donde:

n_h : Tamaño de la muestra de empresas a nivel de agrupación de actividad h.

n_{rhv} : Tamaño de la muestra a nivel de cada estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

- Tamaño muestral a nivel de región o agrupación regional:

$$n_r = \sum_{h=1}^{H} \sum_{v=1}^{V} n_{rhv} \quad (4)$$

Donde:

n_r : Tamaño de la muestra de empresas a nivel de región o agrupación regional r.

n_{rhv} : Tamaño de la muestra a nivel de cada estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

Paso 4. Con la información de la variable número de trabajadores al mes de referencia noviembre 2021, contenida en el marco muestral, se calculan las varianzas, las sumas totales, coeficientes de variación y errores relativos esperados.

Paso 5. Luego de calculados los errores esperados para los niveles de estimación requeridos, se observan dichos errores y en caso de superar el umbral de error establecido, se debe aumentar el tamaño de la muestra en los estratos correspondientes, por lo tanto, el tamaño de la muestra ajustado en el estrato se denomina n'_{rhv}.

Paso 6. Finalmente, se calcula la sobremuestra a partir de la tasa de logro observada en la Encla 2019 y la Encuesta Nacional de Innovación año de referencia 2019-2020 (ENI 2019-2020). Se decide utilizar ambas tasas debido a los cambios observados en el operativo de campo, luego de la contingencia de salud nacional. Considerando que la recolección de la Encla 2019 se realizó durante el año 2019, bajo un escenario de trabajo de campo estable, las tasas de logro obtenidas fueron muy cercanas al 100%. Sin embargo, producto de la emergencia sanitaria ocurrida en el año 2020, se evidencio una disminución considerable en las tasas de logro en diferentes encuestas, debido a la dificultad de encontrar a las unidades en el trabajo de campo. Bajo el contexto descrito anteriormente, se incluye la tasa de la ENI 2019-2020 como fuente primaria, asumiendo un escenario más realista de la situación actual en el país puesto que su recolección se realizó en el año 2021 y se complementa con la tasa de logro de la Encla 2019, la cual corresponde a la fuente secundaria.

Los criterios aplicados para el uso de las tasas de logro en los tamaños muestrales se detallan a continuación:

Paso 1: Para los estratos existentes en ambas encuestas (Encla-ENI), se utiliza la tasa de logro de la ENI 2019-2020 y la Encla 2019, de acuerdo con los siguientes criterios:

- Se incorpora como tasa primaria la obtenida en la ENI 2019-2020.
- En caso de que la tasa de la ENI sea igual a cero o mayor o igual a 1, se utiliza la tasa de la Encla 2019.

Paso 2: Para los estratos en que no se cuente con información de las tasas, se utiliza el 30% correspondiente al valor utilizado en reemplazos para versiones anteriores.

Paso 3: En cuanto a la actividad económica Enseñanza (P) se utiliza la tasa de logro de Encla 2019 y en caso de no contar con dicha información se utiliza el 30%.

Así entonces, el tamaño muestral objetivo con la sobremuestra se calcula de la siguiente forma:

$$n_{rhv}^{s} = \frac{n'_{rhv}}{tl_{rhv}} \quad (5)$$

Donde:

n_{rhv}^{s} : Tamaño de muestra con sobremuestreo ajustado según la tasa de logro de la encuesta primaria o secundaria por región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

n'_{rhv} : Tamaño de muestra ajustado para el estrato según región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

tl_{rhv} : Tasa de logro según ajuste aplicado, en la región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

Finalmente, los tamaños muestrales objetivos con la sobremuestra a nivel más agregado quedan conformados de la siguiente forma:

- Tamaño muestral a nivel de tamaño de empresas:

$$n_{v}^{s} = \sum_{r=1}^{R} \sum_{h=1}^{H} n_{rhv}^{s} \quad (6)$$

Donde:

n_{v}^{s} : Tamaño de la muestra de empresas con sobremuestreo a nivel de tamaño de empresas v.

n_{rhv}^{s} : Tamaño de la muestra con sobremuestra a nivel de cada estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

- Tamaño muestral a nivel de agrupación de actividad económica:

$$n_{h}^{s} = \sum_{r=1}^{R} \sum_{v=1}^{V} n_{rhv}^{s} \quad (7)$$

Donde:

n_{h}^{s} : Tamaño de la muestra de empresas con sobremuestreo a nivel de agrupación de actividad h.

n_{rhv}^{s} : Tamaño de la muestra con sobremuestra a nivel de cada estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

- Tamaño muestral a nivel de región o agrupación regional:

$$n_r^S = \sum_{h=1}^{H} \sum_{v=1}^{V} n_{rhv}^S \quad (8)$$

Donde:

n_r^S : Tamaño de la muestra de empresas con sobremuestreo a nivel de región o agrupación regional r.

n_{rhv}^S : Tamaño de la muestra con sobremuestra a nivel de cada estrato, región o agrupación regional r, agrupación de actividad h y tamaño de empresas v.

Por último, el incremento de la sobremuestra en forma considerable abrió paso a nuevos ejercicios de sobremuestra, los cuales consistieron en rotar las tasas utilizadas previamente: así entonces, se utilizó como fuente primaria la Encla 2019, y como secundaria la XII ENI. Estos ajustes se realizan en las agrupaciones de actividad económica donde la sobremuestra dobla los tamaños muestrales objetivos, con la meta de reducir el total de unidades a gestionar en el trabajo de campo, para ajustarse de mejor forma al presupuesto disponible. Lo anterior, considerando que una parte del proceso de recolección de la Encuesta Laboral se realiza en forma presencial, aumentando la posibilidad de contactar a la empresa, por lo que el cambio de tasas no debiese generar una disminución del logro.

III.3. Selección de unidades muestrales

Posteriormente la selección de unidades se realiza para cada tramo de inclusión en forma independiente: por un lado, las unidades IF quedaron previamente seleccionadas por el umbral de un número de trabajadores mayor o igual a 2.000; por otro, la selección de las unidades de IA se realiza a través del método aleatorio definido con igual probabilidad a nivel del estrato. Al ser una estrategia de muestreo sistemático, la probabilidad de selección de las unidades de cada estrato está definida por la primera unidad en ser seleccionada.

Al igual que el muestreo aleatorio, el muestreo sistemático es un muestreo probabilístico donde cada unidad del marco tiene la misma probabilidad de ser elegida. Al contar con un marco muestral estratificado a los niveles de región, actividad económica y tamaño, el muestreo sistemático permite garantizar la cobertura de unidades en todos los estratos al realizar un recorrido uniforme completo al interior de cada uno.

Las empresas del tramo IA son seleccionadas aleatoriamente al interior de cada estrato dado por el cruce de región, tamaño de la empresa según número de trabajadores y agrupación de actividad económica. El

método aplicado corresponde a un muestreo sistemático aleatorio, a partir de la implementación de los siguientes pasos:

1. Para asegurar un recorrido completo dentro de cada estrato, se ordenan las unidades muestrales en forma ascendente por región, tamaño de empresas y agrupación de actividades económicas, y descendente según la variable número de trabajadores.

2. A través del Software SAS se utiliza el procedimiento surveyselect para realizar muestreos estadísticos complejos, especificando el plan el tamaño muestral por estrato (incluido el sobremuestreo) requerido para la selección.

4. Se especifica el método de muestreo, que en este caso corresponde a el método de muestreo sistemático (method=sys).

5. Se indica que se utilizará la función de generación de números aleatorios de distribución uniforme para seleccionar la muestra (ranuni) . Además, se especifica la semilla inicial para la generación de números aleatorios (seed=2365955).

6. Se solicita la generación de estadísticas resumidas para el proceso de muestreo (stats) y se definen las variables de estratificación que se utilizaran (strata).

7. Se filtran los datos según tramo de inclusión aleatoria (where tramo_final=2).

8. Se guardan las estadísticas resumidas del muestreo en un conjunto de datos llamado tabla resumen (Ods output summary= tabla_resumen)

9. Una vez extraída la selección aleatoria, se genera un archivo que contenga las unidades seleccionadas en forma aleatoria y sus variables de estratificación. Además, se incorporan las unidades IF que fueron seleccionadas previamente.

IV. RESULTADOS DEL TAMAÑO MUESTRAL

En las tablas siguientes se presenta el número de empresas en el marco muestral, el tamaño de muestra objetivo con su respectiva sobremuestra, y la muestra efectivamente lograda en la recolección. Se determina un tamaño de muestra objetivo de 4.267 empresas, de las cuales 202 son del estrato de inclusión forzosa (IF), y 4.065 del estrato de selección aleatoria (IA). Luego, tras el trabajo de campo, se logra una muestra de 173 empresas IF y 4.323 unidades IA, sumando una muestra efectiva de 4.496 unidades.

Tabla IV.1. Tamaño muestral objetivo y efectivo según tamaños de empresas.

