

INSTITUTO NACIONAL DE ESTADÍSTICAS
SUBDIRECCIÓN TÉCNICA
Departamento de Investigación y Desarrollo
Sección de Estadísticas Económicas



**METODOLOGÍA MUESTRAL
ENCUESTA LONGITUDINAL DE EMPRESAS
AÑO CONTABLE 2007**

Octubre, 2009



Departamento de Investigación y Desarrollo
METODOLOGÍA MUESTRAL
" ENCUESTA LONGITUDINAL DE EMPRESAS "

Jefe Departamento Investigación y Desarrollo
Charles Durán Artigas

Coordinadora Sección de Estadísticas Económicas
Grace Palma González

ÍNDICE

1.	Población Objetivo	3
2.	Marco Muestral.....	3
3.	Estratificación	3
4.	Tamaño de la Muestra.....	5
5.	Método de Selección	6
6.	Precisión Muestral.....	7
7.	Descripción de Marco Muestral y Tamaño Muestral Efectivo.....	9
8.	Factores de Expansión.....	12

Diseño Muestral

1. Población Objetivo

La población objetivo de éste estudio esta conformada por empresas formales que desarrollan actividad productiva comercial o de servicio, dentro de los límites territoriales del país y que facturaron más de 0,1 UF en el año 2007.

Para los efectos de éste estudio, se entendió como empresas formales a las personas naturales o jurídicas, cuyo inicio de actividades estuvo registrado por el Servicio de Impuestos Internos, SII y que tributen en primera categoría. Se excluyeron las actividades económicas realizadas por el Gobierno.

2. Marco Muestral

El marco muestral se construyó a partir del Directorio del Instituto Nacional de Estadísticas 2007, conformado con los directorios internos de levantamiento INE y registros del Servicio de Impuestos Internos, año 2007 (Declaración Anual de Impuestos a la Renta, formulario 22 y Declaración Mensual del IVA, formulario 29).

3. Estratificación

La primera estratificación fue realizada a nivel de clasificación económica, pensando en la necesidad de facilitar la comparabilidad de la información y siguiendo normas internacionales, se utilizó la Clasificación Industrial Internacional Uniforme de todas las actividades económicas, CIIU Rev. 3 a nivel de categoría de tabulación.

Cuadro 1. Clasificaciones económicas, según CIIU rev.3, consideradas para el estudio.

A	Agricultura, ganadería, caza y silvicultura.
C	Explotación de minas y canteras.
D	Industrias manufactureras.
E	Suministro de electricidad, gas y agua.
F	Construcción.
G	Comercio al por mayor y por menor; reparación de vehículos automotores, motocicletas, efectos personales y enseres domésticos.
H	Hoteles y restaurantes.
I	Transporte, almacenamiento y comunicaciones.
J	Intermediación financiera.
K	Actividades inmobiliarias, empresariales y de alquiler.
O	Otras actividades de servicios comunitarios, sociales y personales.

Departamento de Investigación y Desarrollo, INE
Sección de Estadísticas Económicas

La segunda estratificación se realizó a nivel de tamaño, considerando el nivel de ventas para segmentar las empresas. Esta segmentación por tamaño, incluye los siguientes 6 estratos:

Cuadro 2. Descripción de los intervalos de corte de ventas para definir tamaño.

Tamaño de ventas	Ventas anuales (UF 2007)	
	Limite Inferior	Limite Superior
Microempresa 1	0,1	800
Microempresa 2	800,1	2.400
Pequeñas empresa 1	2.400,1	5.000
Pequeñas empresa 2	5.000,1	25.000
Mediana empresa	25.000,1	100.000
Empresa grande	100.000,1	Más

Departamento de Investigación y Desarrollo, INE

Adicionalmente, una vez que el marco muestral esta estratificado por tamaño de acuerdo a las ventas anuales de las empresas, se realizó un análisis de estos estratos considerando la variable empleo generado por las empresas al interior de dichos estratos, con el fin de detectar y evaluar qué hacer con posibles casos anómalos, como por ejemplo, una empresa que esté clasificada como micro en

base a ventas, pero grande en base a empleo generado. En cualquier caso, la variable definida para la estratificación por tamaño fue ventas anuales 2007.

