



DISEÑO MUESTRAL
Encuesta Anual de Actividad Económica 2008

Diseño muestral
Encuesta Anual de Actividad Económica 2008

Instituto Nacional de Estadística

Ec. Laura Nalbarte

Directora Técnica

Cr. Manuel Rodríguez

Sub-Director General

División Estadísticas Económicas

Álvaro Fuentes Coiana

Director División Estadísticas Económicas

Marina Fantin

Sub-Directora División Estadísticas Económicas

Graziela Basañez

Jefe de Departamento

Demetrio Sastre

Jefe de Sección

Críticos Analistas

María Moratorio

Jorge Abud

Alina Bedat

Cesar Medero

Diego Fleitas

Mauricio Zunino

Deborah Zak

Diego Garcia

Desarrollo de Diseño Muestral

Juan Pablo Ferreira

Tabla de contenido¹

1. Antecedentes.....	7
2. Universo a investigar.....	7
3. Marco muestral.....	7
4. Diseño muestral.....	8
5. Relevamiento y no respuesta.....	12
6. Cálculo de los ponderadores.....	13
7. Estimadores.....	17
8. Precisión de las estimaciones.....	18
Bibliografía.....	21
Anexo 1.....	23
Anexo 2.....	25

¹ El presente documento fue elaborado por Juan Pablo Ferreira – División Estadísticas Económicas.
Instituto Nacional de Estadística – División Estadísticas Económicas

1. Antecedentes

El antecedente más reciente es la encuesta anual del año 2007. El diseño muestral implementado fue un diseño estratificado aleatorio, en donde los estratos reconocen dos dimensiones, la primera es la clase de actividad económica principal de la empresa y la segunda una medida de tamaño (Personal Ocupado Promedio y/o monto de Ventas en el período).

Dentro de cada estrato se seleccionó una muestra independiente bajo un muestreo sistemático con arranque aleatorio, ordenando las empresas por su Personal Ocupado promedio.

En el diseño del 2007 la variable auxiliar utilizada para calcular los tamaños de muestra (para definir precisiones) fue el Monto Imponible (Remuneraciones), proveniente de Banco de Previsión Social (*BPS*). Lo anterior fue debido a la ausencia de datos provenientes de años anteriores de las variables de interés. Los tamaños de muestra por rama de actividad se calcularon para tener un error relativo de menos de 15% para estimar el total de las remuneraciones.

2. Universo a investigar

El universo a investigar lo compone el conjunto de todas las empresas del país (sector formal), con personal ocupado promedio en el año mayor o igual a diez (según *BPS*), o cuyas ventas para el año de referencia superan los 120 millones de pesos (según datos de la *DGI*) independiente de su personal ocupado, y su clase de actividad económica principal según la clasificación industrial CIIU Rev.4 es:

- B. Explotación de minas y canteras
- C. Industrias Manufactureras
- D. Suministro de electricidad, gas, vapor y aire acondicionado
- E. Suministro de agua; alcantarillado, gestión de desechos y actividades de saneamiento
- G. Comercio al por mayor y al por menor; reparación de los vehículos de motor y de las motocicletas
- H. Transporte y almacenamiento
- I. Alojamiento y servicio de comida
- J. Información y comunicación
- M. Actividades profesionales, científicas y técnicas
- N. Actividades administrativas y servicios de apoyo
- P. Enseñanza
- Q. Servicios sociales y relacionados con la Salud humana.
- R. Artes, entretenimiento y recreación.

3. Marco muestral

El marco muestral, esta compuesto por las unidades provenientes de los registros administrativos de la *Dirección General de Impositiva* y del *Banco de Previsión Social* correspondientes al periodo 2008. Dichos registros han sido depurados en base a encuestas realizadas por el Instituto Nacional de Estadística (*IVF, IMS, EAAE, etc.*), así como relevamientos específicos efectuados en períodos anteriores, con el propósito de mejorar la información del marco muestral.

El marco muestral no posee un *enlace* perfecto con el universo de estudio. Más precisamente posee errores de *sobre* y *sub-cobertura*, debido a la existencia de empresas

mal clasificadas según su actividad económica y no incluye a aquellas empresas que sub-declaren personal (a *BPS*) o ventas (a *DGI*).

Se excluyeron del marco muestral las empresas de Zona Franca, debido a que el *INE* realiza de forma anual un censo específico. Por lo tanto las empresas son incluidas posteriormente de manera forzosa a la muestra (se representan así mismas).

4. Diseño muestral

La muestra se seleccionó al azar mediante un muestreo aleatorio estratificado. Dentro de cada estrato se seleccionó una muestra independiente aplicando un muestreo sistemático con arranque aleatorio, ordenando las empresas por clase de actividad a 4 dígitos y tamaño en términos de su Personal Ocupado promedio.

La Encuesta Anual de Actividad Económica (*EAAE*), tiene como propósito proporcionar estimaciones confiables a nivel de Rama de Actividad, ya sea a dos, tres o cuatro dígitos según la clasificación industrial CIIU Rev4.

Dentro de cada rama de estudio se definieron cuatro estratos de tamaño:

1. Empresas con personal ocupado promedio dependiente entre 10 y 19.
2. Empresas con personal ocupado promedio dependiente entre 20 y 49.
3. Empresas con personal ocupado promedio dependiente entre 50 y 99.
4. Empresas con personal ocupado dependiente mayor o igual a 100, o que declararon ventas ante la *DGI* por más de 120 millones de pesos. Exceptuando a aquellas unidades donde su actividad principal es comercial, las cuales deben tener Ventas mayores a 240 millones de pesos.