Tamaños	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
		Total	IF	IA			Total	IF	IA
Total	**134.035**	**4.267**	**202**	**4.065**	**4.085**	**8.352**	**4.496**	**173**	**4.323**
Grande	3.234	1.028	202	826	289	1.317	981	171	810
Mediana	10.742	972	-	972	811	1.783	1.087	2	1.085
Pequeña	56.483	1.616	-	1.616	2.095	3.711	1.809	-	1.809
Micro	63.576	651	-	651	890	1.541	619	-	619

Elaborado por el Instituto Nacional de Estadísticas.

Tabla IV.2. Tamaño muestral objetivo y efectivo según región.

Región	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
		Total	IF	IA			Total	IF	IA
Total	**134.035**	**4.267**	**202**	**4.065**	**4.085**	**8.352**	**4.496**	**173**	**4.323**
Arica y Parinacota	1.319	185	-	185	133	318	189	-	189
Tarapacá	2.504	217	-	217	194	411	257	-	257
Antofagasta	3.882	221	1	220	192	413	212	1	211
Atacama	1.611	212	1	211	212	424	193	-	193
Coquimbo	4.477	244	-	244	285	529	277	-	277
Valparaíso	12.481	307	7	300	270	577	325	7	318
Metropolitana	65.853	647	174	473	356	1.003	670	148	522
O'Higgins	6.403	273	7	266	358	631	322	6	316
Maule	7.044	270	2	268	295	565	309	2	307
Ñuble	2.691	267	1	266	337	604	258	1	257
Biobío	9.385	284	5	279	298	582	331	4	327
La Araucanía	5.418	273	2	271	305	578	274	2	272
Los Ríos	2.324	228	-	228	236	464	249	-	249
Los Lagos	6.407	280	1	279	272	552	249	1	248
Aysén	690	168	-	168	195	363	174	-	174
Magallanes	1.546	191	1	190	147	338	207	1	206

Elaborado por el Instituto Nacional de Estadísticas.

Tabla IV.3. Tamaño muestral objetivo y efectivo según agrupación de actividad económica[6].

Actividad económica	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
		Total	IF	IA			Total	IF	IA
Total	**134.035**	**4.267**	**202**	**4.065**	**4.085**	**8.352**	**4.496**	**173**	**4.323**
A	11.595	392	2	390	600	992	385	4	381
B	870	210	5	205	156	366	159	7	152
C	16.055	427	23	404	309	736	456	16	440
D-E	999	218	2	216	190	408	226	2	224
F	16.636	450	20	430	519	969	447	18	429
G	29.855	457	29	428	386	843	498	27	471
H-J	13.383	361	17	344	314	675	381	19	362
I	8.077	285	15	270	262	547	309	13	296
K-L	5.675	233	9	224	221	454	202	9	193
M-N	14.946	439	54	385	363	802	466	38	428
P	5.439	317	10	307	279	596	482	7	475
Q	3.309	239	13	226	214	453	316	11	305
R-S	7.196	239	3	236	272	511	169	2	167

Elaborado por el Instituto Nacional de Estadísticas.

Tabla IV.4. Tamaño muestral objetivo y efectivo según región y tamaño.

Región	Tamaño	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
			Total	IF	IA			Total	IF	IA
Total		**134.035**	**4.267**	**202**	**4.065**	**4.085**	**8.352**	**4.496**	**173**	**4.323**
Arica y Parinacota		**1.319**	**185**	**-**	**185**	**133**	**318**	**189**	**-**	**189**
	Grande	12	12	-	12	-	12	8	-	8
	Mediana	69	41	-	41	14	55	39	-	39
	Pequeña	520	88	-	88	81	169	109	-	109
	Micro	718	44	-	44	38	82	33	-	33
Tarapacá		**2.504**	**217**	**-**	**217**	**194**	**411**	**257**	**-**	**257**
	Grande	26	26	-	26	-	26	21	-	21
	Mediana	162	56	-	56	41	97	65	-	65
	Pequeña	1.028	92	-	92	97	189	122	-	122
	Micro	1.288	43	-	43	56	99	49	-	49
Antofagasta		**3.882**	**221**	**1**	**220**	**192**	**413**	**212**	**1**	**211**
	Grande	56	39	1	38	11	50	40	1	39
	Mediana	325	63	-	63	36	99	53	-	53

[6] Para las actividades económicas Agricultura, ganadería, silvicultura y pesca (A), Explotación de minas y canteras (B) y Transporte y almacenamiento e Información y comunicaciones (H-J) se observa que el número de empresas clasificadas como inclusión forzosa en la muestra efectiva son mayores al número de empresas de la muestra objetivo, lo cual se debe a un cambio de actividad económica de estas unidades.

Región	Tamaño	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
			Total	IF	IA			Total	IF	IA
	Pequeña	1.634	78	-	78	89	167	80	-	80
	Micro	1.867	41	-	41	56	97	39	-	39
Atacama		**1.611**	**212**	**1**	**211**	**212**	**424**	**193**	**-**	**193**
	Grande	37	37	1	36	-	37	24	-	24
	Mediana	118	50	-	50	26	76	50	-	50
	Pequeña	676	82	-	82	120	202	84	-	84
	Micro	780	43	-	43	66	109	35	-	35
Coquimbo		**4.477**	**244**	**-**	**244**	**285**	**529**	**277**	**-**	**277**
	Grande	44	38	-	38	4	42	29	-	29
	Mediana	303	64	-	64	60	124	77	-	77
	Pequeña	1.955	103	-	103	163	266	125	-	125
	Micro	2.175	39	-	39	58	97	46	-	46
Valparaíso		**12.481**	**307**	**7**	**300**	**270**	**577**	**325**	**7**	**318**
	Grande	165	94	7	87	37	131	93	7	86
	Mediana	781	64	-	64	65	129	85	-	85
	Pequeña	5.380	110	-	110	121	231	116	-	116
	Micro	6.155	39	-	39	47	86	31	-	31
Metropolitana		**65.853**	**647**	**174**	**473**	**356**	**1.003**	**670**	**148**	**522**
	Grande	2.259	394	174	220	107	501	390	146	244
	Mediana	5.900	92	-	92	85	177	116	2	114
	Pequeña	27.956	122	-	122	123	245	132	-	132
	Micro	29.738	39	-	39	41	80	32	-	32
O'Higgins		**6.403**	**273**	**7**	**266**	**358**	**631**	**322**	**6**	**316**
	Grande	116	70	7	63	34	104	66	6	60
	Mediana	535	67	-	67	91	158	89	-	89
	Pequeña	2.772	97	-	97	165	262	124	-	124
	Micro	2.980	39	-	39	68	107	43	-	43
Maule		**7.044**	**270**	**2**	**268**	**295**	**565**	**309**	**2**	**307**
	Grande	101	54	2	52	25	79	59	2	57
	Mediana	603	71	-	71	107	178	92	-	92
	Pequeña	2.915	106	-	106	113	219	111	-	111
	Micro	3.425	39	-	39	50	89	47	-	47
Ñuble		**2.691**	**267**	**1**	**266**	**337**	**604**	**258**	**1**	**257**
	Grande	46	37	1	36	3	40	29	1	28
	Mediana	184	62	-	62	55	117	58	-	58
	Pequeña	1.065	129	-	129	211	340	125	-	125
	Micro	1.396	39	-	39	68	107	46	-	46
Biobío		**9.385**	**284**	**5**	**279**	**298**	**582**	**331**	**4**	**327**
	Grande	173	88	5	83	33	121	93	4	89
	Mediana	706	63	-	63	62	125	79	-	79

Región	Tamaño	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
			Total	IF	IA			Total	IF	IA
	Pequeña	3.907	94	-	94	140	234	116	-	116
	Micro	4.599	39	-	39	63	102	43	-	43
La Araucanía		**5.418**	**273**	**2**	**271**	**305**	**578**	**274**	**2**	**272**
	Grande	62	47	2	45	13	60	48	2	46
	Mediana	357	73	-	73	59	132	73	-	73
	Pequeña	2.223	114	-	114	175	289	117	-	117
	Micro	2.776	39	-	39	58	97	36	-	36
Los Ríos		**2.324**	**228**	**-**	**228**	**236**	**464**	**249**	**-**	**249**
	Grande	16	16	-	16	-	16	11	-	11
	Mediana	128	55	-	55	35	90	53	-	53
	Pequeña	903	115	-	115	139	254	138	-	138
	Micro	1.277	42	-	42	62	104	47	-	47
Los Lagos		**6.407**	**280**	**1**	**279**	**272**	**552**	**249**	**1**	**248**
	Grande	103	58	1	57	22	80	51	1	50
	Mediana	438	74	-	74	47	121	76	-	76
	Pequeña	2.615	109	-	109	151	260	101	-	101
	Micro	3.251	39	-	39	52	91	21	-	21
Aysén		**690**	**168**	**-**	**168**	**195**	**363**	**174**	**-**	**174**
	Grande	3	3	-	3	-	3	4	-	4
	Mediana	33	33	-	33	-	33	29	-	29
	Pequeña	291	91	-	91	132	223	105	-	105
	Micro	363	41	-	41	63	104	36	-	36
Magallanes		**1.546**	**191**	**1**	**190**	**147**	**338**	**207**	**1**	**206**
	Grande	15	15	1	14	-	15	15	1	14
	Mediana	100	44	-	44	28	72	53	-	53
	Pequeña	643	86	-	86	75	161	104	-	104
	Micro	788	46	-	46	44	90	35	-	35

Elaborado por el Instituto Nacional de Estadísticas.

