

# Propuesta para recuperar la continuidad del MCS 2015 usando máquinas de soporte vectorial

## *A proposal for recovering MCS 2015's continuity using Support Vector Machines*

**José Alejandro Ruiz Sánchez\***

\* Instituto Nacional de Estadística y Geografía (INEGI), jose.ruiz@inegi.org.mx

**Nota:** se agradece la colaboración e ideas proporcionadas por Miriam Romo y Benito Durán.



Robot sitting on chair, reading book/imagewerks/Getty Images

Debido a procesos de mejora llevados a cabo durante el levantamiento de la información del Módulo de Condiciones Socioeconómicas (MCS) 2015, la distribución del ingreso corriente total cambió de manera sustantiva con respecto a lo que se venía observando en levantamientos previos. Esta propuesta intenta recomponer la continuidad, modificando la representatividad que tenía un hogar en el 2015, de acuerdo con su asignación a un decil pronosticado o artificial creado a partir de un conjunto de variables monetarias y no-monetarias. Para asignar un hogar a un decil artificial, se usó una máquina de soporte vectorial. Como conjunto de entrenamiento se usaron los levantamientos

ENIGH-MCS 2010, 2012 y 2014. Una vez realizada la imputación de deciles se ajustaron los factores de expansión de cada hogar, esto permite que hogares con ingresos bajos sean sobreexpandidos y que, para hogares con ingresos altos, su factor de expansión sea reducido.

**Palabras clave:** Módulo de Condiciones Socioeconómicas; MCS; continuidad; SVM; factores de expansión.

Due to improvements in the Socio-economic Module's data collection 2015 (MCS 2015 by its Spanish acronym), the total income distribution shifted from what has been observed in recent years. Consequently, it is no longer possible to make income comparison with previous information. This methodology is intended to recompose continuity by modifying household weights in 2015, according to its classification into some artificial decile generated by monetary and non-monetary variables. To classify a household we use a technique called Support Vector Machine (SVM) and historic data (2010-2014) as a training data set. Once an artificial decile is imputed to a household in the MCS 2015, we modify its weight so that adding them up by decile represent 10 percent. This allows low-income households to be overexpanded and high-income households to be underexpanded.

**Key words:** Socio-economic Module; MCS; Continuity; SVM; weights.

## 1. Introducción

La información relacionada con los ingresos de los hogares captada en el Módulo de Condiciones Socioeconómicas (MCS) 2015 perdió continuidad con levantamientos anteriores del mismo módulo; sin embargo, ello no invalidó ni volvió erróneos o sesgados los datos obtenidos. Es en este sentido que el problema de continuidad histórica puede ser abordado como uno de (re) clasificación de hogares, donde los conjuntos generados estén constituidos por observaciones con características similares.

Una manifestación de la pérdida de continuidad entre levantamientos con respecto a variables de ingresos se dio al realizar análisis por deciles: las tasas de crecimiento del 2014 al 2015 del ingreso corriente total (ICT) de cada decil fueron muy superiores a las que se venían observando históricamente.

El acercamiento metodológico en este documento consiste en pronosticar el decil al cual debería pertenecer cada hogar en el 2015 a partir de información monetaria y no-monetaria, de forma tal que los hogares que pertenezcan a determinado decil sean consistentes con hogares en deciles iguales, pero de años anteriores. Con base en esta clasificación pronosticada o artificial, se hace un ajuste en los factores de expansión para otorgarle mayor o menor peso a cada hogar, de acuerdo con el decil al que haya sido asignado. De esta forma, hogares con ingresos bajos tendrán un mayor peso muestral que el que tenían originalmente, contrario a los de ingresos altos, quienes verán disminuido su peso muestral. Para pronosticar el decil de pertenencia de los hogares en el 2015, se utilizaron técnicas de *Machine Learning* para clasificación, específicamente máquinas de soporte vectorial (SVM, por sus siglas en inglés).

Así, la presente propuesta se desarrolló con el objetivo de generar una base de datos alternativa al MCS 2015 que recomponga la trayectoria histórica del ingreso y que tenga como prioridad el no modificar lo declarado por los hogares. Para lograr cambios, entonces, se adecuarán los factores de expansión de tal forma que se modifique la representatividad original que tiene el ingreso declarado. El cambio en los factores de expansión se hará a partir de asignar un hogar a un decil pronosticado, al que teóricamente debería pertenecer, de acuerdo con valores que este hogar reportó sobre un conjunto de variables monetarias y no-monetarias. Por otra parte, y como consecuencia, cuando se usan los nuevos factores de expansión, es posible que la representatividad de subgrupos poblacionales en los cuales se basó el diseño muestral se vea alterada.

El documento se encuentra organizado en seis capítulos; en el segundo de ellos se presenta una descripción de los datos que dan sustento a la propuesta metodológica. En el tercer apartado se desarrolla la técnica de SVM, que es utilizada para la categorización y pronóstico; además, se muestra la forma para modificar los factores de expansión. En el cuarto capítulo se presentan resultados desglosados por entidad federativa y por decil. En el quinto apartado se realiza una validación de la técnica con años anteriores y en el último se muestran las conclusiones.

## 2. Marco teórico

Características propias al levantamiento del MCS 2015 generaron que la distribución del vector de ingresos se recorriera de manera importante con respecto a años anteriores (ver figura 2.1).

Ello implica que al hacer comparaciones que consideren agrupaciones a través de ingresos (como *decilizaciones*) se estén mezclando grupos con niveles de ingreso muy distintos. En el caso de las comparaciones por deciles del MCS 2014 y del MCS 2015 se están incluyendo hogares en un decil que, bajo otras circunstancias de recolección de información, pertenecerían a otro (ver figura 2.2).

Figura 2.1  
**Gráfica de densidad del ICT (precios del 2015)**

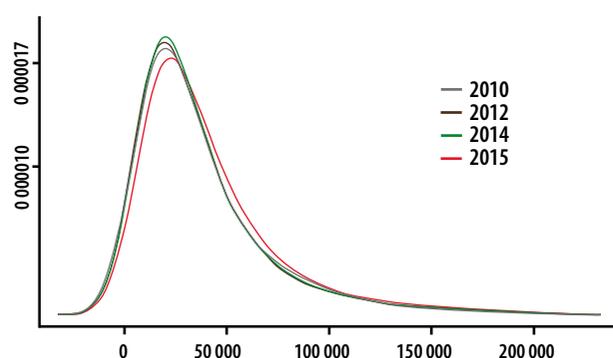
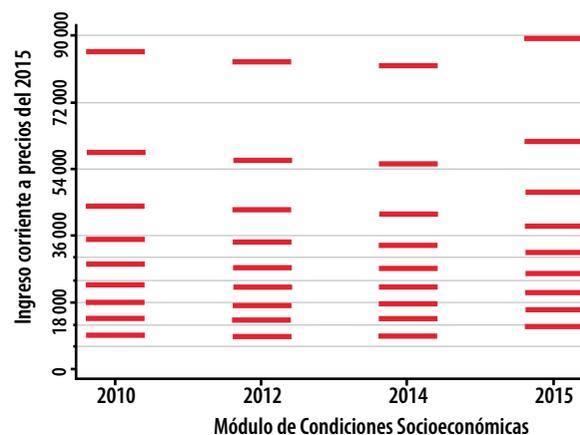


Figura 2.2  
**Puntos de corte para la decilización**



El mecanismo que permita usar información del MCS para generar mediciones consistentes en el tiempo debe incluir el comportamiento histórico de los datos para realizar algún tipo de ajuste en el MCS 2015. Si bien un camino es el ajuste directo al vector de ingresos, la propuesta del presente documento contempla respetar la información proporcionada por los hogares, actuando por medio de la modificación de los factores de expansión. La idea intuitiva del método es generar continuidad a través de la imputación (a cada hogar del 2015) del decil de pertenencia, usando el comportamiento histórico (2010, 2012 y 2014) de *decilización* y la relación que guarda éste con variables monetarias y/o no-monetarias.

Este método será efectivo en la medida en que las técnicas de clasificación sean certeras y en que los ajustes en los factores de expansión hagan que los hogares de determinado decil  $x$  sean representativos de los hogares que, bajo circunstancias similares a las de levantamientos anteriores al 2015, hubiesen pertenecido al mismo decil  $x$ .

