

ENDUTIH

Encuesta Nacional sobre Disponibilidad
y Uso de Tecnologías de la Información
en los Hogares

2023

Diseño muestral



Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares 2023

ENDUTIH

Diseño muestral



Obras complementarias publicadas por el INEGI sobre el tema:

Diseño muestral de la Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2022.

Catalogación en la fuente INEGI:

643.580723 Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (2023).

Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares 2023 : ENDUTIH : diseño muestral / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2023.

vii, 19 p.

1. Computadoras y familia - Encuestas - Metodología. 2. Familia - Recursos en redes de computación. 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

2024, **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 sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2023**, 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	9
2.8 Ajuste a los factores de expansión	10
2.8.1 Ajuste por no respuesta	10
2.8.2 Ajuste por estimación de población	11
2.9 Estimadores	12
2.10 Estimación de errores de muestreo	13
2.11 Homologación de la semaforización para los umbrales de indicadores de precisión estadística	14
Anexos	15
A. Indicadores empleados en la estratificación de las UPM de la Muestra Maestra	17
B. Distribución de la muestra de UPM y viviendas seleccionadas por entidad federativa	18
C. Distribución de la muestra en viviendas seleccionadas por entidad federativa, según ámbito urbano y rural	19

Introducción

La **Encuesta Nacional sobre Disponibilidad y Uso de las Tecnologías de la Información en los Hogares (ENDUTIH) 2023** es un proyecto estadístico destinado a investigar en qué medida, y a través de cuáles medios, la población tiene acceso y hace uso de las tecnologías de la información y comunicación.

Los datos que proporcionará permitirán al **Instituto Federal de Telecomunicaciones (IFT)** realizar análisis para establecer políticas públicas encaminadas a impulsar los cambios necesarios que apunten hacia una Sociedad de la Información y el Conocimiento (SIC).

La encuesta tiene el propósito de medir el acceso y uso que las personas integrantes de los hogares, de seis o más años de edad, tienen a la radio, televisión, teléfono, computadora y consola de video, para el desarrollo de proyectos vinculados a la conectividad, contenidos y sistemas que contribuyan a la adopción de las Tecnologías de la Información y Comunicación (TIC).

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

En virtud de la necesidad de conocer la situación que guarda la tecnología de la información con la población que habita en nuestro país, y debido al progreso tecnológico e incursión de México en el contexto de la Sociedad de la Información y el Conocimiento, se plantean los siguientes objetivos.

General

Generar información estadística que permita conocer la disponibilidad y el uso de las tecnologías de la información y las comunicaciones en los hogares y por las personas de seis o más años a nivel nacional, nacional urbano–rural, por estrato socioeconómico y por entidad federativa.

Específicos

- Aplicar en una muestra de 65 000 viviendas un instrumento de captación (cuestionario) que permita evaluar la disponibilidad y uso de las tecnologías de la información y las comunicaciones en los hogares.
- Obtener información para la generación de indicadores que permita:
- Conocer la condición de disponibilidad de computadora, servicios de telefonía, señal de televisión de paga y de conexión a internet.
- Conocer las condiciones que restringen la disponibilidad de computadora e internet.
- Establecer el volumen de usuarios de computadora, servicio de telefonía, señal de televisión de paga y de conexión a internet, y las características de uso, tipo de aplicaciones utilizadas, la periodicidad y la recurrencia de transacciones electrónicas, así como las características sociodemográficas de los usuarios de dichas tecnologías.

1.2 Referencias metodológicas

Población objetivo. La encuesta está dirigida a las personas de seis o más años de edad que residen permanentemente en viviendas particulares ubicadas en el territorio nacional a la fecha del levantamiento.

Periodo de referencia. La información objeto central de la ENDUTIH corresponde al periodo de levantamiento, última semana, último mes, tres y hasta doce meses anteriores a la entrevista.

Periodo de levantamiento. Del 12 de junio al 4 de agosto de 2023.