Tabla IV.5. Tamaño muestral objetivo y efectivo según agrupación de región y tamaño.

Agrupación regional	Tamaño	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
			Total	IF	IA			Total	IF	IA
Total		**134.035**	**4.267**	**202**	**4.065**	**4.085**	**8.352**	**4.496**	**173**	**4.323**
Norte grande		**3.823**	**402**	**-**	**402**	**327**	**729**	**446**	**-**	**446**
	Grande	38	38	-	38	-	38	29	-	29
	Mediana	231	97	-	97	55	152	104	-	104
	Pequeña	1.548	180	-	180	178	358	231	-	231
	Micro	2.006	87	-	87	94	181	82	-	82

Agrupación regional	Tamaño	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
			Total	IF	IA			Total	IF	IA
Antofagasta		**3.882**	**221**	**1**	**220**	**192**	**413**	**212**	**1**	**211**
	Grande	56	39	1	38	11	50	40	1	39
	Mediana	325	63	-	63	36	99	53	-	53
	Pequeña	1.634	78	-	78	89	167	80	-	80
	Micro	1.867	41	-	41	56	97	39	-	39
Norte chico		**6.088**	**456**	**1**	**455**	**497**	**953**	**470**	**-**	**470**
	Grande	81	75	1	74	4	79	53	-	53
	Mediana	421	114	-	114	86	200	127	-	127
	Pequeña	2.631	185	-	185	283	468	209	-	209
	Micro	2.955	82	-	82	124	206	81	-	81
Valparaíso		**12.481**	**307**	**7**	**300**	**270**	**577**	**325**	**7**	**318**
	Grande	165	94	7	87	37	131	93	7	86
	Mediana	781	64	-	64	65	129	85	-	85
	Pequeña	5.380	110	-	110	121	231	116	-	116
	Micro	6.155	39	-	39	47	86	31	-	31
Metropolitana		**65.853**	**647**	**174**	**473**	**356**	**1.003**	**670**	**148**	**522**
	Grande	2.259	394	174	220	107	501	390	146	244
	Mediana	5.900	92	-	92	85	177	116	2	114
	Pequeña	27.956	122	-	122	123	245	132	-	132
	Micro	29.738	39	-	39	41	80	32	-	32
Centro Sur		**16.138**	**810**	**10**	**800**	**990**	**1.800**	**889**	**9**	**880**
	Grande	263	161	10	151	62	223	154	9	145
	Mediana	1.322	200	-	200	253	453	239	-	239
	Pequeña	6.752	332	-	332	489	821	360	-	360
	Micro	7.801	117	-	117	186	303	136	-	136
Biobío		**9.385**	**284**	**5**	**279**	**298**	**582**	**331**	**4**	**327**
	Grande	173	88	5	83	33	121	93	4	89
	Mediana	706	63	-	63	62	125	79	-	79
	Pequeña	3.907	94	-	94	140	234	116	-	116
	Micro	4.599	39	-	39	63	102	43	-	43
Sur		**14.149**	**781**	**3**	**778**	**813**	**1.594**	**772**	**3**	**769**
	Grande	181	121	3	118	35	156	110	3	107
	Mediana	923	202	-	202	141	343	202	-	202
	Pequeña	5.741	338	-	338	465	803	356	-	356
	Micro	7.304	120	-	120	172	292	104	-	104
Austral		**2.236**	**359**	**1**	**358**	**342**	**701**	**381**	**1**	**380**
	Grande	18	18	1	17	-	18	19	1	18
	Mediana	133	77	-	77	28	105	82	-	82
	Pequeña	934	177	-	177	207	384	209	-	209

Agrupación regional	Tamaño	Marco muestral	Muestra objetivo			Sobremuestreo	Total a gestionar	Muestra efectiva		
			Total	IF	IA			Total	IF	IA
	Micro	1.151	87	-	87	107	194	71	-	71

Elaborado por el Instituto Nacional de Estadísticas.

V. DESARROLLO DEL FACTOR DE EXPANSIÓN

Los factores de expansión, en una muestra probabilística, afectan las magnitudes recabadas en una encuesta, de manera que puedan ser representativas del universo. Este cálculo es, técnicamente, el inverso de la probabilidad de selección de cada unidad, es decir, se obtienen factores de expansión para las unidades incluidas forzosamente y las aleatorias.

Dado que las unidades de inclusión forzosa se auto-representan, se deben realizar ajustes en el factor de expansión de las 8.150 unidades que corresponden a empresas de inclusión aleatoria, que reflejen las incidencias sucedidas en el trabajo de campo que impidieron la recolección total de la muestra seleccionada. Además, para esta versión se incluye la calibración de los factores de expansión para las unidades de inclusión aleatoria, que permite reconstruir los totales poblacionales a partir de un marco auxiliar elaborado de acuerdo a la información proveniente de SUSESO, considerando el total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022 y con las exclusiones consideradas en el punto II.4.3 Construcción del marco muestral.

V.1. Factor de expansión para las empresas del tramo de inclusión aleatoria

V.1.1. Factor de expansión teórico

El factor de expansión teórico corresponde al inverso de la probabilidad de selección de la empresa i perteneciente al estrato región r, tamaño de la empresa v y agrupación de actividad económica h, que se calcula mediante la siguiente fórmula:

$$P_{rvh_i} = \frac{n_{rvh}^{sm}}{N_{rvh}} \quad (9)$$

Donde:

P_{rvh_i} : Probabilidad de selección de la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

n_{rvh}^{sm} : Tamaño de muestra con sobremuestreo ajustado de la Encla 2023 en el estrato región r, tamaño de la empresa v y agrupación de actividad económica h.

N_{rvh} : Número de empresas en el marco muestral de la Encla 2023, en el estrato región r, tamaño de la empresa v y agrupación de actividad económica h.

Por lo tanto, el factor de expansión teórico se obtiene mediante la siguiente fórmula:

$$w_{rvh_i} = \frac{1}{P_{rvh_i}} \quad (10)$$

Donde:

w_{rvh_i} : Factor de expansión teórico de la empresa i en la muestra, perteneciente al estrato región r, tamaño de la empresa v y agrupación de actividad económica h.

P_{rvh_i} : Probabilidad de selección de la empresa i en la muestra, perteneciente al estrato región r, tamaño de la empresa v y agrupación de actividad económica h.

