



---

# VIII Encuesta de Microemprendimiento EME 2025

---

**Documento metodológico del diseño muestral**

**INSTITUTO NACIONAL DE ESTADÍSTICAS**

**Diciembre / 2025**

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

---

VIII Encuesta de Microemprendimiento EME 2025.  
Documento metodológico del diseño muestral.  
Publicación: diciembre 2025

Instituto Nacional de Estadísticas, Chile.  
Diciembre de 2025.

## 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 .....	6
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 .....	7
I.6.1.	Características del MMV 2020 .....	8
I.6.2.	Marco muestral de la Encuesta de Microemprendimiento .....	12
I.7.	Estrategia muestral .....	16
I.8.	Cálculo y distribución del tamaño muestral .....	18
I.8.1.	Metodología de cálculo del tamaño muestral .....	19
I.8.2.	Tamaño calculados y errores esperados .....	21
I.8.3.	Distribución de la muestra según submuestra .....	22
I.8.4.	Tamaños logrados y errores observados .....	23
I.9.	Selección de unidades muestrales .....	24
I.9.1.	Selección de viviendas .....	24
I.9.2.	Selección de microemprendedores .....	25
II.	DESARROLLO DE FACTORES DE EXPANSIÓN .....	26
II.1.	Ponderador Base .....	26
II.1.1.	Probabilidad de selección y entrevista de las viviendas en la ENE- Trimestre MAM 2025 27	
II.1.2.	Probabilidad de selección de los microemprendedores .....	27
II.2.	Ajuste por elegibilidad .....	28
II.2.1.	Ajuste por elegibilidad desconocida .....	29
II.2.2.	Ajuste por no elegibilidad .....	30
II.3.	Ponderador ajustado por falta de respuesta .....	30
II.4.	Suavizamiento del Ponderador ajustado por no respuesta .....	32
II.4.1.	Método de contracción a la media (CM) .....	33
	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: .....	33
II.4.2.	Método r-k .....	33
II.4.3.	Método mixto .....	34
II.5.	Calibración .....	34
III.	ESTIMACIÓN DE LA VARIANZA .....	38
III.1.	Estimación de variables y varianzas en R .....	38

## ÍNDICE DE TABLAS

Tabla I.1. Cantidad de UPM especiales y excluidas, por región .....	14
Tabla I.2. Cantidad de estratos de muestreo por región y área.....	15
Tabla I.3. Total de microemprendedores según ENE trimestre de selección y según Marco VIII EME 2025 .....	17
Tabla I.4. Errores esperados asociados al parámetro de interés según tamaño muestral objetivo.....	22
Tabla I.5. Total de viviendas seleccionadas según región y mes de levantamiento .....	23
Tabla I.6. Errores observados asociados al parámetro de interés según número de viviendas que responden .....	24
Tabla II.1. Distribución de individuos según la clasificación de elegibilidad. ....	30
Tabla II.2. Total de unidades elegibles, unidades que responden y tasas de respuesta.....	32
Tabla II.3. Total de micro emprendedores estimados a partir de la ENE – periodo MAM 2025 .....	36
Tabla III.1. Estructura de la actividad económica de los microemprendedores .....	39
Tabla III.2. Total microemprendedores según códigos de disposición final y estado de elegibilidad. ....	40

## ÍNDICE DE CUADROS

Cuadro I.1. Criterios utilizados en el cálculo del tamaño muestral.....	18
---	----

## PRESENTACIÓN

La subsecretaría de Economía y Empresas de Menor Tamaño —en adelante, la Subsecretaría, ha encargado al Instituto Nacional de Estadísticas (INE) la aplicación, desde el año 2013, de la “Encuesta de Microemprendimiento” (EME). Esta encuesta es una herramienta de enorme valor estadístico para el país, puesto que es el único estudio de este tipo que se realiza a lo largo de todo Chile, abarcando unidades económicas pequeñas, ya sea formales o informales, pertenecientes a todos los sectores económicos.

El objetivo principal de la VIII 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 VIII EME, se obtiene a partir del levantamiento de la “Encuesta Nacional de Empleo” (ENE) en el trimestre móvil marzo, abril, mayo (MAM 2025) 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 microemprendedor<sup>1</sup>; 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.

---

<sup>1</sup> 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. Observar las formas en las que operan y funcionan los micronegocios.
2. Estudiar el grado de formalidad de los micronegocios.
3. Medir y caracterizar el acceso al sistema financiero.
4. Determinar los recursos productivos de los micronegocios.
5. Estimar los ingresos y gastos de los micronegocios.
6. Identificar la generación de empleo de las microempresas y sus características.
7. Estudiar los impactos y medidas de mitigación de la crisis ambiental en los microemprendimientos.
8. Caracterizar a los microempresarios según el estándar de sexo, género y orientación sexual.

#### I.1.3. Objetivo del diseño muestral

Obtener estimaciones del parámetro de interés “Proporción de microempresarios por cuenta propia sobre el total de microempresarios” según los niveles de precisión establecidos<sup>2</sup> a nivel nacional y regional.

---

<sup>2</sup> Los niveles de precisión son: error absoluto de 1,0% y error relativo de 1,1%, a nivel nacional. Para las regiones, el error absoluto máximo es 6,6% y, el error relativo no supera 8,5%. Este valor se obtiene en la Región de Magallanes.

## **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.

## **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 los meses de mayo, junio y julio de 2025, con periodicidad bienal. Mientras que, el periodo de referencia corresponde al trimestre MAM<sup>3</sup> 2025 en el cual se realiza la primera fase de la encuesta (levantamiento de la ENE).

## **I.6. Marco Muestral**

---

Un marco de muestreo se define como un componente material o dispositivo que delimita, identifica de manera unívoca y facilita el acceso observacional a la población finita de interés. Su principal función consiste en permitir la selección de una muestra de unidades mediante un mecanismo que respete las probabilidades establecidas por un diseño de muestreo probabilístico dado (Särndal et al., 1992).

