



Encuesta Nacional de la Dinámica Demográfica 2006

DISEÑO MUESTRAL

Noviembre 2008



Vivir Mejor



Instituto Nacional
de Salud Pública



INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA
GEOGRÁFICA E INFORMÁTICA



CONAPO



SALUD

GOBIERNO
FEDERAL

SEGOB



ESTADOS UNIDOS MEXICANOS

1. Objetivo de la encuesta

Obtener información estadística sobre las características de la población de tipo demográfico, social, de salud, de ocupación y migración, con la finalidad de apoyar el diseño de las políticas sociales y demográficas.

2. Población objetivo

La encuesta está dirigida a la población que reside permanentemente en viviendas particulares dentro del territorio nacional.

3. Cobertura

La encuesta esta diseñada para dar resultados a nivel nacional y por tamaño de localidad.

Tamaño de localidad

1. Localidades de 100 000 y más habitantes.
2. Localidades de 15 000 a 99 999 habitantes.
3. Localidades de 2 500 a 14 999 habitantes.
4. Localidades menores a 2 499 habitantes.

4. Diseño de la muestra

El diseño muestral de la ENADID 2006 se caracteriza por ser probabilístico, por lo cual los resultados obtenidos de la encuesta se generalizan a toda la población. A su vez es bietápico, estratificado y por conglomerados, donde la unidad última de selección es la vivienda y las unidades de observación son las personas.

4.1 Marco de la encuesta

El marco de muestreo que se empleó para la ENADID 2006 es 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; como tal, su diseño es probabilístico, estratificado, unietápico y por conglomerados a los que se denominó unidades primarias de muestreo, pues es en ellas donde se seleccionan en una segunda etapa las viviendas que integran las muestras de las diferentes encuestas.

4.2 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 AGEB¹.
- 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 pero del 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.
- 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 del mismo municipio.

c) 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 parte de otra AGEB colindante del mismo municipio.

4.3 Estratificación

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

¹ Área Geoestadística Básica.

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 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 multivariados².

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

En una tercera etapa, al interior de cada zona y estrato sociodemográfico, algunas 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.

5. Esquema de muestreo

La ENADID 2006 fue diseñada bajo un esquema de muestreo probabilístico, estratificado, bietápico y por conglomerados. A continuación se mencionan las definiciones de cada uno de estos métodos.

a) Probabilístico

Porque las unidades de selección tienen una probabilidad conocida y distinta de cero de ser seleccionadas.

b) Estratificado

Porque las unidades primarias de muestreo con características similares, que pertenecen a un mismo tamaño de localidad, se agrupan para formar estratos.

c) Bietápico

Porque la unidad última de muestreo (vivienda) es seleccionada en dos etapas.

d) Por conglomerados

² Ver documento preliminar del diseño muestral de la muestra maestra.

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

Porque las unidades primarias de muestreo son conjuntos de unidades muestrales.

6. Tamaño de la muestra

El tamaño de muestra se calculó para la proporción de migración internacional, considerada como una de las principales variables de la encuesta, y la cual requiere los tamaños de muestra mayores. Esto garantiza que las estimaciones del resto de las variables de interés queden cubiertas con este tamaño. La expresión empleada para el cálculo es la siguiente:

$$n = \frac{z^2 q \text{DEFF}}{r^2 p (1 - \text{tnr}) \text{PHV}}$$

Donde:

n = tamaño de la muestra.

p = estimación de la proporción de interés.

q = 1- p.

r = error relativo máximo aceptable.

z = valor asentado en las tablas estadísticas de la distribución normal estándar para una confianza prefijada.

DEF = efecto de diseño, definido como el cociente de la varianza obtenida en la estimación con el F diseño utilizado, entre la varianza obtenida, considerando un muestreo aleatorio simple con un mismo tamaño de muestra.

tnr = tasa de No Respuesta máxima esperada.

PHV = promedio de personas de cinco y más años de edad por vivienda.

