



QUINTA ENCUESTA LONGITUDINAL DE EMPRESAS

Instructivo de uso de la base de datos

**INSTITUTO NACIONAL DE ESTADÍSTICAS
Enero/2019**

Departamento de Estudios Económicos Estructurales y Departamento de Investigación y Desarrollo

Subdirección Técnica

Unidad de Proyectos Económicos Especiales

Subdirección de Operaciones

Quinta Encuesta Longitudinal de Empresas - Instructivo de Uso de Base de Datos

Instituto Nacional de Estadísticas

Enero / 2019

ÍNDICE

1	INTRODUCCIÓN	1
2	ACCESO A LOS DATOS	1
3	ESTRUCTURA DE LOS DATOS.....	2
3.1	Descripción de la base de datos.....	2
3.2	Tipos de variables.....	3
3.3	Variables de identificación y estratificación.....	4
4	INDICACIONES PARA EL USO DE LA BASE DE DATOS	7
4.1	Cálculo de estimaciones expandidas.....	7
4.2	Fortalezas de análisis longitudinales.....	8
4.3	Cruce con versiones anteriores de la encuesta.....	8
4.4	Imputaciones realizadas.....	9
5	CONSTRUCCIÓN DE INDICADORES.....	10
5.1	Principales indicadores de la encuesta	10
5.2	Estimación del total de ventas anuales.....	10
5.3	Estimación de la productividad salarial.....	10
5.4	Estimación de la productividad laboral.....	11
5.5	Estimación de la razón de ventas anuales 2017/2015.....	12
6	RECOMENDACIONES DE USOS DE ESTIMACIONES	13
7	ANEXO	14
7.1	Anexo 1: Reseña del proceso de recopilación.....	14
7.2	Anexo 2: Ficha técnica ELE 5.	17
7.3	Anexo 3: Ejemplos prácticos	18
7.4	Anexo 4: Sintaxis en Stata para la aplicación de ejemplos.	22

1 INTRODUCCIÓN

El presente documento tiene por finalidad proporcionar al usuario la información necesaria para utilizar la base de datos de la Quinta Encuesta Longitudinal de Empresas (ELE 5).

La difusión de este documento se enmarca en la política de transparencia y buenas prácticas del Instituto Nacional de Estadística (INE), que busca potenciar el uso de la información estadística, fortalecer su claridad y transparencia hacia el público.

La Encuesta Longitudinal de Empresas (ELE), que es de corte transversal y longitudinal, tiene por objetivo principal la caracterización de las empresas del país, por actividad económica y tamaño según nivel de ventas. De esta forma, la ELE se propone como un instrumento que permite mejorar el análisis de la estructura productiva de Chile, aportando al estudio y diseño de políticas públicas.

Los contenidos del presente documento están estructurados de la siguiente forma: la sección 2 se instruye a los usuarios sobre cómo acceder a las bases de datos de la encuesta; la sección 3 describe la base de datos; la sección 4 presenta las indicaciones de uso necesarias, en atención al diseño muestral de la encuesta, factores de expansión, recomendaciones para el cruce con versiones anteriores de la encuesta, entre otros; la sección 5 muestra la forma de cálculo de los principales indicadores de la encuesta; la sección 6 presenta recomendaciones relativas al uso de las estimaciones. Finalmente, la sección 7 corresponde al anexo del documento.

2 ACCESO A LOS DATOS

Se pone a disposición de los usuarios una base de datos, innominada e indeterminada, de libre acceso, que contiene todos los datos informados por las unidades encuestadas, con la excepción de aquellos que pudieran poner en riesgo el correcto resguardo del secreto estadístico.

Asimismo, se pone a disposición de los usuarios los siguientes documentos de apoyo:

- Informe de diseño muestral
- Antecedentes metodológicos y operativos
- Criterios de comparabilidad de las preguntas a través del tiempo
- Diccionario de la base de datos
- Instructivo de uso de la base de datos
- Formularios

Los documentos se encuentran disponibles en la página web del Ministerio de Economía, Fomento y Turismo (<https://www.economia.gob.cl/>).

Como se mencionó anteriormente, la ELE 5 está diseñada para ser representativa tanto en términos transversales como longitudinales. En este contexto, se debe hacer hincapié en que los niveles de estimación definidos para la ELE 5 varían dependiendo si el análisis se refiere al corte longitudinal o transversal. En el caso del corte transversal, los niveles de estimación son: nacional, sección de actividad económica, tamaño de empresa y el cruce entre actividad económica y tamaño. Por su parte, los niveles de estimación del corte longitudinal son: nacional, sección de actividad económica y tamaño de empresa, excluyendo el cruce entre sección de actividad económica y tamaño. Sin perjuicio de lo anterior, en aquellos niveles de estimación donde el error muestral supera el umbral admisible, no se recomienda realizar inferencia estadística.

Lo anterior quiere decir que las estimaciones serán consideradas representativas a estos niveles de detalle. De la misma forma, no es posible asegurar la calidad, en términos estadísticos, de estimaciones a niveles más desagregados que los recién expuestos.

3 ESTRUCTURA DE LOS DATOS

3.1 Descripción de la base de datos

El diseño muestral de la ELE 5 es probabilístico y estratificado, donde los estratos se encuentran definidos por sección de actividad económica, tamaño de empresa y tramo. La sección de actividad económica se define según la adaptación nacional del Clasificador Internacional Industrial Uniforme de todas las actividades económicas (CIIU4.CL 2012) y el tamaño de la empresa según el total de ventas anuales en UF. El tramo permite diferenciar entre empresas de inclusión forzosa y de inclusión aleatoria, donde las primeras corresponden a aquellas empresas que representan un aporte importante de ventas dentro del estrato. Además, considera la conformación de una muestra panel con la finalidad de realizar análisis de tipo longitudinal. Para mayor información, se recomienda revisar el anexo 1 y documento de metodología de diseño muestral.

