

# Diseño muestral

## ENESEM 2020

Abril, 2022

**Dirección**

Dirección de Infraestructura Estadística y Muestreo

**Unidad**

Gestión de Diseño Muestral

**Elaborado por**

Angel Gaibor

**Revisado por**

Francisco Céspedes

Andrés Albán

**Aprobado por**

Julio Muñoz

<b>1. INTRODUCCIÓN .....</b>	<b>4</b>
<b>2. ANTECEDENTES .....</b>	<b>5</b>
<b>3. PLANIFICACIÓN .....</b>	<b>6</b>
<b>3.1. OBJETIVOS Y DELIMITACIÓN DE LA OPERACIÓN ESTADÍSTICA .....</b>	<b>6</b>
3.1.1. <i>Objetivos</i> .....	6
3.1.2. <i>Delimitación</i> .....	6
<b>4. DISEÑO Y CONSTRUCCIÓN .....</b>	<b>9</b>
<b>4.1. DISEÑAR EL MARCO Y LA MUESTRA .....</b>	<b>9</b>
4.1.1. <i>Tipo y etapas de diseño muestral</i> .....	9
4.1.2. <i>Dominios de estudio</i> .....	9
4.1.3. <i>Marco muestral</i> .....	9
4.1.4. <i>Variables de diseño</i> .....	10
4.1.5. <i>Tamaño muestral</i> .....	10
4.1.6. <i>Selección de la muestra</i> .....	11
<b>4.2. CONTROL DE COBERTURA DE CAMPO Y MUESTRAL .....</b>	<b>11</b>
<b>4.3. CONSTRUCCIÓN DE FACTORES DE EXPANSIÓN .....</b>	<b>13</b>
4.3.1. <i>Factor de expansión teórico</i> .....	13
4.3.2. <i>Factor postestratificado</i> .....	13
<b>4.4. ESTIMACIONES Y ERRORES MUESTRALES.....</b>	<b>15</b>
4.4.1. <i>Estimación de características</i> .....	15
4.4.2. <i>Error de muestreo</i> .....	16
<b>4.5. DECLARACIÓN DEL PLAN DE MUESTREO.....</b>	<b>16</b>
<b>5. BIBLIOGRAFÍA.....</b>	<b>17</b>

# 1. INTRODUCCIÓN

El Instituto Nacional de Estadísticas y Censos (INEC), organismo coordinador de la Producción Estadística en el Ecuador, realiza anualmente encuestas económicas dirigidas a empresas a nivel nacional.

En este sentido, se realiza anualmente la "Encuesta Estructural Empresarial", la cual toma en cuenta las distintas secciones económicas contempladas por el Clasificador Industrial Internacional Uniforme de Actividades Económicas CIIU-4 para empresas grandes y medianas, en esta encuesta en particular se utiliza el CIIU-4 a un dígito.

El proceso que se realiza para la generación del diseño muestral consta de tres etapas:

En la primera etapa se toma en cuenta:

- Determinación del marco de muestreo
- El universo de la investigación
- Las unidades de muestreo y análisis
- Cobertura geográfica
- Cobertura temática

En la segunda etapa:

- Metodología para el diseño muestral
- Determinación de los dominios de estudio
- Determinación del tamaño de la muestra
- Selección de la muestra
- Validación de la cobertura obtenida en campo

Por último, se realiza:

- El cálculo de los factores de expansión
- Postestratificación
- Estimación de características
- Error de muestreo
- Declaración del plan de muestreo

Cabe mencionar que, debido a la dinámica de las empresas, se utiliza el método conocido como postestratificación, con la finalidad de ajustar los cambios producidos en las unidades de observación por actualización del Directorio de Empresas y Establecimientos.

## 2. ANTECEDENTES

En el año de 1955, por disposición del Departamento de Investigaciones Económicas del Banco Central, se realizó el primer Censo Económico del sector industrial de Ecuador. Y a partir de 1962, se empezó a realizar la Encuesta de Manufactura y Minería bajo el nombre de "Encuestas Industriales", la cual se ejecutó hasta 1964.

