

Encuesta Nacional de Ingresos y Gastos de los Hogares 2012 ENIGH

Diseño muestral

Formación de las unidades primarias
de muestreo para el levantamiento



Obras complementarias publicadas por el INEGI sobre el tema:

Encuesta Nacional de Ingresos y Gastos de los Hogares Documento Metodológico, Gastos de Hogares ENIGH, Condiciones Socioeconómicas ENIGH.

Catalogación en la fuente INEGI:

640.4201 Encuesta Nacional de Ingresos y Gastos de los Hogares (2012).
Encuesta Nacional de Ingresos y Gastos de los Hogares 2012 : ENIGH : diseño muestral : formación de las unidades primarias de muestreo para el levantamiento / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2013.

11 p.

ISBN 978-607-494-674-1.

1. Hogar - México - Encuestas - Metodología. 2. Economía doméstica. I. Instituto Nacional de Estadística y Geografía (México).

Si requiere más información sobre esta obra, favor de contactarnos a través de:

Centros de consulta y comercialización (consulte el domicilio en Internet)

Larga distancia sin costo: 01 800 111 46 34

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

Síguenos en:   

Presentación

El **Instituto Nacional de Estadística y Geografía (INEGI)** presenta seis documentos metodológicos en los que se exponen las principales características de la **Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)**.

En el documento de **Diseño muestral**, se describen las actividades que constituyen el proceso de obtención y tratamiento estadístico de la muestra, que inicia con la determinación de la cobertura y dominios de interés para los cuales se pueden obtener estimaciones estadísticamente confiables, en función de los cuales se calcula el tamaño de la muestra y se determina su distribución, seguida de la selección de las unidades de muestreo. Como parte del tratamiento, se describe el cálculo de las probabilidades de selección y su empleo para asignar el peso que debe tener cada unidad seleccionada. Finalmente se presentan las expresiones empleadas para obtener las estimaciones y sus medidas de calidad.

Índice

1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Marco de la encuesta	1
4.1 Formación de las unidades primarias de muestreo (UPM)	1
a) En urbano alto	1
b) En complemento urbano	1
c) En rural	2
4.2 Estratificación	2
5. Tamaño de la muestra	3
6. Afijación de la muestra	3
7. Selección de la muestra	3
7.1 En urbano alto	3
7.2 En complemento urbano	4
7.3 En rural	5
8. Ajuste a los factores de expansión	5
8.1 Ajuste por No respuesta	5
8.2 Ajuste por proyección	6
9. Estimadores	6
10. Estimadores de errores de muestreo	7
Anexo	9
A. Indicadores empleados en la estratificación de la muestra maestra por ámbito de estudio.	11
B. Distribución de la muestra en viviendas por entidad según dominio de estudio para la ENIGH-2012.	12

1. Objetivo de la encuesta

El objetivo de la ENIGH-2012 es obtener información estadística de la distribución, monto y estructura del ingreso y gasto de los hogares, así como las actividades económicas desarrolladas por los miembros de los hogares.

2. Población objetivo

La encuesta está dirigida a los hogares del territorio nacional.

3. Cobertura geográfica

La encuesta está diseñada para dar resultados a nivel nacional, con corte urbano y rural.

4. Marco de la encuesta

El diseño de la muestra para la ENIGH-2012 se caracteriza por ser probabilístico; en consecuencia, los resultados obtenidos de la encuesta se generalizan a toda la población. Para la selección de la muestra se utilizó el Marco Nacional de Viviendas 2002 del INEGI, construido a partir de la información cartográfica y demográfica que se obtuvo del XII Censo General de Población y Vivienda 2000. Este marco es en realidad una muestra maestra a partir de la cual se seleccionan las muestras para todas las encuestas en viviendas que realiza el INEGI. Su diseño es probabilístico, estratificado, unietápico y por conglomerados; estos últimos también se consideran unidades primarias de muestreo, pues es en ellos donde se seleccionan, en etapas subsecuentes, las viviendas que integran las muestras de las diferentes encuestas.

4.1 Formación de las 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 de la misma AGEB.¹
- La unión de dos o más manzanas contiguas de diferentes AGEB de la misma localidad.
- La unión de dos o más manzanas contiguas de diferentes localidades, que pertenezcan al mismo tamaño de localidad.

b) En complemento 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 de la misma AGEB.

