



---

# MANUAL DE USUARIO DE BASES DE DATOS

## SEPTIMA ENCUESTA DE MICROEMPREDIMIENTO

---

LEVANTAMIENTO OFICIAL VII EME 2022

Instituto Nacional de Estadísticas

Junio, 2023

---

**SUBDIRECCIÓN TÉCNICA**

**DEPARTAMENTO DE ESTADÍSTICAS DEL TRABAJO**

**SUBDEPARTAMENTO DE ESTADÍSTICAS ESTRUCTURALES DEL TRABAJO**

**DEPARTAMENTO DE METODOLOGÍA E INNOVACIÓN ESTADÍSTICA**

**SUBDEPARTAMENTO DE IMPLEMENTACIÓN DE DISEÑOS**

---

**MANUAL DE USUARIO DE BASES DE DATOS VII EME**

**Junio de 2023**

**Instituto Nacional de Estadísticas**

*El uso de un lenguaje que no discrimine ni marque diferencias entre hombres y mujeres ha sido una preocupación en la elaboración de este documento. Sin embargo, y con el fin de evitar la sobrecarga gráfica que supondría utilizar en castellano “o/a” para marcar la existencia de ambos sexos, se ha optado por utilizar el masculino genérico, en el entendido de que todas las menciones en tal género representan siempre a hombres y mujeres, abarcando claramente ambos sexos.*

# Indice

<b>1. INTRODUCCIÓN</b>	<b>4</b>
<b>2. PRESENTACIÓN SÉPTIMA ENCUESTA DE MICROEMPENDIMIENTO</b>	<b>5</b>
2.1. Antecedentes . . . . .	5
2.2. Objetivos . . . . .	6
2.2.1. Objetivo general . . . . .	6
2.2.2. Objetivos específicos de la encuesta . . . . .	6
2.3. Ficha técnica . . . . .	7
2.4. Cuestionario . . . . .	9
2.5. Descripción de módulos . . . . .	9
<b>3. ESTRUCTURA DE LA BASE DE DATOS</b>	<b>13</b>
3.1. Base de datos full . . . . .	13
3.2. Base de datos de empleo . . . . .	14
<b>4. INDICACIONES PARA EL USO DE LA BASE DE DATOS</b>	<b>15</b>
4.1. Especificación del diseño muestral complejo . . . . .	15
4.1.1. Descripción del diseño muestral . . . . .	15
4.1.2. Variable que identifican el diseño muestral complejo . . . . .	15
4.1.3. Plan de muestreo . . . . .	16
4.2. Utilización de factores de expansión . . . . .	16
<b>5. INDICADORES Y TABULADOS</b>	<b>18</b>
5.1. Motivación para iniciar el negocio o actividad por cuenta propia . . . . .	18
5.2. Ganancias . . . . .	19
5.3. Tasa de informalidad . . . . .	21
<b>6. PRECISIÓN DE LOS PRINCIPALES INDICADORES</b>	<b>26</b>
6.1. Conceptos utilizados para la evaluación de las medidas de calidad . . . . .	26
6.2. Evaluación de la calidad de las estimaciones . . . . .	28
<b>7. ACCESO, SOLICITUD Y USO DE BASES DE DATOS</b>	<b>30</b>
<b>8. DICCIONARIO DE VARIABLES Y CÓDIGOS</b>	<b>31</b>
<b>9. REFERENCIAS BIBLIOGRÁFICAS</b>	<b>32</b>

# 1. INTRODUCCIÓN

El presente documento está dirigido a los usuarios de las bases de datos de la VII Encuesta de Microemprendimiento (EME) cuyo trabajo de campo se desarrolló entre mayo y agosto de 2022.

El objetivo de este manual es introducir al lector en las bases de datos de la encuesta para que investigadores y usuarios en general puedan replicar algunos de los tabulados oficiales y estudiar la precisión de sus estimaciones. Además, busca entregar herramientas para que las personas puedan adaptar las programaciones aquí ofrecidas y alcanzar sus propios objetivos de investigación.

La asistencia se proporciona mediante el uso del paquete estadístico R y Stata, que permite revisar desde la configuración inicial para trabajar con un diseño muestral complejo hasta la construcción de variables y el desarrollo de gráficos.

Los contenidos del presente documento están estructurados en ocho capítulos, incluida la introducción (capítulo uno). En el capítulo dos se presenta la encuesta, describiendo sus objetivos y el cuestionario utilizado en esta versión. Luego, en el capítulo tres, se describen las dos bases de datos que contienen la información obtenida, mientras en el capítulo cuatro, las indicaciones de uso necesarias, en atención al diseño muestral de la encuesta y sus factores de expansión.

En el capítulo cinco se exhiben algunos ejemplos de programaciones para construir variables y tabular información de interés. En el capítulo seis se aborda las medidas de precisión de los principales indicadores de la encuesta, siguiendo con el capítulo siete donde se describe el proceso de acceso y solicitud de bases de datos, finalizando en el capítulo ocho que contiene el diccionario de variables y códigos de las dos bases de datos de esta versión de la EME.

Si bien esta publicación muestra cómo se desarrolla el cálculo de algunos de los principales resultados de la VII EME, es importante comentar que no busca presentar todos los tabulados posibles de realizar, ni explotar todos los usos de la encuesta, sino que entregar una pequeña guía respecto al uso de los datos con el fin de que las personas puedan utilizar las bases de datos de la mejor forma posible.

## **2. PRESENTACIÓN SÉPTIMA ENCUESTA DE MICRO-EMPRESARIOS**

En esta sección se realizará una presentación de las principales características de la Encuesta de Microempresarios, dentro de las cuales se encuentran los antecedentes y objetivos del estudio, la ficha técnica, el cuestionario y la descripción de sus módulos.

### **2.1. Antecedentes**

La Encuesta de Microempresarios (EME) tiene sus inicios en el año 2006 a partir de una consultoría sobre microempresas informales que permitió realizar una primera caracterización de negocios no registrados ante las autoridades tributarias, los que no son visibilizados por las encuestas tradicionales a empresas.

En el año 2009 se llevó a cabo la primera versión de la EME, que entregó una caracterización de las microempresas a lo largo del país. La segunda versión fue ejecutada por el Centro de Microdatos en el año 2011. A partir del 2013 el INE se hace cargo del proceso estadístico de la encuesta, desarrollando desde el diseño del cuestionario hasta la publicación de los datos.

Entre los diversos cambios implementados por el INE el año 2013, se redujeron considerablemente los contenidos del cuestionario, pasando de un total de 200 preguntas y una hora de duración de la entrevista, a 68 preguntas con tiempo de entrevista de aproximadamente 22 minutos. Esto permitió reducir el rechazo y desgaste de los informantes, mejorando las tasas de respuesta y logro de la encuesta.

Entre las modificaciones al cuestionario, destaca la creación del módulo E “Situación de Formalidad del negocio”, el que permite identificar diferentes grados de informalidad de las unidades económicas y aportar así con insumos para la generación de políticas públicas que incentiven la formalización de los micronegocios. Para ello, se incluyeron once preguntas, que posibilitan realizar un primer análisis de informalidad laboral en base a los lineamientos promovidos por la Organización Internacional del Trabajo (OIT).

Con el fin de enfocar el análisis de la encuesta sólo a los microempresarios, en la V EME (2017) se modificó la población objetivo del estudio de trabajadores independientes a microempresarios, es decir, trabajadores independientes dueños de una unidad económica con hasta 10 empleados o trabajadores (incluyendo al propio dueño), lo que contribuyó con mejorar la precisión de las estimaciones<sup>1</sup> y, al mismo tiempo, vincular el nombre del estudio más estrechamente con la medición que efectivamente se realiza.

