



VII Encuesta de Microemprendimiento EME 2022

Documento metodológico del diseño muestral

INSTITUTO NACIONAL DE ESTADÍSTICAS

Junio, 2023

DEPARTAMENTO DE METODOLOGÍAS E INNOVACIÓN ESTADÍSTICA
SUBDEPARTAMENTO DE DISEÑO DE MARCOS Y MUESTRAS
UNIDAD DE MUESTRAS SOCIALES

VII Encuesta de Microemprendimiento EME 2022.
Documento metodológico del diseño muestral.

Instituto Nacional de Estadísticas, Chile.
Fecha de elaboración: junio de 2023

ÍNDICE

I.	DISEÑO MUESTRAL.....	6
I.1.	Objetivos	6
I.1.1.	Objetivo general del estudio	6
I.1.2.	Objetivos específicos del estudio	6
I.1.3.	Objetivo del diseño muestral	7
I.2.	Población objetivo	7
I.3.	Unidad de información.....	7
I.4.	Dominios de estudio o niveles de estimación	7
I.5.	Período de referencia y periodicidad	7
I.6.	Marco Muestral.....	8
I.6.1.	Características del MMV 2017.....	9
I.6.2.	Marco muestral de la Encuesta de Microemprendimiento	13
I.7.	Estrategia muestral	17
I.8.	Cálculo y distribución del tamaño muestral.....	19
I.8.1.	Metodología de cálculo del tamaño muestral.....	20
I.8.2.	Errores esperados.....	23
I.8.3.	Distribución de la muestra según submuestra	24
I.8.4.	Errores observados.....	25
I.9.	Selección de unidades muestrales	25
I.9.1.	Selección de viviendas	25
I.9.2.	Selección de microemprendedores.....	27
II.	DESARROLLO DE FACTORES DE EXPANSIÓN	28
II.1.	Ponderador Base	29
II.1.1.	Probabilidad de selección y entrevista de las viviendas en la ENE- Trimestre MAM 2022 29	
II.1.2.	Probabilidad de selección de los microemprendedores	29
II.2.	Ajuste por elegibilidad.....	31
II.2.1.	Ajuste por elegibilidad desconocida.....	31
II.2.2.	Ajuste por no elegibilidad.....	32
II.3.	Ponderador ajustado por falta de respuesta	33
II.4.	Suavizamiento del Ponderador ajustado por no respuesta	35
II.4.1.	Método de contracción a la media (CM).....	35
	Este método contrae los factores de expansión, acercándolos al valor promedio de éstos, haciendo que la distribución de los factores sea más homogénea. El factor suavizado o contraído a la media (o promedio) viene dado por:	35
II.4.2.	Método r-k.....	36
II.4.3.	Método mixto	36
II.5.	Calibración.....	37
III.	ESTIMACIÓN DE LA VARIANZA	40
III.1.	Estimación de variables y varianzas en R	41

ÍNDICE DE TABLAS

Tabla I.1. Cantidad de UPM especiales y excluidas, por región	15
Tabla I.2. Cantidad de estratos de muestreo por región y área.....	17
Tabla I.3. Total de microemprendedores según ENE trimestre MAM 2022 y según Marco VII EME 2022	19
Tabla I.4. Errores esperados asociados al parámetro de interés según tamaño muestral objetivo.....	23
Tabla I.5. Total de viviendas seleccionadas según región y mes de levantamiento	24
Tabla I.6. Errores observados asociados al parámetro de interés según número de viviendas que responden	25
Tabla II.1. Distribución de individuos según la clasificación de elegibilidad.	32
Tabla II.2. Total de unidades elegibles, unidades que responden y tasas de respuesta.....	34
Tabla II.3. Total de micro emprendedores estimados a partir de la ENE – periodo MAM 2022	38
Tabla III.1. Estructura de la actividad económica de los microemprendedores	42
Tabla III.2. Total microemprendedores según códigos de disposición final y estado de elegibilidad.	43

ÍNDICE DE CUADROS

Cuadro I.1. Criterios utilizados en el cálculo del tamaño muestral.....	20
---	----

PRESENTACIÓN

La Subsecretaría de Economía y Empresas de Menor Tamaño, responsable de formular e implementar políticas de fomento productivo para las empresas de menor tamaño (EMT) de Chile, ha solicitado al Instituto Nacional de Estadísticas (INE) la aplicación, desde el año 2013, de la “Encuesta de Microemprendimiento” (EME). Esta encuesta es un instrumento de enorme valor ya que permite caracterizar el diverso mundo de los microemprendedores en el país y sus determinantes. En particular, resulta relevante la medición de la economía con características de informalidad, puesto que es el único estudio de este tipo que se realiza a lo largo de todo Chile y a través de todos los sectores económicos.

El objetivo principal de la VII EME, es obtener una caracterización profunda de los microemprendimientos que se desarrollan a nivel nacional, permitiendo conocer las limitantes y los elementos facilitadores que tienen las unidades económicas de menor tamaño para llevar a cabo sus actividades.

La muestra de la VII EME, se obtiene a partir del levantamiento de la “Encuesta Nacional de Empleo” (ENE) en el trimestre móvil marzo, abril, mayo (MAM 2022) por lo que el diseño considera una estrategia de muestreo bifásico, donde la primera fase corresponde al levantamiento de la ENE en el trimestre señalado. Por su parte, la segunda fase corresponde a todas las viviendas que se identificaron en la primera fase, que contienen al menos un microemprendedor¹; conformando así el marco de muestreo.

El presente informe, contiene una descripción metodológica del diseño muestral, que incluye los objetivos del estudio, las características del marco muestral utilizado, la metodología de cálculo del tamaño muestral, los métodos de selección de las unidades muestrales y la metodología utilizada para expandir los datos.

¹ Todos los trabajadores por cuenta propia y empleadores con hasta 10 trabajadores, incluyéndose.

I. DISEÑO MUESTRAL

I.1. Objetivos

I.1.1. Objetivo general del estudio

Caracterizar los microemprendimientos que se desarrollan a nivel nacional, permitiendo conocer las limitantes y los elementos facilitadores que tienen las unidades económicas de menor tamaño para llevar a cabo sus actividades dentro del mercado laboral, a partir de una muestra de viviendas particulares, representativa a nivel nacional y regional y que contienen, al menos, a una persona microempresaria.

I.1.2. Objetivos específicos del estudio

1. Identificar y caracterizar la situación de formalidad de los microemprendimientos bajo distintas dimensiones (registros contables, inscripción en Servicios de Impuestos Internos, declaración de impuestos, organización jurídica, generación de empleo formal e informal, etc.) y sus determinantes.
2. Indagar acerca de la relación que tiene el negocio con el sistema financiero, a través del acceso y trabas al financiamiento, sus características y usos del mismo.
3. Estudiar la motivación y las razones del surgimiento de los emprendimientos. Si estos son motivados por necesidad, por oportunidad, o bien, causados por situaciones del entorno.
4. Identificar los obstáculos que dificultan el desarrollo de las unidades productivas, tales como las restricciones en materia de acceso a tecnología, capacitación, financiamiento, entre otros.
5. Conocer la situación laboral actual del trabajador independiente, así como sus experiencias o fracasos anteriores como emprendedor.
6. Conocer el nivel educacional de los emprendedores, además de las áreas más importantes en donde ha recibido capacitación en los últimos cinco años.
7. Realizar una recopilación de datos que permita comparar los resultados con estadísticas internacionales sobre industrias y emprendimiento.
8. Estudiar el impacto de la pandemia del COVID-19 en el funcionamiento del microemprendimiento, sus ganancias, sus trabajadores y la persistencia en el tiempo.

I.1.3. Objetivo del diseño muestral

Obtener estimaciones del parámetro de interés “Proporción de microemprendedores por cuenta propia sobre el total de microemprendedores” según los niveles de precisión establecidos² a nivel nacional y regional.

I.2. Población objetivo

La población objetivo está conformada por todos los trabajadores por cuenta propia y empleadores con hasta 10 trabajadores incluyéndose, denominados Microemprendedores, que residen en viviendas particulares del territorio nacional.

I.3. Unidad de información

La unidad de información es el Microemprendedor perteneciente a la vivienda particular, que fue clasificado como trabajador independiente en la Encuesta Nacional de Empleo y que mantiene su condición.

I.4. Dominios de estudio o niveles de estimación

Los niveles de estimación, cuyos tamaños muestrales son definidos para garantizar representatividad estadística de la estimación del parámetro de interés “Proporción de microemprendedores por cuenta propia sobre el total de microemprendedores” son:

- Nacional.
- Regional.

I.5. Período de referencia y periodicidad

La información se levanta en el trimestre mayo, junio y julio 2022 (MJJ 2022). El estudio tiene una periodicidad bienal, con excepción de esta versión de la encuesta, cuya periodicidad es trienal (2019 - 2022). El periodo de referencia corresponde al trimestre MAM 2022 en el cual se realiza la primera fase de la encuesta (levantamiento de la ENE).

² Los niveles de precisión son: error absoluto de 1,4% y error relativo de 1,6%, a nivel nacional. Para las regiones, el error absoluto máximo es 8,1% y, el error relativo no supera 10,5%. Este valor se obtiene en la Región de Magallanes.

I.6. Marco Muestral

Un marco de muestreo se define como la lista o los procedimientos que permiten identificar a todos los elementos de una población objetivo (Groves, 2004). Los marcos constituyen el insumo fundamental para el diseño de encuestas por muestreo y, por consiguiente, resulta de suma importancia que su metodología de conformación sea debidamente documentada. Entre las propiedades deseables para un marco muestral, se encuentran la exhaustividad, la exactitud y su actualización (Naciones Unidas, 2009).

La propiedad de exhaustividad se refiere a que el marco muestral cubra efectivamente la totalidad de la población de interés, mientras que la propiedad de exactitud hace referencia a contar con información fidedigna, ausencia de duplicidad de registros, etc. La propiedad de actualización se refiere a la capacidad de un marco muestral de mantener información reciente y oportuna. Si bien, un nivel de actualización perfecto es imposible, es importante que un marco muestral cuente con procedimientos que permitan minimizar los errores no muestrales que surjan de la desactualización.

El marco muestral que se usa con mayor frecuencia en las encuestas sociales está constituido por una lista de unidades geográficas, caracterizadas y actualizadas con base en información censal, desde donde se realiza la selección para que posteriormente, las viviendas seleccionadas al interior de estas áreas pasen a integrar la muestra de cada encuesta. En este sentido, es deseable que un marco cubra en forma exhaustiva a la población de interés, que se encuentre debidamente actualizado y que disponga de información relevante para la construcción de estratos de muestreo adecuados.

Si el marco muestral utilizado para algún estudio no cumple con las propiedades ideales, es muy probable que el estudio presente estimaciones sesgadas, situación generalmente ocasionada por la subestimación de la población objetivo. Si, además, la falta de cobertura no tiene un comportamiento aleatorio, se genera una subrepresentación de determinados grupos de población y, en consecuencia, la sobrerrepresentación de otros. En general, estas situaciones suelen darse cuando los marcos muestrales están obsoletos. Por ende, una solución a este problema es considerar marcos muestrales lo más actualizados posible.

