

### 3. ESTADÍSTICA OFICIAL

#### CALIBRATING A HOUSEHOLD SURVEY BY USING THE CALMAR PROGRAM

Carlos Pérez Arriero\*

Instituto Nacional de Estadística

##### Abstract

This paper addresses the problem of calibrating a survey to some external information. In the first part of the paper, after introducing the key concepts and basic definitions, the calibration techniques are explained. In the second part, a simulation experiment is performed: survey samples are taken from a census population, non-response is also simulated and the basic (Horvitz-Thompson) design estimator of some population totals is compared with a ratio estimator and with a calibration estimator. The conclusions are that calibration estimates are in general more accurate than those of other kind of estimators.

**Keywords:** Sample Survey, Calibration, CALMAR.

#### 1. Introducción

Cualquier usuario que disponga del fichero de microdatos muestrales de una encuesta dirigida a hogares, puede realizar una primera verificación de la misma sin más que sumar los factores de elevación de las personas de la muestra. Como cada factor equivale al número de personas a los que representa la unidad muestral, la suma debe coincidir con el total poblacional. Es decir, la muestra debe estar calibrada, al menos, a la población objetivo.

El siguiente paso puede ser estimar alguna otra variable de la que se conozca aproximadamente su distribución poblacional, por ejemplo el sexo o la edad. Si no se han tenido en cuenta estas variables en el proceso de selección, y no se ha realizado calibrado alguno con ellas, es muy posible que, por la aleatoriedad de la muestra o por falta de respuesta, las distribuciones estimadas sean erráticas, por ejemplo se podría obtener una distribución estimada del 80 por ciento de hombres y el 20 por ciento de mujeres. Entonces se podría concluir, no sin parte de razón, que si esta encuesta no es capaz de representar a la población por sexo y edad, tampoco será válida para proporcionar otro tipo de información social o económica.

La utilización de técnicas de calibrado se basa, de forma intuitiva, en el siguiente supuesto: si una

encuesta es capaz de estimar correctamente los totales conocidos de las variables auxiliares, también proporcionará buenas estimaciones de las variables objetivo. La mejora proporcionada por el calibrado será tanto mayor, cuanto más relacionadas estén las variables auxiliares con las que son objetivo de la encuesta.

#### 2. Antecedentes

En la estadística oficial, prácticamente desde sus inicios, las estimaciones proporcionadas por encuestas de hogares se han obtenido mediante el uso de estimadores de razón, utilizando como fuente auxiliar de información las proyecciones demográficas de población. Este tipo de estimadores recuperan el total de la variable auxiliar utilizada, lo que en el caso señalado supone la estimación correcta de la población actualizada. Por ello se puede decir que desde un primer momento se han empleado, en estadística oficial, técnicas de reponderación o de calibrado.

#### 3. Cambios con el incremento y mejora de la información auxiliar

El estimador de razón resultó útil cuando se disponía de una única variable auxiliar. Pero la presencia de más y mejor información obligó a seleccionar otro tipo de estimadores que, además de utilizar

---

\*Corresponding Author. E-mail: carlospe@ine.es

las nuevas variables auxiliares, permitieran superar algunos inconvenientes de tipo práctico que se resumen a continuación.

Por un lado es frecuente que se disponga de los totales poblacionales marginales de cada variable, pero no de los totales conjuntos. Por ejemplo, se puede disponer de los totales de población por grupos de edad y sexo, y también de los totales de población española y extranjera. Pero puede ocurrir que no se disponga del cruce de nacionalidad por grupos de edad y sexo.

El segundo problema es particular de aquellas encuestas en las que se seleccionan viviendas y, en una etapa posterior, se recoge información de todas las personas que residan en cada vivienda muestral. En este caso, el factor de elevación derivado de las probabilidades de selección, es el mismo para una vivienda y para las personas que habitan en ella, dado que en la vivienda no se realiza ningún proceso de selección. Este hecho resulta útil, pues hay tablas de viviendas y tablas de personas con celdas coincidentes, y si el factor de elevación es el mismo, las estimaciones también lo serán. El ejemplo tipo es el de la estimación del número de viviendas con una sola persona, y la del total de personas que viven solas. Sería útil emplear una técnica que permitiera realizar una reponderación de las personas de la muestra manteniendo el mismo factor para todas aquellas que residan en la misma vivienda. Una interesante propuesta para superar este problema fue elaborada por Lemaître y Dufour (1987).