¹ Área Geoestadística Básica.

- La unión de dos o más manzanas contiguas de diferentes AGEB de la misma localidad.
- La unión de dos o más manzanas contiguas de diferentes AGEB y localidades, pero del mismo municipio.

c) En rural

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

- Una AGEB.
- Parte de una AGEB.
- La unión de dos o más AGEB colindantes del mismo municipio.
- La unión de una AGEB con una parte de otra AGEB colindante del mismo municipio.

4.2 Estratificación

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

En cada entidad federativa se distinguen tres ámbitos, divididos a su vez en siete zonas, como se indica en el siguiente cuadro:

Ámbito	Zona	Tamaño de localidad
Urbano alto	01	32 ciudades autorrepresentadas (C.A.) con 100 000 o más habitantes.
	02	Resto de las ciudades con 100 000 o más habitantes.
Complemento urbano	25	De 50 000 a 99 999 habitantes.
	35	De 15 000 a 49 999 habitantes.
	45	De 5 000 a 14 999 habitantes.
	55	De 2 500 a 4 999 habitantes.
Rural	60	Localidades menores de 2 500 habitantes.

De manera paralela, en una primera etapa se formaron cuatro estratos en los que se agruparon 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, expresadas por medio de 24 indicadores construidos con información del XII Censo General de Población y Vivienda 2000, para lo cual se emplearon métodos estadísticos multivariados.

En una segunda etapa, cada UPM fue asignada de acuerdo con su estrato geográfico (entidad-ámbito-zona).

En una tercera etapa, al interior de cada zona y estrato (sociodemográfico), algunas de las UPM se sometieron a un nuevo proceso de estratificación con el propósito de tener una mayor diferenciación a ese nivel, para esta estratificación se utilizaron indicadores diferenciados por ámbito.* Como resultado se tiene un total de 888 subestratos en todo el ámbito nacional.

*La descripción de estos indicadores se presenta en el cuadro 1.

5. Tamaño de la muestra

Para el cálculo del tamaño de muestra de la ENIGH-2012 se consideró como variable de referencia el promedio del ingreso corriente total por hogar. La expresión utilizada fue la siguiente:

$$n = \frac{z^2 s^2 \text{ DEFF}}{r^2 \bar{x}^2 (1 - \text{tnr}) \text{ PHV}}$$

Donde:

- n = tamaño de la muestra.
- z = valor asentado en las tablas estadísticas de la distribución normal estándar para una confianza prefijada.
- s² = estimación de la varianza poblacional de la variable de interés.
- \bar{x} = estimación del promedio de la variable de interés.
- DEFF = efecto de diseño definido como el cociente de la varianza en la estimación del diseño utilizado, entre la varianza obtenida considerando un muestreo aleatorio simple para un mismo tamaño de muestra.
- r = error relativo máximo aceptable.
- tnr = tasa de No respuesta máxima esperada.
- PHV = promedio de hogares por vivienda.

Fijando un nivel de confianza de 90%, un efecto de diseño de 2.59,** una varianza poblacional de 1534300947** un error relativo máximo aceptable de 3.18%, un promedio de ingreso corriente total por hogar de 34 935.59,** una tasa de No respuesta máxima esperada de 15% y un promedio de hogares por vivienda de 1.02, se determinó una muestra a nivel nacional de 10 062 viviendas.

6. Afijación de la muestra

La muestra se distribuyó aproximadamente igual entre las 32 entidades federativas. Al interior de éstas, se asignó de manera proporcional al tamaño de los estratos del Marco Nacional de Viviendas 2002.

En el cuadro 2, se presenta la distribución de la muestra en viviendas por entidad según dominio de estudio.

