

Ajuste demográfico por imputación

Demographic adjustment by imputation

Benito Durán Romo*

Un aumento inusual en el número de hogares reportado por el Módulo de Condiciones Socioeconómicas (MCS) 2015 se explica en parte por la contracción del tamaño del hogar y el consiguiente ajuste de factores de expansión para aumentar el volumen de la población con el objetivo de acercarse a las proyecciones del Consejo Nacional de Población (CONAPO). Esto también tuvo un efecto: aumentó el volumen de hogares, pero en este caso superó el número proyectado por el CONAPO (32 298 321, por debajo de los 33 218 037 reportados por el MCS 2015). Si en el MCS 2015 la contracción se debió a una omisión poblacional dentro del hogar durante la fase de encuestado, puede restaurarse el tamaño de los hogares con un cuidadoso proceso de imputación automática de individuos en ciertos hogares con base en la técnica de imputación *Cold-Deck*. Para lograrlo, se creó un gran grupo (*pool*) de donantes con las personas registradas en los MCS 2012 hasta 2015. Como resultado final, el proceso produce más individuos que en la base de datos original, requiriendo un reajuste de los factores de expansión, de acuerdo con la nueva estructura de población resultante de la imputación. En general, esto significa que los factores de expansión no tienen que ser tan grandes como los originales y, por lo tanto, se obtienen menos hogares. Con este procedimiento de imputación del MCS 2015, el número de hogares se redujo en poco más de 500 mil, con un tamaño del hogar de 3.71 y no 3.65 correspondiente al proceso original. Al comparar el ingreso actual con el MCS 2014, el monto total crece 14.1% (1.53% menos que el original), aunque el promedio por hogar crece 12.25% (0.32% más que el original).

Palabras clave: *Cold-Deck*; imputación; individuos; hogares; encuesta; factores de expansión.

An unusual increase in the number of households reported by the Socio-Economic Conditions Module 2015 (MCS 2015 by its Spanish acronym) is explained in part by the contraction of the household size and the consequent adjustment of weights to increase the volume of the population with the aim of approaching to the projections of National Council of Population (CONAPO by its Spanish acronym). This also had an effect: increased the volume of households, but in this case exceeded the number projected by CONAPO (32 298 321, below the 33 218 037 reported by the MCS 2015). If in the MCS 2015 the contraction was due to a population omission within the household during the survey phase, we can restore the size of the households with a careful process of automatic imputation of individuals in certain households, based on the *Cold-Deck* imputation technique. To achieve this, a large pool of donors was created with the people recorded in the MCS 2012 until 2015. As a final result, the process produces more individuals than in the original database, requiring a readjustment of the weights, according to the new structure of the population resulting from imputation. In general, this means that the weights do not have to be as large as the original ones and therefore, fewer households are obtained.

* Instituto Nacional de Estadística y Geografía, benito.duran@inegi.org.mx

Nota: el autor agradece la participación de Adriana Pérez Amador, Olinca Páez Domínguez y Lilia Guadalupe Luna Ramírez en el proceso de investigación y de integración de este documento.

With this imputation procedure of MCS 2015, the number of households was reduced by just over 500 thousand, with a household size of 3.71 and not 3.65 corresponding to the original process. When comparing current income with MCS 2014, the total amount grows 14.1% (1.53% less than the original), although the average per household grows 12.25% (0.32% more than the original).

Key words: Cold-Deck; imputation; individuals; households; survey; weights.



Guelagueta festivity in Mexico / Eduardo Gonzalez/CON/Colaborador/Getty Images

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.

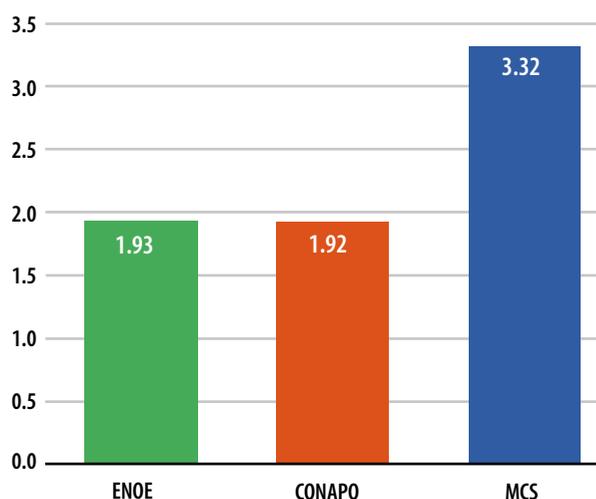
En la gráfica 1.1 se pueden observar las tasas de crecimiento de los hogares para distintas fuentes. En ésta notamos que mientras algunas encuestas del Instituto Nacional de Estadística y

Geografía (INEGI) y las proyecciones del CONAPO reportaron tasas de crecimiento de los hogares menores a 2% entre el 2014 y 2015, el MCS la estimó 3.32% para el mismo periodo, es decir, 1.4% más que la del CONAPO, lo cual es una clara evidencia del crecimiento inusual de los hogares en el MCS 2015.

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.

Gráfica 1.1

Tasa de crecimiento de los hogares 2014-2015 según distintas fuentes



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.

La sección *Marco teórico* se divide en dos partes, en la primera se presenta una de las posibles explicaciones de la baja en el tamaño del hogar del MCS 2015 y en la segunda parte se muestran algunas definiciones del proceso de imputación y su clasificación. Después, en *Metodología* se detalla el proceso completo de imputación aplicado a este caso, desde la forma de determinar el número de personas a imputar hasta llegar a la imputación como tal. Los resultados del procedimiento de imputación se dan en la sección correspondiente y, por último, se presentan comentarios respecto a la metodología y algunas conclusiones.

2. Marco teórico

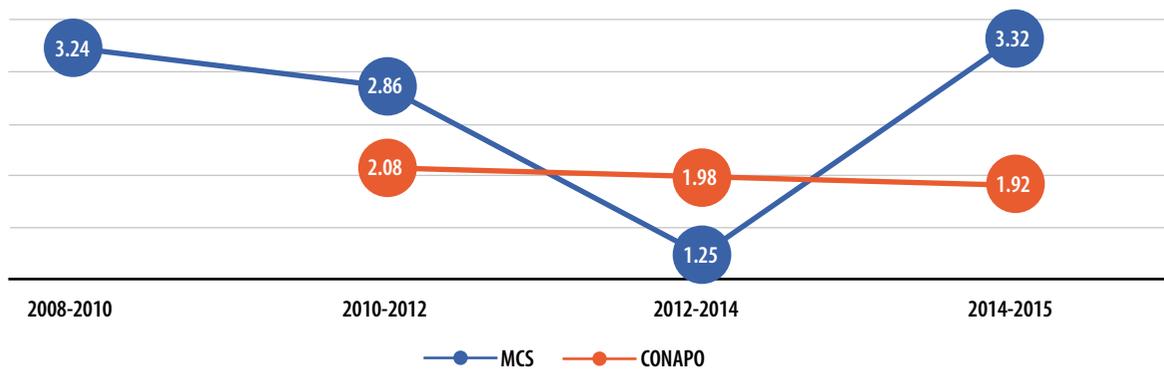
2.1 Módulo de Condiciones Socioeconómicas

El MCS surge en el 2008 como el insumo para la medición multidimensional de la pobreza que por mandato de la *Ley General de Desarrollo Social (LGDS)* tiene que hacer el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) y que, a la vez, obliga al INEGI a llevar a cabo los censos, conteos y encuestas necesarios para su medición. Desde ese año y hasta el 2014 los MCS tuvieron una periodicidad bienal (en sintonía con la LGDS, debido a que mandata que la pobreza categorizada por entidad federativa debe medirse como mínimo cada dos años) y acompañaron a la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), de manera que una parte de la muestra la cubría la propia ENIGH (muestra ENIGH), captando información sobre ingreso y gasto; y otra el MCS (muestra MCS) captando solo información de ingreso. En el 2015 toda la muestra fue cubierta por el operativo del MCS, sin contar con el acompañamiento de la ENIGH. Esto pudo haber influido en que se observaran algunas diferencias en la información con respecto a los eventos anteriores.

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.

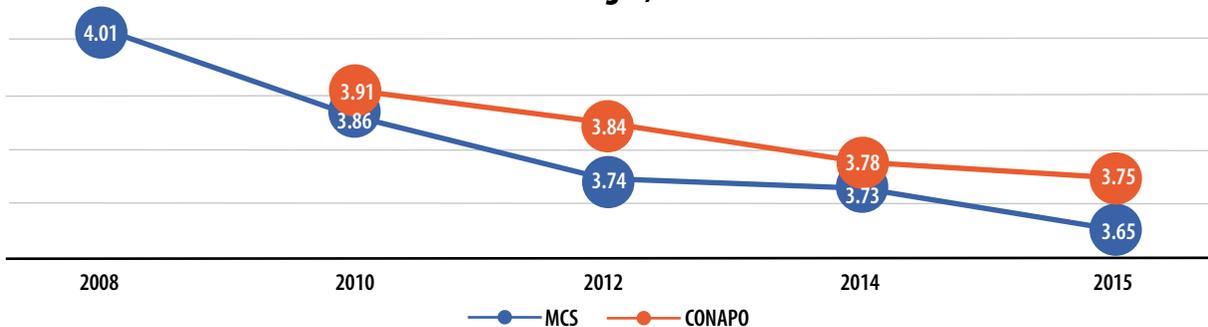
Gráfica 2.1

Tasa de crecimiento media anual de los hogares 2008-2015



Gráfica 2.2

Tamaño del hogar, 2008-2015



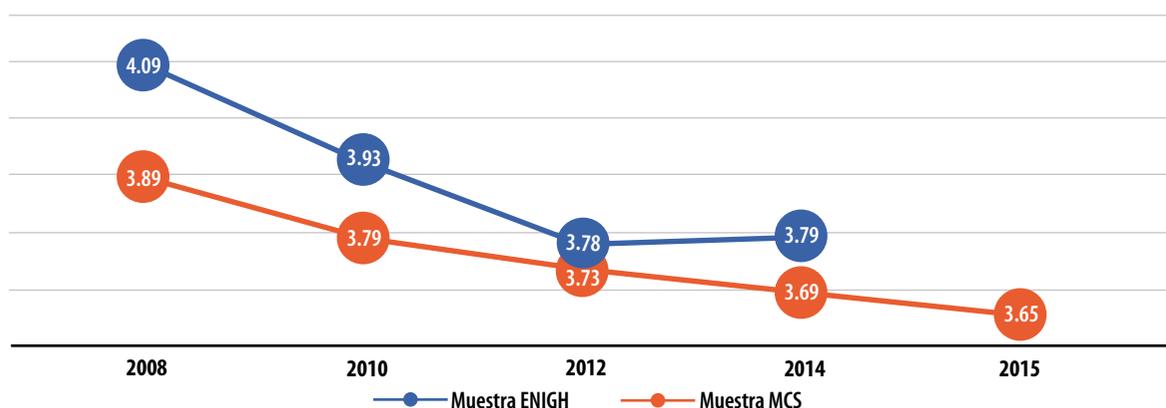
Como se observa en la gráfica 2.1, del 2014 al 2015 la tasa de crecimiento promedio anual de los hogares según los MCS es de 3.3%, mientras que la que resulta de las proyecciones de población del CONAPO es de solo 1.9 por ciento. Este aumento en el crecimiento del número de hogares se debe, en parte, 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.