El estrato 4, es de inclusión Forzosa (censado) a priori.

En el siguiente cuadro se presenta el número de empresas en el marco muestral por estrato de tamaño según la actividad principal económica de la empresa (a nivel de Sección de Actividad).

Cuadro 1 - Número de empresas por estrato-tamaño según sección de actividad CIIU Rev.4

VBPSeccion de Actividad CIIU Rev.4	Total	Estrato tamaño			
		10 a 19	20 a 49	50 a 99	Forzoso
Total	9363	5118	2631	700	914
B. Explotación de minas y canteras	50	35	11	3	1
C. Industrias Manufactureras	2221	1157	621	149	294
D. Suministro de electricidad, gas, vapor y aire acondicionado	9	4	2	-	3
E. Suministro de agua; alcantarillado, gestión de desechos y actividades de saneamiento	32	11	9	6	6
G. Comercio al por mayor y al por menor; reparación de los vehículos	2896	1757	778	178	183
H. Transporte y almacenamiento	820	461	217	54	88
I. Alojamiento y servicio de comida	582	365	173	25	19
J. Información y comunicación	296	146	95	23	32
M. Actividades profesionales, científicas y técnicas	443	259	121	34	29
N. Actividades administrativas y servicios de	522	217	152	54	99
P. Enseñanza	582	230	207	93	52
Q. Servicios sociales y relacionados con la Salud humana.	675	342	187	57	89
R. Artes, entretenimiento y recreación.	235	134	58	24	19

Fuente: Instituto Nacional de Estadística

Los estratos así definidos, se deben a la asociación entre los tramos del personal ocupado y de los tramos de venta con las variables de interés de la EAAE (ver Cuadro 2 y 3). Si bien una estratificación más eficiente se podría dar utilizando como variable auxiliar el monto imponible y utilizando mecanismos más complejos a la hora de definir los límites de los estratos, la misma resultaría poco interpretativa, debido a que los límites de los estratos variarían de una rama a otra.

Cuadro 2 - Número de empresas por estrato-tamaño según quintil de Valor Bruto de Producción

Quintil VBP	Total	Estrato tamaño			
		10 a 19	20 a 49	50 a 99	Forzoso
Total	8086	4227	2329	376	1154
1er	1609	1363	202	33	10
2do	1610	1191	376	34	9
3er	1625	865	618	108	34
4to	1616	645	719	81	172
5to	1625	163	414	120	928

Fuente: Estimaciones EAAE 2007 – Instituto Nacional de Estadística

Cuadro 3 - Número de empresas por estrato-tamaño según quintil de Valor Agregado Bruto

Quintil VAB	Total	Estrato tamaño			
		10 a 19	20 a 49	50 a 99	Forzoso
Total	8086	4227	2329	376	1154
1er	1613	1422	178	3	10
2do	1617	1323	253	31	10
3er	1604	771	735	68	31
4to	1629	586	779	113	151
5to	1623	126	384	162	952

Fuente: Estimaciones EAAE 2007 – Instituto Nacional de Estadística

4.1 Determinación del tamaño de muestra por rama y asignación por estrato tamaño.

El diseño muestral de la *EAAE 2008*, permite desagregar las estimaciones a nivel de rama de actividad. El total de ramas donde se requieren estimaciones por separado asciende a 117.

La variable auxiliar utilizada para determinar el tamaño de la muestra es el Valor Agregado Bruto (*VAB*), el cual es obtenido de la muestra de la *EAAE 2007*. En tanto, para aquellas actividades donde no se tiene información sobre el *VAB* (por ser clases de actividades no investigadas en el año 2007 o por no tener estimaciones estables), se utilizó el Monto Imponible (Remuneraciones) proveniente del *BPS*, al igual que en el diseño muestral del año 2007.

Para definir los tamaños de muestra a nivel de cada rama de actividad, se modificó el criterio utilizado para el diseño muestral del año 2007. En el 2008 se definieron precisiones en base a la importancia de las ramas en el total del Universo, en términos de su Valor Agregado Bruto del año 2007 o sus remuneraciones del año 2008.

Para estimar el total de la variable de interés se exigió una precisión del 5% en aquellas ramas de actividad que acumulan el 50 % del Valor Agregado Bruto total (ordenado de mayor a menor). En tanto para el resto de las ramas de actividad se exigió una precisión del 10 % para estimar el total de la variable interés².

El tamaño de muestra viene dado por la siguiente fórmula:

$$n_i = \frac{1.96^2 \left(\sum_{h=1}^M N_{ih} S_{ih} \right)^2}{\varepsilon^2 + 1.96^2 \sum_{h=1}^M N_{ih} S_{ih}^2} \quad i = 1, \dots, 117 ,$$

donde n_i es el tamaño de la muestra en la rama de actividad i , N_{ih} , es el tamaño del estrato h de la rama de actividad i en el marco muestral, S_{ih} es la varianza de la variable auxiliar utilizada para el cálculo (*VAB* o Remuneraciones) en el estrato de tamaño h de la rama de actividad i , 1.96 es el valor de una variable normal estándar que acumula 0,975 de probabilidad y ε es la precisión deseada (% del total del *VAB* o Remuneraciones).

