

CALCULO DE FACTORES DE EXPANSIÓN

**ENCUESTA NACIONAL SOBRE NIVELES DE VIDA EN LOS
HOGARES**

Índice

Página

1. Factor de expansión 2002	1
1.1 Factor de expansión a nivel vivienda	1
1.1.1 Ajuste a los factores de expansión	3
1.2 Factor de expansión a nivel persona	5
1.2.1 Factor Proxy	5
2. Factor de expansión 2005	6
2.1 Factor de seguimiento longitudinal	6
2.1.1 Factor de seguimiento longitudinal a nivel vivienda	6
2.1.2 Factor de seguimiento longitudinal a nivel persona	6
2.1.3 Proxy de seguimiento longitudinal	7
2.2 Factor de expansión puntual	7
2.2.1 Factor de expansión puntual a nivel vivienda	7
2.2.2 Factor de expansión puntual a nivel persona	8
2.2.3 Factor de expansión puntual proxy	8
3. Factor de expansión 2010	8

**CALCULO DE FACTORES DE EXPANSIÓN
ENCUESTA NACIONAL SOBRE NIVELES DE VIDA EN LOS HOGARES**

1. Factor de expansión 2002

La Encuesta Nacional sobre Niveles de Vida en los Hogares (ENNVH) emplea dos tipos de factores de expansión, el factor vivienda y el factor persona. El procedimiento del cálculo del factor vivienda, no es más que el inverso de la probabilidad conjunta de seleccionar la unidad última de muestreo que es la vivienda, el factor vivienda lleva implícito tres tipos de ajustes, uno es el ajuste por No respuesta debido a la negativa de los residentes de las viviendas de no contestar la encuesta, segundo se debe al ajuste a la proyección de población y el tercero se debe al ajuste por técnicas de calibración. Una vez realizado estos ajustes al factor vivienda se procedió a calcular los factores de expansión a nivel persona para cada uno de los libros, así como el factor del libro proxy a nivel persona.

1.1 Factor de expansión a nivel vivienda

El marco de muestreo de la ENNVH empleo dos tipos de muestreo, el muestreo trietápico y muestreo bietápico, el muestreo trietápico se empleo en las 48 ciudades autorrepresentadas de ENEU y el muestreo bietápico se empleo en el resto urbano y en el rural, el procedimiento de selección de la muestra para cada tipo de muestreo y dominio se presenta a continuación.

- **Ciudades autorrepresentadas**

En el dominio ciudades autorrepresentadas la selección de la muestra se realizó en forma independiente para cada región y estrato, el procedimiento es el siguiente:

1. De las t_r entidades de cada región se seleccionaron t_r^* entidades con igual probabilidad.
2. Dentro de cada región y estrato se seleccionaron n_{rh} UPM con igual probabilidad.
3. En cada UPM seleccionada se seleccionaron 6 USM con probabilidad proporcional al tamaño.
4. En cada USM seleccionada se seleccionaron 'v' viviendas para la ENNVH, con igual probabilidad.

La probabilidad de seleccionar una vivienda de la j-ésima USM, de la i-ésima UPM, del h-ésimo estrato, de la, de la r-ésima región, está dada por la siguiente expresión:

$$P_{v_{rhij}} = \frac{t_r^*}{t_r} \cdot \frac{n_{rh}}{V_{rh}} \cdot \frac{6 V_{rhij}}{V_{rh}} \cdot \frac{v}{V_{rhij}^*} = \frac{6 v t_r^* n_{rh} V_{rhij}}{t_r V_{rh} V_{rhij}^*}$$

Su factor de expansión está dado por:

$$F_{rhij} = \frac{t_r V_{rh} V_{rhij}^*}{6 v t_r^* n_{rh} V_{rhij}}$$

Donde:

