

ENIF

Encuesta Nacional de Inclusión Financiera

2021

Diseño muestral



 **INEGI**

Instituto Nacional de Estadística y Geografía

**Encuesta Nacional
de Inclusión Financiera
2021**

ENIF

Diseño muestral



Obras complementarias publicadas por el INEGI sobre el tema:
Encuesta Nacional de Inclusión Financiera 2018. ENIF. Diseño muestral.

Catalogación en la fuente INEGI:

640.4201 Encuesta Nacional de Inclusión Financiera.
Encuesta Nacional de Inclusión Financiera 2021 : ENIF : diseño muestral /
Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2021.

viii, 19 p.

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

Conociendo México

800 111 4634

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

 **INEGI Informa**  **@INEGI_INFORMA**

Registro en trámite

2021, **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) 2021**, 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	2
2. Diseño estadístico	5
2.1 Marco de la encuesta	5
2.2 Formación de las UPM	5
2.3 Estratificación	6
2.4 Tamaño de la muestra	6
2.5 Distribución de la muestra	7
2.6 Selección de la muestra	7
2.6.1 En urbano alto	7
2.6.2 En complemento urbano	8
2.6.3 En rural	8
2.7 Ajuste a los factores de expansión	9
2.8 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	15
A. Indicadores empleados en la estratificación de la Muestra Maestra	17
B. Distribución de la muestra de viviendas seleccionadas por entidad y tamaño de localidad	18
C. Distribución de la muestra en viviendas seleccionadas por estrato, según tamaño de localidad	19

Introducción

La **Encuesta Nacional de Inclusión Financiera (ENIF) 2021** es un proyecto desarrollado por el **INEGI** en coordinación con la **Comisión Nacional Bancaria y de Valores (CNBV)**, con el propósito de generar información estadística sobre el acceso y uso de servicios financieros entre la población mexicana, así como de la educación financiera y del conocimiento de los mecanismos de protección al consumidor, que tiene la población de 18 años y más.

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

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; asimismo, incorporar cambios y actualizaciones requeridas para dar atención a nuevos requerimientos de información y consideraciones en la Política Nacional de Inclusión Financiera (PNIF).

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, cálculo del interés simple y compuesto, la portabilidad de nómina, el seguro de depósitos, entre otros.
- Identificar las necesidades de la población respecto del 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, nivel socioeconómico, hablantes indígenas, entre otras) de los usuarios de los productos y servicios financieros, así como aquellos que son beneficiarios de programas sociales, no son usuarios y quienes fueron usuarios, pero ya no lo son.
- Identificar las principales barreras para no adquirir un producto o servicio financiero formal, y su baja utilización para aquellos que sí lo tienen. También, entender las razones de las personas que dejaron de tener algún producto financiero, así como el conocimiento de la portabilidad de los productos financieros.
- Conocer la incidencia de ciertos comportamientos financieros de la población adulta, tales como: llevar un registro de los gastos, elaborar un presupuesto, ahorrar, manejar adecuadamente un crédito, comparar productos antes de adquirirlos, realizar aportaciones voluntarias a las Administradoras de Fondos para el Retiro (Afores), utilizar medios de pago diferentes al efectivo, entre otros.
- Medir la incidencia de ciertos fraudes como la clonación, el robo de identidad y la participación en esquemas fraudulentos.
- 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.
- Obtener información sobre la vulnerabilidad financiera y los mecanismos de resiliencia, especialmente durante y posterior a la pandemia por COVID-19.

- Evaluar las actitudes financieras de las personas, tales como comparar productos antes de adquirirlos, conocer su comportamiento de pagos, saber cómo piensan solventar su vejez y cómo hacen frente a sus emergencias económicas.
- Contar con la información necesaria para poder obtener el indicador de alfabetización financiera propuesto por la Red Internacional para la Educación Financiera de la Organización para la Cooperación y el Desarrollo Económicos (OECD/INFE por sus siglas en inglés), el cual se compone de tres subíndices: conocimiento financiero, comportamiento financiero y actitudes financieras, así como del índice de bienestar financiero.

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 y más 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 mayor parte de las variables se refieren a la situación al momento de la entrevista; sin embargo, existen diferentes temporalidades a lo largo de la entrevista, como se muestra a continuación.

