

ENFIH

Encuesta Nacional sobre las Finanzas de los Hogares

2019

Diseño muestral



INEGI

DR © 2020, **Instituto Nacional de Estadística y Geografía**

Edificio Sede

Av. 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.

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

Diseño muestral de la Encuesta Nacional sobre las Finanzas de los Hogares (ENFIH) 2019

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 sobre las Finanzas de los Hogares (ENFIH) 2019**, 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 la operación de futuros proyectos en la materia.

Índice

Introducción	VII
1. Bases metodológicas	1
1.1 Objetivos	1
1.2 Referencias metodológicas	1
2. Diseño estadístico	3
2.1 Marco de la encuesta	3
2.2 Formación de las 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	8
2.8 Ajuste a los factores de expansión	9
2.8.1 Ajuste por no respuesta	9
2.8.2 Ajuste por proyección	10
2.9 Estimadores	10
2.10 Estimación de errores de muestreo	11
2.11 Homologación de la semaforización para los umbrales de indicadores de precisión estadística	12
Anexo	13
A. Indicadores empleados en la estratificación de la Muestra Maestra	15
B. Distribución de la muestra en viviendas seleccionadas por entidad, según tamaño de localidad	16
C. Distribución de la muestra en viviendas seleccionadas por estrato, según tamaño de localidad	17

Introducción

La **Encuesta Nacional sobre las Finanzas de los Hogares (ENFIH) 2019** es un ejercicio estadístico desarrollado por el **INEGI** y **Banco de México (Banxico)**, a iniciativa de este último, que proporcionará información económico-financiera del hogar, tenencia, composición y valor de activos y pasivos, ingresos y egresos, así como acceso de los hogares a mercados financieros; lo que contribuye de forma significativa a profundizar en el conocimiento de la evolución de los activos y pasivos de los hogares mexicanos.

El diseño muestral que se presenta a continuación está organizado en dos capítulos; en el primero, se abordan los objetivos y las 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

Con base en la necesidad de profundizar en el tema, se establecieron los siguientes objetivos:

General

Generar información estadística relevante sobre las finanzas de los hogares en México, en particular sobre los flujos y los acervos de sus activos y pasivos.

Específicos

- Contar con información cuantitativa sobre el balance financiero de los hogares en México.
- Proporcionar datos que permitan satisfacer las necesidades de información que existen sobre el endeudamiento y capacidad de pago de los hogares.
- Generar información que permita analizar y profundizar en el conocimiento sobre la evolución y composición de los activos y pasivos que tienen los hogares.
- Complementar la información estadística existente sobre los pasivos y activos de los hogares que proviene de otras fuentes de información.

1.2 Referencias metodológicas

Unidad de selección. Vivienda particular.

Unidad de observación. Las viviendas particulares seleccionadas, los hogares que en ellas residan y los integrantes del hogar de 18 años y más.

Periodo de levantamiento. Del 7 de octubre al 29 de noviembre de 2019.

Cobertura geográfica. La muestra se encuentra distribuida en todo el país, tanto en localidades urbanas como rurales, de tal manera que es posible generar información a nivel nacional y por tamaño de localidad.

Método de recolección. Se procedió a recolectar la información de forma transversal, cara a cara con el informante adecuado, el administrador y cada persona de 18 o más años, integrante del hogar, a través de un cuestionario electrónico en mini *laptop*. Solamente en caso necesario se utilizó cuestionario impreso para, posteriormente, capturar la información.

Periodo de referencia. La mayor parte de las preguntas se refieren a la situación al momento de la entrevista. Fecha de apertura de crédito; semana anterior a la entrevista respecto a la condición y posición en el trabajo; mes anterior a la entrevista en cantidades pagadas del crédito hipotecario y dinero ahorrado; y en cuanto a solicitud de crédito, mensualidades vencidas e ingresos diferentes al trabajo, de septiembre de 2018 a la fecha de la entrevista.

Informantes:

Adecuado. Persona residente en la vivienda, de 18 o más años de edad, que conoce los datos de los integrantes de su hogar. Responde las secciones *1. Residentes y hogares en la vivienda* y *2. Características sociodemográficas*.

Administrador(a). Persona de 18 o más años de edad, que es quien más conoce sobre el valor y características de la vivienda, el crédito hipotecario, los costos de construcción, o bien, de lo que pagan de renta. Responde la información de las secciones *3. Características generales de la vivienda*, y *4. Vivienda y deuda hipotecaria*.

