



Encuesta Nacional de Inclusión Financiera 2018

ENIF

Diseño muestral



COMISIÓN NACIONAL
BANCARIA Y DE VALORES



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Encuesta Nacional de Inclusión Financiera 2018

ENIF

Diseño muestral



COMISIÓN NACIONAL
BANCARIA Y DE VALORES



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Obras complementarias publicadas por el INEGI sobre el tema:

Encuesta Nacional de Inclusión Financiera 2012. ENIF. Síntesis metodológica; Encuesta Nacional de Inclusión Financiera 2015. ENIF. Documento metodológico.

Catalogación en la fuente INEGI:

640.4201 Encuesta Nacional de Inclusión Financiera.
Encuesta Nacional de Inclusión Financiera 2018 : ENIF : diseño muestral /
Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2018.

vii, 20 p.

1. Hogar - Finanzas - Encuestas - México. 2. Economía doméstica - Estadísticas - México. 3. Hogar - Encuestas - Metodología. I. Instituto Nacional de Estadística y Geografía (México). II. Comisión Nacional Bancaria y de Valores (México).

Conociendo México

01 800 111 4634

www.inegi.org.mx

atencion.usuarios@inegi.org.mx



INEGI Informa



@INEGI_INFORMA

DR © 2018, **Instituto Nacional de Estadística y Geografía**

Edificio Sede

Avenida Héroe de Nacozari Sur 2301

Fraccionamiento Jardines del Parque, 20276 Aguascalientes,

Aguascalientes, Aguascalientes, entre la calle INEGI,

Avenida del Lago y Avenida Paseo de las Garzas.

Presentación

El **Instituto Nacional de Estadística y Geografía (INEGI)**, en el marco de las atribuciones que le confiere la Ley del Sistema Nacional de Información Estadística y Geográfica, desarrolla y aplica estrategias orientadas a consolidar el Sistema Nacional de Información Estadística y Geográfica (SNIEG), entre las cuales se encuentra la documentación de los diferentes proyectos que lleva a cabo.

En tal contexto, presenta el **Diseño muestral de la Encuesta Nacional de Inclusión Financiera (ENIF) 2018**, con el propósito de ofrecer un panorama general sobre la metodología empleada para el diseño y ejecución del proyecto.

De esta forma, el **INEGI** ofrece a los usuarios un documento en el que se muestran, en forma detallada, los diferentes aspectos del diseño estadístico del proyecto.

Asimismo, a través de este tipo de documentos es posible contribuir a transparentar el proceso de generación de información estadística y disponer de un testimonio que permita fortalecer el diseño y operación de futuros proyectos en la materia.

Índice

Introducción	VII
1. Bases metodológicas	1
1.1 Objetivos	1
1.2 Referencias metodológicas	2
2. Diseño estadístico	3
2.1 Marco de la encuesta	3
2.2 Formación de las unidades primarias de muestreo (UPM)	3
2.3 Estratificación	4
2.4 Selección de las UPM de la Muestra Maestra	4
2.5 Tamaño de la muestra	5
2.6 Afijación de la muestra	5
2.7 Selección de la muestra	6
2.7.1 En urbano alto	6
2.7.2 En complemento urbano	7
2.7.3 En rural	9
2.8 Ajuste de los factores de expansión	10
2.8.1 Ajuste por no respuesta	10
2.8.2 Ajuste por proyección	11
2.9 Estimadores	11
2.10 Estimación de errores de muestreo	12
Anexo	15
A. Indicadores empleados en la estratificación de las UPM del Marco Nacional de Viviendas 2012	17
B. Distribución de la muestra en viviendas seleccionadas por entidad, según tamaño de localidad	18
C. Distribución de la muestra en viviendas seleccionadas por estrato, según tamaño de localidad	19
D. Distribución de la muestra en viviendas seleccionadas por región y entidad, según tamaño de muestra	20

Introducción

La **Encuesta Nacional de Inclusión Financiera (ENIF) 2018** es un proyecto estadístico que aborda el tema del nivel de participación de la sociedad en el aprovechamiento de los mecanismos financieros y su percepción sobre la calidad de los mismos, así como el grado de alfabetización financiera de la población. Asimismo, permite continuar con las mediciones periódicas que iniciaron en 2012 sobre el tema, con el propósito de fortalecer las políticas en materia de inclusión y educación financiera. La ENIF 2018 constituye el tercer ejercicio de este tipo que se realiza por iniciativa de la Comisión Nacional Bancaria y de Valores (CNBV).