En atención a lo anterior, desde el segundo semestre del año 2019, el INE pone a disposición el Marco Muestral de Viviendas construido con base en el Censo de Población y Vivienda de 2017 (MMV 2017). Además, cuenta con protocolos para llevar a cabo procesos de actualización continua y permanente, los cuales ya se han puesto en práctica y, gracias a ello, desde el segundo trimestre del año 2021 se encuentra disponible el MMV 2020, versión actualizada del MMV 2017, a partir de enumeraciones en terreno para unidades geográficas priorizadas y de verificaciones en gabinete para aquellas unidades que no han presentado cambios importantes en el último periodo, haciendo uso de todos los registros administrativos con los que cuenta el INE para tal fin.

A continuación, se describen las características del marco muestral utilizado para la selección muestral de la VII EME. Dado que las unidades seleccionadas en la VII EME provienen de la ENE, se deben revisar las características del marco de muestreo asociado a la fase 1 (ENE) y a la fase 2 (EME).

I.6.1. Características del MMV 2017

El MMV 2017 cubre 345 comunas de las 346 que conforman el territorio nacional, continental e insular. La comuna que queda excluida corresponde a la Antártica Chilena perteneciente a la región de Magallanes.

I.6.1.1. Estratificación geográfica del marco muestral

El MMV 2017 contiene variables que permiten la identificación de la división político-administrativa del país³, así como también la identificación de las áreas urbanas y rurales de cada comuna.

La definición de urbano y rural viene dada de las especificaciones del Censo de población y vivienda 2017, donde se define:

- **Entidad urbana:** asentamiento humano con continuidad y concentración de construcciones en un amanzanamiento regular con población mayor a 2.000 habitantes, o entre 1.001 y 2.000 habitantes, donde menos de 50% de la población que declara haber trabajado se dedica a actividades primarias⁴.
- **Entidad rural:** asentamiento humano concentrado o disperso que posee 1.000 o menos habitantes, o entre 1.001 o 2.000 habitantes, donde más de 50% de la población que declara haber trabajado se dedica a actividades primarias.

De las 345 comunas, 319 tienen área urbana, de estas existen 27 que contienen el estrato urbano en forma exclusiva. Por su parte, el estrato rural se encuentra en 319 comunas, de las cuales 27 lo contienen en forma exclusiva.

³ Región, provincia y comuna.

⁴ Actividades primarias: Son todas aquellas actividades económicas que consisten en la extracción de productos naturales, tales como agricultura, pesca, caza, minería, etc.

Figura I.1 Estratificación geográfica del MMV 2017



Fuente: Instituto Nacional de Estadísticas.

I.6.1.2. Conformación de la Unidades Primarias de Muestreo

El MMV 2017 está constituido por unidades primarias de muestreo, en adelante, UPM, que corresponden a áreas geográficas homogéneas en términos del número de “viviendas marco”⁵, observadas en el Censo 2017 o en su actualización más reciente; haciendo distinción del área urbana y rural. Dichas UPM están clasificadas según el área (urbana-rural) a la que pertenezcan. En el área urbana, el tamaño medio de las UPM es de 200 viviendas, con un intervalo de 20%, es decir, la cantidad de viviendas por UPM se encuentra en el rango [160-240]. Por otra parte, en el área rural las UPM tienen un tamaño medio de 90 viviendas con un intervalo de 20%, quedando conformadas por un número de viviendas en el rango [70-110] con algunas excepciones⁶.

⁵ Denominación empleada para las viviendas particulares con moradores presentes o ausentes.

⁶ Estas excepciones se definen con base en límites que fueron determinados en el contexto de la definición de umbrales de crecimiento y decrecimiento para la actualización del MMV. Sus valores pueden ser utilizados de forma referencial.

Para la construcción de las UPM, el criterio fundamental es mantener unidades compactas y homogéneas, con el objetivo de que el muestreo bietápico sea eficiente y que la muestra pueda entrar en operación en forma óptima⁷.

1.6.1.3. Estratificación socioeconómica del marco muestral

El MMV 2017, cuenta con una estratificación socioeconómica a nivel de UPM y se basa en un análisis de diversas variables recopiladas en el Censo 2017.

Esta estratificación de las UPM permite aumentar la eficiencia de la inferencia en las encuestas de hogares que utilizan este marco. En línea con lo anterior, el principal objetivo de realizar esta estratificación es inducir mayor eficiencia y precisión en las estimaciones de las estadísticas oficiales, además de servir como fuente de información para encuestas de hogares que requieren mayor análisis de carácter socioeconómico. Al estratificar se espera que las UPM dentro de un mismo estrato sean similares (disminución de la intra-varianza) y, diferentes a las de otros estratos (aumento de la inter-varianza) (Guerrero & Marín, 2020).

Para la determinación de los niveles o estratos socioeconómicos, se consideraron diversos métodos y niveles de agregación de la información. Los métodos considerados fueron algoritmos de clasificación multivariados, tales como *Principal Component Analysis* (PCA), *K-medias*, y *Principal Component Analysis by means of Alternating Least Squares* (PRINCALS).

A partir de las distintas combinaciones de técnicas, agrupaciones, cantidades de estratos y variables, se obtuvieron 180 vectores de estratificación, que consideraban entre 3 y 5 estratos. Para evaluar el desempeño de estos, se calculó el efecto de diseño de 24 indicadores, con lo cual fue posible escoger la estratificación que redujera la varianza de estos. Los indicadores seleccionados fueron:

- Tasa de ocupación por sexo
- Tasa de desocupación por sexo
- Tasa de inactividad por sexo
- Porcentaje de extranjeros
- Porcentaje de personas dentro de la fuerza de trabajo primaria respecto de la población económicamente activa (PEA) por sexo
- Porcentaje de hogares unipersonales
- Porcentaje de población femenina ocupada por rama
- Porcentaje de población masculina ocupada por rama

⁷ El uso de UPM homogéneas en tamaño, trae bondades desde el punto de vista del diseño muestral, ya que, al seleccionar un número similar de viviendas en cada una de ellas, se reduce la variabilidad de los factores de expansión, disminuyen los posibles valores atípicos y, por tanto, mejora la precisión de las estimaciones.

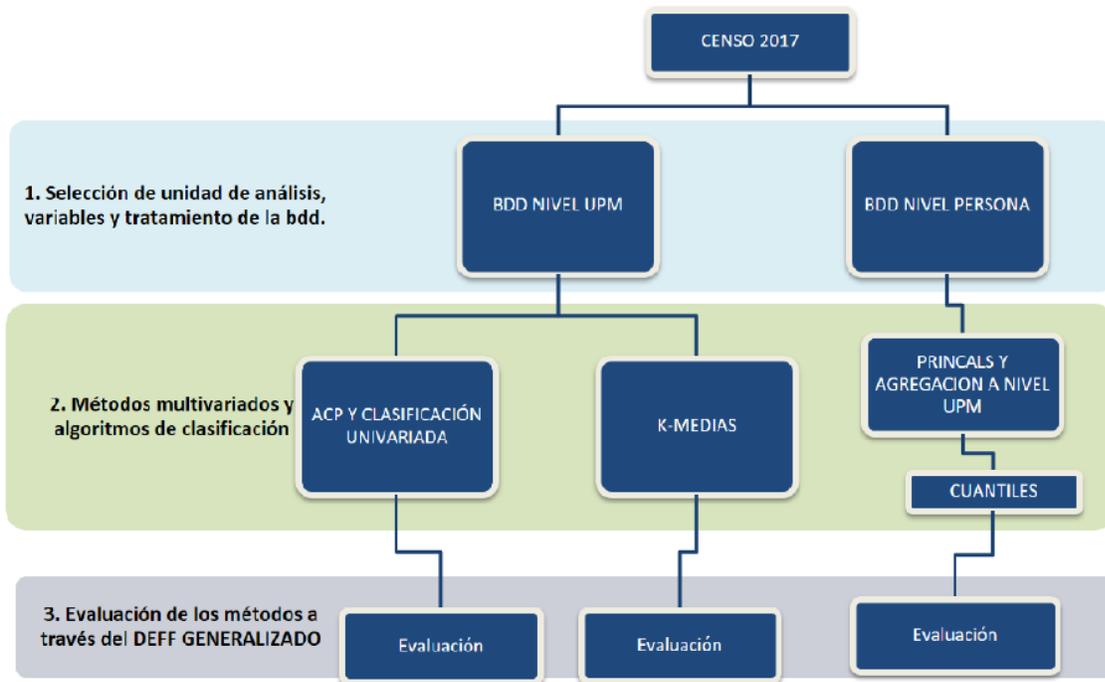
- Porcentaje de personas según nivel educacional por sexo.

Luego de la evaluación, el método utilizado para la estratificación del marco es el de estratificación óptima de la primera componente principal del PCA, que considera los siguientes indicadores:

- Porcentaje de personas en la educación superior
- Tasa de ocupación
- Porcentaje de viviendas con índice de materialidad alto
- Indicador sobre total de hijos nacidos vivos.

La elección del número de estratos (3) se basó en el análisis de las distribuciones de las UPM según el cruce de comuna y área, ya que la baja prevalencia de UPM en regiones extremas, particularmente en el área rural, hace que sea más conveniente considerar solo 3 estratos. Finalmente, se obtiene como resultado una variable de estratificación con tres categorías de nivel socioeconómico, señalada en el MMV 2017 como clasificación socioeconómica de la UPM⁸. En la Figura I.2, se observan gráficamente los pasos seguidos para generar la estratificación socioeconómica del MMV.

Figura I.2 Esquema general de la metodología de estratificación del MMV.



Fuente: Instituto Nacional de Estadísticas.

⁸ Para mayores antecedentes consultar:

[https://www.ine.cl/inicio/documentos-de-trabajo/documento/estratificaci%C3%B3n-socioecon%C3%B3mica-del-marco-muestral-de-viviendas-2017-\(mmv-2017\)-implementaci%C3%B3n-en-r](https://www.ine.cl/inicio/documentos-de-trabajo/documento/estratificaci%C3%B3n-socioecon%C3%B3mica-del-marco-muestral-de-viviendas-2017-(mmv-2017)-implementaci%C3%B3n-en-r)

1.6.1.4. Áreas especiales

En la construcción de un marco muestral, es necesario identificar áreas especiales susceptibles de exclusión. Estas áreas suelen estar afectadas por factores que dificultan su levantamiento y pueden ser de tipo climáticos, de transporte, de costo, de acceso y de acceso restringido, entre otros. Para el caso del MMV 2017, se definieron Áreas Especiales (AE).

Para estas áreas especiales, el MMV 2017 aporta información mediante ocho tipologías que dan cuenta sobre las complejidades presentes en la realización de trabajos operativos, información que resulta de gran utilidad para la coordinación de las actividades en terreno, a saber:

- 1) Tiempo de trayecto
- 2) Clima
- 3) Altitud
- 4) Transporte especial
- 5) Insularidad
- 6) Acceso Pedestre
- 7) Estado de los caminos
- 8) Ingreso restringido

Lo anterior permite calcular a nivel de UPM el porcentaje de viviendas que presentan dificultades de acceso, así como también considerar las razones específicas por las que cada una de estas ha sido identificada como Área Especial. En última instancia, lo anterior permite que la exclusión se lleve a cabo por UPM, en lugar de comunas completas como ocurría con el marco predecesor MM2016, con lo cual el MMV 2017 consigue un mayor nivel de cobertura que dicho marco muestral.