Como una forma de obtener mejor representación y disminuir el error de muestreo que proviene de la dispersión observada en las ventas anuales, se consideró, una vez ordenados los estratos por la variable de interés, integrar las unidades más grandes como Tramo a Censar (TC), y el resto como Tramo a Muestrear (TM).

El punto de corte para estos tramos, se determinó según el aporte de las ventas de las empresas a censar y el error de muestreo que resulta de restar al marco muestral las empresas que afectan notoriamente la precisión de las estimaciones. Estas empresas adquieren carácter censal o de Inclusión Forzosa (IF).

4. Tamaño de la Muestra

La estimación del tamaño muestral a nivel nacional, fue obtenida a partir de un muestreo probabilístico estratificado, según categoría de tabulación CIIU Rev.3, con un nivel de confianza del 95%, alcanzando a 10.213 empresas, que corresponde al 1,37% del total de las unidades del marco muestral. Para la determinación del tamaño muestral de cada estrato, se distribuyó el total de unidades muestrales en forma proporcional al número de empresas de cada uno de ellos. Se aplicaron los siguientes algoritmos:

Tamaño Muestral.

$$n = \frac{\left(\sum_{h=1}^H N_h \cdot S_h \right)^2 * Z^2}{e_r * \sum_{h=1}^H y_h + \left(\sum_{h=1}^H N_h \cdot S_h^2 \right) * Z^2}$$

Donde:

$$N = \sum_h N_h$$

n : Tamaño de la muestra.

Z : Estadístico normal estándar correspondiente a un nivel de confianza 95%.

e_r : Error relativo.

S_h : Desviación estándar en el estrato h .

N_h : Tamaño de la población en el estrato h .

Distribución del tamaño muestral en forma proporcional de acuerdo al nivel de ventas de cada estrato.

$$n_h = n * \frac{\sum_{i=1}^k y_{hi}}{\sum_{h=1}^H \sum_{i=1}^k y_{hi}}$$

Donde:

n_h : Tamaño de la muestra en el estrato h .

n : Tamaño de la muestra.

$\sum_{i=1}^k y_{hi}$: Sumatoria de las ventas de las empresas en el tramo aleatorio del estrato h .

$\sum_{h=1}^H \sum_{i=1}^k y_{hi}$: Sumatoria de las ventas de las empresas del tramo aleatorio de la clasificación económica.

El tamaño definitivo de las agrupaciones, estuvo sujeto a los costos versus precisión, por lo cual se debió aumentar o disminuir el número de empresas de algunos estratos para fortalecer su representación y precisión necesaria para el estudio.

5. Método de Selección

La selección de unidades a encuestar se realizó a través de dos mecanismos.

La primera selección se realizó para las empresas definidas como grandes en cada uno de los estratos, estas empresas son consideradas como inclusión forzosa debido a que son seleccionados en forma directa.

La segunda selección corresponde al tramo a muestrear, esta se realizó en forma independiente para cada sector y tamaño, aplicando una selección sistemática por ventas anuales y utilizando como variable auxiliar la ubicación geográfica de la empresa.

Para cubrir eventuales pérdidas de unidades muestrales, debido a quiebras, paralizaciones, cierre, cambio de giro, etc., se seleccionó un 30% de unidades para reemplazo. En el caso que este 30% de unidades de reemplazo seleccionadas no logró cubrir los requerimientos de tamaño total de la muestra, el INE seleccionó unidades adicionales a entrevistar. Todo esto, en los casos en que el universo de empresas contuvo estas unidades adicionales.