El diseño muestral que se presenta a continuación está organizado en dos capítulos; en el primero se abordan los objetivos y bases metodológicas, mientras que el segundo está dedicado al diseño estadístico.

De esta manera, el **INEGI** da a conocer la metodología empleada y contribuye así a la transparencia del proceso de generación de información estadística.

1. Bases metodológicas

En este apartado se presentan los objetivos del proyecto y las referencias metodológicas que lo distinguen.

1.1 Objetivos

Debido a la necesidad de disponer de información sobre las características de la inclusión financiera, se establecieron los siguientes objetivos:

General

Generar información estadística e indicadores oficiales a nivel nacional y regional que permitan a las autoridades financieras hacer diagnósticos, diseñar políticas públicas y establecer metas en materia de inclusión y educación financiera.

Específicos

- Generar información a nivel nacional y regional, por tamaño de localidad y por sexo, con respecto al acceso y uso de servicios financieros, entre los que destaca el porcentaje de la población con al menos un servicio financiero formal;
- Generar datos sobre el nivel de conocimiento de la población sobre temas financieros, tales como la inflación, el interés simple y compuesto, la portabilidad de nómina, el seguro de depósitos, entre otros;
- Identificar las necesidades de la población respecto al acceso y uso de los servicios financieros, en los ámbitos nacional, regional, urbano y rural, y por sexo;
- Proveer información respecto a la frecuencia y comportamiento de uso referentes a los productos de ahorro, crédito, seguro, ahorro para el retiro, y sus canales de distribución;
- Conocer las características sociodemográficas básicas (sexo, edad, nivel de escolaridad, condición de actividad, entre otras) de los usuarios de los productos y servicios financieros, así como de aquellos que no son usuarios y quienes fueron usuarios, pero ya no lo son;
- Identificar las principales razones para no adquirir un producto o servicio financiero formal, así como su baja utilización por parte de aquellos que sí lo tienen;
- Conocer la incidencia de ciertos comportamientos financieros de la población adulta, tales como: llevar un registro de sus gastos, elaborar un presupuesto, ahorrar, manejar adecuadamente un crédito, comparar productos antes de adquirirlos, realizar aportaciones voluntarias a las Afores, utilizar medios de pago diferentes al efectivo, entre otros;
- Medir la incidencia de ciertos fenómenos como la clonación, el robo de identidad, la participación de esquemas fraudulentos, así como conocer las instituciones a las que han recurrido para levantar sus quejas o reclamos, y el tiempo de resolución de éstos;
- Contar con información para identificar la existencia de brechas de género con respecto a la toma de decisiones financieras sobre el dinero y los activos.

1.2 Referencias metodológicas

Las principales referencias metodológicas asociadas a los objetivos y metas del proyecto son las siguientes:

Población objetivo. La encuesta está dirigida a la población de 18 a 70 años de edad, que reside permanentemente en viviendas particulares dentro del territorio nacional.

Unidad de análisis. Residentes de las viviendas seleccionadas de 18 y más años, y personas seleccionadas para responder sobre los temas que se abordan a partir de la tercera sección del cuestionario de la encuesta.

Periodos de referencia: La referencia al último año se aplica en las secciones:

4. *Administración de gastos y comportamiento financiero*, pregunta 4.5.
5. *Ahorro informal y formal*, preguntas 5.1 y 5.14.
6. *Crédito informal y formal*, preguntas 6.1 y 6.10.
10. *Uso de canales financieros*, preguntas 10.1, 10.8 y 10.13.

La referencia en los últimos 3 años se aplica en la sección 11. Protección de usuarios de servicios financieros, pregunta 11.1.

La referencia al mes anterior a la entrevista se aplica en la sección 3. Características sociodemográficas de la persona elegida, las preguntas 3.5 y 3.6. En un mes sin precisar cuál, en las preguntas 4.3 y 4.4.

El resto de las variables se refieren a la situación al momento de la entrevista.

Cobertura geográfica. La encuesta está diseñada para proporcionar información con representatividad nacional, y por sexo a nivel nacional, desagregada por localidades de 15 000 y más habitantes y menores de 15 000 habitantes. También proporciona información para las regiones Noroeste, Noreste, Occidente y Bajío, Ciudad de México, Centro Sur y Oriente, y finalmente, región Sur.

Periodo de levantamiento. Del 30 de abril al 22 de junio de 2018.

