

ENDIREH

Encuesta Nacional sobre la
Dinámica de las Relaciones
en los Hogares

2021

Diseño muestral



 **INEGI**

Instituto Nacional de Estadística y Geografía

Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021

ENDIREH

Diseño muestral



Obras complementarias publicadas por el INEGI sobre el tema:

Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021. ENDIREH. Marco conceptual;
Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021. ENDIREH. Informe operativo;
Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2016. ENDIREH. Marco conceptual;
Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2016. ENDIREH. Informe operativo;
Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2011. ENDIREH. Manual de la entrevistadora;
Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2011. ENDIREH. Informe operativo;
Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2011. ENDIREH. Síntesis metodológica.

Catalogación en la fuente INEGI:

362.82920723 Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares.
Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021 :
ENDIREH : diseño muestral / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2022.

vii, 18 p.

1. Violencia familiar - México - 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

2022, **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, realiza la documentación de los diferentes proyectos que lleva a cabo.

Por lo anterior presentamos el **Documento Metodológico de la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH) 2021** con la intención de dar a conocer los principales rasgos de la metodología utilizada para el diseño y desarrollo estadístico de este proyecto.

Con la finalidad de transparentar el proceso de generación de información estadística, el **INEGI** pone a disposición de los usuarios este documento y de esta manera se dan a conocer los procedimientos y lineamientos que se siguieron y que hicieron posible la realización del proyecto.

Índice

Introducción	VII
Diseño muestral ENDIREH 2021	1
1. Objetivo general	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Diseño de la muestra	1
4.1 Formación de las unidades primarias de muestreo (UPM)	2
4.2 Estratificación	2
4.3 Selección de las UPM de la muestra maestra	3
5. Tamaño de la muestra	3
6. Afijación de la muestra	4
7. Selección de la muestra	4
7.1 En urbano alto	5
7.2 En complemento urbano	6
7.3 En rural	7
8. Ajuste a los factores de expansión	8
8.1 Ajuste por no respuesta	9
8.2 Ajuste por estimación de población	9
9. Estimadores	10
10. Estimaciones de los errores de muestreo	10
11. Homologación de la semaforización para los umbrales de indicadores de precisión estadística	12
Anexo	15
A. Indicadores empleados en la estratificación de conglomerados de la muestra maestra	17
B. Distribución de la muestra en viviendas por entidad federativa según ámbito de estudio para la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021	18

Introducción

La **Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH) 2021** es un proyecto estadístico estratégicamente diseñado para obtener información sobre la frecuencia y magnitud de la violencia que experimentan las mujeres al interior de las viviendas; así como identificar los eventos de discriminación, agresión y violencia que han padecido en los ámbitos escolar, laboral, familiar y comunitario, por mencionar algunos.

Con el propósito de enmarcar las principales características metodológicas del proyecto, se realiza el Diseño Muestral de la ENDIREH 2021, el cual aborda las principales características del diseño estadístico empleado para el desarrollo de la encuesta, como son el marco de muestreo, el tamaño y la distribución de la muestra, así como el cálculo de las probabilidades de inclusión de las unidades de observación, los factores de expansión y los principales estimadores para la explotación de información.

Diseño muestral ENDIREH 2021

1. Objetivo general

Generar información sobre las experiencias de violencia que han enfrentado las mujeres de 15 años y más, de manera detallada por tipo de violencia, para los distintos ámbitos (de pareja, familiar, escolar, laboral y comunitario) y de acuerdo con el vínculo o relación con las personas agresoras; con la finalidad de disponer de información que permita estimar los principales indicadores sobre la prevalencia y gravedad de la violencia, que sean comparables con los estimados en las ediciones anteriores de la encuesta, y de esta manera contribuir al conocimiento del problema, al desarrollo de investigaciones y al diseño de políticas públicas orientadas a atender y erradicar la violencia contra las mujeres por razones de género.

2. Población objetivo

Mujeres de 15 años o más residentes habituales de las viviendas seleccionadas en la muestra.

3. Cobertura geográfica

La encuesta está diseñada para dar resultados a los siguientes niveles de desagregación:

- Nacional
 - Urbano
 - Rural
- Entidad federativa

4. Diseño de la muestra

El diseño de la muestra para la ENDIREH 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 es la vivienda y la unidad de observación son las mujeres de 15 años y más.

Para la selección de la muestra se utilizó la muestra maestra del Marco Nacional de Viviendas 2012 del INEGI, este último fue 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; el diseño de la muestra maestra es probabilístico, estratificado, unietápico y por conglomerados; estos últimos también se consideran Unidades Primarias de Muestreo (UPM), pues es en ellas donde se seleccionan, en una segunda etapa, las viviendas que integran las muestras de las diferentes encuestas. Las Unidades Primarias de Muestreo (UPM) se forman de la siguiente manera:

4.1 Formación de las unidades primarias de muestreo (UPM)

Las unidades primarias de muestreo (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:

- 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.