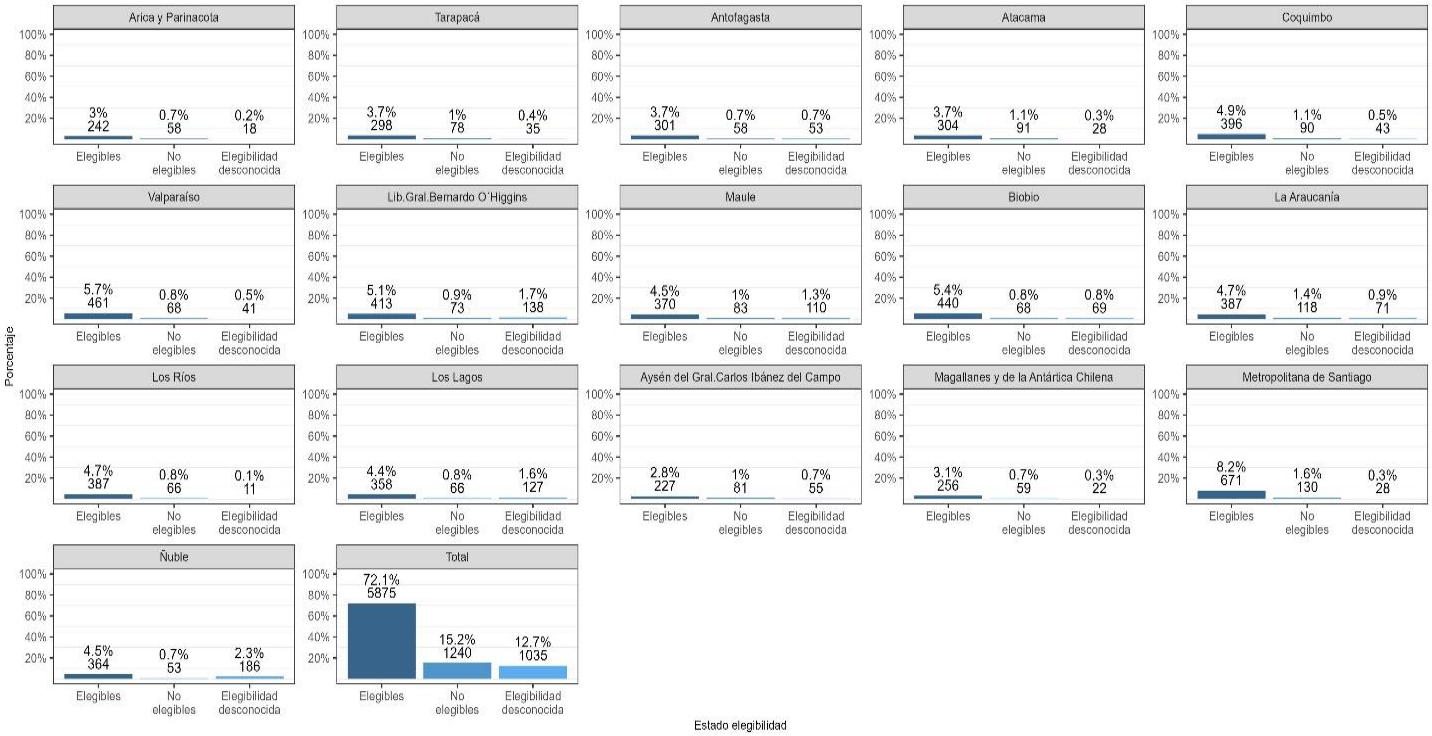
V.1.2. Ajuste por elegibilidad

V.1.2.1. Ajuste por elegibilidad desconocida

Durante el trabajo de campo, las empresas se clasifican según su estado de elegibilidad, diferenciando entre unidades “elegibles”, “no elegibles” o “elegibilidad desconocida” (ver Anexo 1). Tal como se observa en el Gráfico V.1, 72,1% del total de empresas se encuentran clasificadas como elegible, mientras que 15,2% es no elegible y 12,7% corresponde a elegibilidad desconocida.

La región Metropolitana de Santiago presenta el mayor porcentaje de empresas elegibles (8,2%), mientras que la región de Aysén del General Carlos Ibáñez del Campo presenta la menor proporción de empresas elegibles (2,8%). Por otra parte, la región de Ñuble presenta el menor número de empresas no elegibles (0,7%), mientras que la región Metropolitana de Santiago, presentan mayor proporción de empresas no elegibles (1,6%). En cuanto a las empresas con elegibilidad desconocida, la región donde se observa la mayor proporción de estas corresponde a Ñuble (2,3%).

Gráfico V.1. Distribución estado de elegibilidad de las empresas seleccionadas aleatoramiente en la Encla 2023, según región

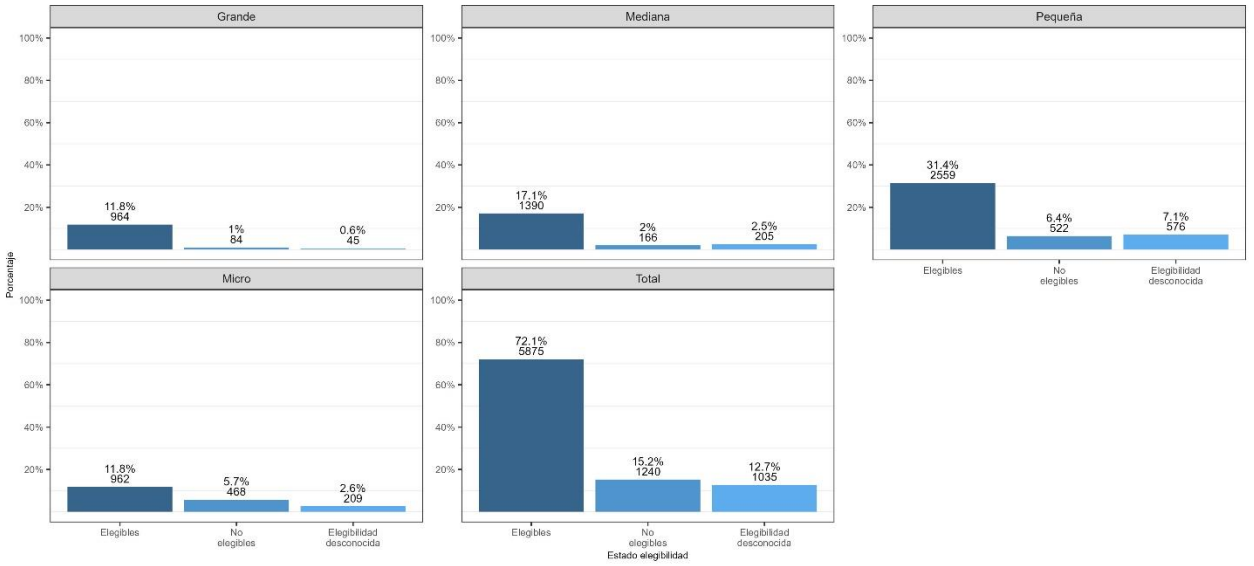


Elaborado por el Instituto Nacional de Estadísticas.

En el Gráfico V.2, se observa la distribución del estado de elegibilidad según tamaño. El tamaño pequeño presenta el mayor porcentaje de empresas elegibles (31,4%), mientras que el menor porcentaje de empresas elegibles se presenta en las microempresas (11,8%). Asimismo, la mayor proporción de empresas con elegibilidad desconocida y de empresas no elegibles se encuentra en el tamaño pequeño (7,1% y 6,4%, respectivamente).

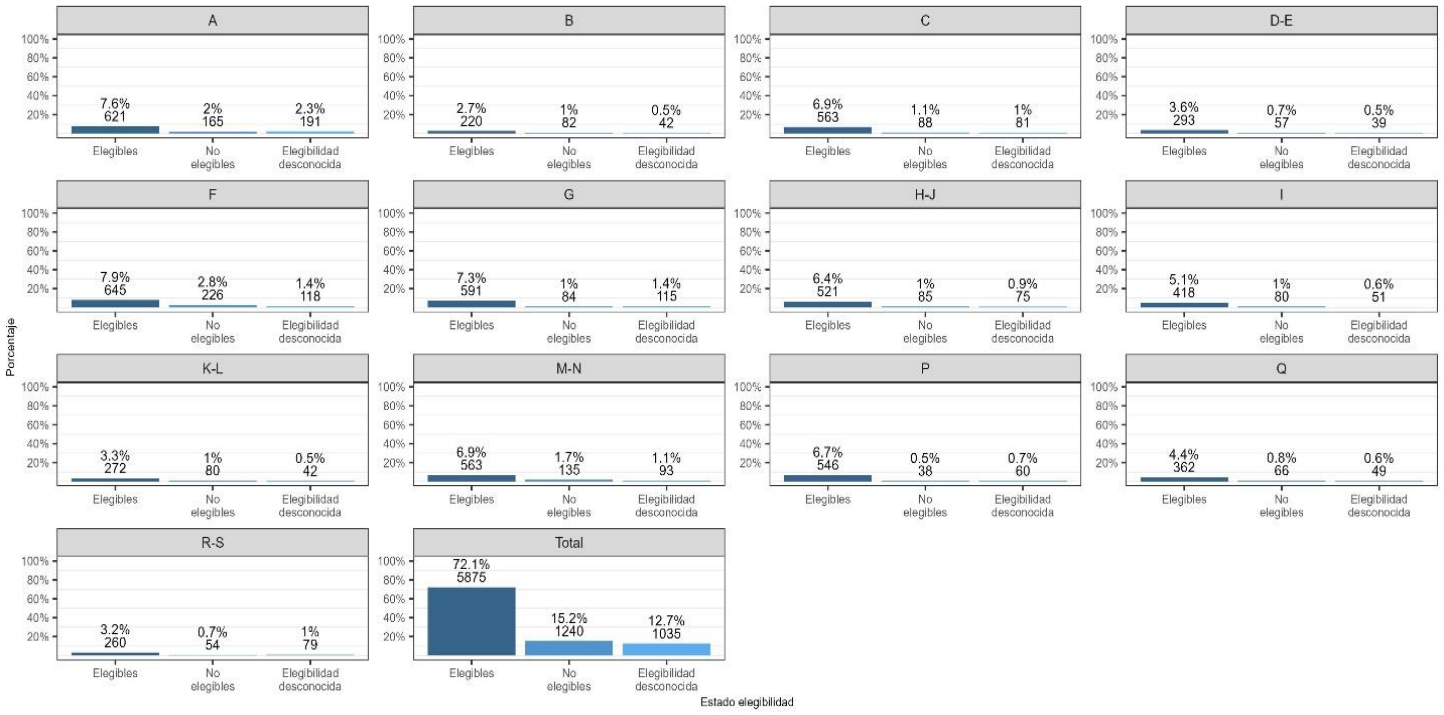
En el Gráfico V.3 se muestra el estado de elegibilidad de las empresas según la agrupación de actividad económica a la que pertenecen. Se observa que, en la actividad de construcción existe un mayor porcentaje de empresas elegibles (7,9%) con relación con el resto de las actividades económicas. Así mismo, dicha actividad presenta la mayor cantidad de las empresas no elegibles (2,8%), y Agricultura, ganadería, silvicultura y pesca evidencia la proporción más alta de empresas con elegibilidad desconocida (2,3 %).