Una vez determinados los tamaños de muestra para cada rama, se procedió a asignar la misma por estrato tamaño de manera *óptima*, utilizando como variable auxiliar el Valor Agregado Bruto o Remuneraciones, dependiendo de la disponibilidad de las mismas

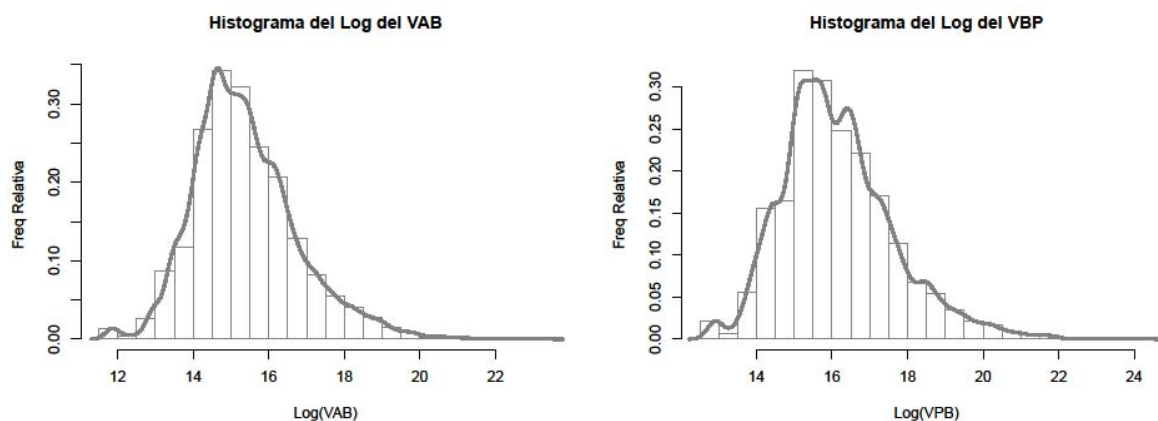
² Monto Imponible (Remuneraciones) ó Valor Agregado Bruto, dependiendo de la disponibilidad de las mismas
10 Instituto Nacional de Estadística – División Estadísticas Económicas

$$n_{ih} = \frac{n_i N_{ih} S_{ih}}{\sum_{h=1}^M N_{ih} S_{ih}} \quad i = 1, \dots, 117,$$

donde n_{ih} es el tamaño de muestra del estrato de tamaño h en la rama i y S_{ih} es la desviación estándar de la variable auxiliar utilizada para el cálculo en el estrato de tamaño h de la rama i .

Debido a la asimetría positiva en las variables de interés (en logaritmo, ver gráfico 1), así como a la naturaleza de la *asignación óptima*, en los estratos pequeños en términos de personal ocupado y ventas, se exigió un mínimo de cinco unidades muestrales (aunque la asignación no lo requiera). De esta manera, se evita estimaciones inestables y se procura preservar un *diseño medible* ante la pérdida de observaciones por causa de la no respuesta. En tanto, para los estratos en donde $n_{ih} \geq N_{ih}$, el estrato fue censado.

Gráfico 1: Histograma del Logaritmo del VAB y del Logaritmo del VBP



Fuente: Estimaciones EAAE 2007 – Instituto Nacional de Estadística

4.2 Tamaño de la muestra.

Bajo los requerimientos anteriormente expuestos el tamaño de muestra es de 2825 unidades (ver cuadro 4). El tamaño de muestra para el año 2008 es un 48 % más grande que para el año 2007, en donde el mismo se situó en 1903. El incremento en el tamaño de muestra se debe principalmente al aumento en las precisiones en las ramas de actividad y en la inclusión de nuevas clases de actividad económica.

Cuadro 4 - Tamaño de muestra por estrato -tamaño según sección de actividad CIU Rev.4

Sección de Actividad CIU Rev.4	Total	Estrato tamaño			
		10 a 19	20 a 49	50 a 99	Forzoso
Total	2825	781	715	415	914
B. Explotación de minas y canteras	20	6	10	3	1
C. Industrias Manufactureras	923	273	233	123	294
D. Suministro de electricidad, gas, vapor y aire acondicionado	9	4	2	-	3
E. Suministro de agua; alcantarillado, gestión de desechos y actividades de saneamiento	25	6	7	6	6
G. Comercio al por mayor y al por menor; reparación de los vehículos	622	190	155	94	183
H. Transporte y almacenamiento	264	72	69	35	88
I. Alojamiento y servicio de comida	83	31	20	13	19
J. Información y comunicación	136	35	48	21	32
M. Actividades profesionales, científicas y técnicas	148	57	39	23	29
N. Actividades administrativas y servicios de apoyo	198	34	44	21	99
P. Enseñanza	99	16	16	15	52
Q. Servicios sociales y relacionados con la Salud humana	193	28	37	39	89
R. Artes, entretenimiento y recreación	105	29	35	22	19

Fuente: Instituto Nacional de Estadística

5. Relevamiento y no respuesta

La muestra original de la *EAAE* se subdividió de forma aleatoria en cuatro submuestras (*olas*). El envío de formularios, su recepción y posterior procesamiento, se realizó siguiendo las *olas*. De esta manera se distribuyeron de forma uniforme los esfuerzos en la recolección de los datos y así evitar que en los estratos de tamaño 10 a 19 y 20 a 49 la tasa de no respuesta fuera excesivamente alta. De todas formas, dado la imposibilidad de ubicar a las empresas “pequeñas” y a la volatilidad de las mismas, las tasas de no respuesta fueron más altas para este tipo de empresas.