6. Precisión Muestral

Para determinar la precisión muestral respectiva, se utilizaron los siguientes algoritmos:

Coefficiente de variación del estrato h .

$$CV_h = \frac{\sqrt{V(\hat{Y})_h}}{\sum_{i=1}^k y_{hi}}$$

Donde:

CV_h : Coeficiente de variación del estrato h .

$V(\hat{Y})_h$: Varianza muestral estimada de las ventas del estrato h .

$\sum_{i=1}^k y_{hi}$: Sumatoria de las ventas del estrato h .

Varianza muestral de las ventas de las empresas

$$V(\hat{y}) = \sum_h^H N_h * (N_h - n_h) * \frac{S_h^2}{n_h}$$

Donde:

$V(\hat{y})$: Varianza muestral estimada de las ventas.

S_h^2 : Cuasi-varianza de las ventas de las empresas del estrato h .

N_h : Número de empresas en el marco muestral en el estrato h .

n_h : Tamaño muestral del estrato h .

Cuasi-varianza de las ventas de las empresas del estrato h .

$$S_h^2 = \frac{N_h}{(N_h - 1)} * \sigma_h^2$$

Donde:

S_h^2 : Cuasi-varianza que mide la dispersión de las ventas de las empresas de la muestra en el estrato h .

N_h : Número de empresas en el estrato h .

σ_h^2 : Varianza de las ventas del estrato h .

Varianza de las ventas del estrato h .

$$\sigma_h^2 = \frac{\sum_{i=1}^k (y_{hi} - \bar{y}_h)^2}{N_h}$$

Donde:

σ_h^2 : Varianza de las ventas del estrato h .

y_{hi} : Venta de la i -ésima empresa del estrato h .

\bar{y}_h : Media aritmética de las ventas en el estrato h .

N_h : Número de empresas en el estrato h .

7. Descripción de Marco Muestral y Tamaño Muestral Efectivo

A partir de la información sobre actividad económica, obtenida en el levantamiento, se procedió a la actualización de la muestra efectiva y marco muestral el que finalmente se emplea en las estimaciones del estudio.

Cuadro 4. Descripción del Marco Muestral, Tamaño Muestral efectivo, aporte en ventas de las unidades IF y coeficiente de variación, según categoría de tabulación.

Categoría de Tabulación	Descripción	Marco Muestral Nº	Muestra Efectiva			Coeficiente de Variación (%)
			Total muestra	Inclusión forzosa	Muestra Aleatoria	
				Nº	Nº	
Total nacional		744.186	10.213	323	9.890	1,25
A	Agricultura, ganadería, caza y silvicultura.	76.360	712	19	693	1,69
C	Explotación de minas y canteras.	2.897	453	21	432	1,23
D	Industrias manufactureras.	57.957	1.220	36	1184	3,57
E	Suministro de electricidad, gas y agua.	1.908	316	19	297	1,41
F	Construcción.	41.518	993	27	966	2,60
G	Comercio al por mayor y por menor; reparación de vehículos automotores, motocicletas, efectos personales y enseres domésticos.	281.706	2.279	45	2234	3,17
H	Hoteles y restaurantes.	33.624	983	17	966	1,36
I	Transporte, almacenamiento y comunicaciones.	53.986	1.123	26	1097	2,80
J	Intermediación financiera.	11.243	583	33	550	3,28
K	Actividades inmobiliarias, empresariales y de alquiler.	93.718	1.048	56	992	2,03
O	Otras actividades de servicios comunitarios, sociales y personales.	89.269	503	24	479	1,85

Departamento de Investigación y Desarrollo, INE
Sección de Estadísticas Económicas

Cuadro 5. Descripción del Marco Muestral, Tamaño Muestral efectivo, aporte en ventas de las unidades IF y coeficiente de variación, según Tamaño de ventas.

