INEGI
Dirección General de Integración, Análisis e Investigación
Dirección General Adjunta de Investigación

MODELO ESTADÍSTICO 2015 PARA LA CONTINUIDAD DEL MCS-ENIGH

NOTA TÉCNICA

INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Antecedentes

En el Módulo de Condiciones Socioeconómicas (MCS) 2015 se observaron circunstancias fuera de su tendencia histórica. Por una parte, se tuvo un aumento por encima de lo esperado en los ingresos de los hogares como consecuencia de mejoras operativas, y por la otra se registró una disminución por debajo de lo esperado en el tamaño promedio de los hogares.

Para recobrar la consistencia histórica, el Modelo Estadístico 2015 para la Continuidad del MCS-ENIGH contiene un ajuste a los ingresos reportados en el MCS 2015, que parte de una versión de los datos corregidos demográficamente.[1]

Modelo estadístico de ajuste a los ingresos

Un gran reto para cualquier ejercicio que se proponga mantener la continuidad histórica del MCS-ENIGH es la selección de elementos de referencia para ajustar el ingreso originalmente reportado. Podemos considerar algunos con base en relaciones propias del MCS-ENIGH, o bien, con base en fuentes externas que no hayan sido afectadas por las mejoras realizadas en el MCS 2015. La metodología que aquí se emplea forma parte del segundo tipo, tomando como fuente externa la Encuesta Nacional de Ocupación y Empleo (ENOE).

A pesar de la diversidad de fuentes de ingreso de los hogares mexicanos, la mayor parte proviene de su ingreso laboral, es decir, de su trabajo como asalariado y/o independiente, que en conjunto han representado cerca del 70 % del ingreso corriente total desde 2010. Este elevado porcentaje, junto con la existencia de un levantamiento regular sobre condiciones laborales como lo es la ENOE, crean una ventana de oportunidad para utilizar el ingreso laboral que reporta la ENOE como ancla para ajustar los ingresos laborales.[2]

Entre las ventajas de usar la ENOE como referencia, podemos enumerar las siguientes:

1) Considera el mismo marco muestral que el MCS-ENIGH.
2) La ENOE es la encuesta continua en hogares más grande de que se dispone en México, con 120 000 hogares al trimestre.
3) La ENOE no fue afectada por los cambios endógenos (capacitación, control del operativo, etc.). En este sentido, los cambios en el ingreso de la ENOE reflejan cambios que son independientes de las mejoras hechas en el MCS 2015.
4) Al igual que el MCS-ENIGH, y si bien no es su propósito principal, la ENOE capta los ingresos provenientes tanto de actividades formales como de informales.
5) El diseño de la ENOE en paneles rotatorios estabiliza la muestra y robustece las comparaciones en el tiempo.
6) La ENOE es sensible a las diferencias del comportamiento de los ingresos entre los estados, además de que es sensible a los efectos del entorno económico y de políticas públicas a nivel nacional.
7) CONEVAL utiliza la ENOE para la elaboración de sus indicadores coyunturales de la pobreza laboral en México.

[1] La corrección demográfica se explica en el Anexo 2.

[2] Salvo en el primer trimestre del año, el ingreso laboral de la ENOE hace referencia al correspondiente originado por el trabajo principal.

8) ENOE es menos volátil que el MCS-ENIGH respecto del trabajo independiente (patrones y cuentas propias).
9) La ENOE es consistente a nivel entidad federativa.

Si bien la ENOE nos ayuda a modificar los ingresos laborales, la trayectoria de los ingresos no laborales puede tener un comportamiento diferente a los primeros y, por tanto, requerir de un ajuste distinto. Para lograr esta diferenciación, hacemos uso de una regularidad empírica entre el ingreso no laboral y el ingreso total, que se ha venido observando desde 2010 en el MCS-ENIGH.

Así, la metodología en su conjunto afecta de manera secuencial el ingreso total de cada hogar: en una primera fase ajusta lo correspondiente al ingreso por trabajo principal (ITP), definido en el Anexo 1, y en una fase posterior, a todo aquel ingreso distinto al ITP (que llamamos ITP^c).

Adicionalmente al ajuste del ingreso, y como se menciona en el Anexo 2, en el Módulo de Condiciones Sociodemográficas (MCS) 2015 se observó un aumento inusual en el número de hogares, lo que, para un determinado volumen de la población, implica una contracción mayor a la esperada del tamaño promedio de los hogares. Para corregir esta situación, se realizó una imputación de individuos dentro de la base de datos original y se calibraron los factores de expansión para ajustar a los totales poblacionales nacionales.

La metodología del Modelo Estadístico 2015 ajusta los ingresos de los hogares de acuerdo con las fases que se describen enseguida y partiendo de una base de datos corregida demográficamente.

Fase 1. Ajuste al Ingreso por Trabajo Principal (ITP)

Para el ajuste del ITP tomamos como punto de referencia la trayectoria observada de los ingresos reportados en la ENOE, para cada una de las entidades federativas. La información del MCS-ENIGH reporta los ingresos de los hogares obtenidos hasta 6 meses anteriores a la fecha del levantamiento, y dado que el levantamiento de la información es de agosto a noviembre, el periodo reportado en la encuesta es de febrero a octubre (de estos 9 meses, cada hogar solo reporta 6 meses, dependiendo del mes en el que es encuestado). Por otra parte, y dado que la realización de la ENOE es trimestral, aquellos periodos donde hay mayor coincidencia con la información del MCS-ENIGH son el segundo y el tercer trimestre; es por ello que, para los datos obtenidos de la ENOE, consideramos dichos periodos.

El modelo (que más adelante se describe), insume valores objetivo para un estadístico deseado, que en nuestro caso es la mediana del ingreso corriente total (ICT). Tales estadísticos servirán como restricciones al momento de realizar el ajuste. Para generar los valores objetivo se siguen estos pasos: en cada entidad federativa por separado, se toman las medianas del ingreso reportado dentro del segundo y tercer trimestre de la ENOE para los años 2014 y 2015. Con ellas se obtienen los promedios entre los dos trimestres del mismo año. Posteriormente se calculan las variaciones porcentuales del promedio de las medianas de 2015 con respecto al de 2014. De esta manera tenemos 32 tasas de crecimiento de las medianas entre 2014 y 2015. Enseguida, y para cada entidad federativa, se calculan las medianas del ICT del MCS-ENIGH 2014; a estas medianas les aplicamos la tasa de crecimiento obtenida con los datos de la ENOE, para finalmente obtener las *medianas objetivo* de cada entidad en 2015.

La idea intuitiva del modelo es crear microdatos a partir del ajuste de una función de distribución de probabilidad; la condición que imponemos a estos nuevos microdatos es que generen el valor de un estadístico deseado (*medianas objetivo* en nuestro caso). Para ello, ajustamos una función de distribución teórica (GB2, función Beta Generalizada del segundo tipo con 4 parámetros) a nuestros datos empíricos, para cada una de las entidades por separado. Este proceso arroja estimadores de los parámetros de dicha función de densidad que determinan una forma funcional específica de la distribución. Posteriormente, realizamos un nuevo ajuste a los datos empíricos, pero ahora imponiendo como restricción que dicho ajuste genere la *mediana objetivo* según la entidad federativa respectiva. Es decir, los nuevos microdatos del ICT son tales que la mediana es igual a la mediana objetivo.

Una vez concluido este proceso, tenemos para cada hogar de la muestra un ingreso total ajustado ($ICT_h^{fase\ 1}$) y un ingreso total original (ICT_h), a partir de los cuales generamos un factor de corrección, f_h, que será aplicado al ingreso por trabajo principal para obtener un ingreso por trabajo principal ajustado ($ITP_h^{ajustado}$):

$$f_h = \frac{ICT_h^{fase\ 1}}{ICT_h}$$

$$ITP_h^{ajustado} = ITP_h * f_h$$

De manera formal, sea $Y^i = \left(y_1^i, y_2^i, \dots, y_{n_i}^i\right)$ el vector de ingreso corriente total para la entidad *i*; sus pesos (factores de expansión) correspondientes están dados por $\boldsymbol{w}^i = (w_1^i, w_2^i \dots, w_{n_i}^i)$. Entonces, el ajuste de la distribución a los datos empíricos que corresponde a maximizar la log-verosimilitud de la función $f(\boldsymbol{Y}|\boldsymbol{\theta}) = \boldsymbol{GB2}(\boldsymbol{\theta})$, se expresa de la siguiente manera**:**

Para cada entidad i = 1,2,..,32:

$$\max\ l\left(\boldsymbol{\theta^i}\middle|\boldsymbol{Y}^i\right) = \sum_{h=1}^{n_i} w_h^i log\, f\left(y_h^i\middle|\boldsymbol{\theta^i}\right)$$

Resultado de estos ajustes obtendremos 32 estimadores de los parámetros de la distribución, los cuales representamos como $\widehat{\boldsymbol{\theta}}_{SR} = (\widehat{\boldsymbol{\theta}}_{SR}^1, \widehat{\boldsymbol{\theta}}_{SR}^2, .., \widehat{\boldsymbol{\theta}}_{SR}^{32})$.