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 observa 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.

Gráfica 2.3

Tamaño del hogar por origen de muestra del MCS, 2008-2015



La baja tan pronunciada en el tamaño del hogar que arroja el MCS 2015 puede deberse a la omisión de personas, en términos del operativo de campo, debido a que en ocasiones el informante olvida reportarlos en la primera visita, sin embargo, se logra integrarlos en alguna de las visitas posteriores; esto es más viable para el levantamiento de la muestra ENIGH, dado que el hogar es visitado durante siete días consecutivos. En la muestra MCS es más factible que ese olvido persista, debido a que los cuestionarios pueden ser llenados en tan solo tres días y si en ese periodo no se consigue integrarlos, se pierde su registro. Esto lo podemos constatar en la gráfica 2.3, donde se observa que históricamente la muestra MCS reporta tamaños de hogar más bajos que la muestra ENIGH. Además, dado que el MCS 2015 solo tuvo muestra MCS, el tamaño del hogar sigue su tendencia a la baja. En cierta forma, se puede decir que el tamaño del hogar de la muestra ENIGH jala hacia arriba el tamaño del hogar de toda la muestra, permitiendo que no sea tan bajo. En el MCS 2015 esto no fue posible por la falta de muestra ENIGH.

2.2 Métodos de imputación

La mayoría de los sistemas de información requiere que le proporcionen datos (entrada o captura de datos) para su funcionamiento, los cuales necesitan ser validados.

En la entrada, los datos deben cubrir ciertas características para poder ser aceptados como correctos al momento de ser capturados. El proceso que permite verificar que esos datos cumplen con las características necesarias para ser aceptados se le llama validación. Para llevar a cabo este proceso, es necesario, primero, identificar el tipo de variable a tratar, ésta puede ser cualitativa (por ejemplo, el género y el nivel educativo de una persona) o cuantitativa (por ejemplo, la edad e ingreso de la persona); y segundo, construir reglas de validación, las cuales pueden involucrar una o más variables.

Por otro lado, no solo los sistemas de información requieren que la entrada de datos sea validada, por ejemplo, el INEGI cada año desarrolla un número importante de proyectos de naturaleza estadística, como censos y encuestas, en los cuales se implementa la validación en varias etapas: comenzando por la captación de los datos, siguiendo con la entrada de éstos a los sistemas de captura y finalizando con la validación de las bases de datos. A medida que el proceso avanza en las etapas, la validación se va haciendo más compleja, sobre todo porque se incrementa el número de variables involucradas en una regla de validación.

Aunado a lo anterior, al llevar a cabo los procesos de validación, pueden detectarse algunas variables con ausencia de información, lo cual puede darse por dos situaciones diferentes; en un primer caso, este escenario puede ser correcto: una persona en la variable *edad* reporta 3 años y el *nivel de instrucción* está vacío; en un segundo caso, esa ausencia puede deberse a una inconsistencia: una persona que reporta en *edad* 15 años y el *nivel de instrucción* está vacío (debería aparecer un código: por lo menos *sin instrucción*). En el segundo caso es necesario imputar (asignar) un dato a la variable *nivel de instrucción* para que pase el proceso de validación.

Al respecto veamos algunas definiciones y/o comentarios.

Palacios y González (2004) nos dice que: "...por imputación entenderemos la asignación de nuevos valores a aquellas variables, en uno o más registros, que necesitan ser corregidos para pasar el proceso de validación".

De acuerdo con Schafer (2006): "...la imputación es la práctica de rellenar datos faltantes con valores probables, es una forma atractiva de analizar datos incompletos. Esto aparentemente resuelve el problema de análisis de datos incompletos. Sin embargo, un nuevo o poco escrupuloso método de imputación podría crear más problemas que los que resuelve, distorsionando estimaciones, errores estándar y pruebas de hipótesis".

Según Juárez (2004): "...las imputaciones son extracciones de una distribución predictiva de los datos faltantes y requiere de un método para crear una distribución predictiva para la imputación basada en datos observados. La imputación es un método general y flexible para manejar problemas de datos faltantes".

Dempster y Rubin (1983) consideran que: "La idea de la imputación es seductora y peligrosa. Seductora porque puede adormecer al usuario en un estado placentero de creer que los datos están completos después de todo y esto es peligroso porque junta situaciones en donde el problema es suficientemente menor que puede ser legítimamente manejado en esta forma y situaciones en donde estimadores estándares aplicados a los datos reales e imputados tienen sesgos substanciales".

En pocas palabras, imputar es un procedimiento para asignar un valor a una variable con no respuesta para que el cuestionario pase el proceso de validación. Además, como señalan los autores mencionados, la imputación puede ser una buena solución al trabajar con datos incompletos, pero también tiene sus riesgos, ya que de no hacerse este procedimiento con el mayor de los cuidados, los datos finales pueden quedar muy sesgados y, por lo tanto, no mostrar la realidad que se pretende descubrir por medio de una encuesta.

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 esta 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*¹ 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 INEGI 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 del mismo evento, 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 éste (3.65) y que parte del objetivo del

¹ Kim y Fuller (2004) mencionan que la imputación *Hot-Deck* aleatoria preserva las propiedades distributivas del conjunto de datos imputado, además de introducir variabilidad al seleccionar aleatoriamente al donante. Esto es aplicable al método *Cold-Deck* aleatorio.

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 , $i = 1, \dots, 32$); 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 $TamHOG_{14,1} = 3.83$ 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).

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).

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

Fuente: cálculos propios con base en las proyecciones de población y hogares del CONAPO.

Tabla 3.2

Coberturas de imputación por entidad federativa

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

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 descendientemente por ese aleatorio.
- Una vez ordenados los hogares por el número aleatorio, se busca el primer hogar que cumpla con las condiciones de 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.

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

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.

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 encontrarán 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. En la tabla 3.7 se puede observar el total de imputados por la clase de hogar y parentesco.

Además, para 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.

Tabla 3.7

Total de imputados por clase de hogar y parentesco

Clase de hogar	Parentesco	Imputados	%
Total		3 759	100.0
Nuclear	Total	2 662	70.8
	Hijo(a), hijo(a) consanguíneo(a), hijo(a) reconocido(a)	2 574	68.5
	Hijo(a) adoptivo(a)	1	0.0
	Hijastro(a), entenado(a)	87	2.3
Ampliado	Total	1 062	28.3
	Hijo(a), hijo(a) consanguíneo(a), hijo(a) reconocido(a)	667	17.7
	Hijastro(a), entenado(a)	10	0.3
	Madre, padre	68	1.8
	Padraastro, madrastra	2	0.1
	Hermano(a)	60	1.6
	Abuelo(a)	3	0.1
	Nieto(a)	215	5.7
	Bisnieto(a)	2	0.1
	Tío(a)	7	0.2
	Sobrino(a)	23	0.6
	Primo(a)	5	0.1
Compuesto	Total	35	0.9
	Hijo(a), hijo(a) consanguíneo(a), hijo(a) reconocido(a)	22	0.6
	Hijastro(a), entenado(a)	1	0.0
	No tiene parentesco	7	0.2
	Nieto(a)	2	0.1
	Bisnieto(a)	1	0.0
	Sobrino(a)	1	0.0
	Primo(a)	1	0.0

- 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.

Además, 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 PEA²).

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

² INEGI. *Encuesta Nacional de Ocupación y Empleo*. Tercer trimestre del 2015.