### 3. Metodología

#### 3.1 Su aplicación

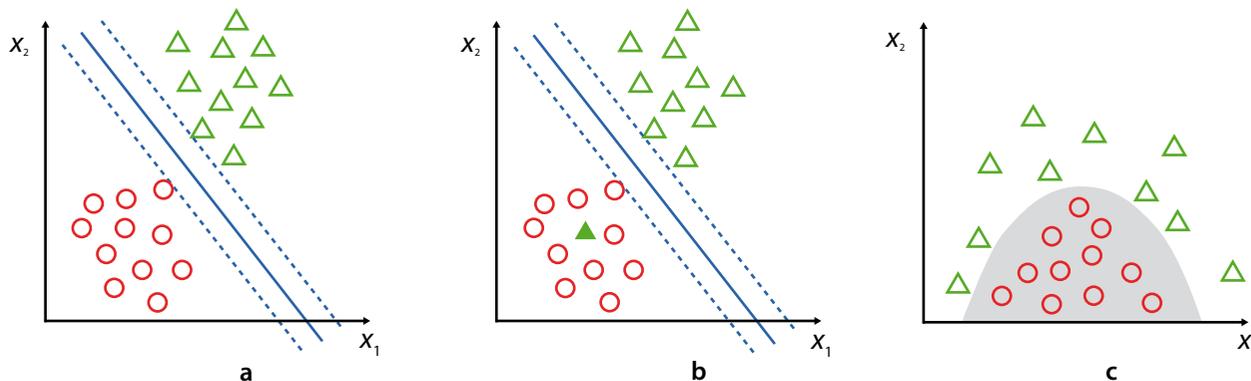
Existen distintos métodos que permiten pronosticar la clasificación de nuevas observaciones con base en parámetros calibrados a partir de una base de datos alterna donde las observaciones ya se encuentran categorizadas. Un conjunto de estos métodos se encuentra en lo que conocemos como *Machine Learning*, así como en la estadística clásica; algunos de ellos son: análisis discriminante lineal (LDA, por sus siglas en inglés), redes neuronales artificiales (ANN, por sus siglas en inglés), árboles de decisión, SVM y regresión logística multinomial, entre otras.

Aunque se utilizaron distintas técnicas para comparar niveles de precisión, solo desarrollaremos la técnica SVM, ya que los resultados que se presentan son con base en ella y por haber sido la que generó resultados aceptables de manera consistente bajo diversas especificaciones. La idea intuitiva de SVM es generar hiperplanos separadores que permitan identificar aquellas observaciones pertenecientes a una misma categoría (ver figura 3.1a).

Para su construcción, los hiperplanos deben ser tales que se tenga el menor error de clasificación; es decir, las clasificaciones pueden no ser perfectas (ver figura 3.1b). Para pronosticar la clasificación de una observación nueva, primero se entrena a la máquina con una base de datos ya clasificada y, a partir de los resultados, la máquina decide el grupo de pertenencia de la nueva observación.

Figura 3.1

#### Representación gráfica de un hiperplano separador



De manera formal,<sup>1</sup> sea un conjunto de  $N$  parejas  $(x_1, y_1), (x_2, y_2) \dots (x_N, y_N)$ , donde  $x_i \in \mathbb{R}^p$  y  $y_i \in \{-1, 1\}$ . La variable  $y_i$  define la categoría de pertenencia y  $x$  el conjunto de covariables.

Sea el hiperplano definido por:

$$\{x: f(x) = x^T \beta + \beta_0 = 0\} \quad (1)$$

donde  $\|\beta\| = 1$ . Una regla de clasificación está dada por la función signo:  $G(x) = \text{signo}[x^T \beta + \beta_0]$ . En el caso de clases totalmente separables, existe una función:

$$f(x) = x^T \beta + \beta_0 = 0, \text{ tal que } y_i * f(x_i) > 0 \forall i \quad (2)$$

y, por lo tanto, podemos encontrar el hiperplano separador que tenga la mayor distancia con respecto a los puntos. A tal distancia se le conoce como margen ( $M$ ).

Entonces, el problema de optimización sería el siguiente:

$$\begin{aligned} & \max M \\ \text{s.a. } & y_i * (x_i^T \beta + \beta_0) \geq M, \quad i = 1, 2, \dots, N \end{aligned} \quad (3)$$

Cuando no puede darse una separación perfecta de las clases, se permite que algunos puntos se encuentren mal clasificados a través de la siguiente especificación:

$$\begin{aligned} & \max M \\ \text{s.a.:} & \\ & y_i * (x_i^T \beta + \beta_0) \geq M(1 - \xi_i), \quad i = 1, 2, \dots, N \end{aligned} \quad (4)$$

$$\xi_i \geq 0 \text{ y } \sum \xi_i \leq C, \text{ donde } C \text{ es una constante}$$

cuya reformulación en términos de  $\beta$  quedaría de la siguiente forma:

$$\begin{aligned} & \min \frac{1}{2} \|\beta\|^2 \\ \text{s.a.:} & \\ & y_i * (x_i^T \beta + \beta_0) \geq (1 - \xi_i), \quad i = 1, 2, \dots, N \end{aligned} \quad (5)$$

$$\xi_i \geq 0 \text{ y } \sum \xi_i \leq C, \text{ donde } C \text{ es una constante.}$$

<sup>1</sup> El presente apartado es un extracto del capítulo 12 del libro *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Para mayor detalle, se recomienda consultarlo directamente.

Formamos la función de Lagrange con sus respectivos multiplicadores  $\alpha_i, \mu_i$ .

$$L = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [y_i (x_i^T \beta + \beta_0) - (1 - \xi_i)] - \sum_{i=1}^N \mu_i \xi_i \quad (6)$$

que en su forma más conocida es representada como:

$$L = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{i'=1}^N \alpha_i \alpha_{i'} y_i y_{i'} \langle h(x_i), h(x_{i'}) \rangle \quad (7)$$

donde  $K(x, x') = \langle h(x), h(x') \rangle$  es la función kernel, la cual puede tomar diferentes especificaciones para ampliar el espacio de las covariables. Las especificaciones más comunes son las siguientes:

$$\begin{aligned} \text{Lineal} \quad K(x, x') &= \langle x, x' \rangle \\ \text{Polinomial de} \\ \text{grado } d \quad K(x, x') &= (\langle x, x' \rangle)^d \\ \text{Radial} \quad K(x, x') &= \exp(-\gamma \|x - x'\|^2) \end{aligned}$$

Dos de estas funciones kernel permiten capturar no linealidades como las de la figura 3.1c.

Para extender la metodología al caso de  $Q$  clases, existen diversos enfoques, dos de los más usados se explican a continuación: enfoque de *clasificación uno vs. uno*, construye  $\binom{Q}{2}$  SVM's, cada uno del cual compara un par de clases. Para clasificar una observación nueva se usan los  $\binom{Q}{2}$  clasificadores y se cuenta el número de veces que dicha observación cae en cada una de las  $Q$  clases. La clasificación final se hace asignando la observación nueva a aquella clase a la cual fue asignada más veces. El segundo es el de *clasificación uno vs. todos*, en el cual se realizan  $Q$  SVM's, comparando una clase con las  $Q-1$  restantes; una nueva observación  $x^*$  será asignada a aquella categoría para la cual  $\beta_{0q} + \beta_{1q} x_1^* + \beta_{2q} x_2^* + \dots + \beta_{pq} x_p^*$  es mayor.

En nuestro caso utilizaremos como datos de entrenamiento los deciles (que tendrán el rol de clases o categorías) y las covariables a nivel hogar de los MCS 2010, 2012 y 2014;<sup>2</sup> con esta información se *entrena* a la máquina para que aprenda a realizar la *decilización* y ese aprendizaje será aplicado a la base de datos del 2015, de tal forma que la máquina de soporte vectorial asignará un decil pronosticado  $x$  a un hogar en el 2015, siempre y cuando los niveles de las covariables para ese hogar sean similares a aquellos hogares pertenecientes al mismo decil  $x$  en años anteriores.

La base de datos de la cual partimos para la *decilización* artificial es aquella proveniente del artículo "Ajuste demográfico por imputación" (publicado en este mismo número especial de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*, y cuyo objetivo es únicamente reconstituir la estructura demográfica del MCS 2015). Una vez obtenido el vector de deciles pronosticados, la siguiente etapa consiste en modificar los factores de expansión, de forma tal que hogares con menores ingresos estén sobrerrepresentados y hogares con ingresos mayores sean subrepresentados.

<sup>2</sup> Las variables monetarias se encuentran en pesos del 2015.

