



METODOLOGÍA DEL DISEÑO MUESTRAL DE LA “ENCUESTA ESTRUCTURAL EMPRESARIAL 2016”

COORDINACION GENERAL TÉCNICA DE PRODUCCION ESTADÍSTICA
Dirección de Infraestructura Estadística y Muestreo
Unidad de Diseño Muestral

Marzo
2018

Tabla de contenido

1. Introducción	3
2. Objetivo	3
3. Determinación del marco de muestreo.	3
4. Universo de investigación.....	3
6. Cobertura geográfica.....	4
7. Cobertura temática.	4
8. Diseño Muestral	5
8.1. Dominios de estudio.....	5
9. Tamaño de la muestra.....	5
10. Selección de la muestra.....	6
11. Cobertura obtenida en campo	6
12. Cálculo de los factores de expansión	7
13. Postestratificación	7
14. Estimación de Características.....	9
15. El error de muestreo.....	10
16. Plan de muestreo.....	10
Bibliografía.....	11

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, la Dirección de Estadísticas Económicas solicita el apoyo de la Dirección de Infraestructura Estadística y Muestreo en la creación del diseño muestral para la “Encuesta Estructural Empresarial 2016”. Esta encuesta se realiza a nivel nacional para las distintas secciones económicas contempladas por el CIU-4 para empresas grandes y medianas.

2. Objetivo

Elaborar el Diseño Muestral de la “Encuesta Estructural Empresarial 2016”, utilizando técnicas de muestreo probabilístico que permitan emitir resultados confiables de las principales variables que se investigará en dicha encuesta.

3. Determinación del marco de muestreo.

El Marco de Muestreo para la “Encuesta Estructural Empresarial 2016” está conformado por las empresas que constan en el Directorio de Empresas y Establecimientos Económicos 2015 (DIEE-2015), 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.

Luego de revisar la base de datos del directorio Empresas y Establecimientos del año 2015 (DIEE-2015) conformado por 844.999 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 a la sección de Clasificación Industrial Internacional Uniforme de Actividades Económicas Revisión (CIU-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 CIU
 - 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 a la variable “empresas_noubicadas” del DIEE.

4. Universo de investigación

El universo para la Encuesta de “Encuesta Estructural Empresarial 2016” queda conformado por 13.773 Empresas.

5. Unidades de muestreo y análisis.

La unidad elemental de muestreo es la empresa.

6. Cobertura geográfica.

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

7. 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).

Como parte de la cobertura temática se tiene los tamaños de empresa: Mediana empresa “A”, Mediana empresa “B” y Grande empresa. Definidos considerando la estratificación de la Comunidad Andina de Naciones (CAN), establecida de la siguiente manera:

Tabla 1. Tamaño de empresas 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

A continuación se detalla el número de empresas del marco de muestreo por tamaño:

Tabla 2. Empresas del marco de muestreo por tamaño de empresa

Tamaño de empresa	Mediana empresa "A"	Mediana empresa "B"	Grande empresa	Total
Número de empresas	6.301	4.249	3.223	13.773

Fuente: Directorio de Empresas y Establecimientos Económicos 2015 (DIEE-2015).

8. Diseño Muestral

El diseño muestral para esta investigación corresponde a un muestreo probabilístico de elementos con selección aleatoria.

8.1. 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 7 *Cobertura temática*. Dentro de cada dominio de estudio se considera como un grupo de inclusión forzosa a las empresas con tamaño "Grande empresa", es decir, estas empresas entran con probabilidad 1 a formar parte de la muestra.

9. Tamaño de la muestra

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 o imprecisos.

En este sentido, la variable de diseño, en base a la cual se calcula el tamaño de la muestra es "**Ventas totales 2015**". Los valores poblacionales (marco muestral) de esta variable se toman de la base de datos del Directorio de Empresas y Establecimientos del año 2015 (DIEE-2015).

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

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

donde:

- n_i = Tamaño de la muestra por dominio
- N_i = Tamaño del dominio i
- S_i = Cuasivarianza del dominio i
- e = Error absoluto máximo admisible
- z = Coeficiente que representa el nivel de seguridad o confianza
- SOB = *Sobremuestreo*

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 13.773 empresas, se ha diseñado una muestra de tamaño igual a 4.003 empresas, con el 10% de error absoluto y un nivel de confianza del 90% para la *variable de diseño: Ventas totales 2015*, con un sobremuestreo del 35% obteniéndose de esta forma una fracción de muestreo global de 0,29064.