Posteriormente al ajuste obtenido, se realiza un nuevo ajuste en el que incluimos como restricción que los nuevos microdatos generen la *mediana objetivo* para la entidad federativa correspondiente. Es decir,

Sea $Y_o^i = \left(y_{(1)}^i, y_{(2)}^i, \dots, y_{(n)}^i\right)$ el vector ordenado de ingresos para la entidad *i* y sus correspondientes pesos muestrales (factores de expansión) denotados por $w_o^i = (w_{(1)}^i, w_{(2)}^i \dots, w_{(n)}^i)$. Y sea el *vect*or de *medianas objetivo* $\mathbf{Me} = (\mathrm{Me}_1, \mathrm{Me}_2, .., \mathrm{Me}_{32})$. Para cada entidad *i* =1,2,..,32 se resuelve el siguiente problema de optimización restringida:

$$\max\ l\left(\boldsymbol{\theta^i}\right) = \sum_{h=1}^{n_i} w_h^i log\, f\left(y_h^i\middle|\boldsymbol{\theta^i}\right)$$

Sujeto a:

a) Restricciones de igualdad

1. $F^{-1}\left(F\left(y_{(k)}^{i}\middle|\widehat{\boldsymbol{\theta}}_{SR}^{i}\right)\middle|\boldsymbol{\theta}^{\boldsymbol{i}}\right)=Me_i$

   donde:

   $y_{(k)}^{i}$ es el ingreso asociado al subíndice *k* tal que $0.5 \leq \frac{W_{(k)}}{W_{(n)}}$ y $0.5 \leq 1-\frac{W_{(k-1)}}{W_{(n)}}$ con $w_{(j)}=\sum_{h=1}^{j} w_{(h)}^{i}$

   Me_i es la mediana de la entidad federativa *i.*
   $\widehat{\boldsymbol{\theta}}_{SR}^{i}$ es el estimador de los parámetros de la función teórica de densidad sin restricciones para la entidad *i*.
   $F(\cdot)$ es la función de probabilidad acumulada de GB2.
   $F^{-1}(\cdot)$ es la función cuantil de GB2 y

2. $\int_0^{\max^i} f(\boldsymbol{Y}|\,\boldsymbol{\theta}^{i})dy = p^{max}(\widehat{\boldsymbol{\theta}}_{SR}^{i})$

   donde:

   $max^i = y_{(n)}^{i}$, es el valor máximo del ingreso de ICT en la entidad *i*.
   $p^{max}\left(\widehat{\boldsymbol{\theta}}_{SR}^{i}\right) = F(y_{(n)}^{i}|\widehat{\boldsymbol{\theta}}_{SR}^{i})$ es la probabilidad acumulada (bajo la función teórica sin restricciones) correspondiente al valor max en la entidad *i*.

b) *Restricciones de desigualdad*: las propias del dominio de los valores de los parámetros de la función densidad, por ejemplo, para $GB2(\mu,\sigma,\upsilon,\tau)$: $\mu,\upsilon,\tau > 0$; $-\infty < \sigma < \infty$; $-\upsilon < \frac{1}{\sigma} < \tau$.

La segunda restricción de igualdad tiene como objetivo controlar el carácter no finito en el dominio de las funciones de distribución. Para ello, tomamos el valor más grande del ICT que se obtuvo en la encuesta en cada entidad federativa y establecemos que los nuevos ingresos estimados estén en [0, max^i]

Derivado de los modelos optimizados, obtenemos un vector que contiene los estimadores de los parámetros restringidos de la función GB2 que denotamos como $\widehat{\boldsymbol{\theta}}_R = (\widehat{\boldsymbol{\theta}}_R^1, \widehat{\boldsymbol{\theta}}_R^2, .., \widehat{\boldsymbol{\theta}}_R^{32})$.

Para encontrar el ICT_fase1 en cada hogar hacemos uso de los vectores $\widehat{\boldsymbol{\theta}}_{SR}^{i}$ Y $\widehat{\boldsymbol{\theta}}_R^i$ de la siguiente manera: tomamos el valor reportado del ingreso corriente total para cada hogar, y de acuerdo con la entidad federativa de pertenencia, se calcula su probabilidad acumulada según la distribución teórica ajustada sin restricción alguna, $\hat{p}_h^i = F\left(y_h^i\middle|\widehat{\boldsymbol{\theta}}_{SR}^{i}\right)\ \forall\ h = 1,2,..n_i$; donde n_i es el número de hogares en la entidad *i*. El valor estimado del ICT_fase1 para cada hogar será entonces $\hat{y}_h^i = F^{-1}\left(\hat{p}_h^i\middle|\widehat{\boldsymbol{\theta}}_R^i\right)\ \forall\ h = 1,2,..n_i$.

De esta manera, obtenemos para cada hogar en toda la muestra, un valor del ICT imputado que está determinado en función de los resultados del ajuste por entidad, $ICT_{fase1\_h} = \hat{y}_h\ \ \forall\ \ h = 1,2,..,n$

Finalmente, para obtener el nuevo vector del Ingreso por Trabajo Principal (ITP_ajustado) realizamos lo siguiente:

i) Para cada hogar de la muestra tenemos dos valores: ICT original y el ICT_fase1. Esto nos permite obtener un factor de corrección para cada hogar *h*,

$$f = \frac{ICT_{fase1}}{ICT_{original}} \triangleq \left(\frac{\hat{y}_1}{y_1}, \frac{\hat{y}_2}{y_2}, .., \frac{\hat{y}_n}{y_n}\right) = (f_1, f_2, .., f_n)$$

ii) Calculamos un nuevo valor del Ingreso por Trabajo Principal para cada hogar (el cual es construido a partir de las claves de ingreso señaladas en el Anexo 1 de este documento) de la siguiente manera:

$$ITP_{ajustado} = ITP_{original} * f \triangleq (y_{ITP_1} f_1, y_{ITP_2} f_2, \dots, y_{ITP_n} f_n)$$

$$ITP_{ajustado} = (\hat{y}_{ITP1}, \hat{y}_{ITP2}, .., \hat{y}_{ITPn})$$

Fase 2. Ajuste al Complemento del Ingreso por Trabajo Principal (ITP^c)

Para la etapa 2 del proceso, nuestro punto de referencia es una regularidad empírica que se ha dado desde 2010. Si calculamos el peso que representa el ITP dentro del ICT, el valor es cercano a 0.615 tanto para 2010 como para 2012 y 2014. De hecho, para 2014 el valor es exactamente 0.615; mientras que para 2012 fue de 0.613 y para 2010 de 0.617 (en promedio 0.615). Por lo tanto, también debe existir regularidad empírica para el ITP^c en este mismo periodo. El objetivo del segundo ajuste es recomponer esa proporción histórica con los nuevos microdatos. Para lograrlo seguimos los siguientes pasos:

a) Generamos una variable complemento del ITP, la cual denominamos ITP^c, y que resulta de la diferencia entre el ICT y el ITP: $ITP^c = ICT_{original} - ITP_{original}$.
b) Ajustamos una función GB2 a nuestros datos empíricos del ITP^c, e imponemos como restricción que dicho ajuste sea tal que reconstituya la proporción del 0.615 entre el $ITP_{ajustado}$ y el $ICT_{ajustado}$. De esta manera obtenemos un nuevo vector del ITP, al cual llamamos $ITP^c_{ajustado}$.
c) El ingreso corriente total ajustado ($ICT_{ajustado}$) para cada hogar es la suma del $ITP_{ajustado} + ITP^c_{ajustado}$

Con ello se logra obtener valores de ingreso consistentes para cada entidad federativa.

Es importante mencionar que dados los procesos computacionales intensivos de optimización que requiere la metodología, los resultados pueden variar ligeramente dependiendo de las especificaciones del equipo de cómputo donde se realicen los cálculos.[3]

[3] Los microdatos correspondientes al *Modelo Estadístico 2015 para la Continuidad del MCS-ENIGH* fueron resultado del procesamiento del algoritmo en una computadora marca Lanix con procesador AMD FX-8370 de 8-núcleos y 4.00 GHz, sistema operativo Windows de 64-bits. La versión de R fue la 3.4.0, a través de su IDE RStudio versión 1.0143.

Anexo 1

Apartados que forman la variable de Ingreso por Trabajo Principal (ITP) del MCS-ENIGH

Código	Descripción
	Ingresos monetarios del trabajo principal para subordinados
P001	Sueldos, salarios o jornal
P002	Destajo
P003	Comisiones y propinas
P004	Horas extras
P005	Incentivos, gratificaciones o premios
P006	Bono, percepción adicional o sobresueldo
P007	Primas vacacionales y otras prestaciones en dinero

Código	Descripción
	Ingresos por negocios del hogar, trabajo principal
P011	Sueldos o salarios
P012	Ganancias/utilidades
P013	Otros ingresos
	Ingresos por negocio propio, trabajo principal
P068	Por negocios con tipo de actividad industrial
P069	Por negocios con tipo de actividad comercial
P070	Por negocios prestadores de servicios
P071	Por negocios con actividades agrícolas
P072	Por negocios con actividades de cría y explotación de animales
P073	Por negocios con actividades de recolección, reforestación y tala de árboles
P074	Por negocios con actividades de pesca, caza y captura de animales

Anexo 2. Ajuste demográfico por imputación

1. Introducción

El Módulo de Condiciones Socioeconómicas (MCS) 2015 reportó un total de hogares casi al nivel del proyectado por el Consejo Nacional de Población (CONAPO) para el 2017, un poco más de 900 mil de lo que se tenía previsto para el 2015. Esto se debió a que el tamaño de hogar reportado por el MCS 2015 es menor de lo esperado y a que los factores de expansión de las encuestas son ajustados para llegar a un total de población, dejando de lado el número de hogares.

