

Instituto Nacional de Estadística y Geografía

Módulo de Movilidad Social Intergeneracional 2016

MMSI

Diseño estadístico



**INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA**

Presentación

El **Módulo de Movilidad Social Intergeneracional (MMSI) 2016** es un ejercicio estadístico destinado a enriquecer la oferta de información vinculada al Subsistema Nacional de Información Demográfica y Social. El propósito de este ejercicio estadístico es obtener información útil para el análisis de los niveles de movilidad social en el país, en particular desde un contexto generacional, mediante la captación de las características sociodemográficas y socioeconómicas retrospectivas de la población de 25 a 64 años, y su contraste con las condiciones actuales. Esto permite obtener una medida de la influencia que tiene el origen socioeconómico en los logros y posición actual dentro de la estructura económica de la sociedad.

Como parte del soporte metodológico del MMSI, el **Instituto Nacional de Estadística y Geografía (INEGI)** presenta el diseño estadístico del Módulo de Movilidad Social Intergeneracional, en el cual se incluye: el marco de muestreo, el tamaño y la distribución de la muestra, el cálculo de las probabilidades de inclusión de las unidades de observación, los factores de expansión y los principales estimadores para la explotación de información. De esta manera, el **INEGI** da a conocer la metodología empleada y contribuye a la transparencia del proceso de generación de información estadística.

Índice

1. Diseño estadístico	1
1.1 Periodo de referencia	1
1.2 Marco de la muestra	1
1.3 Unidades Primarias de Muestreo (UPM)	1
1.4 Estratificación	1
1.5 Tamaño de la muestra	2
1.6 Afijación de la muestra	2
1.7 Selección de la muestra	3
1.7.1 En urbano alto	3
1.7.2 En complemento urbano	4
1.7.3 En rural	5
1.8 Ajuste de los factores de expansión	5
1.8.1 Ajuste por No respuesta	6
1.8.2 Ajuste por proyección	7
1.9 Estimadores de los indicadores	7
1.10 Estimación de muestreo	8

1. Diseño estadístico

1.1 Periodo de referencia

El Módulo de Movilidad Social Intergeneracional se aplicó basado en el diseño estadístico de la Encuesta Nacional de los Hogares llevándose a cabo en el tercer y cuarto trimestre del año 2016.

1.2 Marco de la muestra

El diseño muestral del MMSI-2016 es el empleado en la Encuesta Nacional de los Hogares, ya que el MMSI se levanta en las mismas viviendas del tercer y cuarto trimestre de 2016. Esta encuesta se caracteriza por ser probabilística, por tanto, los resultados obtenidos se generalizan a toda la población; a su vez es bietápico, estratificado y por conglomerados, donde la unidad última de selección es la persona.

1.3 Unidades Primarias de Muestreo (UPM)

Las Unidades Primarias de Muestreo están constituidas por agrupaciones de viviendas con características diferenciadas dependiendo del ámbito al que pertenecen, como se especifica a continuación:

a. En urbano alto

El tamaño mínimo de una UPM es de 80 viviendas habitadas y el máximo es de 160. Pueden estar formadas por:

- Una manzana.
- La unión de dos o más manzanas contiguas del mismo AGEB.¹

b. En complementario urbano

El tamaño mínimo de una UPM es de 160 viviendas habitadas y el máximo es de 300. Pueden estar formadas por:

- Una manzana.
- La unión de dos o más manzanas contiguas del mismo AGEB.

c. En rural

El tamaño mínimo de una UPM es de 160 viviendas habitadas y el máximo es 300. Pueden estar formadas por:

- Una localidad.
- La unión de dos o más localidades cercanas del mismo municipio.

1.4 Estratificación

La división política del país y la agrupación de localidades diferenciadas por su tamaño, forma de manera natural una estratificación geográfica.

De manera paralela, se formaron cuatro estratos sociodemográficos en los que se agrupan todas las UPM del país, esta estratificación considera las características sociodemográficas de los habitantes de las viviendas, así como las características físicas y el equipamiento de las mismas, construidos con información del Censo de Población

¹ Área Geoestadística Básica.

y Vivienda 2010, para lo cual se emplearon métodos estadísticos multivariados. De esta forma, cada UPM fue clasificada en un único estrato geográfico y uno sociodemográfico.

1.5 Tamaño de la muestra

El tamaño de muestra fue de 31 935 viviendas, el cual se determinó con base en las viviendas con respuesta en la ENH-2016 durante el tercer y cuarto trimestre.