En 1965 se llevó a cabo el segundo censo de Manufactura y Minería y el primer censo de Comercio y Servicios. Más adelante, desde 1966 hasta 1979, se investigó formalmente la Encuesta de Manufactura y Minería, Encuesta de Comercio y Encuesta de Hoteles Restaurantes y Servicios, utilizando por primera vez la Clasificación Industrial Internacional Uniforme (CIIU) de las Naciones Unidas.

A partir de 1981, y en base al Censo Económico de 1980, comenzó a ejecutarse la encuesta en dos partes: Tomo I y Tomo II. Cabe mencionar que en 1980 no se realizaron las Encuestas Industriales debido al Censo Económico, puesto que éste incluyó todos los sectores económicos. En el marco de la mejora continua, y de satisfacer las necesidades de los usuarios, a partir del 2008 se cambió el modo de recolección, al pasar de formularios físicos a electrónicos. Así también, se incrementaron, eliminaron o desagregaron algunas variables y capítulos.

En el 2010 el INEC realizó el Censo Nacional Económico (CENEC), el cual fue insumo principal para la actualización del marco muestral, ya que fue realizado en dos etapas. La primera consistió en el empadronamiento a fin de ubicar geográficamente a los establecimientos del universo y contar con un directorio actualizado; la segunda consistió en el levantamiento de la Encuesta Exhaustiva con año de referencia 2011, periodo para el cual no se realizaron las Encuestas Industriales. Es así que, a partir del periodo 2012 se retomó su producción, pero bajo un esquema de marcos muestrales, diseño y estructura de formularios que no la hacía comparable ni con la serie de las Encuestas Industriales de 2010 hacia atrás, ni con la Encuesta Exhaustiva de 2011.

Frente a esta situación, y con el fin de dar continuidad a las series de las principales variables macroeconómicas calculadas en las Encuestas Industriales del 2010 hacia atrás, se procedió a realizar el ejercicio de empalme de encuestas industriales 2010 y 2012 con la Encuesta Exhaustiva 2011, que consistió en analizar la composición de los sectores económicos por cada rama de actividad, con la finalidad de dar continuidad al cálculo de los principales agregados económicos para que sean comparables a través del tiempo. Para la construcción de la serie, se tomó como línea base la información de las Encuestas Industriales (2010). Así, se parte del análisis de la Encuesta Exhaustiva (2011) y se retoman las Encuestas Industriales (2012 -2015).

A partir del año 2016, surge la necesidad de implementar mejoras en las estadísticas empresariales, es por esta razón que se realiza una revisión metodológica de la Encuesta Estructural Empresarial. Esta situación derivó en la implementación de un nuevo diseño muestral, homologándose las Encuestas Industriales (Manufactura, Comercio y Servicios) en un formulario denominado Encuesta Estructural Empresarial, con el fin de reducir la carga de respuesta de los informantes, como una aproximación al uso de registros administrativos, así como, la mejora en los módulos de Información Ambiental, y de uso de Tecnologías de la Información y la Comunicación (TIC), la cual se mantiene hasta la actualidad.

## 3. PLANIFICACIÓN

### 3.1. Objetivos y delimitación de la operación estadística

#### 3.1.1. Objetivos

##### Objetivo General

Elaborar el Diseño Muestral de la “Encuesta Estructural Empresarial 2020”, utilizando técnicas de muestreo probabilístico que permitan emitir resultados confiables de las principales variables que se investigará en dicha encuesta.

##### Objetivos Específicos

- Garantizar la representatividad de la encuesta tanto a nivel geográfico como de los dominios de estudio.
- Calcular factores de expansión y errores asociados al muestreo

#### 3.1.2. Delimitación

##### Tipo de operación estadística

Esta operación estadística corresponde a una encuesta dirigida a empresas.

##### Cobertura temática

A nivel de secciones económicas de la CIIU-4, las empresas a investigar pertenecen a las siguientes ramas:

- B. Explotación de minas y canteras.
- C. Industrias manufactureras.
- D. Suministro de electricidad, gas, vapor y aire acondicionado.
- E. Distribución de agua; alcantarillado, gestión de desechos y actividades de saneamiento
- 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 comunicación.
- K. Actividades financieras y de seguros (No incluye K64 y K66).
- L. Actividades Inmobiliarias
- M. Actividades profesionales científicas y técnicas.
- N. Actividades de servicios administrativos y de apoyo.
- P. Enseñanza.
- Q. Actividades de atención de la salud humana y de asistencia social (No incluye Q88).