Método de recolección. El método para captar la información fue mediante entrevista directa en la que se aplicó un cuestionario en dispositivo electrónico o impreso en papel cuando las condiciones no permitieron el uso del electrónico, estructurado con preguntas ordenadas y secuenciadas, de opciones de respuesta precodificadas y con opción a registrar textualmente, según se requiera, aquellas otras que no estén consideradas en el instrumento.

Informante adecuado. Residente de la vivienda de 15 años o más, quien da respuesta a las preguntas generales sobre vivienda, características sociodemográficas de los individuos y equipamiento de tecnologías de la información en el hogar.

Informante elegido. Fue la persona seleccionada de manera aleatoria, mediante un algoritmo integrado en el programa de captura del Dispositivo Móvil, bajo la característica de tener seis o más años de edad y ser residente habitual de la vivienda. Si la persona informante era menor de edad, debía estar presente durante la entrevista el padre, la madre, el tutor o una persona de su confianza.

En aquellos casos en los que fue necesario aplicar el cuestionario en formato impreso, el criterio de selección de la persona informante consistió en elegir aquella residente de la vivienda de seis o más años de edad cuya fecha de cumpleaños fuera la primera de las próximas y posterior al día de la entrevista.

Cobertura geográfica. La encuesta está diseñada para dar resultados a nivel nacional, ámbito urbano y rural, y por entidad federativa.

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 ENDUTIH 2023 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 en varias etapas, estratificado y por conglomerados, donde la unidad última de selección es la persona que al momento de la entrevista tenga seis años cumplidos o más de edad.

Para la selección de la muestra se utilizó la Muestra Maestra, también conocida como el Marco Nacional de Viviendas 2012 del INEGI, este último construido a partir de la información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2010. A partir de la Muestra Maestra 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. Las UPM se forman de la siguiente manera:

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 AGEB.¹
- La unión de dos o más manzanas contiguas de diferentes AGEB de la misma localidad.
- La unión de dos o más manzanas contiguas de diferentes localidades, 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 lo siguiente:

- 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 lo siguiente:

- 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

Una vez construido el conjunto de UPM, se agruparon aquellas con características similares, lo cual se denomina 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	Descripción
Urbano alto	Áreas urbanas con 100 000 o más habitantes.
Complemento urbano	De 2 500 a 99 999 habitantes.
Rural	Localidades menores de 2 500 habitantes.

De manera paralela, se formaron cuatro estratos socioeconómicos en los que se agruparon todas las UPM del país; esta estratificación considera las características sociodemográficas de los habitantes de las viviendas, así como las características físicas y el equipamiento de las mismas, expresadas por medio de 34 indicadores² contruidos con información del Censo de Población y Vivienda 2010*, para lo cual se emplearon métodos estadísticos multivariados.

De esta forma, cada UPM fue clasificada en un único estrato geográfico y uno sociodemográfico. Como resultado, se 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 seleccionadas² con probabilidad proporcional al tamaño, mediante la siguiente expresión:

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

Donde:

U_{ehi} = i-ésima UPM, del h-ésimo estrato, de 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 captada en el Censo de Población y Vivienda 2010.

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

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

² La distribución de UPM y viviendas seleccionadas por entidad federativa se presenta en el Anexo B..

Las UPM seleccionadas forman la Muestra Maestra, que permite elegir las submuestras de las encuestas en hogares.

2.5 Tamaño de la muestra

Para el cálculo del tamaño de muestra de la ENDUTIH 2023 se empleó la siguiente expresión:

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

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, que garantiza realizar 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.
- PHV = promedio de hogares por vivienda.

Considerando un nivel de confianza de 90 %, un efecto de diseño de 4 observado en experiencias anteriores, un error relativo máximo esperado de 13.8 %, una tasa de no respuesta de 15 %, un promedio de hogares por vivienda de 1.02 y una proporción de 1 por ciento. Con estos parámetros se obtiene un tamaño de muestra aproximado de 64 901 viviendas a nivel nacional, el cual se ajustó a 65 172 viviendas.