10. Selección de la muestra.

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

11. Cobertura obtenida en campo

Tabla 3 Cobertura por dominio de estudio

DOMINIO	UNIVERSO	MUESTRA SIN SOBREMUESTREO	MUESTRA FINAL	EFFECTIVAS	COBERTURA	CUMPLE
3B	50	24	33	27	81,8	●
3C	674	13	18	17	94,4	●
3D	8	3	4	3	75,0	●
3E	18	17	18	16	88,9	●
3F	419	26	36	29	80,6	●
3G	3250	12	16	14	87,5	●
3H	380	14	20	18	90,0	●
3I	180	10	14	13	92,9	●
3J	114	13	18	17	94,4	●
3K	4	4	4	3	75,0	●
3L	138	12	16	16	100,0	●
3M	474	21	28	25	89,3	●
3N	238	20	26	24	92,3	●
3P	168	36	50	49	98,0	●
3Q	127	16	22	22	100,0	●
3R	32	24	32	28	87,5	●
3S	27	12	16	13	81,3	●
4B	48	26	35	33	94,3	●
4C	491	18	24	24	100,0	●
4D	7	4	6	6	100,0	●
4E	21	16	21	19	90,5	●
4F	255	25	34	26	76,5	●
4G	2285	19	26	24	92,3	●
4H	239	19	26	25	96,2	●
4I	81	16	22	20	90,9	●
4J	79	14	20	20	100,0	●
4K	9	7	9	8	88,9	●
4L	66	17	22	19	86,4	●
4M	245	20	28	22	78,6	●
4N	191	23	32	32	100,0	●
4P	86	32	43	37	86,0	●
4Q	103	17	22	22	100,0	●
4R	24	17	23	21	91,3	●
4S	19	12	16	15	93,8	●
5B	92	92	92	82	89,1	●
5C	674	674	674	662	98,2	●
5D	24	24	24	24	100,0	●
5E	26	26	26	25	96,2	●
5F	164	164	164	141	86,0	●
5G	1533	1533	1533	1467	95,7	●
5H	150	150	150	147	98,0	●
5I	50	50	50	50	100,0	●
5J	61	61	61	58	95,1	●
5K	46	46	46	42	91,3	●
5L	41	41	41	39	95,1	●
5M	125	125	125	115	92,0	●
5N	84	84	84	82	97,6	●
5P	60	60	60	59	98,3	●
5Q	73	73	73	73	100,0	●
5R	10	10	10	9	90,0	●
5S	10	10	10	10	100,0	●

Dentro de la etapa de diseño de la encuesta se planificó un sobremuestreo del 35%. Esto significa que la tasa de no respuesta esperada es de 25,9%.

$$TNR = 1 - \frac{1}{1 + SOB}$$

Como se puede observar en la tabla 3 todos los dominios de estudio tienen una cobertura superior al 74,1%, por lo que las estimaciones no presentarán problemas por cobertura. Las coberturas que están con el semáforo amarillo son los dominios que se enviaron a levantar información para todas las empresas del universo pero no se investigó al 100%.

$$cob_i = \frac{nef_i}{n_i}$$

donde:

cob_i = Cobertura para el dominio i.

nef_i = Encuestas efectivas por dominio

La encuesta de manera global obtuvo una cobertura del 94,7%. El 5,3% faltante corresponde a empresas que tienen como novedad que:

- Empresas que no cuentan con características de la población objetivo.
- Inactivas.
- Liquidadas.
- Rechazos.
- Empresas no ubicadas.
- Empresas sin contabilidad en el año de referencia.

12. Cálculo de los factores de expansión

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 estudiantes tanto del marco de muestreo como de la muestra, así:

$$a_{k_i} = \frac{N_k}{n_k}$$

donde:

a_{k_i} = 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.

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

13. Postestratificación

Considerando que la muestra fue seleccionada a partir del DIEE 2015 y el periodo de la información levantada en la encuesta corresponde al año 2016, existen empresas que por el dinamismo económico cambiaron de tamaño y/o sector económico. En consecuencia los dominios de estudio originales sufrieron cambios.

