

**INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA**

**DIRECCIÓN GENERAL
DE ESTADÍSTICAS ECONÓMICAS**

**DIRECCIÓN DE MARCOS
Y MUESTREO**

**Encuesta sobre el Impacto
Económico Generado
por COVID-19
en las Empresas (ECOVID-IE)
2020**

Segunda Ronda

**Documento Metodológico
sobre Diseño Muestral**



Índice

Introducción	VI
1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura	1
3.1. Cobertura geográfica	1
3.2. Cobertura sectorial	1
4. Dominio de estudio	1
5. Unidad de muestreo	1
6. Marco de muestreo	2
7. Estratificación	2
8. Esquema de muestreo	3
9. Tamaño de la muestra	3
10. Afijación de la muestra	4
11. Selección de la muestra	4
12. Cálculo de factores de expansión	4
13. Estimadores	5

Introducción

Como parte de la transparencia de las metodologías utilizadas en las Encuestas Económicas, se presenta este documento metodológico que describe los elementos considerados para la definición y construcción del esquema de muestreo de la Encuesta sobre el Impacto Económico Generado por COVID-19 en las Empresas (ECOVID-IE) 2020 segunda ronda, para obtener información sobre el impacto generado por la contingencia dentro de las empresas de los sectores económicos del país.

Las diferentes secciones ayudan a determinar el tamaño de muestra con su distribución y selección; de tal forma que se tengan los elementos necesarios para definir la precisión de la información, así como el análisis de la captación para la evaluación de las estimaciones finales, mediante el cálculo de varianza y coeficiente de variación.

1. Objetivo de la encuesta

Generar indicadores cualitativos sobre las afectaciones de la contingencia originada por COVID-19 en las empresas del país.

2. Población objetivo

Está definida por todas aquellas empresas del país con instalaciones fijas que realizan actividades económicas correspondientes al sector industria (minería, electricidad, suministro de agua y gas, construcción y manufacturas), comercio (al por mayor y por menor) y servicios (incluye transportes). Por lo que se excluyen las actividades relacionadas a la agricultura, cría y explotación de animales, pesca y acuicultura, y gobierno, según el Sistema de Clasificación Industrial de América del Norte 2018 (SCIAN 2018).

3. Cobertura

3.1. Cobertura geográfica

Nacional.

3.2. Cobertura sectorial

Permite obtener estimaciones para los grandes sectores de actividad, según el SCIAN 2018, que están integrados de la siguiente forma:

- a. Sector industria, conformado por las actividades de minería, energía eléctrica, suministro de agua y gas, construcción y manufacturas.
- b. Sector comercio, integrado por las actividades del comercio al por mayor y comercio al por menor.
- c. Sector servicios, constituido por todos los servicios privados, incluyendo transportes.

4. Dominio de estudio

Se refiere a subconjuntos de la población objeto de estudio, para los que se pretende publicar información y para los cuales se diseña de manera independiente una muestra en cada uno de ellos.

Los que se consideran en la encuesta son:

- Nacional.
- Nacional-gran sector de actividad.
- Nacional-tamaño de empresa.

Donde el tamaño de empresa se refiere a los estratos considerados en la sección 7.

5. Unidad de muestreo

Las empresas que realizan su actividad económica de interés en MIPYME¹, y EMOE² restringiendo a números telefónicos validados, excepto para el sector minería es la unidad minera.

¹ Involucra empresas de tamaño micro, pequeña y mediana.

² Para los sectores comercio, construcción, manufacturas y servicios.

6. Marco de muestreo

Se integró por **1 873 564** empresas con un total de **19 854** registros para EMOE³ y **1 853 710** para MIPYME⁴. La distribución del marco de muestreo se presenta a continuación:

Cuadro 1

Gran sector de actividad SCIAN 2018	Marco
Total	1 873 564
EMOE	19 854
Comercio	6 180
Construcción	1 323
Manufactura	6 077
Servicios	6 274
MIPYME	1 853 710
Comercio	743 113
Industria	199 805
Servicios	910 792

7. Estratificación

Se realizó para la variable de total de personal ocupado de acuerdo con los límites presentados en el siguiente cuadro:

Cuadro 2

Tipo de marco	Gran sector de actividad SCIAN 2018	Estrato de personal ocupado		
		Grande	PYME ^a	Micro
EMOE	Comercio	101 y más	50 a 100	-
	Construcción	101 y más	-	-
	Manufacturas	101 y más	-	-
	Servicio	101 y más	-	-
MIPYME	Comercio	-	11 a 100	1 a 10
	Industria	-	11 a 100	1 a 10
	Servicio	-	11 a 100	1 a 10

^a Involucra empresas de tamaño pequeña y mediana.

