



Encuesta Nacional de Acceso a la Información Pública y Protección de Datos Personales 2016 ENAID Diseño Muestral



Índice	Página
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
4.2. Estratificación	1
5. Tamaño de la muestra	2
6. Afijación de la muestra	2
7. Selección de la muestra	3
8. Ajuste a los factores de expansión	4
8.1. Ajuste por No respuesta a nivel vivienda	4
8.2. Ajuste por proyección	4
9. Estimadores	4
10. Estimaciones de errores de muestreo	5
Anexos	
- Indicadores empleados en la estratificación del marco de la muestra maestra (Cuadro 1).	
- Distribución de la muestra en viviendas por entidad para la Encuesta Nacional de Acceso a la Información Pública y Protección de Datos Personales 2016 (ENAIID-2016) (Cuadro 2).	

1. Objetivo de la encuesta

Obtener información estadística que permita medir el grado de conocimiento, percepciones y actitudes que influyen en el ejercicio de los derechos de acceso a la información y protección de datos personales.

2. Población objetivo

La población objetivo son las personas de 18 años cumplidos o más que residen en viviendas particulares ubicadas en ciudades de 100 000 y más habitantes.

3. Cobertura geográfica

La ENAID-2016 proporciona resultados al agregado de 32 áreas urbanas de interés de 100 000 y más habitantes.

4. Marco de la encuesta

El diseño de la muestra para la ENAID-2016 se caracteriza por ser probabilístico, en consecuencia los resultados obtenidos de la encuesta se generalizan a toda la población. A la vez el diseño es trietápico, estratificado y por conglomerados, donde la unidad última de selección es una persona de 18 años cumplidos o más en cada vivienda seleccionada.

El marco de muestreo que se empleó para la ENAID-2016 es el Marco Nacional de Viviendas 2012 del INEGI, construido a partir de información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2010. 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 ellos donde se seleccionan en una segunda etapa, 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, en el caso del urbano alto es de la siguiente manera:

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 pero del mismo tamaño de localidad.

4.2. Estratificación

En el ámbito urbano alto, se forman 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

¹ Área Geoestadística Básica.

el equipamiento de las mismas, expresadas por medio de 34 indicadores* construidos con información del Censo de Población y Vivienda 2010, para lo cual se emplearon métodos estadísticos multivariados.

5. Tamaño de la muestra

Para el cálculo del tamaño de la muestra se empleó la siguiente expresión:

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

Donde:

- n = tamaño de la muestra.
- p = estimación de la proporción de interés.
- q = (1-p).
- z = valor asentado en las tablas estadísticas que garantiza realizar las estimaciones con una confianza prefijada.
- r = error relativo máximo esperado.
- tnr = tasa de No respuesta máxima esperada.
- 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.

Considerando una confianza de 90%, un efecto de diseño de 2.5, una tasa de No respuesta máxima esperada de 15%, un error relativo máximo esperado de 8%, para una proporción de 7.95%, se obtuvo un tamaño de muestra de 14 400 viviendas a nivel nacional.

6. Afijación de la muestra

La afijación de la muestra se realiza dentro de cada ciudad entre los diferentes estratos de manera proporcional a su tamaño, para lo cual se emplea la siguiente expresión:

$$n_{ch} = \frac{N_{ch}}{N_c} n_c$$

Donde:

- n_{ch} = número de viviendas en muestra en el h-ésimo estrato, en la c-ésima área urbana.
- n_c = número total de viviendas en muestra, en la c-ésima área urbana.
- N_{ch} = número total de viviendas en el h-ésimo estrato, en la c-ésima área urbana.
- N_c = número total de viviendas, en la c-ésima área urbana.

En el cuadro 2, se presenta la distribución de la muestra por entidad.

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

7. Selección de la muestra

La selección de la muestra para la ENAID-2016 se realizó en forma independiente para cada área urbana, el procedimiento es como sigue:

1. De las k_{ch} UPM que integran la muestra maestra² se eligieron k_{ch}^* UPM con igual probabilidad para la ENAID-2016.
 2. En cada UPM se seleccionaron cinco viviendas con igual probabilidad.
 3. En cada vivienda se seleccionó una persona de 18 años cumplidos o más.
- a) La probabilidad de seleccionar una vivienda de la i -ésima UPM, del h -ésimo estrato, de la c -ésima área urbana es:

$$P\{V_{chi}\} = \frac{k_{ch} m_{chi}}{m_{ch}} \frac{k_{ch}^*}{k_{ch}} \frac{5}{m_{chi}^*} = \frac{5 k_{ch}^* m_{chi}}{m_{ch} m_{chi}^*}$$