Los métodos de calibrado desarrollados por Deville y Särndal (1992), y su programación posterior realizada por Deville, Särndal y Sautory (1993), en el software estadístico SAS con el nombre de CALMAR (CALage sur MARGes), intentan superar los problemas señalados. Esta propuesta metodológica, amplió el campo de actuaciones de tipo práctico, basado en la descripción del estimador de regresión generalizada (GREG) de Särndal, Swensson y Wretman (1992).

Un resumen de los desarrollos teóricos realizados hasta el momento en torno al tema del calibrado puede encontrarse en Särndal (2007).

El programa CALMAR se puede descargar gratuitamente, junto con un manual de uso, de la página web del Instituto de Estadística Francés (INSEE) ([www.insee.fr](http://www.insee.fr)), y es actualmente utilizado por numerosas oficinas de estadística públicas y centros

privados de todo el mundo. El éxito en su difusión se debe a que, apoyándose en un desarrollo teórico riguroso, ha sido programado de forma que su utilización no es complicada, y permite aplicar varias opciones de calibrado, tanto sobre variables auxiliares cualitativas como cuantitativas.

También está disponible la programación de la metodología de Deville y Särndal en otros tipos de programas estadísticos, como es el caso del módulo g-CALIB-S programado en SPSS por Vanderhoeft (2001).

#### 4. Características de la información auxiliar

Antes de seguir adelante con la metodología introducida en el párrafo anterior, parece conveniente señalar algunos criterios generales, que deben ser tomados en consideración antes de realizar el calibrado.

- Características de las variables auxiliares

*Disponibilidad:* En la metodología que se describe en este artículo, la información auxiliar debe ser conocida tanto al nivel de unidad muestral, como para el total de la población. No es necesario disponer de esta información en las unidades no incluidas en la muestra.

*Definición:* La definición de las variables auxiliares tiene que ser la misma en la muestra y en la población. Se evitarán, por ejemplo, errores del tipo de utilizar como variable el país de nacimiento de la persona en un caso, y en el otro la nacionalidad.

*Tipo:* En lo posible se utilizarán las variables más correlacionadas con las que son objetivo de la encuesta. Como normalmente las encuestas tienen múltiples variables objetivo, habrá que limitarse a considerar de entre ellas las más importantes.

*Número:* El número de variables auxiliares debe estar controlado. Una desagregación excesiva de la información auxiliar, además de poner en duda la calidad de la misma, puede llegar a incrementar la variabilidad de la estimación, e incluso, impedir la realización del calibrado por falta de muestra. Aunque no existe un criterio objetivo, se puede considerar como razonable no utilizar celdas de calibrado con menos de diez elementos muestrales.

- Características de los totales poblacionales auxiliares

*Referencia temporal:* La fuente externa pobla-

cional debe tener la misma referencia temporal que la encuesta.

*Calidad:* Los totales poblacionales deben ser fiables. De no ser así, se podrían incorporar sesgos de difícil medida en las estimaciones. Puede ser útil en algunos casos definir variables auxiliares más agregadas (por ejemplo territorialmente), para garantizar la calidad de los totales auxiliares.

*Procedencia:* Las fuentes de información auxiliar son diversas: registros administrativos, estudios demográficos o, incluso, otras encuestas de gran tamaño muestral. Las distintas fuentes externas que finalmente se utilicen, deben proporcionar datos homogéneos, tanto en lo relativo a su fiabilidad, como en cuanto a definiciones y referencias temporal y poblacional.

### 5. Reponderación con fuentes externas

Los siguientes párrafos resumen el desarrollo teórico planteado por Deville, Särndal y Sautory (1993).

Se considera una población  $U$  con  $N$  elementos  $u_1 \dots u_N$ , de la que se extrae una muestra  $S$  de tamaño  $n$ .

La variable objetivo es  $y$ . Se pretende estimar el total  $Y = \sum_{k \in U} y_k$

El estimador de partida es el de Horvitz-Thompson, que se puede expresar como:

$$\hat{Y}_\pi = \sum_{k \in s} \frac{1}{\pi_k} y_k = \sum_{k \in s} d_k y_k$$

donde  $\pi_k > 0$  para  $k = 1 \dots n$ , es la probabilidad de que la unidad  $u_k$  pertenezca a la muestra, y su inversa  $d_k$  es el peso de diseño o factor de elevación.