En el caso de cambio de tamaño de las empresas de la muestra, como se puede observar en la tabla 4, se registran 877 variaciones, esto corresponde al 21,91% de la muestra. De estos cambios, 303 pasan a ser micro, pequeñas o sin tamaño, por lo que estas empresas para el 2016 salen de la población objetivo.

Tabla 4 Cambio de tamaño de empresa en la muestra 2015-2016

Tamaño de empresa		DIEE 2016					
		Sin tamaño	Micro empresa	Pequeña empresa	Mediana empresa tipo "A"	Mediana empresa tipo "B"	Grande Empresa
DIEE 2015	Mediana empresa tipo "A"	10	14	117	196	32	2
	Mediana empresa tipo "B"	10	8	34	65	263	29
	Grande Empresa	28	24	58	46	400	2667

De igual forma existen cambios en las secciones económicas, estos se presentan en 229 empresas, es decir, el 5,72% de la muestra.

Tabla 5 Cambio de sector económico en la muestra 2015-2016

Sección Económica		DIEE 2016																		
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
DIEE 2015	B	0	142	6	0	0	4	3	3	0	0	0	0	0	1	1	0	0	0	0
	C	3	0	700	0	0	6	5	1	0	0	0	1	0	0	0	0	0	0	0
	D	0	0	0	33	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	E	0	0	2	0	57	0	4	0	0	0	0	0	1	0	1	0	0	0	0
	F	1	0	0	0	0	227	2	1	0	0	0	1	1	0	1	0	0	0	0
	G	29	3	36	1	0	2	1491	4	1	4	0	0	1	1	0	0	1	0	1
	H	1	0	2	0	0	0	7	184	0	0	0	0	1	0	1	0	0	0	0
	I	0	0	1	0	0	0	0	0	84	0	0	1	0	0	0	0	0	0	0
	J	0	0	1	0	0	0	0	0	0	96	0	0	0	2	0	0	0	0	0
	K	0	0	0	0	0	0	0	0	0	0	58	0	0	0	0	0	1	0	0
	L	1	0	0	0	0	6	4	0	1	0	0	65	1	0	0	0	1	0	0
	M	1	1	1	0	1	17	3	2	0	1	0	0	151	2	0	0	0	0	1
	N	1	0	2	1	0	1	0	1	1	0	0	2	0	133	0	0	0	0	0
	P	2	0	0	0	0	1	0	0	0	1	0	0	2	0	0	147	0	0	0
	Q	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	114	0	0
R	0	0	0	0	1	0	0	0	0	2	0	0	2	2	0	0	0	58	0	
S	1	0	0	0	0	3	3	0	0	0	0	0	0	1	0	0	0	0	34	

Como se puede observar en las tablas anteriores, existen cambios significativos en la distribución de la muestra, lo cual afecta la calidad de los estimadores. Además, es importante considerar que así como existen estos cambios de

dominio en las empresas seleccionadas en la muestra, existen cambios en todo el marco muestral. Para solventar estos inconvenientes se utiliza un método denominado postestratificación.

El método de postestratificación 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 2016, 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.

Un postestrato está conformado por las empresas que tienen un mismo tamaño de empresa y sección económica de acuerdo al DIEE 2016, fragmentando los dominios originales de la muestra.

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

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

$fexp_{kp}$ = 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 al dominio de selección de la muestra original (DIEE 2015).
- Empresas que se convirtieron en grandes de acuerdo al DIEE 2016.

14. 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:

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

donde:

\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 .

15.El 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:

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

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

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

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

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

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

donde:

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

16.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.csplan'  
/PLANVARS ANALYSISWEIGHT=f_exp  
/PRINT PLAN  
/DESIGN STRATA=Pst  
/ESTIMATOR TYPE=EQUAL_WOR  
/POPSIZE VARIABLE=N_fin.
```

Donde la variable “Pst” corresponde al postestrato, la variable “fexp” corresponde al factor de expansión y la variable “N_fin” corresponde al 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
```

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.



www.ecuadorencifras.gob.ec



@ecuadorencifras



INEC/Ecuador



Inec



INECEcuador



INEC Ecuador