La forma de reasignar pesos muestrales está basada en los deciles pronosticados a los que pertenece cada hogar: cuando un hogar es asignado a un decil pronosticado o artificial  $x$ , esa asignación no considera el factor de expansión de dicho hogar y, por lo tanto, no considera el número de hogares a los que representa. Por otra parte, y dado que los montos del ingreso corriente que definen los cortes de la *decilización* son similares durante el periodo del 2010 al 2014 y que son visiblemente superiores en el 2015 (ver figura 2.2), la *decilización* artificial (la cual considera el comportamiento histórico) categorizará en el decil I a un número de hogares que representarán menos del 10% poblacional, de acuerdo con sus factores de expansión; es decir, el número de hogares muestrales que la máquina de soporte vectorial asignará al decil I será menor al número de hogares que caen en el mismo decil I cuando se consideran los datos del MCS 2015. Dado que ese menor número de hogares en el decil I representa menos del 10% poblacional, habrá que incrementar sus factores de expansión; esto permitirá que hogares con ingresos más bajos sean sobreexpandidos, mientras que los de ingresos mayores verán disminuidos sus factores de expansión; es decir, que representarán a una menor cantidad de hogares.

Cabe mencionar que esta *decilización* artificial solo se utiliza en esta etapa de modificación de factores de expansión; una vez creados los nue-vos factores, nos olvidamos por completo de los deciles artificiales. Para la *decilización* final, usada en los análisis de resultados de la cuarta sección, se recurrió al vector de ingreso corriente total, pero incorporando los nuevos factores de expansión.

De manera formal: sea un hogar  $h$  y  $fh$  su factor de expansión, al formar deciles se debe cumplir que:

$$\frac{\sum_{h \in d} fh}{\sum_{d=1}^{10} \sum_{h \in d} fh} = 0.1 \quad (a)$$

para cada decil  $d$ . Sin embargo, dado que los factores de expansión no son considerados en la *decilización* artificial, nada garantiza que se cumpla la igualdad (a) al utilizar la técnica de SVM para categorizar hogares ya que, en este caso, el número de hogares pertenecientes a cada decil se asigna a partir de un criterio de optimización, donde la condición (a) no tiene rol alguno; por lo tanto, para cumplir con la condición del 10% se generará un nuevo vector de factores de expansión  $fh^{(n)}$ , el cual resultará del ajuste a los factores de expansión originales del MCS 2015.

Sea  $h$  un hogar perteneciente al MCS 2015 y sea  $fh$  el factor de expansión original asociado a ese hogar. Denotemos como  $h_{d,svm}$  a un hogar que fue clasificado en el decil  $d$  utilizando la técnica SVM (decil pronosticado).

Entonces, el factor de ajuste para los hogares del decil pronosticado  $d$  se define como:

$$A^d = \frac{\sum_h fh/10}{\sum_{h_{d,svm}} fh}$$

y el nuevo factor de expansión para un hogar que pertenece al decil pronosticado  $d$ , es:

$$fh_d^{(n)} = (A^d) * fh$$

Este proceso se realiza para cada una de las entidades federativas por separado, con lo cual se procura cuidar la representatividad por entidad en el agregado nacional.

### 3.2 Nota técnica

- El *software* estadístico que se utilizó para la metodología descrita en el apartado anterior fue *R*, a través de su IDE *RStudio* versión 1.0.44. Las librerías usadas fueron MASS y NNET. Para el caso particular de la función SVM, el kernel usado fue el lineal, con los parámetros por *default*.
- Se utilizaron los MCS 2010, 2012 y 2014 como base de datos de entrenamiento para el algoritmo de SVM.

Se usaron tres bases de datos dentro de cada MCS: población, vivienda y hogares. De la segunda se utilizaron variables asociadas a las características físicas de la vivienda, como: tipo de piso, paredes, dotación de agua, etcétera. De la tercera se emplearon tanto el ICT como variables sobre la pertenencia de bienes muebles (autos y televisores, por ejemplo). Por último, de la base de población se usaron índices por hogar sobre características sociodemográficas de sus integrantes: proporción de miembros del hogar que se encuentran estudiando primaria, promedio de edad de los integrantes, proporción de integrantes que cuentan con estudios universitarios, etcétera. La tabla del *Anexo* contiene una descripción de las variables usadas en el 2015, así como el tratamiento que se les dio para su incorporación como covariables dentro de las diversas técnicas<sup>3</sup> de *Machine Learning* usadas. Para los MCS 2010, 2012 y 2014 se utilizaron las mismas variables.

## 4. Resultados

### 4.1 Resultados a nivel nacional, por entidad federativa y por decil con la nueva base del MCS 2015

Una deducción de la lectura del apartado 3 es que a nivel microdatos no se afectó el vector de ingresos reportado por los hogares; son los cambios en los factores de expansión por hogar los que permiten tener resultados distintos a los del MCS 2015.

En esta sección se muestran los principales cambios respecto al ingreso y la pobreza bajo distintos niveles de agregación. La gráfica 4.1 muestra los efectos en el ICT por decil. Con respecto al 2014, el MCS 2015 presentó un incremento de 15.6% en el ingreso corriente total, mientras que con la presente propuesta de ajuste, el incremento para el mismo periodo fue de 4.5 por ciento.

<sup>3</sup> Entre las técnicas probadas se encuentran redes neuronales, árboles de decisión y regresión logística multinomial.

Para los resultados a nivel entidad federativa, se presentan las tablas 4.1 y 4.2. La primera tiene información sobre los ingresos en valores absolutos y a precios del 2015; la segunda, información en términos de variaciones porcentuales con respecto al levantamiento del 2014.

En la tabla 4.2 se puede observar que la tasa de variación tanto del ingreso corriente como del promedio por hogar para todas las entidades federativas disminuyó como consecuencia del ajuste a los factores de expansión.

En términos de fuentes de ingreso, la tabla 4.3 muestra sus variaciones porcentuales entre el 2014 y 2015; este último año considerando el ajuste propuesto.

Si bien los ajustes en las tasas de crecimiento de las variables monetarias son a la baja, faltaría verificar si esta propuesta tiene el efecto esperado en pobreza; es decir, un aumento con respecto al MCS 2015 y un alineamiento con los años anteriores; para ello se presenta una serie de tablas a nivel entidad federativa con los distintos tipos de pobreza calculados por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), los cuales ya incorporan el ajuste correspondiente en el 2015.

La tabla 4.4 muestra los resultados de pobreza para cada uno de los estados y para el país. Debido a que esta metodología considera datos no solo del 2014 sino que también incorpora información anterior de entidades federativas, el comportamiento de los nuevos resultados de pobreza (aquellos resultantes de ajustar factores de expansión, con la técnica descrita en apartados anteriores) es más parecido al 2010 y el 2012 que al 2014.