---

<sup>3</sup> Corresponde a los meses de marzo, abril y mayo.

El éxito de un marco de muestreo en abordar eficazmente los temas de estudio radica en su exhaustividad y precisión (ONU, 2009). El marco debe contener el conjunto de unidades pertenecientes a la población objetivo y garantizar que cada una esté representada de manera única. Satisfacer estos criterios reduce significativamente los problemas asociados con la cobertura del marco, permitiendo realizar inferencias válidas sobre la población objetivo y alcanzar los objetivos de la encuesta.

Por el caso contrario, las imperfecciones del marco —como la falta de cobertura, la sobrecobertura y la presencia de unidades duplicadas— conducen a la sub-representación o sobre-representación de segmentos de la población objetivo, generando estimaciones sesgadas. Dado que contar con un marco perfecto rara vez resulta factible en la práctica<sup>4</sup>, una solución efectiva para minimizar estas imperfecciones es emplear un marco de muestreo lo más actualizado posible.

A la luz de estas consideraciones, durante el segundo semestre de 2019, el INE puso a disposición el Marco Muestral de Viviendas 2017 (MMV 2017) construido con base en el Censo de Población y Vivienda del 2017. Este marco fue diseñado con protocolos para realizar procesos de actualización continua y permanente, los cuales se implementaron para dar lugar al Marco Muestral de Viviendas 2020 (MMV 2020), disponible desde el segundo trimestre del 2021. La creación del MMV 2020 involucró enumeraciones en terreno de unidades geográficas prioritarias y enumeraciones de gabinete que aprovechan integralmente los registros administrativos disponibles en el INE.

A continuación, se describen las características del marco muestral utilizado para la selección de la muestra de la VIII EME. Dado que las unidades seleccionadas en esta encuesta provienen de la Encuesta Nacional de Empleo (ENE), es necesario revisar las características del marco muestral correspondiente tanto a la Fase 1 (ENE) como a la Fase 2 (EME).

### **I.6.1. Características del MMV 2020**

#### **I.6.1.1. Cobertura geográfica del marco muestral**

El MMV 2020 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.6.1.2. Estratificación geográfica del marco muestral**

El MMV 2020 contiene variables que permiten la identificación de la división político-administrativa del país<sup>5</sup>, así como también la identificación de las áreas urbanas y rurales de cada comuna.

---

<sup>4</sup> Contar con un marco exhaustivo y exacto resulta un objetivo inalcanzable en las encuestas a hogares, debido al constante cambio de los asentamientos humanos.]

<sup>5</sup> Región, provincia y 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 primarias<sup>6</sup>.
- **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 del país, 295 presentan una configuración mixta, compuestas por áreas urbanas y rurales. Además, 24 comunas están conformadas exclusivamente por área urbana, mientras que 26 contienen únicamente área rural.

**Figura I.1 Estratificación geográfica del MMV 2020**



**Fuente:** Instituto Nacional de Estadísticas.

<sup>6</sup> Actividades primarias: Son todas aquellas actividades económicas que consisten en la extracción de productos naturales, tales como agricultura, pesca, caza, minería, etc.

#### I.6.1.3. Conformación de la Unidades Primarias de Muestreo

El MMV 2020 está conformado por unidades conocidas como Unidades Primarias de Muestreo (UPM), una característica que ya fue introducida en el MMV 2017 y que representa una diferencia fundamental respecto a los marcos anteriores. Las UPM son áreas geográficas homogéneas definidas en función del número de viviendas particulares (con moradores presentes o ausentes) que se encuentran en ellas, las cuales fueron observadas en el Censo 2017 y han sido actualizadas en lo sucesivo.

Las UPM se definen según su pertenencia unívoca al área urbana o rural. En el área urbana, el tamaño medio de las UPM es de 200 viviendas, con un intervalo de 20%. En otras palabras, la cantidad de viviendas por UPM urbana se encuentra en el rango de 160 a 240 viviendas. En el área rural, las UPM tienen un tamaño medio de 90 viviendas, con un intervalo de 20%. Por lo tanto, la cantidad de viviendas por UPM rural se encuentra en el rango de 70 a 110 viviendas.

Para la construcción de las UPM, el criterio fundamental es mantener unidades compactas y homogéneas<sup>7</sup>, con el objetivo de que el muestreo bietápico sea eficiente y que la muestra pueda entrar en operación en forma óptima.

#### I.6.1.4. Estratificación socioeconómica del marco muestral

El MMV 2020, al igual que su antecesor, 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

---

<sup>7</sup> 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

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
- 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 (tres o cuatro) 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 2020 como clasificación socioeconómica de la UPM<sup>8</sup>.

#### 1.6.1.5. Áreas especiales

Al igual que su antecesor, el MMV 2020, aporta información necesaria para identificar áreas especiales susceptibles de ser excluidas, a nivel de UPM. Estas áreas suelen estar afectadas por factores que dificultan su levantamiento, los cuales pueden ser: climáticos, de transporte, costo de acceso y de acceso restringido, entre otros. El MMV 2020 aporta información mediante ocho tipologías que dan cuenta de las

---

<sup>8</sup> 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)

complejidades presentes en la realización de trabajos operativos, información que resulta de gran utilidad para la coordinación de las actividades en terreno, tales como:

- 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 a nivel de UPM, en lugar de comuna, con lo cual el MMV 2020 consigue un mayor nivel de cobertura.

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

#### **I.6.1.6. Conformación definitiva del marco muestral**

Se obtuvo un total de 35.328 UPM separadas en dos estratos urbano y rural para un total de 6.221.695 viviendas marco, obteniendo para el estrato urbano un total de 26.367 UPM en 5.412.216 viviendas marco, por otra parte, para el estrato rural se obtuvieron un total de 8.961 UPM en 809.479 viviendas marco.