- R. Artes, entretenimiento y recreación.
- S. Otras actividades de servicios (No incluye sólo la división S94).

Se excluyen las secciones:

- A. Agricultura, ganadería, silvicultura y pesca.
- O. Administración pública y defensa; planes de seguridad social de afiliación obligatoria.
- T. Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio.
- U. Actividades de organizaciones y órganos extraterritoriales.

En el caso de la sección **A**, se debe a que la forma de investigar dicha sección consta en una encuesta específica de estadísticas agropecuarias. Por otro lado, las secciones **O**, **T** y **U** no hacen referencia a la población objetivo que son las empresas.

Como parte de la cobertura temática se tiene los tamaños de empresa:

- Mediana empresa "A",
- Mediana empresa "B", y,
- Grande empresa.

A los cuales se los han definido en base a la estratificación de la Comunidad Andina de Naciones (CAN), establecida de la siguiente manera:

Tabla 1. Tamaño de empresas según la estratificación de la CAN

Clasificación de la empresa	Personal ocupado	Ventas anuales (USD\$)
Pequeña empresa	10 – 49	100.001 - 1.000.000
Mediana empresa Tipo A	50 – 99	1.000.001 - 2.000.000
Mediana empresa Tipo B	100 – 199	2.000.001 - 5.000.000
Grande empresa	200 y más	5.000.001 y más

Fuente: CAN, Decisión 702, Artículo 3.

### Universo de estudio

Las empresas de tamaño grande, mediana A y mediana B que cumplen con los criterios de clasificación de cobertura temática para el año 2019.

### Unidad de observación y unidad de análisis

Empresas

### Cobertura geográfica

La cobertura geográfica está definida por las empresas que se encuentren ubicadas dentro del territorio ecuatoriano incluyendo la región Insular.

### Población objetivo

Una consideración importante a la hora de estructurar un marco de muestreo, es la relación entre la población objetivo de las encuestas y las unidades de selección. Las unidades de selección deberán estar presentes en el marco de muestreo, así como también la probabilidad de selección de cada una de las unidades de selección en las diferentes etapas del muestreo. Las unidades del marco se constituyen en un elemento básico para el cálculo, asignación y selección de la muestra de unidades poblacionales.

Para la ENESEM 2020 la población objetivo son las empresas que pertenecen a las categorías descritas en la cobertura temática.

### **Continuidad**

Desde el año 2016 hasta la actualidad.

### **Periodicidad**

Anual.

## 4. DISEÑO Y CONSTRUCCIÓN

### 4.1. Diseñar el marco y la muestra

#### 4.1.1. Tipo y etapas de diseño muestral

Muestreo probabilístico de elementos con selección aleatoria.

#### 4.1.2. Dominios de estudio

Los dominios de estudio de la encuesta son los sectores económicos por tamaño de empresa a nivel nacional, detallados en la sección de Cobertura temática. Dentro de cada dominio de estudio se considera como grupo de inclusión forzosa a las empresas con tamaño “Grande empresa”, es decir, estas empresas entran con probabilidad uno (1) a formar parte de la muestra. Cabe recalcar que se incluirá a las empresas que en 2018 eran de tamaño “Grande empresa” y en el 2019 son “Medianas B”, con el fin de llevar el seguimiento de estas empresas en sus características.

#### 4.1.3. Marco muestral

El Marco de Muestreo para la “Encuesta Estructural Empresarial 2020” está conformado por las empresas que constan en el Directorio de Empresas y Establecimientos Económicos 2019 (DIEE-2019), que contiene principalmente variables de identificación; variables de ubicación; variables de estratificación; y, variables de diseño y control de las empresas; indispensables para el diseño muestral.