Método de recolección. El método para captar la información fue a través de la aplicación de un cuestionario electrónico a un integrante de cada hogar que sea residente de alguna de las 14 500 viviendas seleccionadas.

Informante adecuado. Persona residente de la vivienda que tiene 15 años o más y conoce los datos de los integrantes de su hogar. Responde las secciones 1. Residentes y hogares en la vivienda y 2. Características sociodemográficas de los integrantes del hogar.

Informante elegido. Fue seleccionado mediante un algoritmo integrado en el programa de captura del dispositivo móvil, bajo la característica de tener entre 18 y 70 años de edad y ser residente habitual de la vivienda.

En aquellos casos en los que fue necesario aplicar el cuestionario en formato impreso, el criterio de selección del informante adecuado consistió en elegir al residente de la vivienda cuya fecha de cumpleaños fuera la inmediata posterior a la entrevista.

2. Diseño estadístico

Comprende el conjunto de actividades relacionadas con la selección de la muestra, el tamaño suficiente que permita realizar estimaciones para la población objeto de estudio, el marco de muestreo, y la construcción y evaluación de las estimaciones planteadas para la expansión de la información, a partir de los datos obtenidos en campo.

2.1 Marco de la encuesta

El diseño de la muestra para la ENIF 2018 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 la vivienda y las unidades de observación son el hogar y las personas de 18 a 70 años.

Para la selección de la muestra de la ENIF 2018, se utilizó el Marco Maestro de Muestreo 2012 del INEGI, construido a partir de la información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2010. En el marco, se seleccionó una Muestra Maestra a partir de la cual se seleccionan las submuestras 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 una segunda etapa, las viviendas que integran las muestras de las diferentes encuestas. La Muestra Maestra se construye de la siguiente manera:

2.2 Formación de las unidades primarias de muestreo (UPM)

Primeramente, se construye el conjunto de UPM que cubrirá el territorio nacional.

Las UPM 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, 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.
- 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.

¹ Área Geoestadística Básica.

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.

De esta manera, el Marco Maestro de Muestreo quedó conformado por un total de 240 912 UPM.

2.3 Estratificación

Una vez construido el conjunto de UPM, se agrupan aquellas con características similares, es decir, se estratifican, siguiendo la metodología descrita a continuación.

La división política del país y la conformación de localidades diferenciadas por su tamaño, forman de manera natural una estratificación geográfica. En cada entidad federativa se distinguen tres ámbitos, divididos a su vez en zonas, como se indica en el siguiente cuadro:

Estratificación por tamaño de localidad según ámbito geográfico

Ámbito	Zona	Descripción
Urbano alto	01 a 09	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, se formaron cuatro estratos sociodemográficos 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 34 indicadores², contruidos con información del Censo de Población 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. Como resultado, se tiene un total de 683 estratos en todo el ámbito nacional.

2.4 Selección de las UPM de la Muestra Maestra

Las UPM de la Muestra Maestra fueron seleccionadas por medio de un muestreo con probabilidad proporcional al tamaño, esto es:

$$P\{U_{ehi} \in S\} = \frac{n_{eh}m_{ehi}}{m_{eh}}$$

² La descripción de estos indicadores se presenta en el anexo A.

Donde:

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

S = la Muestra Maestra.

n_{eh} = número de UPM en 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 en el Censo de Población y Vivienda 2010.

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

Las UPM seleccionadas forman la Muestra Maestra que permite seleccionar las submuestras de las encuestas en hogares, en las siguientes etapas del diseño estadístico.

2.5 Tamaño de la muestra

El tamaño de la muestra se calculó empleando 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 = $1-p$.

r = error relativo máximo aceptable.

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

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.

tnr = tasa de no respuesta máxima esperada.

Considerando una confianza de 90%, un efecto del diseño de 3.07, un error relativo máximo esperado de 15%, una tasa de no respuesta máxima esperada de 15%, para una proporción mínima de 2.91%, se obtuvo un tamaño de muestra de 14 439 viviendas, mismo que se ajustó a 14 500 viviendas.

2.6 Afijación de la muestra

La afijación de la muestra se realizó dentro de cada entidad federativa, por tamaño de localidad y estrato de manera proporcional a su tamaño, para lo cual se empleó 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.

En los anexos B, C y D se presentan las distribuciones de la muestra por: entidad y tamaño de localidad; tamaño de localidad y estrato; y por región y entidad, respectivamente.

2.7 Selección de la muestra

La selección de la muestra se realizó de manera independiente por entidad, tamaño de localidad y estrato. El procedimiento de selección varió de acuerdo con el dominio.