La base de datos de la ELE 5 contiene 6.480 encuestas terminadas¹ con información de caracterización de la encuesta, más las variables recopiladas en el trabajo de campo. Los datos proporcionados en la base de datos deben cumplir con los protocolos de anonimización del INE, por

¹ Se entenderá por encuesta terminada aquella encuesta completa que ha pasado exitosamente por todos los procesos de revisión y validación.

lo que se elimina toda la información que haga referencia explícita al informante, además de todos los códigos que incorporan texto.

Adicionalmente, todos los campos donde no existe respuesta del informante, a causa de la aplicación de saltos presentes en el formulario, son dejados en vacío y no completados con cero.

Para facilitar la accesibilidad y comprensión de las variables pertenecientes a la base de datos, se pone a disposición de los usuarios un diccionario de base de datos². Este documento contiene la glosa, código y tipo de cada variable recopilada y si está presente en la base de datos publicada o no.

3.2 Tipos de variables

La descripción detallada de las variables recogidas por la encuesta se encuentra en el diccionario de la base de datos. Cada una de ellas, se puede definir dentro de los tipos y rangos de la Tabla 1 a continuación.

Tabla 1. Descripción de tipos de variable y rango en la base de datos de ELE 5.

Tipo	Rango
AÑO	NUMÉRICO AÑO
CATEGORIZACIÓN	1: Más importante / Muy importante 2: Importante 3: Menos importante / Poco o nada importante 1: Siempre 2: Muchas veces 3: Pocas veces 4: Nunca 5: No sabe
NUMÉRICO	[NÚMERO ENTERO POSITIVO] [NÚMERO ENTERO]
PORCENTAJE	0 – 100
PORCENTAJE DECIMAL	0,00 - 100,00
SELECCIÓN MÚLTIPLE	[VACÍO] – True
SELECCIÓN ÚNICA	[VACÍO] – True
TEXTO	[A-Za-zá-úÁ-Ú(),./-]

Fuente: INE.

² Disponible en la página web del Ministerio de Economía, Fomento y Turismo: <https://www.economia.gob.cl/>

3.3 Variables de identificación y estratificación

3.3.1 Variables de identificación

Cada unidad dentro de la base de datos posee un identificador ficticio llamado “rol ficticio” que permite la enumeración de las empresas, manteniendo su anonimato. Para el caso de unidades consideradas como panel fijo³⁴ el identificador asignado se ha mantenido desde la Cuarta Encuesta Longitudinal de Empresas (ELE 4), por lo que es posible realizar el cruce con esta base de datos. Por su parte, las unidades no pertenecientes al panel fijo han recibido identificadores nuevos, es decir, no utilizados en versiones anteriores de la encuesta.

3.3.2 Variables de estratificación

Cada unidad en la base de datos puede ser clasificada dentro de los estratos muestrales de la encuesta, que corresponden a actividad económica, tamaño y tramo. Es importante mencionar que, si bien es posible el análisis a los niveles de estratificación descritos, estos no son equivalentes a los niveles de estimación⁵ de la muestra. Los niveles de estimación dependerán del corte utilizado (transversal o longitudinal). A saber:

- a. Corte transversal:
 - Nacional
 - Sección de actividad económica
 - Tamaño de empresa
 - Sección de actividad económica y tamaño de empresa

- b. Corte longitudinal:
 - Nacional
 - Sección de actividad económica
 - Tamaño de empresa

³ Para la ELE 5, se considera como panel fijo aquellas empresas que fueron encuestadas en la ELE 4.

⁴ Para mayor información, favor referirse al documento “Quinta Encuesta Longitudinal de Empresas – Informe de diseño muestral”.

⁵ Los niveles de estimación son aquellas desagregaciones para las cuales se considera que la muestra utilizada es representativa de la población, en base a los errores calculados. Para mayor información, favor referirse al documento “Quinta Encuesta Longitudinal de Empresas – Informe de diseño muestral”.

Sin perjuicio de lo anterior, se presenta el detalle de las variables de estratificación presentes en la base de datos. La Tabla 2. Sección de actividades CIIU4.CL 2012 en estudio en ELE 5. describe la desagregación de la variable actividad económica.

Tabla 2. Sección de actividades CIIU4.CL 2012 en estudio en ELE 5.

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

Fuente: INE.

Mientras que los valores que toma la variable tamaño de empresas se especifican en la Tabla 3.

Tamaño de las empresas según ventas anuales en UF

Tabla 3. Tamaño de las empresas según ventas anuales en UF.

Tamaño de empresa según ventas	Tramo de ventas anuales (UF)	
	Límite inferior	Límite superior
Grande (G)	100.001	Más
Mediana (M)	25.001	100.000
Pequeña 2 (P2)	5.001	25.000
Pequeña 1 (P1)	2.401	5.000
Micro (M)	500	2.400

Fuente: INE.

Por último, el tramo identifica si la empresa corresponde a inclusión forzosa o fue seleccionada aleatoriamente.

3.3.3 Factores de expansión

El factor de expansión se interpreta como la cantidad de unidades en la población a las que representa una unidad de la muestra. En este contexto, la base de datos cuenta con dos factores de expansión: factor de expansión transversal y factor de expansión longitudinal.