2.6 Distribución de la muestra

La distribución de la muestra se realizó dentro de cada entidad federativa, entre los diferentes estratos 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.
- 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 C se presenta la distribución de la muestra en viviendas seleccionadas para la ENDUTIH 2023 por entidad federativa según ámbito urbano y rural.

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

El procedimiento probabilístico general de selección se describe en tres etapas:

1. De las k_{eh} UPM, seleccionadas en el h-ésimo estrato, en la e-ésima entidad para la Muestra Maestra,³ se eligieron k_{eh}^* con igual probabilidad.
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 seis o más años.

En forma detallada, la probabilidad de selección de las UPM, 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 lo siguiente:

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

- b. La probabilidad de seleccionar, a partir de la Muestra Maestra, la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra de la ENDUTIH 2023 está definida por lo siguiente:

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

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

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

- d. La probabilidad de seleccionar el k-ésimo hogar, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es la siguiente:

$$P_{4ehijk} = 1$$

- e. La probabilidad de seleccionar la ℓ -ésima persona, del k-ésimo hogar, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es la siguiente:

$$P_{5ehijk\ell} = \frac{1}{Q_{ehijk\ell}}$$

³ Del total de UPM que integran el marco de propósitos múltiples, se seleccionaron con probabilidad proporcional a su tamaño k_{eh} UPM para localidades de 100 000 y más habitantes.

Por lo tanto, la probabilidad de selección de una persona en el k-ésimo hogar, 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:

$$P_{ehijk\ell} = P_{1ehi} * P_{2ehi} * P_{3ehij} * P_{4ehijk} * P_{5ehijk\ell}$$

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

Su factor de expansión⁴ está dado por lo siguiente:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad, en 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.
- k_{eh}^* = número de UPM a seleccionar para la ENDUTIH 2023 con igual probabilidad de selección, en el h-ésimo estrato, en la e-ésima entidad.
- $Q_{ehijk\ell}$ = número de personas de seis o más años en el k-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

El procedimiento probabilístico general, de selección, se describe en varias etapas:

1. De las k_{eh} UPM, seleccionadas en el h-ésimo estrato, en la e-ésima entidad para la Muestra Maestra,⁵ se eligieron k_{eh}^* con igual probabilidad.
2. En cada UPM seleccionada, se eligieron 20 viviendas con igual probabilidad.
3. En cada uno de los hogares registrados dentro de la vivienda seleccionada, se seleccionó una persona de seis o más años.

En forma detallada, la probabilidad de selección de las UPM, 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 lo siguiente:

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

⁴ 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.

⁵ Del total de UPM que integran el marco de propósitos múltiples, se seleccionaron con probabilidad proporcional a su tamaño k_{eh} UPM para localidades de 100 000 y más habitantes.

- b. La probabilidad de seleccionar, a partir de la Muestra Maestra, la i -ésima UPM del h -ésimo estrato de la e -ésima entidad para la muestra de la ENDUTIH 2023 está definida por lo siguiente:

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

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

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

- d. La probabilidad de seleccionar el k -ésimo hogar, de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad es la siguiente:

$$P_{4ehijk} = 1$$

- e. La probabilidad de seleccionar la ℓ -ésima persona, del k -ésimo hogar, de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad es la siguiente:

$$P_{5ehijk\ell} = \frac{1}{Q_{ehijk\ell}}$$

Por lo tanto, la probabilidad de selección de una persona en el k -ésimo hogar, 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:

$$\begin{aligned} P_{ehijk\ell} &= P_{1ehi} * P_{2ehi} * P_{3ehij} * P_{4ehijk} * P_{5ehijk\ell} \\ &= \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{k_{eh}^*}{k_{eh}} \frac{20}{m_{ehi}^*} \frac{1}{Q_{ehijk\ell}} = \frac{20 k_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehijk\ell}} \end{aligned}$$

Su factor de expansión está dado por lo siguiente:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h -ésimo estrato, en la e -ésima entidad, en 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.
- k_{eh}^* = número de UPM a seleccionar con igual probabilidad de selección, en el h -ésimo estrato, en la e -ésima entidad.
- $Q_{ehijk\ell}$ = número de personas de seis o más años en el k -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

El procedimiento probabilístico general, de selección, se describe en varias etapas:

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}^* con igual probabilidad.
2. En cada UPM seleccionada, se eligieron cuatro segmentos de aproximadamente cinco viviendas, con igual probabilidad.
3. En cada uno de los hogares registrados dentro de la vivienda seleccionada, se seleccionó una persona de seis o más años.