- P_{rhi}^j = es la probabilidad de seleccionar una vivienda de la j-ésima USM, de la i-ésima UPM, del h-ésimo estrato, de la r-ésima región.
- n_{rh} = es el número de UPM seleccionadas en el h-ésimo estrato, en la r-ésima región para la ENNVH.
- V_{rhi} = es el número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la r-ésima región, según el Censo de Población y Vivienda 1995.
- v = es el número de viviendas seleccionadas en cada USM para la ENNVH.
- V_{rhij}^* = es el total de viviendas en la j-ésima USM, en la i-ésima UPM, en el h-ésimo estrato, en la r-ésima región, según el listado de viviendas.
- V_{rh} = es el número de viviendas en el h-ésimo estrato, en la r-ésima región, según el Censo de Población y Vivienda 1995.
- V_{rhij} = es el número de viviendas en la j-ésima USM, en la i-ésima UPM, en el h-ésimo estrato, en la r-ésima región, según el Censo de Población y Vivienda 1995.
- t_r = es el total de entidades por región.
- t_r^* = es el número de entidades seleccionadas para la ENNVH.

• **Resto urbano y rural**

1. De las t_r entidades por región se seleccionaron t_r^* con igual probabilidad.
2. Al interior de cada región y estrato se seleccionaron n_{rh} UPM con igual probabilidad.
3. En cada UPM seleccionada se seleccionaron “v” viviendas con igual probabilidad.

La probabilidad de seleccionar una vivienda está dada por la siguiente expresión:

$$P_{rhi}^j = \frac{t_r^*}{t_r} \cdot \frac{n_{rh} V_{rhi}}{V_{rh}} \cdot \frac{v}{V_{rhij}^*} = \frac{v t_r^* n_{rh} V_{rhi}}{t_r V_{rh} V_{rhij}^*}$$

Su factor de expansión está dado por:

$$F_{rhi} = \frac{t_r V_{rh} V_{rhij}^*}{v t_r^* n_{rh} V_{rhi}}$$

Donde:

- n_{rh} = es el número de UPM seleccionadas en el h-ésimo estrato, en la r-ésima región para la ENNVH.
- V_{rhi} = es el número de viviendas, en la i-ésima UPM, en el h-ésimo estrato, en la r-ésima región, según el Censo de Población y Vivienda 1995.
- V_{rh} = es el número de viviendas en el h-ésimo estrato, en la r-ésima región, según el Censo de Población y Vivienda 1995.
- v = es el número de viviendas seleccionadas en cada UPM, para la ENNVH.
- V_{rhij}^* = es el número de viviendas, en la i-ésima UPM, en el h-ésimo estrato, en la r-ésima región, según el listado de viviendas.
- t_r = es el total de entidades por región.
- t_r^* = es el número de entidades seleccionadas para la ENNVH.

1.1.1 Ajuste a los factores de expansión

Los factores de expansión elaborados conforme al procedimiento antes descrito, se ajustan para los siguientes conceptos:

a) Ajuste por No respuesta

El ajuste por No Respuesta atribuida al informante se realiza a nivel estrato tanto en áreas autorrepresentadas como en el resto urbano y rural, mediante la siguiente expresión:

$$F'_{rdh} = F_{rdh} \frac{V'_{rdh} + V''_{rdh}}{V'_{rdh}}, \quad d = 1,2,3.$$

Donde:

- F'_{rdh} = factor de expansión corregido por No respuesta en el h-ésimo estrato, en el d-ésimo dominio, de la r-ésima región.
- F_{rdh} = factor de expansión a nivel vivienda en el h-ésimo estrato, en el d-ésimo dominio, de la r-ésima región.
- V'_{rdh} = número de viviendas seleccionadas habitadas con respuesta en el h-ésimo estrato, en el d-ésimo dominio, en la r-ésima región.
- V''_{rdh} = número de viviendas seleccionadas habitadas sin respuesta en el h-ésimo estrato, en el d-ésimo dominio, en la r-ésima región.
- d = dominio (áreas autorrepresentadas=1, resto urbano=2,rural=3).