1.6 Afijación de la muestra

La afijación de la muestra se realizó de manera equitativa a nivel entidad y dentro de cada estrato de manera proporcional, para lo cual se emplea la siguiente expresión:

$$n_{eh} = \frac{N_{eh}}{N_e} n_e$$

Donde:

n_{eh} = número de viviendas en muestra en el h-ésimo estrato, en la e-ésima entidad.

n_e = número total de viviendas en muestra en la e-ésima entidad.

N_{eh} = número total de viviendas en el h-ésimo estrato, en la e-ésima entidad.

N_e = número total de viviendas en la e-ésima entidad.

Asimismo, se distribuyó la muestra de manera equitativa por trimestre y de manera proporcional al tamaño por dominio de estudio, para lo cual se empleó la siguiente expresión:

$$n_{td} = \frac{N_{td}}{N_t} n_t$$

Donde:

n_{td} = número de viviendas en muestra en el d-ésimo dominio, en el t-ésimo trimestre.

n_t = número total de viviendas en muestra en el t-ésimo trimestre.

N_{td} = número total de viviendas en el d-ésimo dominio, en el t-ésimo trimestre.

N_t = número total de viviendas en el t-ésimo trimestre.

En el siguiente cuadro, se presenta la distribución de la muestra por trimestre y dominio de estudio.

Distribución de la muestra en viviendas por trimestre según dominio para el MMSI-2016				
Trimestre	Dominio			Total
	Urbano Alto	Complemento Urbano	Rural	
Total	16 375	8 080	7 480	31 935
Julio-septiembre	8 175	3 960	3 700	15 835
Octubre-diciembre	8 200	4 120	3 780	16 100

1.7 Selección de la muestra

La selección de la muestra para el MMSI-2016 se realizó de manera independiente por entidad y estrato, el procedimiento de selección varió de acuerdo con el dominio.

1.7.1 En urbano alto

En el ámbito urbano alto la selección de la muestra se realizó de forma independiente por cada entidad y estrato de la siguiente manera:

1. Se seleccionaron n_{eh} UPM, con probabilidad proporcional al total de viviendas del estrato.
2. En cada UPM seleccionada, se seleccionaron cinco viviendas con igual probabilidad.
3. En cada vivienda seleccionada se seleccionó una persona de 25 a 64 años de edad con igual probabilidad.

Por lo tanto, la probabilidad de seleccionar una vivienda de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad es:

$$P\{V_{ehi}\} = \frac{n_{eh} m_{ehi}}{m_{eh}} \frac{5}{m_{ehi}^*} = \frac{5n_{eh} m_{ehi}}{m_{eh} m_{ehi}^*}$$

Su factor de expansión² está dado por:

$$F_{ehi} = \frac{m_{eh} m_{ehi}^*}{5n_{eh} m_{ehi}}$$

Por lo tanto, la probabilidad de seleccionar una persona de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad es:

$$P\{V_{ehij}\} = P\{V_{ehi}\} \frac{1}{Q_{ehij}} = \frac{5n_{eh} m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehij}}$$

² El factor de expansión se define como el inverso de la probabilidad de selección.

Su factor de expansión es:

$$F_{ehij} = F_{ehi} * Q_{ehij} = \frac{m_{eh} m_{ehi}^* Q_{ehij}}{5 n_{eh} m_{ehi}}$$

Donde:

n_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad, para la muestra maestra.

m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según el Censo de Población y Vivienda 2010.

m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.

m_{ehi}^* = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según listado de viviendas actualizado.

Q_{ehij} = número de personas de 25 a 64 años de edad en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

1.7.2 En complemento urbano

En el ámbito completo urbano la selección de la muestra se realizó de la siguiente manera:

1. Se seleccionaron n_{eh} UPM con probabilidad proporcional al total de viviendas del estrato.
2. De estas UPM seleccionadas, se seleccionaron 20 viviendas con igual probabilidad.
3. En cada vivienda seleccionada se seleccionó una persona de 25 a 64 años de edad con igual probabilidad.