En la estratificación socioeconómica, se obtuvieron 12.782 UPM en el estrato bajo, 15.621 UPM en el estrato medio y 6.925 UPM en el estrato alto a nivel nacional.

#### **I.6.2. Marco muestral de la Encuesta de Microemprendimiento**

La VIII EME cuenta con un diseño muestral bifásico, por lo que el marco muestral está conformado por dos marcos muestrales complementarios. El primero, corresponde al marco utilizado para la selección de viviendas en la Encuesta Nacional de Empleo (ENE), específicamente en el periodo marzo-abril-mayo (MAM 2025), mientras que el segundo marco corresponde al listado de todas las viviendas que, en la recolección de la ENE resultaron contener una o más personas microemprendedoras.

#### I.6.2.1. Estratificación geográfica

La estratificación geográfica del marco muestral de la ENE, correspondiente a la primera fase, comienza desde la división político-administrativa y la identificación de áreas en sectores urbanos y rurales presentes en el marco muestral maestro. En este caso, el objetivo de la estratificación es 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.

#### I.6.2.2. 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 2020, 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 cobertura.

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 y excluidas.

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

Región	N° UPM	N° UPM especiales excluidas	% UPM Especiales excluidas	N° Viviendas Marco	N° Viviendas marco especiales	% Viv. Marco especiales excluidas
<b>Nacional</b>	<b>35.328</b>	<b>355</b>	<b>1,0%</b>	<b>6.221.695</b>	<b>35.287</b>	<b>0,6%</b>
Arica y Parinacota	414	8	1,9%	74.151	736	1,0%
Tarapacá	580	13	2,2%	111.159	1.195	1,1%
Antofagasta	975	4	0,4%	190.790	359	0,2%
Atacama	609	6	1,0%	106.193	511	0,5%
Coquimbo	1.757	0	0,0%	285.374	0	0,0%
Valparaíso	3.697	28	0,8%	680.833	3.921	0,6%
Metropolitana	12.087	0	0,0%	2.379.740	0	0,0%
O'Higgins	2.155	6	0,3%	333.576	524	0,2%
Maule	2.566	5	0,2%	392.061	457	0,1%
Ñuble	1.236	8	0,6%	182.744	726	0,4%
Biobío	3.138	24	0,8%	558.582	2.064	0,4%
La Araucanía	2.414	49	2,0%	359.601	4.494	1,2%
Los Ríos	984	23	2,3%	144.531	2.023	1,4%
Los Lagos	2.106	167	7,9%	318.562	16.895	5,3%
Aysén	273	11	4,0%	41.844	1.177	2,8%
Magallanes	337	3	0,9%	61.954	205	0,3%

Fuente: Instituto Nacional de Estadísticas.

### I.6.2.3. Estratificación del marco muestral ENE

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 de 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 alguna categoría de nivel socioeconómico (NSE) y, en torno al 60% en otro.
- 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		
	Urbano	Rural	Total
<b>Total</b>	<b>153</b>	<b>79</b>	<b>232</b>
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.

Las viviendas seleccionadas y encuestadas en la ENE para el periodo MAM 2025, son utilizadas como marco de muestreo para la VIII 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 VIII EME, sólo con aquellas viviendas con presencia de microemprendedores.

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

#### I.6.2.4. Cobertura geográfica

La cobertura geográfica de la EME es la misma de la ENE, considera las 16 regiones del país, buscando presencia en todo el territorio nacional y abarcando 97% de las comunas del país. No obstante, se excluyen algunas áreas geográficas que por presentar ciertas características especiales (Insularidad, clima, altitud, tiempo de traslados, estado de los caminos, etc.), son excluidas del marco de muestreo de la encuesta.

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

### I.7. Estrategia muestral

---

La VIII EME posee un diseño muestral bifásico, donde la primera fase consiste en la aplicación de una encuesta para conocer las características las unidades informantes y, a partir de esta información, identificar la población objetivo de la EME. Este proceso permite construir el marco de selección para la segunda fase de la encuesta.

En la primera fase, correspondiente a la Encuesta Nacional de Empleo (ENE), se implementa un muestreo probabilístico, estratificado y bietápico. Las Unidades Primarias de Muestreo (UPM) corresponden a conglomerados geográficos, mientras que las Unidades Secundarias, son las viviendas particulares ocupadas. La selección de las UPM se realiza mediante un muestreo aleatorio sistemático con probabilidad proporcional a su tamaño, en número de viviendas, dentro de cada estrato de muestreo, y las viviendas se seleccionan de forma sistemática con igual probabilidad. Las unidades seleccionadas y encuestadas en la ENE durante el período MAM 2025 conforman el marco de muestreo para la VIII EME, ya que permiten identificar las viviendas en las que reside al menos un microemprendedor.

En la segunda fase, las viviendas se clasifican en dos grupos, según la presencia o no de, al menos, un microemprendedor durante el período de referencia. Aquellas que no cumplen con este criterio se excluyen, conformando así un marco de muestreo compuesto únicamente por viviendas con personas pertenecientes a la población objetivo. Este listado es revisado exhaustivamente para descartar unidades que no cumplen con los criterios técnicos necesarios para ser consideradas efectivamente como microemprendimientos, lo que permite reducir problemas posteriores al levantamiento, debido a la existencia de casos fuera de muestra.

Una vez depurado el listado, se seleccionan las viviendas que integrarán la muestra final. Posteriormente, se listan todos los microemprendedores al interior de cada vivienda y hogar, y se selecciona aleatoriamente un microemprendedor por cada rama de actividad económica dentro del hogar, es decir, en los casos que existan múltiples microemprendedores con la misma actividad económica en un mismo hogar, se selecciona sólo uno de ellos, con el fin de evitar la sobrerrepresentación de dicha actividad.



