

Encuesta Nacional de Bienestar Autorreportado 2021

ENBIARE

Diseño muestral



Instituto Nacional de Estadística y Geografía

**Encuesta Nacional de Bienestar
Autorreportado 2021**

ENBIARE

Diseño muestral



Obras complementarias publicadas por el INEGI sobre el tema:
Módulo de Bienestar Autorreportado (BIARE).

Catalogación en la fuente INEGI:

330.97201 Encuesta Nacional de Bienestar Autorreportado (2021).
Encuesta Nacional de Bienestar Autorreportado 2021 : ENBIARE : diseño muestral / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2021.

vii, 18 p.

1. México - Condiciones económicas - Metodología 2. México - Condiciones sociales - 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

DR © 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 Bienestar Autorreportado (ENBIARE) 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	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 Distribución 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 estimación de la población	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 las UPM de la Muestra Maestra	15
B. Distribución de la muestra de viviendas seleccionadas por entidad	16

Introducción

La **Encuesta Nacional de Bienestar Autorreportado (ENBIARE) 2021** es un ejercicio estadístico desarrollado por el **INEGI**, con el propósito de ampliar y actualizar la producción de información sobre los principales aspectos relacionados con el bienestar de los hogares en México y su vínculo con las condiciones sociales del entorno, de conformidad con las recomendaciones y mejores prácticas internacionales.

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

Debido a la necesidad de disponer de resultados estadísticos sobre el tema del bienestar subjetivo en México y con el fin de avanzar hacia la consolidación, como datos oficiales, de las valoraciones que la propia población adulta está haciendo sobre su calidad de vida y lo que importa en ella, se establecieron los siguientes objetivos:

General

Generar información estadística sobre distintas dimensiones de la medición del bienestar; hacer visible cómo mujeres y hombres perciben y evalúan su propia situación; recopilar información acerca de una amplia gama de circunstancias y situaciones que experimentan las personas, para detectar impulsores y detractores del bienestar; y hacer visibles las desigualdades entre grupos poblacionales. Todo ello respecto al conjunto de la población adulta de 18 años y más años de edad que reside en áreas urbanas y rurales.

Específicos

- Actualizar los indicadores de bienestar subjetivo en el ámbito nacional y por entidad federativa, bajo el estándar OCDE.
- Cubrir las dimensiones del bienestar propuesta por la *Iniciativa Para una Vida Mejor: Midiendo el Bienestar y Progreso (Better Life Initiative: Measuring Well-Being and Progress)*, impulsada por la OCDE.
- Proporcionar información que permita analizar los indicadores de bienestar subjetivo, a partir de eventos y circunstancias en la vida de las personas.
- Proporcionar indicadores sociales para el tablero *How's Life* que demanda la iniciativa OCDE sobre la medición de una vida mejor que considera 11 aspectos específicos del bienestar: ingresos y riqueza; trabajo y calidad del empleo; la vivienda; la salud; la educación, y las habilidades; la calidad del medio ambiente; una de las del bienestar subjetivo: la satisfacción ante la vida; la seguridad; balance entre la vida y el trabajo; la comunidad y los vínculos sociales; el compromiso cívico.
- Dar cuenta de las desigualdades como dimensión transversal del bienestar.
- Generar información sobre otra dimensión transversal del bienestar que es la sostenibilidad, con énfasis en el capital social.

1.2 Referencias metodológicas

Población objetivo. Una persona adulta, de 18 años cumplidos o más, seleccionada en cada uno de los hogares o grupos de personas que comparten un gasto común y residen en las viviendas asignadas para el levantamiento¹.

¹ Al tratarse de una encuesta en hogares, la información proviene de una muestra aleatoria representativa de la población adulta, cuya última unidad de selección es la persona; sin embargo, la estrategia metodológica de entrevista directa requiere habilidades de lecto–escritura en lengua española, lo que podría limitar la elegibilidad de algunos informantes seleccionados. Adicionalmente, el diseño estadístico puede implicar limitaciones para la representación de minorías dentro de esta población, lo que se precisa a través de indicadores que forman parte de la presentación de resultados del proyecto.

Periodo de referencia. En el Cuestionario del hogar, todas las preguntas se refieren a la situación al momento de la entrevista. Para el Cuestionario de bienestar autorreportado, las preguntas ubican al informante en diferentes tiempos, siendo estos el momento actual o presente, el día anterior a la entrevista, la semana pasada, las últimas dos y cuatro semanas, los últimos doce meses, hace un año, hace cinco años, y alguna vez en su vida.

Cobertura geográfica. La encuesta está diseñada para proporcionar resultados a nivel nacional y por entidad federativa.