Por lo tanto, la probabilidad de seleccionar una vivienda de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es:

$$P\{V_{ehi}\} = \frac{n_{eh} m_{ehi}}{m_{eh}} \frac{20}{m_{ehi}^*} = \frac{20 n_{eh} m_{ehi}}{m_{eh} m_{ehi}^*}$$

Su factor de expansión está dado por:

$$F_{ehi} = \frac{m_{eh} m_{ehi}^*}{20 n_{eh} m_{ehi}}$$

Por lo tanto, la probabilidad de seleccionar una persona de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es:

$$P\{V_{ehij}\} = P\{V_{ehi}\} \frac{1}{Q_{ehij}} = \frac{20 n_{eh} m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehij}}$$

Su factor de expansión es:

$$F_{ehij} = F_{ehi} * Q_{ehij} = \frac{m_{eh} m_{ehi}^* Q_{ehij}}{20 n_{eh} m_{ehi}}$$

Donde:

- n_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad para la muestra maestra.
- m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según el Censo de Población y Vivienda 2010.
- m_{ehi}^* = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según listado de viviendas actualizado.
- Q_{ehij} = número de personas de 25 a 64 años de edad en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

1.7.3 En rural

En el ámbito rural la selección de la muestra se realizó mediante el siguiente procedimiento:

1. Se seleccionaron n_{eh} UPM con probabilidad proporcional al total de viviendas del estrato.
2. En cada UPM seleccionada, se seleccionaron cuatro segmentos de cinco viviendas aproximadamente, con igual probabilidad.
3. En cada vivienda seleccionada se seleccionó una persona de 25 a 64 años de edad con igual probabilidad.

Por lo tanto, la probabilidad de seleccionar una vivienda de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es:

$$P\{V_{ehi}\} = \frac{n_{eh} m_{ehi}}{m_{eh}} \frac{4 \cdot 5}{m_{ehi}^*} = \frac{20 n_{eh} m_{ehi}}{m_{eh} m_{ehi}^*}$$

Su factor de expansión está dado por:

$$F_{ehi} = \frac{m_{eh} m_{ehi}^*}{20 n_{eh} m_{ehi}}$$

Por lo tanto, la probabilidad de seleccionar una persona de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es:

$$P\{V_{ehij}\} = P(V_{ehi}) \frac{1}{Q_{ehij}} = \frac{20n_{eh} m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehi}}$$

Su factor de expansión es:

$$F_{ehij} = F_{ehi} * Q_{ehij} = \frac{m_{eh} m_{ehi}^* Q_{ehij}}{20n_{eh} m_{ehi}}$$

Donde:

- n_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad para la muestra maestra.
- m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según el Censo de Población y Vivienda 2010.
- m_{ehi}^* = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según listado de viviendas actualizado.
- Q_{ehij} = número de personas de 25 a 64 años de edad en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

1.8 Ajuste de los factores de expansión

Los factores de expansión elaborados conforme al procedimiento antes descrito, se ajustaron con base en los siguientes conceptos:

1.8.1 Ajuste por No respuesta

El ajuste por No respuesta atribuida al informante se realizó a nivel entidad, en cada uno de los dominios, mediante la siguiente expresión:

- F_D'' = factor de expansión corregido por proyección en el dominio D.
- F_D' = factor de expansión corregido por No respuesta en el dominio D.
- $PROy_D$ = población en el dominio D, según proyección.
- $PEXP_D$ = población total a la que expande el módulo en el dominio D.
- D = dominio de interés para el módulo.

$$F'_{ehi} = F_{ehi} \frac{\sum_{i \in h} F_{ehi} V_{ehi}}{\sum_{i \in h} F_{ehi} V_{ehi}^*}$$

Donde:

F'_{ehi} = factor de expansión corregido por No respuesta de la i-ésima entidad, del h-ésimo estrato, de la e-ésima entidad.

F_{ehi} = factor de expansión de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.

V_{ehi} = número de viviendas seleccionadas en la i-ésima entidad, en el h-ésimo estrato, en la e-ésima entidad.

V_{ehi}^* = número de viviendas con respuesta en la i-ésima entidad, en el h-ésimo estrato, en la e-ésima entidad.

1.8.2 Ajuste por proyección

Los factores de expansión ajustados por la No respuesta se corrigieron con el fin de asegurar que en cada dominio de interés del módulo se obtenga la población total determinada por la proyección de población generada por INEGI referida al punto medio del levantamiento, mediante la siguiente expresión:

Donde:

$$F_D'' = F_D' \frac{PROy_D}{PEXP_D}$$

1.9 Estimadores de los indicadores

El estimador del total de la característica X es:

$$\hat{X} = \sum_e \sum_h \sum_i F_{ehij}^{UA} \left(\sum_s \sum_{\ell} X_{ehis\ell}^{UA} \right) + \sum_e \sum_h \sum_i F_{ehij}^{CU} \left(\sum_s \sum_{\ell} X_{ehis\ell}^{CU} \right) + \sum_e \sum_h \sum_i F_{ehij}^R \left(\sum_s \sum_{\ell} X_{ehis\ell}^R \right)$$

Donde:

F_{ehij}^{UA} = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio urbano alto.