El primero corresponde al componente transversal de la muestra y debe ser utilizado para el análisis de los resultados en este contexto. Denotados por la variable “fe_transversal” en la base de datos.

El segundo factor de expansión corresponde al componente longitudinal de la muestra y se asigna solo a aquellas unidades seleccionadas como parte del panel fijo. Por lo mismo, este factor debe ser utilizado para el análisis de resultados longitudinales, como aquellos que involucren las unidades del panel pertenecientes a ELE 4 y ELE 5. Este factor está denotado por la variable “fe_longitudinal” en la base de datos.

Para más detalle respecto a cuándo utilizar un factor de expansión u otro consultar el subtítulo 4.3.

3.3.4 Descripción de las variables principales contenidas en la base de datos

La base de datos incluye el identificador ficticio asignado a cada unidad, así como también las respectivas variables de estratificación y factores de expansión. Estas variables se encuentran descritas en la Tabla 4.

Tabla 4. Variables principales y su clasificación en la base de datos de la ELE 5.

Campo	Tipo de dato	Descripción
rol_ficticio	Número	Rol de la empresa
dv_rol_ficticio	Número	Dígito verificador de la empresa
Tramo	Texto largo	1: Inclusión forzosa 2: Tramo aleatorio
Tamaño	Texto largo	1: Grande 2: Mediana 3: Pequeña 2 4: Pequeña 1 5: Micro
CIIU_final	Texto largo	Clasificador Internacional Industrial Uniforme (CIIU), revisión 4, en su adaptación nacional (CIIU4.CL 2012), a nivel de sección (letra)
Glosa_CIIU	Texto largo	Glosa de sección CIIU
FE_transversal	Número	Factor de expansión transversal
FE_longitudinal	Número	Factor de expansión longitudinal
Panel	Número	0: Empresa no es panel 1: Empresa es panel

Fuente: INE.

4 INDICACIONES PARA EL USO DE LA BASE DE DATOS

En el presente capítulo se proveen elementos para el uso adecuado de la base de datos. En particular, se entregan recomendaciones sobre el uso de los factores de expansión y se detallan otras consideraciones que se deben tener en cuenta al momento de realizar análisis de los datos.

4.1 Cálculo de estimaciones expandidas

El cálculo de cifras expandidas se realiza a los niveles de estratificación en estudio. En términos genéricos, el cálculo de cifras expandidas para el estrato de tramo t , tamaño s y actividad a se realiza de la siguiente manera:

$$\hat{v}_{t,s,a} = \sum_{i=1}^n w_{t,s,a,i} \cdot v_{t,s,a,i} \quad (1)$$

Donde:

- t : Indicador del estrato de tramo en estudio.
- s : Indicador del estrato de tamaño en estudio.
- a : Indicador del estrato de actividad en estudio.
- i : Indicador de empresa i .
- $\hat{v}_{t,s,a}$: Estimación del total de la variable v^6 para el tramo t , tamaño s y actividad a .
- $w_{t,s,a,i}$: Factor de expansión (transversal o longitudinal, según corresponda) de la empresa i , perteneciente al tramo t , tamaño s y actividad a .
- $v_{t,s,a,i}$: Valor de la variable v de la empresa i , perteneciente al tramo t , tamaño s y actividad a .

Se debe tener en cuenta que este cálculo no es válido para el caso de porcentajes o razones. En el caso de porcentajes, es necesario primero multiplicar estos valores por las variables a las que hacen

⁶ La variable v se utiliza a modo genérico y puede corresponder a cualquier variable disponible en la base de datos.

referencia⁷. Teniendo este nuevo valor, se procede a obtener las cifras expandidas en base a lo explicado en la fórmula (1).

Por su parte, en el caso de la estimación de razones, es necesario primero obtener los valores expandidos del numerador y denominador, en base a lo explicado en la fórmula (1), para luego calcular la razón a nivel del estrato en estudio. En ningún caso se deben calcular las razones a nivel de microdato para luego expandirlas.

El subtítulo 7.3 de este documento provee algunos ejemplos prácticos sobre cálculo de cifras expandidas.

4.2 Fortalezas de análisis longitudinales

Según (Lynn, 2005) las fortalezas de los análisis longitudinales son analíticas e incluyen los siguientes aspectos:

- a. Análisis del cambio bruto
- b. Análisis del cambio medio a nivel de unidad
- c. El análisis de la estabilidad o inestabilidad en cuanto a características
- d. El análisis de las características en términos de tiempo de los sucesos o circunstancias, como son la frecuencia, el momento y la duración;
- e. Análisis de la naturaleza ordinal de los sucesos, lo cual sirve, a menudo, para clarificar problemas de causalidad

En consecuencia, se recomienda la aplicación de análisis longitudinales utilizando el factor de expansión longitudinal cuando se presenten algunos de los casos señalados.

4.3 Cruce con versiones anteriores de la encuesta

Como fue mencionado anteriormente, en esta versión de la ELE se incluye un factor de expansión longitudinal para las unidades del panel fijo, el cual expande la muestra para alcanzar el total de unidades de la población objetivo que existen durante los años de referencia tanto de ELE 4 como de la ELE 5. Por lo tanto, al realizar el cruce de ambas bases de datos, aquellas que cuentan con información en ambos momentos del tiempo tienen asignado un factor de expansión longitudinal, el cual es posible utilizar para el análisis de cambios y diferencias entre ambos períodos. Es decir, el factor de expansión longitudinal es útil para realizar análisis de cambio bruto (o cualquiera de los análisis señalados en el subtítulo 4.2), ya que permite estudiar el comportamiento de una variable para la misma unidad, en distintos momentos del tiempo.