Periodo de levantamiento. Del 3 de junio al 23 de julio de 2021.

Tamaño de la muestra. Fueron 37 000 viviendas a nivel nacional, distribuidas de manera homogénea entre las coordinaciones estatales.

Distribución de la muestra por coordinación estatal

Coordinación estatal	Muestra	Coordinación estatal	Muestra
Nacional	37 000	Michoacán de Ocampo	1 155
Aguascalientes	1 155	Morelos	1 155
Baja California	1 155	Nayarit	1 155
Baja California Sur	1 155	Nuevo León	1 160
Campeche	1 155	Oaxaca	1 155
Ciudad de México	1 155	Puebla	1 160
Coahuila de Zaragoza	1 155	Querétaro	1 155
Colima	1 160	Quintana Roo	1 155
Chiapas	1 155	San Luis Potosí	1 155
Chihuahua	1 160	Sinaloa	1 155
Durango	1 155	Sonora	1 155
Guanajuato	1 160	Tabasco	1 155
Guerrero	1 155	Tamaulipas	1 155
Hidalgo	1 155	Tlaxcala	1 155
Jalisco	1 160	Veracruz de Ignacio de la Llave	1 160
México	1 160	Yucatán	1 155
Oriente	880	Zacatecas	1 155
Poniente	280		

Método de recolección. Entrevista directa a jefa o jefe del hogar o persona mayor de 18 años que conozca la información de la vivienda, por medio de un cuestionario electrónico operado mediante un dispositivo electrónico.

Informante adecuado. Persona de 18 años de edad o más y que resida normalmente en la vivienda.

Persona elegida. Persona adulta seleccionada por un procedimiento aleatorio entre los integrantes del hogar. El procedimiento de selección del informante adecuado inicia con la identificación de los integrantes habituales del hogar que se encuentran dentro del rango de edad establecido de 18 años cumplidos o más, con base en la información recopilada en el Cuestionario del hogar. Además, cumple con los criterios de saber leer, escribir y hablar español.

Cuando la entrevista se realizó mediante cuestionario en papel, se eligió a la persona integrante del hogar de 18 años o más, cuya fecha de cumpleaños fuera la más cercana, posterior a la visita en la vivienda, considerando el día y mes de nacimiento. Este método garantizó de manera sencilla la aleatoriedad de la selección.

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 ENBIARE 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 años o más.

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

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 AGEB².
- 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.

² 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 indicadores³ 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:

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

Donde:

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

S = Muestra Maestra.

k_{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.

³ 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 estatal fue:

$$n = \frac{z^2 s^2 \text{ DEFF}}{r^2 \bar{x} (1 - \text{tnr})}$$

Donde:

- n = tamaño de la muestra.
- \bar{x} = estimación del promedio de la variable de interés a nivel vivienda.
- s^2 = varianza poblacional.
- r = error relativo máximo aceptable.
- z = valor asentado en las tablas estadísticas de la distribución normal estándar que garantiza obtener las estimaciones con 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.

El tamaño de muestra estatal se obtuvo considerando una confianza de 90% y un efecto de diseño de 2.667 observado en los estudios anteriores, así como un error relativo máximo esperado de 2%, un promedio de satisfacción con la vida por vivienda de 8.005, una varianza poblacional de 3.469 y una tasa de no respuesta de 15%, se determinó una muestra de 1 147 viviendas por entidad federativa, el cual se ajustó a un tamaño de muestra de 1 155 o 1 160 viviendas, para llegar a un tamaño nacional acumulado de 37 000 viviendas.

2.6 Distribución de la muestra

La distribución de la muestra se realizó dentro de cada entidad federativa, entre los diferentes estratos de diseño, de manera proporcional a su tamaño, para lo cual se empleó la siguiente expresión:

$$n_{eh} = \frac{N_{eh}}{N_e} n_e$$

El número de UPM a seleccionar se obtiene a partir de la siguiente expresión:

$$k_{eh}^* = \frac{n_{eh}}{b}$$

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, según su distribución proporcional.
- 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.
- k_{eh}^* = número de UPM en muestra, en el h-ésimo estrato, en la e-ésima entidad.
- b = número de viviendas a seleccionar por UPM.

En el anexo B se presenta la distribución de la muestra de viviendas por entidad federativa para la ENBIARE 2021.

2.7 Selección de la muestra

Se realizó de manera independiente por entidad, dominio y estrato; el procedimiento de selección varió de acuerdo con el dominio.

2.7.1 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. En cada uno de los hogares registrados dentro de la vivienda seleccionada, se seleccionó una persona de 18 años o más.