Para la construcción del Marco Muestral, se toma en cuenta los sectores económicos a los que se enfoca la encuesta que están determinados por la cobertura temática y son los siguientes:

- Manufactura
- Minería
- Construcción
- Comercio
- Servicios

Luego de revisar la base de datos del directorio Empresas y Establecimientos del año 2019 (DIEE-2019) conformado por 886.433 empresas. Para la construcción del Marco Muestral se seleccionó empresas que presenten las siguientes características, respondiendo a las necesidades de información planificadas para la encuesta:

- Empresas de tamaño: Grande Empresa, Mediana Empresa Tipo B y Mediana Empresa Tipo A.
- Se excluye empresas cuya actividad Económica Principal de acuerdo con la sección de Clasificación Industrial Internacional Uniforme de Actividades

Económicas Revisión (CIIU-4) corresponde a Agricultura (A), Administración Pública (O), Actividades de hogares (T) y Actividades de organizaciones (U).

- Se excluye de la selección las siguientes divisiones del CIIU
  - Actividades de servicio financiero, excepto de seguros (K64)
  - Actividades auxiliares de servicio financiero (K66)
  - Actividades de Asistencia Social sin alojamiento (Q88)
  - Actividades de Asociaciones (S94)
- Se excluyen las empresas con forma institucional "Institución Pública".
- Se excluye empresas reportadas como no ubicadas de acuerdo con la variable "empresas no ubicadas" del DIEE.
- Se excluye las empresas Grandes y Medianas "B" con forma institucional "Personas naturales no obligadas a llevar contabilidad".

#### 4.1.4. Variables de diseño

El tamaño de la muestra se calcula en función a su variable de diseño "Ventas totales 2019"

#### 4.1.5. Tamaño muestral

Para la determinación del tamaño de la muestra se requiere establecer la característica o características a estimar, el nivel de confianza y la precisión requeridas, de tal manera que los resultados obtenidos no sean demasiado costosos y/o imprecisos.

En este sentido, la variable de diseño, en base a la cual se calcula el tamaño de la muestra es "Ventas totales 2019". Los valores poblacionales (marco muestral) de esta variable se toman de la base de datos del Directorio de Empresas y Establecimientos del año 2019 (DIEE-2019).

La fórmula para la determinación del tamaño de la muestra corresponde a un muestreo probabilístico para el estimador del total.

$$n_i \geq \frac{(N_i \cdot S_i)^2}{\left(\frac{N_i - 1}{N_i}\right) \cdot \left(\frac{e}{z}\right)^2 + (N_i \cdot S_i^2)} \cdot \frac{1}{1 - tnr_i}$$

Dónde:

- $n_i$  = *Tamaño de la muestra por dominio*
- $N_i$  = *Tamaño del dominio i*
- $S_i$  = *Cuasivarianza del dominio i*
- $e$  = *Error relativo máximo admisible*
- $z$  = *Coficiente que representa el nivel de seguridad o confianza*
- $tnr_i$  = *Tasa de no respuesta del dominio i.*

La cuasivarianza representa la variabilidad del dominio a partir de la información de la variable de diseño.

A partir del marco de muestreo constituido por 14.443 empresas, se ha diseñado una muestra de tamaño igual a 4.324 empresas, con el 10% de error relativo y un nivel de confianza del 90% para la variable de diseño: Ventas totales 2019, se consideró una tasa de no respuesta particular para cada dominio de estudio que se presenta en la tabla 4, obteniéndose una fracción de muestreo global de 0,2997.

La muestra se distribuye de la siguiente manera:

Tabla 2. Distribución de la muestra según el tamaño de empresa

Tamaño de empresa	Número de empresas
Mediana empresa Tipo A	438
Mediana empresa Tipo B	619
Grande empresa	3.267
Total	4.324

Fuente: Directorio de Empresas y Establecimientos Económicos 2019 (DIEE 2019)

## 4.1.6. Selección de la muestra

La selección de las empresas que forman parte de la muestra, de acuerdo con el tamaño establecido, se realiza independientemente en cada uno de los dominios de manera aleatoria, asignando a cada empresa igual probabilidad.