El trabajo a nivel de UPM permite también evaluar la exclusión en conjunto con los equipos regionales de modo que se posibilite la coordinación y valorización del esfuerzo operativo con información más detallada para el trabajo de campo.

1.6.2. Marco muestral de la Encuesta de Microemprendimiento

1.6.2.1. Cobertura geográfica

La cobertura es una propiedad estadística que caracteriza al marco muestral. La falta de cobertura denota el error de no incluir ciertos elementos (o unidades completas) de la población objeto de estudio, según el marco muestral que se ha definido. Este error puede ser intencionado o no. Un ejemplo de error no intencionado es la exclusión de ciertas unidades en el proceso de conteo e identificación de las viviendas, previo a la selección.

La VII EME posee un diseño muestral bifásico, por lo tanto, comparte las propiedades de cobertura de dos marcos muestrales: el primero, utilizado para la selección de las viviendas de la ENE (período MAM 2022); y el segundo; utilizado para la selección de los “Microemprendedores”.

La ENE utiliza como insumo la totalidad del marco muestral maestro. Sin embargo, debido a que se realizan exclusiones por motivos operativos, el marco de selección deja fuera 35.287 viviendas, correspondientes a 0,57% del total nacional. El proceso de exclusión se explica con mayor detalle en el apartado I.6.2.3.

Finalmente, en la elaboración del marco muestral de la VII EME, se excluyen intencionadamente todas las viviendas que no poseen un “Microemprendedor”, es decir, que no poseen unidades elegibles.

I.6.2.2. Estratificación geográfica

La estratificación del marco muestral de la ENE, correspondiente a la primera fase, tiene por objetivo crear una división geográfica que cumpla con los requisitos mínimos de UPM para disponer de suficientes unidades al realizar las rotaciones de unidades propias de la encuesta y que refleje, en alguna medida, la aglomeración espacial del mercado laboral.

Se establece como requisito para los estratos un mínimo de 36 UPM en cada uno de ellos, de modo que existan suficientes UPM para completar una rotación completa. Adicionalmente, la estratificación se realiza por separado para el área urbana y para el área rural, de modo que el segmento rural de una comuna podrá compartir estrato con otras, mientras el segmento urbano constituye un estrato por sí mismo.

Finalmente, se obtuvieron 163 estratos geográficos distribuidos en todo el territorio nacional, salvo las zonas de exclusión. Cabe mencionar que existen 11 estratos que no cuentan con el mínimo de UPM requerido, ya que no tenían otras comunas con las cuales agruparse.

En la segunda fase, la VII EME está estratificada de forma natural de acuerdo con las 16 regiones que posee el país.

I.6.2.3. Conformación del marco muestral

Con la inclusión de los estratos geográficos, el marco muestral MMV 2020 ya cuenta con toda la información necesaria para ser utilizado en la ENE, según el diseño muestral definido. Sin embargo, es necesario definir las UPM que se van a excluir del estudio. Para ello, en primer lugar, se identifica a las UPM susceptibles de exclusión, es decir, las áreas especiales del MMV 2017, como también, aquellas que pertenecen a zonas que se encontraban excluidas en el estudio previo. Para decidir cuáles de las UPM identificadas serán finalmente excluidas, el equipo técnico del INE trabajó, en coordinación con los equipos

regionales de recolección, de modo que la decisión incluya también las capacidades instaladas de cobertura⁹.

El resultado fue la exclusión de 355 UPM, que corresponde a 1% de las UPM a nivel nacional, donde diez de las 345 comunas que cubre el marco muestral maestro se ven excluidas en su totalidad, observándose una concentración relevante en la región de Los Lagos, donde la provincia de Palena se ve totalmente excluida. Asimismo, las viviendas del marco excluidas del estudio alcanzan 0,6% a nivel nacional. La Tabla I.1 muestra la distribución regional de UPM y viviendas marco, especiales y excluidas.

Tabla I.1. Cantidad de UPM especiales y excluidas, por región

Región	UPM	UPM especiales	% UPM Especiales	UPM excluidas	% UPM excluidas	Viviendas Marco	Viviendas marco especiales	% Viv. Marco especiales	Viviendas marco excluidas	% Viv. Marco excluidas
Nacional	35.149	973	2,8%	355	1,0%	6.145.493	87.831	1,4%	35.287	0,6%
Arica y Parinacota	405	9	2,2%	8	2,0%	72.602	828	1,1%	736	1,0%
Tarapacá	570	22	3,9%	13	2,3%	107.473	2.046	1,8%	1.195	1,1%
Antofagasta	965	10	1,0%	4	0,4%	186.917	878	0,4%	359	0,2%
Atacama	607	11	1,8%	6	1,0%	105.554	924	0,8%	511	0,5%
Coquimbo	1.738	45	2,6%	0	0,0%	278.060	4.030	1,4%	0	0,0%
Valparaíso	3.687	38	1,0%	28	0,8%	673.541	3.413	0,5%	3.921	0,6%
Metropolitana	12.054	17	0,1%	0	0,0%	2.352.958	1.543	0,1%	0	0,0%
O'Higgins	2.150	71	3,3%	6	0,3%	331.870	6.331	1,9%	524	0,2%
Maule	2.545	171	6,7%	5	0,2%	386.081	15.778	4,1%	457	0,1%
Ñuble	1.225	17	1,4%	8	0,7%	179.827	1.526	0,8%	726	0,4%
Biobío	3.126	75	2,4%	24	0,8%	553.538	6.709	1,2%	2.064	0,4%
La Araucanía	2.398	162	6,8%	49	2,0%	355.092	14.679	4,1%	4.494	1,3%
Los Ríos	977	77	7,9%	23	2,4%	143.005	6.937	4,9%	2.023	1,4%
Los Lagos	2.096	187	8,9%	167	8,0%	316.294	16.847	5,3%	16.895	5,3%
Aysén	272	44	16,2%	11	4,0%	41.682	3.996	9,6%	1.177	2,8%
Magallanes	334	17	5,1%	3	0,9%	60.999	1.366	2,2%	205	0,3%

Fuente: Instituto Nacional de Estadísticas.

⁹ A cada equipo regional le fue remitido el correspondiente listado de UPM susceptibles de exclusión, para que, en función de sus conocimientos del terreno determinaran si era necesario excluir la UPM o no del estudio.

Las viviendas seleccionadas y encuestadas en la ENE para el periodo MAM 2022, son utilizadas como marco de muestreo para la VII EME, pues permite identificar las viviendas donde reside al menos un microemprendedor.

En la segunda fase, se clasifican las viviendas en dos grupos, de acuerdo con si éstas contienen o no, en el período de referencia, al menos un microemprendedor. Las viviendas que no poseen microemprendedores se descartan.

Luego, se revisa detalladamente el listado de microemprendedores para descartar todas las unidades que no cumplen los criterios técnicos para ser clasificadas efectivamente como un microemprendimiento, mitigando así los problemas posteriores de levantamiento debido a la existencia de casos fuera de muestra; formando así el marco de muestreo para la VII EME, sólo con aquellas viviendas con presencia de microemprendedores.

1.6.2.4. Estratificación del marco muestral

El diseño muestral de la ENE considera información tanto de los estratos geográficos, descritos en el apartado anterior, como de la estratificación socioeconómica presente en el MMV 2020. Los estratos de muestreo se conforman a partir del cruce entre la estratificación geográfica y socioeconómica, siempre que los estratos geográficos cumplan ciertas condiciones. Se establece que, para desagregar un estrato geográfico según estratificación socioeconómica (NSE) debe:

- Contener más de 150 UPM.
- Poseer suficiente heterogeneidad en su interior en relación con la condición socioeconómica, esto es:
 - El estrato geográfico no debe agrupar más de 80% de las UPM en un solo NSE o;
 - debe presentar una proporción de UPM en torno a 40% en un NSE y, los dos restantes en conjunto, alrededor de 60%.
- Asegurar que todos los estratos resultantes cumplan con un mínimo de 36 UPM.

Según lo anterior, se obtiene un total de 232 estratos de muestreo, de los cuales 122 provienen directamente desde la estratificación geográfica, y 110 incorporan la información de la estratificación socioeconómica. La distribución regional de los estratos se muestra en la Tabla I.2.

Tabla I.2. Cantidad de estratos de muestreo por región y área

Región	Área		Total
	Urbano	Rural	
Total	153	79	232
Arica y Parinacota	3	1	4
Tarapacá	4	2	6
Antofagasta	5	2	7
Atacama	5	3	8
Coquimbo	8	5	13
Valparaíso	22	9	31
Metropolitana	33	10	43
O'Higgins	8	7	15
Maule	13	10	23
Ñuble	6	4	10
Biobío	15	5	20
Araucanía	8	4	12
Los Ríos	7	5	12
Los Lagos	11	8	19
Aysén	2	2	4
Magallanes	3	2	5

Fuente: Instituto Nacional de Estadísticas.

I.7. Estrategia muestral

La VII EME posee un diseño muestral bifásico, donde la primera fase consiste en levantar una encuesta para conocer las características de las unidades informantes y, con esto, determinar dónde se encuentra la población objetivo para la EME, obteniendo así el marco de selección para la segunda fase de la encuesta.

La primera fase del diseño (ENE) posee un muestreo probabilístico, estratificado y bietápico, donde las unidades primarias de muestreo corresponden a UPM; mientras que, las unidades secundarias de muestreo son las viviendas particulares ocupadas. Para la selección de las UPM o conglomerados, se emplea una selección aleatoria sistemática con probabilidad proporcional al tamaño dentro de cada estrato de muestreo, mientras que las unidades secundarias de muestreo se seleccionan de forma sistemática y con igual probabilidad al interior de cada UPM. Así, las unidades seleccionadas y encuestadas en la ENE para el periodo MAM 2022, son utilizadas como marco de muestreo para la VII EME pues, la información levantada, en la primera fase, permite identificar las viviendas donde reside, al menos, un microempresario.

En la segunda fase, se clasifican las viviendas en dos grupos, según si, contienen o no, en el período de referencia, al menos un microemprendedor. Las viviendas que no poseen microemprendedores se descartan, formando el marco de muestreo con sólo aquellas viviendas con presencia de personas que forman parte de la población objetivo. Este listado se revisa en profundidad para descartar todas las unidades que no cumplen los criterios técnicos para ser clasificadas efectivamente como un microemprendimiento, mitigando así los problemas posteriores de levantamiento, debido a la existencia de casos fuera de muestra.

Posterior a la depuración del listado, se seleccionan con igual probabilidad y de forma sistemática, las viviendas al interior dentro de cada estrato de muestreo y que formarán parte de la muestra. Luego, se listan todos los microemprendedores al interior de la vivienda y del hogar, y se selecciona de forma aleatoria un Microemprendedor por tipo de actividad económica, al interior del hogar. Si al interior de algún hogar dentro de la vivienda existe más de un microemprendedor que desempeña una misma actividad económica, entonces sólo se selecciona a uno de ellos para no sobrerrepresentar la actividad económica. La Tabla I.3 muestra: el total de personas clasificadas en la ENE (MAM 2022) como microemprendedores, el total de microemprendedores, utilizado como marco de selección muestral para la VII EME, luego de la depuración de la base ENE y; finalmente, el porcentaje de microemprendedores que fueron depurados para crear este marco de selección.