La Tabla I.3 muestra: el total de personas clasificadas en la ENE como microemprendedores, el total de microemprendedores, utilizado como marco de selección muestral para la VIII 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 7,1% de casos descartados por ser potenciales unidades no elegibles<sup>9</sup>, donde el mayor porcentaje de depuración se encuentra en la Región de Los Ríos (9,8%) y; el menor, en la Región de Tarapacá (1,1%).

**Tabla I.3. Total de microemprendedores según ENE trimestre de selección y según Marco VIII EME 2025**

Región	Total Microemprendedores	Total Microemprendedores	Porcentaje Depuración
	ENE	EME	
<b>Total Nacional</b>	<b>10.953</b>	<b>10.175</b>	<b>7,1%</b>
Arica y Parinacota	432	424	1,9%
Tarapacá	366	362	1,1%
Antofagasta	322	297	7,8%
Atacama	485	474	2,3%
Coquimbo	646	599	7,3%
Valparaíso	1.222	1.148	6,1%
Metropolitana	2.269	2.080	8,3%
O'Higgins	516	495	4,1%
Maule	770	712	7,5%
Ñuble	478	439	8,2%
Biobío	927	860	7,2%
La Araucanía	707	639	9,6%
Los Ríos	521	470	9,8%
Los Lagos	726	660	9,1%
Aysén	332	303	8,7%
Magallanes	234	213	9,0%

Fuente: Instituto Nacional de Estadísticas.

<sup>9</sup> 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).

I.8. Cálculo y distribución del tamaño muestral

El cálculo de los tamaños muestrales tiene por objetivo la obtención de tamaños de muestra que permitan contar con estimaciones de los parámetros de interés con errores de muestreo tolerables para los distintos niveles de estimación, bajo el estándar de calidad del INE (INE, 2020).

El proceso de cálculo de tamaños implica la generación de diferentes escenarios con errores de estimación similares, entre los cuales se elige la solución óptima. Cada uno de estos escenarios toma como punto de partida las estimaciones del parámetro de interés y sus estadísticos asociados, obtenidos con la información de la ENE levantada en el trimestre marzo, abril y mayo de 2024 (MAM 2024). Utilizando estas estimaciones, se estima la cuasi-varianza poblacional para obtener un tamaño muestral inicial bajo muestreo aleatorio simple monoetápico. A este tamaño se le aplican ajustes para tener en cuenta el efecto de diseño muestral de la encuesta y las correspondientes correcciones por finitud de la población. Además, se ajusta el tamaño calculado según el número de viviendas a encuestar por UPM y se evalúan incrementos al tamaño muestral objetivo para compensar posibles pérdidas de unidades debido a la falta de respuesta en algunas viviendas seleccionadas para la entrevista. 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	$I = \begin{cases} 0, \\ 1, \end{cases}$	<i>si el microemprendedor no es cuenta propia</i> <i>si el microemprendedor es cuenta propia</i>
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 es de 1,0% y; el error relativo 1,1%. Regional: Los errores absolutos no superan 6,6% y; los errores relativos 8,5%.		

Fuente: Instituto Nacional de Estadísticas.

### 1.8.1. 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 AMJ 2025.
$n_{r0}$	: Número de UPM levantadas en trimestre MAM 2025 (ENE) en la región $r$ .
$n_{r0} \cdot \bar{m}_{r0} = m_{r0}$	: Número de viviendas logradas en trimestre MAM 2025 (ENE) en la región $r$ .
$\lambda$	: Porcentaje de selección asignado para cada región (62% a 79%).
$SE(\hat{p}_{r0})$	: 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.
$t_{1-\alpha/2}^v$	: 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 base de datos del levantamiento de la ENE en el trimestre MAM 2024, 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:

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

#### **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 los errores absolutos y la cuasi-varianza del parámetro de interés calculada usando la ecuación (1), se obtienen los tamaños muestrales iniciales a nivel regional, bajo un muestreo aleatorio simple monoetápico. Este cálculo se observa en la ecuación (2):

$$m_{0,r} = \frac{t_{1-\alpha/2}^2 \cdot S(p_{r0})^2}{ea_{r1}^2} \quad (2)$$

Donde,

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

### **Etapas 3**

A partir del efecto de diseño, definido como la razón entre la varianza de la estimación obtenida de una muestra compleja y la varianza de la estimación obtenida de una muestra aleatoria simple con el mismo número de unidades, se calcula el tamaño bajo muestreo complejo. El efecto del diseño se utiliza tanto en la estimación del tamaño de la muestra como en la evaluación de si un plan es eficiente en relación con su costo y complejidad (Cochran, 1977).

En este contexto, el efecto de diseño se utiliza para ajustar el tamaño muestral calculado bajo muestreo aleatorio simple (ecuación 2) ya que el muestreo complejo conlleva una pérdida de eficiencia en las estimaciones en comparación con el muestreo aleatorio simple.

Los valores utilizados como referencia son los efectos de diseño obtenidos en el trabajo de campo de la ENE en el trimestre MAM 2024. La ecuación 3 es utilizada para el cálculo del tamaño muestral  $m_{r1}$  que incorpora dicho efecto:

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

Donde,

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

$Def f$  : Efecto de diseño fijo para cada región.

Donde  $m_{r0}$  representa el número de viviendas estimadas a encuestar en la región  $r$  bajo un muestreo aleatorio simple y  $m_{r1}$  representa el número de viviendas a encuestar en la región  $r$  bajo un diseño complejo.