Considerando una confianza de 90 por ciento, un efecto de diseño de 4.65 observado en experiencias anteriores, un error relativo máximo de 6.5 por ciento, una tasa de No Respuesta máxima esperada de 15 por ciento, un promedio de residentes de 5 años y más por vivienda de 3.66, para una proporción de 2.27 por ciento, se determinó una muestra a nivel nacional de 41 215 viviendas, la cual se ajustó a 42 000.

7. Afijación de la muestra

La afijación de la muestra se realizó, en un primer paso, dentro de cada entidad federativa, entre los tamaños de localidad de manera proporcional y, en un segundo paso, dentro de cada tamaño de localidad entre los estratos de manera proporcional a su tamaño, para lo cual se empleó la siguiente expresión:

$$n_{\text{eth}} = \frac{N_{\text{eth}}}{N_{\text{et}}} n_{\text{et}}$$

Donde:

n_{eth} = número de viviendas en muestra en el h-ésimo estrato, del t-ésimo tamaño de localidad, en la e-ésima entidad.

n_{et} = número total de viviendas en muestra en t-ésimo tamaño de localidad, de la e-ésima entidad.

N_{eth} = número total de viviendas en el h-ésimo estrato, del t-ésimo tamaño de localidad, en la e-ésima entidad.

N_{et} = número total de viviendas del t-ésimo tamaño de localidad, de la e-ésima entidad.

En el cuadro 2 se presenta la distribución de la muestra por entidad y tamaño de localidad.

8. Selección de la muestra

La selección de la muestra para la ENADID 2006 se realizó de manera independiente por entidad, tamaño de localidad y estrato. El procedimiento de selección varió de acuerdo al tamaño de localidad.

8.1 Tamaños de localidad 1, 2 y 3

1. De las n_{eth} UPM seleccionadas en el marco de la muestra maestra, se seleccionaron n_{eth}^* UPM con igual probabilidad para la ENADID 2006.
2. En cada UPM seleccionada, se seleccionaron diez viviendas con igual probabilidad.

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

$$P\{V_{ethi}\} = \frac{n_{eth} m_{ethi} n_{eth}^*}{m_{eth} n_{eth} m_{ethi}^*} = \frac{10 n_{eth}^* m_{ethi}}{m_{eth} m_{ethi}^*}$$

Su factor de expansión³ está dado por:

$$F_{ethi} = \frac{m_{eth} m_{ethi}^*}{10 n_{eth}^* m_{ethi}}$$

Donde:

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

m_{ethi} = número de viviendas en la i-ésima UPM, del h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad, según el XII Censo General de Población y Vivienda 2000.

m_{eth} = número de viviendas en el h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad.

m_{ethi}^* = número de viviendas en la i-ésima UPM, del h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad, según listado de viviendas actualizado.

n_{eth}^* = número de UPM seleccionadas para la ENADID 2006, en el h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad.

³ El factor de expansión se define como el inverso de la probabilidad de selección.

8.2 Tamaño de localidad 4

1. De las n_{eth} UPM seleccionadas para el marco de la muestra maestra, se seleccionaron n_{eth}^* UPM con igual probabilidad para la ENADID 2006.
2. En cada UPM seleccionada, se tomaron dos segmentos de aproximadamente diez viviendas, con igual probabilidad.

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

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

Su factor de expansión está dado por:

$$F_{ethi} = \frac{m_{eth} m_{ethi}^*}{20 n_{eth}^* m_{ethi}}$$

Donde:

- n_{eth} = número de UPM seleccionadas en el h -ésimo estrato, del t -ésimo tamaño de localidad, de la e -ésima entidad, para el marco de la muestra maestra.
- m_{eth} = número de viviendas en el h -ésimo estrato, del t -ésimo tamaño de localidad, de la e -ésima entidad.
- m_{ethi} = número de viviendas en la i -ésima UPM, del h -ésimo estrato, del t -ésimo tamaño de localidad, de la e -ésima entidad, según el XII Censo General de Población y Vivienda 2000.
- m_{ethi}^* = número de viviendas en la i -ésima UPM, del h -ésimo estrato, del t -ésimo tamaño de localidad, de la e -ésima entidad, según listado de viviendas actualizado.
- n_{eth}^* = número de UPM seleccionadas para la ENADID 2006, en el h -ésimo estrato, del t -ésimo tamaño de localidad, de la e -ésima entidad.