Tamaños	Marco Muestral Nº	Muestra Efectiva				Coeficiente de Variación (%)
		Total muestra	Inclusión forzosa	Muestra Aleatoria		
			Nº	Nº		
Total Nacional	744.186	10.213	323	9.890	1,25	
100.000,1 - Más	8.013	1.872	323	0	1549	1,52
25.000,1 - 100.000	15.971	1.371		0	1371	1,25
5.000,1 - 25.000	57.696	1.931		0	1931	1,21
2.400,1 - 5.000	52.565	1.236		0	1236	0,67
800,1 - 2.400	116.000	1.378		0	1378	0,93
0,1 - 800	493.941	2.425		0	2425	2,44

Departamento de Investigación y Desarrollo, INE
Sección de Estadísticas Económicas

Cuadro 6. Descripción del Marco Muestral, Tamaño Muestral efectivo, aporte en ventas de las unidades IF y coeficiente de variación, según Categoría de tabulación y Tamaño.

Categoría de tabulación	Tamaños	Marco Muestral Nº	Muestra Efectiva			Coeficiente de Variación (%)
			Total muestra	Inclusión forzosa	Muestra Aleatoria	
				Nº	Nº	
A	100.000,1 - Más	303	60	19	41	2,70
	25.000,1 - 100.000	1.173	125	0	125	3,40
	5.000,1 - 25.000	7.229	138	0	138	3,84
	2.400,1 - 5.000	6.429	79	0	79	2,38
	800,1 - 2.400	13.184	98	0	98	3,20
	0,1 - 800	48.042	212	0	212	6,16
	Total	76.360	712	19	693	1,69
C	100.000,1 - Más	128	53	21	32	1,25
	25.000,1 - 100.000	171	65	0	65	3,93
	5.000,1 - 25.000	557	103	0	103	4,18
	2.400,1 - 5.000	388	54	0	54	2,57
	800,1 - 2.400	590	67	0	67	3,54
	0,1 - 800	1.063	111	0	111	7,22
	Total	2.897	453	21	432	1,23
D	100.000,1 - Más	1.679	296	36	260	4,03
	25.000,1 - 100.000	2.539	165	0	165	3,03
	5.000,1 - 25.000	7.624	221	0	221	3,13
	2.400,1 - 5.000	5.566	121	0	121	1,89
	800,1 - 2.400	10.584	119	0	119	2,85
	0,1 - 800	29.965	298	0	298	5,25
	Total	57.957	1.220	36	1184	3,57

Cuadro 6 (continuación). Descripción del Marco Muestral, Tamaño Muestral efectivo, aporte en ventas de las unidades IF y coeficiente de variación, según Categoría de tabulación y Tamaño.

Categoría de tabulación	Tamaños	Marco Muestral Nº	Muestra Efectiva			Coeficiente de Variación (%)
			Total muestra	Inclusión forzosa	Muestra Aleatoria	
				Nº	Nº	
E	100.000,1 - Más	112	89	19	70	1,42
	25.000,1 - 100.000	37	20	0	20	5,01
	5.000,1 - 25.000	109	46	0	46	5,32
	2.400,1 - 5.000	137	35	0	35	2,93
	800,1 - 2.400	386	53	0	53	4,05
	0,1 - 800	1.127	73	0	73	7,91
	Total	1.908	316	19	297	1,41
F	100.000,1 - Más	828	259	27	232	3,75
	25.000,1 - 100.000	1.767	142	0	142	3,34
	5.000,1 - 25.000	5.875	178	0	178	3,53
	2.400,1 - 5.000	4.856	118	0	118	1,93
	800,1 - 2.400	9.006	114	0	114	2,92
	0,1 - 800	19.186	182	0	182	6,07
	Total	41.518	993	27	966	2,60
G	100.000,1 - Más	2.821	496	45	451	4,07
	25.000,1 - 100.000	5.530	228	0	228	2,65
	5.000,1 - 25.000	18.317	390	0	390	2,42
	2.400,1 - 5.000	17.157	274	0	274	1,28
	800,1 - 2.400	43.126	341	0	341	1,72
	0,1 - 800	194.755	550	0	550	4,17
	Total	281.706	2.279	45	2234	3,17
H	100.000,1 - Más	130	68	17	50	2,41
	25.000,1 - 100.000	454	151	0	152	2,65
	5.000,1 - 25.000	2.335	235	0	235	3,00
	2.400,1 - 5.000	2.400	141	0	141	1,74
	800,1 - 2.400	5.892	129	0	129	2,79
	0,1 - 800	22.413	259	0	259	5,63
	Total	33.624	983	17	966	1,36
I	100.000,1 - Más	636	186	26	155	3,50
	25.000,1 - 100.000	1.356	149	0	147	3,18
	5.000,1 - 25.000	4.954	182	0	182	3,59
	2.400,1 - 5.000	5.536	117	0	117	1,95
	800,1 - 2.400	11.662	167	0	167	2,42
	0,1 - 800	29.842	332	0	329	4,60
	Total	53.986	1.133	26	1097	2,80

