CALCULATION OF EXPANSION FACTORS

MEXICAN FAMILY LIFE SURVEY

Index   **Page**

CALCULATION OF EXPANSION FACTORS
MEXICAN FAMILY LIFE SURVEY

1. Expansion Factors 2002

The Mexican Family Life Survey (MxFLS) has two types of expansion factors, at the household level and at the individual level. The calculation of expansion factors at the household level is the inverse of the joint probability of selecting the last sampling unit (a household). The expansion factors at the household level imply three types of adjustments. The first one is related with non-response (given that some households did not want to answer the survey); the second one corresponds to the projection of sample to the entire population, and the third one to calibration techniques. Once household level factors are adjusted, individual expansion factors for each book and for the Proxy Book are calculated.

1.1 Expansion factors at the household level

The sampling scheme of the MxFLS is based on two types of sampling, a three-stage and a two-stage sampling. The three-stage sampling is used in 48 self-represented cities of the National Survey of Urban Employment (ENEU, per its name in Spanish). The two-stage sampling is used in the remaining urban areas and in rural areas. Samples are selected as follows:

- **Self-represented cities**

In the self-represented cities domain, sample selection is independent in each region and stratum:

1. Of the t_r entities of each region, t_r^* entities are selected with equal probability.
2. Within each region and stratum, n_{rh} Sample Primary Units (SPU) are selected with equal probability.
3. In each selected SPU, 6 USM are selected with probability proportional to size.
4. In each selected USM, 'v' households are selected for the MxFLS, with equal probability.

The probability of selecting a household in the jth USM, in the ith SPU, in the hth stratum in the rth region is:

$$P\{v_{rhij}\} = \frac{t_r^*}{t_r} \frac{n_{rh}V_{rhi}}{V_{rh}} \frac{6\,V_{rhij}}{V_{rhi}} \frac{v}{V_{rhij}^*} = \frac{6\,v\,t_r^*n_{rh}V_{rhij}}{t_r V_{rh}V_{rhij}^*}$$

The expansion factor is:

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

Where:

$P\{v_{rhij}\}$ = Probability of selecting a household in the jth USM, in the ith SPU, in the hth stratum, in the rth

region.

n_{rh} = Number of selected SPU in the hth stratum, in the rth region for the MxFLS.

V_{rhi} = Number of households in the ith SPU, in the hth stratum, in the rth region, according to the 1995 Mexican Count of Population and Housing[1].

v = Number of selected households in each USM for the MxFLS.

V_{rhij}^{*} = Total number of households in the jth USM, in the ith SPU, in the hth stratum, in the rth region, according to the listing of households.

V_{rh} = Number of households in the hth stratum, in the rth region, according to the 1995 Mexican Count of Population and Housing.

V_{rhij} = Number of households in the jth USM, in the ith SPU, in the hth stratum, in the rth region, according to the 1995 Mexican Count of Population and Housing.

t_r = Number of entities by region.

t_r^{*} = Number of entities selected for the MxFLS.

- **Remaining urban and rural areas**

1. Of the t_r entities per region, t_r^{*} are selected with equal probability.
2. Within each region and stratum, n_{rh} SPU are selected with equal probability.
3. In each selected SPU, "v" households are selected with equal probability.

The following is the probability of selecting a household:

$$P\{v_{rhi}\} = \frac{t_r^{*}}{t_r}\,\frac{n_{rh}V_{rhi}}{V_{rh}}\,\frac{v}{v_{rhi}^{*}} = \frac{v\,t_r^{*}n_{rh}V_{rhi}}{t_r V_{rh}V_{rhi}^{*}}$$

The expansion factor is:

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

Where:

n_{rh} = Number of selected SPU in the hth stratum, in the rth region for the MxFLS.

V_{rhi} = Number of households in the ith SPU, in the hth stratum, in the rth region, according to the 1995 Mexican Count of Population and Housing.

V_{rh} = Number of households in the hth stratum, in the rth region, according to the 1995 Mexican Count of Population and Housing.

v = Number of households selected in each SPU for the MxFLS.

V_{rhi}^{*} = Number of households in the ith SPU, in the hth stratum, in the rth region, according to the listing of households.

t_r = Number of entities by region.

t_r^{*} = Number of entities selected for the MxFLS.

---

[1] In Spanish, *Conteo de Poblacion y Vivienda 1995.*

1.1.1 Adjustment to Expansion Factors

The expansion factors mentioned above are adjusted for the following reasons:

a) Adjustment for NON-response

The adjustment for NON-response attributed to the informant is applied at the stratum level in self-represented areas in the remaining urban and rural areas, as follows:

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

Where:

F'_{rdh} = Expansion factor adjusted for NON-response in the hth stratum, in the dth domain, in the rth region.