Gráfico V.2. Distribución estado de elegibilidad de las empresas seleccionadas aleatoramiente en la Encla 2023, según tamaño de empresas por número de trabajadores



Elaborado por el Instituto Nacional de Estadísticas.

Gráfico V.3. Distribución estado de elegibilidad de las empresas seleccionadas aleatoramiente en la Encla 2023, según agrupación de actividad económica



Elaborado por el Instituto Nacional de Estadísticas.

Tal como se observa en el Gráfico V.1, 72,1% de las unidades se encuentra clasificada como elegible, mientras que 15,2% se clasificaron como no elegibles, y el 12,7% presentan elegibilidad desconocida. Las

regiones del Ñuble (2,3%) y de Los Ríos (0,1%) son las que concentran la mayor y menor proporción de unidades con elegibilidad desconocida, respectivamente.

Para realizar el ajuste por elegibilidad desconocida, dentro de cada estrato se distribuye el peso de las empresas con dicho estado entre las empresas con elegibilidad conocida (elegibles y no elegibles), para lo cual es necesario calcular una razón de ajuste mediante la siguiente fórmula:

$$\hat{R}_{rvh}^{ed} = \frac{\sum_{i \in \Omega_{rvh}} w_{rvh_i}}{\sum_{i \in \Omega_{rvh}^{ec}} w_{rvh_i}} \tag{11}$$

Donde:

$\hat{R}_{rvh}^{ed}$: Razón de ajuste por elegibilidad desconocida, para las empresas pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

w_{rvh_i} : Factor de expansión teórico de la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

Ω_{rvh} : Conjunto de empresas en la muestra pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

Ω_{rvh}^{ec} : Conjunto de empresas en la muestra pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h (elegibles y no elegibles).

A partir de esta etapa, las empresas clasificadas en condición de elegibilidad desconocida no son consideradas en el factor de expansión y se les asigna una ponderación nula, ya que su peso se distribuye en aquellas unidades cuya elegibilidad es conocida (elegibles y no elegibles). Para estas últimas, el factor de expansión ajustado por elegibilidad desconocida corresponde al producto entre la razón de ajuste y el factor de expansión teórico. Por lo tanto, el factor de expansión corregido por elegibilidad desconocida se obtiene mediante la siguiente fórmula:

$$w_{rvh_i}^{ed} = \begin{cases} \hat{R}_{rvh}^{ed} \cdot w_{rvh_i} & , si \quad i \in \Omega_{rvh}^{ec} \\ . & , si \quad i \in \Omega_{rvh}^{ed} \end{cases} \tag{12}$$

Donde:

$w_{rvh_i}^{ed}$: Factor de expansión ajustado por elegibilidad desconocida de la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\hat{R}_{rvh}^{ed}$: Razón de ajuste por elegibilidad desconocida, para las empresas pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

w_{rvh_i} : Factor de expansión teórico de la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

Ω_{rvh}^{ed} : Conjunto de empresas en la muestra pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h (elegibilidad desconocida).

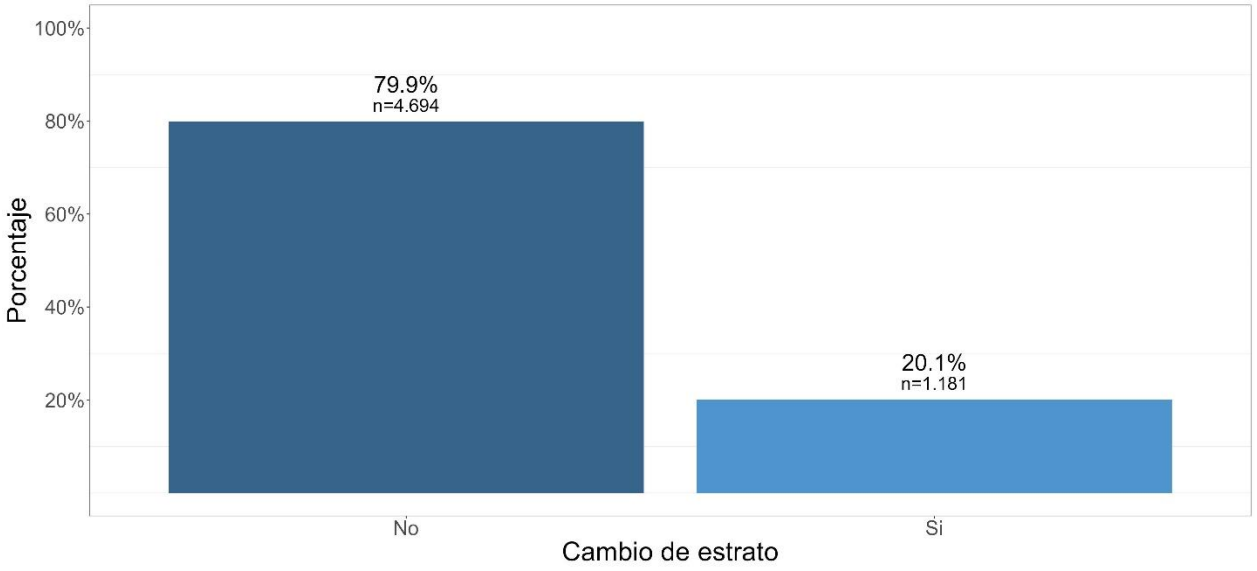
Ω_{rvh}^{ec} : Conjunto de empresas en la muestra pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h (elegibles y no elegibles).

V.1.2.2. Ajuste por cambio de estrato y no elegibilidad

Frecuentemente, en las encuestas a empresas se observa que la información que contienen las variables de estratificación informada en el marco muestral difiere del dato recolectado en el trabajo de campo, de modo que la empresa no pertenece al estrato en el cual fue seleccionada inicialmente. Para la Encla 2023, se observan cambios de estrato, los cuales pueden darse porque la empresa reportó pertenecer a otra región, ejercer otra actividad económica o tamaño.

Siendo así, en el Grafico V.4. se observa que, a nivel nacional 20,1% de las empresas elegibles cambiaron su estrato inicial.

Gráfico V.4. Proporción de empresas elegibles con cambio de estrato, a nivel nacional.



Elaborado por el Instituto Nacional de Estadísticas.

Al analizar los datos de la encuesta se identifican unidades categorizadas como “No elegibles”, lo cual se debe, principalmente, al desfase existente entre la información contenida en el marco muestral y el trabajo

de campo, lo cual se corrige a través del cálculo de los factores de expansión realizando un ajuste por no elegibilidad y cambio de estrato[7]. Por lo tanto, para realizar este ajuste se identifican tres casos:

1. La empresa i es elegible, fue seleccionada en el estrato l, y según la encuesta pertenece al estrato l'.

2. La empresa i es elegible y sigue perteneciendo al estrato en donde fue seleccionada.

3. La empresa i es no elegible.

Este ajuste se expresa en la siguiente fórmula:

$$w^{*}_{rvh_i} = \begin{cases} w^{ed}_{rvh_i} & , si \quad i \in \Omega^{el}_{rvh} \wedge i \in \Omega^{l \to l'}_{rvh} \\ \dfrac{N_{rvh}\left(1 - \dfrac{n^{ne}_{rvh}}{n^{ec}_{rvh}}\right) - \sum_{l \neq l'} \sum_{i=1}^{n^{el,l \to l'}_{rvh}} w^{ed}_{rvh_i}}{n^{el,l \to l}_{rvh}} & , si \quad i \in \Omega^{el}_{rvh} \wedge i \in \Omega^{l \to l}_{rvh} \\ . & , si \quad i \in \Omega^{ne}_{rvh} \end{cases} \tag{13}$$

Donde:

$w^{*}_{rvh_i}$: Factor de expansión ajustado por cambio de estrato y no elegibilidad para la empresa i en la muestra, perteneciente a la región r, tamaño de empresa v y agrupación de la actividad económica h.

$w^{ed}_{rvh_i}$: Factor de expansión ajustado por elegibilidad desconocida de la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

N_{rvh} : Número de empresas en el marco muestral, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

n^{ne}_{rvh} : Número de empresas en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h , categorizadas como no elegibles.

n^{ec}_{rvh} : Número de empresas en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h catalogadas con elegibilidad conocida (elegibles y no elegibles).

$n^{el,l \to l}_{rvh}$: Número de empresas en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h , catalogadas como elegibles y que han sido seleccionadas en el estrato l y después del trabajo de campo se mantienen en el mismo estrato.