Ante esto, el objetivo de este ejercicio es incrementar el tamaño del hogar del MCS 2015 y, como consecuencia, acercar el total de hogares a lo proyectado por el CONAPO correspondiente a ese año. Para llevarlo a cabo, se diseñó un procedimiento de imputación de individuos a cierto número de hogares. De esta forma, se incrementa la población muestral del evento y, dado ese efecto, se reajustan los factores de expansión consiguiendo incrementar el tamaño del hogar y disminuir el total de hogares. Cabe destacar que este proceso de imputación solo pretende hacer un ajuste demográfico al MCS 2015, lo cual no necesariamente se traduce en un ajuste en la distribución del ingreso y otras variables relacionadas con éste.

2. Marco teórico

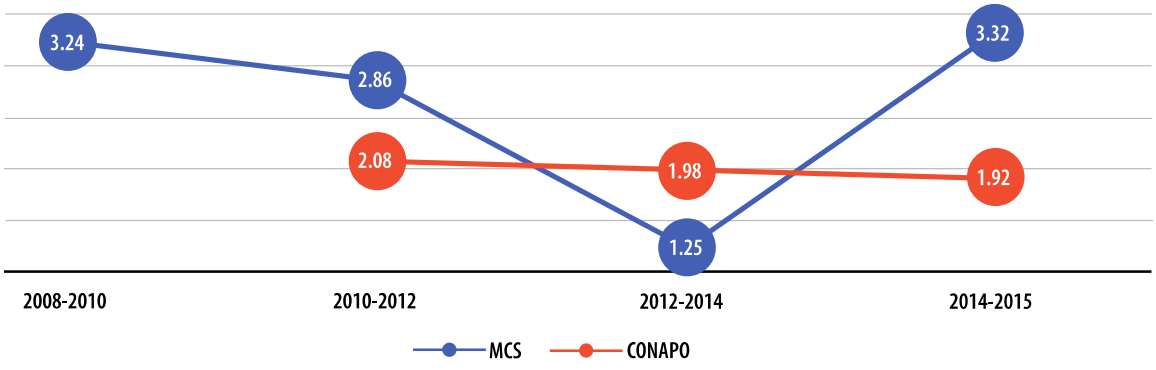
2.1 Módulo de Condiciones Socioeconómicas

Entre las diferencias que se observan en el MCS 2015 respecto a sus antecesores está un cambio en la tasa de crecimiento del total de hogares, que no es consistente con la tendencia observada desde el 2008. Como se observa en la gráfica 2.1, del 2014 al 2015 la tasa de crecimiento promedio anual de los hogares según los MCS es de 3.3%, mientras que la que resulta de las proyecciones de población del CONAPO es de solo 1.9 por ciento. Este aumento en el crecimiento del número de hogares se debe, en parte, a que el tamaño del hogar reportado por el MCS 2015 es menor a lo proyectado y, como puede verse en la gráfica 2.2, muestra una ligera caída respecto a lo observado en el 2012 y 2014.

Como sucede en todas las encuestas sociodemográficas en las que la unidad de selección es la vivienda, un hogar en muestra representa a otros hogares con las mismas características socioeconómicas y demográficas. Así, los datos muestrales se expanden, con base en el inverso de su probabilidad de inclusión en una muestra, para referir no únicamente a la unidad muestral, sino a ésta más las que ella representa en la población de la que deriva dicha muestra. Sin embargo, debe considerarse que los factores de expansión de las encuestas sociodemográficas son ajustados con el fin de alcanzar el monto total de población proyectado por el CONAPO, independientemente del resultado en el número total de los hogares.

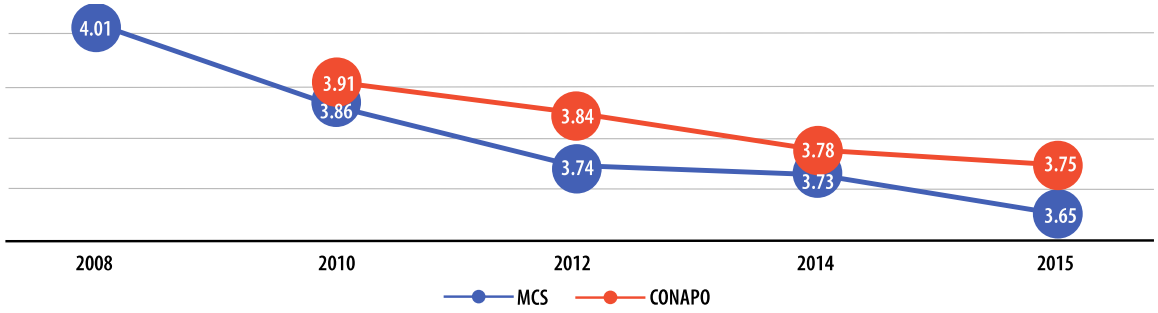
Gráfica 2.1

Tasa de crecimiento media anual de los hogares 2008-2015



Gráfica 2.2

Tamaño del hogar, 2008-2015



Considerando los factores naturales, es decir, los factores de expansión sin la corrección por la no respuesta ni la calibración por proyección de población, se observó que, en comparación con eventos previos, el tamaño promedio del hogar es menor en el MCS 2015 y, por lo tanto, el total de hogares en muestra arroja un menor registro de integrantes. Esto obliga a un ajuste adicional derivado del proceso de ajuste para alcanzar la cifra de población que indica la proyección del CONAPO; de lo contrario, el monto de población quedaría por debajo de lo proyectado para el año del levantamiento. El efecto final después del ajuste es que no solo aumenta el número de personas, sino el volumen de los hogares en los que habitan dichas personas, generando un efecto de mayor crecimiento. Esto significa que un hogar en muestra del MCS 2015 representa más hogares de los que representaría en los levantamientos previos que fueron acompañados con la ENIGH.

2.2 Métodos de imputación

Las imputaciones se pueden hacer de forma manual o automática. En el primer caso, el analista examina los datos y él mismo asigna un valor a la variable en cuestión. En el segundo caso, y de acuerdo con la UNECE (2000), las imputaciones automáticas se pueden clasificar en:

- Imputación determinística. Donde solo existe un valor correcto, por ejemplo, la falta de un total. Un valor como éste es determinado a partir de otros valores en el mismo cuestionario.
- Imputación basada en modelo. Usa una media, mediana, un modelo de regresión, etc., para imputar el valor.
- Imputación Deck. Se determina un cuestionario donador para suplir el valor faltante:
    - Imputación Hot-Deck. El donador es encontrado en la misma encuesta. Pueden distinguirse dos variantes, el determinístico (vecino más cercano), en el que un único donador se identifica basado en una métrica y los valores se imputan de ese caso; y el aleatorio, en el cual el donante se selecciona aleatoriamente de un conjunto de posibles donadores.
    - Imputación Cold-Deck. Con ésta técnica, el donador se encuentra en la misma encuesta, pero de un evento previo. Las dos variantes del Hot-Deck pueden ser aplicadas en este caso.
- Imputación Mixta. Utiliza una combinación de métodos. Primero se hace la imputación determinística y si ésta falla, se intenta una Hot-Deck y si aún sigue fallando lo hace con una imputación basada en modelo, y si de plano todas fallan, entonces se hace una manual.
- Sistemas expertos y redes neuronales. Se pueden desarrollar sistemas expertos o redes neuronales para imputar datos.

Entre los distintos métodos de imputación automática, el método Cold-Deck se considera una opción viable para el problema antes planteado debido a que no se trata de imputar valores ausentes en una variable cualitativa debido a una omisión en campo o captura, que pudiera rescatarse a partir de otros valores presentes en el mismo cuestionario, y para lo cual una imputación determinística resultaría suficiente. Tampoco se trata de predecir valores de una variable cuantitativa a partir del valor dado por la media, mediana, o por un modelo de regresión, y para lo cual la imputación basada en modelo sería la adecuada. En este caso se trata de corregir la omisión de individuos al interior de los hogares y, para ello, es posible aprovechar la información que previamente ha sido generada por el propio Instituto mediante los MCS anteriores. Así, a través de la técnica del Cold-Deck aleatorio es posible encontrar un donador en la misma encuesta, pero de un evento previo. A diferencia de la imputación Hot-Deck, donde el donador proviene de la misma encuesta, la técnica Cold-Deck permite ampliar la base de donadores con la finalidad de mejorar las precisiones estadísticas.

3. Metodología

3.1. Su aplicación

El procedimiento de imputación consistió de tres etapas: 1) determinar el volumen de población expandida a imputar, 2) definir las características sociodemográficas básicas del individuo a imputar y 3) el proceso de imputación como tal.

3.1.1 ¿Cuántos imputar?

Una vez definido el problema, surge la pregunta, ¿cuántos individuos imputar?

Se pueden tomar algunas encuestas del INEGI como punto de partida, de las cuales vale la pena revisar el dato que arrojan y sus características:

- Anclarse a la Encuesta Intercensal (EIC) 2015. Con un tamaño del hogar de 3.74 y una definición de hogar diferente a la de otras encuestas del mismo Instituto, provocando que los hogares no sean comparables debido a que la EIC reporta un hogar por vivienda, mientras que las demás pueden detectar más de uno.
- Basarse en la Encuesta Nacional de Ocupación y Empleo (ENOE). Con un tamaño del hogar de 3.79 (tercer trimestre del 2015) y tasas de variación irregulares. Esto es normal para la ENOE, ya que está diseñada para proporcionar información de la población de 15 años y más de edad sobre ocupación y empleo, como su nombre lo dice; por lo tanto, la información que se pueda obtener de los hogares es secundaria.
- La Encuesta Nacional de la Dinámica Demográfica (ENADID) 2014, con un tamaño del hogar de 3.71.