a) Ajuste por proyección

Los factores ajustados por No respuesta se corrigen a fin de asegurar, que en cada dominio de interés de la encuesta se obtenga la población total; determinada por la proyección de población generada por el CONAPO e INEGI, referida al punto medio del levantamiento mediante la siguiente expresión:

$$F''_d = F'_d \frac{\text{PROY}_d}{\text{PEXP}_d}$$

Donde:

- F''_d = factor de expansión corregido por proyección en el dominio d.
- F'_d = factor de expansión corregido por No respuesta en el dominio d.
- PROY_d = población en el dominio d, según la proyección.
- PEXP_d = población total a la que expande la encuesta en el dominio d.

b) Ajuste por calibración

Al ser la encuesta un estudio estadístico de tipo longitudinal, con un levantamiento de arranque en el primer semestre de 2002, y con dos levantamientos de seguimiento en el tercer trimestre de 2005 y de 2010. Y considerando que la población objetivo presenta un dinamismo demográfico sobre todo en cuanto a movilidad territorial y de migración internacional. Con estas consideraciones, a pesar del ajuste

de factores de expansión con las proyecciones CONAPO-INEGI, no es posible conservar la estructura demográfica final, por lo que es necesario un ajuste adicional a nivel grupo de edad y sexo. Estas adecuaciones adicionales sólo son factibles con técnicas de calibración y la información auxiliar utilizada proveniente de fuentes externas como censos de población, registros administrativos u otras encuestas, la cual se le conoce como variables de calibración.

Formalmente la técnica de calibración consiste en lo siguiente:

Sean x_1, \dots, x_k variables auxiliares cuyos totales poblacionales X_1, \dots, X_k conocemos.

Sean d_1, \dots, d_n los factores de expansión originales de los elementos de la muestra de $k = 1, \dots, n$.

Los w_k son los factores calibrados y se espera que están muy cercanos a los originales.

El problema general de calibración consiste en minimizar $\sum_{k=1}^n d_k F\left(\frac{w_k}{d_k}\right)$ (1)

sujeta a p restricciones de calibración $\sum_{k=1}^n w_k x_{kj} = t_j \quad j = 1, \dots, p$.

La función de distancia F mide la diferencia entre $g_k = \frac{w_k}{d_k}$ y 1.

Esta función debe cumplir las siguientes condiciones de regularidad:

- F es estrictamente convexa y con segunda derivadas continuas en el interior de su dominio.
- F y $F' \geq 0$.
- $F'(0) = 0$.
- F' es una función uno a uno, por tanto F' es estrictamente creciente.
- La inversa de la función F' es llamada función de calibración $g = F'^{-1}$.

Para este ajuste se consideró la función distancia $F(x) = x \ln(x) - x + 1$ conocida como Raking Ratio propuesta por Deville y Sarda (1992).

En la práctica lo que se hace es encontrar los g_k factores lo más cercano posible a 1, y que minimicen la ecuación (1). La restricción para obtener los nuevos w_k , es que cumplan con $t_j = \sum w_k x_{kj}$, donde t_j son los valores de los totales oficiales de la variable de interés.

Finalmente, minimizan la función de distancia, con la restricción t_j la cual se resuelve por medio de multiplicadores de Lagrange:

$$L(w, \lambda) = \sum_{k=1}^n d_k F\left(\frac{w_k}{d_k}\right) - \lambda^T \left(\sum_{k=1}^n w_k x_{kj} - t_j\right)$$

con $\lambda^T = (\lambda_1, \dots, \lambda_p)^T$

En el caso de la ENNVH se tomaron como variables de calibración las siguientes:

1. Población total
2. Población de 15 y más años
3. Población menor de 15 años
4. Población de 5 a 12 años

5. Población de 13 a 65 años
6. Mujeres de 14 a 49 años

Esta información se obtuvo de la ENE del primer semestre del 2002, y de la ENOE tercer trimestre del 2005 y 2010.