Gráfica 4.1

### Variación porcentual del ICT para el periodo 2014-2015, por deciles

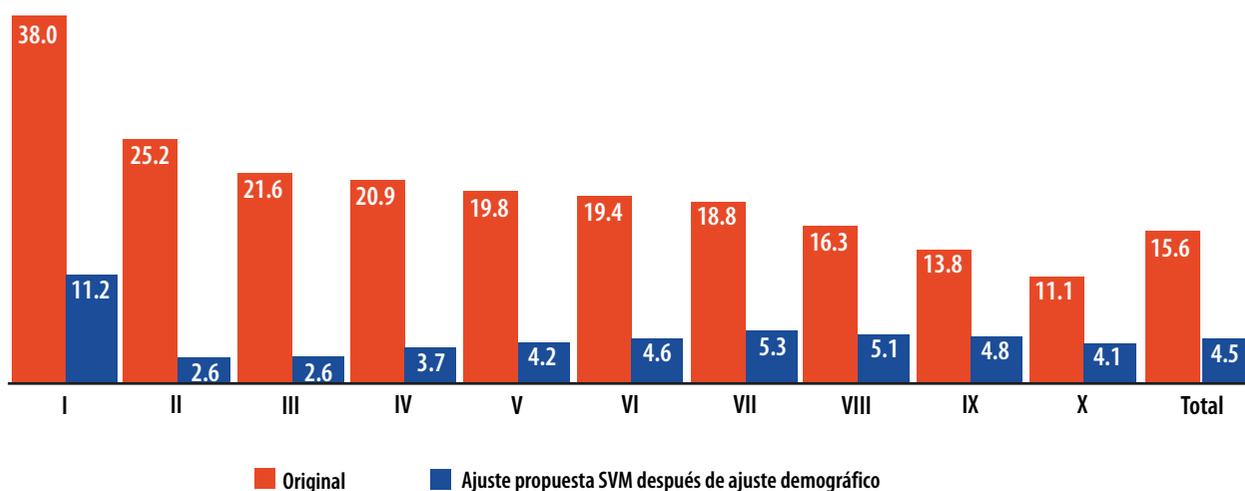


Tabla 4.1

### Ingreso corriente promedio por hogar

Entidad federativa	2012	2014	2015	2015 SVM
1 Aguascalientes	44 149	47 529	59 434	54 942
2 Baja California	49 258	49 017	54 616	50 705
3 Baja California Sur	52 094	49 327	61 361	58 056
4 Campeche	42 933	44 388	50 782	47 178
5 Coahuila de Z.	46 528	51 407	53 301	51 419
6 Colima	43 599	45 443	48 293	44 391
7 Chiapas	24 118	22 417	29 648	25 599
8 Chihuahua	42 886	38 778	56 015	51 596
9 Ciudad de México	56 884	64 907	61 622	58 650
10 Durango	33 552	33 364	41 709	37 473
11 Guanajuato	41 158	35 739	46 457	43 615
12 Guerrero	23 848	26 782	27 584	23 897
13 Hidalgo	32 263	32 792	37 707	33 228
14 Jalisco	46 034	47 369	54 917	52 351
15 México	43 505	39 867	42 114	39 042
16 Michoacán de O.	32 859	30 611	38 112	34 601
17 Morelos	37 761	35 894	43 314	40 062
18 Nayarit	34 639	37 753	44 661	40 077
19 Nuevo León	62 218	61 376	66 836	63 672
20 Oaxaca	27 158	24 629	28 715	24 815
21 Puebla	30 067	36 723	37 176	35 075
22 Querétaro	48 030	48 394	56 048	53 431
23 Quintana Roo	43 276	48 032	45 547	42 309
24 San Luis Potosí	34 364	33 991	40 789	36 695
25 Sinaloa	42 583	43 848	58 056	53 279
26 Sonora	53 234	51 726	58 017	53 874
27 Tabasco	37 267	35 834	42 209	38 127
28 Tamaulipas	41 851	41 170	44 147	40 391
29 Tlaxcala	31 047	30 055	36 888	34 237
30 Veracruz de l. de la Ll.	31 376	28 977	31 328	28 268
31 Yucatán	38 303	41 510	48 698	45 069
32 Zacatecas	31 578	31 061	35 749	32 175
<b>Nacional</b>	<b>40 959</b>	<b>41 000</b>	<b>45 887</b>	<b>41 994</b>

Cifras a precios del 2015.

Tabla 4.2

### Variaciones porcentuales con respecto al MCS inmediato anterior

Entidad federativa	2012-2014		2014-2015		2014-2015 SVM	
	Ingreso corriente	ICT promedio por hogar	Ingreso corriente	ICT promedio por hogar	Ingreso corriente	ICT promedio por hogar
1 Aguascalientes	12.1%	7.7%	28.3%	25.0%	14.8%	15.6%
2 Baja California	1.4%	-0.5%	20.0%	11.4%	2.3%	3.4%
3 Baja California Sur	-2.8%	-5.3%	33.1%	24.4%	16.3%	17.7%
4 Campeche	4.3%	3.4%	16.0%	14.4%	6.7%	6.3%
5 Coahuila de Z.	11.3%	10.5%	7.0%	3.7%	-0.9%	0.0%
6 Colima	7.3%	4.2%	12.4%	6.3%	-2.1%	-2.3%
7 Chiapas	-4.0%	-7.1%	32.7%	32.3%	28.1%	14.2%
8 Chihuahua	-10.8%	-9.6%	48.6%	44.5%	30.6%	33.1%
9 Ciudad de México	12.0%	14.1%	-1.7%	-5.1%	-13.8%	-9.6%
10 Durango	3.2%	-0.6%	29.6%	25.0%	15.6%	12.3%
11 Guanajuato	-8.4%	-13.2%	31.7%	30.0%	20.9%	22.0%
12 Guerrero	10.6%	12.3%	9.4%	3.0%	1.1%	-10.8%
13 Hidalgo	7.6%	1.6%	17.1%	15.0%	8.5%	1.3%
14 Jalisco	11.8%	2.9%	15.0%	15.9%	5.2%	10.5%
15 México	-3.4%	-8.4%	10.3%	5.6%	-1.0%	-2.1%
16 Michoacán de O.	-5.2%	-6.8%	31.6%	24.5%	18.5%	13.0%
17 Morelos	-2.8%	-4.9%	26.6%	20.7%	13.8%	11.6%
18 Nayarit	15.2%	9.0%	22.2%	18.3%	9.6%	6.2%
19 Nuevo León	0.0%	-1.4%	14.2%	8.9%	-0.1%	3.7%
20 Oaxaca	-10.8%	-9.3%	24.6%	16.6%	15.3%	0.8%
21 Puebla	22.2%	22.1%	4.9%	1.2%	-0.7%	-4.5%
22 Querétaro	2.7%	0.8%	21.9%	15.8%	7.4%	10.4%
23 Quintana Roo	18.8%	11.0%	1.6%	-5.2%	-10.5%	-11.9%
24 San Luis Potosí	0.8%	-1.1%	26.0%	20.0%	13.1%	8.0%
25 Sinaloa	5.6%	3.0%	35.9%	32.4%	21.3%	21.5%
26 Sonora	-1.7%	-2.8%	14.6%	12.2%	2.8%	4.2%
27 Tabasco	-2.0%	-3.8%	20.3%	17.8%	9.9%	6.4%
28 Tamaulipas	2.0%	-1.6%	11.0%	7.2%	-0.2%	-1.9%
29 Tlaxcala	0.0%	-3.2%	26.6%	22.7%	16.5%	13.9%
30 Veracruz de I. de la Ll.	-6.5%	-7.6%	8.6%	8.1%	6.5%	-2.4%
31 Yucatán	11.9%	8.4%	23.1%	17.3%	9.8%	8.6%
32 Zacatecas	1.8%	-1.6%	13.6%	15.1%	6.9%	3.6%
<b>Nacional</b>	<b>2.6%</b>	<b>0.1%</b>	<b>15.6%</b>	<b>11.9%</b>	<b>4.5%</b>	<b>2.4%</b>

Tabla 4.3

**Variación porcentual 2014-2015 por fuentes de ingreso**

Decil	Ingreso corriente total		Ingreso por trabajo asalariado		Ingreso por trabajo independiente		Renta de la propiedad	
	2015	2015 SVM	2015	2015 SVM	2015	2015 SVM	2015	2015 SVM
I	38.0%	11.2%	61.5%	-4.4%	54.9%	23.2%	132.1%	70.2%
II	25.2%	2.6%	31.3%	-12.9%	26.9%	8.1%	78.7%	47.5%
III	21.6%	2.6%	17.6%	-10.7%	23.0%	17.1%	79.8%	14.7%
IV	20.9%	3.7%	13.1%	-8.3%	38.5%	18.4%	103.4%	45.6%
V	19.8%	4.2%	12.8%	-6.5%	53.6%	18.0%	76.2%	57.0%
VI	19.4%	4.6%	13.1%	-3.9%	52.3%	12.5%	101.7%	57.0%
VII	18.8%	5.3%	11.8%	-1.6%	36.1%	23.2%	144.1%	112.9%
VIII	16.3%	5.1%	9.7%	-2.5%	20.5%	18.4%	111.4%	69.9%
IX	13.8%	4.8%	6.2%	-2.0%	50.3%	11.1%	102.8%	83.4%
X	11.1%	4.1%	-1.8%	-7.1%	-15.1%	17.0%	63.7%	44.8%
<b>Nacional</b>	<b>15.6%</b>	<b>4.5%</b>	<b>6.4%</b>	<b>-5.0%</b>	<b>22.9%</b>	<b>16.4%</b>	<b>72.3%</b>	<b>50.7%</b>

Tabla 4.4

Continúa

**Porcentaje de población en pobreza**

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	38.1	37.8	34.8	25.9	30.4
Baja California	31.5	30.2	28.6	24.2	30.1
Baja California Sur	31.0	30.1	30.3	22.6	27.7
Campeche	50.5	44.7	43.6	34.2	38.9
Coahuila de Z.	27.8	27.9	30.2	19	22.1
Colima	34.7	34.4	34.3	26	32.4
Chiapas	78.5	74.7	76.2	65.6	70.9
Chihuahua	38.8	35.3	34.4	21.1	25.0
Ciudad de México	28.5	28.9	28.4	25.5	29.1
Durango	51.6	50.1	43.5	32.4	38.3
Guanajuato	48.5	44.5	46.6	30	34.4
Guerrero	67.6	69.7	65.2	59.3	68.1
Hidalgo	54.7	52.8	54.3	41.4	47.5
Jalisco	37.0	39.8	35.4	24.9	27.1