⁷ Por ejemplo: multiplicar el porcentaje informado respecto a la variable “y” por el valor informado de la variable “y”.

Para obtener estimaciones de tipo transversal en la ELE 5, no es necesario realizar el cruce con la base de datos de la ELE 4 y se debe utilizar el respectivo factor de expansión (transversal).

Por otro lado, para realizar estimaciones de tipo transversal para la ELE 4 se deben utilizar los factores de expansión calculados a partir de las ventas anuales y número de empresas⁸, presentes en la base de datos de esa edición de la encuesta. El factor de expansión por ventas fue calculado a partir de las ventas anuales y es empleado para obtener las estimaciones de las variables cuantitativas del estudio. El factor de expansión por número de empresas se utiliza para obtener estimaciones de las variables cualitativas del estudio y fue calculado empleando la cantidad de unidades en el marco muestral y tamaño muestral. Ambos factores de expansión son de tipo transversal (en versiones anteriores de la ELE no se calculaba un factor de expansión para el panel fijo). Si bien la comparación de las estimaciones transversales de ambas encuestas permite construir series de tiempo, el análisis corresponde a la diferencia entre dos estimaciones de tipo transversal, es decir, cambio neto.

4.4 Imputaciones realizadas

En esta edición de la encuesta se realizó imputación de datos para los casos particulares de empresas pertenecientes al tramo de inclusión forzosa cuyas encuestas no fueron terminadas.

Las unidades que pertenecen al tramo de inclusión forzosa tienen un factor de expansión igual a 1, ya que por definición se considera que su comportamiento es único, por lo que no pueden representar ni ser representadas por otras unidades. Al igual que para las empresas del tramo aleatorio, existen unidades para las cuales no se obtiene respuesta, o se obtiene respuesta parcial, por parte de los informantes, sin embargo, debido a sus características, el tratamiento de las unidades de inclusión forzosa es distinto.

En este contexto, se imputaron solo aquellas unidades que contaban con respuesta parcial a la encuesta y cuyos datos faltantes podían ser obtenidos de fuentes complementarias de información. Lo anterior de modo que estas empresas cuenten con la misma información que aquellas que contestaron la encuesta de forma completa. Las fuentes complementarias utilizadas corresponden a otras encuestas del INE que cuentan con datos comparables a los faltantes, como son las encuestas estructurales, de remuneraciones y coyunturales.

⁸ En versiones anteriores de la ELE, se calculaban dos factores de expansión transversales: según ventas y según número de empresas. En la ELE 5, solo se calcula un factor de expansión transversal, el cual expande al número de empresas, y puede ser utilizado para todas las estimaciones de tipo transversal.

5 CONSTRUCCIÓN DE INDICADORES

5.1 Principales indicadores de la encuesta

Los principales indicadores de la encuesta, para el corte transversal, son tres: total de ventas anuales, productividad salarial y productividad laboral. Por otro lado, para el corte longitudinal, se utiliza la razón de ventas anuales. A continuación, se describe la fórmula de cálculo para cada uno de ellos.

5.2 Estimación del total de ventas anuales

La estimación del total de ventas anuales se obtiene a partir de la suma de los apartados “ingresos por actividad principal” y “otros ingresos operacionales”, contenidos en la ELE 5. En el cálculo de la estimación del total de ventas anuales se excluyen las partidas de “corrección monetaria” y “otros ingresos no operacionales”. Este cálculo se expresa en la siguiente fórmula:

$$\hat{Y}^{2017} = \sum_{i=1}^{n_{hv}} (Y_i^{ap,2017} + Y_i^{op,2017}) \cdot w_i^t \quad (2)$$

Donde:

- \hat{Y}^{2017} : Estimación del total de ventas anuales 2017 en la muestra transversal t ELE 5.
- $Y_i^{ap,2017}$: Ingresos por actividad principal 2017 de la empresa i en la muestra transversal t ELE 5.
- $Y_i^{op,2017}$: Otros ingresos operacionales 2017 de la empresa i en la muestra transversal t ELE 5.
- w_i^t : Factor de expansión transversal de la empresa i en la muestra transversal t ELE 5.

5.3 Estimación de la productividad salarial

Las variables asociadas a la productividad se calculan a partir de los datos reportados por las empresas encuestadas en la ELE 5. Es importante aclarar que lo que se está calculando es un proxy del valor agregado dividida por la cantidad de trabajadores o el monto total de remuneraciones. Por definición, el concepto de productividad no es observable, de modo que se recomienda evitar hacer comparaciones entre los cálculos aquí presentados y otros estudios relacionados con la productividad. El proxy del valor agregado se obtiene a partir de la suma de los ingresos

operacionales y la variación de existencias, menos costos operacionales⁹. En la siguiente fórmula se detalla cómo se calcula el valor agregado de la empresa i :

$$VA_i^{2017} = (Y_i^{ap,2017} + Y_i^{op,2017}) + (E_i^{2017} - E_i^{2016}) - CO_i^{2017} \quad (3)$$

Donde:

- VA_i^{2017} : Valor agregado de la empresa i en la muestra transversal t ELE 5.
- $Y_i^{ap,2017}$: Ingresos por actividad principal 2017 de la empresa i en la muestra transversal t ELE 5.
- $Y_i^{op,2017}$: Otros ingresos operacionales 2017 de la empresa i en la muestra transversal t ELE 5.
- E_i^{2017} : Existencias en el año 2017 de la empresa i en la muestra transversal t ELE 5.
- E_{hvi}^{2016} : Existencias en el año 2016 de la empresa i en la muestra transversal t ELE 5.
- CO_i^{2017} : Costos operacionales en el año 2017 de la empresa i en la muestra transversal t ELE 5.