7. Selección de la muestra

La selección de la muestra es bietápico, donde las unidades de selección de segunda etapa fueron las viviendas y la unidad de observación el hogar. Se hizo una selección independiente para cada entidad y estrato, el procedimiento varió dependiendo de la zona como se especifica a continuación:

7.1 En urbano alto

En la zona urbano alto la selección de la muestra se realizó en forma independiente por cada ciudad y estrato mediante el siguiente procedimiento:

1. De las n_{ech} UPM que integran el marco de la muestra maestra, se eligieron n_{ech}^* UPM con igual probabilidad para la ENIGH-2012.
2. En cada UPM se seleccionaron 5 viviendas con igual probabilidad para la ENIGH-2012.

*Estos parámetros fueron obtenidos de la ENIGH-2010.

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

$$P\{V_{echi}\} = \frac{n_{ech} m_{echi}}{m_{ech}} \frac{n_{ech}^*}{n_{ech}} \frac{5}{m_{echi}^*} = \frac{5 n_{ech}^* m_{echi}}{m_{ech} m_{echi}^*}$$

Su factor de expansión² está dado por:

$$F_{echi} = \frac{m_{ech} m_{echi}^*}{5 n_{ech}^* m_{echi}}$$

Donde:

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

n_{ech}^* = número de UPM seleccionadas para la ENIGH-2012, en el h -ésimo estrato, en la c -ésima ciudad, en la e -ésima entidad.

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

m_{echi} = número de viviendas en la i -ésima UPM, en el h -ésimo estrato, en la c -ésima ciudad, en la e -ésima entidad.

m_{echi}^* = número de viviendas en la i -ésima UPM, en el h -ésimo estrato, en la c -ésima ciudad, en la e -ésima entidad al momento de la actualización del listado de viviendas, previo al levantamiento de la ENIGH-2012.

7.2 En complemento urbano

1. De las n_{eh} UPM que se seleccionaron para el marco de la muestra maestra, se eligieron n_{eh}^* UPM con igual probabilidad para la ENIGH-2012.

2. En cada UPM seleccionada, se eligieron 20 viviendas con igual probabilidad.

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

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

Su factor de expansión está dado por:

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

Donde:

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

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

- m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.
- m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- n_{eh}^* = número de UPM seleccionadas para la ENIGH-2012, en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi}^* = número total de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento de levantamiento de la ENIGH-2012.

7.3 En rural

De las n_{eh} UPM que se seleccionaron para el marco de la muestra maestra, se eligieron n_{eh}^* UPM con igual probabilidad para la ENIGH-2012.

En cada UPM seleccionada, se eligieron 2 segmentos de 10 viviendas aproximadamente, con igual probabilidad.

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

$$P\{V_{ehi}\} = \frac{n_{eh} m_{ehi} n_{eh}^* 2 \cdot 10}{m_{eh} n_{eh} m_{ehi}^*} = \frac{20 n_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^*}$$

En consecuencia, su factor de expansión está dado por:

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

Donde:

- n_{eh} = número de UPM seleccionadas del marco de la muestra maestra, 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.
- m_{eh} = número total de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- n_{eh}^* = número de UPM seleccionadas para la ENIGH-2012, en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi}^* = número total de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento del levantamiento de la ENIGH-2012.

8. Ajuste a los factores de expansión

Los factores de expansión elaborados conforme al procedimiento antes descrito se ajustan para los siguientes conceptos:

8.1 Ajuste por No respuesta

El ajuste por No respuesta atribuida al informante se realiza a nivel UPM, mediante la siguiente expresión:

$$F'_{echi} = F_{echi} \frac{nv_{h_{echi}}}{nv_{hcr_{echi}}}$$

Donde:

F'_{echi} = factor de expansión corregido por No respuesta para las viviendas de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad, de la e-ésima entidad.

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

$nv_{h_{echi}}$ = número de viviendas habitadas en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad, en la e-ésima entidad.

$nv_{hcr_{echi}}$ = número de viviendas habitadas con respuesta en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad, en la e-ésima entidad.

8.2 Ajuste por proyección

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

$$F''_D = F'_D \frac{PROY_D}{PEXP_D}$$

Donde:

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 la proyección.

$PEXP_D$ = población total a la que expande la encuesta en el dominio D.

9. Estimadores

El estimador del total de la característica X, a nivel nacional es:

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

Donde:

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

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

F_{ehi}^R = factor de expansión final 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 X en el ℓ -ésimo hogar, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad del dominio rural.