Tabla 4.4

Concluye

**Porcentaje de población en pobreza**

Entidad federativa	2010	2012	2014	2015	2015 SVM
México	42.9	45.3	49.6	40.7	45.8
Michoacán de O.	54.7	54.4	59.2	44.4	52.2
Morelos	43.2	45.5	52.3	36.8	43.5
Nayarit	41.4	47.6	40.5	31.4	37.6
Nuevo León	21.0	23.2	20.4	17.6	21.6
Oaxaca	67.0	61.9	66.8	59.7	68.2
Puebla	61.5	64.5	64.5	50.1	54.1
Querétaro	41.4	36.9	34.2	24	28.5
Quintana Roo	34.6	38.8	35.9	30	35.8
San Luis Potosí	52.4	50.5	49.1	36.5	43.6
Sinaloa	36.7	36.3	39.4	27.5	32.7
Sonora	33.1	29.1	29.4	22.5	26.5
Tabasco	57.1	49.7	49.6	36.7	42.8
Tamaulipas	39.0	38.4	37.9	32	37.5
Tlaxcala	60.3	57.9	58.9	43.3	47.8
Veracruz de I. de la Ll.	57.6	52.6	58.0	50.7	56.8
Yucatán	48.3	48.9	45.9	34.4	40.3
Zacatecas	60.2	54.2	52.3	44.7	49.7
<b>Nacional</b>	<b>46.1</b>	<b>45.5</b>	<b>46.2</b>	<b>36.7</b>	<b>42.5</b>

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.5

Continúa

**Porcentaje de población en pobreza moderada**

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	34.4	34.4	32.7	24.4	28.4
Baja California	28.1	27.5	25.5	22.6	28.0
Baja California Sur	26.4	26.4	26.3	20.9	25.0
Campeche	36.7	34.2	32.5	29.7	32.4
Coahuila de Z.	24.9	24.7	26.4	17.6	20.1
Colima	32.2	30.4	30.9	24.1	29.5
Chiapas	40.2	42.5	44.4	40.3	40.6
Chihuahua	32.2	31.5	28.9	20.3	23.6
Ciudad de México	26.4	26.4	26.7	24.7	27.7
Durango	41.1	42.6	38.2	30.1	34.6

**Porcentaje de población en pobreza moderada**

Entidad federativa	2010	2012	2014	2015	2015 SVM
Guanajuato	40.1	37.6	41.0	28.3	32.0
Guerrero	35.7	38.0	40.8	41.4	42.8
Hidalgo	41.2	42.8	42.0	34.1	37.5
Jalisco	31.8	34.0	32.1	22.5	23.9
México	34.3	39.5	42.4	36.4	39.8
Michoacán de O.	41.2	39.9	45.2	35.8	40.2
Morelos	36.3	39.1	44.4	31.4	35.7
Nayarit	33.0	35.7	32.0	27.1	30.8
Nuevo León	19.2	20.8	19.0	16.6	19.3
Oaxaca	37.7	38.6	38.4	40.7	42.7
Puebla	44.5	46.9	48.4	42.9	45.5
Querétaro	34.0	31.8	30.3	21.7	25.2
Quintana Roo	28.2	30.4	28.9	25.0	28.4
San Luis Potosí	37.1	37.7	39.6	30.5	35.0
Sinaloa	31.2	31.8	34.1	25.7	29.9
Sonora	28.0	24.2	26.1	20.4	23.4
Tabasco	43.5	35.4	38.5	31.0	34.5
Tamaulipas	33.5	33.7	33.6	28.7	32.9
Tlaxcala	50.4	48.8	52.4	40.5	43.9
Veracruz de I. de la Ll.	38.8	38.4	40.9	40.3	43.0
Yucatán	36.6	39.0	35.2	28.6	32.3
Zacatecas	49.4	46.7	46.6	40.1	44.0
<b>Nacional</b>	<b>34.8</b>	<b>35.7</b>	<b>36.6</b>	<b>30.9</b>	<b>34.3</b>

Tabla 4.5

**Porcentaje de población en pobreza moderada**

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	34.4	34.4	32.7	24.4	28.4
Baja California	28.1	27.5	25.5	22.6	28.0
Baja California Sur	26.4	26.4	26.3	20.9	25.0
Campeche	36.7	34.2	32.5	29.7	32.4
Coahuila de Z.	24.9	24.7	26.4	17.6	20.1
Colima	32.2	30.4	30.9	24.1	29.5
Chiapas	40.2	42.5	44.4	40.3	40.6
Chihuahua	32.2	31.5	28.9	20.3	23.6
Ciudad de México	26.4	26.4	26.7	24.7	27.7
Durango	41.1	42.6	38.2	30.1	34.6
Guanajuato	40.1	37.6	41.0	28.3	32.0
Guerrero	35.7	38.0	40.8	41.4	42.8
Hidalgo	41.2	42.8	42.0	34.1	37.5
Jalisco	31.8	34.0	32.1	22.5	23.9
México	34.3	39.5	42.4	36.4	39.8
Michoacán de O.	41.2	39.9	45.2	35.8	40.2
Morelos	36.3	39.1	44.4	31.4	35.7
Nayarit	33.0	35.7	32.0	27.1	30.8
Nuevo León	19.2	20.8	19.0	16.6	19.3
Oaxaca	37.7	38.6	38.4	40.7	42.7
Puebla	44.5	46.9	48.4	42.9	45.5
Querétaro	34.0	31.8	30.3	21.7	25.2
Quintana Roo	28.2	30.4	28.9	25.0	28.4
San Luis Potosí	37.1	37.7	39.6	30.5	35.0
Sinaloa	31.2	31.8	34.1	25.7	29.9
Sonora	28.0	24.2	26.1	20.4	23.4
Tabasco	43.5	35.4	38.5	31.0	34.5
Tamaulipas	33.5	33.7	33.6	28.7	32.9
Tlaxcala	50.4	48.8	52.4	40.5	43.9
Veracruz de I. de la LI.	38.8	38.4	40.9	40.3	43.0
Yucatán	36.6	39.0	35.2	28.6	32.3
Zacatecas	49.4	46.7	46.6	40.1	44.0
<b>Nacional</b>	<b>34.8</b>	<b>35.7</b>	<b>36.6</b>	<b>30.9</b>	<b>34.3</b>

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.6

### Porcentaje de población en pobreza extrema

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	3.8	3.4	2.1	1.5	2.1
Baja California	3.4	2.7	3.1	1.7	2.2
Baja California Sur	4.6	3.7	3.9	1.7	2.7
Campeche	13.8	10.4	11.1	4.6	6.5
Coahuila de Z.	2.9	3.2	3.7	1.4	2.0
Colima	2.5	4.0	3.4	1.9	2.9
Chiapas	38.3	32.2	31.8	25.3	30.3
Chihuahua	6.6	3.8	5.4	0.9	1.4
Ciudad de México	2.2	2.5	1.7	0.9	1.4
Durango	10.5	7.5	5.3	2.2	3.7
Guanajuato	8.4	6.9	5.5	1.7	2.3
Guerrero	31.8	31.7	24.5	17.9	25.3
Hidalgo	13.5	10.0	12.3	7.3	10.0
Jalisco	5.3	5.8	3.2	2.4	3.3
México	8.6	5.8	7.2	4.3	6.0
Michoacán de O.	13.5	14.4	14.0	8.6	12.0
Morelos	6.9	6.3	7.9	5.4	7.8
Nayarit	8.3	11.9	8.5	4.3	6.8
Nuevo León	1.8	2.4	1.3	1	2.3
Oaxaca	29.2	23.3	28.3	19	25.5
Puebla	17.0	17.6	16.2	7.2	8.5
Querétaro	7.4	5.2	3.9	2.4	3.3
Quintana Roo	6.4	8.4	7.0	5	7.4
San Luis Potosí	15.3	12.8	9.5	6	8.6
Sinaloa	5.5	4.5	5.3	1.7	2.8
Sonora	5.1	5.0	3.3	2.1	3.1
Tabasco	13.6	14.3	11.0	5.7	8.2
Tamaulipas	5.5	4.7	4.3	3.3	4.6
Tlaxcala	9.9	9.1	6.5	2.8	4.0
Veracruz de I. de la Ll.	18.8	14.3	17.2	10.4	13.8
Yucatán	11.7	9.8	10.7	5.8	8.0
Zacatecas	10.8	7.5	5.7	4.6	5.7
<b>Nacional</b>	<b>11.3</b>	<b>9.8</b>	<b>9.5</b>	<b>5.8</b>	<b>8.2</b>