No se registró no respuesta por rechazo de las empresas a contestar el formulario de la *EAAE*. Las omisiones se deben a la imposibilidad de ubicar a las empresas por deficiencias en el marco muestral. Dado que el *INE*, mantiene un registro completo para empresas grandes (estrato 3 y 4), debido a que las mismas se encuentran incluidas en la mayoría de las encuestas que realiza el Instituto, la tasa de no respuesta para este conjunto de empresas fue muy “pequeña”.

La no respuesta introduce un posible sesgo a los estimadores así como incrementa la variabilidad de los mismos. Lo anterior se debe a la reducción del tamaño de la muestra el cual fue calculado a priori para obtener estimaciones con una precisión deseada.

La reducción del tamaño de muestra obligó a suprimir las estimaciones para las aperturas inicialmente previstas, optando por publicar a un nivel menos desagregado, mayoritariamente a nivel de división de actividad (ver cálculo de ponderadores).

La tasa de no respuesta se situó en el 19,43 % (ver cuadro 5), en donde se aprecia una relación negativa entre los estratos de tamaño y la tasa de no respuesta.

Cuadro 5 - Tasa de no respuesta por estrato de tamaño

Estrato Tamaño	Total	Situación		Tasa de No respuesta (%)
		Omisiones	Contestaron	
Total	2825	549	2276	19,43
10 a 19	781	235	546	30,09
20 a 49	715	160	555	22,38
50 a 99	415	63	352	15,18
Forzoso	914	91	823	9,96

Fuente: EAAE 2008 – Instituto Nacional de Estadística

La tasa de no respuesta se situó por encima del 30 % en las secciones de Actividad B, E, M, N y R.

Cuadro 6 - Tasa de no respuesta por sección de actividad según CIU Rev4.

Sección de Actividad	Total	Situación		Tasa de No respuesta (%)
		Omisiones	Contestaron	
Total	2825	549	2276	19,43
B. Explotación de minas y canteras	20	7	13	35,00
C. Industrias Manufactureras	923	158	765	17,12
D. Suministro de electricidad, gas, vapor y aire acondicionado	9	1	8	11,11
E. Suministro de agua; alcantarillado, gestión de desechos y actividades de saneamiento	25	9	16	36,00
G. Comercio al por mayor y al por menor; reparación de los vehículos	622	83	539	13,34
H. Transporte y almacenamiento	264	43	221	16,29
I. Alojamiento y servicio de comida	83	17	66	20,48
J. Información y comunicación	136	18	118	13,24
M. Actividades profesionales, científicas y técnicas	148	49	99	33,11
N. Actividades administrativas y servicios de apoyo	198	63	135	31,82
P. Enseñanza	99	14	85	14,14
Q. Servicios sociales y relacionados con la Salud humana	193	50	143	25,91
R. Artes, entretenimiento y recreación	105	37	68	35,24

Fuente: EAAE 2008 – Instituto Nacional de Estadística

6. Cálculo de los ponderadores

El uso de información auxiliar permite bajo ciertas condiciones obtener estimadores de mayor precisión. Para el caso de la EAAE, se utilizaron estimadores calibrados, los cuales generan un nuevo sistema de ponderadores en base a la información auxiliar disponible.

A continuación, se hace una breve reseña metodológica sobre la calibración. Se trata de un procedimiento para calcular sistemas de ponderadores que estimen de forma más eficiente los parámetros poblacionales. Dicho procedimiento es de vital importancia cuando se está ante casos de no respuesta y las variables auxiliares utilizadas se encuentran en mayor o menor medida correlacionadas con las variables de interés. Si las variables auxiliares utilizadas son cualitativas (ejemplo tramo de personal ocupado) a la calibración se le denomina post-estratificación.

La post-estratificación puede aplicarse de dos formas dependiendo de la información auxiliar disponible. El primer caso es si se dispone de información a nivel de las celdas de una tabla de contingencia, por ejemplo, si se conoce la distribución de las empresas por tramo de personal ocupado y ventas para las empresas del universo de interés, se pueden ajustar los ponderadores muestrales con dicha información.

Sea X y Z dos variables cualitativas con I y J categorías respectivamente, la tabla de contingencia que surge de cruzar estas dos variables tiene $I \times J$ celdas. Para cada una de ellas se obtiene una estimación del total poblacional N_{ij} sumando los ponderadores muestrales de las empresas pertenecientes a la celda, o sea $\hat{N}_{ij} = \sum_{k \in s_{ij}} a_k$, donde a_k es el ponderador muestral original de la empresa k y s_{ij} es el subconjunto de las empresas incluidas en la muestra que pertenecen a la celda (ij) .

El ajuste de los ponderadores muestrales a_k se realiza haciendo el cociente entre el total poblacional N_{ij} y el total estimado \hat{N}_{ij} . El ponderador muestral obtenido luego de la post-estratificación se calcula como

$$w_k = \frac{N_{ij}}{\hat{N}_{ij}} a_k.$$

En donde, para aquellas celdas donde se este sobreestimando ($\hat{N}_{ij} / N_{ij} > 1$), el ponderador calibrado w_k es menor que el ponderador original a_k y en aquellas donde se esté subestimando ($\hat{N}_{ij} / N_{ij} < 1$), el ponderador modificado w_k es mayor que a_k .

El nuevo sistema de ponderadores a_k estima sin error los totales poblacionales de las celdas por las cuales se ajustó. A su vez, si las variables utilizadas se encuentran asociadas a la no respuesta de las empresas, se obtienen mejores resultados en las estimaciones.

La post-estratificación es un caso especial de los estimadores de regresión en donde se utiliza un modelo de análisis de varianza o de clasificación a una vía.