Informante adecuado de la temática de finanzas. Cada una de las personas del hogar de 18 años y más, las cuales proporcionarán la información de la sección *4a. Características personales* en adelante, relativo a los activos y pasivos en la entrevista individual.

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 ENFIH 2019 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 la unidad de observación es el hogar.

Para la selección de la muestra se utilizó el Marco Nacional de Viviendas 2012 del INEGI, construido a partir de la información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2010. Esta muestra es 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 (UPM), pues es en ellos donde se seleccionan, en una segunda etapa, las viviendas que integran las muestras de las diferentes encuestas.

2.2 Formación de las UPM

En principio, se construye el conjunto de UPM que cubren 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 AGE¹.
- La unión de dos o más manzanas contiguas de diferentes AGE¹ 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 AGE¹.
- La unión de dos o más manzanas contiguas de diferentes AGE¹ de la misma localidad.
- La unión de dos o más manzanas contiguas de diferentes AGE¹ 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, la Muestra Maestra quedó conformada por un total de 240 912 UPM a nivel nacional.

2.3 Estratificación

Una vez construido el conjunto de UPM, se agruparon aquellas con características similares de acuerdo a lo siguiente:

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 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 construidos 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 obtuvo un total de 683 estratos en todo el territorio 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:

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

Donde:

U_{ehi} = UPM i -ésima, en el h -ésimo estrato, en 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.

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

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

La expresión para calcular el tamaño de muestra fue:

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

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 de diseño de 3.24, un error relativo máximo esperado de 15%, una tasa de no respuesta máxima esperada de 15%, para una proporción mínima de 1.96%, se obtuvo un tamaño de muestra de 22 931 viviendas, mismo que se ajustó a 23 000 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 el anexo B se presenta la distribución de la muestra de viviendas por entidad según tamaño de localidad, mientras que el C contiene la distribución de la muestra de viviendas según tamaño de localidad por estrato.

2.7 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

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

La probabilidad de selección de las viviendas y las 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 en la Muestra Maestra a la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra de ENFIH 2019 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 dada 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 de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de 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 expansión³ está dado por:

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

³ El factor de expansión se define como el inverso de la probabilidad de selección. En la *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, el factor de expansión se nombra Ponderador.

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 cifras del 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 y más 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

1. De las k_{eh} UPM que integran la Muestra Maestra, se eligieron k_{eh}^* UPM con igual probabilidad.
2. En cada UPM seleccionada, se eligieron 20 viviendas con igual probabilidad.
3. En cada vivienda seleccionada, se eligió una persona de 18 o más 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 en la Muestra Maestra a la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra de la ENFIH 2019 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 dada 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 de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es el producto de las probabilidades de selección:

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

Su factor de expansión está dado por:

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

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 para la ENFIH 2019 con igual probabilidad de selección, en el h-ésimo estrato, en la e-ésima entidad.
- Q_{ehij} = número de personas de 18 y más 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

1. De las k_{eh} UPM, se seleccionaron k_{eh}^* UPM con igual probabilidad.
2. En cada UPM seleccionada, se eligieron cuatro segmentos de cinco viviendas aproximadamente, con igual probabilidad.
3. En cada vivienda seleccionada, se eligió una persona de 18 o más años con igual probabilidad.

La probabilidad de selección de 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 en la Muestra Maestra a la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra de la ENFIH 2019 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 dada por:

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

- 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 de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es el producto de las probabilidades de selección:

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

Su factor de expansión está dado por:

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

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 cifras del 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 para la ENIF 2019, en el h-ésimo estrato, en la e-ésima entidad.
- Q_{ehij} = número de personas de 18 y más años en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

2.8 Ajuste a 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

Se realizó a nivel de las viviendas, los hogares y las personas seleccionadas a nivel UPM, en cada uno de los dominios a nivel vivienda y a nivel persona.

A NIVEL VIVIENDA

El ajuste por no respuesta atribuida a la vivienda 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; esta función toma el valor 1 si la j-ésima vivienda seleccionada, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad tiene respuesta completa, y toma el valor de cero en caso contrario.

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 persona 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 persona 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; esta función toma el valor 1 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 completa, y toma el valor de cero en caso contrario.

2.8.2 Ajuste por proyección

Los factores de expansión ajustados por la no respuesta se corrigieron, con el propósito de asegurar que en cada dominio de interés de la encuesta se obtenga la población total determinada por la proyección generada por el Consejo Nacional de Población (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:

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

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.
- $X_{ehis,\lambda}^{15+}$ = valor observado de la característica de interés X en la λ -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.
- $X_{ehis,\lambda}^{15-}$ = valor observado de la característica de interés X en la λ -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, la variable \hat{Y} es definida 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 últimos⁴, 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 UPM.