## 4.2. Control de cobertura de campo y muestral

La cobertura hace referencia a la información obtenida en campo y proporcionada por los informantes de las empresas, puede estar constituida de dos formas:

Tabla 3. Estructura de condición de levantamiento de información

Tamaño de empresa	Número de empresas
Efectiva	Cambio de rama de actividad
	Cambio de sector económico
	Cambio de razón social
	Cambio de RUC
	Cambio de Provincia de la empresa
	Empresas que producen para terceros
	Fusionadas
	Absorbidas
	Desintegración
	Escisión
No efectiva	No ubicadas
	Rechazos
	Liquidadas
	Sin características
	Inactivas
	Sin contabilidad en el año de referencia

Fuente: Encuesta Estructural Empresarial (ENESEM 2017)

Se debe considerar que la cobertura está en función de cada uno de los dominios, y se determina a partir del número de empresas efectivas y el total de la muestra.

Debido a esto, a lo largo del levantamiento se realizan seguimientos continuos de cobertura parcial, con el fin de garantizar la calidad de los estimadores.

Como se puede observar en la tabla 4, del total de la muestra de la Encuesta Estructural Empresarial 2020 (4.324 empresas) se obtuvo una cobertura de 4.065 casos (efectivas), es decir, una efectividad del 94,01%.

Tabla 4. Seguimiento de cobertura por dominio de estudio

Dominio	Universo 2019	TNR	Muestra	Universo 2020	Efectivas	Cobertura
3B	62	39,29	45	53	39	86,67
3C	719	38,89	20	581	16	80,00
3D	13	25,00	13	14	12	92,31
3E	38	33,33	38	34	35	92,11
3F	407	68,42	64	279	51	79,69
3G	3.367	38,89	19	3.074	16	84,21
3H	440	33,33	17	399	17	100,00
3I	197	20,00	16	104	15	93,75
3J	109	52,17	25	109	20	80,00
3K	8	33,33	5	7	5	100,00
3L	124	57,89	23	120	21	91,30
3M	422	50,00	28	320	26	92,86
3N	283	39,39	27	208	22	81,48
3P	186	12,50	33	141	32	96,97
3Q	147	18,75	24	139	24	100,00
3R	31	41,67	31	24	27	87,10
3S	34	23,53	10	30	8	80,00
4B	52	23,33	28	40	24	85,71
4C	539	21,74	69	460	56	81,16
4D	6	20,00	6	4	6	100,00
4E	23	14,29	16	29	15	93,75
4F	239	31,48	63	199	45	71,43
4G	2.529	20,18	161	2.244	142	88,20
4H	279	17,86	35	234	34	97,14
4I	97	23,81	25	54	25	100,00
4J	76	9,52	21	74	21	100,00
4K	6	40,00	6	8	6	100,00
4L	87	25,93	30	58	27	90,00
4M	222	33,33	33	162	30	90,91
4N	169	22,58	37	121	29	78,38
4P	110	10,00	38	91	35	92,11
4Q	124	11,43	29	129	28	96,55
4R	15	16,67	12	16	11	91,67
4S	16	38,46	10	17	10	100,00
5B	79	0,00	79	72	65	82,28
5C	674	0,00	674	658	648	96,14
5D	24	0,00	24	25	21	87,50
5E	28	0,00	28	24	28	100,00
5F	139	0,00	139	105	128	92,09
5G	1.581	0,00	1.581	1.444	1.530	96,77
5H	185	0,00	185	145	181	97,84
5I	54	0,00	54	29	53	98,15
5J	62	0,00	62	57	58	93,55
5K	42	0,00	42	43	42	100,00
5L	39	0,00	39	30	37	94,87
5M	82	0,00	82	71	80	97,56
5N	100	0,00	100	71	92	92,00
5P	68	0,00	68	49	64	94,12
5Q	83	0,00	83	86	82	98,80
5R	16	0,00	16	9	15	93,75
5S	11	0,00	11	9	11	100,00
<b>Total</b>	<b>14.443</b>		<b>4.324</b>	<b>12.503</b>	<b>4.065</b>	<b>94,01</b>

Fuente: Diseño Muestral, INEC 2022.

Nota: La variable "Dominio" está constituido por un número y una letra, el número describe el tamaño de la empresa, es decir, 3: Mediana tipo "A", 4: Mediana tipo "B" y 5: Grande. Por otro lado, la letra adjunta representa la rama de actividad a la que pertenece y está anclada con el CIU-4, el cual está especificado en el apartado "Cobertura temática".