F_{rdh} = Expansion factor at the household level in the hth stratum, in the dth domain, in the rth region.

v'_{rdh} = Number of selected occupied households with response in the hth stratum, in the dth domain, in the rth region.

v''_{rdh} = Number of selected occupied households with no response in the hth stratum, in the dth domain, in the rth region.

d = Domain (self-reported areas=1, remaining urban=2, rural=3).

a) Adjustment for projection

The adjusted expansion factors for NON-response are corrected in order to guarantee that the total population of the MxFLS is obtained in each domain of the survey. This is determined by the projection of the population generated by National Council Population (CONAPO, per its name in Spanish) and the National Institute of Geography and Statistics (**INEGI**, per its name in Spanish**)**. The reference is the middle point of data collection, as follows:

$$F''_d = F'_d \frac{PROY_d}{PEXP_d}$$

Where:

F''_d = Expansion factor adjusted due to projection in the d domain.

F'_d = Expansion factor adjusted due to NON-response in the d domain.

$PROY_d$ = Population in the d domain, according to the projection.

$PEXP_d$ = Total population to which the survey is expanded in the d domain.

b) Adjustment for calibration

Given the longitudinal nature of the MxFLS, and considering that the baseline survey was conducted in the first semester of 2002 and the two follow-up surveys were conducted in the third trimester of 2005 and 2010, and that the targeted population migrates (within the country and abroad), it is not possible to

maintain the demographic structure—despite the adjustments to the expansion factors based on the projections of CONAPO and INEGI. Therefore, it is necessary to apply a final adjustment by groups of age and gender. These final adjustments are applied based on calibration techniques that use external information from population census, administrative records, and other surveys to construct calibration variables.

The calibration technique consists on the following:

Given that $x_1,...,x_k$ are auxiliary variables with known population totals $X_1,...,X_k$.

Given the original expansion factors $d_1,...,d_n$ of the elements in the sample $k=1,...,n$.

The w_k are the calibrated factors, which are expected to be very close to the original factors.

The general calibration problem consists on minimizing $\sum_{k=1}^{n} d_k F\left(\dfrac{w_k}{d_k}\right)$ 　　　　　　(1)

given p calibration restrictions $\sum_{k=1}^{n} w_k x_{kj} = t_j$ 　　　 $(j=1,...,p)$.

The distance function F measures the differences between $g_k = \dfrac{w_k}{d_k}$ and 1.

This function should comply with the following constraints:

- $F(\bullet)$ is strictly convex and with second derivatives continuous within their domain.
- $F(\bullet)$ And $F(\bullet) \geq 0$.
- $F'(1) = 0$.
- F' is a one to one function, therefore F' is strictly increasing.
- The inverse of function F' is called calibration function $g(\bullet) = F'^{-1}(\bullet)$

The distance function $F(x) = x\ln(x) - x + 1$ known as Raking Ratio—proposed by Deville and Sardnal (1992) —is considered for these adjustments.

In practice, we find g_k factors as close as possible to 1 and that minimize equation (1). The restriction to obtain the new w_k is $t_j = \sum w_k x_{kj}$, where t_j are the official total values of the variable of interest.

Finally, we minimize the distance function with restriction t_j, which can be solved based on Lagrange multipliers:

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

with 　　　 $\lambda^{T} = (\lambda_1,...,\lambda_P)^{T}$

The following calibration variables are used in the MxFLS:

1. Total population
2. Population aged 15 years and older
3. Population younger than 15 years
4. Population aged 5 to 12 years

5. Population aged 13 to 65 years
6. Women aged 14 to 49 years

This information is obtained from the National Employment Survey (ENE, per its name in Spanish) conducted in the first semester of 2002, and from the National Survey of Occupation and Employment (ENOE, per its name in Spanish) of the third trimester of 2005 and 2010.

The calibration technique is available in the Software SAS, macro CALMAR, the command Calib in R and the function g-calib in SPSS.

1.2 Expansion factors at the individual level

The calibrated household factor and the information from Book C is used to generate the expansion factors at the individual level for each Book of the MxFLS, as follows:

$$F(k)_{rd} = F_{rd} \frac{m_{rd}}{m_{rd}^*}$$

Where:

k = id of the Book of interest, as described in Table 1.
F_{rd} = Expansion factor of Book C at the household level calibrated in the dth domain, in the rth region.
m_{rd} = Number of people in Book C by group of age in the dth domain, in the rth region.
m_{rd}^* = Number of people who answered Book k in the dth domain, in the rth region.