---

<sup>1</sup>Producto de que la EME es una encuesta mixta que se ancla a una encuesta de hogares como es la Encuesta Nacional de Empleo (ENE), no es posible estimar de manera precisa el número de unidades económicas de mayor tamaño, las que típicamente son obtenidas a través de encuestas a establecimientos cuya muestra es extraída del directorio de empresas. Esto ocurre debido a que las viviendas donde residen los dueños de estas unidades económicas de unidades de mayor tamaño (mediana o grandes) tienen menor probabilidad de ser seleccionadas en la EME, ya que el marco de muestreo proveniente de la ENE tiene limitantes de acceso al estrato socioeconómico alto donde comúnmente residen estos empresarios.

El cambio de la población objetivo facultó la incorporación de una grilla de trabajadores, que da la posibilidad de aproximarse a las principales características del personal empleado por la persona microempresaria. La estructura de la información recopilada se repite para cada uno de los empleos generados por el informante, por ello desde la V versión, se decidió publicar la información de estas preguntas en una base de datos independiente, en la que cada fila representa la información de un empleado y no de un microempresario.

En la VI EME, se incorporó preguntas para capturar con mayor detalle la información de las ganancias de las unidades económicas, el valor agregado de éstas y los salarios de los trabajadores. Además, se incluyó un módulo de trabajo no remunerado que permite conocer el tiempo destinado a las actividades de trabajo de cuidado y doméstico no remunerado.

En la versión actual se incorporaron preguntas, referidas a la pandemia del COVID-19 y al uso de tecnologías de la información. Lo que permite poder capturar y estudiar el impacto de la pandemia del COVID-19 en el funcionamiento del microempresariado, sus ganancias, sus trabajadores y la persistencia en el tiempo, y además poder estimar la propagación del uso de tecnologías de información, la prevalencia de éstas y el efecto que tienen en el encadenamiento productivo de los microempresarios.

## 2.2. Objetivos

La Encuesta de Microempresariado (EME) tiene como objetivo principal realizar una caracterización profunda de los microempresarios 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.

### 2.2.1. Objetivo general

El objetivo general de la encuesta es profundizar en el conocimiento de las características que definen los microempresarios del país, limitando el análisis a aquellos de menor escala. Es decir, busca caracterizar la heterogénea realidad de los Microempresarios del país, sus negocios y su evolución en el tiempo. Esto considerando, una muestra representativa de viviendas particulares que contienen al menos a un microempresario, a nivel nacional y regional.

### 2.2.2. Objetivos específicos de la encuesta

- 1. Observar las formas en las que operan y funcionan las unidades económicas:** la encuesta permite identificar la escala de funcionamiento de los negocios. Por tanto, se puede estudiar la organización y planificación del trabajo de los microempresarios, así como también los factores utilizados en el proceso productivo y la capacidad de poder realizar una separación de ellos.
- 2. Estudiar el grado de formalidad de los micronegocios:** la principal virtud de la encuesta es que permite caracterizar a los micronegocios per-

tenecientes al sector informal de la economía. Para ello, el módulo E, contiene una serie de preguntas enfocadas en conocer la situación de formalidad de los negocios.

**3. Cuantificar el acceso al sistema financiero y cómo se vinculan con éste:** la encuesta permite identificar cómo los microemprendedores financian sus negocios, por ejemplo, si es por medio de créditos y el tipo de ellos. Además, posibilita la estimación del nivel de deuda de los micronegocios.

**4. Determinar los recursos productivos de los micronegocios:** otra virtud de la encuesta es que permite determinar el patrimonio de los microemprendedores, debido a que captura información de herramientas, equipos y vehículos que los micronegocios destinan para fines de su actividad económica. De tal manera que se puede saber sus valores y posibles inversiones futuras en nuevos insumos.

**5. Estimar los ingresos y gastos de los micronegocios:** parte importante de las brechas de los salarios o ganancias obtenidas por los trabajadores o dueños se explican por las brechas en productividad de los micronegocios, en este sentido, la encuesta considera una serie de preguntas que permiten aproximarse al cálculo del valor agregado de los micronegocios y con ello a la estimación general de algunas medidas de productividad.

**6. Identificar la generación de empleo de las microempresas y sus características:** algunos de los aspectos relevantes que permite estudiar la encuesta, son la existencia de empleo ocasional, las relaciones entre los empleados y dueños, la distribución de la carga laboral y si los trabajadores tienen un contrato con garantías legales.

**7. Estudiar el impacto del COVID-19 en los micronegocios:** la encuesta permite indagar sobre impactos del COVID-19 en el funcionamiento del microemprendimiento, sus ganancias, sus trabajadores y la persistencia en el tiempo.

### 2.3. Ficha técnica

La siguiente sección muestra un resumen de los principales ámbitos de la ficha técnica del producto. Para más detalles, se puede descargar la versión completa, desde la sección de metadatos del micrositio de la encuesta<sup>2</sup>.

<sup>2</sup>Link <https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral/microemprendimiento>.

Tabla 1: Ficha técnica VII Encuesta de Microemprendimiento

<b>Dimensión</b>	<b>Descripción</b>
<b>Población objetivo</b>	Trabajadores por cuenta propia o personas empleadoras dueñas de una microempresa con hasta 10 trabajadores y/o trabajadoras (incluyendo a la o el informante dueño de la unidad económica), denominadas en su conjunto como personas microempendedoras, que residen en viviendas particulares ocupadas dentro del territorio nacional.
<b>Fuente de información</b>	Encuesta a hogares aplicada al informante directo
<b>Diseño muestral</b>	<p>Diseño muestral bifásico: en la primera fase se identificaron las viviendas donde reside al menos un trabajador independiente, a través de la Encuesta Nacional de Empleo (ENE). Luego, en la segunda fase, se realiza una encuesta a trabajadores por cuenta propia y empleadores dueños de una unidad económica con hasta 10 trabajadores (incluyendo al dueño de la unidad económica).</p> <p>Período de referencia: se obtiene de la muestra: trimestre marzo-mayo de 2022 de la Encuesta Nacional de Empleo.</p> <p>Representatividad a nivel regional y nacional. error muestral teórico: 1,4 % para el estimador principal (proporción de personas que trabajan por cuenta propia sobre el total de microempendedores)</p> <p>Muestra teórica: 7.850 viviendas y 8.576 informantes.</p>
<b>Tamaño de la muestra</b>	<p>Muestra objetivo: 6.479 viviendas.</p> <p>Muestra lograda: 6.772 viviendas y 6.942 informantes.</p>
<b>Periodo de levantamiento</b>	La encuesta fue levantada entre la segunda semana de mayo hasta la primera de agosto de 2022 (3 meses).
<b>Cobertura geografica</b>	Cobertura del área urbana y rural a nivel nacional
<b>Modo de aplicación</b>	Mixta, modalidad telefónica o presencial, dependiendo de la disponibilidad y preferencia del informante
<b>Medio de aplicación</b>	Dispositivo de captura móvil (DMC)
<b>Duración promedio</b>	31,8 minutos por cuestionario
<b>Cobertura temática</b>	Encuesta compuesta por 12 módulos, con 107 preguntas en total.
<b>Otros</b>	<p>Entrevista voluntaria</p> <p>No se solicitan datos personales</p> <p>Bases de datos de acceso público</p>

Fuente: INE, Ficha Técnica VII EME



## 2.4. Cuestionario

El cuestionario de la VII Encuesta de Microemprendimiento, cuenta con 107 preguntas distribuidas en 12 módulos. La actualización del cuestionario con respecto a la versión anterior de la encuesta (VI EME 2019) mantuvo en gran parte la estructura y contenido. El número de módulos y el contenido de ellos no cambió significativamente (ver Tabla 2). Debido a la identificación de nuevas necesidades, el número total de preguntas aumentó de 97 a 107. Estas necesidades tienen que ver con tres dimensiones: (1) Impacto de la pandemia covid-19 en microemprendimientos, (2) Tecnologías de información (uso de canales digitales) y (3) Ajuste en la medición de trabajo no remunerado.