Departamento de Investigación y Desarrollo, INE
Sección de Estadísticas Económicas

Cuadro 6 (continuación). Descripción del Marco Muestral, Tamaño Muestral efectivo, aporte en ventas de las unidades IF y coeficiente de variación, según Categoría de tabulación y Tamaño.

Categoría de tabulación	Tamaños	Marco Muestral Nº	Muestra Efectiva			Coeficiente de Variación (%)
			Total muestra	Inclusión forzosa	Muestra Aleatoria	
				Nº	Nº	
J	100.000,1 - Más	384	117	33	83	3,46
	25.000,1 - 100.000	467	85	0	86	3,86
	5.000,1 - 25.000	1.542	115	0	115	4,30
	2.400,1 - 5.000	1.467	79	0	79	2,31
	800,1 - 2.400	2.601	73	0	72	3,61
	0,1 - 800	4.782	115	0	115	7,81
	Total	11.243	584	33	550	3,28
K	100.000,1 - Más	895	197	56	141	3,22
	25.000,1 - 100.000	2.254	160	0	162	3,11
	5.000,1 - 25.000	7.754	207	0	207	3,34
	2.400,1 - 5.000	6.974	147	0	147	1,73
	800,1 - 2.400	13.590	152	0	152	2,55
	0,1 - 800	62.251	186	0	183	7,96
	Total	93.718	1.049	56	992	2,03
O	100.000,1 - Más	97	57	24	34	1,43
	25.000,1 - 100.000	223	81	0	79	3,76
	5.000,1 - 25.000	1.400	116	0	116	4,25
	2.400,1 - 5.000	1.655	71	0	71	2,40
	800,1 - 2.400	5.379	66	0	66	3,94
	0,1 - 800	80.515	114	0	113	11,39
	Total	89.269	505	24	479	1,85
Total Nacional		744.186	10.213	323	9.890	1,25

Departamento de Investigación y Desarrollo, INE
Sección de Estadísticas Económicas

8. Factores de Expansión

Como se ha indicado, la información recolectada se hace a través de una muestra de empresas, por lo tanto, para obtener una visión completa de las actividades que componen los sectores se realiza una expansión de los datos estimados a partir de la información levantada. Para realizar la mencionada expansión, se considera en primer lugar las empresas de inclusión forzosa que son unidades muestrales auto representadas, ya que tienen carácter censal, en segundo lugar se encuentran las empresas seleccionadas en forma aleatoria.

8.1. Factor de Expansión por ventas anuales

El factor de expansión calculado a partir de las ventas anuales, se emplea para obtener una mejor estimación de las variables cuantitativas del estudio. Para las unidades de inclusión forzosa en casos de existir no respuesta, se aplica el siguiente algoritmo:

Factor de expansión de las empresas de inclusión forzosa (IF).

$$FE(IF)_h = \frac{\sum_{i=1}^K y(IF)_{h,i}}{\sum_{i=1}^K y(IFe)_{h,i}}$$

Donde:

$FEV(IF)_h$: Factor de expansión de ventas de las empresas IF en el estrato h .