Por lo tanto, la probabilidad de seleccionar una vivienda de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, se define de la siguiente manera:

$$P\{V_{ehi}\} = \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{5}{m_{ehi}^*} = \frac{5 k_{eh} m_{ehi}}{m_{eh} m_{ehi}^*}$$

Su factor de expansión está dado por:

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

La probabilidad de seleccionar una persona en el ℓ -ésimo hogar, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, se define de la siguiente manera:

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

Su factor de expansión⁴ está dado por:

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

⁴ El factor de expansión se define como el inverso de la probabilidad de selección.

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, 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.
- $Q_{ehij\ell}$ = número de personas de 18 años o más en el ℓ -ésimo hogar, 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. Se seleccionaron k_{eh} UPM, con probabilidad proporcional al número de viviendas del estrato.
2. En cada UPM seleccionada, se eligieron 20 viviendas con igual probabilidad.
3. En cada vivienda seleccionada y por cada uno de los hogares registrados, se seleccionó una persona de 18 años o más.

Por lo tanto, la probabilidad de seleccionar una vivienda de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, se define de la siguiente manera:

$$P\{V_{ehi}\} = \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{20}{m_{ehi}^*} = \frac{20 k_{eh} m_{ehi}}{m_{eh} m_{ehi}^*}$$

Su factor de expansión está dado por:

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

La probabilidad de seleccionar una persona en el ℓ -ésimo hogar, de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, se define de la siguiente manera:

$$P\{V_{ehij\ell}\} = \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{20}{m_{ehi}^*} \frac{1}{Q_{ehij\ell}} = \frac{20 k_{eh} m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehij\ell}}$$

Su factor de expansión está dado por:

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

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, 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 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.
- $Q_{ehij\ell}$ = número de personas de 18 años o más en el ℓ -ésimo hogar, 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. 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. En cada vivienda seleccionada y por cada uno de los hogares registrados, se seleccionó una persona de 18 años o más.

Por lo tanto, la probabilidad de seleccionar una vivienda de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, se define de la siguiente manera:

$$P\{V_{ehi}\} = \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{4 \times 5}{m_{ehi}^*} = \frac{20 k_{eh} m_{ehi}}{m_{eh} m_{ehi}^*}$$

Su factor de expansión está dado por:

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

La probabilidad de seleccionar una persona en el ℓ -ésimo hogar, de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, se define de la siguiente manera:

$$P\{V_{ehij\ell}\} = \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{4 \times 5}{m_{ehi}^*} \frac{1}{Q_{ehij\ell}} = \frac{20 k_{eh} m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehij\ell}}$$

Su factor de expansión está dado por:

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

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_{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, 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.
- $Q_{ehij\ell}$ = número de personas de 18 años o más en el ℓ -ésimo hogar, 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

El ajuste por no respuesta, atribuida al informante, se realizó a nivel estrato, en cada uno de los dominios para cada una de las unidades de observación como se describe a continuación:

AJUSTE POR NO RESPUESTA DE LAS VIVIENDAS

$$F_{ehi}^* = F_{ehi} \frac{\sum_{1 \in h} F_{ehi} V_{ehi}}{\sum_{i \in h} F_{ehi} V_{ehi}^*}$$

AJUSTE POR NO RESPUESTA DE LOS HOGARES

$$F_{ehij}^{**} = F_{ehi}^* \frac{\sum_{1 \in h} F_{ehi}^* V_{ehij\ell}}{\sum_{i \in h} F_{ehi}^* V_{ehij\ell}^*}$$

Donde:

- F_{ehi}^* = factor de expansión corregido por no respuesta de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- F_{ehi} = factor de expansión de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- V_{ehi} = número de viviendas seleccionadas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.
- V_{ehi}^* = número de viviendas cuyos elegidos proporcionaron respuesta en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

2.8.2 Ajuste por estimación de población

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 el monto de la estimación de población generada por el INEGI⁵ a partir del primer trimestre del año 2021, referida al punto medio del levantamiento, mediante la siguiente expresión:

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

Donde:

- F''_D = factor de expansión corregido por estimación de población en el dominio D.
- F'_D = factor de expansión corregido por no respuesta en el dominio D.
- $PEST_D$ = población en el dominio D, según estimación de población del INEGI.
- $PEXP_D$ = población total a la que expande la encuesta en el dominio D.
- D = nivel de desagregación en se realiza el ajuste.