2.7.1 En urbano alto

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

1. De las k_{eh} UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad para la Muestra Maestra, se seleccionaron k_{eh}^* UPM con igual probabilidad.
2. En cada UPM elegida, se seleccionaron cinco viviendas con igual probabilidad.
3. En cada vivienda elegida, se seleccionó una persona de 18 a 70 años con igual probabilidad.

La probabilidad de selección de las viviendas y personas se calculó de la siguiente manera:

- a. La probabilidad de selección de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra maestra está definida por:

$$P_{1ehi} = \frac{k_{eh} m_{ehi}}{m_{eh}}$$

- b. La probabilidad de seleccionar de la muestra maestra a la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra está definida por:

$$P_{2ehi} = \frac{k_{eh}^*}{k_{eh}}$$

- c. La probabilidad de seleccionar la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad está definida por:

$$P_{3ehij} = \frac{5}{m_{ehi}^*},$$

- d. 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:

$$P_{4ehij} = \frac{1}{Q_{ehij}},$$

Por lo tanto, la probabilidad total de selección de una persona en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad es el producto de las probabilidades de selección:

$$P_{ehij} = P_{1ehi} * P_{2ehi} * P_{3ehij} * P_{4ehij}$$

$$= \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{k_{eh}^*}{k_{eh}} \frac{5}{m_{ehi}^*} \frac{1}{Q_{ehij}} = \frac{5 k_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehij}}$$

Su factor de expansión está dado por:

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

Donde:

- k_{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.
- k_{eh}^* = número de UPM a seleccionar con igual probabilidad de selección, en el h-ésimo estrato, en la e-ésima entidad.
- Q_{ehij} = número de personas de 18 a 70 años en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

2.7.2 En complemento urbano

En el ámbito complemento urbano la selección de la muestra se realizó en forma independiente por cada entidad y estrato mediante el siguiente procedimiento:

1. De las k_{eh} UPM que integran la Muestra Maestra, se seleccionaron k_{eh}^* UPM con igual probabilidad.
2. En cada UPM elegida, se seleccionaron 20 viviendas con igual probabilidad.
3. En cada vivienda elegida, se seleccionó una persona de 18 a 70 años con igual probabilidad.

La probabilidad de selección de las viviendas y personas se calculó de la siguiente manera:

- a. La probabilidad de selección de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la Muestra Maestra está definida por:

$$P_{1ehi} = \frac{k_{eh} m_{ehi}}{m_{eh}}$$

- b. La probabilidad de seleccionar de la Muestra Maestra a la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra de ENIF-2018 está definida por:

$$P_{2ehi} = \frac{k_{eh}^*}{k_{eh}}$$

- c. La probabilidad de seleccionar la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad está definida por:

$$P_{3ehij} = \frac{20}{m_{ehi}^*},$$

- d. 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:

$$P_{4ehij} = \frac{1}{Q_{ehij}},$$

Por lo tanto, la probabilidad total de selección de una persona en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad es el producto de las probabilidades de selección:

$$\begin{aligned} P_{ehij} &= P_{1ehi} * P_{2ehi} * P_{3ehij} * P_{4ehij} \\ &= \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{k_{eh}^*}{k_{eh}} \frac{20}{m_{ehi}^*} \frac{1}{Q_{ehij}} = \frac{20 k_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehij}} \end{aligned}$$

Su factor de expansión está dado por:

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

Donde:

- k_{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.
- k_{eh}^* = número de UPM a seleccionar con igual probabilidad de selección, en el h-ésimo estrato, en la e-ésima entidad.
- Q_{ehij} = número de personas de 18 a 70 años en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

2.7.3 En rural

En el ámbito rural la selección de la muestra se realizó en forma independiente por cada entidad y estrato mediante el siguiente procedimiento:

1. De las k_{eh} UPM se seleccionaron k_{eh}^* UPM con igual probabilidad.
2. En cada UPM elegida, se seleccionaron cuatro segmentos de aproximadamente cinco viviendas cada uno, con igual probabilidad.
3. En cada vivienda elegida, se seleccionó una persona de 18 a 70 años con igual probabilidad.