Al revisar estas encuestas no se dejó de lado el conocimiento de que el tamaño del hogar del MCS 2014 es de 3.73 y que el del MCS 2015 es menor que ese (3.65) y que parte del objetivo del ejercicio es acercarlo mas no dejarlo por arriba; por lo tanto, se descartó considerarlas para el ejercicio.

Otra opción consistía en tomar como referencia las proyecciones de población y hogares del CONAPO, específicamente las tasas de variación implícitas del tamaño del hogar. Que al final fue la vía que se tomó.

Entonces, se decidió que para ajustar demográficamente el MCS 2015, el tamaño del hogar de este evento debía ser menor que el del MCS 2014 para todas las entidades federativas, dada la tendencia a la baja que a través del tiempo muestran las proyecciones del CONAPO (ver tabla 3.1).

Con base en lo anterior, tomamos la tasa de variación implícita 2014-2015 del tamaño del hogar que reportan las proyecciones del CONAPO por entidad federativa (TCh_i , i = 1,...,32); así, por ejemplo, la tasa de variación de Aguascalientes es de -0.91% (TCh_1 = -0.0091).

Y calculamos el tamaño del hogar esperado en la i-ésima entidad para el 2015 como sigue:

$$TamHOG_{15,i}^{(es)} = TamHOG_{14,i} * (1 + TCh_i),$$

donde $TamHOG_{14,i}$ es el tamaño del hogar reportado por el MCS 2014.

Por ejemplo, para Aguascalientes que presenta $TamHOG_{14,i}$= 3.83 se tiene que:

$$TamHOG_{15,1}^{(es)} = 3.83 * (1 + (-0.0091))$$

$$TamHOG_{15,1}^{(es)} = 3.79$$

Con este dato, se calcula el total de hogares esperado para MCS 2015 en la i-ésima entidad:

$$TotHOG_{15,i}^{(es)} = \frac{TotPOB_{15,i}}{TamHOG_{15,i}^{(es)}},$$

siendo $TotPOB_{15,i}$ la población total de la i-ésima entidad del MCS 2015.

Siguiendo con el ejemplo para Aguascalientes, la población reportada por el MCS 2015 fue $TotPOB_{15,1}$= 1 292 721, entonces, el total de hogares esperado es:

$$TotHOG_{15,1}^{(es)} = \frac{1\ 292\ 721}{3.79}$$

$$TotHOG_{15,1}^{(es)} = 340\ 989$$

Sin embargo, con el total de hogares esperado y el tamaño del hogar del MCS 2015 para la mayoría de las entidades, no es posible llegar al total de población que reporta este evento, por lo que la población preliminar por entidad está dada por:

$$TotPOB_{15,i}^{(*)} = TotHOG_{15,i}^{(es)} * TamHOG_{15,i}$$

donde $TamHog_{15,i}$ representa el tamaño del hogar que arroja el MCS 2015 para la i-ésima entidad.

Luego, para Aguascalientes:

$$TotPOB_{15,1}^{(*)} = 340\ 989 * 3.78$$

$$TotPOB_{15,1}^{(*)} = 1\ 290\ 080$$

Por lo tanto, es necesario imputar población de forma expandida, que se obtiene por:

$$ImpPOB_{15,i}^{(*)} = TotPOB_{15,i} - TotPOB_{15,i}^{(*)}$$

Siguiendo el caso de Aguascalientes, nos da:

$$ImpPOB_{15,1}^{(*)} = 1\ 292\ 721 - 1\ 290\ 080$$

$$ImpPOB_{15,1}^{(*)} = 2\ 641$$

Esos 2 641 casos representan la cobertura de individuos que se deben imputar de forma expandida para Aguascalientes; cabe aclarar que en algunas entidades el resultado de este ajuste fue negativo; ante ello, se tomó la decisión de no hacer nada en esos casos (ver tabla 3.2).

Tabla 3.1

Tasa de crecimiento media anual del tamaño del hogar

Entidad	2010-2012	2012-2014	2014-2015
Aguascalientes	-0.93	-0.92	-0.91
Baja California	-1.08	-0.99	-0.95
Baja California Sur	-0.71	-0.68	-0.66
Campeche	-0.83	-0.84	-0.83
Coahuila de Zaragoza	-0.78	-0.76	-0.75
Colima	-0.81	-0.78	-0.76
Chiapas	-0.87	-0.93	-0.97
Chihuahua	-0.73	-0.66	-0.63
Ciudad de México	-1.00	-0.94	-0.90
Durango	-0.90	-0.87	-0.87
Guanajuato	-0.87	-0.88	-0.88
Guerrero	-0.73	-0.81	-0.85
Hidalgo	-0.79	-0.80	-0.81
Jalisco	-0.89	-0.87	-0.84
México	-1.00	-0.99	-0.96
Michoacán de Ocampo	-0.76	-0.77	-0.75
Morelos	-0.74	-0.74	-0.71
Nayarit	-0.57	-0.54	-0.54
Nuevo León	-0.85	-0.80	-0.77
Oaxaca	-0.53	-0.58	-0.62
Puebla	-0.75	-0.79	-0.81
Querétaro	-1.07	-1.05	-1.02
Quintana Roo	-1.06	-0.92	-0.84
San Luis Potosí	-0.65	-0.70	-0.73
Sinaloa	-0.93	-0.87	-0.85
Sonora	-0.75	-0.72	-0.71
Tabasco	-1.00	-1.02	-1.00
Tamaulipas	-0.80	-0.76	-0.75
Tlaxcala	-0.85	-0.86	-0.87
Veracruz de Ignacio de la Llave	-0.75	-0.76	-0.76
Yucatán	-0.78	-0.75	-0.73
Zacatecas	-0.66	-0.69	-0.70

Tabla 3.2

Coberturas de imputación por entidad federativa

Entidad	Cobertura
Total	**1 956 473**
Aguascalientes	2 641
Baja California	165 844
Baja California Sur	23 583
Campeche	0
Coahuila de Zaragoza	35 612
Colima	20 877
Chiapas	0
Chihuahua	40 888
Ciudad de México	200 560
Durango	29 222
Guanajuato	0
Guerrero	157 513
Hidalgo	0
Jalisco	0
México	307 343
Michoacán de Ocampo	182 359
Morelos	53 027
Nayarit	9 639
Nuevo León	119 344
Oaxaca	207 278
Puebla	92 396
Querétaro	46 780
Quintana Roo	48 339
San Luis Potosí	78 436
Sinaloa	25 861
Sonora	412
Tabasco	0
Tamaulipas	50 154
Tlaxcala	9 777
Veracruz de Ignacio de la Llave	0
Yucatán	48 587
Zacatecas	0

3.1.2 ¿Qué y a quién imputar?

Una vez que se tiene el número de pobladores a imputar en cada entidad, ahora sigue determinar las características de los individuos donadores. Para ello, requerimos cuidar ciertos aspectos:

- La proporción de hombres y mujeres en las entidades federativas.
- La proporción de los grupos de edad (0 a 11, 12 a 64 y 65 y más), los tamaños de localidad y los estratos socioeconómicos tanto en hombres como en mujeres, por entidad federativa.

Al cuidar esas proporciones buscamos evitar que el proceso de imputación sesgue los resultados hacia una o varias categorías de esas variables, es decir, que al final del proceso haya más hombres que mujeres, por ejemplo.

Para obtener las proporciones, usamos información de la EIC y la ENOE, no con el fin de tratar de acercarse a su estructura poblacional (el MCS 2015 ya trae la propia y modificarla sería demasiado ostentoso) sino, como ya se dijo, para evitar el sesgo hacia alguna categoría.

Para decidir sobre las características que debe tener el donador, se usaron las proporciones ya mencionadas. En el caso del sexo se utilizaron las de hombres y mujeres por entidad federativa. Para grupos de edad, tamaño de localidad y estrato socioeconómico, se construyeron intervalos de decisión para cada categoría de manera que fuera de una longitud igual a la proporción observada ya sea en hombres o bien en mujeres (ver tablas 3.3 a 3.5 para el caso de Aguascalientes).

La segregación de sexo por las variables entidad federativa, grupos de edad, tamaño de localidad y estrato socioeconómico, realizada de forma independiente para obtener las proporciones, se hizo pensando que al hacer inferencia estadística a esos niveles se obtienen coeficientes de variación aceptables para un gran número de variables. Hacerlo de forma diluida, es decir, desagregando sexo dentro de grupos de edad, estrato socioeconómico, tamaño de localidad y entidad federativa complicaría el

procedimiento (dada la alta posibilidad de no encontrar donadores), además de que la inferencia estadística a esos niveles de desagregación arroja coeficientes de variación no satisfactorios.

El proceso de asignación de características se lleva a cabo entidad por entidad, generando números aleatorios de la distribución uniforme, siendo de la siguiente forma:

1. Para determinar el sexo, se genera un primer número aleatorio, si este es menor o igual a la proporción de hombres, se asigna un hombre, de lo contrario, se asigna una mujer.

2. Ya que se tiene el sexo del individuo, se genera un segundo número aleatorio para asignarle el grupo de edad condicionado al sexo determinado con anterioridad. Para decidir sobre éste, se verifica en qué intervalo de decisión se incluye.

3. Para saber qué tamaño de localidad y estrato socioeconómico es asignado, se generan un tercer y cuarto números aleatorios, y se revisa el intervalo de decisión en el que se incluyen, condicionado también al sexo.