Tabla 2: Cambios cuestionario VII EME

AREAS DE ESTUDIO	VI EME	VII EME	Eliminadas	Agregadas
Situación Laboral	12	13	0	1
Historia Laboral	9	9	1	1
Características del Negocio o Actividad por Cuenta propia	9	9	0	0
Gastos, Ingresos y Ganancias	12	13	1	2
Situación de Formalidad	10	11	0	1
Generación de empleo	4	5	0	1
Relación con el sistema Financiero	13	10	4	1
Herramientas y Equipos	4	4	0	0
Tecnología de la información	4	9	1	6
Capacitación	4	4	0	0
Entorno del negocio	13	15	0	2
Trabajo No Remunerado	3	5	1	3
<b>Total</b>	<b>97</b>	<b>107</b>	<b>8</b>	<b>18</b>

Fuente: INE

## 2.5. Descripción de módulos

### Módulo A "Situación laboral"(13 preguntas)

Este módulo es fundamental para la encuesta, ya que permite clasificar y filtrar a la población objetivo de la EME. Esto lo hace a través de la pregunta principal (A1), que permite conocer si el/la informante continúa con la actividad por cuenta propia

o de empleador/a declarada en el 2022. Adicionalmente, en esta versión, se han incluido nuevas preguntas, las que refieren a la caracterización de aquellos informantes que han dejado sus empleos como trabajadores por cuenta propia y si el cambio de situación laboral se relaciona con la pandemia del COVID-19 (A2).

### **Módulo B “Historia laboral” (9 preguntas)**

Este módulo indaga en la historia laboral del o la informante, incluyendo preguntas relacionadas con su(s) actividad(es) económica(s) actual(es) y pasada(s). De tal forma, se obtiene información sobre la antigüedad laboral, actividades económicas paralelas, trabajo asalariado, entre otras. Adicionalmente, en esta versión, se puede indagar si los motivos de inicio de la actividad económica tienen relación con la pandemia del COVID-19 (B1).

### **Módulo C “Características del negocio o actividad por cuenta propia” (9 preguntas)**

El módulo cuenta con 9 preguntas, que permiten caracterizar de manera general el negocio. Entre las preguntas principales, se incluyen las dirigidas a conocer la actividad principal (C1), el lugar donde se realiza, la cantidad de socios/as o dueños/as según sexo (C3) y los horarios y días de trabajo (C6).

### **Módulo D “Gastos, ingresos y ganancias” (13 preguntas)**

Este módulo permite caracterizar de manera específica y desagregada los gastos, ingresos y ganancias de los negocios. Revisando, por ejemplo, en los gastos de arriendo (D1), servicios básicos, materias primas, etc. (D2). Conjuntamente se indaga en características de los ingresos como periodicidad, monto promedio y días que se recibe esos ingresos, así como también, en las ganancias obtenidas de la venta de productos o servicios del negocio, con quién se comparte esta utilidad y el porcentaje que le corresponde al hogar del o la informante. Adicionalmente, en esta versión, se permite indagar en efectos de la pandemia COVID-19 en ganancias (D8).

### **Módulo E “Situación de formalidad del negocio o actividad por cuenta propia” (11 preguntas)**

Este módulo está formado por 11 preguntas, a través de las cuales se caracteriza la situación de formalidad del negocio. Es decir, se pregunta principalmente si existen registros contables (E1), si se ha iniciado actividades en el Servicio de Impuestos Internos (SII) (E3), si se declara IVA, impuesto a la renta y/o permisos sanitarios (E7) y si se cotiza en salud, AFP y/o seguro de accidentes, entre otros aspectos.

### **Módulo F “Generación de empleo” (5 preguntas)**

Este módulo consta de 5 preguntas de gran importancia para la encuesta, ya que a través de ellas se puede identificar, cuantificar y caracterizar el empleo que generan los microemprendimientos del país. De tal manera, se obtiene información acerca de trabajadores/as empleados/as o contratados/as en cada negocio, permitiendo clasificarlos/as según sexo, categoría ocupacional, tipo de acuerdo, entre otros. Conjuntamente se indaga por variaciones internas producto de la pandemia COVID-19, los efectos producidos a nivel del personal, salarios, horarios y teletrabajo (F4)

### **Módulo G “Relación con el sistema financiero” (10 preguntas)**

En este módulo cuenta con 10 preguntas, que permiten conocer y caracterizar la relación del negocio/actividad con distintas instituciones financieras, específicamente en lo relacionado a solicitudes de préstamos o créditos para fines del negocio. Dentro de las principales preguntas, se encuentran aquellas que hacen referencia a la solicitud del préstamo/crédito (G1), las deudas que actualmente mantiene (G3), solicitud de crédito producto de pandemia COVID-19 (G9) y el financiamiento de los gastos regulares del negocio (G10).

### **Módulo H “Herramientas y equipos” (4 preguntas)**

Módulo conformado por 4 preguntas que permiten cuantificar y caracterizar los equipos, herramientas o vehículos con los que cuenta el negocio para su funcionamiento. Además se indaga en posibles inversiones futuras en insumos.

### **Módulo I “Tecnologías de la información” (9 preguntas)**

Este módulo indaga acerca de las tecnologías de información utilizadas para el desarrollo de la actividad económica. Captura información acerca del uso y acceso a internet, si se utiliza como medio de comunicación y de difusión del negocio o actividad por cuenta propia. En esta versión producto de la pandemia COVID-19, se incorporaron nuevas preguntas, que permiten profundizar en el estudio de la digitalización de los negocios.

### **Módulo J “Capacitación” (4 preguntas)**

Módulo que indaga en las capacitaciones y/o talleres de perfeccionamiento que ha recibido el negocio para la actividad económica que realiza. Cuenta con 4 preguntas, que se basan en caracterizar dicha capacitación, considerado, por ejemplo, el principal beneficio obtenido de esta (J2) y la fuente de financiamiento (J3).

### **Módulo K “Entorno del negocio o actividad por cuenta propia” (15 preguntas)**

Este módulo indaga en la relación de negocios con distintas instituciones de financiamiento y/o capacitación del entorno, y con actores involucrados en el proceso de venta de sus productos o servicios, tales como clientes y proveedores. Cuenta con 15 preguntas; dentro de las cuales, destacan las referidas a conocer los beneficios y limitantes de trabajar de manera independiente (K1 y K2), la relación con los/as clientes/as y proveedores/as principales (K4, K5 y K8) y el conocimiento y utilización de servicios/beneficios y programas/beneficios de distintas instituciones (K12, K13, K14 y K15).

### **Módulo L “Trabajo no remunerado” (5 preguntas)**

Este módulo está destinado a indagar sobre la realización de una serie de actividades, sin pago, relacionadas con el trabajo doméstico. Consta de cinco preguntas, las cuales permiten caracterizar las actividades no remuneradas según la cantidad de días que se realiza a la semana y el tiempo que demora el/la informante en realizarlas.

### 3. ESTRUCTURA DE LA BASE DE DATOS

La información recolectada durante el trabajo de campo de la VII EME, está dividida en dos bases de datos. La primera es la denominada “Base de datos full”, en ella se encuentran todas las preguntas del cuestionario con la excepción de la pregunta F2. La segunda base de datos corresponder a la “Base de datos de empleo” y contiene la información recabada a través de la variable F2, es decir, la caracterización general de los trabajadores que se desempeñan en los negocios o actividades de los microemprendedores.

#### 3.1. Base de datos full

La base de datos full contiene 425 variables, con un total de observaciones que asciende hasta las 6.942 personas. Las variables de esta base son un reflejo de la información entregada por el informante entrevistado al encuestador, con la ya mencionada ausencia de las variables asociadas a la pregunta F2, se detallan las variables de esta base de datos además de su estructura (ver Tabla 3)

**Variables sociodemográficas:** esta sección cuenta con un identificador único de personas y las variables de caracterización sociodemográficas del informante como su nivel de educación y la región a la que pertenece.