9. Ajuste de los factores de expansión

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

9.1 Ajuste por No Respuesta

El ajuste por No Respuesta de la persona objeto de estudio al interior de la vivienda atribuida al informante se realiza a nivel UPM, en cada uno de los dominios, mediante la siguiente expresión:

$$F'_{ethi} = F_{ethi} \frac{nv_{ethi}}{nvhcR_{ethi}}$$

Donde:

- F'_{ethi} = factor de expansión corregido por No Respuesta para las viviendas de la i-ésima UPM, del h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad.
- nvh_{ethi} = número de viviendas habitadas seleccionadas en la i-ésima UPM, del h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad.
- $nvhcR_{ethi}$ = número de viviendas habitadas seleccionadas con respuesta en la i-ésima UPM, del h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad.

9.2 Ajuste a cifras del II Censo Nacional de Población y Vivienda

Los factores de expansión ajustados por la 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:

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

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.
- POB_D = población en el dominio D, según II Censo.
- $PEXP_D$ = población total a la que expande la encuesta en el dominio D.

9.3 Ajuste a nivel persona para las mujeres de 15 a 54 años

Los factores de expansión ajustados a la población del Coneto, se estiman a nivel vivienda para la población de mujeres de 15 a 54 años de edad, mediante la siguiente expresión:

$$F'''_{ethi} = F''_{ethi} \frac{M_{15-54}}{M^*_{15-54}}$$

Donde:

- F'''_{ethi} = factor de expansión ajustado a nivel persona de las mujeres de 15 a 54 años de edad para las viviendas de la i-ésima UPM, del h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad.
- F''_{ethi} = factor de expansión ajustado a la población del Coneto para las viviendas de la i-ésima UPM, del h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad.
- M_{15-54} = total de mujeres de 15 a 54 años de edad en la vivienda seleccionada.
- M^*_{15-54} = total de mujeres de 15 a 54 años de edad que contestaron el cuestionario del módulo de fecundidad en la vivienda seleccionada.

10. Estimadores

El estimador del total de la característica X es:

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

Donde:

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

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

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

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

F_{ethi}^R = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, del t-ésimo tamaño de localidad, del e-ésimo entidad del dominio rural.

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

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

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

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

11. Estimación de las precisiones

Para la evaluación de los errores de muestreo de las principales estimaciones estatales y nacionales se usó el método de Conglomerados Últimos⁴, basado en que la mayor contribución a la varianza de un estimador, en un diseño bietápico, es la que se presenta entre las 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, 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} :

⁴ Vease Hasen, M. H. Horwitz, W.N. y Madow, W.G., *Sample Survey Methods and Theory*, (1953) Vol. 1 pág. 242.

$$\hat{V}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_{e=1}^{32} \left\{ \sum_h^{L_e} \frac{n_{eth}}{n_{eth}-1} \sum_i^{n_{eth}} \left[\left(\hat{X}_{ethi} - \frac{1}{n_{eth}} \hat{X}_{eth} \right) - \hat{R} \left(\hat{Y}_{ethi} - \frac{1}{n_{eth}} \hat{Y}_{eth} \right) \right]^2 \right\}$$

Donde:

\hat{X}_{ethi} = total ponderado de la variable de estudio X para la i-ésima UPM, en el h-ésimo estrato, del t-ésimo tamaño de localidad, de la e-ésima entidad.