El segundo caso, se da cuando no se conocen los totales poblacionales a nivel de las celdas, pero sí se conocen al nivel de las marginales de una tabla. Esto se puede dar por ejemplo, cuando la información auxiliar proviene de distintas fuentes, lo cual imposibilita confeccionar una tabla cruzada. Otro motivo puede ser que haya pocas observaciones a nivel de las celdas (caso de la EAAE), generando que los estimadores obtenidos no sean estables, en este caso se realiza un ajuste iterativo proporcional denominado raking.

Al igual que en el caso anterior, se parte de las estimaciones de los totales de las celdas, para así obtener una estimación de las marginales: $\hat{N}_{.i} = \sum_j \hat{N}_{ij}$ y $\hat{N}_{.j} = \sum_i \hat{N}_{ij}$.

El algoritmo puede comenzar ajustando las filas o las columnas. Si se comienza ajustando las filas se obtendrá un primer sistema de ponderadores que reproduzca la marginal de la fila. Se obtiene entonces un ponderador $w_k^1 = a_k N_{.i} / \hat{N}_{.i}$ donde $N_{.i}$ es el total poblacional de la marginal. Con los nuevos ponderadores w_k^1 se obtiene nuevas estimaciones para los totales de las filas y de las columnas. Los de las filas ajustan sin error, por la forma en que fueron calculados, pero no tiene por que suceder lo mismo para las columnas. Se obtiene así una nueva estimación para los totales de las celdas $\hat{N}_{ij}^1 = \sum_{k \in s_{ij}} w_k^1$ y por lo tanto para las columnas $\hat{N}_{.j}^1 = \sum_i \hat{N}_{ij}^1$.

El segundo ajuste se realiza obteniendo los ponderadores $w_k^2 = w_k^1 N_{.j} / \hat{N}_{.j}^1$, con lo cual se estima de forma exacta las columnas, pero cambia la estimación de las filas. El algoritmo

continúa este procedimiento de ajuste por filas y columnas hasta que la diferencia entre las marginales estimadas y las poblacionales sean menores a un margen de error fijado.

Se obtiene entonces un nuevo sistema de ponderadores calibrados w_k , los cuales estiman sin error las marginales de las variables auxiliares utilizadas.

El raking es un caso especial de los estimadores de regresión en donde se utiliza un modelo Log-Lineal.

6.1 Grupos de calibración.

Debido a la pérdida de unidades (resultando estratos vacíos o tasas de no respuesta alta) y a la necesidad de obtener estimaciones con una buena precisión a un nivel más desagregado de la población, se calcularon sistemas de ponderadores para obtener estimaciones confiables a nivel de división de actividad según la CIIU Rev4.

Para el caso de la post-estratificación y del raking descrito anteriormente, se utiliza información auxiliar a nivel de toda la población para calcular un sistema único de ponderadores. Dicho sistema de ponderadores es utilizado para estimar todas las variables y para todas las subpoblaciones (divisiones) de interés.

De esta forma, se obtiene un sistema de ponderadores que estiman de forma más eficiente los totales poblacionales, pero no tiene porque suceder lo mismo con las estimaciones a nivel de las divisiones de actividad si las mismas presentan diferentes estructuras.

Una manera de obtener estimadores más eficientes, es calculando sistemas de ponderadores para cada una de las divisiones de actividad, los cuales son determinados utilizando únicamente información auxiliar específica de la división o de un subconjunto de la población que contenga a la división y se encuentre próximo a la misma (o sea, que tenga aproximadamente la misma cantidad de elementos y que sus totales poblacionales sean parecidos). A estos subconjuntos se les denomina grupos de calibración³.

Por ello se utilizó información auxiliar a nivel de los grupos de calibración, en donde cada división de actividad se encuentra contenida solamente en un grupo. Por otro lado, algunas divisiones de actividad pueden tener muy pocas observaciones, lo cual puede generar estimadores inestables si se utiliza información auxiliar solo de la división de actividad. Con los grupos de calibración, se obtiene compromiso entre la información auxiliar a nivel de toda la población y de la información específica de la división.

Se conformaron dieciocho grupos de calibración, en donde el tamaño de muestra efectivo (en el grupo) es mayor que cincuenta y presentan estructuras lo más homogéneas posibles. A su vez no participaron en el proceso de estimación aquellos estratos de tamaño forzosos, en donde no se registró no respuesta.

Para la calibración se utilizaron los ponderadores efectivos $a_{ek} = a_k f_{rih}^{-1} = N_{ih} / n_{ih} f_{rih}^{-1}$, donde a_{ek} es el ponderador efectivo de la empresa k perteneciente a la rama i y al estrato de tamaño h y f_{rih} es la tasa de respuesta en la rama i - estrato de tamaño h .

La utilización de los ponderadores efectivos a_{ek} , a diferencia de los ponderadores originales a_k , se debe al hecho de no "forzar" tanto la calibración debido a la no respuesta.

³ Estevao, Victor M.; Särndal C.E. (2004) "Borrowing Strength Is Not the Best Technique Within a Wide Class of Design - Consistent Domain Estimators". Journal of Official Statistics, Vol 20, No 4.

Una vez conformados los grupos de calibración, se utiliza la post-estratificación o raking, dependiendo del grupo de calibración. La información auxiliar utilizada para el cálculo es: remuneraciones, ventas, personal ocupado y división de Actividad⁴.

En el cuadro 7 se presenta la conformación de los grupos de calibración, las divisiones que comprenden, la técnica y la información auxiliar utilizada.