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:

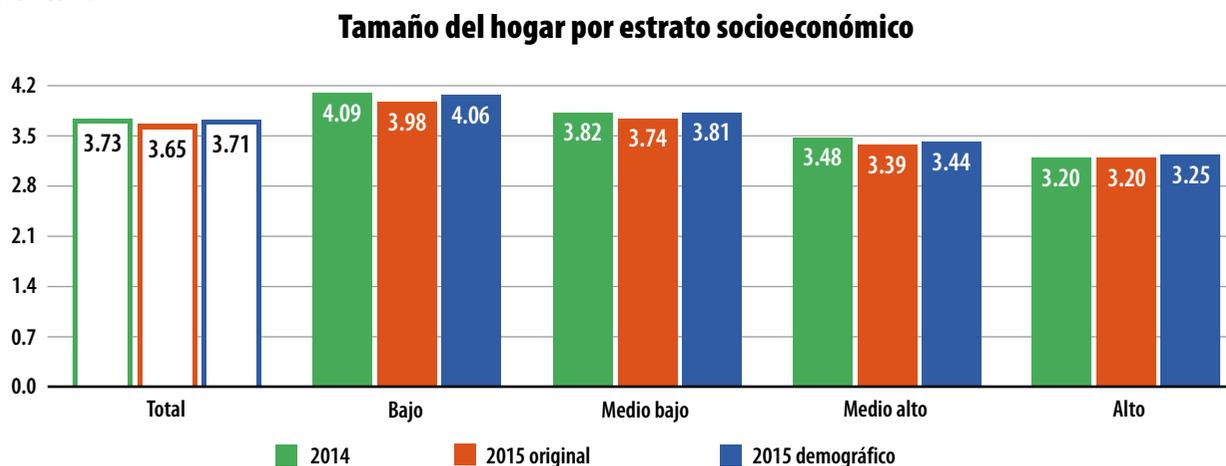
- INEGI. *Módulo de Condiciones Socioeconómicas 2012*.
- INEGI. *Módulo de Condiciones Socioeconómicas 2014*.
- INEGI. *Módulo de Condiciones Socioeconómicas 2015*.
- INEGI. *Encuesta Intercensal 2015*.
- INEGI. *Encuesta de Ocupación y Empleo*. Tercer trimestre del 2015.
- CONAPO. Proyecciones de población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

4. Resultados

4.1 Resultados en el ámbito nacional, por entidad federativa y por decil con la nueva base del MCS 2015

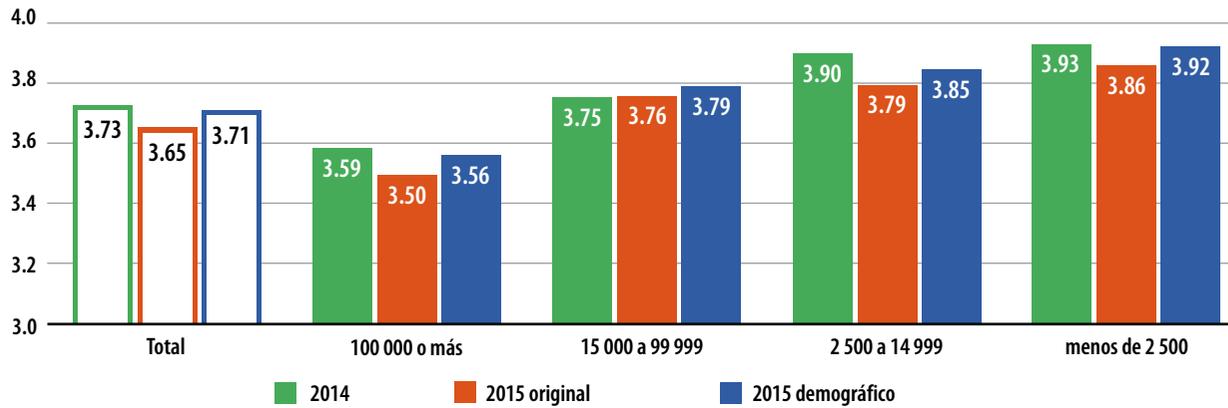
4.2 Comparación de los resultados con los publicados para el MCS 2015 y ENIGH-MCS 2010,