Se dispone de información sobre los totales poblacionales de  $J$  variables auxiliares. Es decir, para la variable auxiliar  $j$  se conoce el valor de:

$$X_j = \sum_{k \in U} x_{kj}$$

donde  $x_{kj}$  es el valor de la variable auxiliar  $j$  en la unidad  $k$ .

El objetivo del proceso de reponderación o calibrado, es encontrar unos pesos nuevos  $w_k$  que veri-

fiquen las ecuaciones de equilibrio:

$$\hat{X}_j = \sum_{k \in s} w_k x_{kj} = X_j \quad (j = 1, \dots, J)$$

Es decir, por calibrado se entiende la obtención de unos nuevos factores de elevación que permitan estimar correctamente los totales poblacionales conocidos.

Se tiene por tanto, un sistema de  $J$  ecuaciones de equilibrio y  $n$  incógnitas. Como  $J$  debe ser bastante menor que  $n$ , se trata de un sistema de ecuaciones lineales compatible indeterminado, es decir, con infinitas soluciones. Por tanto, habrá que establecer un criterio para seleccionar la más adecuada.

Los factores de elevación de partida corresponden a un estimador con buenas propiedades, como es el de Horvitz-Thompson (H-T), por tanto un criterio que parece razonable es elegir entre todos los conjuntos de pesos muestrales que verifiquen las condiciones de equilibrio, aquél en el que los nuevos factores se diferencien lo menos posible de los antiguos.

Para formalizar este procedimiento, se define una función de distancia que permita medir la proximidad de los posibles conjuntos de factores de calibrado a los pesos iniciales.

De acuerdo con todo lo anterior, el planteamiento del problema queda como sigue:

$$\min_{w_k} \sum_{k \in s} d_k G \left( \frac{w_k}{d_k} \right) \quad (k = 1, \dots, n)$$

Condicionado a:

$$\sum_{k \in s} w_k x_k = X$$

$$x'_k = (x_{k1}, x_{k2}, \dots, x_{kJ}) \quad X' = (X_1, X_2, \dots, X_J)$$

Donde  $G(x)$  es una función positiva y estrictamente convexa que verifica  $G(1) = G'(1) = 0$  y  $G''(1) = 1$ , con objeto de garantizar la existencia de solución.

Para resolver este problema de optimización se utiliza el método de los multiplicadores de Lagrange:

$$L = \sum_{k \in s} d_k G \left( \frac{w_k}{d_k} \right) - \lambda' \left( \sum_{k \in s} w_k x_k - X \right)$$

con:

$$\lambda' = (\lambda_1 \dots \lambda_J)$$

cuya expresión desarrollada sería:

$$L = \sum_{k \in s} d_k G \left( \frac{w_k}{d_k} \right) - (\lambda_1 \dots \lambda_J) \begin{pmatrix} \sum_{k \in s} w_k x_{k1} - X_1 \\ \vdots \\ \sum_{k \in s} w_k x_{kJ} - X_J \end{pmatrix}$$

Derivando respecto de  $w_k$  e igualando a cero se tiene:

$$\begin{aligned} d_k \frac{1}{d_k} G' \left( \frac{w_k}{d_k} \right) - (\lambda_1 \dots \lambda_J) \begin{pmatrix} x_{k1} \\ \vdots \\ x_{kJ} \end{pmatrix} &= 0 \\ \Rightarrow G' \left( \frac{w_k}{d_k} \right) - x'_k \lambda &= 0 \\ \Rightarrow w_k &= d_k \cdot F(x'_k \lambda) \quad \forall k = 1 \dots n \end{aligned}$$

Donde  $F$  es la función inversa de  $G'$ , que existe según las hipótesis realizadas. En esta expresión se observa con claridad que los nuevos pesos ( $w_k$ ) se obtienen multiplicando los factores de diseño ( $d_k$ ) por un valor que depende de la función de distancia que se elija.

El vector  $\lambda$  es la solución de un sistema de  $J$  ecuaciones con  $J$  incógnitas, que resulta de sustituir la expresión de los  $w_k$  en las ecuaciones de equilibrio:

$$\sum_{k \in s} w_k x_k = X \Rightarrow \sum_{k \in s} d_k F(x'_k \lambda) x_k = X$$

Este sistema, de  $J$  ecuaciones con  $J$  incógnitas se resuelve utilizando el método iterativo de Newton.

