



INFORME METODOLÓGICO MUESTRA EFECTIVA

VII Encuesta de Innovación

INSTITUTO NACIONAL DE ESTADÍSTICAS

Febrero / 2012

INFORME METODOLÓGICO MUESTRA EFECTIVA
Departamento de Investigación y Desarrollo

VII ENCUESTA DE INNOVACIÓN

Instituto Nacional de Estadísticas.

Febrero / 2012.

Jefe Depto. Investigación y Desarrollo: Charles Duran A.

Coordinadora Sección Estadísticas Económicas: Carolina Morales L.

Analista Investigador: Joselin Jeldes P.

ÍNDICE

1. Introducción.....	3
2. Marco Conceptual.....	3
2.1. Objetivo.....	3
2.2. Clasificación de Actividades.....	3
2.3. Cobertura.....	3
2.3.1. Cobertura geográfica.....	3
2.3.2. Cobertura temática.....	4
2.3.3. Unidad estadística y de información.....	4
3. Diseño Muestral.....	4
3.1. Población Objetivo.....	4
3.2. Marco Muestral.....	4
3.3. Estratificación.....	5
3.4. Unidad de Muestreo e Información.....	7
4. Estimación Muestral.....	7
4.1. Tamaño Muestral.....	7
4.2. Precisión Muestral.....	8
4.3. Método de Selección.....	13
5. Factores de Expansión.....	13
5.1. Factores de Expansión Ventas.....	13
5.2. Factores de Expansión Número de Empresas.....	14
6. Niveles de Estimación.....	15

1. Introducción

Obtener estadísticas oficiales y de calidad es fundamental para la generación de políticas públicas informadas y eficientes para nuestro país. Con esta finalidad es que el Ministerio de Economía en conjunto con el Instituto Nacional de Estadísticas desarrolla la Séptima Encuesta de Innovación, la cual tiene como objetivo obtener información cuantitativa y cualitativa sobre Innovación en las empresas, de acuerdo a los lineamientos entregados por la OCDE, mediante el manual de Oslo. De esta manera, se logra fortalecer la información existente para el Sistema Nacional de Innovación.

Los resultados de esta encuesta, permiten efectuar análisis de la evolución de las variables que describen los niveles de innovación en las empresas, construir indicadores internacionales y compilados nacionales de las variables.

El presente documento explica la Metodología Muestral Efectiva utilizada en la Séptima Encuesta de Innovación, año de referencia 2009-2010.

2. Marco Conceptual

2.1. Objetivo

Describir la metodología muestral efectiva utilizada en la Séptima Encuesta de Innovación. La cual, detalla desde los sectores involucrados hasta los niveles de estimación necesarios para generar estadísticas oficiales y representativas.

2.2. Clasificación de Actividades

Para facilitar la comparación internacional y de la información obtenida con respecto al tiempo, se utiliza la Clasificación Industrial Uniforme de todas las actividades económicas, tercera Revisión (CIIU Rev. 3.), a nivel de categoría de tabulación.

2.3. Cobertura

2.3.1. Cobertura geográfica

El presente estudio tiene una cobertura geográfica a nivel nacional, es decir, la información es recopilada a través de una muestra representativa en todo el territorio nacional.

2.3.2. Cobertura temática

La información obtenida durante el período de referencia, se selecciona a través del Directorio INE, año contable 2009, la cual incluye los contribuyentes situados dentro de los límites territoriales del país.

El presente estudio considera los contribuyentes que pertenecen a los sectores siguientes sectores económicos: de Agricultura, Ganadería, Caza y Silvicultura (A); Pesca (B); Explotación de minas y canteras (C); Industria Manufacturera (D); Suministro de Electricidad, Gas y Agua (E); Construcción (F); Comercio al por mayor y al por menor; Reparación de vehículos automotores, motores, motocicletas, efectos personales y enseres domésticos (G); Hoteles y restaurantes (H); Transporte, almacenamiento y comunicaciones (I); Intermediación financiera (J); Actividades inmobiliarias, empresariales y de alquiler (K); Servicios sociales y de salud (N); Otras actividades de servicios comunitarias, sociales y personales (O).

2.3.3. Unidad estadística y de información

La unidad estadística es la empresa, organización que tiene iniciación de actividades independiente, es decir, un RUT y contabilidad propia, y cuyo giro lo puede realizar en uno o más establecimientos, que desarrollan alguna actividad principal señaladas en la cobertura temática.