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.7

**Porcentaje de población vulnerable por carencias sociales**

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	26.2	24.5	25.3	33.1	30.1
Baja California	37.9	37.6	38.3	38.8	34.2
Baja California Sur	32.5	30.0	34.4	40.1	36.6
Campeche	24.9	28.6	32.1	39.8	37.1
Coahuila de Z.	25.6	24.4	24.2	30.9	28.9
Colima	33.8	31.7	33.2	41.0	36.4
Chiapas	13.0	17.2	15.3	24.1	20.6
Chihuahua	22.8	27.4	25.8	34.7	32.7
Ciudad de México	34.4	32.4	28.0	28.9	26.8
Durango	21.1	21.8	24.3	35.0	30.9
Guanajuato	28.9	32.6	27.7	43.7	40.8
Guerrero	23.0	21.7	26.2	32.5	25.1
Hidalgo	27.5	30.7	25.8	39.3	34.4
Jalisco	33.7	28.3	29.6	38.7	37.9
México	32.2	29.5	23.7	27.9	25.0
Michoacán de O.	28.6	30.7	25.0	38.6	32.2
Morelos	33.6	32.0	25.7	38.5	33.1
Nayarit	33.4	28.2	29.4	38.8	35.0
Nuevo León	31.6	29.1	28.0	30.2	28.3
Oaxaca	22.2	26.1	23.3	28.5	22.0
Puebla	21.6	22.0	19.8	32.8	29.6
Querétaro	31.7	32.6	33.1	40.4	36.9
Quintana Roo	36.1	30.4	34.2	39.8	35.6
San Luis Potosí	20.9	24.7	24.3	34.4	30.0
Sinaloa	33.3	33.5	30.4	40.4	36.9
Sonora	31.6	36.6	32.1	36.8	34.7
Tabasco	27.2	34.0	37.5	50.5	45.3
Tamaulipas	26.9	26.8	24.4	31.7	28.4
Tlaxcala	19.3	23.9	22.6	33.7	30.0
Veracruz de I. de la Ll.	23.6	30.6	24.8	33.3	28.0
Yucatán	26.0	27.0	27.7	36.9	32.6
Zacatecas	18.4	20.4	24.4	30.5	28.0
<b>Nacional</b>	<b>28.1</b>	<b>28.6</b>	<b>26.3</b>	<b>33.9</b>	<b>30.3</b>

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.8

**Porcentaje de población vulnerable por ingresos**

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	8.1	10.1	8.9	7.7	8.9
Baja California	6.3	8.6	6.6	5.9	7.6
Baja California Sur	4.5	7.9	5.6	4.6	6.0
Campeche	4.3	5.6	4.0	2.9	3.0
Coahuila de Z.	12.9	12.7	11.1	10.5	11.7
Colima	4.9	6.3	6.4	5	6.7
Chiapas	2.4	1.7	2.5	1.9	1.7
Chihuahua	13.0	10.7	12.0	8.1	8.7
Ciudad de México	5.4	6.6	8.0	6.9	8.9
Durango	8.8	11.1	10.5	7.6	9.0
Guanajuato	5.7	4.9	8.6	4.7	5.1
Guerrero	2.0	2.3	2.6	1.8	2.0
Hidalgo	4.2	3.0	5.1	3.2	4.2
Jalisco	6.2	8.1	7.9	5.4	5.4
México	5.6	7.8	9.3	8.6	9.2
Michoacán de O.	4.3	3.5	4.1	3	3.9
Morelos	5.8	4.6	6.1	3.9	4.5
Nayarit	4.3	5.6	6.5	5.5	5.8
Nuevo León	8.2	8.4	9.1	8	9.2
Oaxaca	1.3	1.7	2.1	1.6	1.9
Puebla	5.6	4.2	5.1	4.3	4.5
Querétaro	5.0	6.4	7.8	4.9	6.0
Quintana Roo	4.7	6.2	6.3	4.8	5.8
San Luis Potosí	7.2	6.6	7.6	4.9	6.0
Sinaloa	7.7	6.4	6.9	4.9	5.9
Sonora	6.8	4.7	7.3	5.7	6.4
Tabasco	4.1	3.0	2.3	1.6	1.9
Tamaulipas	9.4	8.8	11.2	6.9	7.8
Tlaxcala	7.4	6.0	7.7	6.4	7.7
Veracruz de I. de la Ll.	4.5	4.0	5.0	2.9	3.1
Yucatán	6.4	6.3	7.0	4.8	5.7
Zacatecas	6.9	6.4	7.3	5	5.0
<b>Nacional</b>	<b>5.9</b>	<b>6.2</b>	<b>7.1</b>	<b>5.4</b>	<b>6.1</b>

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.9

**Porcentaje de población no pobre y no vulnerable**

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	27.6	27.6	31.0	33.3	30.6
Baja California	24.2	23.5	26.5	31.1	28.0
Baja California Sur	31.9	32.0	29.8	32.7	29.8
Campeche	20.3	21.2	20.4	23.1	20.9
Coahuila de Z.	33.7	34.9	34.5	39.6	37.3
Colima	26.7	27.6	26.1	27.9	24.5
Chiapas	6.1	6.4	6.0	8.4	6.8
Chihuahua	25.4	26.6	27.8	36.1	33.6
Ciudad de México	31.7	32.1	35.7	38.7	35.2
Durango	18.6	17.0	21.8	25	21.9
Guanajuato	16.9	18.1	17.2	21.6	19.7
Guerrero	7.5	6.4	6.0	6.3	4.8
Hidalgo	13.7	13.5	14.8	16.1	13.9
Jalisco	23.0	23.8	27.1	31	29.6
México	19.3	17.4	17.4	22.8	20.0
Michoacán de O.	12.3	11.5	11.7	14	11.8
Morelos	17.3	17.9	15.8	20.8	18.9
Nayarit	20.9	18.6	23.5	24.3	21.6
Nuevo León	39.1	39.2	42.5	44.2	40.9
Oaxaca	9.5	10.3	7.9	10.1	7.9
Puebla	11.3	9.4	10.5	12.8	11.8
Querétaro	21.8	24.0	24.9	30.6	28.6
Quintana Roo	24.6	24.6	23.7	25.4	22.8
San Luis Potosí	19.5	18.3	19.0	24.2	20.4
Sinaloa	22.3	23.8	23.4	27.2	24.6
Sonora	28.4	29.6	31.3	35	32.4
Tabasco	11.5	13.4	10.6	11.2	10.0
Tamaulipas	24.8	26.0	26.5	29.3	26.4
Tlaxcala	13.0	12.2	10.8	16.5	14.4
Veracruz de I. de la LI.	14.3	12.8	12.2	13.1	12.2
Yucatán	19.2	17.9	19.5	23.9	21.5
Zacatecas	14.5	19.0	16.0	19.8	17.3
<b>Nacional</b>	<b>19.9</b>	<b>19.8</b>	<b>20.5</b>	<b>23.9</b>	<b>21.2</b>

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

## 5. Validación y evaluación de la metodología

Una de las ventajas que tiene la metodología desarrollada en el apartado 3 consiste en que puede ser fácilmente evaluada con otras bases de datos. El mecanismo de validación comprende la información de los MCS 2010 y 2012 como datos de entrenamiento; como datos de prueba se usaron los del MCS 2014.

**Durango**  
Decil pronosticado

	I	II	III	IV	V	VI	VII	VIII	IX	X
Decil original	I	86.8%	13.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	II	8.1%	72.5%	19.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	III	0.0%	5.7%	84.4%	9.9%	0.0%	0.0%	0.0%	0.0%	0.0%
	IV	0.0%	0.0%	19.8%	76.3%	4.0%	0.0%	0.0%	0.0%	0.0%
	V	0.0%	0.0%	0.6%	22.1%	71.2%	6.1%	0.0%	0.0%	0.0%
	VI	0.0%	0.0%	0.0%	0.0%	23.0%	71.7%	5.3%	0.0%	0.0%
	VII	0.0%	0.0%	0.0%	0.0%	0.0%	22.4%	75.1%	2.5%	0.0%
	VIII	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	23.7%	74.2%	2.2%
	IX	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	20.1%	77.0%
	X	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	17.1%

**Zacatecas**  
Decil pronosticado

	I	II	III	IV	V	VI	VII	VIII	IX	X
Decil original	I	91.9%	8.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	II	6.5%	86.2%	7.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	III	0.0%	11.1%	76.8%	12.2%	0.0%	0.0%	0.0%	0.0%	0.0%
	IV	0.0%	0.0%	17.0%	76.0%	7.0%	0.0%	0.0%	0.0%	0.0%
	V	0.0%	0.0%	0.0%	15.5%	72.3%	12.3%	0.0%	0.0%	0.0%
	VI	0.0%	0.0%	0.0%	0.0%	14.9%	81.8%	3.4%	0.0%	0.0%
	VII	0.0%	0.0%	0.0%	0.0%	0.0%	32.3%	60.2%	7.5%	0.0%
	VIII	0.0%	0.0%	0.0%	0.0%	0.0%	1.4%	25.7%	65.0%	7.9%
	IX	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	17.5%	80.0%
	X	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.4%