La estimación de la productividad salarial se define de la siguiente forma:

$$\hat{R}^s = \frac{\sum_{i=1}^n VA_i^{2017} \cdot w_i^t}{\sum_{i=1}^n SAL_i^{2017} \cdot w_i^t} \quad (4)$$

Donde:

- \hat{R}^s : Estimación de la productividad salarial en la muestra transversal t ELE 5.
- VA_i^{2017} : Valor agregado de la empresa i en la muestra transversal t ELE 5.
- SAL_i^{2017} : Total gasto en salarios de la empresa i en la muestra transversal t ELE 5.
- w_i^t : Factor de expansión transversal de la empresa i en la muestra transversal t ELE 5.

5.4 Estimación de la productividad laboral

La estimación de la productividad laboral se calcula mediante la siguiente fórmula:

⁹ Los costos operacionales corresponden a la suma los costos de venta y/o producción, gastos en electricidad, gastos en combustible, gastos en servicios de comunicación y TICs, gastos generales, gastos de administración y finanzas.

$$\hat{R}^l = \frac{\sum_{i=1}^n VA_i^{2017} \cdot w_i^t}{\sum_{i=1}^n EMP_i^{2017} \cdot w_i^t} \quad (5)$$

Donde:

- \hat{R}^l : Estimación de la productividad laboral en la muestra transversal t ELE 5.
- VA_i^{2017} : Valor agregado de la empresa i perteneciente en la muestra transversal t ELE 5.
- EMP_i^{2017} : Número de empleados de la empresa i en la muestra transversal t ELE 5.
- w_i^t : Factor de expansión transversal de la empresa i en la muestra transversal t ELE 5.

5.5 Estimación de la razón de ventas anuales 2017/2015

A partir de los datos de la muestra panel, se obtiene la estimación de la razón de ventas anuales 2017/2015 según la siguiente fórmula:

$$\hat{R}^Y = \frac{\sum_{i=1}^n (Y_i^{ap,2017} + Y_i^{op,2017}) \cdot w_i^f}{\sum_{i=1}^n (Y_i^{ap,2015} + Y_i^{op,2015}) \cdot w_i^f} \quad (6)$$

Donde:

- \hat{R}^Y : Estimación de la razón entre las variables total de ventas anuales 2017 y 2015 en la muestra panel fijo f ELE 5.
- $Y_i^{ap,2017}$: Ingresos por actividad principal 2017 de la empresa i en la muestra panel fijo f ELE 5.
- $Y_i^{op,2017}$: Otros ingresos operacionales 2017 de la empresa i en la muestra panel fijo f ELE 5.
- $Y_i^{ap,2015}$: Ingresos por actividad principal 2015 de la empresa i en la muestra panel fijo f ELE 5.
- $Y_i^{op,2015}$: Otros ingresos operacionales 2015 de la empresa i en la muestra panel fijo f ELE 5.
- w_i^f : Factor de expansión longitudinal de la empresa i en la muestra panel fijo f ELE 5.

6 RECOMENDACIONES DE USOS DE ESTIMACIONES

Dados los tamaños muestrales (considerando como unidad de análisis las empresas), el que una estimación sea o no aceptable dependerá del coeficiente de variación obtenido y de la desagregación del nivel de estimación, ya que al analizar un subgrupo de la muestra se debe considerar la menor cantidad de observaciones disponibles. Por tanto, es relevante considerar la cantidad de unidades con las que se busca realizar alguna inferencia estadística.

Para interpretar el error muestral, usualmente se selecciona el coeficiente de variación asociado a una estimación, para lo cual se recomienda considerar, a modo de referencia, los siguientes rangos:

- Hasta un 5% se considera que la estimación es óptima en términos de precisión.
- Sobre un 5% y hasta un 15% se considera una buena estimación en términos de precisión.
- Sobre un 15% la estimación se considera deficiente en términos de precisión y, por ende, no es recomendable realizar análisis o inferencias, sino solo para fines descriptivos.

Aun cuando el coeficiente de variación presente rangos adecuados en términos de precisión, las estimaciones podrían derivar en errores de interpretación por parte de los usuarios de las bases de datos debido a la disparidad de tamaños muestrales cuando se realizan estimaciones utilizando desagregaciones. Si bien no existen recomendaciones oficiales respecto al tamaño de la muestra, debido a que los totales poblacionales de las variables de interés difieren entre sí, el analista debe definir un límite mínimo de observaciones que aseguren representatividad y precisión del estimador.

Por consiguiente, para definir el uso que se le puede dar a las estimaciones, se debe considerar un análisis conjunto del número de observaciones con información de la variable de interés y del coeficiente de variación efectivo, al nivel de desagregación evaluado.

La Tabla 5 entrega criterios relativos a la calidad de los estimadores según el coeficiente de variación calculado.

Tabla 5. Clasificación de estimaciones por número de observaciones según coeficiente de variación.

Coeficiente de variación (%)	Número de observaciones	
	Bajo	Alto
(15 , 100]	Estimador poco confiable	Descriptivo (nota técnica)
(5 , 15]	Descriptivo (nota técnica)	Estimador confiable
[0 , 5]	Estimador confiable	Estimador confiable

Fuente: INE.