3. Diseño Muestral

3.1. Población Objetivo

La población objetivo se compone de las empresas naturales o jurídicas, que desarrollen su actividad dentro de los límites territoriales del país, que cuenten con declaración en el Servicio de Impuestos Internos en el año 2009 y con un nivel de ventas anuales superiores a UF 2.400.

3.2. Marco Muestral

El marco muestral se construye a partir del Directorio INE, año contable 2009, conformado con los registros del Servicio de Impuestos Internos (SII) y directorios internos de levantamiento INE.

Para algunos sectores se utilizan bases internas INE, año contable 2009. Las cuales, se describen a continuación:

Industria Manufacturera

Directorio de Encuesta Nacional Industrial Anual, año 2010 (ENIA 2010), con período de referencia año 2009. El cual se conforma de empresas de 10 y más trabajadores, desagregado por establecimientos. Para la Séptima Encuesta de Innovación, el directorio se construye a nivel de empresas, con venta anual superior a UF 2.400.

Explotación de minas y canteras

Directorio del Índice de Minería, INE año 2010, con período de referencia año 2009. Corresponde a un censo de empresas, de la mediana y gran minería consideradas como tales por el INE y SERNAGEOMIN.

Suministro de Electricidad, Gas y Agua

Directorio del Índice de Electricidad, Gas y Agua (EGA), INE año 2010, con período de referencia año 2009. Corresponde a un censo de empresas que realizan generación y distribución de electricidad, gas y agua. Las empresas Generadoras (productoras y autoproductoras) deben contar con una producción de más de 2 M.W.H. a nivel nacional.

3.3. Estratificación

De acuerdo a los objetivos del estudio se realiza una estratificación del Sector de Industrias Manufactureras, según CIIU Rev.3 (2 dígitos), del cual resultan ocho estratos determinados por el Ministerio de Economía Fomento y Turismo (MINECON), de acuerdo a su valor agregado y tasa de innovación presentada en la sexta encuesta, estos estratos son presentados en el cuadro siguiente:

Cuadro 1. Actividades consideradas en la Industria Manufacturera, CIIU Rev.3

Categoría	División	Descripción
D		Industria Manufacturera
	15	Elaboración de productos alimenticios y bebidas.
	20	Producción de madera y fabricación de productos de madera y de corcho, exoptos muebles, fabricación de artículos de paja y de materiales trenzables.
	21	Fabricación de papel y productos de papel.
	24	Fabricación de sustancias y productos químicos.
	27	Fabricación de metales comunes.
	28	Fabricación de productos elaborados de metal excepto maquinaria y equipo.
	31	Fabricación de maquinaria y aparatos eléctricos N.C.P.
	99 ¹	Resto de Industria Manufacturera.

1/ Esta división contiene todas las actividades de Industria manufacturera no consideradas en la estratificación

La estratificación para el resto de los sectores económicos considerados es a nivel de categoría de tabulación de la CIIU Rev. 3, de acuerdo a los requerimientos del estudio. Los cuales se exponen a continuación:

Cuadro 2: Estratificación del resto de sectores económicos considerados.

Categoría	Descripción
A	Agricultura, Ganadería, Caza y Silvicultura
B	Pesca
C	Explotación de Minas y Canteras
E	Suministro de Electricidad, Gas y Agua
F	Construcción
G	Comercio al por mayor y al 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 y de alquiler
N	Actividades de servicios sociales y de salud
O	Otras actividades de servicios comunitarios, sociales y personales del tipo servicio

Se realiza una segunda estratificación por tamaño de venta al interior de cada sector económico determinado.

Cuadro 3. Intervalos de tamaño de acuerdo al valor de venta

Tamaño Empresa	Ventas anuales (UF)	
	Límite Inferior	Límite Superior
Grande	100.000,10	Más
Mediana	25.000,10	100.000
Pequeña	2.400,10	25.000

Además el presente estudio considera la ubicación geográfica de las empresas, la cual permite obtener estimaciones regionales. Adicionalmente, para fortalecer estas estimaciones, se emplea una pregunta dentro del cuestionario que permite desagregar los gastos realizados por la empresa en actividades innovativas a nivel regional.

Una vez estratificado el marco muestral, se incluyen las unidades muestrales con mayores ventas anuales como "tramo a censar" (TC), y el resto se define como "tramo a muestrear" (TM).

Para estos dos tramos se determina el punto de corte a partir del aporte de las ventas en cada unidad, considerando para ello el 2,0% respecto del total acumulado y un segundo corte, según el error de muestreo resultante y el número de unidades por estrato. Estas empresas adquieren carácter censal o inclusión forzosa (IF).