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

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

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

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

Donde, la variable \hat{Y} es definida en forma análoga a \hat{X} .

10. Estimadores de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones se usó el método de “Conglomerados Últimos”³, basado en que la mayor contribución a la varianza de un estimador, en un diseño polietápico, 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 de los estimadores de razón, se aplicó el método de Conglomerados Últimos conjuntamente con el método de series de Taylor, obteniéndose la siguiente fórmula para estimar la precisión de \hat{R} .

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

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.

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

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.

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

Las estimaciones del error estándar (E.E.), coeficiente de variación o error relativo del estimador (C.V.) y el efecto de diseño (DEFF) se calculan mediante las siguientes expresiones:

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

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

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

$$DEFF = \frac{\hat{V}(\hat{\epsilon})}{\hat{V}(\hat{\epsilon})_{mas}}$$

Donde:

$\hat{\epsilon}$ = estimador del parámetro poblacional.

$\hat{V}(\hat{\epsilon})_{mas}$ = estimador de la varianza, bajo un muestreo aleatorio simple.

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

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

Anexo

A. Indicadores empleados en la estratificación de la muestra maestra por ámbito de estudio.

Descripción del indicador	Ámbito de estudio
	Nacional
% De vivienda	
Que disponen de agua entubada dentro de la vivienda	
Con drenaje	X
Con electricidad	
Que disponen de agua, luz y drenaje	X
Con piso diferente de tierra	X
Con paredes de material sólido	
Con cocina exclusiva	X
Sin hacinamiento	X
Con servicio sanitario exclusivo con conexión de agua	
Con servicio sanitario exclusivo con admisión de agua	X
Que utilizan gas para cocinar	X
Con radio o radiograbadora	X
Con televisión	X
Con refrigerador	X
Con licuadora	X
Con automóvil o camioneta propios	X
Con video casetera	
Con lavadora	X
Con teléfono	
Con calentador de agua	
Con cuatro bienes (teléfono, refrigerador, lavadora y boiler)	
Con cuatro bienes (radio, televisión, licuadora y refrigerador)	X
Con el mínimo equipamiento (radio o televisión y licuadora)	
% De población	
Derechohabiente a servicio de salud	
De 6 a 17 años que asiste a la escuela	X
De 6 a 14 años que asiste a la escuela	
De 15 años y más alfabeto	X
De 15 años y más con postprimaria	X
Grado promedio de escolaridad	X
Ocupada que gana más de 2.5 salarios mínimos	X
Ocupada que gana más de 5 salarios mínimos	X
Femenina de 12 años y más económicamente activa	X
Económicamente activa de 20 a 49 años	X
% Otros	
De hogares de los deciles 8, 9 y 10 a nivel nacional	X
Relación de dependencia económica	X
Total de indicadores	24

B. Distribución de la muestra en viviendas por entidad según dominio de estudio para la (ENIGH-2012).

Cve.	Entidad	Urbano	Rural	Total
01	Aguascalientes	200	101	301
02	Baja California	260	39	299
03	Baja California Sur	200	103	303
04	Campeche	180	126	306
05	Coahuila de Zaragoza	220	79	299
06	Colima	180	122	302
07	Chiapas	100	204	304
08	Chihuahua	220	80	300
09	Distrito Federal	400	0	400
10	Durango	160	141	301
11	Guanajuato	160	137	297
12	Guerrero	120	178	298
13	Hidalgo	120	186	306
14	Jalisco	300	101	401
15	México	320	84	404
16	Michoacán de Ocampo	140	164	304
17	Morelos	180	122	302
18	Nayarit	160	139	299
19	Nuevo León	275	25	300
20	Oaxaca	100	198	298
21	Puebla	180	118	298
22	Querétaro	160	142	302
23	Quintana Roo	200	102	302
24	San Luis Potosí	140	164	304
25	Sinaloa	180	124	304
26	Sonora	240	61	301
27	Tabasco	120	185	305
28	Tamaulipas	240	63	303
29	Tlaxcala	200	102	302
30	Veracruz de Ignacio de la Llave	180	230	410
31	Yucatán	220	81	301
32	Zacatecas	140	166	306
Total		6 195	3 867	10 062