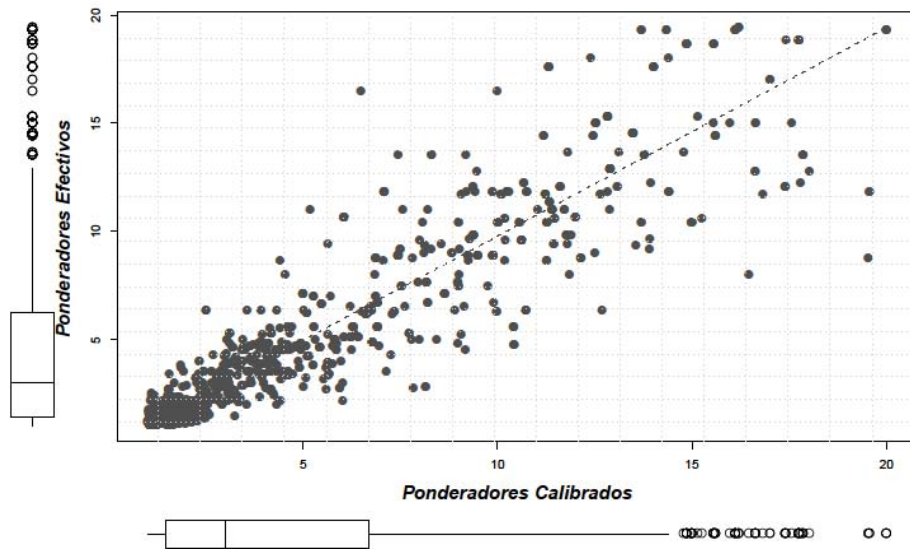
Cuadro 7 - Grupos de calibración.

Grupo Calibración	de	División Actividad	de	Técnica	Información Auxiliar
1		10		Raking	Remuneraciones, Personal Ocupado
2		11-13-14-15		Post-Estratificación	Remuneraciones
3		16-17-18		Raking	Remuneraciones, Personal Ocupado
4		20-21-22-23		Raking	Remuneraciones, Tramo de Ventas
5		24-25		Post-Estratificación	Remuneraciones
6		26-27-28-29-31-32-33-35-38		Raking	Remuneraciones, Tramo de Ventas
7		45		Post-Estratificación	Remuneraciones
8		46		Post-Estratificación	Remuneraciones
9		47		Post-Estratificación	Remuneraciones
10		49		Raking	Remuneraciones, Tramo de Ventas
11		50-51-52-53		Raking	Remuneraciones, Personal Ocupado
12		55-56		Raking	Remuneraciones, Personal Ocupado
13		58-59-60-61-62-63		Post-Estratificación	Remuneraciones
14		69-70-71-72-73		Raking	División de Actividad, Remuneraciones
15		77-78-79		Raking	Remuneraciones, Personal Ocupado
16		80-81		Raking	División de Actividad, Remuneraciones
17		85		Post-Estratificación	Remuneraciones
18		86-87-88		Post-Estratificación	División de Actividad, Remuneraciones

Para las divisiones *B* – Explotación de Minas y Canteras y *R* – Artes, entretenimiento y recreación no se realizaron ajustes y se emplearon los ponderadores originales corregidos por las tasas de respuesta.

En el siguiente gráfico se presenta la modificación de los ponderadores efectivos obtenidos a partir de los grupos de calibración.

⁴ Para el caso del Personal Ocupado Promedio y las Remuneraciones al ser variables cuantitativas, se codificaron en tramos, en donde los límites de los mismos dependen de la asimetría de las variables en el grupo de calibración.



Fuente: Instituto Nacional de Estadística.

7. Estimadores

Los sistemas de ponderadores obtenidos de los Grupos de Calibración conforman un sistema único (debido a las características del diseño muestral), los cuales son utilizados para estimar todas las variables y todos los subconjuntos de interés de la EAAE.

Estimador de un total.

Sea y_k el valor de la variable de interés y que toma la empresa k y w_{Ck} el ponderador de la empresa k perteneciente al grupo de calibración C .

El estimador para el total de la variable t_y , se define como:

$$\hat{t}_y = \sum_{k \in s} w_{Ck} y_k$$

Estimador de una media.

El estimador de una media poblacional se obtiene como el cociente entre el estimador del total y la estimación del tamaño de la población:

$$\hat{y} = \frac{\hat{t}_y}{\hat{N}} = \frac{\sum_{k \in s} w_{Ck} y_k}{\sum_{k \in s} w_{Ck}}$$

Estimador de una razón.

El estimador de una razón se emplea cuando el objetivo es estimar el cociente de dos totales desconocidos, sea $R_{xy} = t_y / t_x$ el cociente entre la variable y - x , su estimación viene dada como:

$$\hat{R}_{xy} = \frac{\hat{t}_y}{\hat{t}_x} = \frac{\sum_{k \in s} w_{Ck} Y_k}{\sum_{k \in s} w_{Ck} X_k}$$

8. Precisión de las estimaciones

Los errores muestrales ocurren debido a que las inferencias acerca de la población son basadas en información obtenida de una muestra de la misma. El diseño muestral, la variabilidad de los datos y el tamaño de muestra determinan el error muestral. Adicionalmente, diferentes métodos de estimaciones conllevan a diferentes errores muestrales dado el diseño muestral implementado.

Una medida común del error es la desviación estándar del estimador. La desviación estándar es el grado de variación en las estimaciones como resultado de la selección de una muestra en particular en lugar de otra del mismo tamaño y diseño. El verdadero valor del parámetro de interés se encuentra dentro de un rango de +/- una desviación estándar en el 68 % de las muestras y de +/- dos desviaciones estándar para el 95 % aproximado de todas las muestras.