Periodos de referencia por sección temática

Sección temática	Periodos de referencia en las preguntas					
	Última vez	Alguna vez	Mes pasado	Últimos tres meses	Último año	Últimos tres años
3 Características sociodemográficas de la persona elegida			3.5 y 3.6			
4 Actitudes, comportamientos, vulnerabilidad y bienestar financieros	4.4				4.3	
5 Ahorro informal y formal					5.1 a 5.3, 5.7, 5.9 y 5.10	
6 Crédito informal y formal		6.14 y 6.17			6.1 y 6.4	
7 Pagos				7.7	7.5	
8 Seguros		8.3 y 8.10				
10 Uso de canales financieros					10.1, 10.4 y 10.7	
11 Confianza y protección de personas usuarias de servicios financieros						11.2

Cobertura geográfica. La encuesta está diseñada para proporcionar información con representatividad nacional, desagregada por localidades de 15 000 y más habitantes y menores de 15 000 habitantes. También proporciona información para las regiones que se detallan en el diseño estadístico.

Periodo de levantamiento. Del 28 de junio al 13 de agosto de 2021.

Método de recolección. Entrevista directa a persona residente de la vivienda que tiene 15 años o más que conozca la información de la vivienda, por medio de un cuestionario electrónico, operado mediante un dispositivo móvil.

Informante adecuado. Persona residente de la vivienda que tiene 15 años o más, que conoce los datos sociodemográficos de todas las personas integrantes de su hogar. Responde las secciones 0. *Características de la vivienda*, 1. *Residentes y hogares en la vivienda*, y 2. *Características sociodemográficas de las personas del hogar*.

Persona elegida. Para las entrevistas que se levantan en papel, la persona elegida es la integrante del hogar de 18 años y más, cuya fecha de cumpleaños fue la inmediata posterior a la fecha de la entrevista. En el caso de las entrevistas realizadas con el dispositivo electrónico, la persona elegida fue seleccionada aleatoriamente mediante un algoritmo integrado en el programa de captura.

La persona elegida proporcionó información sobre sus características sociodemográficas e inclusión financiera.

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 2021 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 son las personas de 18 y más años.

El marco de muestreo que se utilizó fue 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. 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; 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

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.

2.3 Estratificación

La división política del país y la conformación de localidades diferenciadas por su tamaño forman, de manera natural, una primera estratificación geográfica. En cada entidad federativa se distinguen tres ámbitos, divididos a su vez en 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 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 asignada a su estrato geográfico entidad-ámbito-zona.

2.4 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.

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

Considerando una confianza de 90% y un efecto de diseño de 3.05, un error relativo máximo esperado de 15%, una tasa de no respuesta máxima esperada de 15%, para una proporción mínima de 2.75%, se obtuvo un tamaño de muestra de 15 259 viviendas, misma que se ajustó a 15 291.

2.5 Distribución de la muestra

La distribución 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 por entidad y tamaño de localidad; mientras que en el anexo C, se describe la distribución de la muestra por tamaño de localidad y estrato.

2.6 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.

2.6.1 En urbano alto

1. Se seleccionaron k_{eh} UPM, con probabilidad proporcional al número de viviendas del estrato.
2. En cada UPM seleccionada, se eligieron cinco viviendas con igual probabilidad.
3. Al interior de cada vivienda, se seleccionó de manera aleatoria a una persona de 18 años o más, que fuera habitante permanente de la vivienda.

Por lo tanto, la probabilidad de seleccionar la k-ésima persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, es:

$$P_{ehijk} = \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{5}{m_{ehi}^*} \frac{1}{q_{ehij}} = \frac{5 k_{eh} m_{ehi}}{m_{eh} m_{ehi}^* q_{ehij}}$$

Su factor de expansión está dado por:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad, para el marco de la Muestra Maestra.
- 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_{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 listado de viviendas actualizado.
- q_{ehij} = número de personas de 18 años o más encontradas en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

2.6.2 En complemento urbano

1. Se seleccionaron k_{eh} UPM, con probabilidad proporcional al total de viviendas del estrato.
2. En cada UPM seleccionada, se eligieron 20 viviendas con igual probabilidad.
3. Al interior de cada vivienda, se seleccionó de manera aleatoria a una persona de 18 años o más, que fuera habitante permanente de la vivienda.

Por lo tanto, la probabilidad de seleccionar la k-ésima persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, es:

$$P_{ehijk} = \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{20}{m_{ehi}^*} \frac{1}{q_{ehij}} = \frac{20 k_{eh} m_{ehi}}{m_{eh} m_{ehi}^* q_{ehij}}$$

Su factor de expansión está dado por:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad, para el marco de 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.
- q_{ehij} = número de personas de 18 años o más encontradas en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

2.6.3 En rural

1. Se seleccionaron k_{eh} UPM, con probabilidad proporcional al total de viviendas del estrato.
2. En cada UPM seleccionada, se eligieron cuatro segmentos de cinco viviendas, aproximadamente, con igual probabilidad.
3. Al interior de cada vivienda, se seleccionó de manera aleatoria a una persona de 18 años o más, que fuera habitante permanente de la vivienda.