$X_{ehis\ell}^{UA}$ = valor observado de la característica de interés X en la ℓ -ésima persona, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el dominio urbano alto.

F_{ehij}^{CU} = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, en el dominio complemento urbano.

$X_{ehis\ell}^{CU}$ = valor observado de la característica de interés X en la ℓ -ésima persona, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el dominio complemento urbano.

F_{ehij}^R = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad del dominio rural.

$X_{ehis\ell}^R$ = valor observado de la característica de interés en la ℓ -ésima persona, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, del dominio rural.

Para la estimación de proporciones, tasas y promedios se utilizó el estimador de razón:

$$\hat{R} = \frac{\hat{X}}{\hat{Y}}$$

Donde \hat{Y} se define en forma análoga a \hat{X}

1.10 Estimación de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones estatales y nacionales se usó el método de Conglomerados Últimos³, basado en que la mayor contribución a la varianza de un estimador en un diseño bietápico, es la que se presenta entre las Unidades Primarias de Muestreo (UPM). El término “Conglomerados Últimos” se utiliza para denotar el total de unidades en muestra de una unidad primaria de muestreo.

Para obtener las precisiones estadísticas de los estimadores de razón, conjuntamente al método de Conglomerados Últimos se aplicó el método de series de Taylor, obteniéndose la siguiente fórmula para estimar la precisión de \hat{R} a nivel nacional:

$$\hat{V}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_{e=1}^{32} \left\{ \sum_{h=1}^{L_e} \frac{k_{eh}}{k_{eh} - 1} \sum_{i=1}^{k_{eh}} \left[\left(\hat{X}_{ehi} - \frac{1}{k_{eh}} \hat{X}_{eh} \right) - \hat{R} \left(\hat{Y}_{ehi} - \frac{1}{k_{eh}} \hat{Y}_{eh} \right) \right]^2 \right\}$$

Donde:

\hat{X}_{ehi} = total ponderado de la variable de estudio X en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

\hat{X}_{eh} = total ponderado de la variable de estudio X en el h-ésimo estrato, en la e-ésima entidad.

K_{eh} = número de UPM en el h-ésimo estrato, en la e-ésima entidad.

L_e = número de estratos en la e-ésima entidad.

\hat{Y}^2 = el cuadrado del total ponderado de la variable de estudio Y.

Estas definiciones son análogas para la variable de estudio Y.

La estimación de la varianza del estimador de un total, se calcula con la siguiente expresión:

$$\hat{V}(\hat{X}_{NAL}) = \sum_{e=1}^{32} \sum_{h=1}^{L_e} \frac{k_{eh}}{k_{eh} - 1} \sum_{i=1}^{k_{eh}} \left(\hat{X}_{ehi} - \frac{1}{k_{eh}} \hat{X}_{eh} \right)^2$$

³Vease Hansen, M. H. Horwitz, W.N. y Madow, W.G., Sample Survey Methods and Theory, (1953) Vol. 1 pág. 242.

Las estimaciones de la desviación estándar (D.E.), efecto de diseño (DEFF) y coeficiente de variación (C.V.) se calculan mediante las siguientes expresiones:

$$D.E. = \sqrt{\hat{V}(\hat{\theta})}$$

$$DEFF = \frac{\hat{V}(\hat{\theta})}{\hat{V}(\hat{\theta})_{MAS}}$$

$$C.V. = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}}$$

Donde:

$\hat{\theta}$ = estimador del parámetro poblacional θ .

$\hat{V}(\hat{\theta})_{MAS}$ = estimador de la varianza bajo muestreo aleatorio simple.

$\hat{V}(\hat{\theta})$ = estimador de la varianza bajo el diseño de muestreo descrito en este documento.

Finalmente, el intervalo de confianza $I_{1-\alpha}$ al $100(1-\alpha)\%$, se calcula de la siguiente forma:

$$I_{1-\alpha} = \left(\hat{\theta} - z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\theta})} \right)$$

Donde α es el nivel de significancia.