En total, la depuración del marco muestral corresponde a 4,8% de casos descartados por ser potenciales unidades no elegibles¹⁰, donde el mayor porcentaje de depuración se encuentra en la Región de Ñuble (7,1%) y; el menor, en la Región de Coquimbo (2,4%).

¹⁰ En la EME son unidades no elegibles aquellos individuos que en la ENE fueron clasificados como microemprendedores, según información proporcionada por informante idóneo y que, al momento de realizar el trabajo de campo se observa que la persona seleccionada no era un microemprendedor, o también en el caso que el individuo haya cambiado de estado (dejó de ser microemprendedor).

Tabla I.3. Total de microemprendedores según ENE trimestre MAM 2022 y según Marco VII EME 2022

Región	Total Microemprendedores ENE	Total Microemprendedores EME	Porcentaje Depuración
Total Nacional	9.954	9.474	4,8%
Arica y Parinacota	403	384	4,7%
Tarapacá	283	267	5,7%
Antofagasta	375	364	2,9%
Atacama	355	336	5,4%
Coquimbo	582	568	2,4%
Valparaíso	1.133	1.082	4,5%
Metropolitana	1.842	1.723	6,5%
O'Higgins	473	455	3,8%
Maule	740	709	4,2%
Ñuble	467	434	7,1%
Biobío	922	878	4,8%
La Araucanía	572	548	4,2%
Los Ríos	484	469	3,1%
Los Lagos	736	699	5,0%
Aysén	340	323	5,0%
Magallanes	247	235	4,9%

Fuente: Instituto Nacional de Estadísticas.

I.8. Cálculo y distribución del tamaño muestral

El tamaño muestral se realiza con el objetivo de obtener un tamaño que garantice estimaciones que representen eficientemente la población objetivo del estudio y que, además, sean confiables, es decir, generen márgenes de error tolerables, según el estándar de calidad INE. Para esta versión de la encuesta, se busca representatividad a nivel nacional y regional.

Para realizar el cálculo del tamaño muestral se toma la información de la ENE levantada en los trimestres MAM 2019 y MAM 2021. La información de ambos trimestres se fusiona; conformando un escenario híbrido, el cual surge de la necesidad de incorporar en el estudio un escenario en condiciones normales y otro con efecto “Pandemia”, desde donde es posible obtener una estimación de la cuasi-varianza poblacional, para obtener un tamaño muestral inicial bajo muestreo aleatorio simple monoetápico. Luego, el tamaño inicial calculado recibe una serie de ajustes asociados al efecto de diseño, ajuste por finitud de la población y el aumento del tamaño muestral objetivo para cubrir la posible pérdida de unidades por concepto de no logro.

Dado que este escenario considera dos trimestres en distintos años, el tamaño objetivo que se asigna a nivel regional (corregido por finitud) es aproximadamente el promedio entre los dos tamaños definidos en cada trimestre. Luego, el tamaño muestral objetivo a nivel nacional es de 5.959 viviendas, que permite realizar una estimación del parámetro de interés con un error absoluto esperado de 1,4% y; un error relativo de 1,6%. Con este tamaño y utilizando el promedio entre las tasas de no respuesta entre la VI EME (considera un escenario normal de levantamiento) y la ENE MAM 2021 (considera escenario con efecto pandemia) se obtiene un tamaño con sobremuestreo de 7.850 viviendas.

Para el cálculo del tamaño muestral se utiliza como parámetro de interés la proporción de microemprendedores por cuenta propia sobre el total de microemprendedores y sus estadísticos asociados, obtenidos desde la ENE en los trimestres mencionados anteriormente. En el Cuadro I.1, se presentan definiciones y consideraciones utilizadas en la obtención del tamaño muestral.

Cuadro I.1. Criterios utilizados en el cálculo del tamaño muestral

Parámetro	Descripción
Variable de diseño	I : Variable Bernoulli que considera los siguientes valores $I = \begin{cases} 0, & \text{si el microemprendedor no es cuenta propia} \\ 1, & \text{si el microemprendedor es cuenta propia} \end{cases}$
Parámetro asociado	Proporción de microemprendedores por cuenta propia sobre el total de microemprendedores en la región r (p_{r0})
Estimador asociado	Estimador de razón: $r = \frac{\text{Estimación del número de microemprendedores por cuenta propia en la región}}{\text{Estimación del número de microemprendedores en la región}}$
Niveles de estimación	Nacional. Regional.
Errores de muestreo	Nacional: El error absoluto no debe superar 1,4% y; el error relativo 1,6%. Regional: Los errores absolutos no superan 8,1% y; los errores relativos 10,5%.

Fuente: Instituto Nacional de Estadísticas.

I.8.1. Metodología de cálculo del tamaño muestral

Para calcular el tamaño muestral se ejecutan varios pasos que utilizan una serie de nomenclaturas las cuales, se definen a continuación:

- r : Subíndice que identifica a la región.
- p_{r0} : Parámetro de interés en la región r .
- \hat{p}_{r0} : Estimación del parámetro de interés en la región r .
- \bar{m}_{r0} : Número promedio de viviendas a encuestar por unidad primaria de muestreo en la región r en la ENE trimestre MAM 2019 y MAM 2021.

n_{r0}	: Número de UPM levantadas en trimestre MAM 2019 y MAM 2021 (ENE) en la región r .
$n_{r0} \cdot \bar{m}_{r0} = m_{r0}$: Número de viviendas logradas en trimestre MAM 2019 Y MAM 2021 (ENE) en la región r .
λ	: Porcentaje de selección asignado para cada región (70,0% o 80,0%).
$SE(\hat{p}_{r0})$: Error estándar de la estimación de la proporción de microemprendedores por cuenta propia en la región r . Corresponde a la raíz cuadrada de la varianza de la estimación.
$t_{1-\alpha/2, v-1}$: Cuantil de una distribución t de Student con nivel de confianza $1-\alpha$ (para efectos de la simulación se utiliza un valor fijo $t=2$).

A continuación, se describe en detalle cada una de las etapas realizadas para definir el número de viviendas a encuestar en cada región r .

Etapas 1:

Utilizando la fusión de las bases de datos del levantamiento de la ENE en los trimestres MAM 2019 y MAM 2021, se obtienen las estimaciones a nivel regional de: parámetro de interés regional (\hat{p}_{r0}), error estándar asociado ($SE(\hat{p}_{r0})$) y total de unidades logradas (m_{r0}). La cuasi-varianza poblacional $S(p_{r0})^2$ se calcula a partir del error estándar obtenido, utilizando la siguiente ecuación:

$$S(p_{r0})^2 = \frac{n \cdot \bar{m}_{r0}}{Def(p_{r0})} \cdot SE(p_{r0})^2 \quad (1)$$

Etapas 2:

A partir del error estándar, se obtienen los errores absolutos por región (ea_{0r}) y luego, los errores relativos (e_{0r}). Posteriormente, se establece el nivel de confianza de 95% para obtener el cuantil de una distribución t de Student que, para efectos de la simulación, se utiliza un valor de $t=2$. Con estos y la cuasi-varianza del parámetro de interés derivada en (1), se obtienen los tamaños muestrales iniciales a nivel regional considerando un muestreo aleatorio simple monoetápico. Este cálculo se observa en la ecuación (2):

$$m_{0,r} = \frac{2^2 \cdot S_r^2}{ea_{r1}^2} \quad (2)$$

Donde,

$m_{0,r}$: Número de viviendas a nivel regional, bajo muestreo aleatorio simple.

Etapa 3

Debido a que el diseño muestral es complejo, es necesario considerar la incorporación del efecto diseño, dado que el muestreo complejo implica una pérdida de eficiencia en las estimaciones con respecto al muestreo aleatorio simple. El efecto del diseño utilizado corresponde al efecto considerado para el cálculo de las simulaciones de tamaño muestral del nuevo diseño de la ENE (ENE 2020) donde se propuso un efecto fijo (*Def*) de 1,7 para cada región. Por esta razón, el tamaño de muestra inicial, bajo muestreo aleatorio simple, se incrementa a partir del efecto de diseño según la siguiente ecuación (3):

$$m_{1,r} = m_{0,r} \cdot Def \quad (3)$$

Donde,

$m_{1,r}$: Número de viviendas a nivel regional ajustado por efecto de diseño.

Def : Efecto de diseño fijo para cada región.

Este tamaño corresponde a 70,0% de las viviendas disponibles en el marco de selección, donde reside al menos un microempresario, con excepción de algunas regiones, donde se asigna 80,0%. Este aumento de la muestra se realiza con el objetivo de disminuir el error muestral.

Etapa 4

Los tamaños muestrales regionales obtenidos, deben corregirse por un ponderador que da cuenta de la condición de población finita (fracción de muestreo, *FRm_r*). Esta fracción se obtiene a partir de la ecuación (4):

$$FRm_r = \frac{m_{1r}}{MM_r} \quad (4)$$

Etapa 5

Obtenida la fracción de muestreo a nivel regional, se determina el tamaño muestral “corregido por finitud”, según la ecuación (5):

$$m_{2,r} = m_{1,r} \cdot \left(\frac{1}{1 + \frac{m_{1,r}}{M_r}} \right) \quad (5)$$

Etapa 6

Una vez obtenidos los tamaños objetivos regionales, se aplica un factor que corresponde a la tasa de no respuesta (*tnr*). Dicha tasa se obtiene a partir del promedio entre la tasa de no respuesta de la VI EME y el trimestre ENE MAM 2021, y es considerada con el propósito de contrarrestar la pérdida de muestra por no respuesta por parte de los encuestados.

La fórmula para calcular el tamaño de muestra con sobremuestreo viene dada por la siguiente ecuación (6):

$$m_{3,r} = \frac{m_{2,r}}{1 - tnr} \quad (6)$$

$m_{3,r}$: Número de viviendas con sobremuestreo a nivel regional.

1.8.2. Errores esperados

La Tabla I.4 presenta los errores esperados, asociados a la estimación del parámetro de interés, para un tamaño muestral objetivo de 7.850 viviendas¹¹ a nivel nacional.

Tabla I.4. Errores esperados asociados al parámetro de interés según tamaño muestral objetivo

Nivel	Estimación Microemprendedores MAM 2019-2021	Efecto diseño propuesto	Error absoluto propuesto VII EME	Error relativo propuesto VII EME	Tamaño objetivo ajustado VII EME	Tamaño con sobremuestreo VII EME
Nacional	86,9%	1,7	1,4%	1,6%	5.959	7.850
Arica y Parinacota	88,8%	1,7	4,9%	5,5%	239	298
Tarapacá	86,5%	1,7	7,8%	9,1%	197	229
Antofagasta	87,7%	1,7	5,4%	6,2%	202	251
Atacama	86,9%	1,7	5,2%	6,0%	214	268
Coquimbo	85,6%	1,7	4,9%	5,8%	370	493
Valparaíso	85,2%	1,7	3,6%	4,2%	681	938
Metropolitana	87,2%	1,7	2,9%	3,3%	1.030	1.472
O'Higgins	86,2%	1,7	5,0%	5,8%	316	415
Maule	86,7%	1,7	4,0%	4,6%	418	535
Ñuble	81,8%	1,7	6,3%	7,7%	240	292
Biobío	89,3%	1,7	2,9%	3,3%	602	775
La Araucanía	89,0%	1,7	4,8%	5,4%	360	483
Los Ríos	86,4%	1,7	5,3%	6,2%	262	315
Los Lagos	86,3%	1,7	5,0%	5,8%	461	619
Aysén	86,1%	1,7	4,4%	5,1%	222	286
Magallanes	77,3%	1,7	8,1%	10,5%	145	181

Fuente: Instituto Nacional de Estadísticas.