Table 1.

k	Book	Age Range
1	3a	15 years and older
2	3b	15 years and older
3	4	14 to 49 years
4	5	0 to 14 years
5	ea	13 to 65 years
6	en	5 to 12 years

d = can be one of the following domains: metropolitan areas (1), urban complement (2) or rural (3).
r = can belong to any of the five regions.

1.2.1 Factor Proxy

The calculation of the Proxy Book expansion factor is obtained based on the calibrated household factor and information from Book C:

$$F^P(k)_{rd} = F_{rd} \frac{m_{rd}}{m_{rd}' + m_{rd}''}$$

Where:

k   =   id of the Book of interest, as described in Table 1.

F_{rd} =   Book C household level expansion factor calibrated in the dth domain, in the rth region.

m_{rd} =   Number of people in Book C by group of age in the dth domain, in the rth region.

m'_{rd} =   Number of people who answered Book k in the dth domain, in the rth region.

m''_{rd} =   Number of people who answered Book k as indirect informants in the dth domain, in the rth region.

2. Expansion factors 2005

In 2005, the MxFLS followed the sample selected in 2002 (households and people), and interviewed the population that was incorporated to that sample, like offspring and new partners. For this reason, it is necessary to have follow-up longitudinal expansion factors applicable to the sample selected in 2002 in addition to factors that estimates the total population at the moment of the survey.

2.1 Follow-up longitudinal factor

The follow-up longitudinal factors are applicable only to households and individuals that were selected in the 2002 sample and re-interviewed in 2005.

2.1.1 Follow-up longitudinal factor at the household level

The follow-up longitudinal factors are generated based on the 2002 household level expansion factors adjusted for NON-response:

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

Where:

F_{rd}^{*} =   2005 household level expansion factor adjusted for non-response in the dth domain, in the rth region.

F_{rd} =   2002 household level expansion factor in the dth domain, in the rth region.

V_{rd}^{*} =   Number of occupied selected households with response in 2005 in the dth domain, in the rth region.

V_{rd}^{**} =   Number of occupied selected households with non-response in 2005 in the dth domain, in the rth region.

2.1.2 Follow-up longitudinal factor at the individual level

The follow up longitudinal factors at the individual level for each Book are based on the household level longitudinal factor adjusted for NON-response and on the information from Book C.

$$F(k)_{rd} = F_{rd} \frac{m_{rd}}{m_{rd}^{*}}$$

6

Where:

K = id of the Book of interest, as described in Table 1.

F_{rd} = Household level expansion factor of Book C in the dth domain, in the rth region.

m_{rd} = Number of people in Book C by group of age in the dth domain, in the rth region.

m^{*}_{rd} = Number of people by group of age in Book C that answered Book k in the dth domain, in the rth region.

2.1.3 Follow-up longitudinal Proxy factor

The calculation of the Proxy factor is based on the household longitudinal factor adjusted for NON-response and on the information from Book C:

$$F^{p}(k)_{rd} = F_{rd} \; \frac{m_{rd}}{m'_{rd} + m''_{rd}}$$

Where:

F_{rd} = Household level expansion factor of Book C in the dth domain, in the rth region.

m_{rd} = Number of people in Book C by group of age in the dth domain, in the rth region.

m'_{rd} = Number of people who answered Book k in the dth domain, in the rth region.

m''_{rd} = Number of people who answered Book k as indirect informants in the dth domain, in the rth region.

2.2 Wave specific expansion factors

The punctual expansion factors at the household level and at the individual level are obtained based on the adjusted factor for NO response described in 2.1.1. The following is the calculation process for each expansion factor:

2.2.1 Wave specific household level expansion factor

In order to obtain the punctual factor at the household level that incorporates the new population in 2005 is necessary to project the adjusted factor for NO response according to the procedure described in 1.1.1. b). The projection of the population is situated in the middle point of the third trimester of 2005. Once the factor at the household level is generated and adjusted for the projection of the population, the adjustment for calibration is done following the procedure described in 1.1.1 c).

2.2.2 Wave specific individual level expansion factor

The individual level expansion factor is generated based on the calibrated household factor using the procedure described in 1.2.

2.2.3 Wave specific expansion Proxy factor

The Proxy Book expansion factor is generated based on the procedure described in 1.2.1.

3. Expansion factor 2010

The follow-up longitudinal expansion factors and the wave specific expansion factors are obtained based on the same procedure used to generate the expansion factors in 2005. The only difference is the projection of the population used in the wave specific expansion factors, which is based on the middle point of the third trimester of 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.