Ω^{el}_{rvh} : Conjunto de empresas en la muestra perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h , catalogadas como elegibles.

[7] Para la Encla 2023, se adopta el ajuste realizado por el Instituto Nacional de Estadísticas de España en el diseño muestral de las encuestas dirigidas a empresas. Para más información visitar: https://www.ine.es/daco/daco43/metoite2016.pdf

$\Omega_{rvh}^{l \to l}$: Conjunto de empresas en la muestra que fueron seleccionadas en el estrato l y según la encuesta siguen perteneciendo al estrato l.

$n_{rvh}^{el,l \to l'}$: Número de empresas en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h , catalogadas como elegibles y que han sido seleccionadas en el estrato l y después del trabajo de campo pertenecen a otro estrato l'.

$\Omega_{rvh}^{l \to l'}$: Conjunto de empresas en la muestra que fueron seleccionadas en el estrato l y según la encuesta pertenecen al estrato l' (cambió de estrato).

Ω_{rvh}^{ne} : Conjunto de empresas en la muestra perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h, catalogadas como no elegibles.

. : Indica valor perdido o *missing*.

Es importante resaltar que, al realizar el ajuste por no elegibilidad, se redujo el stock poblacional al que se expande antes del ajuste, debido a que se descuenta el peso de las 1.240 empresas no elegibles. Por lo tanto, después de aplicar este procedimiento, la población de empresas es de 98.054, permitiendo corregir los efectos que pudieran causar estas empresas sobre la estimación.

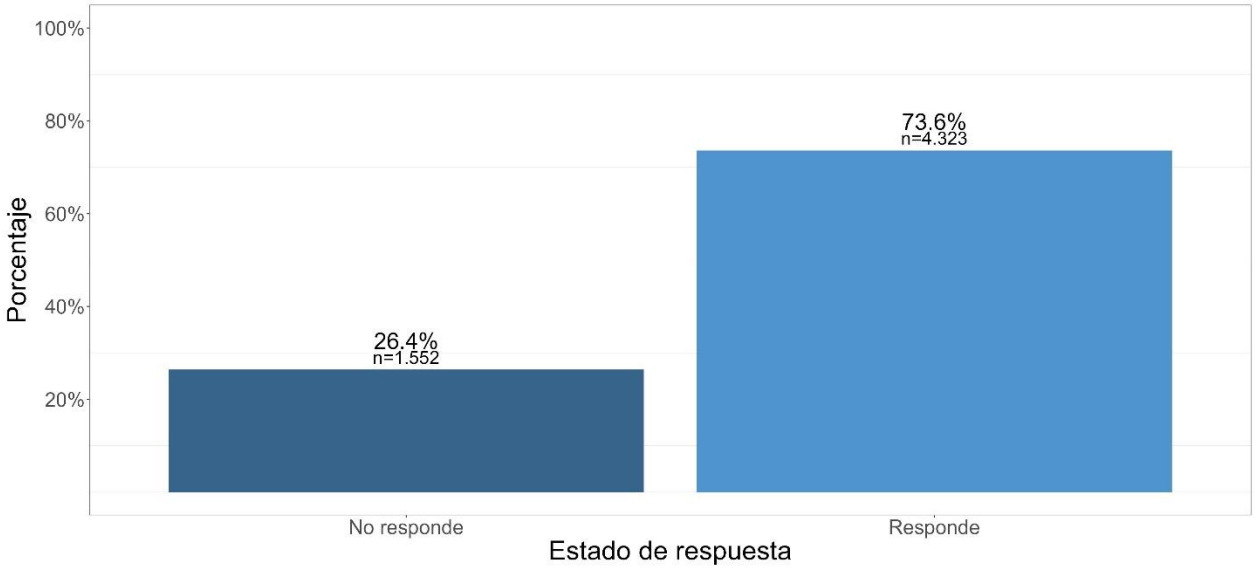
Por otra parte, en cuanto a los cambios de estrato, existe una compensación de empresas dentro de cada uno de ellos, dado que, mientras unas empresas se van del estrato, otras llegan al mismo.

V.1.3. Ajuste por no respuesta

Existen algunas unidades muestrales que, aun siendo elegibles, no proporcionan la información solicitada o lo hacen de forma incompleta, lo que se denomina “no respuesta”. Esto tiene como consecuencia que, al momento de realizar los análisis con el factor de expansión definido hasta el paso anterior, no se estaría dando cuenta de la totalidad de las empresas elegibles de la población.

Por lo tanto, en el Grafico V.5, se observa que a nivel nacional 73,6% de las empresas elegibles respondieron la encuesta y 26,4% no lo realizó.

Gráfico V.5. Proporción de empresas elegibles que no respondió la Encla 2023, a nivel nacional.



Elaborado por el Instituto Nacional de Estadísticas.

Por lo tanto, una vez aplicadas las estrategias operativas necesarias para conseguir la información sin mayor éxito, se aplican metodologías para corregir o mitigar los efectos en los estimadores. Esto implica realizar un ajuste por no respuesta en los factores de expansión. Se opta por la estrategia de ajuste por celda, que consiste en distribuir los pesos de las unidades que no responden entre las unidades que sí responden de forma independiente al interior de cada estrato. Para realizar este ajuste, primero se calcula una razón de ajuste por estrato, correspondiente a la suma de los factores de expansión de las unidades elegibles sobre la suma de los factores de expansión de las unidades que respondieron la encuesta, tal como se observa en la siguiente fórmula:

$$\hat{R}_{rvh}^{nr} = \frac{\sum_{i \in \Omega_{rvh}^{el}} w_{rvh_i}^{*}}{\sum_{i \in \Omega_{rvh}^{el,R}} w_{rvh_i}^{*}} \qquad (14)$$

Donde:

$\hat{R}_{rvh}^{nr}$: Razón de ajuste por no respuesta para las empresas pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$w_{rvh_i}^{*}$: Factor de expansión ajustado por cambio de agrupación de actividad económica o región y no elegibilidad para la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

Ω_{rvh}^{el} : Conjunto de empresas en la muestra pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h, catalogadas como elegibles.

$\Omega_{rvh}^{el,R}$: Conjunto de empresas en la muestra pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h, catalogadas como elegibles y que respondieron la encuesta.

Posteriormente, para las unidades que respondieron la encuesta se multiplica el factor de expansión ajustado por cambio de estrato y no elegibilidad por la razón de ajuste por no respuesta, mientras que para el resto de las unidades se deja un valor perdido o *missing*.

$$w_{rvh_i}^{nr} = \begin{cases} \hat{R}_{rvh}^{nr} \cdot w_{rvh_i}^{*} & , si \quad i \in \Omega_{rvh}^{el,R} \\ . & , si \quad i \in \Omega_{rvh}^{el,NR} \end{cases} \tag{15}$$

Donde:

$w_{rvh_i}^{nr}$: Factor de expansión ajustado por no respuesta para la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\hat{R}_{rvh}^{nr}$: Razón de ajuste por no respuesta para las empresas pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$w_{rvh_i}^{*}$: Factor de expansión ajustado por cambio de agrupación de actividad económica o región y no elegibilidad para la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\Omega_{rvh}^{el,R}$: Conjunto de empresas en la muestra pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h, catalogadas como elegibles y que respondieron la encuesta.

$\Omega_{rvh}^{el,NR}$: Conjunto de empresas en la muestra pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h, catalogadas como elegibles y que no respondieron la encuesta.

Al realizar el ajuste por no respuesta, se redujo el marco muestral a 96.935 empresas, debido a que se encontraron estratos sin ninguna empresa lograda, lo que trae como consecuencia no poder distribuir los pesos muestrales.

V.1.4. Suavizamiento del factor de expansión

Después de realizar los ajustes descritos anteriormente, se observa la presencia de valores atípicos en los factores de expansión, de modo que surge la necesidad de realizar un suavizamiento[8].