A continuación se entregan definiciones de las clasificaciones presentadas en la Tabla 5:

- Estimador confiable: el estimador se puede utilizar según los objetivos propuestos.
- Uso descriptivo: se recomienda que el estimador se utilice para estadística descriptiva. Si se desea utilizar para otro tipo de análisis, se recomienda agruparlo a un nivel superior de la clasificación. En todos estos casos se recomienda incluir una nota técnica advirtiendo sobre las restricciones de la estimación que mencione el valor del coeficiente de variación.
- Estimador poco confiable: no se recomienda utilizar la estimación obtenida. Se debe reevaluar la estimación a un nivel superior de la clasificación, para allí evaluar su calidad y determinar su aptitud para un uso inferencial o solamente descriptivo.

En cuanto al número de observaciones y para el contexto de la ELE 5, se sugiere categorizar el umbral 30 empresas o menos como “Bajo” y el umbral más de 30 empresas como “Alto”. Cabe mencionar que es un umbral sugerido que debería someterse a consideración dependiendo de los análisis llevados a cabo.

7 ANEXO

7.1 Anexo 1: Reseña del proceso de recopilación

El proceso de recopilación se inicia con la verificación de las unidades seleccionadas en la muestra. En esta fase se confirma la información de contacto y estado de la unidad (en funcionamiento, cerrada, paralizada, etc.). La recopilación de la encuesta se realiza mediante formulario físico y web.

Posterior a la verificación, se envía el formulario a los informantes. A cada informante se le proporciona un usuario y clave de acceso al sistema web, medio por el cual se centraliza la digitación y procesamiento. Sumado a esto, se captura información a través de correo electrónico o en forma física en los casos que así lo requieran.

Dadas las características del sistema informático, la recolección de datos se realiza de forma simultánea a la clasificación, codificación y análisis de consistencia de datos. El sistema informático utilizado posee un set de validadores que tiene por objetivo detectar posibles inconsistencias o errores en los datos, con el fin de minimizarlos.

Los validadores utilizados se separan en tres niveles. Los validadores de primer y segundo nivel se encuentran incorporados en el sistema informático. Los de primer nivel, consisten en un conjunto de reglas de completitud y consistencia numérica y lógica aplicados en la fase de respuesta del formulario electrónico.

Los validadores de segundo nivel toman como base los de primer nivel y se complementan con validadores de rangos de tolerancias para así identificar eventuales datos atípicos. Los validadores de primer y segundo nivel se activan en caso de detectar advertencias o errores, conforme avance el informante en el formulario electrónico.

Una vez que la encuesta es completada y enviada en el sistema, o enviada por correo electrónico o en forma física, comienza la supervisión directa por parte del analista que recibe la encuesta. El analista revisa la consistencia de la información en base a lo arrojado por los validadores incorporados al sistema informático y emite el correspondiente certificado de recepción.

Por su parte, los validadores de tercer nivel no se encuentran incorporados en el sistema informático de la encuesta y consisten en análisis de contraste con otras fuentes, revisión de comportamiento intertemporal y análisis agregados de los datos.

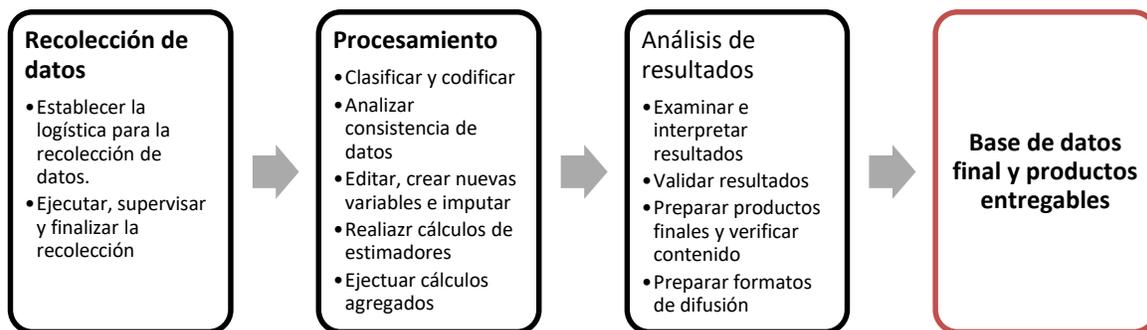
Posterior al procesamiento de los datos por parte del analista, se inicia el análisis de los resultados agregados. En este contexto, el coordinador del producto revisa las bases de datos, con el fin de obtener una visión general de las cifras e identificar eventuales inconsistencias.

En los casos donde se identifican errores, se procede a contactar al informante por teléfono o correo electrónico con el fin de corroborar o corregir la información, según sea el caso.

Una vez validados los datos, el analista se encuentra en condiciones de dejar la encuesta como terminada. Asimismo, se procede a la generación de los productos finales de la encuesta, como, por ejemplo, bases de datos innominadas, informes, tablas de resultados, entre otros.

El Diagrama 1 presenta un resumen del proceso de recopilación de datos de la ELE 5.

Diagrama 1. Resumen de proceso de recopilación de datos de la ELE 5.