### **Etapas 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):

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

### **Etapla 5**

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

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

### **Etapla 6**

Una vez obtenidos los tamaños objetivos regionales, se aplica un factor que corresponde a la tasa de no logro (*tnl*). Dicha tasa se obtiene a partir de la tasa de no logro de la VII EME, 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):

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

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

### **I.8.2. Tamaño calculados y errores esperados**

La Tabla I.4 presenta los errores esperados, asociados a la estimación del parámetro de interés, para un tamaño muestral objetivo de 8.381 viviendas<sup>10</sup> a nivel nacional.

---

<sup>10</sup> La estimación de los errores se obtiene a partir del tamaño objetivo y no del tamaño con sobremuestreo, pues dadas las pérdidas naturales de unidades muestrales, se espera obtener un tamaño de 6.459 viviendas.

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

Región	Estimación	Efecto Diseño Propuesto	Error Absoluto Esperado	Error Relativo Esperado	Tamaños ajustados por finitud	Tamaño ajustado por tnl de VII EME
<b>Nacional</b>	<b>89,0%</b>	<b>2,2</b>	<b>1,0%</b>	<b>1,1%</b>	<b>6.459</b>	<b>8.381</b>
Arica y Parinacota	92,2%	1,1	3,5%	3,8%	253	317
Tarapacá	88,6%	1,0	4,2%	4,8%	227	284
Antofagasta	89,8%	1,4	4,9%	5,5%	211	264
Atacama	88,2%	1,2	4,0%	4,6%	305	382
Coquimbo	89,2%	1,9	4,3%	4,8%	398	531
Valparaíso	87,5%	1,2	2,8%	3,2%	688	917
Metropolitana	90,0%	1,2	1,9%	2,1%	1.174	1.565
O'Higgins	87,0%	1,3	4,2%	4,9%	329	411
Maule	87,9%	1,3	3,4%	3,9%	479	599
Ñuble	86,5%	1,1	4,1%	4,8%	301	377
Biobío	87,7%	1,5	3,4%	3,8%	568	757
La Araucanía	92,2%	0,9	2,5%	2,7%	426	569
Los Ríos	86,0%	2,3	5,8%	6,7%	328	409
Los Lagos	89,0%	1,2	3,4%	3,8%	401	535
Aysén	82,4%	1,0	5,1%	6,2%	216	270
Magallanes	78,1%	1,0	6,6%	8,5%	155	194

Fuente: Instituto Nacional de Estadísticas.

### I.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 atrición<sup>11</sup>, la muestra de la VIII 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.

<sup>11</sup> 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).

**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
<b>Nacional</b>	<b>2.722</b>	<b>2.826</b>	<b>2.833</b>	<b>8.381</b>
Arica y Parinacota	92	107	118	317
Tarapacá	76	96	112	284
Antofagasta	89	79	96	264
Atacama	95	137	150	382
Coquimbo	180	173	178	531
Valparaíso	306	306	305	917
Metropolitana	522	522	521	1.565
O'Higgins	137	137	137	411
Maule	200	200	199	599
Ñuble	129	125	123	377
Biobío	249	256	252	757
La Araucanía	190	192	187	569
Los Ríos	129	141	139	409
Los Lagos	166	191	178	535
Aysén	93	98	79	270
Magallanes	69	66	59	194

Fuente: Instituto Nacional de Estadísticas.

#### **I.8.4. Tamaños logrados y errores observados**

Para un tamaño logrado de 6.735 viviendas a nivel nacional, se obtuvo una estimación del parámetro de interés de 87,3% con un error absoluto de 1,0% y un error relativo de 1,2%.

A nivel regional los tamaños permiten obtener estimaciones del parámetro de interés con errores absolutos que fluctúan entre 2,0% para la región Metropolitana y 5,5% para la región de Aysén y Magallanes.

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

Nivel	Estimación VIII EME	Tamaño efectivo	Error absoluto	Error relativo
<b>Total nacional</b>	<b>87,3%</b>	<b>6.735</b>	<b>1,0%</b>	<b>1,2%</b>
Arica y Parinacota	86,5%	264	4,0%	4,6%
Tarapacá	85,0%	268	4,9%	5,7%
Antofagasta	90,6%	219	4,1%	4,5%
Atacama	86,1%	289	4,5%	5,2%
Coquimbo	87,7%	409	4,0%	4,6%
Valparaíso	87,4%	681	2,6%	2,9%
Metropolitana	87,3%	1.217	2,0%	2,3%
O'Higgins	87,8%	332	4,1%	4,6%
Maule	89,2%	486	3,0%	3,3%
Ñuble	87,0%	331	3,9%	4,5%
Biobío	85,6%	652	2,9%	3,4%
La Araucanía	88,3%	442	3,3%	3,8%
Los Ríos	83,0%	329	4,3%	5,2%
Los Lagos	87,9%	421	3,8%	4,4%
Aysén	88,2%	226	5,5%	6,2%
Magallanes	82,4%	169	5,5%	6,7%

Fuente: Instituto Nacional de Estadísticas.

## I.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.

### I.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: " $A$ ", " $A + k$ ", " $A + 2k$ ", " $A + 3k$ ", ..., " $A + (m_{3,r} - 1)k$ ".

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 " $A + 2k$ " 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:

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

### 1.9.2. 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 VIII EME, las probabilidades de selección asociadas a los microemprendedores presentan varias componentes:

1. Probabilidad de que la vivienda haya sido seleccionada y contestara en la ENE en el trimestre MAM 2025.
2. Probabilidad de seleccionar una vivienda para la EME, dado que la vivienda posee al menos un microemprendedor.
3. Probabilidad de seleccionar un microemprendedor, 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 2025.

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 2025

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 VIII EME es el factor de expansión ajustado por no respuesta para aquellas viviendas que contestan la ENE en el periodo MAM 2025 llamado  $w_{hij}^{NR12}$ .