Gráfica 4.1



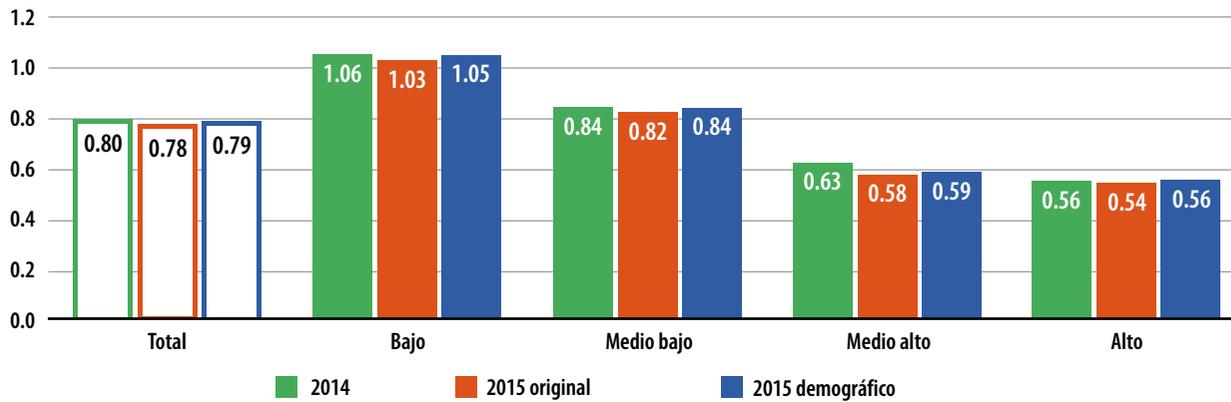
Gráfica 4.2

Tamaño del hogar por tamaño de localidad



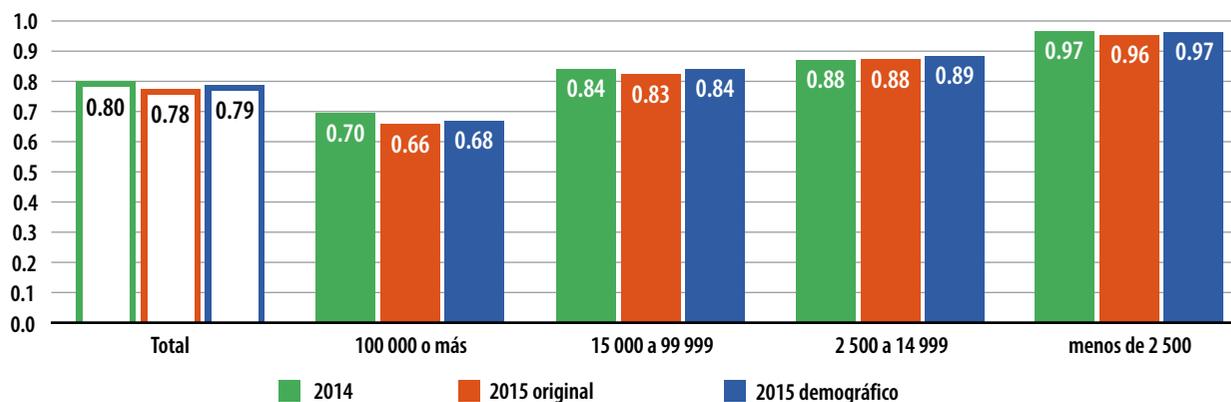
Gráfica 4.3

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



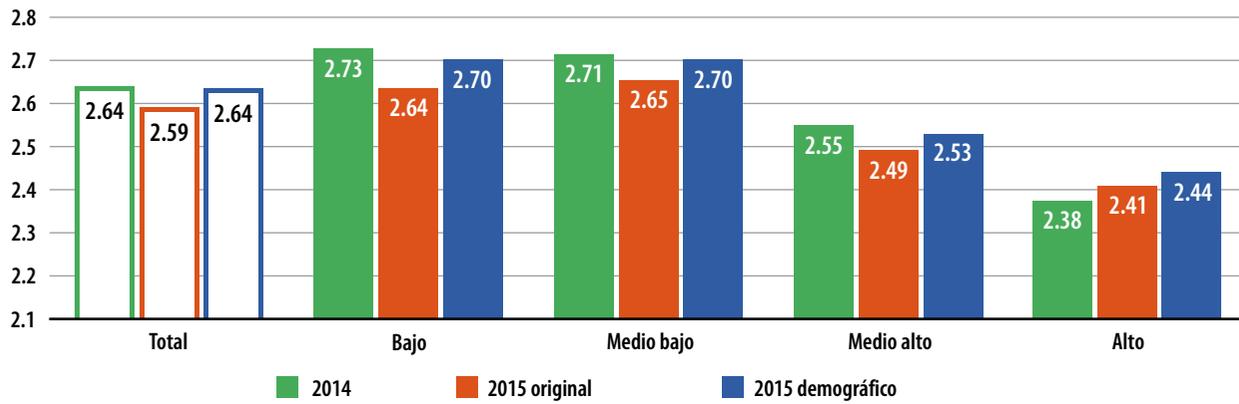
Gráfica 4.4

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



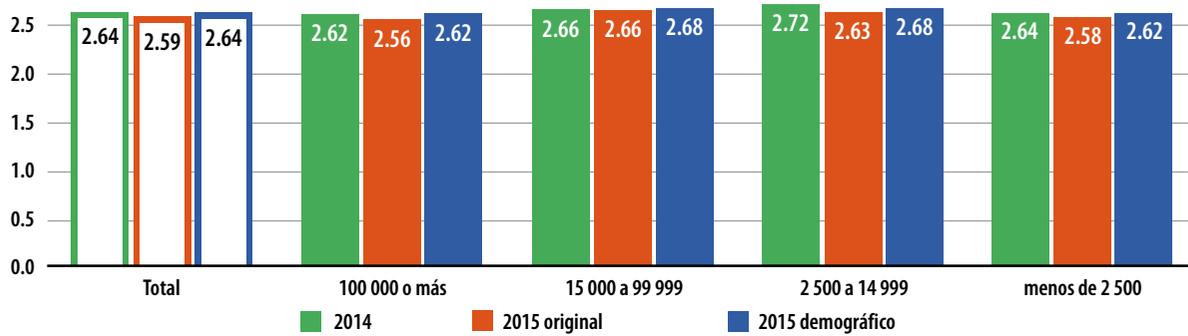
Gráfica 4.5

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



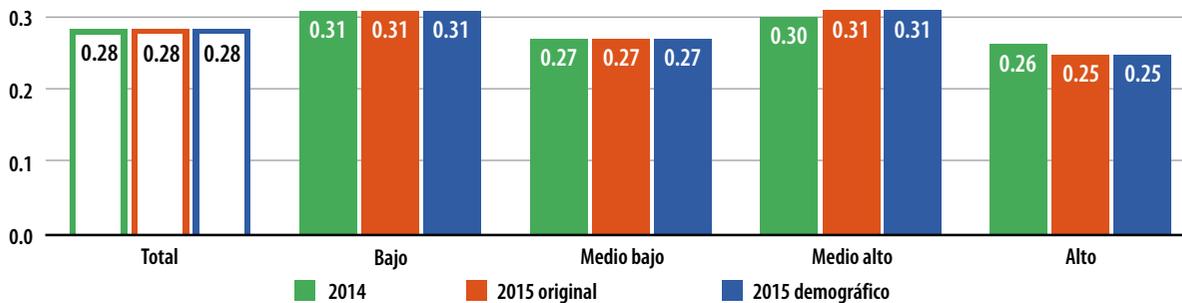
Gráfica 4.6

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



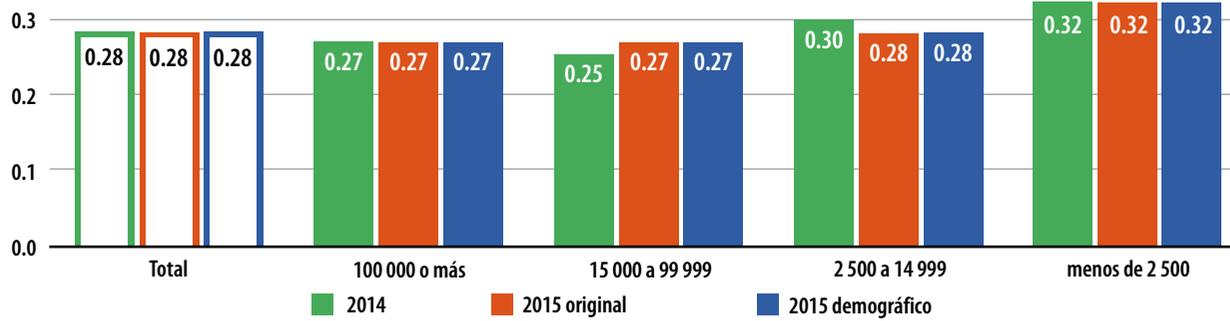
Gráfica 4.7

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



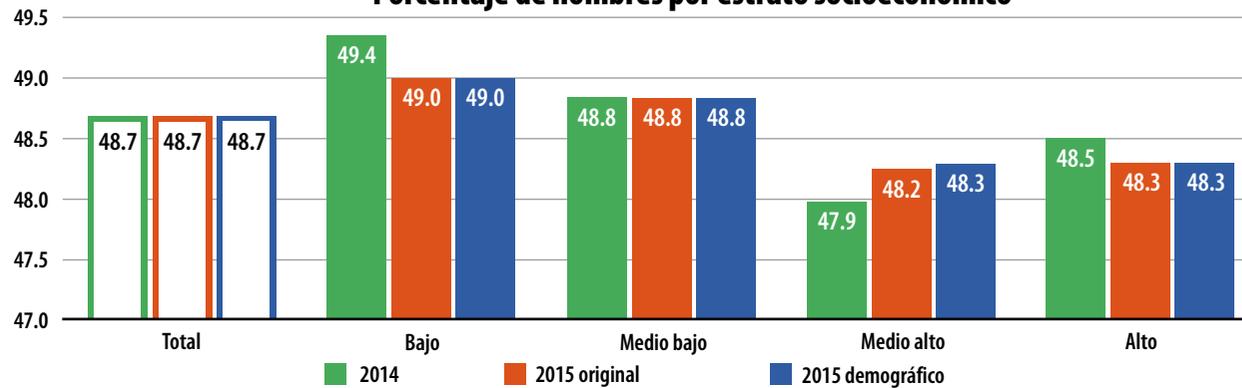
Gráfica 4.8

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



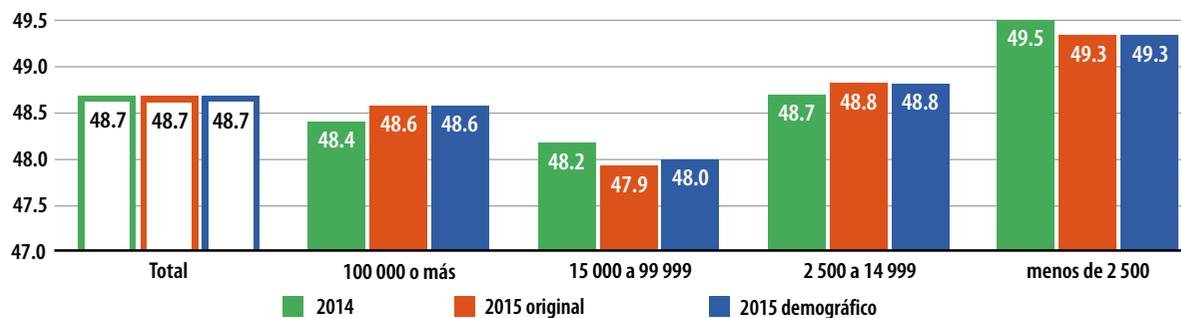
Gráfica 4.9

Porcentaje de hombres por estrato socioeconómico



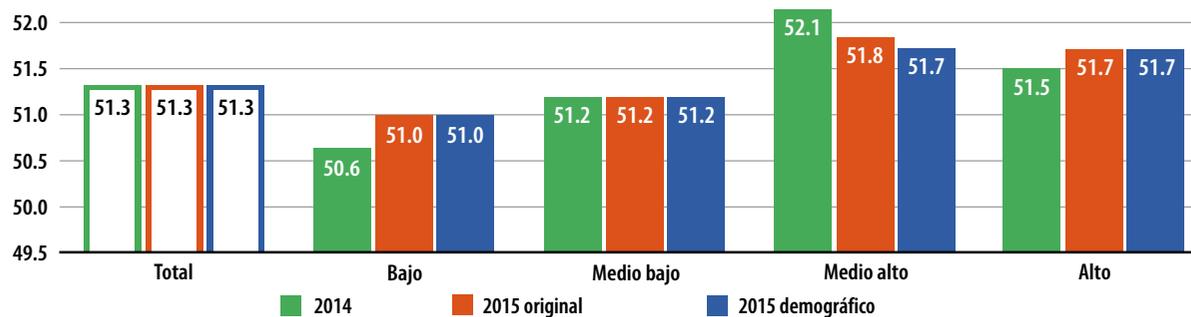
Gráfica 4.10

Porcentaje de hombres por tamaño de localidad



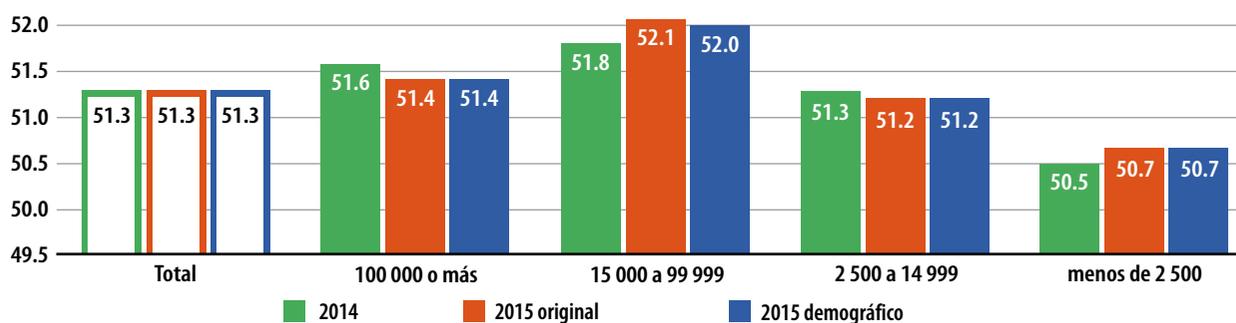
Gráfica 4.11

Porcentaje de mujeres por estrato socioeconómico



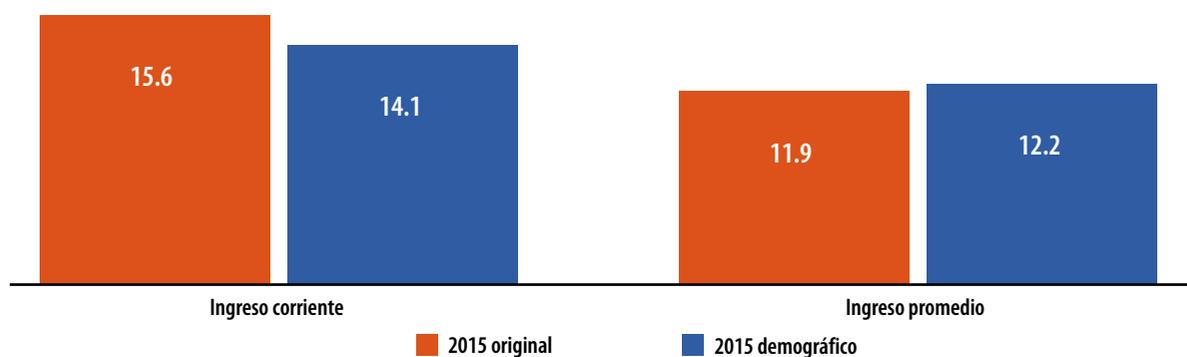
Gráfica 4.12

Porcentaje de mujeres por tamaño de localidad



Gráfica 4.13

Variación porcentual del ingreso corriente 2014-2015



2012 y 2014. A nivel nacional, por entidad federativa y por decil

Gráfica 4.14

Promedio por hogar del ingreso corriente total 2010-2015

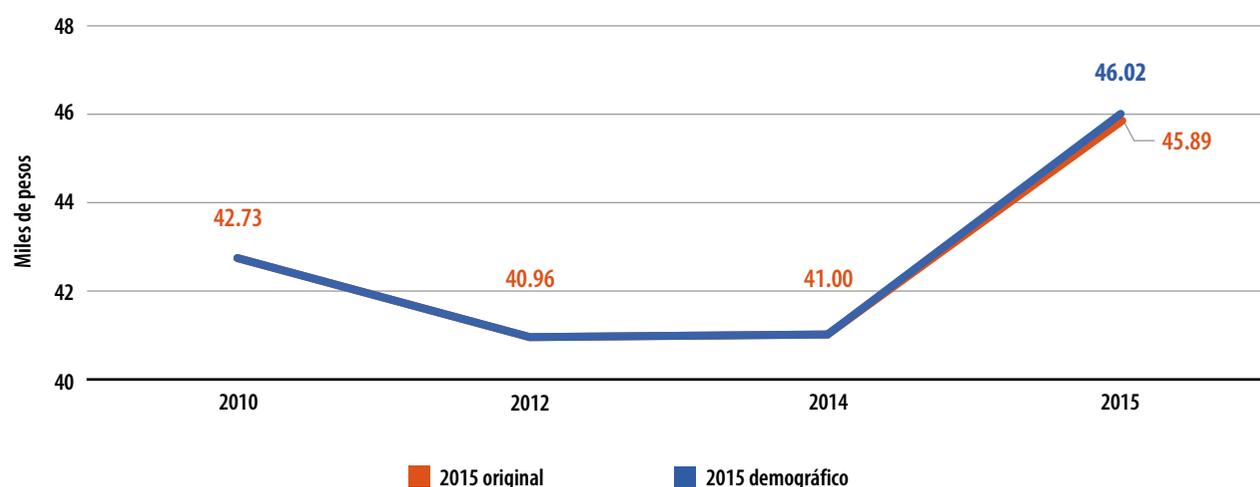


Tabla 4.1

Continúa

Variaciones porcentuales del tamaño de los hogares, del ingreso total y del ingreso promedio por hogar para cada entidad federativa