La probabilidad de selección de las viviendas y personas se calculó de la siguiente manera:

- a. La probabilidad de selección de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad para la Muestra Maestra está definida por:

$$P_{1ehi} = \frac{k_{eh} m_{ehi}}{m_{eh}}$$

- b. La probabilidad de seleccionar de la Muestra Maestra a la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad para la muestra de ENIF-2018 está definida por:

$$P_{2ehi} = \frac{k_{eh}^*}{k_{eh}}$$

- c. La probabilidad de seleccionar la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad está definida por:

$$P_{3ehij} = \frac{4 \cdot 5}{m_{ehi}^*},$$

- d. 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:

$$P_{4ehij} = \frac{1}{Q_{ehij}},$$

Por lo tanto, la probabilidad total de selección de una persona en la j -ésima vivienda, en la i -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad es el producto de las probabilidades de selección:

$$\begin{aligned} P_{ehij} &= P_{1ehi} * P_{2ehi} * P_{3ehij} * P_{4ehij} \\ &= \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{k_{eh}^*}{k_{eh}} \frac{20}{m_{ehi}^*} \frac{1}{Q_{ehij}} = \frac{20 k_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehij}} \end{aligned}$$

Su factor de expansión³ está dado por:

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

Donde:

- k_{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.
- k_{eh}^* = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad.
- Q_{ehij} = número de personas de 18 a 70 años en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

2.8 Ajuste de los factores de expansión

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

2.8.1 Ajuste por no respuesta

El ajuste por no respuesta se realizó a nivel de las viviendas, los hogares y las personas seleccionadas en cada UPM, de cada uno de los dominios, mediante las siguientes expresiones:

AJUSTE POR NO RESPUESTA A NIVEL VIVIENDA

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

$$F'_{ehij} = F_{ehij} \frac{\sum_{i \in h} \sum_{j \in i} F_{ehij}}{\sum_{i \in h} \sum_{j \in i} F_{ehij} I_{ehij}}$$

Donde:

- F'_{ehij} = factor de expansión corregido por no respuesta de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- F_{ehij} = factor de expansión de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- I_{ehij} = función indicadora con valor 1 o 0. Si la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad tiene respuesta, entonces esta función tomará el valor 1; y tomará el valor 0 en caso contrario.

³ El factor de expansión está definido como el inverso de la probabilidad de selección.

AJUSTE POR NO RESPUESTA A NIVEL PERSONA

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

$$F''_{ehijk} = F'_{ehijk} \frac{\sum_{i \in h} \sum_{j \in i} \sum_{k \in j} F'_{ehijk}}{\sum_{i \in h} \sum_{j \in i} \sum_{k \in j} F'_{ehijk} I_{ehijk}}$$

Donde:

- F''_{ehijk} = factor de expansión corregido por no respuesta de la k-ésima vivienda seleccionada, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- F'_{ehijk} = factor de expansión para la k-ésima vivienda seleccionada, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- I_{ehijk} = función indicadora con valor 1 o 0. Si la k-ésima persona seleccionada, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad tiene respuesta, entonces esta función tomará el valor 1; y tomará el valor 0 en caso contrario.

2.8.2 Ajuste por proyección

Los factores ajustados por la no respuesta se corrigieron, con el 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 el CONAPO, 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 de la subpoblación en que se hace el ajuste por proyección.

2.9 Estimadores

El estimador del total de la característica X es:

$$\hat{X} = \sum_e \sum_h \sum_i \left(\sum_s F_{ehis}^{15+} + \sum_{\ell} X_{ehis\ell}^{15+} \right) + \sum_e \sum_h \sum_i \left(\sum_s F_{ehis}^{15-} - \sum_{\ell} X_{ehis\ell}^{15-} \right)$$

Donde:

F_{ehis}^{15+} = factor de expansión final, de la s-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en las localidades de 15 000 y más habitantes.

X_{ehis}^{15+} = 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 las localidades de 15 000 y más habitantes.

F_{ehis}^{15-} = factor de expansión final de la s-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, en las localidades menores de 15 000 habitantes.

X_{ehis}^{15-} = valor observado de la característica X de 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 las localidades menores de 15 000 habitantes.

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} .

2.10 Estimación de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones, 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 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} :

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

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 característica 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:

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

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

Las estimaciones del error estándar (EE), efecto de diseño (DEFF) y coeficiente de variación (CV) se calculan mediante las siguientes expresiones:

$$EE = \sqrt{\hat{V}(\hat{\theta})} \quad DEFF = \frac{\hat{V}(\hat{\theta})}{\hat{V}(\hat{\theta})_{MAS}} \quad CV = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}}$$

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- α)%, 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.