¹¹ La estimación de los errores se obtiene a partir del tamaño objetivo y no del tamaño con sobremuestreo, pues dadas las pérdidas naturales de unidades muestrales, se espera obtener un tamaño de 5.959 viviendas.

1.8.3. Distribución de la muestra según submuestra

Una vez obtenido el tamaño muestral requerido para cada región, se debe distribuir a nivel de submuestra. Dado que la muestra ENE está subdividida en tres meses o períodos de levantamiento, con el objetivo de disminuir el tiempo transcurrido entre el levantamiento de información de la ENE y la EME y, con ello tener una menor atrición¹², la muestra de la VII EME se distribuye independientemente en tres meses de levantamiento: mayo, junio y julio, de acuerdo con el mes de levantamiento de la ENE, marzo, abril y mayo, respectivamente.

Según lo anterior, el tamaño regional se distribuye en tres partes iguales, o lo más similares posibles. En la Tabla I.5 se observa la distribución de la muestra según mes de levantamiento y región.

Tabla I.5. Total de viviendas seleccionadas según región y mes de levantamiento

Mes de levantamiento VII EME				
Nivel	Mayo	Junio	Julio	Total viviendas seleccionadas
Nacional	2.683	2.575	2.592	7.850
Arica y Parinacota	92	107	99	298
Tarapacá	77	78	74	229
Antofagasta	84	84	83	251
Atacama	90	88	90	268
Coquimbo	172	139	182	493
Valparaíso	340	319	279	938
Metropolitana	519	464	489	1.472
O'Higgins	143	138	134	415
Maule	179	178	178	535
Ñuble	98	97	97	292
Biobío	259	243	273	775
La Araucanía	165	159	159	483
Los Ríos	105	105	105	315
Los Lagos	209	215	195	619
Aysén	90	101	95	286
Magallanes	61	60	60	181

Fuente: Instituto Nacional de Estadísticas.

¹² Acumulación de pérdida de información por la no respuesta que se presenta en estudios sobre unidades en el tiempo por parte de los participantes (Jones, Koolman, & Rice, 2006).

1.8.4. Errores observados

La Tabla I.6 presenta los errores observados, a nivel nacional y regional, asociados a la prevalencia del parámetro de interés, para un tamaño muestral efectivo de 6.492 viviendas, según levantamiento de la VII EME.

Tabla I.6. Errores observados asociados al parámetro de interés según número de viviendas que responden

Nivel	Estimación VII EME	Tamaño efectivo	Error absoluto	Error relativo
Total nacional	88,8%	6.492	1,1%	1,2%
Arica y Parinacota	88,4%	240	5,4%	6,1%
Tarapacá	85,2%	203	6,4%	7,5%
Antofagasta	92,9%	218	3,5%	3,8%
Atacama	88,0%	221	5,1%	5,8%
Coquimbo	85,0%	410	4,6%	5,4%
Valparaíso	87,7%	698	2,9%	3,3%
Metropolitana	89,4%	1.201	2,1%	2,3%
O'Higgins	85,4%	344	4,4%	5,1%
Maule	89,3%	442	3,6%	4,1%
Ñuble	86,9%	268	4,3%	5,0%
Biobío	89,9%	639	3,0%	3,3%
La Araucanía	91,8%	383	3,3%	3,6%
Los Ríos	85,0%	271	4,8%	5,6%
Los Lagos	87,9%	540	3,6%	4,1%
Aysén	88,8%	242	4,6%	5,2%
Magallanes	81,9%	172	6,6%	8,0%

Fuente: Instituto Nacional de Estadísticas.

1.9. Selección de unidades muestrales

La selección de unidades muestrales se realiza en 2 etapas. En la primera etapa se seleccionan las viviendas que contienen al menos un microemprendedor y luego, en la segunda etapa se seleccionan los microemprendedores al interior de los hogares que contienen las viviendas seleccionadas.

1.9.1. Selección de viviendas

La selección de viviendas se realiza de forma sistemática con igual probabilidad de selección, para todas las viviendas al interior de cada región.

La selección de viviendas es implementada mediante el paquete “Sampling” del software R, fijando una semilla aleatoria, a fin de que pueda ser replicado en cualquier momento.

Para calcular la probabilidad de inclusión de la j – ésima vivienda dentro de la r – ésima región, se definen:

$m_{3,r}$: Número de viviendas a seleccionar en la región r .

M_r : Número de viviendas que contiene el marco de selección en la región r .

Y para la selección de viviendas ($m_{3,r}$) el *software* ejecuta los siguientes pasos:

Paso 1

En primera instancia, se ordenan en forma ascendente todas las viviendas según región, estrato, comuna, id_upm, y número de orden de vivienda dentro de la upm.

Paso 2

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

$$k = M_r / m_{3,r}.$$

Paso 3

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

Paso 4

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

La primera vivienda seleccionada es "A" y es un número entero, la segunda es el redondeo de "A + k ", la tercera es el redondeo de "A + 2 k " y así sucesivamente, hasta la $m_{3,r}$ selección, dada por el redondeo de "A + ($m_{3,r} - 1$) k ".

Luego, la probabilidad de inclusión de la j – ésima vivienda dentro de la r – ésima región, es igual a:

$$P_r(j|r) = \frac{m_{3,r}}{M_r} \quad (7)$$

I.9.2. Selección de microemprendedores

Una vez seleccionadas las viviendas, dentro de cada una de ellas y al interior de sus hogares, se identifica a los microemprendedores y las actividades económicas en que estos se desenvuelven. Luego, se seleccionan de forma aleatoria y con igual probabilidad, tantos microemprendedores como actividades distintas identificadas dentro del hogar, es decir, en caso de encontrar más de un microemprendedor en el hogar ejecutando la misma actividad económica, se debe seleccionar sólo a un representante por actividad dentro del hogar.

II. DESARROLLO DE FACTORES DE EXPANSIÓN

El factor de expansión se interpreta como la cantidad de unidades en la población que representa una unidad de la muestra, y es calculado como el inverso de la probabilidad de selección de las unidades de muestreo. Atendiendo al diseño muestral de la VII EME, las probabilidades de selección asociadas a los microempresarios presentan varias componentes:

1. Probabilidad de que la vivienda haya sido seleccionada y contestara en la ENE en el trimestre MAM 2022.
2. Probabilidad de seleccionar una vivienda para la EME, dado que la vivienda posee al menos un microempresario.
3. Probabilidad de seleccionar un microempresario, dado que su vivienda fue seleccionada.

La metodología de cálculo de los factores de expansión consiste en la aplicación secuencial de seis ponderadores o ajustes:

1. Ponderador Base.
2. Ajuste por elegibilidad desconocida.
3. Ajuste por no elegibilidad.
4. Ajuste por no respuesta.
5. Suavizamiento
6. Calibración.

II.1. Ponderador Base

El ponderador Base se define como el factor de expansión obtenido de las probabilidades de selección, sin ajustes ni correcciones, de las viviendas en la fase 1, y la selección de los microemprendedores en la fase 2, condicional a que la vivienda de residencia fue seleccionada y lograda en la ENE en el trimestre MAM 2022.

Según lo anterior, un insumo fundamental del ponderador base son los factores de expansión de las viviendas de la ENE, que dan cuenta de la probabilidad de que una vivienda haya sido seleccionada y entrevistada en la encuesta.

En la sección II.1.1 se muestra las probabilidades de selección y respuesta de la ENE; mientras que la sección II.1.2 se muestra la fórmula explícita de la probabilidad condicional de selección de un microemprendedor.

II.1.1. Probabilidad de selección y entrevista de las viviendas en la ENE- Trimestre MAM 2022

El diseño muestral de la Encuesta Nacional de Empleo corresponde a un muestreo probabilístico, estratificado y bietápico, donde las unidades primarias son las UPM; mientras que, las unidades de segunda etapa son las viviendas particulares.

Las unidades primarias de muestreo son seleccionadas en forma proporcional al tamaño, mientras que al interior de cada UPM las unidades secundarias se seleccionaron de forma sistemática y con igual probabilidad de selección. El factor de expansión que se extrae desde la ENE para la muestra de la VII EME es el factor de expansión ajustado por no respuesta para aquellas viviendas que contestan la ENE en el periodo MAM 2022 llamado w_{hij}^{NR} ¹³.

II.1.2. Probabilidad de selección de los microemprendedores

La selección de los microemprendedores se realiza en dos etapas. Primero, se seleccionan con igual probabilidad las viviendas que contienen al menos un microemprendedor, según submuestra (mes de levantamiento en la ENE) al interior de cada región.

Así, la probabilidad de seleccionar una vivienda que posee al menos un microemprendedor viene dada por la ecuación (8):

¹³ Mas información en https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/publicaciones-y-anuarios/separatas/contingencia-covid-19/separata-t%C3%A9cnica-n-4-contingencia-covid-19.pdf?sfvrsn=37f5a108_6.

$$p_{Rj}^v = \frac{m_R^{micro}}{M_R^{micro}} \quad (8)$$

Donde:

R : Subíndice que representa la región de pertenencia. $R = 1, \dots, 16$.

j : Subíndice que representa la vivienda j .

p_{Rj}^v : Probabilidad de seleccionar la vivienda j perteneciente a la región R , según el listado de viviendas del marco EME, que poseen al menos un microempresario.

M_R^{micro} : Total de viviendas seleccionadas del marco EME, provenientes del periodo MAM de la ENE, con al menos un microempresario en la región R .

m_R^{micro} : total de viviendas seleccionadas desde el marco de la EME con al menos un microempresario en la región R .

Una vez seleccionada la vivienda se seleccionan los microempresarios. La probabilidad de seleccionar al microempresario k al interior de la vivienda j , del hogar l y rama de actividad m , perteneciente a la región R , dado que la vivienda fue seleccionada, está determinada por la ecuación (9):

$$p_{Rjklm}^{micro|v} = \frac{S_{Rjlm}^{micro}}{T_{Rjlm}^{micro}} \quad (9)$$

Donde:

T_{Rjlm}^{micro} : Total de microempresarios identificados en la EME, en la vivienda j , hogar l , rama de actividad m , perteneciente a la región R .

S_{jlm}^{micro} : Total de microempresarios seleccionados, en la vivienda j , hogar l , rama de actividad m , perteneciente a la región R .

Luego, la probabilidad condicional de seleccionar el microempresario k , en la vivienda j , de la región R , viene dada por la ecuación (10):

$$p_{Rjk}^{micro} = p_{Rj}^v \cdot p_{Rjklm}^{micro|v} \quad (10)$$

Luego, el ponderador base es obtenido a partir del factor ajustado por no respuesta en el trimestre MAM de la ENE y la probabilidad de seleccionar un microempresario desde la EME. El ponderador base está definido por la ecuación (11):

$$F_{Rjk}^{base} = w_{hij}^{NR} \cdot \left(\frac{1}{p_{Rjk}^{micro}} \right) \quad (11)$$

II.2. Ajuste por elegibilidad

A partir de la hoja de ruta obtenida del levantamiento, es posible visualizar diferentes estados de elegibilidad de las unidades seleccionadas (elegibles, no elegibles y elegibilidad desconocida). Estos estados se obtienen según el código de disposición final (CDF¹⁴) al momento del levantamiento, y se definen de la siguiente manera:

Unidades elegibles: Corresponde a unidades en que el encuestador pudo determinar el estado de elegibilidad de la unidad, incluyendo tanto entrevistas logradas como no logradas.