3.4. Unidad de Muestreo e Información

La unidad de muestreo y de información es la empresa que cuentan con iniciación de actividades dentro de los límites territoriales del país y declaradas en el SII en el año 2009, pertenecientes a los sectores económicos antes mencionados.

4. Estimación Muestral

4.1. Tamaño Muestral

Debido a la atomización existente en los sectores de Explotación de Minas y Canteras y Suministro de Electricidad, Gas y Agua, fueron consideradas todas las empresas como censo.

Para el resto de los sectores económicos de éste estudio, se calculan los tamaños muestrales, provenientes de los marcos muestrales de Industria Manufacturera y Otros Sectores. El diseño muestral utilizado, consiste en aplicar un muestreo probabilístico estratificado (actividad-tamaño-región), con un nivel de confianza del 95%, distribuyendo la muestra en forma proporcional al volumen de ventas de cada uno de los estratos.

Tamaño Muestral

$$n = \frac{\left(Z_{\alpha/2}^2 N Q \right)}{\left((N - 1) \varepsilon_r^2 P \right) + \left(Z_{\alpha/2}^2 (1 - P) \right)}$$

donde,

n : Tamaño de la muestra.

$Z_{\alpha/2}$: Percentil del modelo normal estándar correspondiente a un nivel de confianza del 95%.

ε_r : Error relativo.

p : Valor 0,5, asumiendo varianza máxima.

N : Tamaño de la población.

La distribución del tamaño muestral se realiza en forma proporcional de acuerdo al nivel de ventas de cada estrato.

$$n_h = n \left(\frac{\sum_{i=1}^k y_{hi}}{\sum_{h=1}^H \sum_{i=1}^k y_{hi}} \right)$$

donde,

n_h : Tamaño de la muestra del estrato h .

n : Tamaño de la muestra.

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

$\sum_{h=1}^H \sum_{i=1}^k y_{hi}$: Suma total de las ventas a través de todos los estratos

4.2. Precisión Muestral

El tamaño de muestra definitivo para cada estrato está sujeto a los costos versus precisión, por lo cual se procede a aumentar o disminuir el número de empresas de algunos estratos para fortalecer su representatividad y precisión necesaria para el estudio.

Para determinar la precisión muestral, se calculan los diferentes coeficientes de variación, utilizando el siguiente algoritmo:

Coeficiente de variación

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

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.

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

N_h : Número de empresas en el estrato h .

n_h : Tamaño de la muestra 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 del estrato h .

N_h : Número de empresas en el estrato h .

σ_h^2 : Varianza de las ventas en el 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 ventas en el estrato h .

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

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

N_h : Número de empresas en el estrato h .

De esta forma, el tamaño total de la muestra efectiva, es de 3.653 unidades, con un coeficiente de variación de 2,46%.

Cuadro 4. Estimación muestral según categoría de tabulación (CIIU rev3)

Categoría de Tabulación CIIU Rev 3.	Marco Muestral	Muestra Efectiva			
		Total Muestra	Inclusión Forzosa	Inclusión Aleatoria	Coefficiente de Variación (%)
Total	126.819	3.653	1.218	2.435	2,46
A	15.205	198	46	152	4,41
B	1.011	121	61	60	1,26
C	56	56	56	-	-
D	3.671	905	204	701	4,15
E	118	118	118	-	-
F	12.789	302	104	198	5,80
G	44.514	492	180	312	5,95
H	5.449	185	52	133	4,11
I	12.807	354	92	262	5,40
J	3.884	194	98	96	6,15
K	19.299	406	95	311	5,18
N	4.611	189	63	126	4,57
O	3.405	133	49	84	5,70

Cuadro 5. Tamaños muestrales según tamaño de ventas anuales en UF.

Tamaño	Marco Muestral	Muestra Efectiva			
		Total Muestra	Inclusión Forzosa	Inclusión Aleatoria	Coefficiente de Variación (%)
Total	126.819	3.653	1.218	2.435	2,46
Empresa Grande	7.217	1.724	1.029	695	3,01
Empresa Mediana	14.661	776	152	624	3,28
Empresa Pequeña	104.941	1.153	37	1.116	3,72

Cuadro 6. Tamaños muestrales según sector económico y tamaño de ventas.