Antes de realizar el suavizamiento, se debe establecer o definir algún criterio o mecanismo para identificar valores atípicos en los factores de expansión. Para ello, se establece como regla que, al interior de cada estrato, todo aquel factor de expansión que se encuentre a dos o más desviaciones estándar respecto al factor de expansión promedio del estrato es considerado atípico. Es decir, el factor de expansión $w_{rvh_i}^{nr}$ se considera atípico si está a una distancia superior a 2 desviaciones estándares respecto al promedio de su estrato. Si denotamos al factor atípico como A_{rvh_i}, entonces A_{rvh_i} no pertenece al intervalo I_{rvh}, tal como se expresa en la siguiente fórmula:

[8] El método para el suavizamiento del factor de expansión corresponde al mismo aplicado en ENCLA 2019.

$$A_{rvh_i} \notin \mathrm{I}_{rvh} = \left[\bar{w}^{nr}_{rvh} - 2 \cdot \sigma^{nr}_{w_{rvh}} \ , \bar{w}^{nr}_{rvh} + 2 \cdot \sigma^{nr}_{w_{rvh}}\right] \quad (16)$$

Donde:

A_{rvh_i} : Factor de expansión de la empresa i en la muestra, considerado atípico, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\bar{w}^{nr}_{rvh}$: Promedio de los factores de expansión ajustados por no respuesta del conjunto de empresas pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\sigma^{nr}_{w_{rvh}}$: Desviación estándar de los factores de expansión del conjunto de empresas, pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

Una vez identificados los estratos con presencia de valores atípicos, se suavizan todos los factores de expansión al interior de éstos, mientras que en aquellos estratos donde no hay presencia de valores atípicos, se mantiene el factor de expansión anterior. La fórmula para realizar el suavizamiento es la siguiente[9]:

$$w^{s}_{rvh_i} = \begin{cases} \dfrac{\left(w^{nr}_{rvh_i} + \bar{w}^{nr}_{rvh}\right)}{2} & , si \quad i \in \left\{\Omega^{el,R}_{rvh} \supset A_{rvh_i}\right\} \\ w^{nr}_{rvh_i} & , si \quad i \in \left\{\Omega^{el,R}_{rvh} \not\supset A_{rvh_i}\right\} \end{cases} \quad (17)$$

Donde:

$w^{s}_{rvh_i}$: Factor de expansión suavizado de la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$w^{nr}_{rvh_i}$: Factor de expansión ajustado por no respuesta de la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\bar{w}^{nr}_{rvh}$: Promedio de los factores de expansión ajustados por no respuesta del conjunto de empresas de la región r, tamaño de la empresa v y agrupación de actividad económica h.

A_{rvh_i} : Factor de expansión de la empresa i en la muestra, considerado atípico, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\Omega^{el,R}_{rvh}$: Conjunto de empresas en la muestra que son categorizadas como elegibles y respondieron la encuesta, pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

La fórmula de suavizamiento utilizada garantiza que el factor de expansión suavizado expanda a la misma población que el factor de expansión ajustado por no respuesta. Asimismo, garantiza mantener el orden de los factores de expansión previo y posterior al suavizamiento, es decir, si antes de realizar el suavizamiento $w^{nr}_{rvh_{i=1}} > w^{nr}_{rvh_{i=2}}$, una vez realizado el suavizamiento $w^{s}_{rvh_{i=1}} > w^{s}_{rvh_{i=2}}$.

[9] La fórmula utilizada para realizar el suavizamiento del factor de expansión consiste en obtener un promedio simple entre el factor de expansión y el promedio de los factores de expansión de su estrato. Esta fórmula puede ser modificada dependiendo si se desea dar más ponderación al promedio o al valor del factor de expansión de la empresa. En este caso, se entrega igual ponderación al promedio y al factor de expansión.

Para esta versión de la encuesta, 4,5% de las unidades contaban con un factor de expansión ajustado por no respuesta fuera del umbral I_{rvh}. Por lo tanto, se realizó el suavizamiento a los estratos donde se encontraban estos valores. La fórmula de suavizamiento utilizada garantiza que el factor expanda a la misma población que el factor de expansión ajustado por no respuesta, es decir, 96.935 empresas.

V.1.5. Calibración del factor de expansión

La calibración de los factores de expansión implica ajustar los pesos de la muestra para que reflejen de manera más precisa las características de la población que se está estudiando. Esto se logra mediante la comparación de las características de la muestra y la población en diferentes variables.

Los métodos de calibración utilizan algoritmos iterativos que ajustan los factores de expansión con la finalidad de que alcancen totales poblacionales de variables auxiliares, minimizando la diferencia entre los factores de expansión originales y los calibrados. Esto permite asociar, por aproximación, cualquier propiedad óptima de los ponderadores iniciales a los calibrados. Asimismo, la calibración garantiza que las estimaciones de la encuesta sean coherentes con las estimaciones que ya han sido publicadas.

Tal como se mencionó anteriormente, no todas las empresas seleccionadas respondieron la encuesta, por lo tanto, la suma de los factores de expansión de las unidades que sí respondieron no alcanza el mismo stock que el tamaño estimado de la población objetivo de la encuesta. Para corregir esto, se realiza un ajuste mediante un método de calibración por *raking*[10].

El método de calibración por *raking*, el cual se utiliza para este ajuste, asegura que los factores de expansión calibrados sean positivos, debido a que la función de distancia no se encuentra definida por valores negativos.

La función de distancia $G(w^c, w^s)$, se define mediante la siguiente fórmula:

$$G(w^c, w^s) = w^c \log(w^c/w^s) - w^c + w^s \tag{18}$$

Donde:

[10] Para más información, revisar: Deville & Särndal, (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association.* 87(418), 376-382.

$G(w^c, w^s)$: Función de distancia entre el factor de expansión calibrado y el factor de expansión suavizado.

w^c : Factor de expansión calibrado.

w^s : Factor de expansión suavizado.

Lo anterior permite obtener un factor de expansión calibrado a nivel de la marginal correspondiente, el cual se obtiene mediante la siguiente fórmula:

$$w^c = w^s \cdot F(w^c, w^s) \tag{19}$$

Donde:

w^c : Factor de expansión calibrado.
w^s : Factor de expansión suavizado.
$F(w^c, w^s)$: Función inversa de $\partial G(w^c, w^s)/\partial w^c$, también denotada por $g^{-1}(w^c, w^s)$.

Las marginales definidas para la calibración son las siguientes:

- Total de empresas para cada variable: región, actividad económica y tamaño de empresa, según estimación del marco muestral reconstruido con periodo de referencia noviembre 2022 a partir del Registro Administrativo de SUSESO.
- Total de número de trabajadores para cada variable: región, actividad económica, tamaño de empresa), según estimación del marco muestral reconstruido con periodo de referencia noviembre 2022 a partir del Registro Administrativo de SUSESO.

Para realizar la calibración por *raking*, se utilizó el paquete ReGenesees en el software estadístico R. Luego del proceso de calibración, el factor expande a 129.058 empresas de inclusión aleatoria.

V.2. Factor de expansión para las empresas del tramo de inclusión forzosa

Las 173 unidades que pertenecen al tramo de inclusión forzosa poseen factores de expansión iguales a 1, ya que, por su relevancia para el sector en estudio, se considera que su comportamiento es único, por lo que no pueden representar ni ser representadas por otras unidades.

Finalmente, el factor expande a 129.231 empresas, esto luego de todos los pasos de ajustes y considerando las empresas incluidas forzosa y aleatoriamente.

VI. ESTIMADORES Y MEDIDAS DE PRECISIÓN

La estimación estadística corresponde al procedimiento mediante el cual se calculan parámetros de la población a partir de datos muestrales obtenidos. En el contexto de la Encla 2023 se realiza la estimación del total del número de trabajadores, variable que se extrae del formulario autoaplicado, específicamente de la pregunta A1: Indique el total de personas contratadas directamente por esta empresa al 30 de noviembre de 2022, según región del país y sexo. Esta estimación se calcula para cada nivel de dominio del estudio. Para exponer la fórmula de cálculo, esta se expresará en el nivel más desagregado de variables.

La estimación se obtiene mediante la siguiente fórmula:

$$\hat{t}_{rvh} = \sum_{i=1}^{n_{rvh}} y_{rvh_i}^{A1} \cdot w_{rvh_i}^{c} \qquad (20)$$

Donde:

$\hat{t}_{rvh_i}$: Estimación del total de trabajadores contratados por las empresas directamente al 30 de noviembre del 2022, pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$y_{rvh_i}^{A1}$ Número de trabajadores contratados directamente por la empresa i al 30 de noviembre del 2022, pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$w_{rvh_i}^{c}$: Factor de expansión calibrado de la empresa i en la muestra, perteneciente a la región r, tamaño de la empresa v y agrupación de actividad económica h.

Por otro lado, para obtener las medidas de precisión, se debe estimar la varianza a nivel de estrato, tal como se especifica en la siguiente fórmula:

$$\hat{v}(\hat{t}_{rvh}) = \sum_{i=1}^{n_{rvh}} \widehat{N}_{rvh} \cdot (\widehat{N}_{rvh} - n_{rvh}) \cdot \frac{S^2(y)_{rvh}}{n_{rvh}} \qquad (21)$$

Donde:

$\hat{v}(\hat{t}_{rvh})$: Varianza de la estimación del total de trabajadores contratados por las empresas directamente al 30 de noviembre del 2022, pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\widehat{N}_{rvh}$: Estimación del número de empresas en la población de la región r, tamaño de empresa v y agrupación de actividad económica h.

n_{rvh} : Número de empresas en la muestra lograda, pertenecientes a la región r, tamaño de empresa v y agrupación de actividad económica h.

$S^2(y)_{rvh}$: Cuasi-varianza del número de trabajadores de la región r, tamaño de empresa v y agrupación de actividad económica h en la muestra.