El coeficiente de variación (CV) es la desviación estándar expresada como un porcentaje de la estimación. Se utiliza para indicar el grado de incertidumbre asociado a una estimación, por ejemplo, si la estimación puntual del Personal Ocupado para una división de actividad es 1000, y el CV correspondiente es del 5%, entonces el verdadero valor se encuentra entre 950 y 1050, en el 68 % de las veces y entre 900 y 1100 en el 95 % de las veces.

Los coeficientes de variación para la *EAAE* se estimaron utilizando el paquete *Survey* del programa *R*. Los coeficientes de variación de las estimaciones provenientes de la *EAAE* dependen de la desagregación. Para las principales variables de interés (*VBP*, *VAB*, *CI*, *PO*, *REM*) los coeficientes de variación para el total de la población son menor que el 1 % (ver cuadro 8). En tanto, las precisiones a nivel de sección de actividad se encuentran por debajo del 5 % en la mayoría de los casos a excepción de la sección B – Explotación de Minas y Canteras, I – Alojamiento y servicio de comida y R- Artes, entretenimiento y recreación.

Cuadro 8 - Coeficientes de Variación para las principales variables por sección de actividad según CIU Rev4.

Sección de Actividad CIU Rev.4	Coeficientes de Variación (%)				
	VBP	VAB	CI	PO	REM
Total	0,8%	0,9%	1,0%	0,7%	0,8%
B. Explotación de minas y canteras	22,0%	22,0%	24,9%	4,3%	16,4%
C. Industrias Manufactureras	1,0%	1,0%	1,2%	1,3%	1,1%
D. Suministro de electricidad, gas, vapor y aire acondicionado	0,0%	0,0%	0,0%	0,0%	0,0%
E. Suministro de agua; alcantarillado, gestión de desechos y actividades de saneamiento	1,5%	1,5%	5,3%	1,9%	1,5%
G. Comercio al por mayor y al por menor; reparación de los vehículos	2,7%	2,7%	2,6%	1,0%	2,4%
H. Transporte y almacenamiento	4,8%	4,8%	5,8%	2,2%	2,3%
I. Alojamiento y servicio de comida	7,2%	7,2%	6,8%	5,9%	5,7%
J. Información y comunicación	0,9%	0,9%	1,1%	1,3%	1,9%
M. Actividades profesionales, científicas y técnicas	5,3%	5,3%	6,8%	3,3%	6,1%
N. Actividades administrativas y servicios de apoyo	3,8%	3,8%	4,8%	2,7%	3,5%
P. Enseñanza	3,9%	3,9%	7,3%	3,0%	4,4%
Q. Servicios sociales y relacionados con la Salud humana	2,5%	2,5%	2,7%	2,4%	2,0%
R. Artes, entretenimiento y recreación	8,9%	8,9%	13,0%	5,6%	9,0%

Fuente: Instituto Nacional de Estadística

Bibliografía

Estevao, Victor M.; Särndal C.E (2004) "Borrowing Strength Is Not the Best Technique Within a Wide Class of Design – Consistent Domain Estimators". Journal of Official Statistics, Vol 20, No 4.

Fuller, Wayne A. (2009). Sampling Statistics. John Wiley & Sons.

Lumley, T. (2004) Analysis of complex survey samples. Journal of Statistical Software 9(1): 1- 19.

Lumley, T. (2009) "Survey: analysis of complex survey samples". R packaged version 3.11-2.

Särndal, C.E.; Swensson, B.; Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag.

R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.Rproject.org>.

Anexo 1

Anexo 1 - Coeficientes de variación por división de actividad.