En el programa CALMAR el parámetro de entrada  $M$  permite seleccionar una función de distancia entre cuatro opciones. La primera de ellas,  $M = 1$ , corresponde al denominado 'método lineal', y es la siguiente:

$$G(x) = \frac{1}{2} (x - 1)^2$$

En este caso se tiene:

$$F(u) = 1 + u$$

Sustituyendo en las ecuaciones de equilibrio se llega a:

$$\begin{aligned} \sum_{k \in s} d_k F(x'_k \lambda) x_k = X &\Rightarrow \sum_{k \in s} d_k (1 + x'_k \lambda) x_k = X \\ \sum_{k \in s} d_k x_k x'_k \lambda &= X - \hat{X}_\pi \\ \Rightarrow \lambda &= \left( \sum_{k \in s} d_k x_k x'_k \right)^{-1} (X - \hat{X}_\pi) \end{aligned}$$

donde el subíndice  $\pi$  refiere al estimador obtenido con los pesos de diseño (H-T).

Finalmente se alcanza la expresión de los nuevos pesos, sustituyendo el valor obtenido para el vector  $\lambda$  en la fórmula general alcanzada anteriormente para los  $w_k$ :

$$w_k = d_k \left( 1 + x'_k \left( \sum_{k \in s} d_k x_k x'_k \right)^{-1} (X - \hat{X}_\pi) \right)$$

La estimación de la variable de interés  $y$ , obtenida con estos factores de elevación calibrados, vendrá dada por:

$$\begin{aligned} \hat{Y}_{Cal} &= \sum_{k \in s} w_k y_k = \sum_{k \in s} d_k y_k + \\ &+ \sum_{k \in s} d_k y_k x'_k \left( \sum_{k \in s} d_k x_k x'_k \right)^{-1} (X - \hat{X}_\pi) \end{aligned}$$

Que es la expresión del estimador de regresión generalizada (Särndal, Swensson y Wretman, 1992):

$$\hat{Y}_{Cal} = \hat{Y}_\pi + \hat{B}'_s (X - \hat{X}_\pi) = \hat{Y}_{GREG}$$

con:

$$\hat{B}_s = \left( \sum_s d_k x_k x'_k \right)^{-1} \sum_s d_k x_k y_k$$

Llegar a un estimador conocido y con buenas propiedades, como es el caso, ratifica el proceso de calibrado realizado con la distancia  $G$ .

Este hecho permite, además, el cálculo de la varianza de la estimación mediante el uso de las fórmulas propias del GREG:

$$\hat{V}(\hat{Y}_{GREG}) = \sum_s \sum_s \left( \frac{\pi_{kl} - \pi_k \pi_l}{\pi_{kl}} \right) (w_k e_k) (w_l e_l)$$

con:

$\pi_{kl}$  : Probabilidad de inclusión en la muestra de  $u_k$  y  $u_l$

$\pi_k$  : Probabilidad de inclusión en la muestra de  $u_k$

$$e_k = y_k - x'_k \hat{B}_s$$

El programa CALMAR permite utilizar, además de la función de distancia que da lugar al estimador GREG ( $M = 1$ ), tres distancias más ( $M = 2, 3$  y  $4$ ), alguna de las cuales también proporciona una estimación conocida, como es el caso de  $M = 2$ , que da lugar al conocido método de raking ratio (Deming y Stephan, 1940), en el caso de variables auxiliares cualitativas. No es inmediato decidir cuál de las cuatro distancias es la más adecuada en cada caso. Tranquiliza saber, no obstante, que todas ellas convergen a una misma varianza de la estimación con tamaños muestrales suficientemente grandes.

## 6. Aplicación práctica: Simulación

Con objeto de comprobar en un caso práctico las consecuencias de realizar un proceso de calibrado en las estimaciones de una encuesta, se realiza a continuación un ejercicio de simulación.

Con información procedente del último Censo de Población realizado en 2001, se ha elaborado un marco de viviendas correspondiente a la comunidad autónoma de Murcia. Como todos los marcos muestrales, contiene la identificación anonimizada de cada vivienda, y los valores de algunas variables susceptibles de ser utilizadas, tanto en el proceso de elaboración del diseño (estratos, conglomerados, etc...), como en el proceso de estimación.