En forma detallada, la probabilidad de selección de las UPM, 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 lo siguiente:

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

- b. La probabilidad de seleccionar, a partir de la Muestra Maestra, la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad para la muestra de la ENDUTIH 2023 está definida por lo siguiente:

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

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

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

- d. La probabilidad de seleccionar el k -ésimo hogar, de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad es la siguiente:

$$P_{4ehijk} = 1$$

- e. La probabilidad de seleccionar la ℓ -ésima persona, del k -ésimo hogar, de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad es la siguiente:

$$P_{5ehijk\ell} = \frac{1}{Q_{ehijk\ell}}$$

Por lo tanto, la probabilidad de selección de una persona en el k -ésimo hogar, 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:

$$\begin{aligned} P_{ehijk\ell} &= P_{1ehi} * P_{2ehi} * P_{3ehij} * P_{4ehijk} * P_{5ehijk\ell} \\ &= \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{k_{eh}^*}{k_{eh}} \frac{4 * 5}{m_{ehi}^*} \frac{1}{Q_{ehijk\ell}} = \frac{20 k_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehijk\ell}} \end{aligned}$$

Su factor de expansión está dado por lo siguiente:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad, en 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.
- k_{eh}^* = número de UPM a seleccionar con igual probabilidad de selección, en el h-ésimo estrato, en la e-ésima entidad.
- $Q_{ehijk\ell}$ = número de personas de seis o más años en el k-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 se realizó a nivel de las viviendas, los hogares y las personas seleccionadas a nivel estrato, en cada uno de los dominios, mediante las siguientes expresiones:

A NIVEL VIVIENDA

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

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

Donde:

- F'_{ehij} = factor de expansión corregido por no respuesta de la j-ésima vivienda seleccionada, 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 uno si la j-ésima vivienda seleccionada, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad proporcionó respuesta completa, y toma el valor de cero en caso contrario.

A NIVEL HOGAR

El ajuste por no respuesta atribuida al informante se realiza 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 del k-ésimo hogar en muestra, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- F'_{ehijk} = factor de expansión del k-ésimo hogar, 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 uno si en el k-ésimo hogar en muestra, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad proporcionó respuesta completa, y toma el valor de cero en caso contrario.

A NIVEL PERSONA

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

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

Donde:

- $F'''_{ehijk\ell}$ = factor de expansión corregido por no respuesta de la ℓ -ésima persona seleccionada, del k-ésimo hogar, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- $F''_{ehijk\ell}$ = factor de expansión para la ℓ -ésima persona seleccionada, del k-ésimo hogar, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- $I_{ehijk\ell}$ = función indicadora con valor 1 o 0; esta función toma el valor uno si la ℓ -ésima persona seleccionada, en el k-ésimo hogar, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad proporcionó respuesta completa, y toma el valor de cero en caso contrario.

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 la población total determinada por la estimación de población 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.
 $PEXP_D$ = población total a la que expande la encuesta en el dominio D.
 D = nivel de desagregación de la población al que se realiza la estimación de población.

Una vez ajustados los factores de expansión de acuerdo a los criterios antes descritos, se aplican a cada una de las observaciones obtenidas en la encuesta. De manera natural, se obtiene la estimación de población de seis años o más de edad, o cualquier otro grupo de edad, a partir de la propia distribución muestral de residentes, por lo que no se consideran ajustes adicionales a los ponderadores por grupos de edad específicos.