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} :

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

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.

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:

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

Las estimaciones del error estándar (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 construye 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.

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

2.11 Homologación de la semaforización para los umbrales de indicadores de precisión estadística

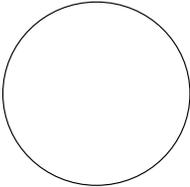
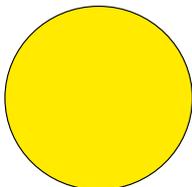
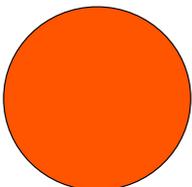
Para facilitar la interpretación de las precisiones estadísticas de la información pública en tabulados, el Comité de Aseguramiento de la Calidad, en la cuarta sección celebrada el 1 de noviembre de 2018, aprobaron los siguientes umbrales y especificaciones para los coeficientes de variación en la publicación de tabulados, así como la semaforización de estos.

Umbrales aprobados para la precisión de las estimaciones

Interpretación	Semaforización	Nivel de precisión
Alta	Blanco	[0%, 15%)
Moderada	Amarillo	[15%, 30%)
Baja	Naranja oscuro	>=30%

A partir del segundo trimestre de 2018, se publican los siguientes indicadores de precisión estadística en la presentación de resultados en tabulados de todas las encuestas con muestreo probabilístico del INEGI: error estándar, intervalo de confianza y coeficiente de variación. Adicionalmente, se estandariza la coloración en los tabulados para indicar el nivel de precisión de las estimaciones con base en el CV. A continuación, se presenta el código RGB (sigla en inglés de *red, green, blue*; en español: rojo, verde y azul) de los colores utilizados en la semaforización.

Parámetros RGB para la semaforización del CV

Código	Interpretación		
	Alta	Moderada	Baja
			
Rojo	255	255	255
Verde	255	234	84
Azul	255	0	0

El siguiente texto explicativo aparece en cada uno de los tabulados publicados de encuestas por muestreo probabilístico.

Las estimaciones que aparecen en este cuadro están coloreadas, de acuerdo con su nivel de precisión, en alta, moderada y baja, tomando como referencia el coeficiente de variación CV (%). Una precisión baja requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.

Anexo

A. Indicadores empleados en la estratificación de la Muestra Maestra

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.
PVDRENER	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.
PVPHCBIE	todos los bienes.

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

Entidad federativa	Total	Tamaño de localidad	
		15 000 y más habitantes	Menos de 15 000 habitantes
Nacional	23 041^A	15 980	7 061
01 Aguascalientes	711	585	126
02 Baja California	721	620	101
03 Baja California Sur	739	555	184
04 Campeche	698	495	203
05 Coahuila de Zaragoza	710	630	80
06 Colima	725	585	140
07 Chiapas	702	385	317
08 Chihuahua	731	585	146
09 Ciudad de México	725	700	25
10 Durango	710	490	220
11 Guanajuato	715	515	200
12 Guerrero	707	445	262
13 Hidalgo	735	290	445
14 Jalisco	720	580	140
15 México	722	575	147
16 Michoacán de Ocampo	726	440	286
17 Morelos	720	470	250
18 Nayarit	732	430	302
19 Nuevo León	719	680	39
20 Oaxaca	723	215	508
21 Puebla	725	450	275
22 Querétaro	721	460	261
23 Quintana Roo	726	600	126
24 San Luis Potosí	726	485	241
25 Sinaloa	712	510	202
26 Sonora	717	595	122
27 Tabasco	720	390	330
28 Tamaulipas	720	620	100
29 Tlaxcala	709	265	444
30 Veracruz de Ignacio de la Llave	746	415	331
31 Yucatán	710	525	185
32 Zacatecas	718	395	323

^A El tamaño de la muestra calculada fue de 23 000 viviendas; sin embargo, la selección final fue de 23 041. Tal diferencia se debe a que en el dominio rural, en cada UPM, se seleccionan segmentos conformados por conjuntos de cuatro a seis viviendas cercanas entre sí hasta completar 20 viviendas o un número cercano a 20, de tal manera que el número final de viviendas seleccionadas puede ser marginalmente inferior o superior al calculado, esta diferencia 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	
		15 000 y más habitantes	Menos de 15 000 habitantes
Total	23 041	15 980	7 061
1 Bajo	3 289	315	2 974
2 Medio bajo	8 592	4 940	3 652
3 Medio alto	5 605	5 445	160
4 Alto	5 555	5 280	275