Tabla 3.3

Intervalos de decisión para grupos de edad. Aguascalientes

Grupos de edad	Hombres			Mujeres		
	Proporción	**LI**	**LS**	**Proporción**	**LI**	**LS**
1. De 0 a 11 años	0.241	0.000	0.241	0.224	0.000	0.224
2. De 12 a 64 años	0.705	0.241	0.946	0.714	0.224	0.938
3. De 65 y más años	0.054	0.946	1.000	0.062	0.938	1.000

Fuente: INEGI. Encuesta Intercensal 2015.

Tabla 3.4

Intervalos de decisión para tamaño de localidad. Aguascalientes

Tamaño de localidad	Hombres			Mujeres		
	Proporción	**LI**	**LS**	**Proporción**	**LI**	**LS**
1. De 100 mil y más habitantes	0.595	0.000	0.595	0.602	0.000	0.602
2. De 15 mil a 99 999 habitantes	0.113	0.595	0.708	0.113	0.602	0.715
3. De 2 500 a 14 999 habitantes	0.089	0.708	0.797	0.088	0.715	0.803
4. De menos de 2 500 habitantes	0.203	0.797	1.000	0.197	0.803	1.000

Fuente: INEGI. Encuesta Intercensal 2015.

Tabla 3.5

Intervalos de decisión para estrato económico. Aguascalientes

Estrato socioeconómico	Hombres			Mujeres		
	Proporción	**LI**	**LS**	**Proporción**	**LI**	**LS**
1. Bajo						
2. Medio bajo	0.553	0.000	0.553	0.540	0.000	0.540
3. Medio alto	0.327	0.553	0.880	0.334	0.540	0.874
4. Alto	0.120	0.880	1.000	0.126	0.874	1.000

Fuente: INEGI. Encuesta Nacional de Ocupación y Empleo. Tercer trimestre del 2015.

3.1.3 Proceso de imputación

El proceso de imputación se hace también entidad por entidad.

- Para seleccionar a los hogares receptores de individuos, se usa la técnica Permanent Random Numbers (PRN). Ésta consiste en asignar un número aleatorio (de la distribución uniforme) a cada hogar y después ordenarlos descendentemente por ese aleatorio.

- Una vez ordenados los hogares por el número aleatorio, se busca el primer hogar que cumpla con las condiciones del mismo tamaño de localidad y mismo estrato socioeconómico como los ya designados.

- Se asigna el individuo a ese hogar y se acumula su factor de expansión, luego el siguiente y acumulando su factor de expansión y así hasta que el factor acumulado supere la cobertura de la entidad, como se muestran en la tabla 3.1.

Una vez que se determinó lo que ha de imputarse y se seleccionaron los receptores, se procede al proceso de imputación de valores para el resto de variables aún no determinadas.

Para asegurarnos de encontrar posibles donadores, se construyó un agregado (pool) de pobladores de los MCS 2012 a 2015 con el fin de tener un conjunto de individuos mucho más grande de dónde obtener dichos donadores. En la tabla 3.6 se presentan el total de imputados y su origen.

Tabla 3.6

Total de imputados por MCS de origen

MCS	Imputados	%
Total	**3 759**	**100.0**
2012	1 039	27.6
2014	1 139	30.3
2015	1 581	42.1

El método de imputación usado fue Cold-Deck aleatorio, considerando las siguientes variables de empate para encontrar posibles donadores:

- Entidad federativa.
- Tamaño de localidad.
- Estrato socioeconómico.
- Clase de hogar (nuclear, ampliado y compuesto).
- Sexo, edad y nivel educativo del jefe del hogar.
- Sexo y grupo de edad del individuo a imputar.

A pesar de que se pudo haber elegido un conjunto más amplio de variables, no se hizo, puesto que con dicha decisión se corría el riesgo de que en un gran número de casos no encontraran posibles donadores.

Es importante mencionar que, aunque la clase de hogar se incluyó como variable de empate, se determinó no modificarla. Para ello, se filtraron los parentescos que pueden ser parte de cada clase; por ejemplo, para hogar nuclear solo se buscaron donadores cuyo parentesco fuera hija o hijo, para hogares ampliados solo se buscaron a hijas o hijos y otros parientes, y para compuestos se buscó a hijas o hijos, otros parientes y no parientes.

Adicionalmente, a los parentescos se aplicaron ciertos filtros:

- Si el donador era hija o hijo, se cuidó que la edad de la jefa o del jefe del hogar receptor fuera 13 o más años mayor a dicho donador, con el fin de evitar imputar un hijo mayor que el jefe.
- Si el donador era hija o hijo y el hogar receptor era encabezado por una mujer, se buscó que la diferencia máxima de edades entre la jefa o el jefe y el donador fuera de 45 años, para evitar imputar hijas o hijos menores de 12 años a jefas de hogar de la tercera edad.
- Si el donador era nieta o nieto, bisnieta o bisnieto o bien tataranieta o tataranieto, se cuidó que la edad de la jefa o el jefe del hogar receptor fuera 26, 39 y 52 años mayor que aquéllos (como mínimo), respectivamente.
- Si el donador era abuela o abuelo, bisabuela o bisabuelo o bien tatarabuela o tatarabuelo, se cuidó que la edad del jefe del hogar receptor fuera 26, 39 y 52 años menor que aquéllos (como mínimo), respectivamente.

Cabe aclarar que solo se donó la información sociodemográfica del individuo, excluyendo toda aquella relacionada con ingresos, gastos y trabajos. Así que, cuando se imputó a un individuo cuyo donador estaba clasificado como población no económicamente activa (PNEA) o buscador de trabajo, se le respetó la categoría. Sin embargo, cuando el donador era una persona con trabajo, se cambió la condición de actividad a PNEA o buscador de trabajo del individuo agregado. Lo anterior en el entendido de no incrementar el número de trabajadores.

Los siguientes puntos enuncian la forma en que quedaron clasificados aquellos registros en los que sus donadores eran parte de la población económicamente activa (PEA):

- Como estudiante si asistía a la escuela.
- Dedicado a quehaceres domésticos si era menor de 15 años de edad y no asistía a la escuela.
- Buscador de trabajo si era hombre de 15 años y más y no asistía a la escuela.
- 46% de las mujeres se clasificaron como dedicadas a quehaceres domésticos si eran de 15 años y más y no asistía a la escuela; las restantes (44%) se catalogaron como buscadoras de trabajo (en virtud de los porcentajes de mujeres de 15 años y más que se clasifican como PNEA y PEA).

Debido a que son pocas las características de empate entre la persona que se busca y los posibles donadores, al ir por estos últimos se pueden encontrar más de uno. Si esto ocurre, se vuelve a usar la técnica PRN para elegir solo uno.

Una vez concluido el proceso de imputación, dado que hay más pobladores, se reajustaron los factores de expansión (ajuste de razón por entidad y tamaño de localidad) usando como base los factores de expansión del propio MCS 2015 (como fueron publicados) para que el total de la población expandida fuera congruente con las proyecciones del CONAPO para el 2015. En general, ello significa que los factores de expansión no tienen que ser tan grandes como los originales y, por ende, se obtienen menos hogares.

3.2 Nota técnica

3.2.1 Software usado

Visual FoxPro 9.0 de Microsoft fue el software usado para llevar a cabo el procedimiento de imputación. Esta pieza de software es un lenguaje de programación procedural y orientado a objetos que incluye un Sistema Gestor de Bases de Datos (DBMS, por sus siglas en inglés). Para llevar a cabo el trabajo, se hizo uso tanto del lenguaje de programación como del SQL que este software incluye.

3.2.2 Fuentes utilizadas

La primera fuente de información usada fueron las proyecciones del CONAPO con el fin de obtener las variaciones 2014-2015 del tamaño del hogar. Una segunda fue la Encuesta Intercensal 2015, de la cual se consiguieron las proporciones de hombres y mujeres por entidad federativa, las proporciones de hombres por grupos de edad y tamaño de localidad, así como las proporciones de mujeres por grupos de edad y tamaño de localidad, en cada entidad federativa. Otra más fue la Encuesta Nacional de Ocupación y Empleo, de la que se obtuvieron las proporciones de hombres por estrato socioeconómico y mujeres por los mismos estratos en cada entidad federativa.

Ahora bien, para incrementar las posibilidades de encontrar donadores para todos los casos a imputar se construyó un pool de pobladores con los MCS 2012 a 2015, donde se incluyeron todas las variables sociodemográficas (solo la tabla Población de la base de datos).

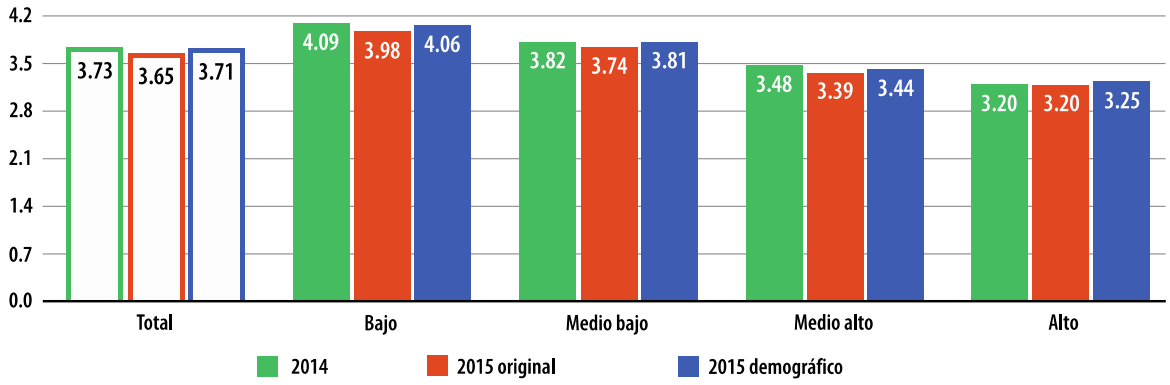
Entonces, enlistando las fuentes, serían:

- INEGI. Módulo de Condiciones Socioeconómicas 2012.
- INEGI. Módulo de Condiciones Socioeconómicas 2014.
- INEGI. Módulo de Condiciones Socioeconómicas 2015.
- INEGI. Encuesta Intercensal 2015.
- INEGI. Encuesta de Ocupación y Empleo. Tercer trimestre del 2015.
- CONAPO. Proyecciones de población nacional y entidades federativas, 2010-2030.
- CONAPO. Proyecciones de los hogares en México y las entidades federativas, 2010-2030.