**Variables sobre la historia laboral:** estas recogen el tiempo que se dedicó, en caso de haber cambiado de condición, o el que se ha dedicado el microemprendedor a su negocio o actividad propia, así como también las motivaciones para tener un negocio o actividad por cuenta propia, el financiamiento inicial, entre otros elementos.

**Caracterización del negocio o actividad por cuenta propia:** esta sección cuenta con variables que permiten perfilar el tipo de negocio o actividad por cuenta propia llevada a cabo por los informantes, describiendo el rubro al que pertenecen, el lugar donde trabaja, así como el detalle de los gastos, ingresos y ganancias registrados en los últimos doce meses. También se encuentran variables, que permiten describir el nivel de formalidad del negocio o actividad por cuenta propia, así como su capacidad para generar empleo y su relación con el sistema financiero.

**Trabajo no remunerado:** estas variables contienen información respecto a una serie de actividades que el informante podría o no realizar en su hogar sin un pago a cambio.

Tabla 3: Ejemplo estructura Base de datos FULL VII EME

Enc_rphSEXO	TRAMO_ ETARIO	REGION	a1	...	ganancia_ final	ESTRATO	ID_UPM	FACTOR_ EME
13170C	2	4	12	1	1560000	12000220	1240120008	99.7383
226235	1	3	15	1	12000000	15000120	1510120034	154.8417
291011	2	3	13	1	6000000	13100120	1312420009	53.4691

Fuente: Base de datos Full VII EME

### 3.2. Base de datos de empleo

La base de empleo se compone de once variables, con un total de 2.432 observaciones correspondientes a personas que trabajan en los negocios o actividades por cuenta propia que realizan los informantes de la VII EME. La base contiene variables de identificación que permiten asociar a los trabajadores con los entrevistados que declararon sus datos en la encuesta. Las variables y estructura de esta base de datos se detallan a continuación:

- Identificador único de la persona encuestada
- Número de línea en f2 de persona seleccionada
- Miembro del hogar
- Sexo trabajador
- Edad trabajador
- Nivel educacional alcanzado
- Categoría ocupacional
- Tipo de acuerdo
- Monto total bruto recibido
- Horas semanales trabajadas
- Indicador de submuestra de recolección

Tabla 4: Ejemplo estructura Base de datos de Empleo VII EME

Enc_rph	f2_nlinea	f2_b	f2_c	f2_d	f2_e	f2_f	f2_g	f2_h	f2_i	SUBMUESTRA
131704	1	1	1	36	3	2	3	99	8	1
137068	2	2	2	66	2	1	2	450000	24	1
86581	1	2	2	21	4	1	1	300000	86	2
200288	2	2	2	30	3	1	2	700000	45	2
220826	1	2	1	50	2	1	1	400000	45	3
291011	1	1	2	18	3	1	1	400000	25	3

Fuente: Base de datos de Empleo VII EME

## **4. INDICACIONES PARA EL USO DE LA BASE DE DATOS**

En el presente capítulo se entrega al usuario elementos para hacer un adecuado uso de las bases de datos de la VII EME. Es importante destacar, que en esta versión de la encuesta se incluyen, sólo en la base full, las variables de Estrato de varianza y Conglomerado de varianza al que pertenece cada hogar, por lo que sólo a partir de la información de esta base, se pueden realizar estimaciones considerando el muestreo complejo utilizado en la encuesta. A continuación, se presenta la especificación del diseño muestral complejo, junto con la definición y utilización del factor de expansión de la encuesta.

### **4.1. Especificación del diseño muestral complejo**

El objetivo principal de la EME es lograr una caracterización profunda de los microemprendimientos que se desarrollan a nivel nacional, ya sean formales o informales. Para esto, se realiza una encuesta a hogares que toma una muestra representativa de viviendas particulares ocupadas a nivel nacional y regional. Este tipo de encuestas permite observar unidades económicas de menor tamaño, las cuales en muchos casos no están inscritas en el Servicios de Impuestos Internos (SII) por lo que quedan excluida de las encuestas tradicionales a establecimientos.

#### **4.1.1. Descripción del diseño muestral**

El diseño muestral de la VII EME es bifásico, es decir, cuenta de dos fases. En la primera, se levanta una encuesta a hogares denominada Encuesta Nacional de Empleo (ENE), para capturar información de una primera población objetivo, que sirve para identificar a los individuos que compondrán la segunda fase. Esta primera fase del diseño posee un muestro probabilístico, estratificado y bietápico, donde las unidades primarias de muestreo corresponden a manzanas en el área urbana y secciones en el área RAU - rural: mientras que, las unidades secundarias de muestreo son las viviendas particulares ocupadas. En la segunda fase, se clasifican las viviendas en dos grupos, de acuerdo a si éstas 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 microemprendedores.

#### **4.1.2. Variable que identifican el diseño muestral complejo**

Para evaluar la precisión de las estimaciones se requiere incorporar las características del diseño muestral en el cálculo de la varianza, ya que un cálculo genérico (tipo muestreo aleatorio simple) introducirá sesgos en la estimación del error. La VII EME presenta un diseño muestral probabilístico, estratificado y bietápico. Para identificar las distintas etapas del muestreo se presentan en la base full las variables ESTRATO y ID\_UPM.(Tabla 3).

Tabla 5: Variables para la declaración del diseño muestral

Variable	Nombre	Etiqueta
Conglomerado	ID_UPM	Conglomerados
Estrato	ESTRATO	Estratos

Fuente: INE, Base de datos VII EME

### 4.1.3. Plan de muestreo

Al identificar el plan de muestreo es posible generar estimaciones adecuadas de los parámetros de interés y sus respectivas varianzas. Esto puede realizarse utilizando las variables que identifican el diseño muestral complejo (ID\_UPM y ESTRATO) y el factor de expansión (FACTOR\_EME), a través del siguiente comando:

#### Programación 1: Identificación del diseño muestral complejo

##### Programa Stata

```
svyset ID_UPM[w=FACTOR_EME], strata(ESTRATO) single(certainty)
```

##### Programa R

```
options(survey.lonely.psu = "certainty")

emevii_srvy <- data %>%
  as_survey_design(
    ID_UPM,
    strata = ESTRATO,
    weight = FACTOR_EME)
```

De no realizarse este paso, el programa entenderá que se trata de un muestreo aleatorio simple.

## 4.2. Utilización de factores de expansión

El factor de expansión representa la cantidad de unidades en la población a las que representa cada unidad presente en la muestra. De acuerdo con el diseño muestral de la encuesta, el factor de expansión se calcula en función del total de manzanas y viviendas seleccionadas en cada estrato, así como también del número de manzanas totales en el estrato (según información reportada por el Marco Muestral de Viviendas) y del número de viviendas particulares enumeradas. La variable que contiene el factor de expansión elaborado para esta versión de la EME está presente en la base full y su nombre se muestra a continuación:



Tabla 6: Factor de expansión VII EME

Variable	Nombre	Etiqueta
Factor de Expansión	FACTOR_EME	Factor de Expansión VII EME

Fuente: INE, Base de datos VII EME

El factor de expansión debe estar especificado en cada análisis que se programe con el software Stata, con el fin de conseguir resultados a nivel poblacional de la encuesta. Si es que no se considera el factor de expansión en los análisis, los resultados obtenidos equivaldrían a realizar análisis considerando sólo los resultados de la muestra. Para ejemplificar esto, en la programación 2 se muestra la especificación del factor de expansión, a través de un tabulado para la variable sexo. Es importante considerar que al programar en Stata, que si no se agrega la indicación entre corchetes, el programa no pondera los casos por el factor. Al programar en el software R, para conseguir resultados a nivel poblacional de la encuesta, se debe especificar en cada análisis la función correspondiente del paquete “survey”, tal como se observa a continuación, para calcular el número total de microemprendedores por sexo.