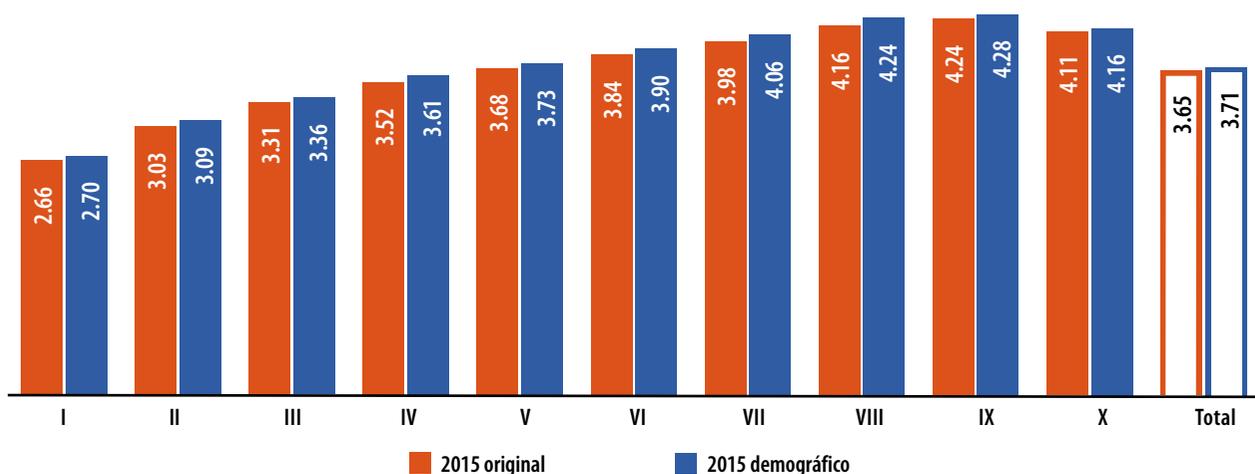
Entidad	2014-2015 original			2014-2015 demográfico		
	Hogares	Ingreso corriente	Ingreso promedio por hogar	Hogares	Ingreso corriente	Ingreso promedio por hogar
Estados Unidos Mexicanos	3.32	15.64	11.92	1.65	14.10	12.25
Aguascalientes	2.61	28.31	25.05	2.45	27.74	24.68
Baja California	7.67	19.96	11.42	2.23	13.96	11.47
Baja California Sur	7.02	33.13	24.40	3.50	28.64	24.29
Campeche	1.42	16.03	14.40	1.39	15.88	14.30
Coahuila de Zaragoza	3.17	6.98	3.68	2.10	7.44	5.23
Colima	5.81	12.45	6.27	2.54	9.00	6.29
Chiapas	0.31	32.66	32.26	0.33	32.75	32.31
Chihuahua	2.84	48.55	44.45	1.50	46.38	44.22
Ciudad de México	3.56	-1.68	-5.06	1.05	-4.16	-5.15
Durango	3.64	29.56	25.01	1.84	27.20	24.90
Guanajuato	1.31	31.69	29.99	1.14	33.86	32.35
Guerrero	6.21	9.39	2.99	1.18	3.85	2.64
Hidalgo	1.80	17.05	14.99	1.96	16.53	14.29

Variaciones porcentuales del tamaño de los hogares, del ingreso total y del ingreso promedio por hogar para cada entidad federativa

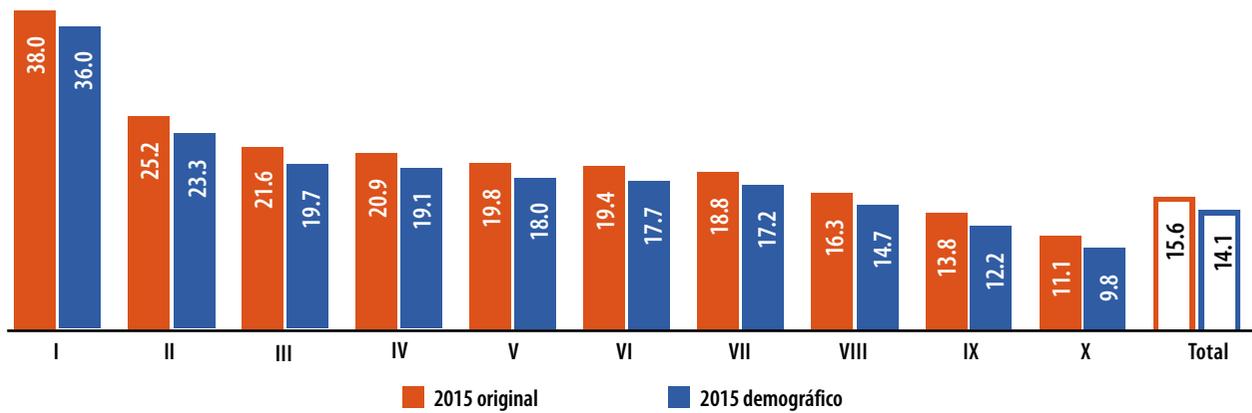
Entidad	2014-2015 original			2014-2015 demográfico		
	Hogares	Ingreso corriente	Ingreso promedio por hogar	Hogares	Ingreso corriente	Ingreso promedio por hogar
Jalisco	-0.78	15.03	15.94	-0.54	16.58	17.21
México	4.44	10.33	5.64	2.35	8.27	5.79
Michoacán de Ocampo	5.67	31.56	24.50	1.56	26.77	24.83
Morelos	4.94	26.64	20.67	1.81	25.28	23.05
Nayarit	3.29	22.19	18.30	2.41	20.99	18.15
Nuevo León	4.91	14.24	8.90	2.37	11.56	8.98
Oaxaca	6.91	24.65	16.59	1.25	18.40	16.93
Puebla	3.64	4.92	1.24	1.92	2.79	0.85
Querétaro	5.24	21.88	15.82	2.75	20.05	16.83
Quintana Roo	7.16	1.62	-5.17	3.85	-0.71	-4.39
San Luís Potosí	5.01	26.02	20.00	1.80	22.10	19.94
Sinaloa	2.68	35.95	32.40	1.67	34.68	32.46
Sonora	2.20	14.63	12.16	2.12	14.57	12.18
Tabasco	2.11	20.27	17.79	2.09	20.05	17.59
Tamaulipas	3.52	11.01	7.23	2.06	9.38	7.17
Tlaxcala	3.18	26.63	22.74	2.52	26.29	23.19
Veracruz de Ignacio de la Llave	0.49	8.65	8.11	2.10	11.74	9.43
Yucatán	4.94	23.11	17.32	2.23	19.90	17.29
Zacatecas	-1.32	13.57	15.09	-1.27	13.90	15.36

Gráfica 4.15

Tamaño promedio del hogar por deciles



Gráfica 4.16

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

Gráfica 4.17

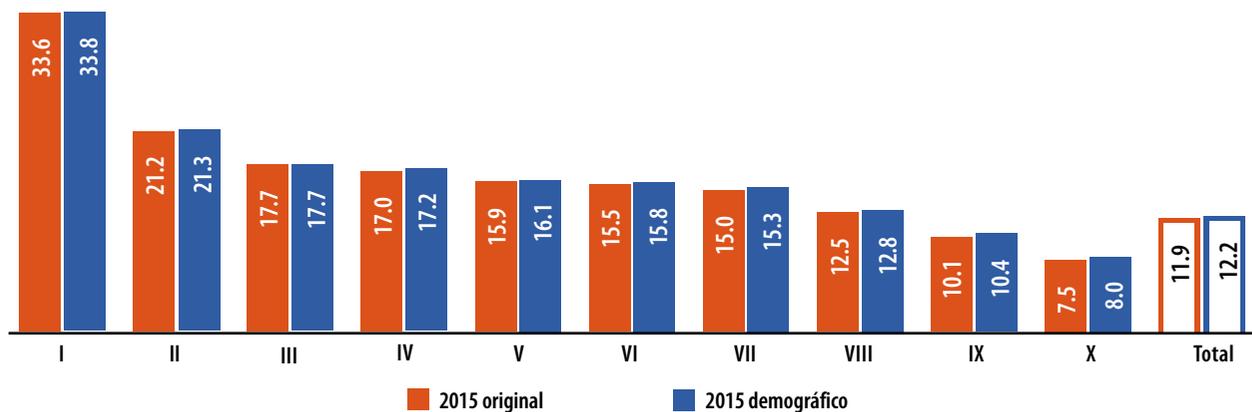
Promedio por hogar de ingreso corriente total por deciles. Variación porcentual 2014-2015

Tabla 4.2

Descomposición del ingreso corriente total (ICT)

Decil	2015 original				2015 demográfico			
	ICT ₁₅ / ICT ₁₄	IPP	PH	Hog	ICT ₁₅ / ICT 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.