División - Agrup div	Descripción	Coeficiente de Variación (%)				
		VBP	VAB	CI	PO	REM
B	Explotación de minas y canteras	22,0%	22,0%	24,9%	4,3%	16,4%
10	Elaboración de productos alimenticios	2,4%	2,3%	2,6%	2,5%	2,4%
11	Elaboración de bebidas	0,8%	0,9%	0,8%	5,8%	0,8%
12	Elaboración de productos de tabaco	0,0%	0,0%	0,0%	0,0%	0,0%
13	Fabricación de productos textiles	3,6%	6,6%	3,9%	7,9%	7,3%
14	Fabricación de prendas de vestir	5,9%	5,2%	8,0%	2,6%	3,7%
15	Fabricación de cueros y productos conexos	4,3%	4,1%	4,5%	3,1%	3,5%
16	Producción de madera y fabricación de productos de madera y corcho, excepto muebles; fabricación de artículos de paja y de materiales trenzables	1,4%	4,2%	1,5%	3,9%	4,0%
17	Fabricación de papel y de los productos de papel	0,9%	0,2%	1,6%	3,5%	1,5%
18	Actividades de impresión y reproducción de grabaciones	9,5%	12,8%	8,4%	5,1%	8,1%
19	Fabricación de coque y de productos de la refinación del petróleo	0,0%	0,0%	0,0%	0,0%	0,0%
20	Fabricación de sustancias y productos químicos	1,3%	0,7%	1,8%	1,7%	2,0%
21	Fabricación de productos farmacéuticos, sustancias químicas medicinales y de productos botánicos	2,6%	2,7%	2,8%	2,0%	2,3%
22	Fabricación de productos de caucho y plástico	2,8%	2,8%	3,7%	2,9%	3,2%
23	Fabricación de otros productos minerales no metálicos	5,3%	4,1%	6,6%	6,9%	6,9%
24	Fabricación de metales comunes	8,5%	1,5%	11,7%	17,3%	3,0%
25	Fabricación de productos derivados del metal, excepto maquinaria y equipo	6,5%	9,1%	6,5%	6,1%	6,4%
26	Fabricación de los productos informáticos, electrónicos y ópticos	13,2%	13,5%	14,1%	7,9%	14,8%
27	Fabricación de equipo eléctrico	10,0%	3,0%	14,3%	0,5%	3,0%
28	Fabricación de maquinaria y equipo n.c.p.	13,0%	16,2%	12,1%	17,6%	26,2%
29	Fabricación de vehículos automotores, remolques y semiremolques	1,2%	1,1%	1,5%	1,4%	1,8%
30	Fabricación de otros tipos de equipo de transporte	0,0%	0,0%	0,0%	0,0%	0,0%
31	Fabricación de muebles	6,4%	3,5%	8,2%	5,7%	3,9%
32	Otras industrias manufactureras	6,2%	7,4%	8,3%	10,4%	8,7%
33	Reparación e instalación de la maquinaria y equipo	4,0%	6,5%	4,6%	2,4%	4,7%
35	Suministro de electricidad, gas, vapor y aire acondicionado	0,0%	0,0%	0,0%	0,0%	0,0%
36 y 37	Captación, tratamiento y suministro de agua. Alcantarillado	0,0%	0,0%	0,0%	0,0%	0,0%
38	Recolección, tratamiento y eliminación de desechos, recuperación de materiales	10,8%	16,9%	18,8%	8,6%	13,2%
45	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	10,9%	13,8%	9,5%	4,6%	5,7%
46	Comercio al por mayor, excepto de los vehículos de motor y de las motocicletas	3,7%	4,4%	3,9%	1,7%	4,6%
47	Comercio al por menor, excepto el comercio de vehículos automotores y motocicletas	3,8%	4,8%	3,3%	1,4%	2,3%
49	Transporte por vía terrestre; transporte por tuberías	6,3%	3,3%	9,1%	3,2%	1,0%
50	Transporte por vía acuática	9,3%	20,7%	7,1%	3,2%	2,2%
51	Transporte por vía aérea	2,5%	0,9%	2,9%	1,2%	1,0%
52	Depósito y actividades de transporte complementarias	12,2%	7,6%	17,8%	4,4%	9,9%
53	Correo y servicios de mensajería	9,3%	11,9%	9,6%	2,9%	1,9%
55	Alojamiento	7,9%	11,5%	5,9%	5,1%	6,2%
56	Servicios de alimentos y bebidas	13,0%	17,8%	12,7%	10,0%	11,4%
58	Actividades de publicación	3,4%	4,6%	3,2%	3,2%	3,8%
59	Actividades de producción de películas, de video, de programas de televisión, grabación y publicación de música y sonido	2,9%	4,2%	2,6%	3,1%	3,0%
60	Actividades de programación y distribución	8,8%	7,7%	10,9%	5,5%	7,8%

Diseño muestral Encuesta Anual de Actividad Económica 2008

61	Telecomunicaciones	0,4%	0,3%	0,8%	0,7%	0,4%
62 y 63	Actividades de la tecnología de información y del servicio informativo. Actividades del servicio informático.	7,1%	8,3%	6,8%	5,7%	9,4%
69	Actividades jurídicas y de contabilidad	5,0%	8,4%	3,3%	8,9%	14,9%
70	Actividades de oficinas centrales, actividades de administración de empresas y de consultoría sobre administración de empresas	14,4%	15,9%	19,2%	4,1%	11,3%
71	Actividades de arquitectura e ingeniería; ensayos y análisis técnicos	4,7%	5,1%	4,8%	6,5%	5,6%
72	Investigación y desarrollo científico	3,4%	4,0%	2,7%	13,1%	6,9%
73	Publicidad e investigación de mercado	15,1%	11,5%	17,5%	3,8%	9,7%
77	Actividades de alquiler y arrendamiento	12,8%	10,3%	17,3%	9,4%	12,4%
78	Actividades de las agencias de empleo	5,5%	5,6%	7,1%	3,9%	5,7%
79	Actividades de las agencias de viajes, operadores turísticos y servicios de reserva relacionados.	7,1%	14,0%	7,0%	7,5%	11,7%
80	Actividades de seguridad e investigación	7,9%	9,0%	9,1%	6,9%	8,2%
81	Actividades de servicio a edificios y paisajes (jardines, áreas verdes, etc.)	19,8%	16,7%	29,9%	7,0%	9,6%
82	Actividades de oficinas administrativas, soporte de oficinas y otras actividades de soporte de negocios	6,7%	10,1%	7,9%	7,4%	6,0%
85	Enseñanza	4,0%	3,6%	7,3%	3,0%	4,4%
86	Actividades relacionadas con la salud humana	2,6%	3,0%	2,7%	2,6%	2,1%
87	Instituciones residenciales de cuidado	25,0%	24,4%	30,7%	25,5%	21,3%
88	Servicios sociales sin alojamiento	24,6%	24,4%	37,9%	22,9%	21,6%
92	Actividades de juego y apuestas	15,1%	7,1%	18,8%	2,2%	1,5%
93	Actividades deportivas, de diversión y esparcimiento	9,4%	10,6%	12,4%	7,1%	11,0%

Fuente: Instituto Nacional de Estadística.

Anexo 2

Modificación de los ponderadores efectivos según grupo de calibración.