La población considerada para el cálculo de los ponderadores fue la estimación de población que se genera de manera trimestral a través de la Muestra Maestra, ajustada a la referencia de agosto de 2023.

2.9 Estimadores

El estimador del total de la característica X es la siguiente:

$$\hat{X} = \sum_e \sum_h \sum_i F_{ehij}^{UA} \left(\sum_s \sum_{\ell} X_{ehis\ell}^{UA} \right) + \sum_e \sum_h \sum_i F_{ehij}^{CU} \left(\sum_s \sum_{\ell} X_{ehis\ell}^{CU} \right) + \sum_e \sum_h \sum_i F_{ehij}^R \left(\sum_s \sum_{\ell} X_{ehis\ell}^R \right)$$

Donde:

- F_{ehij}^{UA} = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio urbano alto.
 $X_{ehis\ell}^{UA}$ = valor observado de la característica de interés X en la persona elegida, en el ℓ -ésimo hogar, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el dominio urbano alto.
 F_{ehij}^{CU} = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, en el dominio complemento urbano.
 $X_{ehis\ell}^{CU}$ = valor observado de la característica de interés X en la persona elegida, en el ℓ -ésimo hogar, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el dominio complemento urbano.
 F_{ehij}^R = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad del dominio rural.
 $X_{ehis\ell}^R$ = valor observado de la característica de interés X en la persona elegida, en el ℓ -ésimo hogar, en la s-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} .

2.10 Estimación de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones del agregado nacional, se usó el método de *Conglomerados Últimos*,⁶ 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 los errores de muestreo de los estimadores de razón, conjuntamente al método de *Conglomerados Últimos* se aplicó el método de *Serie de Taylor*, obteniéndose la siguiente fórmula para estimar la precisión de \hat{R} :

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

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}_{NAL}^2 = estimador del total de la característica Y, al cuadrado.

Estas definiciones son análogas para la variable de estudio \hat{Y} .

La estimación de la varianza del estimador de un total se calcula con la siguiente expresión:

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

Las estimaciones de la desviación estándar (DE), efecto de diseño (DEFF) y coeficiente de variación (CV) se calculan mediante las siguientes expresiones:

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

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 $(1 - \alpha)\%$, 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.

⁶ 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 estadística⁷

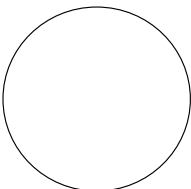
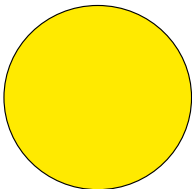
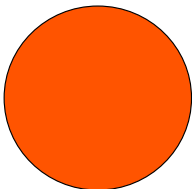
Para facilitar la interpretación de las precisiones estadísticas de la información pública en tabulados, el Comité de Aseguramiento de la Calidad, en la cuarta sección celebrada el 1 de noviembre de 2018, aprobó los siguientes umbrales y especificaciones para los coeficientes de variación en la publicación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

Nivel de precisión de las estimaciones:

Alta, CV en el rango de [0, 15)

Moderada, CV en el rango de [15, 30)

Baja, CV de 30 % en adelante

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_umbrals.pdf.