Tabla 4.3

Pobreza multidimensional en MCS 2015 original

Entidad federativa	Pobreza	Pobreza moderada	Pobreza extrema	Población vulnerable por carencias	Población vulnerable por ingresos	Población no pobre y no vulnerable
Total	36.69%	30.85%	5.85%	33.94%	5.45%	23.91%
Aguascalientes	25.88%	24.38%	1.50%	33.05%	7.72%	33.35%
Baja California	24.19%	22.52%	1.67%	38.80%	5.94%	31.08%
Baja California Sur	22.63%	20.92%	1.71%	40.11%	4.60%	32.66%
Campeche	34.16%	29.66%	4.50%	39.89%	2.89%	23.06%
Coahuila de Zaragoza	19.01%	17.62%	1.40%	30.92%	10.49%	39.57%
Colima	25.99%	24.08%	1.91%	41.03%	5.03%	27.95%
Chiapas	65.54%	40.22%	25.32%	24.11%	1.94%	8.41%
Chihuahua	21.14%	20.25%	0.89%	34.73%	8.07%	36.07%
Ciudad de México	25.54%	24.68%	0.86%	28.92%	6.98%	38.56%
Durango	32.37%	30.12%	2.25%	35.01%	7.59%	25.03%
Guanajuato	30.04%	28.31%	1.73%	43.74%	4.65%	21.56%
Guerrero	59.35%	41.44%	17.91%	32.51%	1.80%	6.33%
Hidalgo	41.41%	34.09%	7.32%	39.34%	3.20%	16.06%
Jalisco	24.91%	22.50%	2.41%	38.70%	5.35%	31.04%
México	40.69%	36.37%	4.31%	27.86%	8.62%	22.83%
Michoacán de Ocampo	44.42%	35.79%	8.64%	38.64%	2.98%	13.96%
Morelos	36.78%	31.36%	5.42%	38.47%	3.93%	20.82%
Nayarit	31.40%	27.12%	4.28%	38.81%	5.47%	24.32%
Nuevo León	17.49%	16.52%	0.96%	30.29%	8.05%	44.18%
Oaxaca	59.73%	40.74%	18.99%	28.54%	1.60%	10.14%
Puebla	50.11%	42.87%	7.24%	32.76%	4.30%	12.83%
Querétaro	24.03%	21.67%	2.36%	40.37%	4.95%	30.65%
Quintana Roo	30.01%	25.01%	4.99%	39.76%	4.80%	25.43%
San Luis Potosí	36.50%	30.47%	6.03%	34.40%	4.88%	24.22%
Sinaloa	27.47%	25.72%	1.74%	40.37%	4.94%	27.22%
Sonora	22.46%	20.39%	2.07%	36.84%	5.68%	35.02%
Tabasco	36.68%	31.01%	5.67%	50.50%	1.64%	11.17%
Tamaulipas	31.99%	28.70%	3.29%	31.75%	6.92%	29.34%
Tlaxcala	43.39%	40.59%	2.80%	33.70%	6.39%	16.52%
Veracruz de Ignacio de la Llave	50.72%	40.26%	10.46%	33.28%	2.89%	13.11%
Yucatán	34.36%	28.59%	5.76%	36.94%	4.80%	23.91%
Zacatecas	44.69%	40.09%	4.60%	30.49%	5.01%	19.81%

Tabla 4.4

Pobreza multidimensional en MCS 2015 demográfico

Entidad federativa	Pobreza	Pobreza moderada	Pobreza extrema	Población vulnerable por carencias	Población vulnerable por ingresos	Población no pobre y no vulnerable
Total	37.32%	31.34%	5.98%	33.25%	5.63%	23.80%
Aguascalientes	26.12%	24.61%	1.51%	32.75%	7.81%	33.32%
Baja California	26.33%	24.67%	1.66%	36.63%	7.00%	30.04%
Baja California Sur	24.37%	22.45%	1.93%	38.47%	5.35%	31.81%
Campeche	34.20%	29.67%	4.53%	39.92%	2.89%	22.99%
Coahuila de Zaragoza	18.96%	17.57%	1.39%	30.65%	10.11%	40.28%
Colima	27.95%	25.63%	2.32%	39.29%	5.84%	26.92%
Chiapas	65.51%	40.21%	25.30%	24.18%	1.93%	8.38%
Chihuahua	21.57%	20.70%	0.88%	34.43%	8.03%	35.97%
Ciudad de México	26.33%	25.37%	0.96%	28.23%	7.50%	37.94%
Durango	32.62%	30.22%	2.40%	34.70%	7.76%	24.92%
Guanajuato	29.65%	27.93%	1.72%	44.20%	4.56%	21.59%
Guerrero	61.77%	42.52%	19.24%	30.29%	1.88%	6.06%
Hidalgo	41.13%	33.79%	7.34%	38.93%	3.54%	16.40%
Jalisco	24.66%	22.34%	2.33%	38.78%	5.26%	31.30%
México	41.63%	37.17%	4.45%	27.10%	8.78%	22.49%
Michoacán de Ocampo	46.66%	37.66%	9.00%	36.14%	3.56%	13.64%
Morelos	37.89%	32.22%	5.68%	36.92%	4.27%	20.91%
Nayarit	31.75%	27.39%	4.37%	38.52%	5.50%	24.23%
Nuevo León	18.16%	16.92%	1.24%	29.85%	8.62%	43.37%
Oaxaca	61.74%	41.60%	20.15%	26.44%	1.82%	9.99%
Puebla	51.00%	43.62%	7.38%	31.94%	4.36%	12.70%
Querétaro	24.91%	22.41%	2.50%	38.96%	5.29%	30.84%
Quintana Roo	31.35%	26.33%	5.02%	38.16%	5.48%	25.00%
San Luis Potosí	37.57%	31.33%	6.24%	33.60%	5.36%	23.46%
Sinaloa	28.04%	26.15%	1.89%	39.86%	5.06%	27.05%
Sonora	22.50%	20.42%	2.08%	36.86%	5.67%	34.96%
Tabasco	36.75%	31.06%	5.69%	50.47%	1.65%	11.13%
Tamaulipas	32.55%	29.33%	3.22%	31.28%	7.06%	29.11%
Tlaxcala	43.83%	40.95%	2.87%	33.10%	6.48%	16.60%
Veracruz de Ignacio de la Llave	49.82%	39.73%	10.09%	33.00%	2.74%	14.44%
Yucatán	35.78%	29.70%	6.08%	35.74%	4.89%	23.58%
Zacatecas	44.63%	40.03%	4.60%	30.51%	5.00%	19.86%

Tabla 4.5

Pobreza multidimensional. Población con carencias en MCS 2015 original

Entidad federativa	Población con al menos una carencia	Población con al menos tres carencias	Población con carencia por rezago educativo	Población con carencia por acceso a servicios de salud	Población con carencia por acceso a la seguridad social	Población con carencia por calidad y espacios de la vivienda	Población con carencia de acceso a servicios básicos de la vivienda	Población con carencia por acceso a la alimentación
Total	70.64%	20.53%	17.95%	16.87%	56.59%	12.01%	20.48%	21.75%
Aguascalientes	58.93%	8.56%	13.76%	13.77%	42.93%	4.90%	4.70%	18.79%
Baja California	62.99%	12.45%	13.57%	18.10%	46.39%	9.34%	12.59%	14.95%
Baja California Sur	62.74%	14.61%	14.17%	12.76%	43.01%	14.12%	11.57%	21.72%
Campeche	74.05%	27.45%	20.14%	12.35%	57.60%	18.36%	34.04%	23.17%
Coahuila de Zaragoza	49.93%	7.50%	12.29%	13.75%	31.04%	5.31%	4.59%	16.33%
Colima	67.02%	13.68%	17.98%	13.36%	51.94%	9.09%	9.56%	19.53%
Chiapas	89.66%	47.06%	29.96%	18.84%	81.21%	28.94%	54.57%	25.04%
Chihuahua	55.86%	7.36%	15.80%	11.94%	39.64%	4.96%	4.21%	12.92%
Ciudad de México	54.46%	5.79%	8.91%	19.07%	45.19%	4.85%	1.67%	9.92%
Durango	67.37%	11.29%	16.01%	15.37%	50.49%	5.70%	11.67%	17.81%
Guanajuato	73.79%	16.52%	20.28%	14.17%	56.88%	8.77%	12.77%	26.14%
Guerrero	91.86%	50.22%	25.40%	16.44%	76.68%	33.44%	59.36%	40.65%
Hidalgo	80.75%	21.47%	17.60%	18.29%	69.22%	9.71%	24.49%	21.57%
Jalisco	63.61%	12.98%	18.78%	17.13%	49.06%	7.74%	9.55%	15.23%
México	68.55%	16.00%	13.19%	19.91%	57.70%	9.96%	10.21%	20.19%
Michoacán de Ocampo	83.07%	32.84%	27.20%	23.12%	69.97%	16.24%	26.47%	32.04%
Morelos	75.25%	21.76%	17.18%	16.06%	61.21%	13.75%	22.16%	26.19%
Nayarit	70.21%	15.43%	15.95%	15.97%	56.57%	7.54%	16.84%	19.50%
Nuevo León	47.78%	6.57%	11.55%	12.34%	30.94%	4.74%	4.22%	15.55%
Oaxaca	88.27%	45.78%	27.92%	16.86%	76.94%	20.89%	59.46%	32.49%
Puebla	82.87%	26.90%	21.91%	18.92%	71.25%	13.83%	26.96%	23.95%
Querétaro	64.40%	14.63%	14.92%	12.87%	50.51%	8.62%	17.66%	17.46%
Quintana Roo	69.77%	21.00%	15.31%	16.48%	52.11%	18.06%	18.85%	22.24%
San Luis Potosí	70.90%	20.93%	17.80%	9.83%	56.22%	14.09%	29.92%	16.85%
Sinaloa	67.84%	14.06%	15.78%	14.69%	49.11%	8.23%	14.05%	23.31%
Sonora	59.30%	11.87%	12.55%	13.33%	37.71%	8.78%	9.93%	23.67%
Tabasco	87.18%	35.17%	17.80%	17.10%	70.47%	11.64%	46.13%	39.79%
Tamaulipas	63.73%	12.61%	16.21%	15.02%	46.12%	8.94%	10.44%	19.61%
Tlaxcala	77.09%	14.92%	13.84%	14.84%	66.95%	10.76%	11.06%	21.51%
Veracruz de Ignacio de la Llave	84.00%	35.60%	26.03%	19.21%	68.33%	19.48%	40.88%	30.09%
Yucatán	71.30%	27.66%	21.81%	14.91%	52.70%	18.24%	38.86%	16.87%
Zacatecas	75.18%	13.22%	19.10%	12.37%	61.59%	4.29%	14.43%	20.03%