## Programación 2: Uso de factor de expansión

### Programa Stata

```
*Total de microemprendedor por sexo
tab sexo [iw=FACTOR_EME]
```

### Programa R

```
# Microemprendedores por Sexo ----
tab_eme_region <- emevii_srvy %>%
  group_by(SEX0) %>%
  dplyr::summarize(
    micro = round(survey_total(vartype = NULL)),
    micro_porc = round(survey_prop(vartype = "se")*100, 1)
  ) %>% print()
```

## 5. INDICADORES Y TABULADOS

A continuación, se ofrecen alternativas para la construcción de algunos tabulados que se pueden obtener con las bases de datos de la VII EME. Es importante señalar que el siguiente es un ejercicio que busca entregar ejemplos de lo que es posible construir con algunas de las variables contenidas en esta versión de la encuesta, pero la utilidad y alcance de la información recogida no se acaba en estos. En este sentido, este manual busca orientar a los usuarios que usen las bases de datos en el manejo y tratamiento de la información, con el fin de reducir las complejidades que esto les pueda implicar. Se advierte que si bien gran parte de las programaciones que se ofrecen a lo largo del documento pueden ejecutarse a través de la consola de los software Stata y R (interfaz principal del programa), están programadas en un do file y Script. Por ende, se recomienda ejecutarlas a través de dicho ambiente de programación.

### 5.1. Motivación para iniciar el negocio o actividad por cuenta propia

Una de las variables interesantes de revisar es la motivación que tuvieron los informantes para ser empleadores o trabajadores por cuenta propia. En la Base Full se puede encontrar la variable b2 que contempla las respuestas a la pregunta. “¿Cuál fue la motivación principal por la cual inició su actual negocio o actividad por cuenta propia?”. Esta variable cuenta con doce alternativas más las opciones “Otro, especifique”, “No sabe” y “No responde” siendo presentada usualmente en una versión con un número reducido de categorías. A continuación, se expone la programación con la que se puede lograr una tabla que describa los resultados de la encuesta clasificando la motivación en 6 grupos: 1 “Tradición familiar”, 2 “Necesidad”, 3 “Oportunidad”, 4 “Formación”, 77 “Otro, especifique” y 96 “Sin dato”.

#### Programación 3: Motivación para iniciar el negocio o actividad por cuenta propia por sexo

##### Programa Stata

```
* Generar variable motivación
gen motivacion_eme=.
replace motivacion_eme = 1 if (b2==1)
replace motivacion_eme = 2 if (b2==2) | (b2==3) | (b2==4) | (b2==7) |
(b2==10) | (b2==11)
replace motivacion_eme = 3 if (b2==5) | (b2==6) | (b2==8) | (b2==9)
replace motivacion_eme = 4 if (b2==12)
replace motivacion_eme = 77 if (b2==77)
replace motivacion_eme = 96 if (b2==88) | (b2==99) | (b2==96)
*Etiquetar variable
lab var motivacion_eme "Motivación principal para emprender"
label define motivacion_eme 1 "Tradición familiar" 2 "Necesidad"
3 "Oportunidad" 4 "Formación" 77 "Otro, especifique" 96 "Sin dato"
label values motivacion_eme motivacion_eme
```

```
label variable motivacion_eme "Motivación principal para emprender"
*Tabulado
tab motivacion_eme SEXO [iw=FACTOR_EME] if inlist(a1,1,2)
```

## Programa R

```
# Generar variable motivación
emevii_srvy$motivacion <- car::recode(emevii_srvy$b2, "
    1 = 1;
    c(2,3,4,7,10,11) = 2;
    c(5, 6, 8, 9) = 3;
    12 = 4;
    c(88,99,96) = 96;
    77 = 77")

emevii_srvy$motivacion <- factor(emevii_srvy$motivacion,
    levels = c(1,2,3,4,77,96), labels = c("Tradición Familiar",
    "Necesidad",
    "Oportunidad",
    "Formación",
    "Otro",
    "Sin Dato"))

### Etiquetar

eme$motivacion <- sjlabelled::set_label(eme$motivacion,
    label = "Motivacion para el Negocio")

# Tabulado
tab_eme_sexo_motivacion <- emevii_srvy %>%
    group_by(SEXO, motivacion) %>%
    dplyr::summarize(
    micro = round(survey_total(vartype = NULL)),
    micro_porc = round(survey_prop(vartype = "se")*100, 2)
    ) %>% print()
```

## 5.2. Ganancias

En la Base Full es posible encontrar la variable que expone las ganancias de los microemprendedores encuestados en la variable ganancia\_final. A continuación, se presenta una programación que permite hacer una tabla que muestra los resultados de esta variable según sexo.

### Programación 4: Ganancias mensuales expresadas tramos de salarios mínimos por sexo

#### Programa Stata

```

* Generar variable ganancia
***Para crear variable ganancia
*Primer grupo hasta 1 salario mínimo
*Segundo grupo de 1 a 2 salarios mínimos (SM)
*Tercer grupo de 2 a 3 SM
*Cuarto grupo de 3 a 4 SM
*Quinto grupo de 4 a 5 SM
*Sexto más de 5 SM
**Ganancias respecto al salario mínimo
gen salario_minimo = 380000
gen ganancia_mens= ganancia_final/12
label variable ganancia_mens "Ganancia mensualizada"
gen tramos_sm = 1 if ganancia_mens <= 380000
replace tramos_sm = 2 if inrange(ganancia_mens,380001,760000)
replace tramos_sm = 3 if inrange(ganancia_mens,760001,1140000)
replace tramos_sm = 4 if inrange(ganancia_mens,1140001,1520000)
replace tramos_sm = 5 if inrange(ganancia_mens,1520001,1900000)
replace tramos_sm = 6 if ganancia_mens >= 1900001
label define tramos_sm 1 "Hasta 1 S.M." 2 "Más de 1 S.M.
hasta 2 S.M." 3 "Más de 2 S.M. hasta 3 S.M." 4 "Más de 3 S.M.
hasta 4 S.M." 5 "Más de 4 S.M. hasta 5 S.M." 6 "Más de 5 S.M."
label values tramos_sm tramos_sm
**Tabulado de ganancia
tab tramos_sm SEXO [iw=FACTOR_EME], row

```

## Programa R

```

## Mensualizar
eme$ganancia_final_men <- eme$ganancia_final/12

## Crear variables de tramos de ganancia
ganancias_minimas<-min(eme$ganancia_final_men, na.rm = T)
ganancias_maximas<-max(eme$ganancia_final_men, na.rm = T)
eme$tramos_ganancias <- car::recode(eme$ganancia_final_men,"
ganancias_minimas:0 = 1;
0:380000 = 1;
380001:760000 = 2;
760001:1140000 = 3;
1140001:1520000 = 4;
1520001:1900000 = 5;
1900000:ganancias_maximas= 6")

eme$tramos_ganancias <- factor(eme$tramos_ganancias,
levels = c( 1, 2, 3, 4, 5,
6), labels=c("Hasta 1 SM",
"Entre 1 SM y 2 SM",

```

```

"Entre 2 SM y 3 SM",
"Entre 3 SM y 4 SM",
"Entre 4 SM y 5 SM",
"Más de 5 SM"))

### Etiquetar

eme$tramos_ganancias <- sjlabelled :: set_label(eme$tramos_ganancias,
label="Tramos de Ganancias")

# Tabulado
##Tramo ganancia por sexo
tab_eme_sexo_ganancias <- emevii_srvy %>%
  group_by(SEXO, tramos_ganancias) %>%
  dplyr::summarize(
    micro = round(survey_total(vartype = NULL)),
    micro_porc = round(survey_prop(vartype = "se")*100, 1))%>% print()

```

### 5.3. Tasa de informalidad

Un indicador relevante que se puede calcular con la Base Full es el número de microempresarios informales. A continuación, se presenta una programación que permite hacer una tabla que muestra los resultados de esta variable según sexo.