### 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):

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

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 microemprendedor.

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

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

Una vez seleccionada la vivienda se seleccionan los microemprendedores. La probabilidad de seleccionar al microemprendedor  $k$  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):

<sup>12</sup> 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](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).

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

Donde:

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

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

Luego la probabilidad condicional de seleccionar el microemprendedor  $k$ , 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 microemprendedor 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 (CDF<sup>13</sup>) 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).

---

<sup>13</sup> Detalle de los códigos de disposición final en Anexo Nro. 1 del presente documento.

**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 2025.

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,

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

Donde,

- $\Theta_j$  Es el conjunto de micro emprendedores  $k$  seleccionados al interior de cada vivienda  $j$  en la región  $R$ .
- $\Theta_{j,known}$  Es el conjunto microemprendedores seleccionados en la vivienda  $j$  clasificada como elegibles o no elegibles.
- $\Omega_R$  Es el conjunto de estratos  $h$  de la EME (regiones).
- $F_{Rjk}^{base}$  Es el ponderador de selección del  $k$  – ésimo microemprendedor 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 microemprendedor 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 9.094 personas residentes en viviendas particulares. De este total, solo 3 personas resultaron con elegibilidad desconocida, una vez finalizado el trabajo de campo. De las 9.091 personas restantes 8.661 resultaron elegibles (logradas y no logradas) y; 430, 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	8.661	95,24%
No elegible	430	4,73%
Elegibilidad desconocida	3	0,03%
<b>Total</b>	<b>9.094</b>	<b>100,00%</b>

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.

En la VIII EME, de un total de 8.381 viviendas seleccionadas, se seleccionaron 9.094 microemprendedores. De éstos, 8.661 fueron clasificados como elegibles (95,24%), de los cuales 7.170 respondieron la encuesta. Por lo tanto, la tasa de respuesta de la EME, ajustada por elegibilidad, es de 82,8%.

El método utilizado para compensar la falta de respuesta es el de estratificación mediante “propensity score”. De acuerdo, a lo indicado por Valliant (Valliant, Dever, & Kreuter, 2013). Este método consiste en modelar la probabilidad de respuesta en la VIII 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 ajustes<sup>14</sup>, se procede a estimar el ponderador ajustado por falta de respuesta, dado por la ecuación (14):

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

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.
- $F'_{Rjk}$  : Estimación del total de microemprendedores que responden, utilizando el ponderador ajustado por elegibilidad.

<sup>14</sup> 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.

$F'_{Rjk}$  : 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):

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

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	$\hat{R}_c^{NR}$
<b>Total</b>	8.661	7.170	82,8%	1,22
1	1.444	765	53,0%	1,90
2	1.444	1.067	73,9%	1,34
3	1.444	1.235	85,5%	1,17
4	1.443	1.314	91,1%	1,10
5	1.443	1.373	95,1%	1,05
6	1.443	1.416	98,1%	1,02

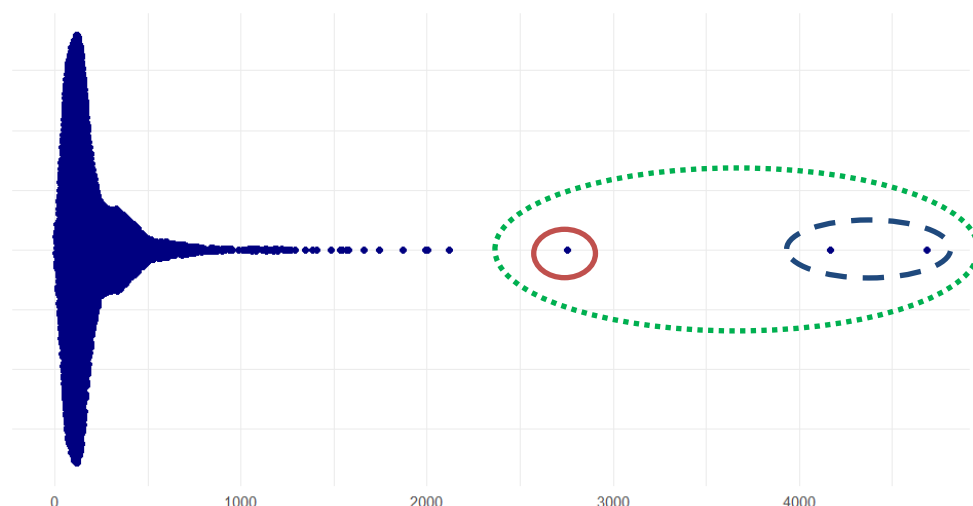
Fuente: Instituto Nacional de Estadísticas.

#### 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 no 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  $\beta$  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.**

Factores		Aumento o disminución	
$W_1$	Atípicos	$W_1 \cdot (1 - r)$	
$W_2$		$W_2 \cdot (1 - r)$	
$\vdots$		$\vdots$	
$\vdots$		$\vdots$	
$W_n$		$W_n \cdot (1 - r)$	
$w_1$	No atípicos	$w_1 \cdot (1 + k)$	
$w_2$		$w_2 \cdot (1 + k)$	
$\vdots$		$\vdots$	
$\vdots$		$\vdots$	
$w_m$		$w_m \cdot (1 + k)$	

$$t_1 = \left( \sum_{i=1}^n W_i \right) - \left( \sum_{i=1}^n W_i \right) \cdot r$$

$$t_2 = \left( \sum_{j=1}^m w_j \right) + \left( \sum_{j=1}^m w_j \right) \cdot k$$

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  $\beta$  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, 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 VIII EME, está anclada a la población de referencia del trimestre MAM 2025 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 2025, obtenido de la ENE, actualizada al período del trabajo de campo de la VIII EME. Para esto, se utiliza el crecimiento proyectado para el mes central del período de levantamiento de la encuesta, es decir junio 2025.