## 4.3. Construcción de factores de expansión

### 4.3.1. Factor de expansión teórico

El objetivo de una encuesta por muestreo es hacer inferencias acerca de la población a partir de la información contenida en la muestra. En este sentido se utilizan factores de expansión para llevar la información de la muestra a niveles poblacionales. Se define como el inverso de la probabilidad de escoger una empresa en un determinado estrato, considerando el informe de cobertura que registra las novedades de la empresa.

El factor de expansión para cada dominio de estudio se determina en base al número de empresas tanto del marco de muestreo como de la muestra, así:

$$a_{ki} = \frac{N_k - n_{fk}}{n_k - n_{fk}}$$

Dónde:

$a_{ki}$  = Factor de expansión de la empresa  $i$  perteneciente al dominio  $k$ .

$N_k$  = Tamaño de la población del dominio  $k$ .

$n_k$  = Tamaño de la muestra del dominio  $k$ .

$n_{fk}$  = Tamaño de empresas de inclusión forzosa en el dominio  $k$ .

A partir de ahora, se notará por  $a_k$  al factor de expansión de la empresa  $i$  perteneciente al dominio  $k$   $a_{ki}$ .

**Nota:** La empresa con identificación 14596814175 se categorizó según el directorio 2019 como Mediana A, sin embargo, se realiza un análisis histórico que determina el tamaño de la empresa como Grande; por lo que, para el cálculo de factores de expansión se la considera como de inclusión forzosa (Grande empresa).

### 4.3.2. Factor postestratificado

Considerando que la muestra fue seleccionada a partir del DIEE 2019 y el periodo de la información levantada en la encuesta corresponde al año 2020, existen empresas que por el dinamismo económico cambiaron de tamaño y/o sector económico. En consecuencia, los dominios de estudio originales sufrieron cambios.

En la tabla 5, se evidencia el cambio del tamaño de las empresas en la muestra, donde se registran 1.064 variaciones, lo cual corresponde al 26,2% del total de empresas efectivas.

Tabla 5. Cambio de tamaño de las empresas de la muestra

DIEE 2019	DIEE 2020							
	Tamaño de empresa	Micro Empresa	Pequeña Empresa	Mediana empresa tipo "A"	Mediana empresa tipo "B"	Grandes empresas	Sin tamaño	Total cambios
Mediana Tipo "A"		6	177	172	28	3	0	214
Mediana Tipo "B"		7	47	128	316	44	2	228
Grandes Empresas		5	33	53	525	2.513	6	622
<b>Total cambio</b>		<b>18</b>	<b>257</b>	<b>181</b>	<b>553</b>	<b>47</b>	<b>8</b>	<b>1.064</b>

Fuente: Diseño Muestral, INEC 2022

Adicional, para el 2020 se dan de baja 283 empresas de la población objetivo, ya que se distribuyen entre sin tamaño (8), microempresa (18) y pequeñas empresas (257).

De igual forma existen cambios en las secciones económicas, estos se presentan en 98 empresas, es decir, el 2,41% del total de empresas efectivas. Por otro lado, es necesario destacar que 8 empresas (0,20%) pasaron a no tener ninguna rama de actividad.