Anexo

A. Indicadores empleados en la estratificación de las UPM del Marco Nacional de Viviendas 2012

Mnemónico	Descripción
Proporción de 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 o más años de edad que aprobaron algún grado de escolaridad diferente al nivel preescolar.
PP15PRI_CO	de 15 o más años de edad que tienen como máxima escolaridad 6 grados aprobados en primaria.
PP15SEC_CO	de 15 o más años de edad que tienen como máxima escolaridad 3 grados aprobados en secundaria.
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.
TOCU12A17	no ocupada de 12 a 17 años entre la población total de este rango.
PPOMAYED	ocupada de 18 años y más entre la población total ocupada.
PGDO_ESC	Grado promedio de escolaridad.
PTASAOcupa	Tasa de ocupación.
Proporción de 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.
PVPH3YMASD	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 ella, 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.
PVDRERED	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.
Proporción de 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.
PVPHCBIEN	todos los bienes.

B. Distribución de la muestra en viviendas seleccionadas por entidad, según tamaño de localidad

Entidad federativa	Total	Tamaño de localidad	
		De 15 000 y más habitantes	Menos de 15 000 habitantes
Nacional	14 537^A	9 305	5 232
Aguascalientes	335	250	85
Baja California	453	370	83
Baja California Sur	450	330	120
Campeche	385	220	165
Coahuila de Zaragoza	674	575	99
Colima	340	240	100
Chiapas	386	130	256
Chihuahua	451	350	101
Ciudad de México	999	970	29
Durango	450	270	180
Guanajuato	335	195	140
Guerrero	391	165	226
Hidalgo	442	130	312
Jalisco	335	255	80
México	500	380	120
Michoacán de Ocampo	340	160	180
Morelos	444	280	164
Nayarit	345	180	165
Nuevo León	677	615	62
Oaxaca	385	95	290
Puebla	441	230	211
Querétaro	336	165	171
Quintana Roo	385	325	60
San Luis Potosí	678	375	303
Sinaloa	452	270	182
Sonora	450	350	100
Tabasco	385	140	245
Tamaulipas	677	555	122
Tlaxcala	440	165	275
Veracruz de Ignacio de la Llave	443	190	253
Yucatán	386	240	146
Zacatecas	347	140	207

^A De una muestra nacional de 14 500 viviendas seleccionadas, al cierre del operativo de campo se captó información de 14 537. Tal diferencia se debe a que en el dominio rural se seleccionan segmentos conformados por conjuntos de cuatro a seis viviendas cercanas entre sí, de las que se obtiene información, de tal manera que se recolectan datos de un número de viviendas que puede ser marginalmente inferior o superior al calculado, el cual no influye en las precisiones estadísticas.

C. Distribución de la muestra en viviendas seleccionadas por estrato, según tamaño de localidad

Estrato	Total	Tamaño de localidad	
		De 15 000 y más habitantes	Menos de 15 000 habitantes
Total	14 537	9 305	5 232
Bajo	2 518	130	2 388
Medio bajo	7 299	4 525	2 774
Medio alto	3 300	3 250	50
Alto	1 420	1 400	20

D. Distribución de la muestra en viviendas seleccionadas por región y entidad, según tamaño de muestra

Región		Entidad federativa		Muestra
Clave	Nombre	Clave	Nombre	
Total				14 537
1	Noroeste	Total		2 706
		02	Baja California	453
		03	Baja California Sur	450
		08	Chihuahua	451
		10	Durango	450
		25	Sinaloa	452
		26	Sonora	450
2	Noreste	Total		2 706
		05	Coahuila de Zaragoza	674
		19	Nuevo León	677
		24	San Luis Potosí	678
		28	Tamaulipas	677
3	Occidente y Bajío	Total		2 713
		01	Aguascalientes	335
		06	Colima	335
		11	Guanajuato	335
		14	Jalisco	340
		16	Michoacán de Ocampo	345
		18	Nayarit	336
		22	Querétaro	347
		32	Zacatecas	340
4	Ciudad de México	Total		999
		09	Ciudad de México	999
5	Centro, Sur y Oriente	Total		2 710
		13	Hidalgo	500
		15	México	442
		17	Morelos	444
		21	Puebla	441
		29	Tlaxcala	440
		30	Veracruz de Ignacio de la Llave	443
6	Sur	Total		2 703
		04	Campeche	385
		07	Chiapas	386
		12	Guerrero	391
		20	Oaxaca	385
		23	Quintana Roo	385
		27	Tabasco	386
		31	Yucatán	385