Su factor de expansión³ está dado por:

$$F_{chi} = \frac{m_{ch} m_{chi}^*}{5 k_{ch}^* m_{chi}}$$

- b) La probabilidad de seleccionar una persona de 18 años cumplidos o más de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la c -ésima área urbana, H_{chij} , es:

$$P\{H_{chij}\} = \frac{k_{ch} \cdot m_{chi}}{m_{ch}} \frac{k_{ch}^*}{k_{ch}} \frac{5}{m_{chi}^*} \frac{1}{H} = \frac{5 k_{ch}^* m_{chi}}{m_{ch} m_{chi}^* H}$$

Su factor de expansión está dado por:

$$F_{chij} = \frac{H m_{ch} m_{chi}^*}{5 k_{ch}^* m_{chi}}$$

Donde:

- k_{ch} = número de UPM seleccionadas para el marco de la muestra maestra en el h -ésimo estrato, en la c -ésima área urbana.
- m_{chi} = número de viviendas en la i -ésima UPM, en el h -ésimo estrato, en la c -ésima área urbana en el Censo de Población y Vivienda 2010.
- m_{ch} = total de viviendas en el h -ésimo estrato, en la c -ésima área urbana en el Censo de Población y Vivienda 2010.
- m_{chi}^* = total de viviendas en la i -ésima UPM, en el h -ésimo estrato, en la c -ésima área urbana al momento del levantamiento de la ENAID-2016.

² Del total de UPM que integran el marco de propósitos múltiples se seleccionaron con probabilidad proporcional a su tamaño k_{ch} UPM para localidades de 100 000 y más habitantes.

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

k_{ch}^* = número de UPM seleccionadas, en el h-ésimo estrato, en la c-ésima área urbana.

H = número de personas de 18 años cumplidos o más en la vivienda seleccionada.

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 a nivel vivienda

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

$$F'_{ch} = F_{ch} \frac{V_{ch}}{V_{ch}^*}$$

Donde:

F'_{ch} = factor de expansión corregido por No respuesta para las viviendas del h-ésimo estrato, de la c-ésima área urbana.

F_{ch} = factor de expansión del h-ésimo estrato, de la c-ésima área urbana.

V_{ch} = número de viviendas habitadas seleccionadas en el h-ésimo estrato, en la c-ésima área urbana.

V_{ch}^* = número de viviendas habitadas seleccionadas en el h-ésimo estrato, en la c-ésima área urbana para las cuales se obtuvo respuesta.

8.2. Ajuste por proyección

Los factores de expansión ajustados anteriormente se corrigen, a fin de asegurar que se obtenga la población determinada por la proyección de población generada por el 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 proyección.

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

D = nivel de desagregación en que se efectuó el ajuste.

9. Estimadores

El estimador del total de la característica X es:

$$\hat{X} = \sum_c \sum_h \sum_i \sum_j F_{chij} X_{chij}$$

Donde:

F_{chij} = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima área urbana.

X_{chij} = valor observado de la característica de interés X de personas seleccionadas, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima área urbana.

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} .

10. Estimaciones de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones nacionales se usó el método de Conglomerados Últimos,⁴ basado en que la mayor contribución a la varianza de un estimador, en un diseño multietápico 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} :

$$\hat{V}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_{c=1}^{32} \left\{ \sum_{h=1}^{L_c} \frac{k_{ch}}{k_{ch}-1} \sum_{i=1}^{k_{ch}} \left[\left(\hat{X}_{chi} - \frac{1}{k_{ch}} \hat{X}_{ch} \right) - \hat{R} \left(\hat{Y}_{chi} - \frac{1}{k_{ch}} \hat{Y}_{ch} \right) \right]^2 \right\}$$

Donde:

\hat{X}_{chi} = total ponderado de la variable de estudio X en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima área urbana.

\hat{X}_{ch} = total ponderado de la variable de estudio X en el h-ésimo estrato, en la c-ésima área urbana.

k_{ch} = número de UPM en el h-ésimo estrato, en la c-ésima área urbana.

L_c = es el número de estratos en la c-ésima área urbana.

Estas definiciones son análogas para la variable de estudio Y.