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 2025, 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 2025 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 2025 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 2025.

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 2025; y según el total de personas estimadas con la información levantada en MAM 2025, 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 2.108.428 personas. Sin embargo, al actualizar las proyecciones de población este total asciende a 2.114.567, lo que equivale a un incremento de 0,29% a nivel nacional.

**Tabla II.3. Total de micro emprendedores estimados a partir de la ENE – periodo MAM 2025**

Región	Sexo	Total Micro emprendedores	
		Factor Expansión Oficial ENE - MAM 2025	Factor Expansión Información ENE - ajustado a junio 2025
Total Nacional		2.108.428	2.114.567
Total Nacional	Hombre	1.250.829	1.254.138
	Mujer	857.598	860.429
Arica y Parinacota	Hombre	17.908	18.446
	Mujer	15.304	16.108
Tarapacá	Hombre	26.670	27.385
	Mujer	24.026	24.272
Antofagasta	Hombre	33.665	33.718
	Mujer	28.055	28.104
Atacama	Hombre	20.243	20.569
	Mujer	15.387	15.979
Coquimbo	Hombre	53.234	53.381
	Mujer	38.904	39.002
Valparaíso	Hombre	120.778	120.984
	Mujer	82.137	82.290
Metropolitana	Hombre	516.444	517.187
	Mujer	375.080	375.539
O´Higgins	Hombre	51.072	51.167
	Mujer	44.756	44.848
Maule	Hombre	77.807	77.942
	Mujer	47.897	47.993
Ñuble	Hombre	34.550	34.576
	Mujer	18.206	18.234
Biobío	Hombre	80.221	80.292
	Mujer	54.792	54.857
La Araucanía	Hombre	86.444	86.524
	Mujer	51.451	51.511
Los Lagos	Hombre	76.399	76.515
	Mujer	30.689	30.745
Los Ríos	Hombre	31.692	31.726
	Mujer	17.414	17.442
Aysén	Hombre	10.237	10.245
	Mujer	7.014	7.016
Magallanes	Hombre	13.464	13.481
	Mujer	6.486	6.489

Fuente: Instituto Nacional de Estadísticas

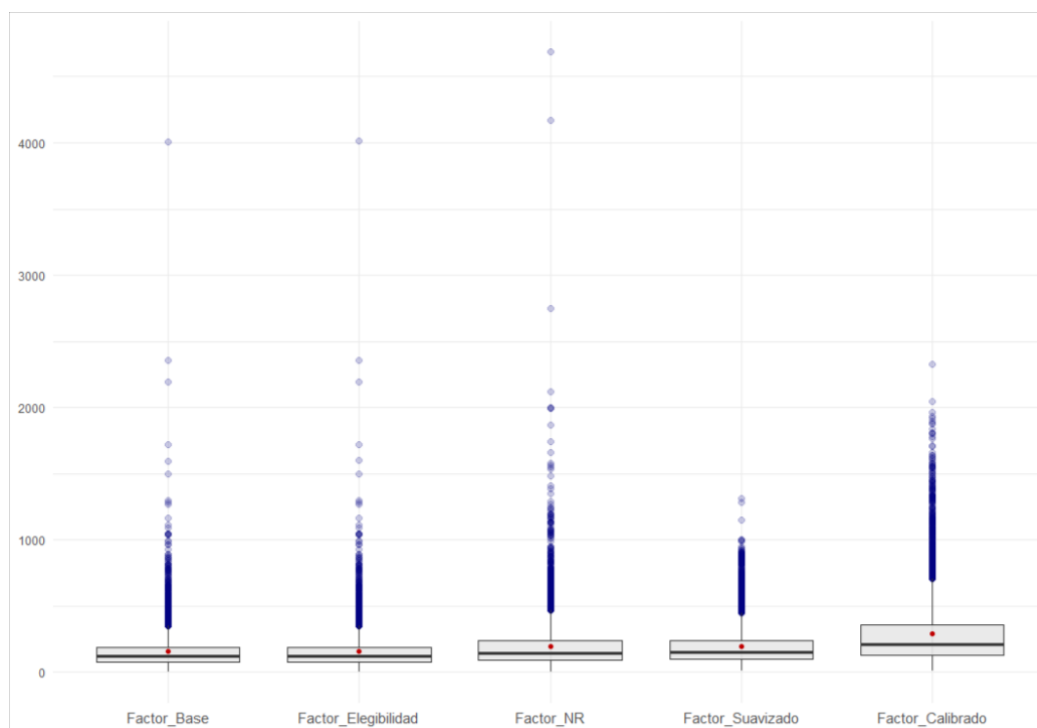
Finalmente, el ponderador calibrado actualizado a las proyecciones de junio 2025 es un insumo para la calibración del factor. Esta información auxiliar se utiliza como un vector de totales poblacionales, llamados marginales de calibración, y para esta encuesta fueron consideradas las siguientes desagregaciones:

- 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 12,9%<sup>15</sup>. A partir del proceso de calibración se obtiene una estimación de 2.114.567 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.

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



Fuente: Instituto Nacional de Estadísticas

<sup>15</sup> El umbral de calidad del coeficiente de variación se definió en 15% al interior de cada marginal.

### III. ESTIMACIÓN DE LA VARIANZA

Como se mencionó en capítulos precedentes, el diseño muestral de la VIII EME, 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

---

Diversos paquetes estadísticos permiten estimar errores muestrales bajo diseños complejos, utilizando métodos como la linealización de Taylor, técnicas de replicación (Jackknife, Bootstrap), entre otros. Para facilitar su implementación, es necesario considerar ciertos supuestos: se asume que la selección de unidades en las distintas etapas del muestreo es independiente, lo que simplifica tanto los cálculos como las expresiones matemáticas. Además, aunque el diseño muestral pueda incluir múltiples etapas, generalmente se considera solo la primera, ya que esta aporta la mayor parte de la variabilidad al error total.