7.2 Anexo 2: Ficha técnica ELE 5.

Nombre de la encuesta		Quinta Encuesta Longitudinal e Empresas
Objetivo		Caracterizar a través de la implementación de una encuesta a empresas, la heterogénea realidad empresarial del país según sector de actividad económica y tamaño de las unidades.
Cobertura	Geográfica	Nacional.
	Temática	Actividades o agrupación de actividades, seleccionadas para la realización del estudio, definidas en base a la Clasificación Internacional Industrial Uniforme de Naciones Unidas (CIU Rev. 4), en su adaptación nacional (CIU4.CL 2012).
Unidad de muestreo		Empresa formal que desarrolla su actividad productiva dentro del territorio nacional, y que es clasificada en alguno de los sectores económicos en estudio.
Población objetivo		Empresas formales que cuentan con iniciación de actividades en el Servicio de Impuestos Internos (SII), que desarrollan alguna actividad económica en el ámbito productivo, comercial o de servicios, según la Clasificación Internacional Industrial Uniforme de todas las Actividades Económicas (CIU4.CL 2012), dentro de los límites geográficos del país, y que durante el año contable 2016 alcanzaron ventas anuales desde 800,01 UF según el Formulario N°29 (Declaración Mensual de IVA del SII). Este último requisito se flexibiliza para las microempresas pertenecientes al panel fijo, disminuyendo el límite de inclusión a 500 UF.
Marco Muestral		Se conforma a partir del DNE año contable 2016, que contiene información de las empresas formales que desarrollan sus actividades económicas de los sectores en estudio, así como la venta anual en UF declarada en el Formulario N°29 (declaración mensual de IVA) del SII. Dadas las características de la encuesta, es necesario trabajar con dos marcos muestrales, a saber: <ul style="list-style-type: none"> - Marco transversal (unidades presentes en DNE 2016). - Marco longitudinal (unidades presentes en el marco muestral de ELE 4 y en el DNE 2016).
Estratificación		Según sección de actividad económica, tamaño según ventas y tramo.
Tamaño Muestral		6.480 empresas.
Niveles de estimación		Los niveles de estimación dependen del corte utilizado (transversal o longitudinal). A saber: <ol style="list-style-type: none"> 1. Corte transversal: <ul style="list-style-type: none"> - Nacional - Sección de actividad económica - Tamaño de empresa - Sección de actividad económica y tamaño de empresa 2. Corte longitudinal: <ul style="list-style-type: none"> - Nacional - Sección de actividad económica - Tamaño de empresa

Fuente: INE.

7.3 Anexo 3: Ejemplos

Debido a que el diseño muestral de la ELE 5 es probabilístico y estratificado, se deben tener ciertas consideraciones al momento de calcular estimaciones y errores de estimación. Para esto, los softwares estadísticos cuentan con módulos de muestras complejas, en los cuales es posible especificar el diseño muestral.

Para utilizar el módulo de muestras complejas es necesario definir un plan de muestreo, el cual resume las principales características de la muestra. En el caso de la ELE 5, estas son:

- Variables de estratificación: actividad económica (CIU_final), tamaño de empresa (Tamaño), tramo (Tramo).
- Ponderación muestral (factor de expansión): fe_transversal (para análisis transversales) o fe_longitudinal (para análisis longitudinales).
- Tamaño del marco muestral: esto se especifica para realizar corrección por muestras finitas el cálculo de la varianza. Para calcular el tamaño del marco muestral, es necesario sumar los factores de expansión (fe_transversal o fe_longitudinal, según si se realiza análisis transversal o longitudinal, respectivamente) para cada estrato (actividad económica, tamaño y tramo). Esto se puede obtener mediante estadísticas descriptivas.

Se debe recalcar que, si bien la estimación de estadísticos descriptivos obtenidos utilizando el factor de expansión tiene el mismo valor que la obtenida mediante el módulo de muestras complejas, el error de muestreo es distinto en cada caso. El uso de uno u otro error calculado puede llevar a diferentes conclusiones al momento de realizar inferencia estadística. Esto se debe a que el primero calcula la varianza bajo el supuesto de muestreo aleatorio simple, sin considerar que el diseño muestral de la ELE 5 es estratificado¹⁰.

En vista de lo anterior, se presentan algunos ejemplos de cálculos a partir de los datos de la ELE 5, que tienen por fin mostrar la forma correcta de realizar estimaciones y obtener los errores de estimación respectivos.

7.3.1 Cálculo del total ventas 2017 según tamaño (análisis transversal)

Paso 1. Calcular el total de ventas anuales 2017 para la empresa i , según la siguiente fórmula:

$$Y_i^{2017} = Y_i^{ap,2017} + Y_i^{op,2017} \quad (1)$$

Donde:

¹⁰ Si se desea profundizar en las implicancias de usar ambos métodos de estimación, se recomienda consultar el siguiente documento: Cassell, D.L., (2006). Wait Wait, Don't Tell Me... You're Using the Wrong Proc!

- Y_i^{2017} : Estimación del total de ventas anuales 2017 de la empresa i en la muestra transversal t ELE 5.
- $Y_i^{ap,2017}$: Ingresos por actividad principal 2017 de la empresa i en la muestra transversal t ELE 5.
- $Y_i^{op,2017}$: Otros ingresos operacionales 2017 de la empresa i en la muestra transversal t ELE 5.

Paso 2. Definir un plan de muestreo mediante el módulo de encuestas complejas, identificando las variables de estratificación (actividad económica, tamaño de empresa y tramo), factor de expansión y tamaño del marco muestral.

Paso 3. Definir la variable (ventas anuales 2017), el tipo de estimación (total) y el nivel de estimación (tamaño de empresa). Solicitar que el software calcule: estimación, y si es considerado necesario, coeficiente de variación, grados de libertad.

Paso 4. Si se requiere calcular el error relativo, calcular el cuantil de la distribución T-Student correspondiente a gl grados de libertad y un nivel de confianza $(1 - \alpha)$ ($t_{gl,1-\frac{\alpha}{2}}$). Con este valor, es posible calcular el error relativo.

A partir del procedimiento descrito, se obtienen los siguientes resultados:

Tabla 6. Estimación, coeficiente de variación y error relativo de la variable total ventas 2017. Total y según tamaño.