Tabla 4.6

Pobreza multidimensional. Población con carencias en MCS 2015 demográfico

Entidad federativa	Población con al menos una carencia	Población con al menos tres carencias	Población con carencia por rezago educativo	Población con carencia por acceso a servicios de salud	Población con carencia por acceso a la seguridad social	Población con carencia por calidad y espacios de la vivienda	Población con carencia de acceso a servicios básicos de la vivienda	Población con carencia por acceso a la alimentación
Total	70.57%	20.43%	17.77%	16.87%	56.55%	12.05%	20.38%	21.61%
Aguascalientes	58.86%	8.52%	13.74%	13.68%	42.90%	4.84%	4.69%	18.80%
Baja California	62.96%	12.25%	13.29%	17.94%	46.58%	9.26%	12.15%	14.71%
Baja California Sur	62.84%	14.54%	14.00%	12.84%	43.17%	14.07%	11.49%	21.76%
Campeche	74.12%	27.54%	20.17%	12.35%	57.69%	18.39%	34.17%	23.23%
Coahuila de Zaragoza	49.61%	7.35%	12.08%	13.62%	30.87%	5.09%	4.49%	16.51%
Colima	67.24%	13.77%	17.61%	13.57%	52.35%	9.17%	9.63%	19.53%
Chiapas	89.69%	47.07%	29.95%	18.87%	81.25%	28.98%	54.56%	25.03%
Chihuahua	56.00%	7.32%	15.72%	11.94%	39.75%	4.94%	4.21%	12.97%
Ciudad de México	54.56%	5.82%	8.86%	19.09%	45.33%	4.85%	1.72%	9.85%
Durango	67.32%	11.31%	15.93%	15.38%	50.54%	5.67%	11.63%	17.79%
Guanajuato	73.85%	16.56%	20.34%	14.49%	57.00%	8.94%	12.61%	25.85%
Guerrero	92.06%	50.59%	25.08%	16.60%	77.01%	33.63%	60.07%	40.56%
Hidalgo	80.06%	21.34%	17.42%	18.33%	68.66%	9.56%	24.29%	21.40%
Jalisco	63.44%	12.84%	18.67%	17.14%	48.91%	7.69%	9.33%	15.25%
México	68.73%	15.95%	13.16%	19.82%	57.86%	10.09%	10.14%	20.22%
Michoacán de Ocampo	82.80%	32.13%	26.45%	22.99%	69.81%	15.98%	26.04%	31.44%
Morelos	74.82%	21.52%	16.73%	16.21%	60.88%	13.59%	21.49%	26.07%
Nayarit	70.27%	15.46%	15.92%	15.92%	56.68%	7.58%	16.87%	19.55%
Nuevo León	48.01%	6.55%	11.33%	12.55%	31.46%	4.49%	4.35%	15.50%
Oaxaca	88.18%	45.51%	27.22%	17.04%	76.98%	20.69%	59.20%	32.23%
Puebla	82.94%	26.89%	21.73%	18.87%	71.26%	13.82%	27.30%	23.91%
Querétaro	63.87%	14.47%	14.71%	13.03%	50.14%	8.47%	16.95%	17.54%
Quintana Roo	69.52%	20.78%	14.97%	16.35%	52.18%	18.00%	18.75%	21.75%
San Luis Potosí	71.18%	20.91%	17.68%	10.17%	56.45%	14.05%	29.90%	16.86%
Sinaloa	67.90%	14.04%	15.71%	14.64%	49.19%	8.26%	14.05%	23.24%
Sonora	59.36%	11.94%	12.57%	13.37%	37.77%	8.79%	9.98%	23.71%
Tabasco	87.22%	35.21%	17.81%	17.08%	70.48%	11.66%	46.21%	39.82%
Tamaulipas	63.83%	12.52%	16.12%	14.96%	46.22%	8.92%	10.44%	19.57%
Tlaxcala	76.92%	14.81%	13.72%	14.86%	66.81%	10.61%	11.02%	21.30%
Veracruz de Ignacio de la Llave	82.82%	34.97%	25.44%	18.91%	66.95%	20.33%	40.14%	29.06%
Yucatán	71.52%	27.83%	21.77%	14.84%	53.05%	18.33%	39.22%	16.87%
Zacatecas	75.13%	13.21%	19.08%	12.39%	61.55%	4.28%	14.44%	20.01%

5. Validación y evaluación de la metodología

Aunque el ajuste demográfico no tiene gran impacto en la disminución del ingreso y aumento de la pobreza 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 arrojará ingresos más altos que lo que se venían captando en los eventos anteriores. Así, pone el piso base para que se realicen ejercicios de alineación a los eventos anteriores.

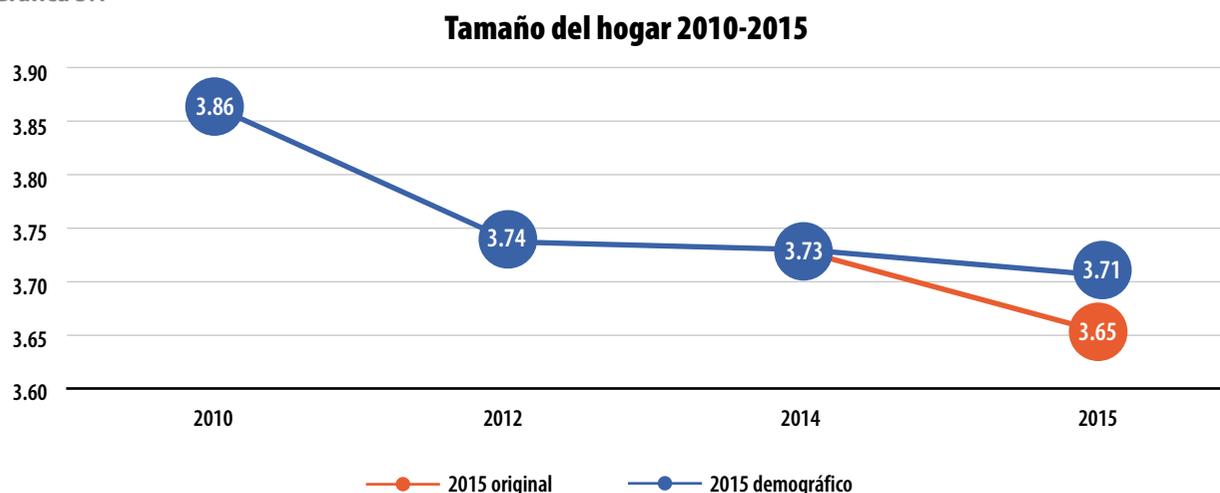
Sin duda, los métodos de imputación *Deck* arrojan resultados satisfactorios cuando existe un conjunto muy grande de posibles donadores como en este ejercicio. Sin embargo, cuando el conjunto de posibles donadores es pequeño, para un gran número de casos existe la posibilidad de que se encuentren el 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.

Así como en el ingreso, en las carencias no hay grandes efectos de que preocuparse. Como se observa en el cuadro 5.3, las diferencias en las entidades federativas entre el MCS 2015 original y el MCS 2015 demográfico van desde -1.38 hasta 0.86 por ciento. Aunque estas diferencias mínimas no necesariamente pueden ser aplicables a todas las variables, para conocer los efectos a profundidad del ajuste demográfico es necesario hacer un análisis sobre aquellas que se puedan considerar las más importantes del MCS 2015.

Gráfica 5.1



Gráfica 5.2

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

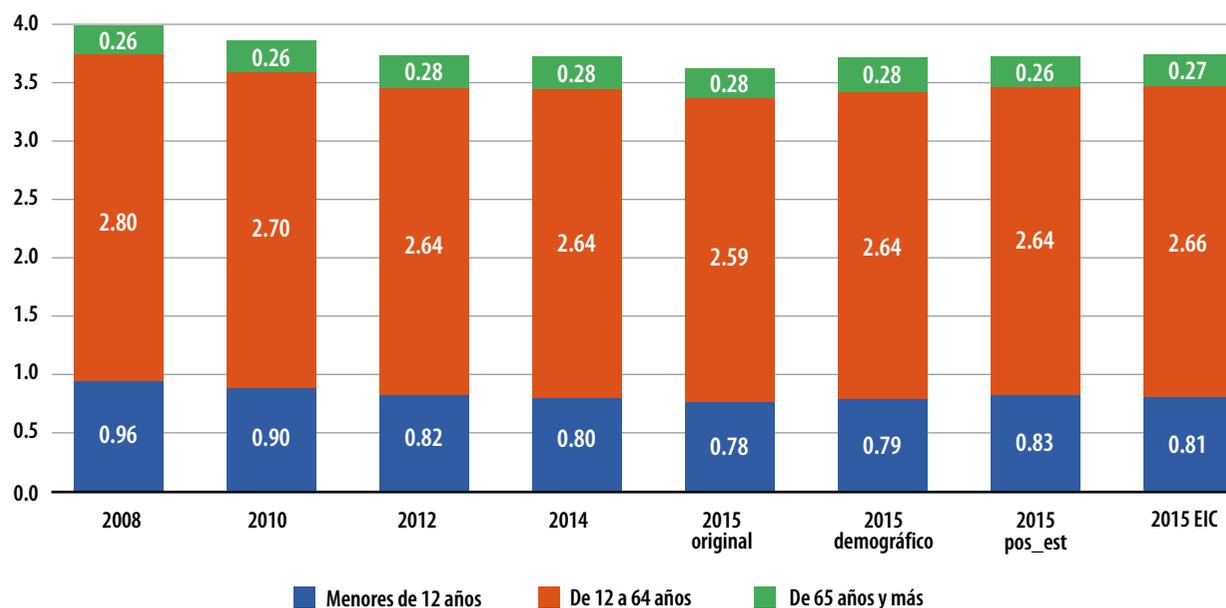


Tabla 5.1

Pobreza multidimensional. Diferencias 2015 demográfico vs. 2015 original, medidas entre las entidades federativas

Medida	Población con al menos una carencias	Población con al menos tres carencias	Población con carencia por rezago educativo	Población con carencia por acceso a servicios de salud	Población con carencia por acceso a la seguridad social	Población con carencia por calidad y espacios de la vivienda	Población con carencia de acceso a servicios básicos de la vivienda	Población con carencia por acceso a la alimentación
Mínimo	-1.18	-0.71	-0.74	-0.30	-1.38	-0.26	-0.73	-1.02
Máximo	0.27	0.37	0.06	0.33	0.52	0.86	0.71	0.18
Media	-0.06	-0.07	-0.18	0.02	0.00	-0.01	-0.07	-0.10
Desviación estándar	0.30	0.20	0.20	0.14	0.33	0.19	0.29	0.23
Mediana	0.05	-0.03	-0.11	0.01	0.06	-0.03	-0.02	-0.02

6. Conclusión y comentarios finales

Los resultados del ejercicio logran el objetivo de incrementar el tamaño del hogar de acuerdo con las tasas de variación implícitas calculadas a partir de las proyecciones de hogares y población del CONAPO y como consecuencia disminuir el total de hogares que reportó el MCS 2015.