4. Resultados

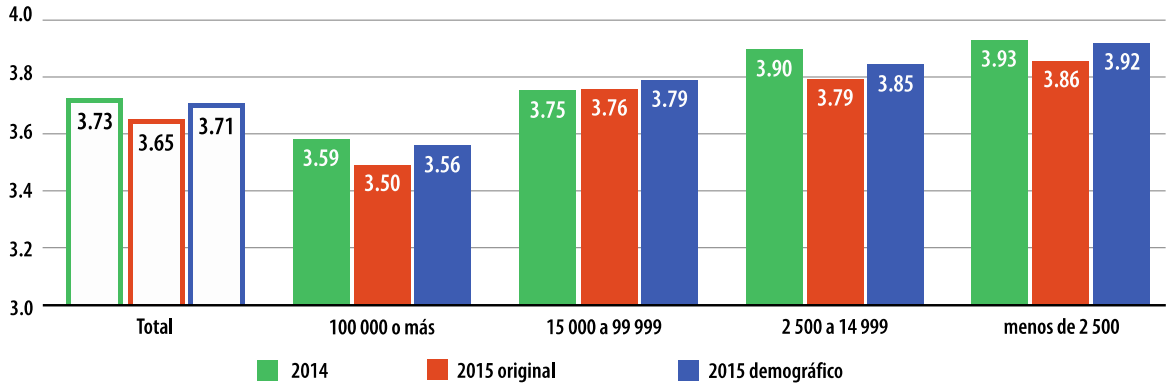
Gráfica 4.1

Tamaño del hogar por estrato socioeconómico



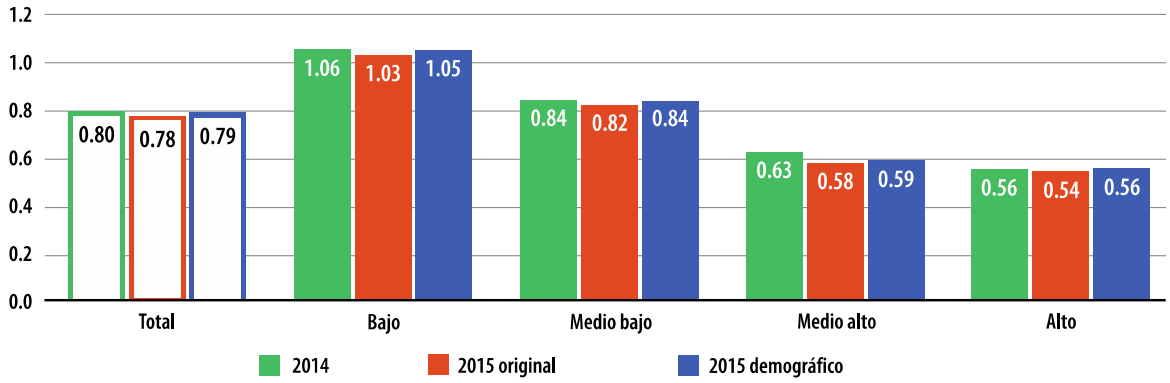
Gráfica 4.2

Tamaño del hogar por tamaño de localidad



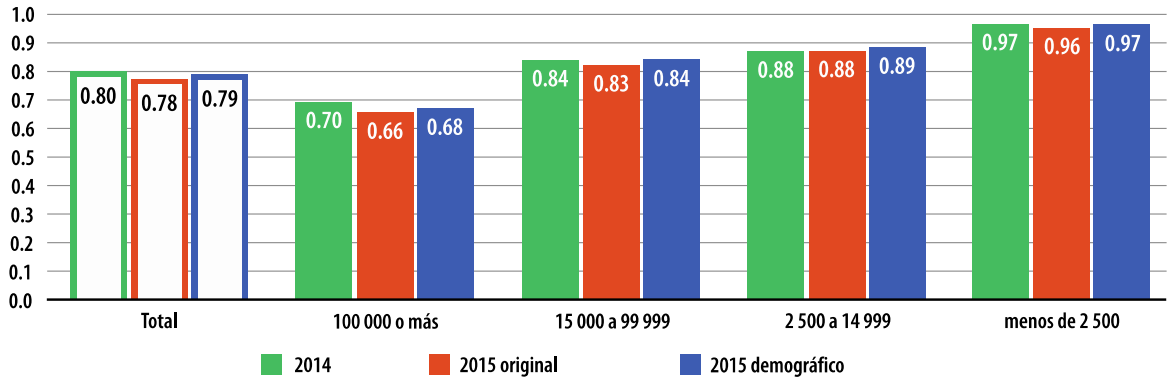
Gráfica 4.3

Promedio de integrantes del hogar menores de 12 años por estrato socioeconómico



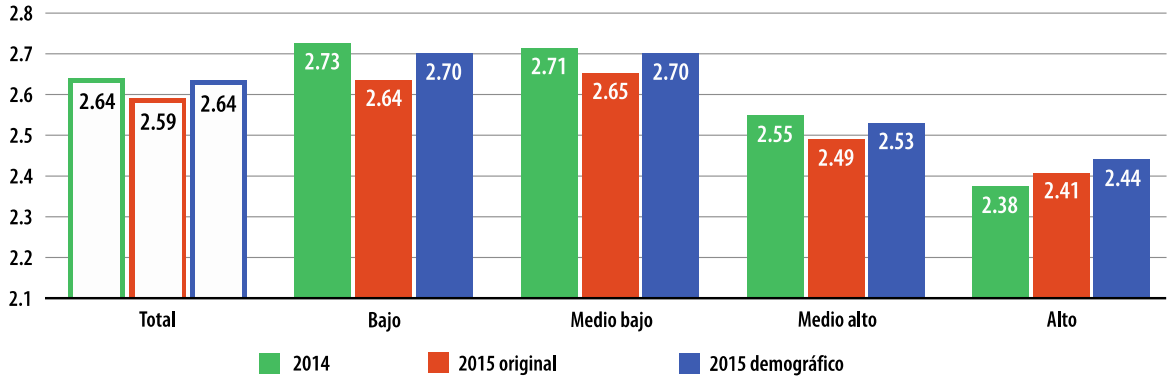
Gráfica 4.4

Promedio de integrantes del hogar menores de 12 años por tamaño de localidad



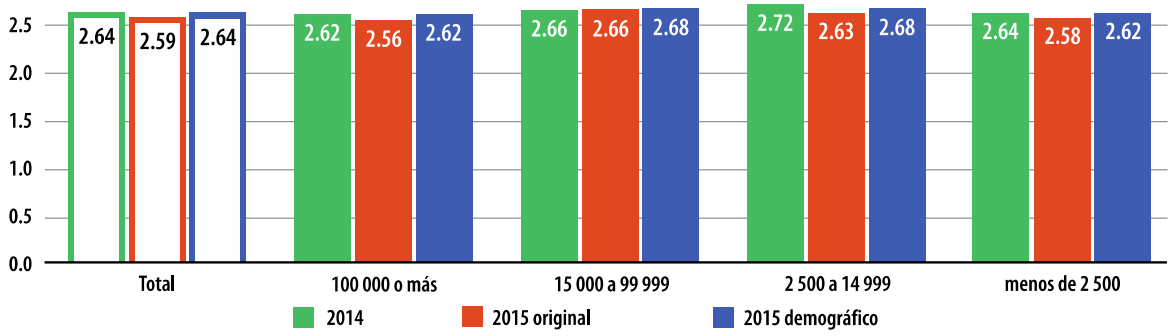
Gráfica 4.5

Promedio de integrantes del hogar de 12 a 64 años por estrato socioeconómico



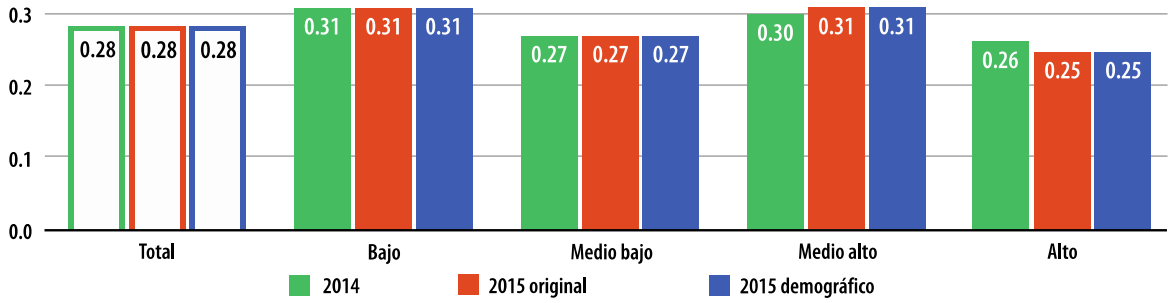
Gráfica 4.6

Promedio de integrantes del hogar de 12 a 64 años por tamaño de localidad



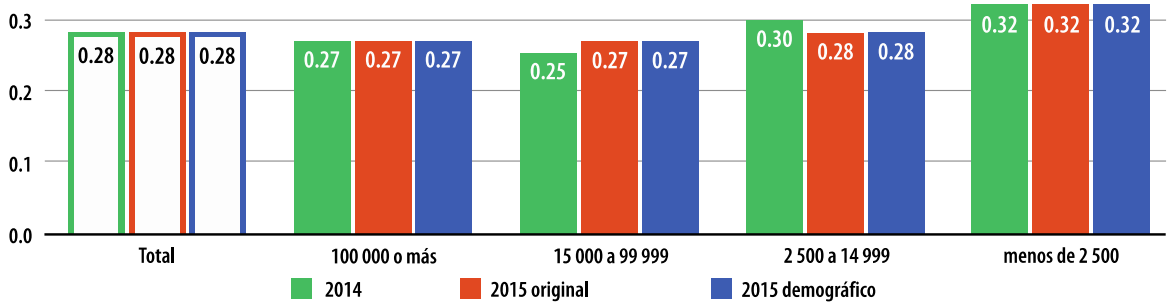
Gráfica 4.7