En el entorno de R, antes de estimar la variable de estudio —que en esta encuesta corresponde a la estructura de las actividades en las que se desempeñan los microemprendedores— y sus errores

asociados, es fundamental definir el diseño muestral sobre el cual se realizarán las estimaciones. Las variables presentes en la base de datos que definen el diseño muestral de la VIII EME son:

1. *fact\_cal*: Factor de expansión que da cuenta de las probabilidades de selección en las fases 1 y 2, ajustado por elegibilidad y no respuesta, suavizado por región y rama de actividad económica, y calibrado.
2. *Estrato*: Estrato de la Encuesta Nacional de Empleo (ENE), identificado como ESTRATO\_ENE.
3. Conglomerado: Unidad primaria de muestreo, identificada como id\_upm.

Para analizar la estructura de las actividades de los microemprendedores, el investigador debe seguir los siguientes pasos:

1. Determinar y construir la variable de interés, en caso de que no esté previamente definida.
2. Especificar las variables que definen el diseño muestral complejo.
3. Realizar la estimación correspondiente utilizando las funciones adecuadas en R.

En la Tabla III.1 se observa la estructura de la rama de actividad económica (caenes1\_final) para los microemprendedores.

**Tabla III.1. Estructura de la actividad económica de los microemprendedores**

Rama de actividad económica	Estimación	Error estándar	Intervalo de confianza al 95%		Coeficiente de variación	Efecto de diseño
			Inferior	Superior		
<b>Total Nacional</b>	<b>100,0%</b>	<b>0,0%</b>	<b>100,0%</b>	<b>100,0%</b>	<b>0,0</b>	
Sin clasificación	5,5%	0,3%	4,9%	6,2%	0,1	1,7
Agricultura, ganadería, silvicultura y pesca	0,1%	0,0%	0,0%	0,2%	0,6	1,4
Explotación de minas y canteras	6,0%	0,3%	5,5%	6,6%	0,0	1,1
Industrias manufactureras	0,1%	0,0%	0,0%	0,2%	0,5	1,3
Suministro de agua; evacuación de aguas residuales	11,1%	0,5%	10,3%	12,1%	0,0	1,5
Construcción	0,4%	0,1%	0,2%	0,6%	0,3	1,8
Comercio al por mayor y al por menor	10,2%	0,5%	9,4%	11,2%	0,0	1,7
Transporte y almacenamiento	27,9%	0,7%	26,5%	29,2%	0,0	1,7
Actividades de alojamiento y de servicio de comidas	8,0%	0,4%	7,2%	8,9%	0,1	1,8
Información y comunicaciones	4,2%	0,3%	3,6%	4,8%	0,1	1,8
Actividades financieras y de seguros	1,1%	0,2%	0,8%	1,6%	0,2	2,9
Actividades inmobiliarias	0,3%	0,1%	0,1%	0,5%	0,3	1,8
Actividades profesionales, científicas y técnicas	0,9%	0,2%	0,7%	1,4%	0,2	2,4
Actividades de servicios administrativos y de apoyo	6,7%	0,5%	5,9%	7,7%	0,1	2,4
Enseñanza	3,4%	0,3%	2,9%	4,0%	0,1	1,7
Actividades de atención de la salud humana y de asistencia social	2,0%	0,3%	1,6%	2,6%	0,1	2,4
Actividades artísticas, de entretenimiento y recreativas	3,4%	0,3%	2,8%	4,0%	0,1	2,1
Otras actividades de servicios	1,3%	0,2%	0,9%	1,8%	0,2	2,7
Actividades de los hogares como empleadores	7,4%	0,4%	6,6%	8,3%	0,1	1,8
	0,0%	0,0%	0,0%	0,1%	1,0	0,7

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
11. Lograda	6.760	74,3%	SI
12. Entrevista parcial	410	4,5%	SI
21. Informante del hogar rechazó la ent	183	2,0%	SI
22. Informante directo rechazó la entre	641	7,0%	SI
31. Vivienda ocupada, sin moradores pre	240	2,6%	SI
32. Se impide el acceso a la vivienda	9	0,1%	SI
33. Informante no ubicable, se contacta	158	1,7%	SI
34. Informante ausente en período de le	51	0,6%	SI
41. Informante impedido/a física y/o me	23	0,3%	SI
42. Informante no habla idioma español	2	0,0%	SI
43. Muerte de el/la informante o algún	4	0,0%	SI
45. Información falseada	9	0,1%	SI
49. Otra razón de no entrevista	64	0,7%	SI
52. No fue posible localizar la dirección	3	0,0%	NO
64. Vivienda en demolición, incendiada,	1	0,0%	NO
65. Vivienda desocupada (por ejemplo: e	24	0,3%	NO
68. No existe informante elegible	397	4,4%	NO
69. Otra razón para ser considerada no	8	0,1%	NO
71. Marca tono, pero sin respuesta	28	0,3%	SI
72. Responde buzón de voz	13	0,1%	SI
73. Suena Ocupado	2	0,0%	SI
74. Se encuentra apagado o fuera del área	16	0,2%	SI
75. Número telefónico no existe	2	0,0%	SI
76. Número telefónico equivocado	1	0,0%	SI
81. Calidad del audio mala o intermiten	1	0,0%	SI
82. Lugar o actividad del informante no	2	0,0%	SI
83. Contesta y luego cuelga	42	0,5%	SI

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_VIII_EME_fact_cal)
```

## BIBLIOGRAFÍA

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

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.

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](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.

Guerrero, J., Marín, A. (2020) *Estratificación socioeconómica del marco muestral de viviendas 2017 (MMV 2017) –Implementación en R*