Tamaño	Estimación total ventas 2017 (millones de UF)	CV (%) total ventas 2017	Error relativo (%) total ventas 2017
Total	38.334,5	3,1	6,1
Grande	33.146,2	2,2	4,3
Mediana	2.017,9	18,9	37,1
Pequeña 2	1.138,8	5,4	10,6
Pequeña 1	1.244,5	65,5	128,6
Micro	787,0	30,3	59,4

Fuente: INE.

A modo de ejemplo, en el Anexo 4 se incorpora la sintaxis del cálculo del total de ventas anuales 2017 en el software Stata.

7.3.2 Cálculo de la razón de ventas según tamaño (análisis longitudinal)

Paso 1. Cruzar la base de datos de la ELE 4 y la ELE 5, mediante la variable “rol_ficticio”, manteniendo solamente las empresas que cuentan con información en ambas bases de datos.

Paso 2. Calcular el total de ventas 2017 según la fórmula 1, y el total de ventas anuales 2015 según la fórmula 2 a continuación:

$$\hat{Y}^{2015} = Y_i^{ap,2015} + Y_i^{op,2015} \quad (2)$$

Donde:

\hat{Y}^{2015} : Estimación del total de ventas anuales 2015 en la muestra longitudinal l ELE 5.

$Y_i^{ap,2015}$: Ingresos por actividad principal 2015 de la empresa i en la muestra longitudinal l ELE 5.

$Y_i^{op,2015}$: Otros ingresos operacionales 2015 de la empresa i en la muestra longitudinal l ELE 5.

Paso 3. Definir un plan de muestreo mediante el módulo de muestras complejas, identificando las variables de estratificación (actividad, tamaño y tramo), factor de expansión y tamaño del marco muestral.

Paso 4. Definir la variable numerador (ventas anuales 2017) y denominador (ventas anuales 2015), el tipo de estimación (razón) y el nivel de estimación (tamaño de empresa). Si se considera necesario, solicitar que el software calcule: estimación, coeficiente de variación¹¹.

Paso 5. Si se requiere calcular el error relativo, calcular el cuantil de la distribución T-Student correspondiente a gl grados de libertad y un nivel de confianza $(1 - \alpha)$ ($t_{gl,1-\frac{\alpha}{2}}$). Con este valor, es posible calcular el error relativo.

A partir del procedimiento descrito, se obtienen los siguientes resultados:

Tabla 7. Estimación, coeficiente de variación y error relativo de la variable razón ventas 2017/2015. Total y según tamaño.

Tamaño	Estimación razón ventas 2017/2015	CV (%) razón ventas 2017/2015	Error relativo (%) razón ventas 2017/2015
Total	1,9	9,2	18,0
Grande	2,0	10,3	20,2
Mediana	1,2	26,0	51,1
Pequeña 2	0,8	10,6	20,9
Pequeña 1	0,9	32,4	63,8
Micro	6,4	67,3	132,4

Fuente: INE.

¹¹ En algunos softwares estadísticos, el cálculo de los grados de libertad que realiza de forma automática al momento de obtener una estimación de razón no es correcto. Por lo tanto, los grados de libertad se deben calcular de forma manual.

A modo de ejemplo, el Anexo 4 se incorpora la sintaxis de la razón de ventas anuales 2017/2015 en el software Stata. En la sintaxis se especifica la forma de cálculo de los grados de libertad.

7.4 Anexo 4: Sintaxis en Stata para la aplicación de ejemplos.

Ejemplo 1: Cálculo del total de ventas según tamaño (análisis transversal)

```
*Cálculo del total de ventas 2017 según tamaño.

//Obtención de las ventas operacionales totales en miles de UF (Tomando valor promedio de la
UF durante el año):

gen venta_2017=(c037+c038)/26571.9261

replace venta_2017=0 if venta_2017==.

//Definición del plan de muestreo y especificación de corrección por muestras finitas.

egen estrato=concat(Tramo CIU_final Tamaño)

bys estrato : egen muestra_finita=total(FE_transversal)

svyset, clear

svyset rol_ficticio [pw=FE_transversal] , strata(estrato) singleunit(certainty) fpc(muestra_finita)

//Estimación

*Con grados de libertad bien calculados

forvalues x = 1(1)5 {

svy : total venta_2017 , subpop(if Tamaño==`x')

}

}
```

Ejemplo 2: Cálculo de la razón de ventas según tamaño (análisis longitudinal)

```
*Cálculo de la razón de ventas 2017 según tamaño.

//Obtención de las ventas operacionales totales en miles de UF para 2017 (Tomando valor promedio de la UF durante el año):

gen venta_2017=(c037+c038)/26571.9261

replace venta_2017=0 if venta_2017==.

//Incorporar base de datos ELE 4

rename rol_ficticio ID

merge 1:1 ID using "BBDD ELE4"

keep if _merge==3

//Obtención de las ventas operacionales totales en miles de UF para 2015 (Tomando valor promedio de la UF durante el año):

replace C037_ELE4=0 if C037_ELE4==.

replace C038_ELE4=0 if C038_ELE4==.

gen venta_2015=(C037_ELE4+C038_ELE4)/25021.9963

//Definición del plan de muestreo

egen estrato=concat(Tramo CIIU_final Tamaño)

bys estrato : egen muestra_finita_2=total(FE_longitudinal)

svyset, clear

svyset ID [pw=FE_longitudinal] , strata(estrato) singleunit(certainty) fpc(muestra_finita_2)

//Estimación

forvalues x = 1(1)5 {

svy: ratio razon_ventas : venta_2017 / venta_2015 , subpop(if Tamaño==`x')

}

}
```