Promedio de integrantes del hogar de 65 años y más por estrato socioeconómico



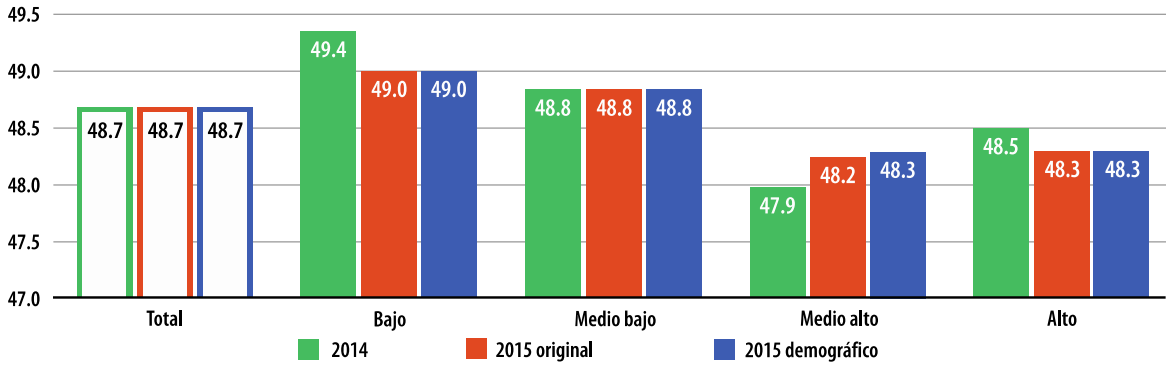
Gráfica 4.8

Promedio de integrantes del hogar de 65 años y más por tamaño de localidad



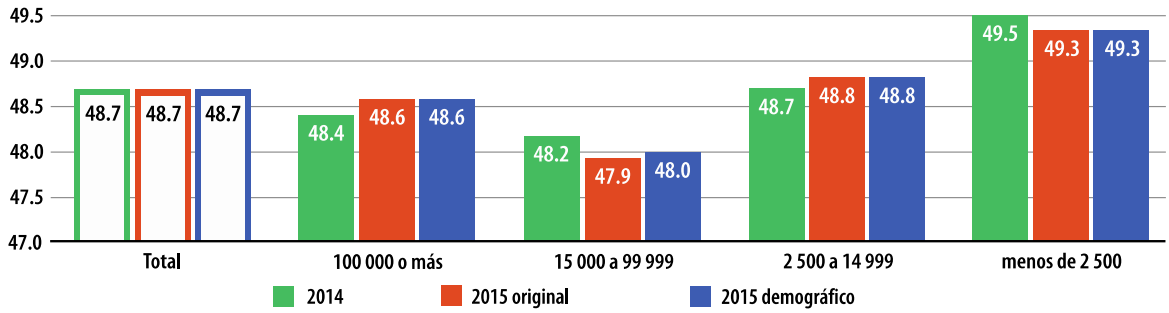
Gráfica 4.9

Porcentaje de hombres por estrato socioeconómico



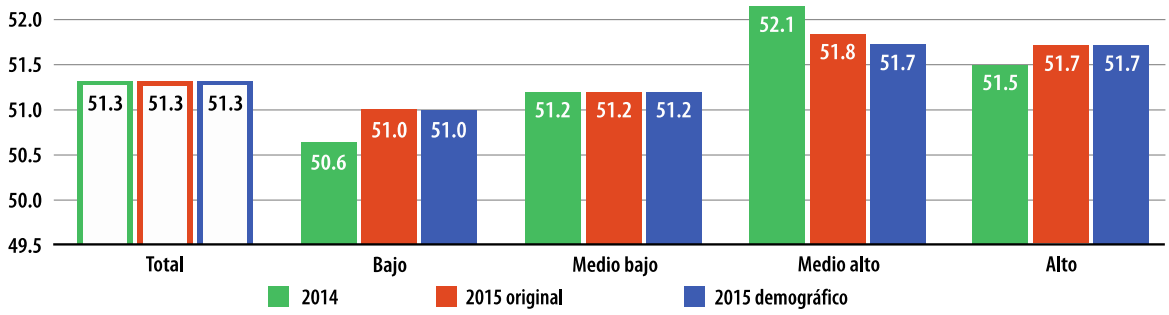
Gráfica 4.10

Porcentaje de hombres por tamaño de localidad



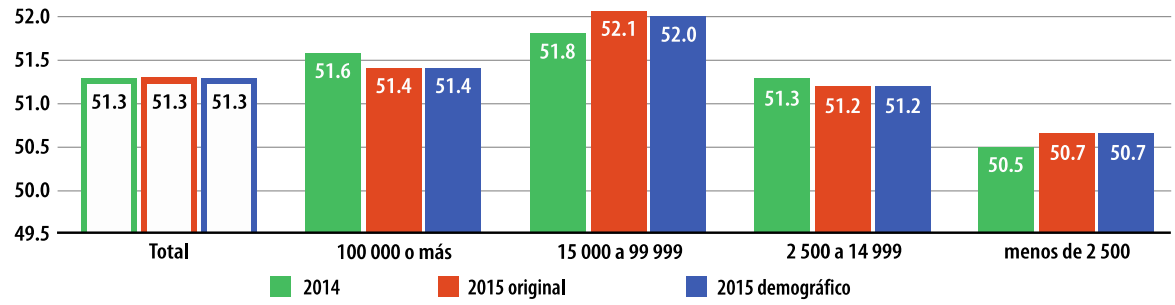
Gráfica 4.11

Porcentaje de mujeres por estrato socioeconómico



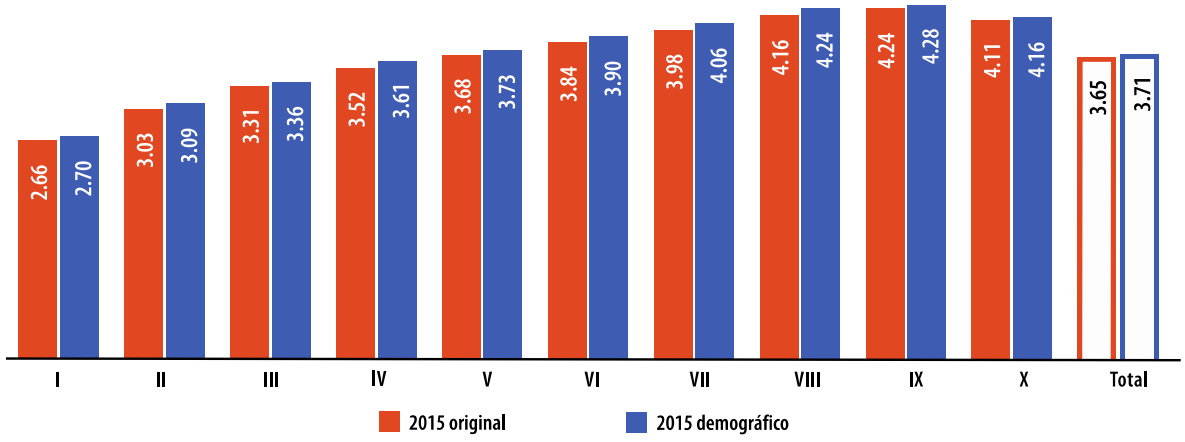
Gráfica 4.12

Porcentaje de mujeres por tamaño de localidad



Gráfica 4.13

Tamaño promedio del hogar por deciles



Gráfica 4.14

Promedio por hogar de ingreso corriente total por deciles. Variación porcentual 2014-2015