Tabla 6. Cambio de sector económico de las empresas pertenecientes a la muestra

DIEE 2019	Sección Económica	Sin Rama	DIEE 2020																	Total
			B	C	D	E	F	G	H	I	J	K	L	M	N	P	Q	R	S	
	B	0	122	4	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	6
	C	2	0	702	0	0	2	13	0	0	0	0	0	0	1	0	0	0	0	18
	D	1	0	0	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	E	0	0	0	0	75	1	2	0	0	0	0	0	0	0	0	0	0	0	3
	F	0	0	1	0	0	219	2	0	0	1	0	0	0	1	0	0	0	0	5
	G	4	2	12	0	0	1	1.667	0	0	1	0	1	0	0	0	0	0	0	21
	H	0	0	0	0	0	0	1	231	0	0	0	0	0	0	0	0	0	0	1
	I	0	0	0	0	0	0	3	0	90	0	0	0	0	0	0	0	0	0	3
	J	0	0	0	0	0	0	3	0	0	96	0	0	0	0	0	0	0	0	3
	K	0	0	0	0	0	0	0	0	0	0	53	0	0	0	0	0	0	0	0
	L	0	0	1	0	0	5	2	0	1	0	0	74	0	2	0	0	0	0	11
	M	1	1	1	0	0	8	5	1	0	0	0	0	118	1	0	0	0	0	18
	N	0	0	1	0	0	1	1	0	0	0	0	0	0	140	0	0	0	0	3
	P	0	0	0	0	0	0	1	0	0	0	0	0	0	0	130	0	0	0	1
	Q	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	133	0	0	1
	R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	52	1	1
	S	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	27	2
	<b>Total</b>	<b>8</b>	<b>3</b>	<b>21</b>	<b>0</b>	<b>0</b>	<b>18</b>	<b>35</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>6</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>98</b>

Fuente: Diseño Muestral, INEC 2022

Como consecuencia de los cambios de tamaño y sección económica, no todas las empresas que fueron efectivas contarán con factor de expansión, esto se debe a que estas empresas están fuera de la población objetivo. Para la ENESSEM 2020, las empresas que cuentan con factor de expansión son 3.782, es decir, una efectividad de 87,47%.

Ahora bien, estos cambios de dominio ocurren tanto en la muestra como en todo el marco muestral. Con el fin de solventar estos inconvenientes se utiliza un método denominado postestratificación, el cual, permite realizar un ajuste a los factores de expansión mediante la incorporación de información auxiliar. En este caso, la información auxiliar utilizada proviene del DIEE 2020, desde el cual se realiza un seguimiento a las empresas de la población objetivo. Esta información se convierte

en la base para realizar actualizaciones al universo de investigación y asignar el dominio de estudio al que pertenece cada una de las empresas investigadas.

De esta manera, un postestrato está conformado por las empresas que tienen un mismo tamaño de empresa y sección económica de acuerdo con el DICE 2020, fragmentando los dominios originales de la muestra.

La fórmula de cálculo del factor de expansión por empresa es:

$$fexp_{kp} = a_k \times \frac{N_p}{\hat{N}_p}$$

Dónde:

$fexp_{kp}$  = Factor de expansión de la empresa  $k$  perteneciente al postestrato  $p$ .

$\hat{N}_p$  = Tamaño estimado de la población del postestrato  $p$ .

$N_p$  = Tamaño de la población del postestrato  $p$ .

Es necesario mencionar que las empresas que tienen factor 1 corresponden a:

- Empresas grandes de acuerdo con el dominio de selección de la muestra original (DICE 2019).
- Empresas grandes según DICE 2018, que se convirtieron en mediana B según el DICE 2019, con el fin de realizar seguimiento de dichas empresas.

### Recorte del factor de expansión

Una vez postestratificados los factores de expansión se realiza un control de las empresas cuyo factor de corrección  $\left(\frac{N_p}{\hat{N}_p}\right)$  es superior a 3 y aquellas empresas grandes según dominio 2020 que tienen el factor superior o igual a 5. Estas empresas se identifican y se reportan a la Gestión de Estadísticas Estructurales para constatar que han sido categorizadas correctamente de acuerdo al tamaño. Una vez pasada esta verificación se procede a calcular  $tam5$  que es igual al número de veces que una empresa ha sido grande en los últimos 5 años. Finalmente, se realiza un recorte a los factores de expansión de acuerdo a los siguientes criterios.

$$\widetilde{fexp}_{kp} = \begin{cases} 1 & \text{si } tam_{2020} = 5, fexp_{kp} > 5 \text{ y } tam5 \geq 3 \\ a_k & \text{si } tam_{2020} = 5, fexp_{kp} > 5 \text{ y } tam5 < 3 \text{ o } fexp_{kp} < 1 \\ fexp_{kp} & \text{Caso contrario} \end{cases}$$

## 4.4. Estimaciones y errores muestrales

### 4.4.1. Estimación de características

Debido a la postestratificación que se realiza, la estimación del Total Poblacional para determinada característica o variable  $Y$ , se obtiene a partir de la siguiente fórmula:

$$\hat{Y}_{POST} = \sum_{p=1}^P \sum_{sp} fexp_{kp} y_k$$

Dónde:

- $\hat{Y}_{POST}$  = Estimador postestratificado de la característica  $y$ .
- $fexp_{kp}$  = Factor de expansión de la empresa  $k$  perteneciente al postestrato  $p$ .
- $y_k$  = Característica  $y$  de la empresa  $k$ .
- $s_p$  = Empresas que pertenecen al postestrato  $p$ .