#### Programación 5: Microempresarios informales por sexo<sup>3</sup>

##### Programa Stata

```

*La "Informalidad" se construye a partir de dos criterios: 1)No llevar
*un registro de cuentas y 2) No estar inscrito en el SII.
*Primero se construirán estas variables, para luego construir la
*variable de interés.
*Proc Variable registro_UE
gen registro_UE = .
replace registro_UE = 1 if inrange(e3,1,4)
replace registro_UE = 0 if inrange(e3,5,6)
replace registro_UE = 999 if inrange(e3,88,99)
label define registro_UE 1 "Con inscripción" 0 "Sin inscripción"
999 "No Sabe/No responde/Sin dato"
label values registro_UE registro_UE
cap la var registro_UE `'"Registro de la unidad económica ante SII"'

```

<sup>3</sup>Para aquellos trabajadores independientes a quienes no se les pueda asignar un "tipo de sector", al que pertenece su unidad económica, mediante la variable registro en SII o existencia contabilidad, se utilizará como primera variable proxy "el lugar de trabajo". Que consiste en analizar la correlación entre el Sector Formal (Informal) y la alternativa de la pregunta c8. En concreto, se observan las prevalencias de las alternativas de respuesta de la variable lugar\_trabajo a quienes sí se les asignó la formalidad del sector sobre la base de las preguntas de informalidad. De esta forma, fue posible obtener, para cada alternativa de la pregunta c8, la proporción de ocupados del Sector Formal e Informal, respectivamente.

```

*''Proc Variable cuentas_UE
gen cuentas_UE = .
replace cuentas_UE = 1 if inrange(e1,1,2)
replace cuentas_UE = 2 if e1==3
replace cuentas_UE = 3 if e1==4
replace cuentas_UE = 999 if inrange(e1,88,99)
label define cuentas_UE 1 "Si, a través de contabilidad formal"
2 "Si, manteniendo registros personales"
3 "No, ningún tipo de contabilidad"
999 "No Sabe/No responde/Sin dato"
label values cuentas_UE cuentas_UE
cap la var cuentas_UE `'"Contabilidad de la unidad económica"'
*''Proc Variable separa_gasto
gen cuentas_UE = .
replace separa_gasto = 1 if e2==1
replace separa_gasto = 2 if e2==2
replace separa_gasto = 999 if inrange(e2,88,99)
label define separa_gasto 1 "Si" 2 "No" 999 "No Sabe/No responde/Sin dato"
label values separa_gasto separa_gasto
cap la var separa_gasto `'"Separación de gastos del negocio con los del hogar"'
*''Crear variable sector_vf
gen sector_vf = .
replace sector_vf = 1 if registro_UE==1
replace sector_vf = 1 if registro_UE==0 & cuentas_UE==1 &
separa_gasto==1
replace sector_vf = 1 if registro_UE==999 & cuentas_UE==1 &
separa_gasto==1
replace sector_vf = 2 if registro_UE==0 & cuentas_UE==1 &
separa_gasto==2
replace sector_vf = 2 if registro_UE==999 & cuentas_UE==1 &
separa_gasto==2
replace sector_vf = 2 if registro_UE==0 & cuentas_UE==2
replace sector_vf = 2 if registro_UE==999 & cuentas_UE==2
replace sector_vf = 2 if registro_UE==0 & cuentas_UE==3
replace sector_vf = 2 if registro_UE==999 & cuentas_UE==3
replace sector_vf = 3 if registro_UE==0 & cuentas_UE==999
replace sector_vf = 3 if registro_UE==999 & cuentas_UE==999
replace sector_vf = 3 if registro_UE==0 & cuentas_UE==1 &
separa_gasto==999
replace sector_vf = 3 if registro_UE==999 & cuentas_UE==1 &
separa_gasto==999
label define sector_vf 1 "Sector Formal" 2 "Sector Informal"
3 "Usar proxy".
label values sector_vf sector_vf
cap la var sector_vf `'"Separacion de gastos del negocio con
los del hogar"'

```

```

*''Luego de incorporar al código el ajuste para las observaciones
*en que se utilizó la variable proxy lugar de trabajo.
*Se crea variable informalidad
gen informalidad = .
replace informalidad = 1 if sector_vf==2
replace informalidad = 0 if sector_vf==1
label define informalidad 1 "Informalidad" 0 "Formalidad"
label values informalidad informalidad
cap la var informalidad `Informalidad del Microempendedor''
*''Tabulado de informalidad por sexo
tab informalidad SEXO [iw=FACTOR_EME], row

```

## Programa R

```

#La "Informalidad" se construye a partir de dos criterios: 1) No llevar
#un registro de cuentas y 2) No estar inscrito en el SII.
#Primero se construirán estas variables, para luego construir la
#variable de interés.
#Creando la variable registro_UE
frq(eme$e3) # Ver frecuencias inscripción en el SII
eme$registro_UE <- car::recode(eme$e3, "
    1 = 1;
    2 = 1;
    3 = 1;
    4 = 1;
    5 = 0;
    6 = 0;
    c(88,96,99) = 999")

#Recodificar
eme<-eme %>% set_value_labels(registro_UE =
    c("Con inscripcion "=1,
      "Sin inscripcion "=0,
      "No Sabe/No responde/Sin dato "=999))

#"Etiquetar
eme$registro_UE <- sjlabelled::set_label(eme$registro_UE,
    label = "Registro de la unidad economica ante SII")

#Creando la variable cuentas_UE
frq(eme$e1) # Ver frecuencias registro de cuentas
eme$cuentas_UE <- car::recode(eme$e1,
    "c(1,2) = 1;
    3 = 2;
    4 = 3;
    c(88,96,99) = 999")

#Recodificar
eme<-eme %>% set_value_labels(cuentas_UE = c(
    "Si, a través de contabilidad formal "=1,
    "Si, manteniendo registros personales "=2,

```

```

      "No, ningún tipo de contabilidad"=3,"No Sabe/No responde"=999))
#Etiquetar
eme$cuentas_UE <- sjlabelled::set_label(eme$cuentas_UE,
      label = "Contabilidad de la unidad economica")
#Creando la variable separa_gastos -----
frq(eme$e2) # Ver frecuencias
eme$separa_gasto <- car::recode(eme$e2, "
      1 = 1;
      2 = 2;
      c(88,96,99) = 999")
#Recodificar
eme<-eme %>% set_value_labels(separa_gasto = c("Si"=1,
      "No"=2,
      "No Sabe/No responde"=999))
#Etiquetar
eme$separa_gasto <- sjlabelled::set_label(eme$separa_gasto,
      label = "Separación de gastos del negocio con los del hogar")
#Creando la variable sector_vf -----
eme <- eme %>% mutate(sector_vf=
      case_when(registro_UE == 1 ~1,
      registro_UE == 0 & cuentas_UE == 1 &
      separa_gasto == 1 ~1,
      registro_UE == 999 & cuentas_UE == 1
      & separa_gasto == 1 ~1,
      registro_UE == 0 & cuentas_UE == 1 &
      separa_gasto == 2 ~2,
      registro_UE == 999 & cuentas_UE == 1 &
      separa_gasto == 2 ~2,
      registro_UE == 0 & cuentas_UE == 2 ~2,
      registro_UE == 999 & cuentas_UE == 2 ~2,
      registro_UE == 0 & cuentas_UE == 3 ~2,
      registro_UE == 999 & cuentas_UE == 3 ~2,
      registro_UE == 0 & cuentas_UE == 999 ~3,
      registro_UE == 999 & cuentas_UE == 999 ~3,
      registro_UE == 0 & cuentas_UE #no está == 1 &
      separa_gasto == 999 ~3,
      registro_UE == 999 & cuentas_UE == 1
      & separa_gasto == 999 ~3))
#Recodificar
eme<-eme %>% set_value_labels(sector_vf = c("Sector Formal"=1,
      "Sector Informal"=2,
      "Usar proxy"=3))
#Etiquetar
eme$sector_vf <- sjlabelled::set_label(eme$sector_vf,
      label = "Unidad según sector")
#Crear variable informalidad
eme$informalidad <- car::recode(eme$sector_vf,