Por lo tanto, la distribución del marco de muestreo por tamaño de empresa al que denominaremos estrato, se presenta a continuación:

Cuadro 3
(1a. parte)

Tipo de marco	Gran sector de actividad SCIAN 2018	Total	Estrato		
			Grande	PYME	Micro
Total		1 873 564	16 893	128 261	1 728 410
EMOE	Comercio	6 180	3 219	2 961	-
	Construcción	1 323	1 323	-	-
	Manufactura	6 077	6 077	-	-
	Servicios	6 274	6 274	-	-

(Continúa)

³ Versión RENEM 04 octubre 2018.

⁴ Versión RENEM 02 de abril 2020.

Cuadro 3
(2a. parte y última)

Tipo de marco	Gran sector de actividad SCIAN 2018	Total	Estrato		
			Grande	PYME	Micro
Total		1 873 564	16 893	128 261	1 728 410
MIPYME	Comercio	743 113	-	29 954	713 159
	Industria	199 805	-	25 053	174 752
	Servicios	910 792	-	70 293	840 499

8. Esquema de muestreo

Es probabilístico y estratificado, con selección independiente en cada estrato por dominio de estudio:

- Probabilístico
Las unidades de muestreo tienen una probabilidad conocida y distinta de cero de ser seleccionadas.
- Estratificado
Las unidades de muestreo con características similares se agrupan para formar estratos.

9. Tamaño de la muestra

Se calculó de manera independiente para cada dominio de estudio, considerando los siguientes parámetros:

Cuadro 4

Parámetros (%)		
Nivel de confianza		95.00
Promedio de empresas afectadas por el COVID-19		50.00
Error relativo	MIPYME Comercio	5.63
	MIPYME Industria	5.66
	MIPYME Servicios	5.84
	EMOE Construcción	15.64
	EMOE Manufacturas	6.86
	EMOE Comercio	10.17
	EMOE Servicios	7.04

La expresión que se utilizó para calcular el tamaño de la muestra en cada dominio es la siguiente:

$$n_d = \frac{N_d z_{1-\alpha/2}^2 p q}{N_d r^2 p^2 + z_{1-\alpha/2}^2 p q}$$

Donde:

- n_d = tamaño de la muestra en el d-ésimo dominio⁵
- N_d = total de unidades económicas en el marco en el d-ésimo dominio
- $z_{1-\alpha/2}$ = valor de tablas estadísticas de la distribución normal estándar para una determinada confianza
- p = promedio de empresas afectadas por el COVID-19
- q = 1-p
- r = error relativo

⁵ Véase el apartado 4.

El tamaño de la muestra resultante es de **5 671** empresas considerando el máximo tamaño de muestra para los dominios de interés.

10. Afijación de la muestra

Se realizó una distribución proporcional al número de personal ocupado de acuerdo con el sector de actividad, mediante la siguiente expresión, donde el dominio considerado fue **nacional-sector de actividad** y **nacional-gran sector de actividad**:

$$n_{dh} = n_d * \frac{PO_{dh}}{PO_d}$$

Donde:

n_{dh} = tamaño de la muestra en el h-ésimo estrato, en el d-ésimo dominio

n_d = tamaño de la muestra en el d-ésimo dominio

PO_{dh} = total de personal ocupado en el marco en el h-ésimo estrato, en el d-ésimo dominio

PO_d = total de personal ocupado en el marco en el d-ésimo dominio

El tamaño de la muestra total afijado es de **5 671** empresas y la distribución por gran sector-estrato se presenta a continuación:

Cuadro 5

Tipo de marco	Gran sector de actividad SCIAN 2018	Estrato			
		Total	Grande	PYME	Micro
Total		5 671	2 098	1 564	2 009
EMOE	Comercio	373	343	30	-
	Construcción	159	159	-	-
	Manufactura	818	818	-	-
	Servicios	778	778	-	-
MIPYME	Comercio	1 214	-	345	869
	Industria	1 201	-	687	514
	Servicios	1 128	-	502	626

11. Selección de la muestra

Se realizó de forma aleatoria e independiente en cada dominio de estudio de acuerdo con la afijación.

12. Cálculo de factores de expansión

Se define como el inverso de la probabilidad de seleccionar una empresa del h-ésimo estrato, del d-ésimo dominio.