## Veracruz de Ignacio de la Llave

### Decil pronosticado

	I	II	III	IV	V	VI	VII	VIII	IX	X
Decil original	I	96.1%	3.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	II	10.9%	77.1%	11.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	III	0.0%	14.2%	70.9%	14.9%	0.0%	0.0%	0.0%	0.0%	0.0%
	IV	0.0%	0.0%	17.0%	63.5%	19.5%	0.0%	0.0%	0.0%	0.0%
	V	0.0%	0.0%	0.0%	10.1%	70.4%	19.6%	0.0%	0.0%	0.0%
	VI	0.0%	0.0%	0.0%	0.0%	8.5%	69.5%	22.0%	0.0%	0.0%
	VII	0.0%	0.0%	0.0%	0.0%	0.0%	9.3%	66.7%	24.0%	0.0%
	VIII	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.8%	80.0%	16.2%
	IX	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.1%	75.5%
	X	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.5%

## Oaxaca

### Decil pronosticado

	I	II	III	IV	V	VI	VII	VIII	IX	X
Decil original	I	95.4%	4.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	II	7.8%	83.3%	8.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	III	0.0%	8.5%	82.2%	9.4%	0.0%	0.0%	0.0%	0.0%	0.0%
	IV	0.0%	0.0%	11.7%	82.2%	6.1%	0.0%	0.0%	0.0%	0.0%
	V	0.0%	0.0%	0.8%	17.1%	75.2%	7.0%	0.0%	0.0%	0.0%
	VI	0.0%	0.0%	0.0%	0.0%	27.5%	69.6%	2.9%	0.0%	0.0%
	VII	0.0%	0.0%	0.0%	0.0%	0.9%	30.2%	66.4%	2.6%	0.0%
	VIII	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	22.4%	74.8%	2.8%
	IX	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	18.6%	73.2%
	X	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	15.9%

A continuación, se presentan los resultados de validación para cuatro entidades seleccionadas de forma aleatoria. La diagonal principal indica el porcentaje de hogares para quienes la *decilización* artificial y la original coincidieron; es decir, que la clasificación artificial no cambió el decil de pertenencia del hogar. Como se observa, el porcentaje de hogares para los cuales la clasificación resultó acertada es elevado.

## 6. Conclusión y comentarios finales

Como consecuencia de las mejoras realizadas en el levantamiento del MCS 2015 se observa un corrimiento del ingreso reportado por los hogares, rompiendo la continuidad con levantamientos del MCS-ENIGH anteriores; sin embargo, ello no implica que la información recabada por el INEGI sea errónea o sesgada.

Una manifestación del corrimiento del vector de ingreso se da al comparar resultados por deciles del MCS 2015 con deciles del MCS 2014: las tasas de crecimiento del 2014 al 2015 del ingreso corriente total de cada decil fueron muy superiores a las que se venían observado.

La metodología que aquí se propone para recomponer la continuidad del MCS 2015 con los años anteriores consiste en la modificación de los factores de expansión, dejando inalterada la declaración de ingresos por parte de los hogares.

La forma de lograr continuidad consiste en modificar la representatividad que tiene un hogar de acuerdo con su asignación a un decil pronosticado o artificial, creado para ese propósito a partir de un conjunto de variables monetarias y no-monetarias. La asignación de un hogar a un decil pronosticado se realiza a través de una técnica de *Machine Learning*, conocida como máquina de soporte vectorial; esta técnica consiste en entrenar una máquina con datos de MCS anteriores al 2015 con el fin de pronosticar el decil de pertenencia de un hogar en el 2015. Para hacer comparaciones válidas con los resultados de los MCS anteriores, el supuesto en el que se basa la propuesta es que los hogares que pertenecen a determinado decil, considerando sus nuevos factores de expansión, representan a aquellos hogares que bajo circunstancias similares a las históricas hubiesen pertenecido a ese decil.

Una ventaja de este método es que utiliza información histórica para hacer pronósticos en el 2015 sin modificar la información reportada por los hogares, además de que trata de minimizar las clasificaciones erróneas y sus resultados son fácilmente validables y comparables. Sin embargo, debe considerarse que al ser una técnica de optimización, los resultados dependerán de las covariables incluidas y de los parámetros de ajuste especificados dentro del paquete computacional, aunados a las posibles consecuencias en el diseño muestral a causa de la modificación de los factores de expansión.

En términos de resultados, la variación porcentual entre el 2014 y 2015 del ICT a nivel nacional pasó de 15.6 a 4.5% con la presente propuesta de ajuste. Respecto a la variación porcentual en el decil I, ésta pasó de 38 a 11.2%; mientras que en el decil X fue de 11.1 a 4.1 por ciento. En términos de medición de pobreza, el porcentaje de población en pobreza según el MCS 2015 fue de 36.7, en tanto que con el ajuste aumentó a 42.5 por ciento. La pobreza extrema creció de 5.8 a 8.2%, mientras que la moderada fue de 30.9 a 34.3 por ciento.

### Fuentes

- Cortes, Corina y Vladimir Vapnik. "Support vector networks", en: *Machine Learning*. Boston, United States of America, Kluwer Academic Publishers, 1995.
- Hastie, Trevor; Robert Tibshirani y Jerome Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. New York, United States of America, Springer, 2009.
- Schölkopf, Bernhard y Alexander J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, England, MA, MIT Press, 2002.
- Witten, Ian y Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA, United States of America, Morgan Kaufmann, 2011.

## Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
Población	folioviv	Identificador de la vivienda compuesto por dos dígitos con la clave de la entidad federativa, uno con el ámbito (urbano, código diferente a 6; rural, código 6), cuatro dígitos del número consecutivo de la upm, un dígito de la decena de levantamiento y dos dígitos con un número consecutivo para la vivienda seleccionada.	Ninguno.
	foliohog	El código 1 identifica al hogar principal y del 2 al 5, los hogares adicionales.	Ninguno.
	numren	Número consecutivo en el registro de personas integrantes del hogar.	Ninguno.
	edad	Años transcurridos entre la fecha de nacimiento de la persona y la fecha de la entrevista.	Promedio de edad de los integrantes del hogar.
	hablaind	Personas de 3 años o más que hablan alguna lengua indígena o dialecto.	Proporción de integrantes del hogar que habla alguna lengua indígena.
	alfabetism	Situación que distingue a las personas de 3 años o más según sepan leer y escribir un recado.	Proporción de integrantes del hogar mayores de 12 años que no sabe escribir/ leer un recado.
	nivel	Nivel al que asiste a la escuela el integrante del hogar de 3 años o más dentro del Sistema Educativo Nacional.	Proporción de integrantes del hogar que asiste a cada uno de los niveles escolares.
	tipoesc	Tipo de escuela a la que asiste el integrante del hogar de 3 años o más dentro del Sistema Educativo Nacional, según sea pública, privada o de otro tipo.	Proporción de integrantes del hogar que asiste a cada uno de los tipos de escuela.
	tiene_b	Apoyo en dinero o especie que se proporciona a los integrantes del hogar que asisten a la escuela.	Proporción de integrantes del hogar que tiene beca.
	tiene_c	Aportación que recibe el integrante del hogar que asiste a la escuela para el pago de colegiatura, hospedaje, alimentación, materiales, etcétera. Quien lo recibe está comprometido a pagarlo en parte o en su totalidad al concluir sus estudios.	Proporción de integrantes del hogar que tiene crédito educativo.
nivelaprob	Año máximo aprobado en la escuela por el integrante del hogar de 3 años o más dentro del Sistema Educativo Nacional.	Proporción de integrantes del hogar para cada nivel de educación.	
trabajo_mp	Es la situación que distingue a la población de 12 años o más de acuerdo con el desempeño o no de una actividad económica en el periodo de referencia, que permite clasificarla como parte de la Población Económicamente Activa (PEA) o como parte de la Población No Económicamente Activa (PNEA).	Proporción de integrantes del hogar que trabaja.	
act_pnea1	Actividad o situación de las personas no económicamente activas.	a) Proporción de integrantes que se dedicó a buscar trabajo.	
act_pnea1	Actividad o situación de las personas no económicamente activas.	b) Proporción de integrantes que se dedicó a quehaceres del hogar.	
Vivienda	folioviv	Identificador de la vivienda compuesto por dos dígitos con la clave de la entidad federativa, uno con el ámbito (urbano, código diferente a 6; rural, código 6), cuatro dígitos del número consecutivo de la upm, un dígito de la decena de levantamiento y dos dígitos con un número consecutivo para la vivienda seleccionada.	Ninguno.

## Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
	tipo_viv	Clase de vivienda particular.	Generación de una variable dicotómica por cada una de las categorías.
	mat_pared	Material predominante en las paredes de la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	mat_techos	Material predominante en el techo de la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	cocina	Cuarto de la vivienda dispuesto para preparar o calentar alimentos, aunque también tenga otros usos.	Generación de una variable dicotómica por cada una de las categorías.
	cocina_dor	Cocina utilizada de manera habitual también para dormir.	Generación de una variable dicotómica por cada una de las categorías.
	cuart_dorm	Número de cuartos de la vivienda que son usados habitualmente para dormir, aunque también tengan otros usos.	Ninguno.
	num_cuarto	Número total de cuartos que tiene la vivienda, independientemente de su uso.	Ninguno.
	excusado	Instalación sanitaria o excusado en la vivienda para el desalojo de los desechos humanos.	Generación de una variable dicotómica por cada una de las categorías.
	uso_compar	Uso compartido de la instalación sanitaria con los habitantes de otra vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	sanit_agua	Funcionamiento de la instalación sanitaria con o sin conexión de agua.	Generación de una variable dicotómica por cada una de las categorías.
	drenaje	Disponibilidad de un sistema de drenaje para desalojar de la vivienda los desechos humanos y las aguas utilizadas.	Generación de una variable dicotómica por cada una de las categorías.
	disp_elect	Fuente de donde se obtiene la energía eléctrica en la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	combustible	Combustible más utilizado para preparar o calentar alimentos en la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	eli_basura	Formas de eliminar la basura de la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	tenencia	Dominio o posesión legal de la vivienda por sus ocupantes.	Generación de una variable dicotómica por cada una de las categorías.
	lavadero	Lavadero.	Generación de una variable dicotómica por cada una de las categorías.
	fregadero	Fregadero o tarja.	Generación de una variable dicotómica por cada una de las categorías.
	regadera	Regadera.	Generación de una variable dicotómica por cada una de las categorías.

## Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
	tinaco_azo	Tinaco en la azotea.	Generación de una variable dicotómica por cada una de las categorías.
	cisterna	Cisterna o aljibe.	Generación de una variable dicotómica por cada una de las categorías.
	pileta	Pileta, tanque o depósito de agua.	Generación de una variable dicotómica por cada una de las categorías.
	calent_gas	Calentador de gas u otros.	Generación de una variable dicotómica por cada una de las categorías. Se renombró como calentador.
	medidor_luz	Medidor de luz.	Generación de una variable dicotómica por cada una de las categorías.
	bomba_agua	Bomba de agua.	Generación de una variable dicotómica por cada una de las categorías.
	tanque_gas	Tanque de gas estacionario.	Generación de una variable dicotómica por cada una de las categorías.
	aire_acond	Aire acondicionado.	Generación de una variable dicotómica por cada una de las categorías.
	calefacc	Calefacción.	Generación de una variable dicotómica por cada una de las categorías.
	ubica_geo	Contiene la ubicación geográfica de la vivienda. Los dos primeros dígitos representan la clave de la entidad y los tres últimos dígitos la clave del municipio. Éstos corresponden al <i>Catálogo único de claves de áreas geostatísticas estatales, municipales y localidades</i> , que está disponible en el sitio del INEGI.	Se extrajo la entidad y se formó una nueva variable llamada edo.
Hogar	folioviv	Identificador de la vivienda compuesto por dos dígitos con la clave de la entidad federativa, uno con el ámbito (urbano, código diferente a 6; rural, código 6), cuatro dígitos del número consecutivo de la upm, un dígito de la decena de levantamiento y dos dígitos con un número consecutivo para la vivienda seleccionada.	Ninguno.
	foliohog	El código 1 identifica al hogar principal y del 2 al 5, los hogares adicionales.	Ninguno.
	telefono	Servicio de línea telefónica en el hogar ya sea con instalación alámbrica o satelital (antena).	Generación de una variable dicotómica por cada una de las categorías.
	celular	Hogares que cuentan con el servicio de teléfono móvil aunque sea un solo integrante del hogar.	Generación de una variable dicotómica por cada una de las categorías.
	tv_paga	Hogares que cuentan con los servicios de televisión por pago, considerando que el servicio está en funcionamiento.	Generación de una variable dicotómica por cada una de las categorías.
	conex_inte	Hogares que cuentan con el servicio de línea telefónica, tarjeta prepagada o telecable.	Generación de una variable dicotómica por cada una de las categorías.
	num_auto	Número de automóviles con los que cuenta el hogar.	Ninguno.

## Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
	anio_auto	Año de adquisición del último automóvil.	Se generó una nueva variable con categorías de acuerdo con el rango de antigüedad del auto: 1. El vehículo tiene a lo más un año de antigüedad. 2. El vehículo tiene entre 2 y 5 años de antigüedad. 3. El vehículo tiene entre 6 y 10 años de antigüedad. 4. El vehículo tiene más de 10 años de antigüedad. 5. Tienen vehículo pero no respondieron sobre la antigüedad. 6. No tienen vehículo.
	num_van	Número de camionetas cerradas o con cabina ( <i>van</i> o <i>minivan</i> ) con las que cuenta el hogar.	Ninguno.
	anio_van	Año de adquisición de la última camioneta cerrada o con cabina.	Se generó una nueva variable con categorías de acuerdo con el rango de antigüedad del auto: 1. El vehículo tiene a lo más un año de antigüedad. 2. El vehículo tiene entre 2 y 5 años de antigüedad. 3. El vehículo tiene entre 6 y 10 años de antigüedad. 4. El vehículo tiene más de 10 años de antigüedad. 5. Tienen vehículo pero no respondieron sobre la antigüedad. 6. No tienen vehículo.
	num_pickup	Número de camionetas con caja con las que cuenta el hogar.	Ninguno.
	anio_pickup	Año de adquisición de la última camioneta de caja.	Se generó una nueva variable con categorías de acuerdo con el rango de antigüedad del auto: 1. El vehículo tiene a lo más un año de antigüedad. 2. El vehículo tiene entre 2 y 5 años de antigüedad. 3. El vehículo tiene entre 6 y 10 años de antigüedad. 4. El vehículo tiene más de 10 años de antigüedad. 5. Tienen vehículo pero no respondieron sobre la antigüedad. 6. No tienen vehículo.
	num_ester	Número de estéreos, modulares o minicomponentes con los que cuenta el hogar.	Ninguno
	num_grab	Número de grabadoras con las que cuenta el hogar.	Ninguno.

## Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
	num_radio	Número de radios con las que cuenta el hogar.	Ninguno.
	num_tva	Número de televisores analógicos con los que cuenta el hogar.	Su sumó con la variable num_tvd y se formó una nueva variable llamada num_tv.
	num_tvd	Número de televisores digitales con los que cuenta el hogar.	Su sumó con la variable num_tva y se formó una nueva variable llamada num_tv.
	num_dvd	Número de DVD, <i>Blue-ray</i> (reproductores de discos de video) con los que cuenta el hogar.	Ninguno.
	num_video	Número de videocaseteras con las que cuenta el hogar.	Ninguno.
	num_licua	Número de licuadoras con las que cuenta el hogar.	Ninguno.
	num_tosta	Número de tostadores con los que cuenta el hogar.	Ninguno.
	num_micro	Número de hornos de microondas con los que cuenta el hogar.	Ninguno.
	num_refri	Número de refrigeradores con los que cuenta el hogar.	Ninguno.
	num_estuf	Número de estufas de gas o eléctricas con las que cuenta el hogar.	Ninguno.
	num_lavad	Número de lavadoras con las que cuenta el hogar.	Ninguno.
	num_planc	Número de planchas eléctricas con las que cuenta el hogar.	Ninguno.
	num_maqui	Número de máquinas de coser con las que cuenta el hogar.	Ninguno.
	num_venti	Número de ventiladores con los que cuenta el hogar.	Ninguno.
	num_aspir	Número de aspiradoras con las que cuenta el hogar.	Ninguno.
	num_compu	Número de computadoras con las que cuenta el hogar.	Ninguno.
	num_impre	Número de impresoras con las que cuenta el hogar.	Ninguno.
	num_juego	Número de videojuegos ( <i>Wii, Playstation, Xbox</i> , otros) con los que cuenta el hogar.	Ninguno.
Concentrado	folioviv	Identificador de la vivienda compuesto por dos dígitos con la clave de la entidad federativa, uno con el ámbito (urbano, código diferente a 6; rural, código 6), cuatro dígitos del número consecutivo de la upm, un dígito de la decena de levantamiento y dos dígitos con un número consecutivo para la vivienda seleccionada.	Ninguno.
	foliohog	El código 1 identifica al hogar principal y del 2 al 5, los hogares adicionales.	Ninguno.
	ing_cor	Suma de los ingresos por trabajo, los provenientes de rentas, de transferencias, de estimación del alquiler y de otros ingresos.	Ninguno.

Fuente: INEGI. *Módulo de Condiciones Socioeconómicas 2015*. Descripción de la base de datos.