Por otro lado, el error relativo de la estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre de 2022 en la Encla 2023, se obtiene a partir de la siguiente fórmula:

$$e\left(\hat{t}_{rvh}^{A1}\right) = \frac{\sqrt{v(\hat{t}_{rvh}^{A1})}}{\hat{t}_{rvh}^{A1}} \cdot t_{\text{gl},1-\frac{\alpha}{2}} \qquad (22)$$

Donde:

$e\left(\hat{t}_{rvh}^{A1}\right)$: Error relativo de la estimación del total de trabajadores contratados por las empresas directamente al 30 de noviembre del 2022, pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$\hat{t}_{rvh}^{A1}$: Estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$v\left(\hat{t}_{rvh}^{A1}\right)$[11] : Varianza de la estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, pertenecientes a la región r, tamaño de la empresa v y agrupación de actividad económica h.

$t_{gl,1-\frac{\alpha}{2}}$: Cuantil de la distribución T-Student correspondiente a gl grados de libertad y un nivel de confianza $(1 - \alpha)$.

VII. ESTIMACIONES DE LOS PARÁMETROS DE INTERÉS

Una vez generados los factores de expansión, y junto con la base de datos de levantamiento, se calculan los errores de estimación asociados a la variable de interés. En las siguientes tablas se presenta la estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, además del error relativo, según región, tamaño de empresa, agrupación de actividad económica y el cruce de región y tamaños de empresa y cruce de agrupación regional y tamaños de empresa según número de trabajadores, respectivamente.

El tamaño muestral logrado finalmente corresponde a 4.496 empresas a nivel nacional con un error relativo de 3,10%.

[11] Los cálculos realizados para la varianza de la estimación del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022 fueron realizados en el software R-Studio.

Tabla VII.1. Estimación y error relativo del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, según agrupación de actividad económica

Agrupación de actividad económica	Estimación de empresas en la población	Número de trabajadores (NT)		
		Estimación trabajadores	CV (%)	Error relativo (%)
Total	**129.231**	**4.923.727**	**1,58**	**3,10**
Agricultura, ganadería, silvicultura y pesca	11.076	377.307	4,00	7,87
Explotación de minas y canteras	831	92.271	2,62	5,19
Industrias manufactureras	15.269	587.583	3,83	7,54
Suministro de electricidad, gas, vapor y aire acondicionado y Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación	994	60.859	4,21	8,31
Construcción	15.130	600.979	4,90	9,63
Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	28.374	844.790	4,56	8,96
Transporte y almacenamiento e Información y comunicaciones	12.878	461.308	4,37	8,59
Actividades de alojamiento y de servicio de comidas	7.738	257.421	7,21	14,21
Actividades financieras y de seguros y Actividades inmobiliarias	5.455	225.234	5,67	11,20
Actividades profesionales, científicas y técnicas y Actividades de servicios administrativos y de apoyo	14.522	773.652	4,80	9,43
Enseñanza	6.063	328.169	6,44	12,65
Actividades de atención de la salud humana y de asistencia social	3.301	148.353	3,54	6,97
Actividades artísticas, de entretenimiento y recreativas y Otras actividades de servicios	7.600	165.801	8,39	16,63

Elaborado por el Instituto Nacional de Estadísticas.

Tabla VII.2. Estimación y error relativo del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, según tamaño de empresa

Tamaño de empresa	Estimación de empresas en la población	Número de trabajadores (NT)		
		Estimación trabajadores	CV (%)	Error relativo (%)
Total	**129.231**	**4.923.727**	**1,58**	**3,10**
Grande	3.349	2.417.672	1,90	3,72
Mediana	11.113	1.027.261	4,52	8,86
Pequeña	55.324	1.091.654	2,81	5,52
Micro	59.445	387.140	7,57	14,88

Elaborado por el Instituto Nacional de Estadísticas.

Tabla VII.3. Estimación y error relativo del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, según agrupación de región y tamaños

Agrupación de región	Tamaño	Estimación de empresas en la población	Número de trabajadores (NT)		
			Estimación trabajadores	CV (%)	Error relativo (%)
Total		**129.231**	**4.923.727**	**1,58**	**3,10**
Norte grande	Grande	34	17.396	6,21	13,10
	Mediana	199	18.609	3,08	6,11
	Pequeña	1.757	33.539	4,18	8,24
	Micro	1.607	10.012	13,51	26,96
Antofagasta	Grande	81	31.378	3,33	6,80
	Mediana	331	26.307	6,80	13,74
	Pequeña	1.673	35.868	8,13	16,22
	Micro	1.775	11.377	5,45	11,21
Norte chico	Grande	90	30.941	4,14	8,35
	Mediana	447	39.605	4,26	8,44
	Pequeña	2.530	52.893	3,92	7,73
	Micro	2.729	18.175	9,30	18,55
Centro Sur	Grande	276	160.694	3,45	6,82
	Mediana	1.286	118.756	2,75	5,42
	Pequeña	6.155	133.718	4,18	8,22
	Micro	7.674	50.184	10,19	20,17
Valparaíso	Grande	129	83.491	3,24	6,46
	Mediana	879	81.363	4,70	9,36
	Pequeña	6.188	120.302	5,85	11,61
	Micro	4.774	32.644	11,77	24,48
Biobío	Grande	206	106.468	8,05	16,03
	Mediana	657	61.497	4,92	9,82
	Pequeña	3.455	65.058	5,23	10,36
	Micro	4.768	31.093	13,68	27,94
Sur	Grande	171	95.434	2,53	5,02
	Mediana	1.063	94.720	4,72	9,31
	Pequeña	5.637	116.705	2,97	5,84
	Micro	6.837	41.139	9,68	19,23
Zona austral	Grande	22	11.176	0,54	1,21
	Mediana	140	12.017	4,21	8,39
	Pequeña	998	19.776	3,34	6,59
	Micro	996	6.268	10,02	20,06
Metropolitana	Grande	2.341	1.880.695	2,36	4,65
	Mediana	6.110	574.386	7,96	15,79
	Pequeña	26.931	513.794	5,59	11,06
	Micro	28.287	186.249	14,98	31,26

Elaborado por el Instituto Nacional de Estadísticas.

Tabla VII.4. Estimación y error relativo del total de trabajadores contratados directamente por las empresas al 30 de noviembre del 2022, según región.

Región	Estimación de empresas en la población	Número de trabajadores (NT)		
		Estimación trabajadores	CV (%)	Error relativo (%)
Total	**129.231**	**4.923.727**	**1,58**	**3,10**
Tarapacá	2.375	53.179	2,77	5,46
Antofagasta	3.860	104.931	3,46	6,83
Atacama	1.557	44.087	3,16	6,25
Coquimbo	4.239	97.526	3,19	6,29
Valparaíso	11.970	317.799	2,92	5,75
Libertador General Bernardo O'Higgins	6.267	205.284	3,28	6,45
Maule	6.583	190.240	3,66	7,20
Biobío	9.086	264.116	4,01	7,90
La Araucanía	5.285	131.334	3,71	7,30
Los Lagos	6.163	173.051	2,98	5,87
Aysén del General Carlos Ibáñez del Campo	690	12.239	2,64	5,21
Magallanes y La Antártica Chilena	1.465	36.998	2,68	5,30
Metropolitana	63.669	3.155.123	2,39	4,69
Los Ríos	2.259	43.613	4,24	8,37
Arica y Parinacota	1.222	26.378	6,70	13,24
Ñuble	2.541	67.829	3,34	6,59

Elaborado por el Instituto Nacional de Estadísticas.

VIII. ANEXOS

Anexo N°1. Clasificación de empresas en estados de disposición final

Previo al cálculo de los factores de expansión, se debe clasificar a las empresas seleccionadas para la muestra según el resultado del proceso de levantamiento. Para esto, se utilizan los códigos de disposición final de la AAPOR (2016) los cuales además se pueden clasificar según estado de elegibilidad. En el siguiente cuadro se presenta la descripción de los códigos de disposición final.

Cuadro VIII.1. Definición códigos de disposición final AAPOR

Clasificación AAPOR		
Códigos de disposición final	**Estado Elegibilidad**	**Estados de Disposición Final**
Empresa lograda	Elegible	Entrevista completa
Logro parcial	Elegible	Entrevista parcial
Rechazo	Elegible	Elegible no entrevistado
Elegible no contactada	Elegible	Elegible no entrevistado
Elegibilidad desconocida	Elegibilidad desconocida	Elegibilidad desconocida
No elegible	No elegible	No elegible

Elaborado por el Instituto Nacional de Estadísticas.

IX. BIBLIOGRAFIA

AAPOR. (2016). *Standard Definitions. Final dispositions case codes and outcome rates for surveys.* Recuperado el 20 de Noviembre de 2018, de http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

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

Instituto Nacional de Estadística de España. (2016). *Encuesta sobre innovación a las empresas.* Recuperado de https://www.ine.es/daco/daco43/metoite2016.pdf

Lohr, S. L. (2000). *Muestreo Diseño y análisis.* México: International Thomson Editores.