La técnica de calibración está disponible en el Software SAS, macro CALMAR, la función Calib en R y la función g-calib en SPSS.

1.2 Factor de expansión a nivel persona

El factor vivienda calibrado junto con la información del libro C se empleó para generar los factores de expansión a nivel persona de cada libro de la encuesta, el procedimiento de generación de los factores de expansión es el siguiente:

$$F_{rd}^k = F_{rd} \frac{m_{rd}}{m_{rd}^*}$$

Donde:

- k = es el identificador del libro de interés, como se describe en el cuadro 1.
- F_{rd} = factor de expansión del libro C a nivel vivienda calibrado en el d-ésimo dominio, de la r-ésima región.
- m_{rd} = número de personas en el libro C por grupo de edad en el d-ésimo dominio, en la r-ésima región.
- m_{rd}^* = número de personas que contestaron el libro k en el d-ésimo dominio, en la r-ésima región.

Cuadro 1.

k	Libro	Rango de edad
1	3a	15 y más años
2	3b	15 y mas años
3	4	14 a 49 años
4	5	0 a 14 años
5	ea	13 a 65 años
6	en	5 a 12 años

- d = se refiere a que puede ser uno de los siguientes tres dominios: áreas metropolitanas (1), complemento urbano (2) o rural (3).
- r = se refiere que puede pertenecer a cualquiera de las cinco regiones.

1.2.1 Factor Proxy

El cálculo del factor de expansión del libro Proxy, se obtuvo considerando como base el factor vivienda calibrada y la información del libro C, el procedimiento es el siguiente:

$$F_{rd}^P = F_{rd} \frac{m_{rd}}{m'_{rd} + m''_{rd}}$$

Donde:

- k = es el identificador del libro de interés como se describe en el cuadro 1.

- F_{rd} = factor de expansión calibrado del libro C a nivel vivienda en el d-ésimo dominio, de la r-ésima región.
- m_{rd} = número de personas en el libro C por grupo de edad en el d-ésimo dominio, en la r-ésima región.
- m'_{rd} = número de personas que contestaron el libro k en el d-ésimo dominio, en la r-ésima región.
- m''_{rd} = número de personas que contestaron el libro k como informantes indirectos en el d-ésimo dominio, en la r-ésima región.

2. Factor de Expansión 2005

La ENNVH en el 2005 dio seguimiento a la muestra seleccionada en el 2002, tanto a los hogares como a las personas, también incorpora a la nueva población que entró a la muestra, como nuevos nacimientos o porque se hayan unido como parejas a los seleccionados en el 2002, por esta razón es necesario elaborar factores de expansión de seguimiento longitudinales aplicables solo a la muestra seleccionada en el 2002, además de otro factor nuevo o puntual que estime la población total en el momento del levantamiento de la encuesta.

2.1 Factor de seguimiento longitudinal

Los factores de seguimiento longitudinal solo se aplican a los hogares y las personas seleccionadas en el 2002 y que se encontraron en este segundo levantamiento de la encuesta.

2.1.1 Factor de seguimiento longitudinal a nivel vivienda

Los factores de seguimiento longitudinal se generaron a partir de los factores vivienda del 2002 ajustados por No respuesta, la expresión para generar los factores de expansión es:

$$F_{rd}^* = F_{rd} \frac{V_{rd}^* + V_{rd}^{**}}{V_{rd}^*}$$

Donde:

- F_{rd}^* = factor expansión a nivel vivienda del 2005 ajustado por No respuesta en el d-ésimo dominio, de la r-ésima región.
- F_{rd} = factor de expansión a nivel vivienda del 2002 en el d-ésimo dominio, en la r-ésima región.
- V_{rd}^* = número de viviendas seleccionadas habitadas con respuesta en el 2005 en el d-ésimo dominio, en la r-ésima región.
- V_{rd}^{**} = número de viviendas seleccionadas habitadas sin respuesta en el 2005 en el d-ésimo dominio, de la r-ésima región.