\hat{X}_{eth} = total ponderado de la variable de estudio X para el h-ésimo estrato, del t-ésimo tamaño de localidad, en la e-ésima entidad.

n_{eth} = número de UPM en el h-ésimo estrato, del t-ésimo tamaño de localidad, para 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:

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

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

$$D.E. = \sqrt{\hat{V}(\hat{\theta})} \quad DEFF = \frac{\hat{V}(\hat{\theta})}{\hat{V}(\hat{\theta})_{MAS}} \quad C.V. = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}}$$

Donde:

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

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

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

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

Indicadores empleados en la estratificación del marco de la muestra maestra por ámbito de estudio

Cuadro 1

Descripción del indicador	Ámbito de estudio			
	Nacional	Urbano alto	Complemento urbano	Rural
% De vivienda				
Que disponen de agua entubada dentro de la vivienda		X		
Con drenaje	X	X		X
Con electricidad				X
Que disponen de agua, luz y drenaje	X	X	X	X
Con piso diferente de tierra	X			X
Con paredes de material sólido				X
Con cocina exclusiva	X	X	X	X
Sin hacinamiento	X	X	X	X
Con servicio sanitario exclusivo con conexión de agua		X		
Con servicio sanitario exclusivo con admisión de agua	X		X	
Que utilizan gas para cocinar	X			X
Con radio o radiograbadora	X			X
Con televisión	X			
Con refrigerador	X	X	X	
Con licuadora	X			X
Con automóvil o camioneta propios	X	X	X	
Con videocasetera			X	
Con lavadora	X		X	
Con teléfono		X	X	
Con calentador de agua		X	X	
Con cuatro bienes (teléfono, refrigerador, lavadora y boiler)		X		
Con cuatro bienes (radio, televisión, licuadora y refrigerador)	X		X	
Con el mínimo equipamiento (radio o televisión y licuadora)				X
% De población				
Derechohabiente a servicio de salud		X	X	
De 6 a 17 años que asiste a la escuela	X	X	X	
De 6 a 14 años que asiste a la escuela				X
De 15 años y más alfabeto	X			X
De 15 años y más con postprimaria	X	X	X	X
Grado promedio de escolaridad	X	X	X	X
Ocupada que gana más de 2.5 salarios mínimos	X	X	X	X
Ocupada que gana más de 5 salarios mínimos	X	X	X	
Femenina de 12 años y más económicamente activa	X	X	X	
Económicamente activa de 20 a 49 años	X	X	X	
% De otros				
Hogares de los deciles 8, 9 y 10 a nivel nacional	X	X	X	
Relación de dependencia económica	X	X	X	
Total de indicadores	24	21	21	16

Distribución de la muestra en viviendas por entidad y tamaño de localidad para la ENADID-2006

Cuadro 2

Entidad		Distribución de Viviendas				
Cve	Nombre	TLOC1	TLOC2	TLOC3	TLOC4	TOTAL
01	Aguascalientes	670	100	70	360	1 200
02	Baja California	710	60	70	360	1 200
03	Baja California Sur	540	180	120	360	1 200
04	Campeche	560	90	190	360	1 200
05	Coahuila de Zaragoza	610	180	50	360	1 200
06	Colima	220	490	130	360	1 200
07	Chiapas	420	180	240	360	1 200
08	Chihuahua	600	160	80	360	1 200
09	Distrito Federal	1 700	20	20	60	1 800
10	Durango	600	70	170	360	1 200
11	Guanajuato	650	350	120	480	1 600
12	Guerrero	540	140	160	360	1 200
13	Hidalgo	340	340	160	360	1 200
14	Jalisco	810	240	210	540	1 800
15	Estado de México	940	140	180	540	1 800
16	Michoacán de Ocampo	410	200	230	360	1 200
17	Morelos	390	190	260	360	1 200
18	Nayarit	440	150	250	360	1 200
19	Nuevo León	950	100	70	480	1 600
20	Oaxaca	140	290	410	360	1 200
21	Puebla	520	240	360	480	1 600
22	Querétaro de Arteaga	480	170	190	360	1 200
23	Quintana Roo	600	150	90	360	1 200
24	San Luis Potosí	560	110	170	360	1 200
25	Sinaloa	610	100	130	360	1 200
26	Sonora	600	140	100	360	1 200
27	Tabasco	360	210	270	360	1 200
28	Tamaulipas	640	130	70	360	1 200
29	Tlaxcala		420	420	360	1 200
30	Veracruz de Ignacio de la Llave	540	380	340	540	1 800
31	Yucatán	440	190	210	360	1 200
32	Zacatecas	240	400	200	360	1 200
Total		17 830	6 310	5 740	12 120	42 000