Unidades no elegibles: Son unidades que, en el trabajo de campo han sido identificadas como: negocios, viviendas colectivas, deshabitadas, de veraneo, destruidas, etc. y, por lo tanto, no forman parte de las edificaciones definidas en la población objetivo (viviendas particulares).

Unidades de elegibilidad desconocida: Corresponden a edificaciones donde no se pudo determinar su estado de elegibilidad. Casos como este pueden ser: Unidades que nunca fueron enviadas a terreno o viviendas a las cuales no se pudo llegar o encontrar su dirección u otros casos similares.

El ponderador base contiene valores válidos para las unidades elegibles, no elegibles y de elegibilidad desconocida. No obstante, se hace necesario realizar ajustes de elegibilidad desconocida y de no elegibilidad con el fin de obtener sólo valores para los microemprendedores que residen en viviendas elegibles.

II.2.1. Ajuste por elegibilidad desconocida

Inicialmente, todas las unidades tienen un estatus conocido de elegibilidad debido a que el marco de selección es el conjunto de microemprendedores entrevistados en la ENE en el periodo marzo, abril y mayo 2022.

No obstante, durante el trabajo de campo, este estado no se puede comprobar en algunas unidades, clasificándolas como elegibilidad desconocida. Los pesos de las personas residentes en estas viviendas son eliminados y redistribuidos en las personas que residen en las viviendas que, si se conoce su elegibilidad, proporcionalmente al interior de cada estrato.

El ponderador que contiene los pesos a redistribuir $R_{h,known}$, se calcula como la razón entre la estimación del total de personas que residen en viviendas de la población y la estimación del total de personas que residen en viviendas con elegibilidad conocida (viviendas elegibles y no elegibles). Ambas estimaciones se realizan con el ponderador de selección de microemprendedor, es decir,

¹⁴ Detalle de los códigos de disposición final en Anexo Nro. 1 del presente documento.

$$R_{h, know} = \frac{\sum_{j \in \Omega_R} \sum_{k \in \Theta_j} F_{Rjk}^{base}}{\sum_{j \in \Omega_R} \sum_{k \in \Theta_{j, know}} F_{Rjk}^{base}} \quad (12)$$

Donde,

- Θ_j Es el conjunto de micro emprendedores k seleccionados al interior de cada vivienda j en la región R .
- $\Theta_{j, know}$ Es el conjunto microemprendedores seleccionados en la vivienda j clasificada como elegibles o no elegibles.
- Ω_R Es el conjunto de estratos h de la EME (regiones).
- F_{Rjk}^{base} Es el ponderador de selección del k – ésimo microempendedor de la vivienda j ubicado en la región R .

Esta razón es equivalente para todas las unidades muestrales pertenecientes a un mismo estrato.

Luego, el ponderador de selección de microempendedor F'_{Rjk} corregido por elegibilidad desconocida F'_{Rjk} está dado por:

$$F'_{Rjk} = R_{R, know} \cdot F_{Rjk}^{base} \quad (13)$$

II.2.2. Ajuste por no elegibilidad

Originalmente, la muestra de la Encuesta de Microemprendimiento contempla 8.577 personas residentes en viviendas particulares. De este total, resultaron 17 personas con elegibilidad desconocida, una vez finalizado el trabajo de campo. De las 8.560 personas restantes 8.251 resultaron elegibles (logradas y no logradas) y; 309, no elegibles (ver Tabla II.1).

Dado el objetivo de generar inferencias hacia la población de microemprendedores que reside en viviendas particulares (elegibles) en esta etapa, se realiza un ajuste que consiste en descartar aquellas personas que no conforman la población objetivo de la encuesta; asignándoles un valor blanco o “missing” en el ponderador.

Tabla II.1. Distribución de individuos según la clasificación de elegibilidad.

Clasificación elegibilidad	Cantidad personas	Porcentaje
Elegible	8.251	96,20%
No elegible	309	3,60%
Elegibilidad desconocida	17	0,20%
Total	8.577	100,00%

Fuente: Instituto Nacional de Estadísticas.

II.3. Ponderador ajustado por falta de respuesta

A partir de esta etapa se trabaja con una base que contiene sólo microemprendedores (unidades elegibles) donde es posible observar casos que manifiestan su voluntad de no participar en la encuesta o no se encuentran en la vivienda al momento del levantamiento, entre otras cosas, dando origen a la no respuesta.

La ausencia de respuestas debe ser corregida con la finalidad de reducir sesgos provocados por este tipo de errores no muestrales. Sin embargo, esta corrección debe realizarse sólo para algunos casos, tales como: el informante rechazó la entrevista; la vivienda de residencia del informante se encuentra sin moradores presentes en todas las visitas efectuadas; entre otro.

El ajuste consiste en aplicar un ponderador con el objetivo de lograr que las unidades que no responden sean representadas por las que sí, previendo no introducir sesgo debido a la posibilidad de que exista relación entre la no respuesta y la variable de interés. De un total de 7.850 viviendas seleccionadas, se seleccionaron 8.577 microemprendedores. De éstos, 8.251 fueron clasificados como elegibles (96,2%) de los cuales 6.942 respondieron la encuesta. Por lo tanto, la tasa de respuesta de la EME ajustada por elegibilidad es 84,1%.

El método utilizado para compensar la falta de respuesta es el de estratificación mediante “propensity score”. De acuerdo, a lo indicado por Valliant, Dever, & Kreuter (2013), este método consiste en modelar la probabilidad de respuesta en la VII EME como la realización de un proceso de variables latentes ($R_i^* = x_i^T \beta + u_i$), es decir, un conjunto de variables que inciden en la “motivación” (R^*) de participar o no en la encuesta. Así, mediante un conjunto de variables conocidas para quienes responden y quienes no responden se busca estimar la probabilidad de responder en la encuesta ($P(R_i^* > \theta)$) y posteriormente, ordenar de menor a mayor estas probabilidades conformando 6 grupos de igual frecuencia denominados sextiles.

El modelo que se utiliza para realizar el ajuste por no respuesta es el modelo logístico y se resume según los siguientes pasos:

1. Determinar las variables que se incluirán en el modelo de regresión logística con el cual se realizará la predicción de la probabilidad de respuesta de una persona elegible.
2. A través de este modelo, calcular la probabilidad de responder de cada una de las unidades elegibles que fueron utilizadas.
3. Ordenar las unidades de menor a mayor, según la probabilidad estimada.
4. Crear los estratos o “celdas de ajustes” donde se realizarán las correcciones de no respuesta.

Una vez creadas las celdas de ajustes¹⁵, se procede a estimar el ponderador ajustado por falta de respuesta, dado por la ecuación (14):

$$\hat{R}_c^{NR} = \frac{\sum_{k \in S_c} F'_{Rjk}{}^r}{\sum_{k \in S_{c,R}} F'_{Rjk}{}^r} \quad (14)$$

Donde:

- c : Subíndice de la celda de ajuste por falta de respuesta. $c = 1, \dots, 6$.
- S_c : Total de microemprendedores seleccionados y elegibles en la celda c .
- $S_{c,R}$: Total de microemprendedores seleccionados en la celda c y que responde la encuesta.
- $F'_{Rjk}{}^r$: Estimación del total de microemprendedores que responden, utilizando el ponderador ajustado por elegibilidad.
- $F'_{Rjk}{}^r$: Corresponde a la estimación de microemprendedores utilizando el ponderador ajustado por elegibilidad para la persona k , de la vivienda j , en la región R .

Luego, la expresión del ponderador ajustado por no respuesta, está dada por la ecuación (15):

$$F'_{Rjk}{}^{NR} = F'_{Rjk}{}^r \cdot \hat{R}_c^{NR} \quad (15)$$

De acuerdo con la metodología antes expuesta, son 6 las celdas en las cuales se realizan los ajustes por falta de respuesta. En la Tabla II.2 se presentan las tasas de respuesta para cada una de estas celdas, así como también el factor de ajuste por no respuesta (\hat{R}_c^{NR}).

Tabla II.2. Total de unidades elegibles, unidades que responden y tasas de respuesta

Celda Ajuste	Total Elegibles	Total Responde	Tasa de Respuesta	\hat{R}_c^{NR}
Total	8.251	6.942	84,1%	1,04
1	1.376	768	55,8%	1,82
2	1.375	1.099	79,9%	1,27
3	1.375	1.189	86,5%	1,14
4	1.375	1.267	92,1%	1,08
5	1.375	1.290	93,8%	1,07
6	1.375	1.329	96,7%	1,04

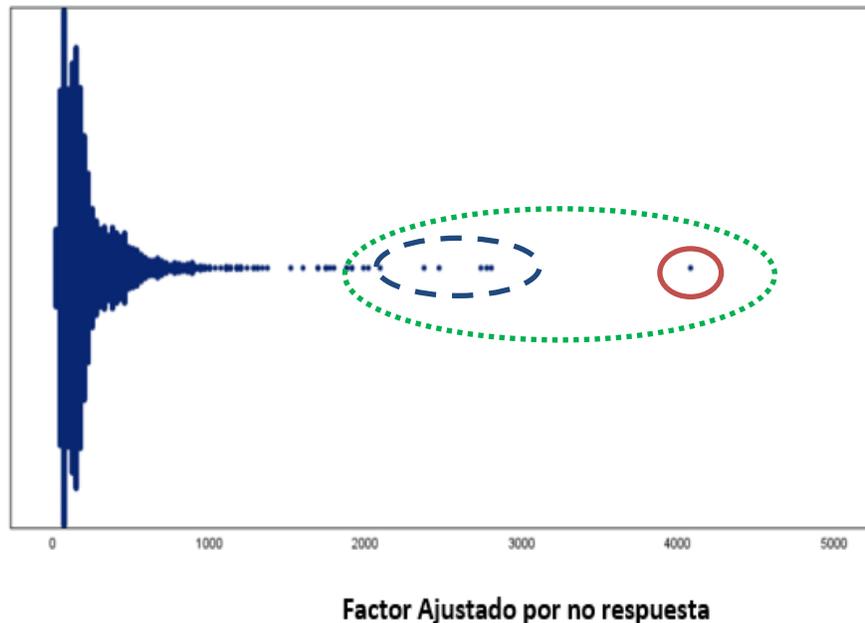
Fuente: Instituto Nacional de Estadísticas.

¹⁵ Las celdas de ajuste son agrupaciones de individuos que poseen características similares asociadas a responder la encuesta. En el caso específico de la EME, se generaron 6 celdas de ajuste denominadas "sextiles" de respuesta, mediante un modelo logístico, en que se predice la probabilidad de responder, y en base a esta probabilidad ordenada en forma ascendente, se generan seis grupos de igual cantidad de individuos.

II.4. Suavizamiento del Ponderador ajustado por no respuesta

Obtenido el factor de microemprendedores ajustado por no respuesta, se evalúa la pertinencia de realizar suavizamiento. En la Figura II.1, al observar los factores ordenados, permite identificar valores más altos que el resto de los ponderadores.