Categoría de Tabulación CIU Rev 3.	Tamaño	Marco Muestral	Muestra Efectiva			
			Total Muestra	Inclusión Forzosa	Inclusión Aleatoria	Coefficiente de Variación (%)
Total		126.819	3.653	1.218	2.435	2,46
A	Total	15.205	198	46	152	4,41
	Grande	300	68	41	27	6,16
	Mediana	1.161	51	5	46	6,92
	Pequeña	13.744	79	-	79	8,83
B	Total	1.011	121	61	60	1,26
	Grande	54	50	50	-	-
	Mediana	109	29	11	18	8,76
	Pequeña	848	42	-	42	12,37
C	Total	56	56	56	-	-
	Grande	56	56	56	-	-
D 15	Total	1.116	212	39	173	6,24
	Grande	335	115	27	88	6,60
	Mediana	261	41	11	30	7,13
	Pequeña	520	56	1	55	5,65
D 20	Total	193	81	19	62	2,07
	Grande	44	29	12	17	2,21
	Mediana	74	25	6	19	6,95
	Pequeña	75	27	1	26	9,40
D 21	Total	123	56	19	37	2,89
	Grande	49	24	16	8	2,95
	Mediana	43	17	1	16	7,43
	Pequeña	31	15	2	13	8,68
D 24	Total	204	88	27	61	5,59
	Grande	120	56	18	38	5,72
	Mediana	51	15	2	13	8,91
	Pequeña	33	17	7	10	7,51
D 27	Total	55	27	16	11	2,65
	Grande	40	17	12	5	2,87
	Mediana	13	8	2	6	10,33
	Pequeña	2	2	2	-	-
D 28	Total	366	118	43	75	7,93
	Grande	87	37	21	16	5,18
	Mediana	117	33	14	19	9,06
	Pequeña	162	48	8	40	7,15
D 31	Total	48	31	14	17	2,59
	Grande	8	8	8	-	-
	Mediana	23	13	2	11	7,06
	Pequeña	17	10	4	6	8,14

Cuadro 6. Tamaños muestrales según sector económico y tamaño de ventas.

Categoría de Tabulación CIU Rev 3.	Tamaño	Marco Muestral	Muestra Efectiva			
			Total Muestra	Inclusión Forzosa	Inclusión Aleatoria	Coefficiente de Variación (%)
D 99	Total	1.566	292	27	265	7,63
	Grande	356	119	18	101	4,26
	Mediana	507	59	5	54	6,63
	Pequeña	703	114	4	110	6,45
E	Total	118	118	118	-	-
	Grande	118	118	118	-	-
F	Total	12.789	302	104	198	5,80
	Grande	758	172	104	68	7,48
	Mediana	1.568	49	-	49	9,18
	Pequeña	10.463	81	-	81	13,71
G	Total	44.514	492	180	312	5,95
	Grande	2.622	308	177	131	7,29
	Mediana	5.529	71	3	68	7,41
	Pequeña	36.363	113	-	113	8,18
H	Total	5.449	185	52	133	4,11
	Grande	131	47	34	13	5,48
	Mediana	492	50	18	32	7,89
	Pequeña	4.826	88	-	88	8,89
I	Total	12.807	354	92	262	5,40
	Grande	621	152	86	66	6,43
	Mediana	1.378	84	6	78	5,93
	Pequeña	10.808	118	-	118	8,40
J	Total	3.884	194	98	96	6,15
	Grande	346	84	67	17	6,48
	Mediana	434	47	27	20	9,23
	Pequeña	3.104	63	4	59	10,84
K	Total	19.299	406	95	311	5,18
	Grande	929	165	92	73	7,64
	Mediana	2.342	97	3	94	6,44
	Pequeña	16.028	144	-	144	7,38
N	Total	4.611	189	63	126	4,57
	Grande	132	59	44	15	6,29
	Mediana	308	50	19	31	7,67
	Pequeña	4.171	80	-	80	9,02
O	Total	3.405	133	49	84	5,70
	Grande	111	40	28	12	8,14
	Mediana	251	37	17	20	9,46
	Pequeña	3.043	56	4	52	11,03

Cuadro 7. Tamaños muestrales según región.