Por lo tanto, la probabilidad de seleccionar la k -ésima persona, de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, es:

$$P_{ehijk} = \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{4 \cdot 5}{m_{ehi}^*} \frac{1}{q_{ehij}} = \frac{20 k_{eh} m_{ehi}}{m_{eh} m_{ehi}^* q_{ehij}}$$

Su factor de expansión está dado por:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h -ésimo estrato, en la e -ésima entidad, para el marco de 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.
- q_{ehij} = número de personas de 18 años o más encontradas en la j -ésima vivienda, en la i -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad.

2.7 Ajuste a los factores de expansión

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

Ajuste por no respuesta de las viviendas

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

Donde:

- F'_{ehi} = factor de expansión corregido por no respuesta para las viviendas de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad.
- nv_{ehi} = número de viviendas habitadas seleccionadas en la i -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad.
- $nvhcR_{ehi}$ = número de viviendas habitadas seleccionadas con respuesta en la i -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad.
- F_{ehi} = factor de expansión de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad.

Ajuste por no respuesta de las personas

$$F_{ehijk}^* = F_{ehijk} \frac{q_{ehi}}{q_{ehi}^*}$$

Donde:

- F_{ehijk}^* = factor de expansión corregido por no respuesta para la k-ésima persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- q_{ehi} = número de personas de 18 años o más seleccionadas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.
- q_{ehi}^* = número de personas de 18 años o más seleccionadas con respuesta en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

2.8 Ajuste por proyección

Los factores de expansión ajustados por la no respuesta se corrigieron, a fin de asegurar que en cada dominio de interés de la encuesta se obtenga la población total determinada por la estimación 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 de la subpoblación en que se realiza 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.

$X_{ehis\ell}^{15+}$ = 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.

$X_{ehis\ell}^{15-}$ = valor observado de la característica de interés 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, 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 Últimos*³, basado en que la mayor contribución a la varianza de un estimador, en un diseño trietápico, 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:

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

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

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

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

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

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

Donde α es el nivel de significancia.

2.11 Homologación de la semaforización para los umbrales de indicadores de precisión estadística⁴

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, aprobó los siguientes umbrales y especificaciones para los coeficientes de variación en la publicación de tabulados, así como la semaforización de estos.

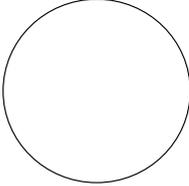
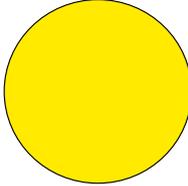
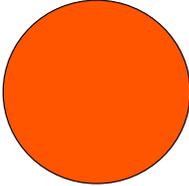
Umbrales aprobados para la precisión de las estimaciones

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.

⁴ La fuente de esta información está basada en el documento del Comité de Aseguramiento de la Calidad, depositado en el siguiente sitio http://intranet.inegi.org.mx/calidad/wp-content/uploads/2017/02/Homologacion_de_umbrales.pdf.

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 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 de viviendas seleccionadas por entidad y tamaño de localidad

Entidad federativa	Total	Tamaño de localidad	
		De 15 000 y más habitantes	Menos de 15 000 habitantes
Nacional	15 291	9 800	5 491
Aguascalientes	355	250	105
Baja California	475	395	80
Baja California Sur	476	355	121
Campeche	405	240	165
Coahuila de Zaragoza	711	610	101
Colima	357	255	102
Chiapas	412	150	262
Chihuahua	476	355	121
Ciudad de México	1 050	1 020	30
Durango	477	275	202
Guanajuato	355	195	160
Guerrero	409	185	224
Hidalgo	463	145	318
Jalisco	358	275	83
México	526	385	141
Michoacán de Ocampo	362	180	182
Morelos	463	280	183
Nayarit	356	195	161
Nuevo León	710	650	60
Oaxaca	407	95	312
Puebla	463	230	233
Querétaro	358	185	173
Quintana Roo	407	340	67
San Luis Potosí	711	390	321
Sinaloa	476	275	201
Sonora	475	375	100
Tabasco	405	160	245
Tamaulipas	710	570	140
Tlaxcala	463	165	298
Veracruz de Ignacio de la Llave	460	205	255
Yucatán	405	260	145
Zacatecas	355	155	200

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
Nacional	15 291	9 800	5 491
1	2 631	175	2 456
2	7 640	4 680	2 960
3	3 500	3 450	50
4	1 520	1 495	25