$\sum_{i=1}^K y(IF)_{h,i}$: Sumatoria de las ventas del marco muestral correspondiente a las empresas IF del estrato h .

$\sum_{i=1}^k y(IFe)_{h,i}$: Sumatoria de las ventas del marco muestral que corresponden a las empresas IF efectivamente logradas, del estrato h .

Recuérdese que para homogenizar la variable en estudio se realizó una segunda estratificación en el tramo aleatorio, de acuerdo al tamaño según el nivel de ventas que declararon las empresas. Esto cobra relevancia en el momento de realizar la expansión, debido a que debe efectuarse con la misma desagregación, así se obtendrá una mejor estimación de las ventas totales, pues las empresas de distintos tamaños tendrán su propio tratamiento.

La expresión algebraica del factor de expansión de los establecimientos pertenecientes al Tramo Aleatorio (TA), tiene la siguiente forma:

Factor de expansión de las empresas de inclusión aleatoria (TA).

$$FEV(TA)_h = \frac{\sum_{i=1}^K y(TA)_{h,i}}{\sum_{i=1}^K y(TAe)_{h,i}}$$

Donde:

$FEV(TA)_h$: Factor de expansión de ventas del tramo aleatorio del estrato h .

$\sum_{i=1}^K y(TA)_{h,i}$: Sumatoria de las ventas del marco muestral correspondiente a las empresas del TA del estrato h .

$\sum_{i=1}^k y(TAe)_{h,i}$: Sumatoria de las ventas del marco muestral que corresponden a las empresas efectivamente logradas en el TA del estrato h .

8.2. Factor de Expansión por número de empresas

En esta oportunidad para obtener una mejor estimación de las variables cualitativas del estudio, se obtiene un factor de expansión a partir del número de empresas, los que son calculados empleando la cantidad de ellas en el marco muestral y muestra, a nivel de estratos por tipo de inclusión, clasificación económica y tamaño. Para las unidades de inclusión forzosa en caso de que exista no respuesta, se aplica el siguiente algoritmo:

Factor de expansión de las empresas de inclusión forzosa (IF).

$$FEN(IF)_h = \frac{\sum_{i=1}^K y(IF)_{h,i}}{\sum_{i=1}^K y(IFe)_{h,i}}$$

Donde:

$FEN(IF)_h$: Factor de expansión por número de empresas de las unidades IF en el estrato h .

$\sum_{i=1}^K y(IF)_{h,i}$: Número de empresas del marco muestral, correspondiente a las empresas IF, del estrato h .

$\sum_{i=1}^k y(IFe)_{h,i}$: Número de empresas del marco muestral, corresponden a las empresas IF efectivamente logradas, del estrato h .

La expresión algebraica del factor de expansión de las empresas pertenecientes al Tramo Aleatorio (TA), tiene la siguiente forma:

Factor de expansión de las empresas de inclusión aleatoria (TA).

$$FEN(TA)_h = \frac{\sum_{i=1}^K y(TA)_{h,i}}{\sum_{i=1}^k y(TAe)_{h,i}}$$

Donde:

$FEN(TA)_h$: Factor de expansión de número de empresas del tramo aleatorio del estrato h .

$\sum_{i=1}^K y(TA)_{h,i}$: Número de empresas del marco muestral correspondiente a las empresas del TA del estrato h .

$\sum_{i=1}^k y(TAe)_{h,i}$: Número de empresas efectivamente logradas en el TA del estrato h .

9. Niveles de Estimación

De acuerdo a los objetivos del estudio, la validez estadística de las variables a estimar tiene representatividad para los siguientes niveles de estimación:

- Nacional según actividad económica.
- Nacional según tamaño.
- Nacional según actividad económica y tamaño.