



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

MODELO ESTADÍSTICO 2016 PARA LA CONTINUIDAD DEL MCS-ENIGH

NOTA TÉCNICA 1

28 DE AGOSTO DE 2017



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

El modelo estadístico que aquí se presenta es el resultado de un proyecto de investigación coordinado por José Alejandro Ruiz y Ana Miriam Romo, investigador y subdirectora de la Dirección General Adjunta de Investigación, respectivamente. En el proyecto también participaron Lorenzo Cecilio, subdirector de investigación “A”, y Benito Durán, subdirector de investigación en indicadores sociales, demográficos y económicos.¹

¹ El equipo agradece los comentarios, sugerencias y ayuda proporcionada durante el proceso.

Antecedentes

En un proceso continuo de mejora, el INEGI identificó áreas de oportunidad alrededor de la capacitación, supervisión y control operativo del MCS-ENIGH². A partir de ello, se llevaron a cabo acciones de mejora operativas en el levantamiento de la ENIGH 2016, en un esfuerzo por alcanzar un mayor apego al diseño de la encuesta.

Con estas medidas implementadas, se inicia una nueva serie histórica para las encuestas de ingresos y gastos de los hogares. Sin embargo, y partiendo del reconocimiento de la necesidad de preservar la continuidad de sus ejercicios estadísticos, el INEGI presenta el *Modelo Estadístico 2016 para la Continuidad del MCS-ENIGH*.

El objetivo del Modelo Estadístico es proveer a los usuarios de la información con los insumos necesarios para la medición de la pobreza multidimensional, de tal manera que se mantenga la consistencia con los resultados de la serie bienal 2008-2014, tanto a nivel nacional como por entidad federativa.

Modelo Estadístico

Un gran reto para cualquier ejercicio que se proponga mantener la continuidad histórica del MCS-ENIGH, es la selección de elementos de referencia para ajustar el ingreso originalmente reportado. Podemos considerar algunos con base en relaciones propias del MCS-ENIGH, o bien, con base en fuentes externas que no hayan sido afectadas por las mejoras realizadas en la ENIGH 2016. La metodología que aquí se emplea forma parte del segundo tipo, tomando como fuente externa la Encuesta Nacional de Ocupación y Empleo (ENOE).

A pesar de la diversidad de fuentes de ingreso de los hogares mexicanos, la mayor parte proviene de su ingreso laboral, es decir, de su trabajo como asalariado y/o independiente, que en conjunto han representado cerca del 70 % del ingreso corriente total desde 2010. Este elevado porcentaje, junto con la existencia de un levantamiento regular sobre condiciones laborales como lo es la ENOE, crean una ventana de oportunidad para utilizar el ingreso laboral que reporta la ENOE como ancla para ajustar los ingresos laborales³.

Entre las ventajas que tiene usar la ENOE como referencia podemos enumerar las siguientes:

- 1) Considera el mismo marco muestral que el MCS-ENIGH.
- 2) La ENOE es la encuesta continua en hogares más grande de que se dispone en México, con 120,000 hogares al trimestre.
- 3) La ENOE no fue afectada por los cambios endógenos (capacitación, control del operativo, etc.) que hacen diferente a la ENIGH 2016 de los levantamientos anteriores. En este sentido, los cambios en el ingreso de la ENOE reflejan cambios que son independientes de las mejoras hechas en la ENIGH 2016.

² MCS-ENIGH se refiere al Módulo de Condiciones Socioeconómicas de la ENIGH

³ Salvo en el primer trimestre del año, el ingreso laboral de la ENOE hace referencia al correspondiente originado por el trabajo principal.

- 4) Al igual que el MCS-ENIGH, y si bien no es su propósito principal, la ENOE capta los ingresos provenientes tanto de actividades formales como de informales.
- 5) El diseño de la ENOE en paneles rotatorios estabiliza la muestra y robustece las comparaciones en el tiempo.
- 6) La ENOE es sensible a las diferencias del comportamiento de los ingresos entre los estados, además de que es sensible a los efectos del entorno económico y de políticas públicas a nivel nacional.
- 7) CONEVAL utiliza la ENOE para la elaboración de sus indicadores coyunturales de la pobreza laboral en México.
- 8) ENOE es menos volátil que el MCS-ENIGH respecto del trabajo independiente (patrones y cuentas propias).
- 9) La ENOE es consistente a nivel entidad federativa.