```



```

"
1 = 0;
2 = 1
")

#Factorizar
eme$informalidad <- factor(eme$informalidad, levels = c(1,0),
                          labels = c("Informalidad", "Formalidad"))

#Etiquetar
eme$informalidad <- sjlabelled::set_label(eme$informalidad,
                                          label = "Informalidad del Microemprededor")

#Tabulado de informalidad por sexo
tab_eme_sexo_informalidad2 <- emevii_srvy %>%
  group_by(SEX0) %>%
  dplyr::summarize(
    micro = round(survey_total(informalidad == "Informalidad",
                              vartype = NULL)),
    micro_porc = round(survey_mean(informalidad == "Informalidad",
                                   proportion = T, vartype = "se")*100, 3)
  ) %>% print()

```

## 6. PRECISIÓN DE LOS PRINCIPALES INDICADORES

El siguiente apartado tiene como objetivo orientar al usuario al momento de evaluar la precisión de las estimaciones realizadas, pudiendo determinar si éstas tienen la calidad necesaria para realizar inferencias sobre las características reales de la población. Para ello, la estructura del capítulo contiene una descripción de los principales indicadores que permiten estimar la variabilidad y confiabilidad de las estimaciones, y una presentación del estándar Institucional para evaluar la calidad de las estimaciones. Este apartado fue realizado a partir del documento “Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares”. Para más detalle sobre la implementación de este estándar, en la evaluación de los tabulados publicados por la Institución, se recomienda revisar el documento disponible en la web institucional.<sup>4</sup>

### 6.1. Conceptos utilizados para la evaluación de las medidas de calidad

A continuación, se detallan los conceptos utilizados para la evaluación de las medidas de calidad de las estimaciones realizadas. Antes de describir los conceptos o medidas de variabilidad, es importante describir los tipos de estimadores que están sujetos a esta variabilidad, los cuales se pueden descomponer en los siguientes grupos:

- **Total:** corresponde al conjunto de todos los elementos sobre los cuales se observa una o más características. Por ejemplo, total de personas microempresarias. El cual se obtiene como el total de los individuos de la muestra que no cambiaron de situación ocupacional ( $a_1=1$  o  $2$ ) ponderado por los factores de expansión. Donde  $M$  corresponde a los microempresarios que cambiaron de categoría ocupacional y  $N$  al total de microempresarios.

$$t = \hat{T}_y = \sum_{i=1}^M F_i \quad i \in M \quad y \quad M \in N$$

- **Media:** corresponde a la suma de los atributos de una característica o variable continua ponderada por el factor de expansión de la unidad estudiada (individuos, empresas, etc.), la cual luego es dividida por el número total de elementos estudiados. Por ejemplo, ganancia promedio de los microempresarios. ( $y_i$ )

$$\bar{y} = \hat{y} = \frac{\sum_{i=1}^M y_i * F_i}{\sum_{i=1}^M F_i} \quad i \in M \quad y \quad M \in N$$

- **Proporción:** Fracción de unidades que estén clasificados dentro de alguna clase definida, por ejemplo, proporción de microempresarios por cuenta propia.

<sup>4</sup><https://www.ine.gob.cl/docs/default-source/institucionalidad/buenas-pr%C3%A1cticas/clasificaciones-y-est%C3%A1ndar-evaluaci%C3%B3n-de-calidad-de-estimaciones-publicaci%C3%B3n-27022020.pdf>

$$p = \hat{p} = \frac{\sum_{i=1}^C F_i}{\sum_{i=1}^M F_i} \quad i \in C \quad y \quad c \in M$$

- Razón: definido como el cociente de dos totales o medias, por ejemplo, promedio de tiempo destinado al trabajo no remunerado de mujeres sobre el promedio del tiempo destinado al trabajo no remunerado de los hombres. Donde  $\hat{Y}$  corresponde al promedio de tiempo destinado al TNR por mujeres y  $\hat{X}$  al de hombres.

$$r = \hat{r} = \frac{\hat{Y}}{\hat{X}} = \frac{\hat{Y}}{\hat{X}} \quad i \in M \quad y \quad M \in N$$

- Percentiles: indica la posición de un valor respecto a los demás individuos del grupo de interés, por ejemplo, la media de la ganancia de las personas microempendedoras

Para evaluar la variabilidad de estas estimaciones, se deben considerar algunos conceptos como el tamaño de muestra de la estimación el error estándar, entre otros. A continuación, se detallan los conceptos que luego sostendrán la evaluación de los criterios de calidad de la Institución:

### 1) Tamaño muestra

En el análisis de la precisión de la estimación se entiende como “tamaño muestral” al total de unidades de análisis consideradas para obtener la estimación. Este cálculo depende de la unidad de análisis sobre la cual se está obteniendo la estimación, las cuales pueden ser: viviendas, hogares y/o personas<sup>5</sup>.

### 2) Grados de libertad

Los grados de libertad ( $g_l$ ) se definen como el número de observaciones independientes que son libres de variar (número de individuos en los datos) menos el número de parámetros estimados (número de relaciones impuestas a los datos), o también como el número de valores que podemos elegir libremente (Levin, 1996).

### 3) Error estándar

El error estándar corresponde a la raíz cuadrada de la varianza de muestreo. Esta medida de dispersión resulta ser más fácil de interpretar, ya que utiliza la misma escala de medición de la estimación. Las expresiones para obtener el error estándar se presentan en la siguiente tabla:

<sup>5</sup>El recuento del tamaño muestral se realiza dependiendo del tipo de estimador. Para proporciones y razones entre 0 y 1, se realiza el conteo sobre las unidades de análisis del denominador; mientras que para el resto de las estimaciones se hace sobre el total de unidades de análisis que nutren la estimación.

Tabla 7: Error estándar poblacional y muestral

Error estándar Poblacional	Estimación del error estándar
$EE(\hat{\theta}) = \sqrt{V(\hat{\theta})}$	$ee(\hat{\theta}_s) = \sqrt{v(\hat{\theta}_s)}$

Fuente: Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares (INE, 2020)

#### 4) Coeficiente de variación.

El coeficiente de variación, como puede verse en la Tabla 8, se obtiene mediante el cociente entre el error estándar y la propia estimación. Suele expresarse en porcentaje y resulta ser útil para comparar la precisión de estimaciones entre encuestas con tamaño muestral diferente o en los casos en donde la escala de medición de las estimaciones a comparar es diferente, ya que no tiene unidad de medida.

Tabla 8: Coeficiente de variación poblacional y muestral

Coeficiente de variación Poblacional	Estimación del coeficiente de variación
$CV(\hat{\theta})(\%) = \frac{EE(\hat{\theta})}{\hat{\theta}} * 100\%$	$cv(\hat{\theta}_s)(\%) = \frac{ee(\hat{\theta}_s)}{\hat{\theta}_s} * 100\%$

Fuente: Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares (INE, 2020)