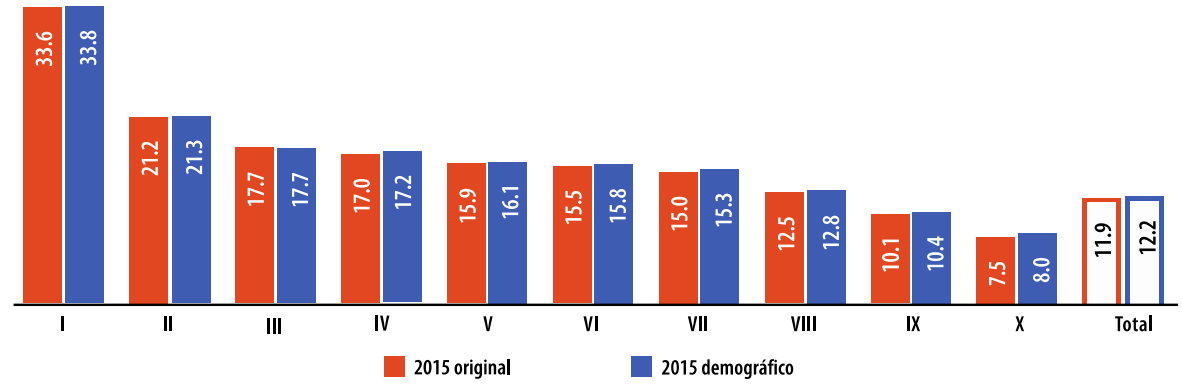


Tabla 4.1

Descomposición del ingreso corriente total (ICT)

Decil	2015 original				2015 demográfico			
	$ICT1_{15}$ / $ICT1_{14}$	**IPP**	**PH**	**Hog**	**$ICT1_{15}$ / $ICT1_{14}$**	**IPP**	**PH**	**Hog**
Total	**1.156**	**1.077**	**1.039**	**1.033**	**1.141**	**1.082**	**1.037**	**1.017**
I	1.380	1.313	1.017	1.033	1.360	1.318	1.015	1.017
II	1.252	1.243	0.976	1.033	1.233	1.243	0.976	1.017
III	1.216	1.155	1.019	1.033	1.197	1.155	1.019	1.017
IV	1.209	1.146	1.021	1.033	1.191	1.147	1.022	1.017
V	1.198	1.105	1.049	1.033	1.180	1.113	1.044	1.017
VI	1.194	1.123	1.028	1.033	1.177	1.127	1.028	1.017
VII	1.188	1.094	1.051	1.033	1.172	1.097	1.051	1.017
VIII	1.163	1.072	1.049	1.033	1.147	1.076	1.049	1.017
IX	1.138	1.030	1.069	1.033	1.122	1.036	1.066	1.017
X	1.111	0.993	1.082	1.033	1.098	1.000	1.080	1.017

Nota: IPP = variación de ingreso corriente por perceptor, PH = variación de perceptores por hogar y Hog = variación de hogares.

5. Validación y evaluación de la metodología

Aunque el ajuste demográfico no tiene gran impacto en la disminución del ingreso del MCS 2015, sí elimina el efecto que el problema demográfico pudiera tener en éstos, es decir, este ejercicio elimina los posibles efectos que el comportamiento demográfico podría representar para que el MCS 2015 arrojara ingresos más altos que los que se venían captando en los eventos anteriores. Así, pone el piso base para que se realicen ejercicios de alineación a los eventos anteriores.

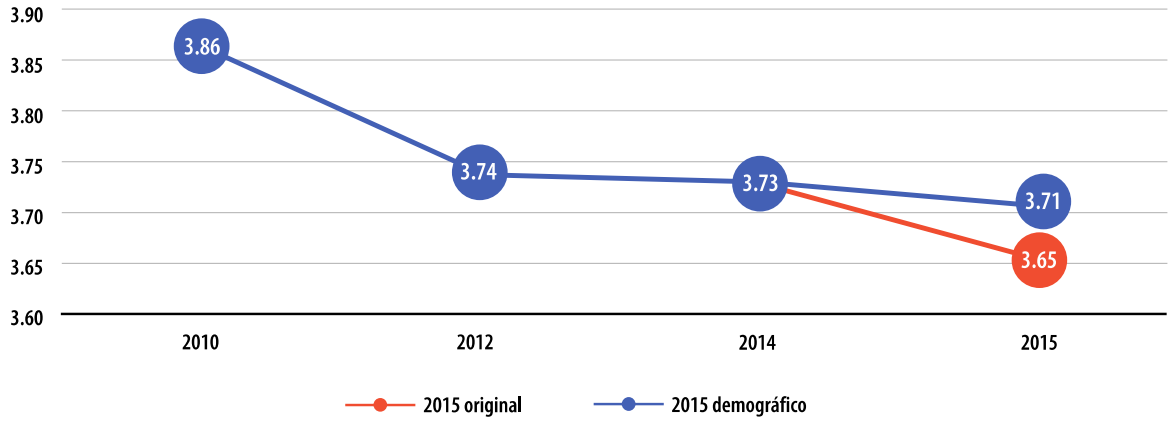
Sin duda, los métodos de imputación Deck arrojan resultados satisfactorios cuando existe un conjunto muy grande de posibles donadores como en este ejercicio. Sin embargo, cuando el conjunto de posibles donadores es pequeño, para un gran número de casos existe la posibilidad de que se encuentren el o los mismos donadores, eliminando poco a poco la variabilidad en las observaciones y como consecuencia compactar las varianzas.

Ahora bien, para este ejercicio en específico, los resultados fueron muy satisfactorios. Si observamos la gráfica 5.1, notamos que se cumple con el objetivo de incrementar el tamaño del hogar del MCS 2015, al pasar de 3.65 a 3.71. Además, el total de hogares lo deja en 32 681 856 y no los 33 218 037 del MCS 2015 original (536 181 hogares menos), dejando la tasa de crecimiento de éstos entre el 2014 y 2015 en 1.65, muy por debajo del 3.3 del original.

Por otro lado, en la gráfica 5.2 se presenta la descomposición del tamaño del hogar en grandes grupos de edad, pudiendo observar que el ejercicio se acerca a la tendencia histórica del MCS, sobre todo en el grupo de 12 a 64 años, que es donde se presenta el desajuste del MCS 2015 original. También se comporta mejor que la EIC 2015 y que un ejercicio de posestratificación del MCS 2015 con Stata (teniendo como base las variables total de integrantes del hogar, edad y sexo de la jefa o del jefe del hogar que reportó la EIC 2015), ya que estos dos últimos rejuvenecen la población.

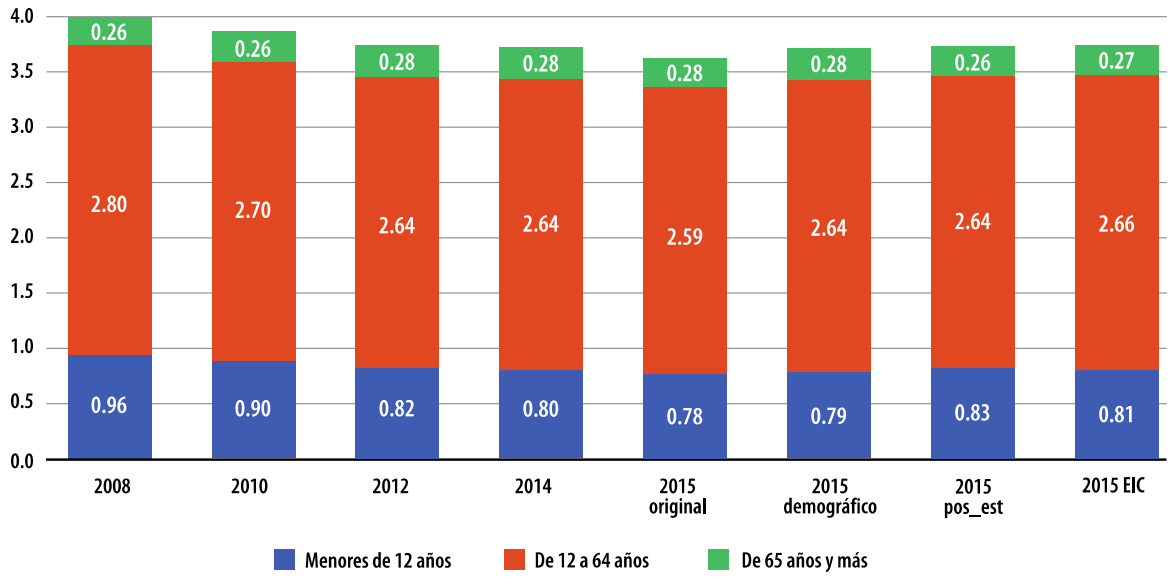
Gráfica 5.1

Tamaño del hogar 2010-2015



Gráfica 5.2

Descomposición del tamaño del hogar por grandes grupos de edad 2008-2015



6. Referencias

Dempster, A. P. y D. B. Rubin. Incomplete data in sample surveys. Academic Press, New York, NY, pp.3-10, 1983.

Juárez Alonso, Carlos Alberto. Fusión de datos: imputación y validación. Tesis doctoral, Universidad Politécnica de Cataluña, 2004.

Kim, J. K. y W. Fuller. "Fractional hot deck imputation", en: Biometrika. 91(3):559-578, 2004.

Palacios Ostria, Margot Alejandra y Edgar Javier González Liceaga. Metodología de Fellegi y Holt: validación e imputación de datos. Tesis profesional, UNAM, 2004.

Schafer, J. L. The multiple imputation FAQ page. (DE) consultada el 9 de enero de 2017 en http://www.stat.ufl.edu/~athienit/STA6167/Missing%20Data/MI_FAQ.pdf

United Nations Statistical Commission and Economic Commission for Europe. Glossary of terms on statistical data editing, UNECE, 2000. (DE) consultada el 9 de enero de 2017 en https://webgate.ec.

europa.eu/fpfis/mwikis/essvalidserv/images/3/37/UN_editing_glossary.pdf