El cálculo de la varianza del estimador de un total, se obtiene con la siguiente expresión:

$$\hat{V}(\hat{X}_{NAL}) = \sum_{c=1}^{32} \sum_{h=1}^{L_c} \frac{k_{ch}}{k_{ch}-1} \sum_{i=1}^{k_{ch}} \left(\hat{X}_{chi} - \frac{1}{k_{ch}} \hat{X}_{ch} \right)^2$$

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

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

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

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

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

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 a $I_{1-\alpha}$ 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.

Mnemónico	Descripción
Población	
PPSSNOSP	Que tiene derecho a recibir servicios médicos en alguna institución de salud pública o privada excepto seguro popular.
PPDER_SS	Derechohabiente a servicios de salud.
PDP3A14A	De 3 a 14 años de edad que asiste a la escuela.
PDP15A24A	De 15 a 24 años de edad que asiste a la escuela.
PDP8A14ALF	De 8 a 14 años de edad que saben leer o escribir.
PDP15YM_SE	De 15 años o más de edad que aprobaron algún grado de escolaridad diferente al nivel preescolar.
PP15PRI_CO	De 15 años o más de edad que tienen como máxima escolaridad 6 grados aprobados en primaria.
PP15SEC_CO	De 15 años o más de edad que tienen como máxima escolaridad 3 grados aprobados en secundaria.
PGDO_ESC	Grado promedio de escolaridad.
PPEA	De 12 años y más que trabajaron; tenían trabajo pero no trabajaron o; buscaron trabajo en la semana de referencia.
PPEA_F	Femenina de 12 años y más que trabajaron; tenían trabajo pero no trabajaron o; buscaron trabajo en la semana de referencia.
PTASAOcupa	Tasa de ocupación.
TOCU12A17	No ocupada de 12 a 17 años de edad entre la población de 12 a 17 años de edad.
PPOMAYED	Ocupada de 18 y más años de edad entre la población ocupada.
Viviendas Particulares Habitadas	
PVIVSINH	Que no tienen hacinamiento.
PVPH_PISDT	Que tienen piso de cemento o firme, madera, mosaico u otro material.
PVPH2YMASD	Que usan para dormir entre 2 y 25 cuartos.
PVPH_2MASC	Que tienen más de un cuarto.
PVPH3YMASC	Que tienen entre 3 y 25 cuartos.
PVPH_C_ELE	Que disponen de luz eléctrica.
PVPHAGUADV	Que tienen disponibilidad de agua entubada dentro de la vivienda, o fuera de la vivienda pero dentro del terreno.
PVPH_EXCSA	Que tienen excusado, retrete, sanitario, letrina u hoyo negro.
PVPHDRENAJ	Que tienen drenaje conectado a la red pública, fosa séptica, barranca, grieta, río, lago o mar.
PVDREDERED	Que disponen de drenaje conectado a la red pública.
PVEXCAGU	Que disponen de excusado con descarga directa de agua.
PVPH_CSERV	Que disponen de luz eléctrica, agua entubada dentro o fuera de la vivienda, pero dentro del terreno, así como drenaje.
PSIN_HASIN	Que no se encuentran en situación de hacinamiento a nivel manzana.
Viviendas Particulares Habitadas que disponen de:	
PVPH_TV	Televisor
PVPH_AUTOM	Automóvil o camioneta
PVPH_CEL	Teléfono celular
PVCELFIJ	Teléfono celular y teléfono fijo
PV4ELEC	Radio, televisor, refrigerador y lavadora
PVRADTEL	Radio y televisor
PVPHCBIEEN	Todos los bienes

Distribución de la muestra en viviendas por entidad para la Encuesta Nacional de Acceso a la Información Pública y Protección de Datos Personales 2016 (ENAID-2016) Cuadro 2

Entidad		Viviendas
Clave	Nombre	
01	Aguascalientes	450
02	Baja California	450
03	Baja California Sur	450
04	Campeche	450
05	Coahuila de Zaragoza	450
06	Colima	450
07	Chiapas	450
08	Chihuahua	450
09	Ciudad de México	450
10	Durango	450
11	Guanajuato	450
12	Guerrero	450
13	Hidalgo	450
14	Jalisco	450
15	México	450
16	Michoacán de Ocampo	450
17	Morelos	450
18	Nayarit	450
19	Nuevo León	450
20	Oaxaca	450
21	Puebla	450
22	Querétaro	450
23	Quintana Roo	450
24	San Luis Potosí	450
25	Sinaloa	450
26	Sonora	450
27	Tabasco	450
28	Tamaulipas	450
29	Tlaxcala	450
30	Veracruz de Ignacio de la Llave	450
31	Yucatán	450
32	Zacatecas	450
Total		14 400