Los efectos en el ingreso y la pobreza son marginales, ya que logra disminuir muy poco el monto del ingreso corriente y aumentar, también muy poco, al promedio de ingreso por hogar (debido a la disminución de los hogares) y la pobreza, esto respecto al original.

Las restricciones en cuanto a sexo, edad, tamaño de localidad y estrato socioeconómico, aunado a la aleatoriedad del ejercicio, logran una distribución consistente de la población para esas variables.

Probablemente existan otros métodos de imputación más efectivos que los *Deck*, pero es un hecho que éstos son los más simples y fáciles de implementar, y sus resultados se pueden considerar satisfactorios. Sin embargo, es posible explorar en la literatura otros métodos más efectivos y compararlos con éstos, en estudios futuros.

7. 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 Ostría, 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

8. Anexos

Tabla 8.1

Continúa

Variables demográficas

	Total	Estrato				Tamaño de localidad			
		Bajo	Medio bajo	Medio alto	Alto	100 mil o más	15 mil a 99 999	2 500 a 14 999	Menos de 2 500
Integrantes del hogar									
2014	119 897 652	23 889 861	60 786 821	24 819 266	10 401 704	57 207 724	17 627 470	17 260 987	27 801 471
2015 original	121 381 730	24 017 970	62 122 995	24 673 826	10 566 939	57 932 027	17 866 706	17 470 197	28 112 800
2015 demográfico	121 381 494	24 025 184	61 861 654	24 823 977	10 670 679	57 931 828	17 866 097	17 470 402	28 113 167

Variables demográficas

	Total	Estrato				Tamaño de localidad			
		Bajo	Medio bajo	Medio alto	Alto	100 mil o más	15 mil a 99 999	2 500 a 14 999	Menos de 2 500
Menores de 12 años									
2014	25 865 041	6 159 128	13 433 923	4 463 723	1 808 267	11 140 730	3 961 036	3 883 818	6 879 457
2015 original	25 874 659	6 222 485	13 607 517	4 253 593	1 791 064	10 948 270	3 928 672	4 036 002	6 961 715
2015 demográfico	25 951 155	6 211 846	13 625 007	4 290 008	1 824 294	10 980 741	3 981 272	4 032 474	6 956 668
12 a 64 años									
2014	84 970 122	15 944 673	43 088 074	18 200 206	7 737 169	41 785 936	12 469 412	12 065 644	18 649 130
2015 original	86 079 732	15 926 893	44 018 306	18 170 193	7 964 340	42 510 060	12 642 329	12 129 427	18 797 916
2015 demográfico	86 138 251	15 968 263	43 839 374	18 299 244	8 031 370	42 520 713	12 631 671	12 154 212	18 831 655
65 años y más									
2014	9 062 489	1 786 060	4 264 824	2 155 337	856 268	4 281 058	1 197 022	1 311 525	2 272 884
2015 original	9 427 339	1 868 592	4 497 172	2 250 040	811 535	4 473 697	1 295 705	1 304 768	2 353 169
2015 demográfico	9 292 088	1 845 075	4 397 273	2 234 725	815 015	4 430 374	1 253 154	1 283 716	2 324 844
Hombres									
2014	58 366 324	11 790 126	29 653 461	11 879 887	5 042 850	27 695 664	8 492 877	8 406 539	13 771 244
2015 original	59 099 914	11 773 368	30 319 001	11 904 056	5 103 489	28 136 511	8 563 786	8 529 336	13 870 281
2015 demográfico	59 119 189	11 768 571	30 211 132	11 986 030	5 153 456	28 146 732	8 577 120	8 524 750	13 870 587
Mujeres									
2014	61 531 328	12 099 735	31 133 360	12 939 379	5 358 854	29 512 060	9 134 593	8 854 448	14 030 227
2015 original	62 281 816	12 244 602	31 803 994	12 769 770	5 463 450	29 795 516	9 302 920	8 940 861	14 242 519
2015 demográfico	62 262 305	12 256 613	31 650 522	12 837 947	5 517 223	29 785 096	9 288 977	8 945 652	14 242 580

Tabla 8.2a

Insumos para la descomposición del ICT

Decil	Ingreso corriente total ^a			Hogares (Hog)			Perceptores			Ingreso por perceptor (IPP)			Perceptores por hogar (PH)		
	2014	2015 original	Razón	2014	2015 original	Razón	2014	2015 original	Razón	2014	2015 original	Razón	2014	2015 original	Razón
	Total	1 318 162	1 524 263	1.156	32 150 400	33 218 037	1.033	75 381 871	80 911 445	1.073	17 486	18 839	1.077	2.34	2.44
I	19 662	27 137	1.380	3 215 040	3 321 804	1.033	5 925 336	6 226 511	1.051	3 318	4 358	1.313	1.84	1.87	1.017
II	36 680	45 939	1.252	3 215 040	3 321 804	1.033	6 725 440	6 778 853	1.008	5 454	6 777	1.243	2.09	2.04	0.976
III	50 224	61 079	1.216	3 215 040	3 321 804	1.033	6 856 385	7 219 234	1.053	7 325	8 461	1.155	2.13	2.17	1.019
IV	63 588	76 890	1.209	3 215 040	3 321 804	1.033	7 166 736	7 563 063	1.055	8 873	10 167	1.146	2.23	2.28	1.021
V	78 976	94 598	1.198	3 215 040	3 321 804	1.033	7 320 377	7 934 187	1.084	10 789	11 923	1.105	2.28	2.39	1.049
VI	97 024	115 818	1.194	3 215 040	3 321 804	1.033	7 720 394	8 203 636	1.063	12 567	14 118	1.123	2.40	2.47	1.028
VII	120 349	142 938	1.188	3 215 040	3 321 804	1.033	8 046 727	8 739 751	1.086	14 956	16 355	1.094	2.50	2.63	1.051
VIII	154 661	179 838	1.163	3 215 040	3 321 804	1.033	8 492 589	9 207 690	1.084	18 211	19 531	1.072	2.64	2.77	1.049
IX	213 831	243 329	1.138	3 215 040	3 321 804	1.033	8 650 216	9 558 109	1.105	24 720	25 458	1.030	2.69	2.88	1.069
X	483 167	536 696	1.111	3 215 040	3 321 804	1.033	8 477 671	9 480 410	1.118	56 993	56 611	0.993	2.64	2.85	1.082

Nota: se incluye solo a los perceptores de ingreso monetario, excluyendo a quienes reportan ingreso no monetario.

^a Millones de pesos constantes a precios del 2015.

Tabla 8.2b

Insumos para la descomposición del ICT

Decil	Ingreso corriente total ^a			Hogares (Hog)			Perceptores			Ingreso por perceptor (IPP)			Perceptores por hogar (PH)		
	2014	2015 demográfico	Razón	2014	2015 demográfico	Razón	2014	2015 demográfico	Razón	2014	2015 demográfico	Razón	2014	2015 demográfico	Razón
	Total	1 318 162	1 504 051	1.141	32 150 400	32 681 856	1.017	75 381 871	79 496 030	1.055	17 486	18 920	1.082	2.34	2.43
I	19 662	26 747	1.360	3 215 040	3 268 186	1.017	5 925 336	6 115 158	1.032	3 318	4 374	1.318	1.84	1.87	1.015
II	36 680	45 238	1.233	3 215 040	3 268 186	1.017	6 725 440	6 674 012	0.992	5 454	6 778	1.243	2.09	2.04	0.976
III	50 224	60 103	1.197	3 215 040	3 268 186	1.017	6 856 385	7 101 789	1.036	7 325	8 463	1.155	2.13	2.17	1.019
IV	63 588	75 756	1.191	3 215 040	3 268 186	1.017	7 166 736	7 445 925	1.039	8 873	10 174	1.147	2.23	2.28	1.022
V	78 976	93 221	1.180	3 215 040	3 268 186	1.017	7 320 377	7 765 773	1.061	10 789	12 004	1.113	2.28	2.38	1.044
VI	97 024	114 213	1.177	3 215 040	3 268 186	1.017	7 720 394	8 064 370	1.045	12 567	14 163	1.127	2.40	2.47	1.028
VII	120 349	141 016	1.172	3 215 040	3 268 186	1.017	8 046 727	8 595 184	1.068	14 956	16 406	1.097	2.50	2.63	1.051
VIII	154 661	177 390	1.147	3 215 040	3 268 186	1.017	8 492 589	9 051 806	1.066	18 211	19 597	1.076	2.64	2.77	1.049
IX	213 831	240 004	1.122	3 215 040	3 268 186	1.017	8 650 216	9 372 792	1.084	24 720	25 606	1.036	2.69	2.87	1.066
X	483 167	530 363	1.098	3 215 040	3 268 186	1.017	8 477 671	9 309 222	1.098	56 993	56 972	1.000	2.64	2.85	1.080

Nota: se incluye solo a los perceptores de ingreso monetario, excluyendo a quienes reportan ingreso no monetario.

^a Millones de pesos constantes a precios del 2015.