## 6.2. Evaluación de la calidad de las estimaciones

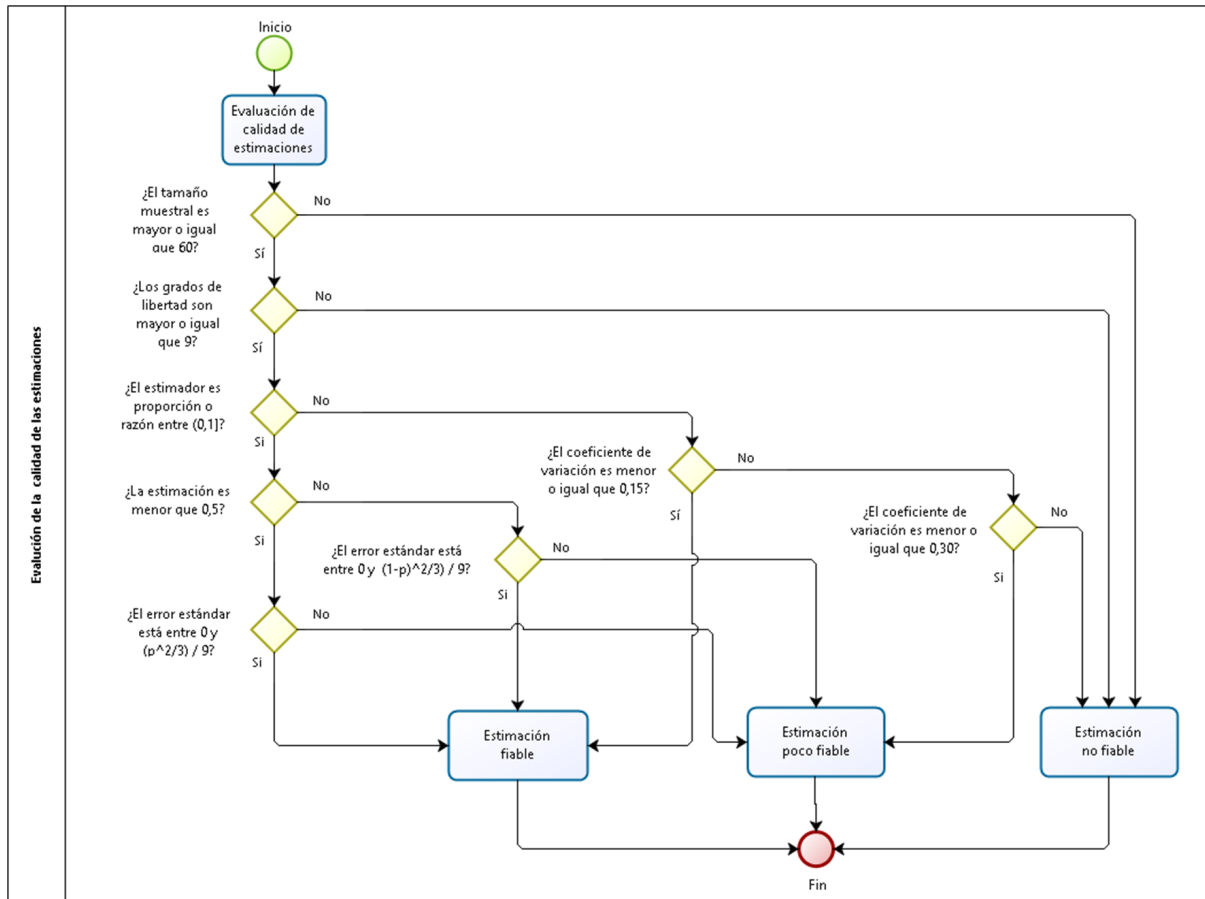
De acuerdo con el estándar definido por la institución, las estimaciones pueden ser clasificadas en tres grupos: estimación fiable, estimación poco fiable y estimación no fiable. La Figura 1 muestra el flujo con el detalle de los filtros para ser clasificado en cada uno de los grupos. Este comienza con dos condiciones excluyentes para cualquier tipo de estimación, las cuales establecen que, para poder publicar una estimación, ésta debe tener un tamaño muestral de al menos 60 casos y un número de grados de libertad mayor o igual a 9. En caso contrario, el flujo termina y las estimaciones deben ser clasificadas como “estimación no fiable”. Si cumplen las dos condiciones anteriores, el flujo se divide dependiendo de si el tipo de estimador es de proporción (o razón entre 0 y 1) o no. Si el estimador es una proporción o razón entre 0 y 1, el error estándar debe cumplir la siguiente condición, dependiendo del nivel de la estimación asociada de  $p$ :

$$\text{Máximo } ee \text{ tolerable} = \begin{cases} \frac{\sqrt[3]{p^2}}{9} & ; p < 0,5 \\ \frac{\sqrt[3]{(1-p^2)}}{9} & ; p \geq 0,5 \end{cases}$$

En caso de que el error estándar ( $ee$ ) se encuentre en dicho intervalo, el estimador debe considerarse como estimación fiable. Pero si el estimador no cumple con dicho criterio, debe consignarse como estimación poco fiable dada su baja confiabilidad. El resto de las estimaciones se evalúan mediante el coeficiente de variación, el cual debe ser 15% o menos para ser clasificado como estimación fiable. Cuando el coeficiente de variación sea mayor a 15% y menor o igual a 30%, la estimación debe ser

catalogada como poco fiable; finalmente, si el coeficiente de variación supera 30 % la estimación no debe ser publicada dada su baja precisión, siendo identificada como estimación no fiable.

Figura 1: Flujo para la evaluación de la calidad de las estimaciones



## 7. ACCESO, SOLICITUD Y USO DE BASES DE DATOS

Se pone a disposición del público las bases de datos innominadas e indeterminadas, las cuales pasaron por los procesos de edición, imputación y anonimización evitando la posible identificación de los informantes. Al igual que la versión anterior de la Encuesta se pone a disposición del público general dos Bases de datos: la Base Full, que cuenta con toda la información del cuestionario salvo de la pregunta F2 y la Base de Empleo que almacena la información correspondiente a la pregunta F2 sobre la caracterización de los trabajadores del negocio. La información de las bases de datos y la documentación de la encuesta queda disponible al público general. Se recomienda que el uso de estimaciones realizadas con los datos sea citado a través del uso de referencia o pie de la fuente ya sea un gráfico o un tabulado, de acuerdo con formatos estandarizados y adecuados al carácter de cada publicación. La información que haya sido procesada por parte de investigadores de instituciones privadas o públicas debe acreditar, en primer lugar, a la persona o institución responsable de la elaboración de dicho análisis y luego mencionar a la Encuesta de Microemprendimiento y su versión. Por ejemplo: “Fuente: Elaboración propia a partir de la base de datos de la VII Encuesta de Microemprendimiento, 2022”. En el caso de citarse resultados o información procesada de la Encuesta de Microemprendimiento, que haya sido publicada por el Instituto Nacional de Estadísticas o el Ministerio de Economía en tablas o documentos de resultados de carácter oficial, se debe reconocer a esta institución como fuente de origen de los datos e identificar, además, el año correspondiente a la versión de la Encuesta Microemprendimiento en la que está basada la información publicada (por ejemplo: “Fuente: Ministerio de Economía, VII Encuesta de Microemprendimiento”). Finalmente, al momento de citar documentos de resultados, metodológicos o información de la Encuesta de Microemprendimiento, que hayan sido elaborados por el Ministerio de Economía o el INE o por otras instituciones, se sugiere emplear formato de cita de obra completa para autor corporativo, en conformidad con norma APA:

Nombre de autor corporativo (año publicación).

Título de la obra en cursiva. Lugar

Publicación: nombre editorial.

Ejemplo: Instituto Nacional de Estadística (2022). Síntesis de Resultados VII Encuesta de Microemprendimiento. Santiago, Chile.

## **8. DICCIONARIO DE VARIABLES Y CÓDIGOS**

El diccionario de variables y códigos de las bases Full y Base de Empleo, se encuentra en el micrositio de la encuesta Link <https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral/microemprendimiento>.

## 9. REFERENCIAS BIBLIOGRÁFICAS

- Levin r, rubin d. Estadística para Administradores. México: prentice-hall. 1996.
- Instituto Nacional de Estadística (2020). Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares. Santiago, Chile. Disponible en: <https://www.ine.gob.cl/docs/default-source/institucionalidad/buenas-pr%C3%A1cticas/clasificaciones-y-estandares/est%C3%A1ndar-evaluaci%C3%B3n-de-calidad-de-estimaciones-publicaci%C3%B3n-27022020.pdf>