2.9 Estimadores

El estimador del total de la característica X es:

$$\hat{X} = \sum_e \sum_h \sum_i F_{ehi}^{UA} \left(\sum_j \sum_k X_{ehijk}^{UA} \right) + \sum_e \sum_h \sum_i F_{ehi}^{CU} \left(\sum_j \sum_k X_{ehijk}^{CU} \right) + \sum_e \sum_h \sum_i F_{ehi}^R \left(\sum_j \sum_k X_{ehijk}^R \right)$$

Donde:

- F_{ehi}^{UA} = factor de expansión final para la ENBIARE 2021 de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio urbano alto.
- X_{ehijk}^{UA} = valor observado de la característica de interés X en la k-ésima persona, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el dominio urbano alto.
- F_{ehi}^{CU} = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, en el dominio complemento urbano.
- X_{ehijk}^{CU} = valor observado de la característica de interés X en la k-ésima persona, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el dominio complemento urbano.
- F_{ehi}^R = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad del dominio rural.
- X_{ehijk}^R = valor observado de la característica de interés X en la k-ésima persona, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad del dominio rural.

Para la estimación de proporciones, tasas y promedios se utiliza el estimador de razón: $\hat{R} = \frac{\hat{X}}{\hat{Y}}$

Donde, la variable \hat{Y} es definida en forma análoga a \hat{X} .

⁵ El INEGI estará generando una estimación de población por entidad federativa, con base en la propia actualización del Marco de Muestreo de Viviendas del INEGI y de la información del CPV 2020, a través de una muestra aleatoria de viviendas que serán visitadas y contabilizadas su número de personas. La estimación de población se comenzó a implementar a partir del primer trimestre de 2021.

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*⁶, basado en que la mayor contribución a la varianza de un estimador, en un diseño bietápico, es la que se presenta entre las UPM. El término conglomerados últimos se utiliza para denotar el total de unidades en muestra de una 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} a nivel nacional:

$$\hat{V}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_e^{32} \left\{ \sum_h^{L_e} \frac{k_{eh}}{k_{eh} - 1} \sum_i^{k_{eh}} \left[\left(\hat{X}_{ehi} - \frac{1}{k_{eh}} \hat{X}_{eh} \right) - \hat{R} \left(\hat{Y}_{ehi} - \frac{1}{k_{eh}} \hat{Y}_{eh} \right) \right]^2 \right\}$$

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:

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

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.

⁶ Véase Hansen, M. H., Horwitz, W. N. y Madow, W.G., *Sample Survey Methods and Theory*, (1953), Vol. 1, pág. 242.

Finalmente, el intervalo de confianza $I_{1-\alpha}$ al 100 $(1 - \alpha)\%$, 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.

2.11 Homologación de la semaforización para los umbrales de indicadores de precisión estadística⁷

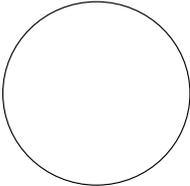
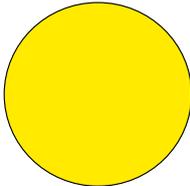
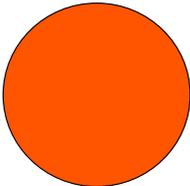
Para facilitar la interpretación de las precisiones estadísticas de la información pública en tabulados, el Comité de Aseguramiento de la Calidad, en la cuarta sección celebrada el 1 de noviembre de 2018, aprobó los siguientes umbrales y especificaciones para los coeficientes de variación en la publicación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

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.

⁷ 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.

Anexo

A. Indicadores empleados en la estratificación de las UPM 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

Entidad federativa	Total
Nacional	36 978 ^A
01 Aguascalientes	1 157
02 Baja California	1 155
03 Baja California Sur	1 151
04 Campeche	1 152
05 Coahuila de Zaragoza	1 154
06 Colima	1 155
07 Chiapas	1 155
08 Chihuahua	1 151
09 Ciudad de México	1 160
10 Durango	1 152
11 Guanajuato	1 160
12 Guerrero	1 154
13 Hidalgo	1 156
14 Jalisco	1 162
15 México	1 162
16 Michoacán de Ocampo	1 153
17 Morelos	1 158
18 Nayarit	1 156
19 Nuevo León	1 160
20 Oaxaca	1 149
21 Puebla	1 160
22 Querétaro	1 157
23 Quintana Roo	1 152
24 San Luis Potosí	1 157
25 Sinaloa	1 157
26 Sonora	1 150
27 Tabasco	1 158
28 Tamaulipas	1 155
29 Tlaxcala	1 153
30 Veracruz de Ignacio de la Llave	1 159
31 Yucatán	1 155
32 Zacatecas	1 153

^A El tamaño de la muestra calculada fue de 37 000 viviendas; sin embargo, la selección final fue de 36 978. 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.