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

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

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

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

El modelo (que más adelante se describe), insume valores objetivo para un estadístico deseado, que en nuestro caso es la mediana del ingreso corriente total (ICT). Tales estadísticos servirán como restricciones al momento de realizar el ajuste. Para generar los valores objetivo se siguen estos pasos: en cada entidad federativa por separado, se toman las medianas del ingreso reportado dentro del segundo y tercer trimestre de la ENOE para los años 2014 y 2016. Con ellas se obtienen los promedios entre los dos trimestres del mismo año. Posteriormente se calculan las variaciones porcentuales del promedio de las medianas de 2016 con respecto al de 2014. De esta manera tenemos 32 tasas de crecimiento de las medianas entre 2014 y 2016. Enseguida, y para cada entidad federativa, se calculan las medianas del ICT del MCS-ENIGH 2014; a estas medianas les aplicamos la tasa de crecimiento obtenida con los datos de la ENOE, para finalmente obtener las *medianas objetivo* de cada entidad en 2016.

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

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

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

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

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

Para cada entidad $i = 1, 2, \dots, 32$:

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

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

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

Sea $Y_o^i = (y_{(1)}^i, y_{(2)}^i, \dots, y_{(n)}^i)$ el vector ordenado de ingresos para la entidad i y sus correspondientes pesos muestrales (factores de expansión) denotados por $w_o^i = (w_{(1)}^i, w_{(2)}^i, \dots, w_{(n)}^i)$. Y sea el vector de *mediana objetivo* $\mathbf{Me} = (Me_1, Me_2, \dots, Me_{32})$. Para cada entidad $i = 1, 2, \dots, 32$ se resuelve el siguiente problema de optimización restringida

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

Sujeto a:

a) Restricciones de igualdad

$$1. F^{-1}(F(y_{(k)}^i | \hat{\theta}_{SR}^i) | \theta^i) = Me_i$$

donde:

$y_{(k)}^i$ es el ingreso asociado al subíndice k tal que $0.5 \leq \frac{W_{(k)}}{W_{(n)}}$ y $0.5 \leq 1 - \frac{W_{(k-1)}}{W_{(n)}}$

con $w_{(j)} = \sum_{h=1}^j w_{(h)}^i$

Me_i es la mediana de la entidad federativa i .

$\hat{\theta}_{SR}^i$ es el estimador de los parámetros de la función teórica de densidad sin restricciones para la entidad i .

$F(\cdot)$ es la función de probabilidad acumulada de GB2.

$F^{-1}(\cdot)$ es la función cuantil de GB2 y

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

donde:

$\max^i = y_{(n)}^i$, es el valor máximo del ingreso de ICT en la entidad i .

$p^{\max}(\hat{\theta}_{SR}^i) = F(y_{(n)}^i | \hat{\theta}_{SR}^i)$ es la probabilidad acumulada (bajo la función teórica sin restricciones) correspondiente al valor max en la entidad i .

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

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

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

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

De esta manera, obtenemos para cada hogar en toda la muestra, un valor del ICT imputado que está determinado en función de los resultados del ajuste por entidad, $ICT_{fase1_h} = \hat{y}_h \forall h = 1, 2, \dots, n$

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

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

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

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

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

$$ITP_{ajustado} = (\hat{y}_{ITP_1}, \hat{y}_{ITP_2}, \dots, \hat{y}_{ITP_n})$$

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

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

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

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

Es importante mencionar que dado los procesos computacionales intensivos de optimización que requiere la metodología, los resultados pueden variar ligeramente dependiendo de las especificaciones del equipo de cómputo donde se realicen los cálculos⁴.

⁴ Los microdatos correspondientes al *Modelo Estadístico 2016 para la Continuidad del MCS-ENIGH* fueron resultado del procesamiento del algoritmo en una computadora marca Lanix con procesador AMD FX-8370 de 8-núcleos y 4.00 GHz, sistema operativo de 64-bits. La versión de R fue la 3.4.0, a través de su IDE RStudio versión 1.0143.

Anexo

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

Código	Descripción	Código	Descripción
	Ingresos monetarios del trabajo principal para subordinados		Ingresos por negocios del hogar, trabajo principal
P001	Sueldos, salarios o jornal	P011	Sueldos o salarios
P002	Destajo	P012	Ganancias/utilidades
P003	Comisiones y propinas	P013	Otros ingresos
P004	Horas extras		Ingresos por negocio propio, trabajo principal
P005	Incentivos, gratificaciones o premios	P068	Por negocios con tipo de actividad industrial
P006	Bono, percepción adicional o sobresueldo	P069	Por negocios con tipo de actividad comercial
P007	Primas vacacionales y otras prestaciones en dinero	P070	Por negocios prestadores de servicios
		P071	Por negocios con actividades agrícolas
		P072	Por negocios con actividades de cría y explotación de animales
		P073	Por negocios con actividades de recolección, reforestación y tala de árboles
		P074	Por negocios con actividades de pesca, caza y captura de animales