2.1.2 Factor de seguimiento longitudinal a nivel persona

El factor vivienda longitudinal ajustado por No repuesta y la información del libro C, son la base para generar los factores a nivel persona de cada libro.

$$F_{rd}^{\leftarrow} = F_{rd} \frac{m_{rd}}{m'_{rd}}$$

Donde:

- K = es el identificador del libro de interés, como se describe en el cuadro 1.
- F_{rd} = factor de expansión a nivel vivienda del libro C en el d-ésimo dominio, de la r-ésima región.
- m_{rd} = número de personas en el libro C por grupo de edad en el d-ésimo dominio, en la r-ésima región.
- m^*_{rd} = número de personas por grupo de edad en el libro C que contestaron el libro k en el d-ésimo dominio, en la r-ésima región.

2.1.3 Proxy de Seguimiento longitudinal

El cálculo del factor Proxy se realizó tomando de base al factor vivienda longitudinal ajustado por No respuesta y la información del libro C.

$$F^p_{rd} = F_{rd} \frac{m_{rd}}{m'_{rd} + m''_{rd}}$$

Donde:

- F_{rd} = factor de expansión a nivel vivienda del libro C en el d-ésimo dominio, de la r-ésima región.
- m_{rd} = número de personas del libro C por grupo de edad en el d-ésimo dominio, en la r-ésima región.
- m'_{rd} = número de personas que contestaron el libro k en el d-ésimo dominio, en la r-ésima región.
- m''_{rd} = número de personas que contestaron el libro k como informantes indirectos en el d-ésimo dominio, en la r-ésima región.

2.2 Factor de expansión puntual

El factor de expansión puntual a nivel vivienda y el factor persona, se obtuvieron a partir del factor corregido por No respuesta descrito en el inciso 2.1.1, a continuación se describe el procedimiento del cálculo de cada uno de los factores de expansión.

2.2.1 Factor de expansión puntual a nivel vivienda

Para obtener el factor puntual a nivel vivienda y que incorpore a la nueva población en el 2005, el inciso es necesario proyectar el factor corregido por No respuesta de acuerdo con el procedimiento descrito en el inciso 1.1.1. b). La proyección de la población se situó en el punto medio del tercer trimestre de 2005. Una vez generado el factor a nivel vivienda y ajustado por proyección de población, se realizó el ajuste por calibración siguiendo el procedimiento descrito en el inciso 1.1.1 c).

2.2.2 Factor de expansión puntual a nivel persona

El factor de expansión a nivel persona se generó a partir del factor vivienda calibrado empleando el procedimiento descrito en el inciso 1.2.

2.2.3 Factor de expansión puntual proxy

El factor de los libros proxy se generó aplicando el mismo procedimiento descrito en el inciso 1.2.1.

3. Factor de expansión 2010

Los factores de expansión longitudinales de seguimiento y puntuales, se obtuvieron empleando el mismo procedimiento que se utilizó para generar los factores de expansión del 2005, la única diferencia es la proyección de población empleada para los factores de expansión puntuales situada en el punto medio del tercer trimestre de 2010.

Bibliografía

1. J.C. Deville y C.E. Sarndal. "Calibration Estimators in Survey", Junio 1992.
2. C.E. Sarndal "The Calibration approach in survey theory and practice", December 2007.
3. Camille Venderhaeft "Generalised Calibration at Statistics Belgium", Mayo 2001.
4. Mónica Tinajero Bravo y Guillermina Eslava Gómez. "Calibración en muestreo: Una aplicación a la Encuesta Nacional de Ingreso y Gastos en los Hogares 1992 y 1996. Series Monográficas. Junio del 2000.