Figura II.1 Dispersión del Ponderador ajustado por falta de respuesta



Fuente: Instituto Nacional de Estadísticas

Según la figura II.1 se decide realizar el suavizamiento del ponderador ajustado por falta de respuesta. Este suavizamiento se realiza por región y rama de actividad económica.

El método utilizado para el suavizamiento del factor corresponde a un método mixto (MM) conformado por el método de contracción a la media (CM) y el método R-K.

II.4.1. Método de contracción a la media (CM).

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

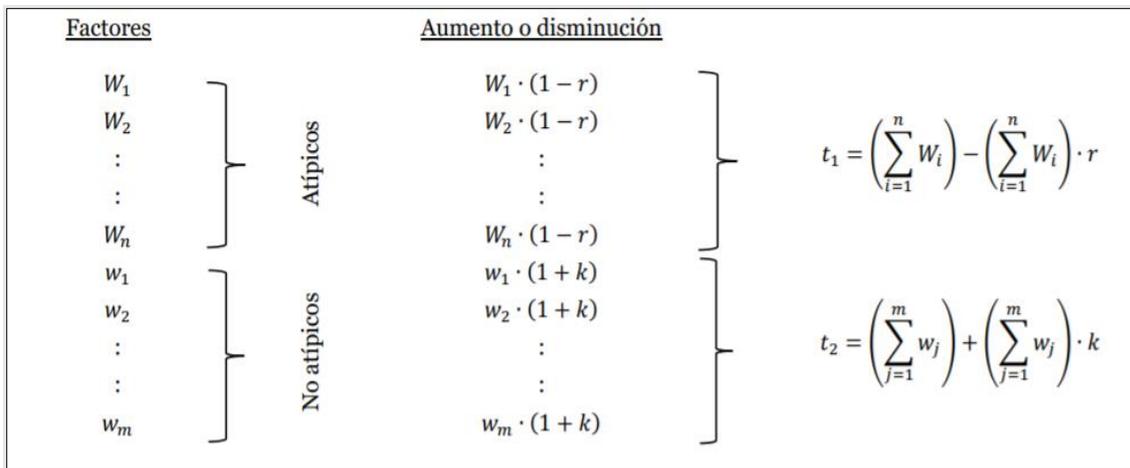
$$w_i^S = \beta \cdot w_i + (1 - \beta) \cdot \bar{w}; \quad 0 \leq \beta \leq 1 \quad (16)$$

Donde β es un parámetro real que permite controlar la intensidad de la contracción, w_i el factor i – *esimo* y \bar{w} el promedio de los factores a nivel de estrato región, basado en la existencia de valores atípicos o no.

II.4.2. Método r-k

Este método se caracteriza por dar tratamiento diferenciado a los conjuntos de datos atípicos y no atípicos, ya que genera un aumento proporcional en los factores no atípicos, mientras que, origina una disminución, también proporcional, de los factores atípicos. Por lo que, este método, disminuye en una proporción r los valores de los factores de expansión atípicos y, por otro lado, aumenta en una proporción k a los valores de los factores de expansión no atípicos, según muestra la siguiente figura:

Figura II.2 Factores de expansión w_i ordenados de mayor a menor.



Fuente: Instituto Nacional de Estadísticas

II.4.3. Método mixto

Para aplicar el método mixto se separan los factores de expansión en dos subconjuntos (factores no atípicos y factores atípicos). Para el conjunto de los factores no atípicos, se aplica el método de contracción a la media considerando β relativamente cerca de la unidad ($\beta = 0,8$). Para el conjunto de los factores atípicos, se considera $\beta = 0,2$ para que, en caso de tener un solo valor atípico al interior de un estrato, el método CM no tenga efecto sobre este valor, conservando el valor del factor atípico. Posterior a este cálculo, se aplica el método R-k, en donde se juntan ambas distribuciones de factores (no atípicos y atípicos) considerando valores del multiplicador cercano a la unidad ($\lambda = 0,99$) para conservar una pequeña distancia entre el mayor de los factores no atípicos, y el menor de los factores atípicos, después del suavizamiento. Posteriormente, el ponderador ajustado por falta de respuesta suavizado, se utiliza como insumo para realizar la calibración de este factor.

II.5. Calibración

En general, en todas las encuestas de hogares el ponderador final o factor de expansión es calibrado, con el objetivo de alcanzar algún stock poblacional obtenido de una fuente externa a la encuesta. Por ejemplo, los factores de expansión de la Encuesta Nacional de Empleo son calibrados, cada trimestre móvil, al total de población según ciertas características al día 15 del mes central correspondiente a cada periodo de levantamiento.

En el ejemplo expuesto, la población objetivo corresponde a personas que poseen ciertos atributos demográficos, cuantificados en los Censos de Población y Vivienda, lo que permite obtener una estimación de la población desagregada a esos niveles. Para la EME en cambio, existe un inconveniente, ya que no existe una estimación “oficial” o de referencia, respecto a los “microemprendedores” (formales e informales) a nivel del país.

Por otro lado, la muestra seleccionada en la VII EME, está anclada a la población de referencia del trimestre MAM 2022 de la ENE, lo cual implica que la EME hace un seguimiento a los microemprendedores que se encontraban en ese período en particular clasificados en esa categoría, sin tomar en cuenta los flujos de entrada a esa condición laboral.

Dado lo anterior, se decide utilizar la estimación del total de microemprendedores del trimestre MAM 2022, obtenido de la ENE, actualizada al período del trabajo de campo de la VII EME. Para esto, se utiliza el crecimiento proyectado para el mes central del período de levantamiento de la encuesta, es decir junio 2022.

En definitiva, la estimación utilizada en la calibración del ponderador de la EME se obtiene a partir de los siguientes pasos:

1. Primero, se considera toda la información (referente a microemprendedores) levantada para la ENE en el período MAM 2022, es decir, todos los microemprendedores que fueron clasificados como tales, sin importar que, a futuro puedan cambiar de condición.
2. Se debe obtener la estimación del total de los microemprendedores a partir del levantamiento de la ENE al mes de abril y proyectarlas al mes de junio de 2022 que corresponde al mes central del levantamiento de la EME a nivel de región, área y sexo.
3. Se calcula un nuevo factor de expansión F_{hij}^{2cal} , obtenido según el punto 2, como se observa en la ecuación (17):

$$F_{hij}^{2cal} = F_{hijk}^{cal} \cdot \frac{P_{hs}^{junio}}{P_{hs}^{abril}} \quad (17)$$

Donde:

F_{hijk}^{cal} : Factor calibrado del periodo MAM 2022 de la ENE.

P_{hs}^m : Total de población por sexo y tramo de edad, del estrato ENE h , proyectada al 15 del mes m (abril) y a junio de 2022.

Cabe destacar, que, para efectos de generar estos nuevos stocks, F_{hij}^{2cal} , estos poderadores son redondeados a números enteros.

La Tabla II.3, presenta el total de microemprendedores estimado a partir de la publicación de la ENE en el periodo MAM 2022; y según el total de personas estimadas con la información levantada en MAM 2022, pero con proyecciones actualizadas a la fecha de levantamiento de la EME. En esta tabla, se observa que, el total de “microemprendedores” estimados y publicados oficialmente son 2.086.024 personas. Sin embargo, al actualizar las proyecciones de población este total asciende a 2.089.383, lo que equivale a un incremento de 0,16% a nivel nacional.

Tabla II.3. Total de micro emprendedores estimados a partir de la ENE – periodo MAM 2022

Región	Sexo	Total Micro emprendedores	
		Factor Expansión Oficial ENE - MAM 2022	Factor Expansión Información ENE - ajustado a junio 2022
Total Nacional		2.086.024	2.089.383
Total Nacional	Hombre	1.236.463	1.238.447
	Mujer	849.561	850.936
Arica y Parinacota	Hombre	18.865	26.991
	Mujer	12.360	22.686
Tarapacá	Hombre	26.912	37.882
	Mujer	22.623	31.398
Antofagasta	Hombre	37.805	18.449
	Mujer	31.336	13.691
Atacama	Hombre	18.433	45.502
	Mujer	13.667	32.493
Coquimbo	Hombre	45.395	119.667
	Mujer	32.406	86.856
Valparaíso	Hombre	119.451	50.573
	Mujer	86.705	33.787
Metropolitana	Hombre	530.427	71.750
	Mujer	390.762	45.583
O´Higgins	Hombre	50.481	87.180
	Mujer	33.721	56.444
Maule	Hombre	71.629	81.153
	Mujer	45.484	45.273
Ñuble	Hombre	34.398	66.474
	Mujer	16.621	30.420
Biobío	Hombre	87.101	9.678
	Mujer	56.385	8.002
La Araucanía	Hombre	81.076	12.386
	Mujer	45.226	7.229

Región	Sexo	Total Micro emprendedores	
		Factor Expansión Oficial ENE - MAM 2022	Factor Expansión Información ENE - ajustado a junio 2022
Los Lagos	Hombre	66.381	531.342
	Mujer	30.369	391.339
Los Ríos	Hombre	26.069	26.085
	Mujer	16.688	16.706
Aysén	Hombre	9.671	18.907
	Mujer	7.988	12.380
Magallanes	Hombre	12.369	34.428
	Mujer	7.220	16.649

Fuente: Instituto Nacional de Estadísticas

Finalmente, el ponderador calibrado actualizado a las proyecciones de junio 2022 es un insumo para la calibración del factor. Esta información auxiliar se utiliza como un vector de totales poblacionales, llamados marginales de calibración, y para esta encuesta fueron consideradas las siguientes desagregaciones:

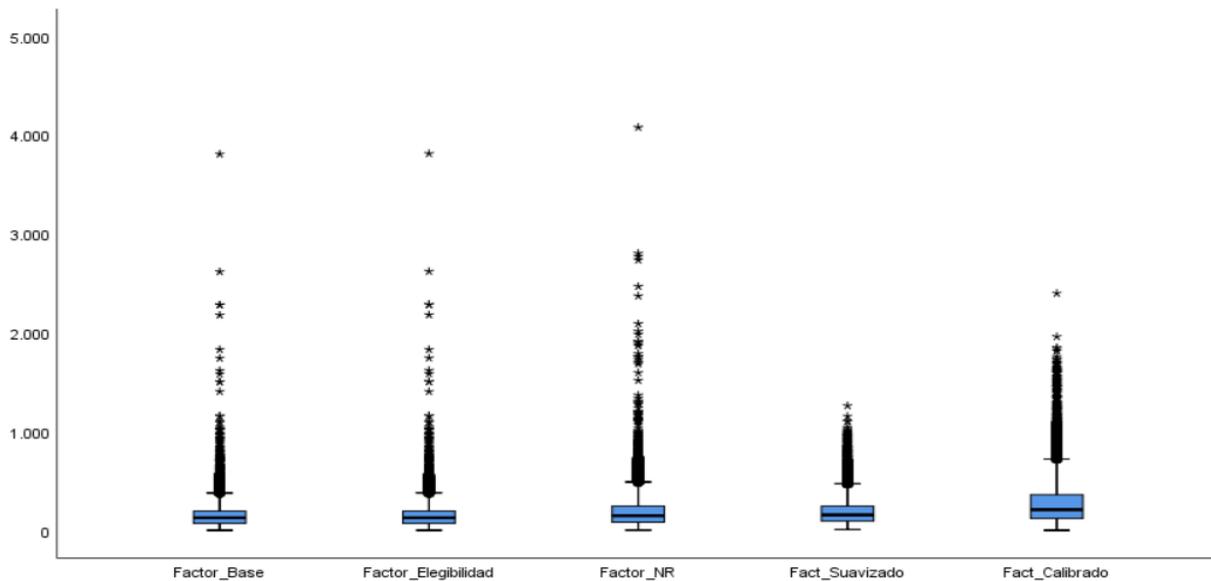
- A nivel nacional: Sexo y 11 tramos etarios: [15-24], [25-29], [30-34], [35-39], [40-44], [45-49], [50-54], [55-59], [60-64], [65-69] [70-100].
- A nivel regional: Población total según sexo.
- A nivel regional: Población total área urbana y sexo.