Región	Marco Muestral	Muestra Efectiva			Coeficiente de Variación (%)
		Total Muestra	Inclusión Forzosa	Inclusión Aleatoria	
Total	126.819	3.653	1.218	2.435	2,46
Arica y Parinacota	1.246	112	51	61	5,18
Tarapacá	2.237	137	62	75	5,38
Antofagasta	3.812	152	68	84	4,23
Atacama	1.507	119	56	63	4,88
Coquimbo	3.707	142	46	96	5,78
Valparaíso	10.889	241	68	173	3,89
Libertador General Bernardo O' Higgins	6.002	199	73	126	2,88
Del Maule	6.802	191	55	136	5,05
Biobío	11.727	298	87	211	3,78
Araucanía	5.071	186	44	142	5,54
Los Ríos	2.242	135	47	88	4,05
Los Lagos	6.249	197	67	130	8,96
Aysén del General Carlos Ibañez del Campo	771	76	32	44	6,66
Magallanes y Antártica Chilena	1.527	137	43	94	3,54
Metropolitana	63.030	1.331	419	912	3,27

4.3. Método de Selección

La primera selección de unidades muestrales se realiza para aquellas que pertenecen al tramo censado, las cuales se incluyen directamente en la muestra.

La segunda selección corresponde al tramo aleatorio, la cual se realiza en forma independiente para cada estrato, (actividad, tamaño y región), aplicando una selección sistemática, con el objetivo de asegurar un recorrido completo dentro de cada estrato.

Se considera un 30% de selección adicional de empresas para eventuales reemplazos, como una forma de cubrir los posibles cambios de actividad, situaciones de quiebras, paralizaciones, cierres de empresas y de otro tipo, que pudieran presentar algunas unidades seleccionadas, entre la fecha de referencia del marco muestral y la fecha del levantamiento.

5. Factores de Expansión

5.1. Factores de Expansión Ventas

Una vez levantada la encuesta, se procede a realizar el cálculo de los factores de expansión respectivos, los que son calculados utilizando la variable ventas en el marco y la muestra, a nivel de estratos por tipo de inclusión y región. El factor de expansión calculado a partir de las ventas, se emplea para obtener una mejor estimación de las variables cuantitativas del estudio.

Para las unidades de inclusión forzosa en el caso que alguna empresa no conteste, se aplica el siguiente algoritmo:

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

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

donde,

$FEV(IF)_h$: Factor de expansión según ventas de las unidades de *IF* del estrato *h*.

$\sum_{i=1}^{n_k} y(IF)_{h,i}$: Suma de las ventas de las unidades de *IF* del estrato *h*.

$\sum_{i=1}^{n_k} y(IFe)_{h,i}$: Suma de las ventas de las unidades de *IF* efectivamente logradas del estrato *h*.

La expresión algebraica del factor de expansión de las unidades muestrales pertenecientes a la Inclusión Aleatoria (IA), es la siguiente:

$$FEV(IA)_h = \frac{\sum_{i=1}^{n_k} y(IA)_{h,i}}{\sum_{i=1}^{n_k} y(IAe)_{h,i}}$$

Donde,

$FEV(IA)_h$: Factor de expansión según ventas de las unidades de *IA* del estrato *h*.

$\sum_{i=1}^{n_k} y(IA)_{h,i}$: Suma de las ventas de las unidades de *IA* del estrato *h*.

$\sum_{i=1}^{n_k} y(IAe)_{h,i}$: Suma de las ventas de las unidades de *IA* efectivamente logradas del estrato *h*.

5.2. Factores de Expansión Número de Empresas

En esta oportunidad para obtener una mejor estimación de las variables cualitativas del estudio, se calculan los factores de expansión a partir del número de empresas, los que son calculados empleando la cantidad de ellas en el marco y muestra, a nivel de estratos por tipo de inclusión y región.

Para las unidades de inclusión forzosa en el caso que alguna empresa no conteste, se aplica el siguiente algoritmo:

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

$$FEN(IF)_h = \frac{N(IF)_h}{N(IFe)_h}$$

donde,

$FEN(IF)_h$: Factor de expansión según número de empresas de IF del estrato h .

$N(IF)_h$: Número de empresas de IF del estrato h .

$N(IFe)_h$: Número de empresas de IF efectivamente logradas del estrato h .

La expresión algebraica del factor de expansión de las unidades muestrales pertenecientes a la inclusión aleatoria (IA), es la siguiente:

$$FEN(IA)_h = \frac{N(IA)_h}{N(IAe)_h}$$

donde,

$FEN(IA)_h$: Factor de expansión según número de empresas de IA del estrato h .

$N(IA)_h$: Número de empresas de IA del estrato h .

$N(IAe)_h$: Número de empresas de IA efectivamente logradas del estrato h .

6. Niveles de Estimación

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

- Representatividad a nivel de sector económico.
- Representatividad a nivel de tamaño de ventas.
- Representatividad a nivel de sector económico y tamaño de ventas.
- Representatividad a nivel regional.