En resumen, las variables disponibles para cada registro del fichero de viviendas son:

- Identificación: Provincia, municipio, distrito, sección y numeración (correlativa) de la vivienda.
- Información auxiliar: Personas por grupos quinquenales de edad, personas de nacionalidad española y extranjera en la vivienda.
- Variables objetivo: Ocupados, parados e inactivos en la vivienda.

De este marco se extrae un conjunto de dos mil muestras independientes, con un diseño similar al habitualmente utilizado en las encuestas de hogares, y se estudia la media y la variabilidad de diferentes estimadores.

### *Diseño muestral utilizado*

Muestreo bietápico con estratificación de las unidades primarias.

Las unidades primarias son las secciones censales, áreas geográficas en las que se divide cada municipio.

Las unidades secundarias son las viviendas familiares principales.

La estratificación de las unidades primarias se realiza en función del tamaño poblacional del municipio al que pertenecen.

La selección se realiza como sigue:

Unidades primarias: Se eligen con probabilidad proporcional a su tamaño, medido éste por el número de viviendas.

Unidades secundarias: Se seleccionan con probabilidades iguales.

El tamaño muestral es de 91 unidades primarias, en cada una de las cuales se eligen 18 unidades secundarias, lo que proporciona una muestra total de 1638 viviendas.

En cada vivienda muestral se obtiene información de todas las personas de 16 o más años que allí residan.

El objetivo es estimar el total de parados, ocupados e inactivos en la comunidad autónoma.

Se consideran tres estimadores diferentes:

1. Estimador de diseño (Horvitz-Thompson), basado en las probabilidades de pertenencia a la muestra.

$$\hat{Y} = \sum_s \frac{V_h}{v_h} y_h$$

donde:

$V_h$  : Viviendas en el estrato  $h$

$v_h$  : Muestra teórica de viviendas en el estrato  $h$

2. Estimador de razón separado

$$\hat{Y} = \sum_s \frac{P_h}{p_h} y_h$$

donde:

$P_h$  : Población de 16 o más años en el estrato  $h$

$p_h$  : Población de 16 o más años en la muestra efectiva del estrato  $h$

3. Estimador calibrado

Se ha aplicado el método lineal ( $M = 1$ ).

Como factores de entrada a CALMAR se han tomado los correspondientes al estimador de razón. No se han utilizado los de diseño por considerar que la razón corrige en alguna medida el sesgo de no respuesta y por ello constituye un mejor punto de partida para el calibrado.

Variables auxiliares (24):

- a. Grupos quinquenales de edad y sexo (22)
- b. Población española y extranjera (2)

En el fichero de datos muestrales sobre el que se realiza el proceso de calibrado, cada registro representa a una vivienda con los siguientes campos:

Identificación

Factor de entrada

Variables objetivo (3): Ocupados, parados e inactivos en la vivienda

Variables auxiliares (24):

G01-G22: personas de la vivienda en cada grupo de edad-sexo

ESP y EXT: personas de la vivienda españolas y extranjeras

Se ha incorporado a la simulación una cierta falta de respuesta, definida en función de la composición de la vivienda. Las probabilidades de respuesta (PR) consideradas intentando reproducir una situación real, son las siguientes:

Tamaño de la vivienda mayor que dos y sin extranjeros: PR=0,8

Tamaño de la vivienda menor o igual que dos y sin extranjeros: PR=0,7

Tamaño de la vivienda mayor que dos y con extranjeros: PR=0,6

Tamaño de la vivienda menor o igual que dos y con extranjeros: PR=0,5

En las siguientes tablas se resumen, en términos de media y precisión, las estimaciones obtenidas para cada una de las tres variables objetivo, con las dos mil muestras independientes seleccionadas según los criterios señalados.

TABLA I. Resumen de los resultados de la simulación (dos mil muestras)

*Sin falta de respuesta*

	Ocupados		Parados		Inactivos	
	Media	CV(%)	Media	CV(%)	Media	CV(%)
Estimador de diseño (H-T)	480.216	2,45	62.460	7,29	427.255	2,24
Estimador de razón	480.266	1,89	62.457	6,98	427.561	2,04
Estimador calibrado	480.243	1,44	62.480	6,95	427.559	1,41
Población	480.657		62.602		427.024	

CV es el coeficiente de variación de las estimaciones.