Una vez calibrado el factor se procede a revisar los coeficientes de variación de cada una de las marginales, cuyo valor máximo es 13,6%¹⁶. A partir del proceso de calibración se obtiene una estimación de 2.089.383 personas a nivel nacional.

En la Figura II.3 se observan, de manera descriptiva, los ajustes del ponderador de selección de microemprendedor, hasta llegar al proceso de calibración.

¹⁶ El umbral de calidad del coeficiente de variación se definió en 15% al interior de cada marginal.

Figura II.3 Gráfico de caja para los diferentes ponderadores.



Fuente: Instituto Nacional de Estadísticas

III. ESTIMACIÓN DE LA VARIANZA

El diseño muestral de la VII EME, como ya fue mencionado, contempla un diseño muestral complejo. En general, cuanto más complejo es el diseño bajo el cual se implementa una encuesta, más compleja se vuelve la forma de determinar los errores muestrales. Tanto así que, no existen ecuaciones exactas y/o explícitas para esto. Sin embargo, paquetes estadísticos en softwares especializados, facilitan los cálculos a través de aproximaciones realizadas mediante distintos modelos o métodos de estimación, donde se deben identificar las variables que definen el diseño muestral (estratos, conglomerados) y el factor de expansión apropiado (considerando todos los ajustes pertinentes).

En este contexto, en los siguientes apartados se exponen las variables que identifican el diseño muestral, así como también su implementación en R Studio. Para ello, se definió como variable de análisis la estructura de la rama de actividad de los microemprendedores, sobre la cual se realizan las estimaciones.

En ocasiones, pueden existir algunas dificultades en la implementación de la estimación de los errores mediante un paquete estadístico, originadas por las características del diseño muestral, por ejemplo: más de una fase de muestreo; muestreo multietápico de las unidades muestrales, selección de unidades sin reemplazo, estratos de muestreo con solo una unidad primaria con unidades elegibles, variabilidad de los tamaños de los conglomerados, etc.

A fin de minimizar problemas del diseño complejo y, siguiendo las recomendaciones internacionales (Valliant, Dever, & Kreuter, 2013) los errores son estimados a partir de modelos que buscan dar cuenta, lo más fielmente posible del diseño muestral.

III.1. Estimación de variables y varianzas en R

La estimación de los errores muestrales provenientes de encuestas cuyo diseño es complejo, se puede calcular en diversos paquetes estadísticos a través de métodos como la linealización de Taylor o métodos de replicación repetido (Jackknife, Bootstrap) entre otros. Sin embargo, para que éstos sean más simples de implementar, se deben considerar algunos supuestos como, por ejemplo: Se asume que la selección de las unidades en las distintas etapas se realiza de forma independiente (lo que simplifica los cálculos y las expresiones matemáticas) o cuando el diseño muestral de la encuesta posea muchas etapas, sólo se da cuenta de la primera etapa, pues es esta la que aporta la mayor variabilidad al error total.

En R, previo a la estimación de la variable de estudio y los errores asociados a ella, se debe definir el diseño muestral bajo el cual se realizarán las estimaciones. Las variables que se encuentran en la base de datos y que definen el diseño muestral de la VII EME son:

1. *fact_cal*: corresponde al factor de expansión que da cuenta de las probabilidades de selección, de la fase 1 y 2, ajustado por elegibilidad y falta de respuesta, suavizado a nivel de cada región y rama de actividad económica; y calibrado.
2. *Estrato*: Corresponde al estrato de la ENE, llamado *ESTRATO_ENE*.
3. *Conglomerado*: Corresponde a las unidades primarias de muestreo, llamado *id_upm*.

Así, para revisar la estructura de la actividad en la cual se desenvuelven los microemprendedores, previamente, el investigador debiera hacer lo siguiente:

1. Determinar y construir la variable de interés, si ésta no está definida.
2. Especificar las variables que definen el diseño complejo
3. Realizar la estimación correspondiente

En la Tabla III.1 se observa la estructura de la rama de actividad económica (*b14_1_rev4cl_caenes*) para los microemprendedores.

Tabla III.1. Estructura de la actividad económica de los microempresarios

Rama de actividad económica	Estimación	Error estándar	Intervalo de confianza al 95%		Coeficiente de variación	Efecto de diseño
			Inferior	Superior		
Total Nacional	100,0%	0,0%	100,0%	100,0%	0,0	
Agricultura, ganadería, silvicultura y pesca	5,9%	0,3%	5,4%	6,4%	0,0	0,8
Explotación de minas y canteras	0,1%	0,0%	0,1%	0,2%	0,3	0,6
Industrias manufactureras	11,5%	0,5%	10,6%	12,5%	0,0	1,6
Suministro de electricidad, gas, vapor y aire acondicionado	0,0%	0,0%	0,0%	0,1%	0,5	0,5
Suministro de agua	0,3%	0,1%	0,2%	0,5%	0,2	1,1
Construcción	13,7%	0,5%	12,7%	14,8%	0,0	1,8
Comercio al por mayor y al por menor	29,9%	0,7%	28,4%	31,4%	0,0	1,8
Transporte y almacenamiento	7,5%	0,4%	6,7%	8,3%	0,1	1,8
Actividades de alojamiento y de servicio de comidas	4,6%	0,3%	4,0%	5,2%	0,1	1,7
Información y comunicaciones	1,3%	0,2%	0,9%	1,8%	0,2	2,4
Actividades financieras y de seguros	0,4%	0,1%	0,2%	0,8%	0,3	2,8
Actividades inmobiliarias	0,9%	0,2%	0,6%	1,4%	0,2	2,8
Actividades profesionales, científicas y técnicas	5,3%	0,4%	4,6%	6,2%	0,1	2,3
Actividades de servicios administrativos y de apoyo	4,4%	0,3%	3,8%	5,1%	0,1	1,8
Administración pública y defensa	0,0%	0,0%	0,0%	0,0%	1,0	0,1
Enseñanza	1,8%	0,3%	1,3%	2,5%	0,2	3,5
Actividades de atención de la salud humana y de asistencia social	2,4%	0,3%	1,9%	3,0%	0,1	2,2
Actividades artísticas, de entretenimiento y recreativas	1,3%	0,2%	1,0%	1,7%	0,1	2,0
Otras actividades de servicios	8,7%	0,5%	7,8%	9,7%	0,1	2,0
Actividades de los hogares como empleadores	0,0%	0,0%	0,0%	0,1%	1,0	0,5

Fuente: Instituto Nacional de Estadísticas

ANEXOS

Anexo N°1. Códigos de disposición final de casos. Última visita

En la Tabla III.2 se muestran las categorías en las que son clasificadas las unidades luego del trabajo de campo. La variable “elegible” indica sí, en unidades elegibles y sobre las cuales se realizan los ajustes por falta de respuesta. Las restantes son clasificadas como no elegibles

Tabla III.2. Total microemprendedores según códigos de disposición final y estado de elegibilidad.

Códigos de disposición final	Frecuencia	%	elegible
Entrevista completa	6312	73,59%	SI
Entrevista parcial	630	7,35%	SI
Informante de la vivienda rechazó la en	146	1,70%	SI
Informante directo rechazó la entrevista	436	5,08%	SI
Vivienda ocupada, sin moradores presente	164	1,91%	SI
Se impide el acceso a la vivienda	10	0,12%	SI
Informante no ubicable, se contacta a o	106	1,24%	SI
Informante ausente en período de levantamiento	70	0,82%	SI
Encuesta anulada	1	0,01%	SI
Marca tono, pero sin respuesta	166	1,94%	SI
Responde buzón de voz	65	0,76%	SI
Suena ocupado	11	0,13%	SI
Se encuentra apagada fuera área de cobertura	25	0,29%	SI
Número de teléfono no existe	13	0,15%	SI
Número telefónico equivocado	7	0,08%	SI
Sin número telefónico	3	0,03%	SI
Otra razón de no contacto	20	0,23%	SI
Lugar o actividad del informante no per	2	0,02%	SI
Informante impedido/a física y/o mental	23	0,27%	SI
Informante no habla idioma español	1	0,01%	SI
Muerte de el/la informante o algún familiar	2	0,02%	SI
Concertó cita con el/la informante	24	0,28%	SI
Otra razón de no entrevista	14	0,16%	SI
No fue posible localizar la dirección	9	0,10%	NO
Inaccesibilidad por razones climáticas	3	0,03%	NO
Área peligrosa	2	0,02%	NO
Otra razón de elegibilidad desconocida	3	0,03%	NO
Vivienda en demolición, incendiada, des	2	0,02%	NO
Vivienda desocupada (por ejemplo: en ve	6	0,07%	NO
Vivienda de veraneo o de uso temporal	2	0,02%	NO
Edificación no calificable (por ejemplo	1	0,01%	NO
No existe informante elegible (Informan	291	3,39%	NO
Otra razón para ser considerada no elegible	7	0,08%	NO

Fuente: Instituto Nacional de Estadísticas.

Anexo N°2. Estimación de varianzas. Código para aplicar en R.

#Diseño Complejo

```
DC<-svydesign(id=~id_upm, #conglomerado
             weights=~fact_cal, #Factores de expansión
             strata=~ESTRATO_ENE,
             data=Base_CDF_VII_EME_v2b_R)
prop.table(svytable(~b14_1_rev4cl_caenes,DC))
```

BIBLIOGRAFÍA

- Cochran, W. (1998). *Técnicas de Muestreo*. México D.F.: Compañía Editorial Continental, S.A. de C.V.
- Groves, R. (2004). *Survey Errors and Survey Cost*. Nueva York: John Wiley & Sons.
- Guerrero, J., & Marín, A. (2020). *Estratificación socioeconómica del marco muestral de viviendas 2017 (MMV 2017)-Implementación en R*. Santiago de Chile: INE.
- Jones, A., Koolman, X., & Rice, N. (2006). Health Related Non Response in The British Household Panel Survey and European Community Household Panel: Using Inverse Probability Weighted Estimators in NonLinear Models. *Journal of The Royal Statistical Society. Series A (Statistics in Society)*, Vol. 169, No. 3.
- Kish, L. (1963). Changing strata and selection probabilities. *Proceedings of the Social Statistics Section, American Statistical Association*, 124-131.
- Naciones Unidas. (2009). *Diseño de muestras para encuestas de hogares. Directrices prácticas*. Nueva York.
- ONU. (2009). *Diseño de muestras para encuestas de hogares. Directrices prácticas*. Recuperado el 15 de Noviembre de 2017, de https://unstats.un.org/unsd/publication/seriesf/Seriesf_98s.pdf
- Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.