Anexos

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 trabajó; tenía trabajo pero no trabajó; o buscó trabajo en la semana de referencia.
TOCU12A17	no ocupada de 12 a 17 años entre la población total de este rango.
PPOMAYED	ocupada de 18 años y más entre la población total ocupada.
PGDO_ESC	Grado promedio de escolaridad.
PTASAOcupa	Tasa de ocupación.
Proporción de viviendas particulares habitadas	
PVIVSINH	que no tienen hacinamiento.
PVPH_PISDT	que tienen piso de cemento o firme, madera, mosaico u otro material.
PVPH2YMASD	que usan para dormir entre 2 y 25 cuartos.
PVPH_2MASC	que tienen más de un cuarto.
PVPH3YMASD	que tienen entre 3 y 25 cuartos.
PVPH_C_ELE	que disponen de luz eléctrica.
PVPHAGUADV	que tienen disponibilidad de agua entubada dentro de la vivienda, o fuera de ella, pero dentro del terreno.
PVPH_EXCSA	que tienen excusado, retrete, sanitario, letrina u hoyo negro.
PVPHDRENAJ	que tienen drenaje conectado a la red pública, fosa séptica, barranca, grieta, río, lago o mar.
PVDRERED	que disponen de drenaje conectado a la red pública.
PVEXCAGU	que disponen de excusado con descarga directa de agua.
PVPH_CSERV	que disponen de luz eléctrica, agua entubada dentro o fuera de la vivienda, pero dentro del terreno, así como drenaje.
PSIN_HASIN	que no se encuentran en situación de hacinamiento a nivel manzana.
Proporción de viviendas particulares habitadas que disponen de	
PVPH_TV	televisor.
PVPH_AUTOM	automóvil o camioneta.
PVPH_CEL	teléfono celular.
PVCELFIJ	teléfono celular y teléfono fijo.
PV4ELEC	radio, televisor, refrigerador y lavadora.
PVRADTEL	radio y televisor.
PVPHCBIEN	todos los bienes.

B. Distribución de la muestra de UPM y viviendas seleccionadas por entidad federativa

Entidad federativa	UPM	Viviendas
Nacional	8 236	65 172
Aguascalientes	307	2 034
Baja California	331	2 032
Baja California Sur	207	2 033
Campeche	256	2 041
Coahuila de Zaragoza	325	2 033
Colima	289	2 031
Chiapas	173	2 051
Chihuahua	322	2 030
Ciudad de México	405	2 035
Durango	268	2 033
Guanajuato	239	2 041
Guerrero	199	2 043
Hidalgo	163	2 046
Jalisco	269	2 044
México	329	2 037
Michoacán de Ocampo	181	2 039
Morelos	265	2 037
Nayarit	208	2 032
Nuevo León	344	2 036
Oaxaca	154	2 044
Puebla	227	2 044
Querétaro	265	2 038
Quintana Roo	325	2 032
San Luis Potosí	238	2 033
Sinaloa	253	2 030
Sonora	289	2 033
Tabasco	160	2 032
Tamaulipas	322	2 034
Tlaxcala	277	2 034
Veracruz de Ignacio de la Llave	206	2 042
Yucatán	259	2 031
Zacatecas	181	2 037

C. Distribución de la muestra en viviendas seleccionadas por entidad federativa, según ámbito urbano y rural

Entidad federativa	Total	Ámbito	
		Urbano	Rural
Nacional	65 172	49 800	15 372
Aguascalientes	2 034	1 670	364
Baja California	2 032	1 870	162
Baja California Sur	2 033	1 750	283
Campeche	2 041	1 550	491
Coahuila de Zaragoza	2 033	1 830	203
Colima	2 031	1 810	221
Chiapas	2 051	1 095	956
Chihuahua	2 030	1 730	300
Ciudad de México	2 035	2 015	20
Durango	2 033	1 430	603
Guanajuato	2 041	1 435	606
Guerrero	2 043	1 230	813
Hidalgo	2 046	1 130	916
Jalisco	2 044	1 755	289
México	2 037	1 795	242
Michoacán de Ocampo	2 039	1 410	629
Morelos	2 037	1 710	327
Nayarit	2 032	1 430	602
Nuevo León	2 036	1 915	121
Oaxaca	2 044	990	1 054
Puebla	2 044	1 495	549
Querétaro	2 038	1 470	568
Quintana Roo	2 032	1 810	222
San Luis Potosí	2 033	1 330	703
Sinaloa	2 030	1 510	520
Sonora	2 033	1 750	283
Tabasco	2 032	1 210	822
Tamaulipas	2 034	1 770	264
Tlaxcala	2 034	1 670	364
Veracruz de Ignacio de la Llave	2 042	1 295	747
Yucatán	2 031	1 730	301
Zacatecas	2 037	1 210	827