La expresión de cálculo que se utilizó para el **dominio-estrato** donde el dominio considerado fue **nacional-sector de actividad** y **nacional-gran sector de actividad** es la siguiente:

$$f_{dh} = \frac{N_{dh}}{n_{dh}}$$

Donde:

- f_{dh} = factor de expansión del h-ésimo estrato, del d-ésimo dominio
 N_{dh} = total de unidades económicas en el marco, en el h-ésimo estrato, en el d-ésimo dominio
 n_{dh} = tamaño de la muestra en el h-ésimo estrato, en el d-ésimo dominio.

13. Estimadores

El estimador total nacional de la característica X se calcula con base en la fórmula:

$$\hat{t}_d = \sum_{h=1}^H \sum_{i=1}^{n_{dh}} f_{dh} x_{dhi}$$

Para la estimación de proporciones, tasas y promedios se utiliza el estimador de razón:

$$\hat{R} = \frac{\hat{X}}{\hat{Y}}$$

\hat{Y} se define en forma análoga a \hat{X} .

Donde:

- \hat{t}_d = total estimado en el d-ésimo dominio
 H = número de estratos en cada dominio
 n_{dh} = tamaño de la muestra en el h-ésimo estrato, en el d-ésimo dominio
 f_{dh} = factor de expansión del h-ésimo estrato, del d-ésimo dominio
 x_{dhi} = valor de la característica de interés de la i-ésima unidad económica, del h-ésimo estrato, del d-ésimo dominio
 \hat{R} = estimador de razón
 \hat{X} = estimador total de la característica X
 \hat{Y} = estimador total de la característica Y

El estimador de la varianza del estimador del total de la característica X es:

$$\hat{V}(\hat{t}_d) = \sum_{h=1}^H N_{dh}^2 \left(1 - \frac{n_{dh}}{N_{dh}}\right) \frac{s_{dh}^2}{n_{dh}} \quad \text{donde} \quad s_{dh}^2 = \frac{1}{n_{dh}-1} \sum_{i=1}^{n_{dh}} [x_{dhi} - \bar{x}_{dh}]^2$$

Donde:

- $\hat{V}(\hat{t}_d)$ = varianza estimada del total estimado del d-ésimo dominio
 \hat{t}_d = total estimado en el d-ésimo dominio
 H = número de estratos en cada dominio
 N_{dh} = total de unidades económicas en el marco en el h-ésimo estrato, en el d-ésimo dominio
 n_{dh} = tamaño de la muestra en el h-ésimo estrato, en el d-ésimo dominio
 s_{dh}^2 = varianza muestral del h-ésimo estrato, del d-ésimo dominio
 x_{dhi} = valor de la característica de interés en la i-ésima unidad económica, en el h-ésimo estrato, en el d-ésimo dominio
 \bar{x}_{dh} = media muestral del h-ésimo estrato, del d-ésimo dominio

El estimador de la varianza del estimador de razón \hat{R} de las características X y Y por dominio está dado por:

$$\widehat{V}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_{h=1}^H \frac{N_h(N_h - n_h)}{n_h(n_h - 1)} \sum_{i=1}^{n_h} [(x_{hi} - \bar{x}_h) - \hat{R}(y_{hi} - \bar{y}_h)]^2$$

Donde:

- $\widehat{V}(\hat{R})$ = varianza del estimador de razón \hat{R}
- \hat{R} = estimador de razón de las características X y Y
- \hat{Y} = estimación de total de la característica Y
- H = número de estratos
- N_h = total de unidades económicas en el marco, en el h-ésimo estrato
- n_h = tamaño de la muestra en el h-ésimo estrato
- x_{hi} = valor de la característica de interés en la i-ésima unidad económica, en el h-ésimo estrato
- \bar{x}_h = media muestral del h-ésimo estrato
- y_{hi} = característica Y a estimar reportada en la i-ésima unidad económica, en el h-ésimo estrato
- \bar{y}_h = media muestral del h-ésimo estrato

Error de estimación (EE): $\widehat{EE} = \sqrt{\widehat{V}(\hat{t}_d)}$

Coefficiente de variación (CV): $\widehat{CV} = \frac{\widehat{EE}}{\hat{t}_d}$

Intervalo de confianza: $\left(\hat{t}_d - Z_{1-\alpha/2} \sqrt{\widehat{V}(\hat{t}_d)}, \hat{t}_d + Z_{1-\alpha/2} \sqrt{\widehat{V}(\hat{t}_d)} \right)$

Donde:

- $\widehat{V}(\hat{t}_d)$ = varianza estimada del total estimado del d-ésimo dominio
- \hat{t}_d = total estimado en el d-ésimo dominio
- $Z_{1-\alpha/2}$ = valor de tablas estadísticas de la distribución normal estándar para una determinada confianza.