TABLA II. Resumen de los resultados de la simulación (dos mil muestras)

*Con falta de respuesta*

	Ocupados		Parados		Inactivos	
	Media	RMSE(%)	Media	RMSE(%)	Media	RMSE(%)
Estimador de diseño (H-T)	356.146	35,10	46.860	34,65	320.682	33,30
Estimador de razón	477.420	2,26	62.763	8,26	430.100	2,38
Estimador calibrado	480.488	1,66	62.747	8,39	427.048	1,61
Población	480.657		62.602		427.024	

RMSE es la raíz cuadrada del error cuadrático relativo medio, cuya expresión es:

$$RMSE(\%) = \frac{100}{Y} \sqrt{\frac{1}{2000} \sum_{k=1}^{2000} (\hat{Y}_{(k)} - Y)^2}$$

$\hat{Y}_{(k)}$ : estimación obtenida con la muestra  $k$

En la tabla 1, la primera consideración es la ausencia de sesgo apreciable en los tres estimadores. En el caso del estimador de Horvitz-Thompson esta situación era de esperar pues se trata de un estimador insesgado y otro resultado habría puesto en duda los cálculos realizados. En los otros dos casos, razón y calibrado (GREG), la situación es parecida, pues aunque no son estimadores insesgados, sí está probado que su sesgo tiende a cero conforme aumenta el tamaño de la muestra. De hecho son habitualmente admitidos dentro del grupo de los aproximadamente insesgados, en contraposición a los estimadores basados en modelos, que presentan sesgos difíciles de valorar.

La segunda reflexión se refiere a la precisión de

las estimaciones, medida por el coeficiente de variación (CV). En dos de las variables objetivo, ocupados e inactivos, se observa una clara mejora de la precisión con el uso de información auxiliar, mientras que en la tercera, parados, permanece prácticamente inalterada. Por ello se puede deducir, con todas las limitaciones lógicas de un trabajo como el presente, que el uso de estimadores que utilizan información auxiliar, en general aporta ganancia en la precisión de algunas de las estimaciones, y difícilmente empeora la situación de partida. Este hecho es relevante, pues es habitual que las encuestas sean multiobjetivo, y por ello no es fácil encontrar un conjunto de información auxiliar que sea idóneo para todos ellos.

En la tabla 2, se ha considerado el RMSE debido al evidente sesgo en el estimador directo (H-T) causado por la falta de respuesta. El estimador de razón corrige completamente el sesgo en la variable paro, y lo reduce considerablemente en inactividad y ocupación. El estimador calibrado proporciona estimaciones insesgadas, lo cual era de esperar dado que una de las variables de calibrado, la nacionalidad española o extranjera, intervino junto con el tamaño de la vivienda en la simulación de falta de respuesta. En cuanto a la precisión, en todos los casos es menor que la que se obtuvo sin falta de respuesta, aunque se mantiene el mismo orden de los estimadores en cuanto a su variabilidad.

## 7. Conclusiones

La mejora en la calidad y facilidad de acceso a información auxiliar procedente de fuentes administrativas, unida al desarrollo de procedimientos informáticos, permite aplicar, de forma relativamente sencilla, técnicas de calibrado, que en general mejoran la precisión de las estimaciones de una encuesta, ayudan a corregir el sesgo introducido por la falta de respuesta, y proporcionan consistencia entre las cifras presentadas y las procedentes de otras fuentes.

No obstante, es necesario mantener un control riguroso de la información auxiliar que se considere, para evitar posibles sesgos derivados de la utilización de datos externos inadecuados.

## Referencias

- [1] Deming W.E. and Stephan F.F. (1940). On a least squares adjustment of sampled frequency table when the expected marginal totals are known. *Annals of mathematical statistics*, **11**, 427-444.
- [2] Deville, J. C., Särndal, C. E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, **87**(418), 376-382.
- [3] Deville, J. C., Särndal, C. E., Sautory, O. (1993). Generalised Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, **88**(423), 1013-1020.
- [4] Lemaitre G. and Dufour, J. (1987). An integrated method for weighting persons and families. *Survey Methodology*, **13**, 199-207.
- [5] Särndal, C.E., Swenson, Wretman. (1992). *Model Assisted Survey Sampling*. Springer-Verlag.
- [6] Särndal, C.E. (2007). *The calibration approach in survey theory and practice*. *Survey Methodology*, **33**(2), 99-119.
- [7] Vanderhoeft C. (2001). Generalised Calibration at Statistics Belgium. SPSS Module g-Calib-S and Current Practices. (Descarga gratuita en: <http://www.statbel.fgov.be/studies/paper03.pdf>).