## 4.4.2. Error de muestreo

El error de muestreo cometido para la estimación de características podrá ser calculado a partir de la estimación de la varianza del estimador del total postestratificado  $\hat{Y}_{POST}$ , cuya fórmula es la siguiente:

$$CV(\hat{Y}_{POST}) = \frac{\sqrt{\widehat{Var}(\hat{Y}_{POST})}}{\hat{Y}_{POST}}$$

Con:

$$\widehat{Var}(\hat{Y}_{POST}) = \sum_s \sum_{kl} \frac{\Delta_{kl} g_k e_k g_l e_l}{\pi_{kl} \pi_k \pi_l}$$

$$\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$$

$$\pi_k = \frac{1}{ak}$$

$$e_k = y_k - \tilde{y}_{sp}$$

$$\tilde{y}_{sp} = \frac{1}{N_p} \sum_{sp} fexp_{kp} y_k$$

$$g_k = \frac{N_p}{\tilde{N}_p}$$

Dónde:

- $CV(\hat{Y}_{POST})$  = Coeficiente de variación del estimador de total postestratificado.
- $\widehat{Var}(\hat{Y}_{POST})$  = Estimador de la varianza del estimador de total postestratificado.
- $\pi_{kl}$  = Probabilidad de inclusión conjunta de las empresas  $k$  y  $l$ .
- $\tilde{y}_{sp}$  = Estimador de la media en el postestrato  $p$ .

## 4.5. Declaración del Plan de muestreo

Para el cálculo de los estimadores y sus errores, varianzas, coeficientes de variación, intervalos de confianza y efecto de diseño considerando el diseño muestral, se puede utilizar el software estadístico SPSS, en el que se construye un plan de

muestreo para muestras complejas. Aunque el paquete no tiene las facilidades para realizar un diseño con postestratificación, se puede estimar con un diseño aproximado, para el cual la sintaxis del plan de muestreo es la siguiente:

```

CSPLAN ANALYSIS
/PLAN FILE='C:\Users\Estructurales\plan_muestreo_final.csaplan'
/PLANVARS ANALYSISWEIGHT=f_exp
/PRINT PLAN
/DESIGN STRATA=Pst
/ESTIMATOR TYPE=EQUAL_WOR
/POPSIZE VARIABLE=N_fin.

```

Dónde las variables corresponden a:

- “Pst”: postestrato,
- “fexp”: factor de expansión; y,
- “N\_fin”: total de la población por postestrato.

De igual manera, también se puede calcular utilizando el software Stata con las siguientes líneas de programación:

```

svyset [pweight=f_exp], strata(Pst) fpc(N_fin) vce(linearized)
svyset, singleunit(certainty) noclear

```

Por último, en el software R se utiliza la siguiente sintaxis:

```

plan_muestreo <- factores_fin_V1_1 %>%
  as_survey_design (ids = "id_empresa",
                    strata="Pst",
                    fpc= "N_fin",
                    weights = f_exp,
                    nest = T)
options(survey.lonely.psu = "certainty")

```

## 5. BIBLIOGRAFÍA

- Hidiroglu, M., Särndal, C., & Binder, D. (1995). Weighting and Estimation in Business Surveys. En Cox, Binder, Chinnappa, Christianson, Colledge, & Kott, Business Survey Methods (págs. 477-502). Jhon Wiley & Sons.
- Kish, L. (1972). Muestreo de Encuestas. México: Trillas.
- Särndal, C., Swensson, B., & Wretman, J. (2003). Model Assisted Survey Sampling, Nueva York, USA. New York: Springer.



 Buenas cifras,  
mejores vidas