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.

4.2 Estratificación

Una vez construido el conjunto de UPM, se agrupan aquellas con características similares, a lo cual se le denomina estratificación.

¹ Área Geoestadística Básica.

² INEGI. Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021. SNIEG. Información de Interés Nacional

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

Ámbito	Descripción
Urbano alto	Ciudades 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 sociodemográficos en los que se agruparon todas las UPM del país, esta estratificación considera las características sociodemográficas de los habitantes de las viviendas, así como las características físicas y el equipamiento de las mismas, expresadas por medio de 34 indicadores² construidos con información del Censo de Población y Vivienda 2010, para lo cual se emplearon métodos estadísticos multivariados.

De esta forma, cada UPM fue clasificada en un único estrato geográfico y uno sociodemográfico. Como resultado, se tienen un total de 683 estratos en todo el ámbito nacional.

4.3 Selección de las UPM de la muestra maestra

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

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

Donde:

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

S = muestra maestra m-ésima.

n_{eh} = número de UPM en la muestra maestra, en el h-ésimo estrato, en la e-ésima entidad.

m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad 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.

Las UPM seleccionadas forman la muestra maestra que permite seleccionar las submuestras de las encuestas en hogares.

5. Tamaño de la muestra

El tamaño de la muestra para la ENDIREH 2021, se calculó tomando como variable de referencia a la proporción de mujeres con violencia de pareja del tipo sexual, la expresión para calcular el tamaño de muestra fue la siguiente.

$$n = \frac{z^2 q DEFF}{r^2 \rho (1 - tnr) P_{VM}}$$

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

Donde:

n = tamaño de la muestra en viviendas.

p = estimación de la proporción de interés.

q = $(1-p)$.

r = error relativo máximo aceptable.

z = valor asentado en las tablas estadísticas 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.

P_{VM} = proporción de viviendas con mujeres de 15 años y más.

Considerando una confianza del 90%, un efecto de diseño de 1.32, una tasa neta de No respuesta máxima esperada del 15%, un error relativo máximo esperado del 15%, una proporción de viviendas con mujeres de 15 años y más del 91%, se podrán estimar proporciones de al menos 4.46%, con un tamaño de muestra de 4 400 viviendas por entidad federativa; para algunas entidades federativas se aplicaron ajustes siendo 4 371 y 4 426 viviendas, los mínimos y máximos, respectivamente.

A nivel nacional resultó el tamaño de muestra en 140 784 viviendas, que nos permite estimar proporciones mínimas de 2.099% relacionadas a la proporción de mujeres con violencia de pareja del tipo sexual en el último año, bajo los parámetros de error relativo máximo esperado de 5.23%, nivel de confianza del 90%, un efecto de diseño de 2.36, una tasa neta de no respuesta máxima esperada del 15% y una proporción de viviendas con mujeres de 15 años y más de 91 por ciento.

6. Afijación de la muestra

La afijación 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:

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

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 2, se presenta la distribución de la muestra, por entidad y ámbito de estudio.

7. Selección de la muestra

La selección de la muestra para la ENDIREH 2021, se realizó de manera independiente por entidad, dominio y estrato, el procedimiento de selección varió de acuerdo con el dominio.

7.1 En urbano alto

1. De las n_{eh} UPM que integran la muestra maestra en el h-ésimo estrato, en la e-ésima entidad se eligieron n_{eh}^* UPM con igual probabilidad para la ENDIREH 2021.
2. En cada UPM se seleccionaron cinco viviendas con igual probabilidad para la ENDIREH 2021.
3. En cada vivienda se seleccionó una mujer de 15 o más años de manera aleatoria.

La probabilidad de seleccionar la i-ésima UPM con probabilidad proporcional al tamaño para la muestra maestra del h-ésimo estrato, de la e-ésima entidad, está dada por:

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

La probabilidad de seleccionar una UPM de la muestra maestra del h-ésimo estrato, de la e-ésima entidad para la ENDIREH 2021 está dada por:

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

La probabilidad de seleccionar una vivienda en la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad está dada por:

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

La probabilidad de selección de una mujer de 15 o más años de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad está dada por:

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

Por lo tanto, la probabilidad total de selección de las viviendas de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es el producto de las probabilidades de selección, definida por:

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

Su factor de expansión³ está dado por:

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

³ El factor de expansión se define como el inverso de la probabilidad de selección.

Donde:

- n_{eh} = número de UPM seleccionadas para la muestra maestra en el h-ésimo estrato, en la e-ésima entidad.
- n_{eh}^* = número de UPM seleccionadas para la ENDIREH 2021, en el h-ésimo estrato, en la e-ésima entidad.
- 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.
- m_{eh}^* = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento de la actualización del listado de viviendas, previo al levantamiento de la ENDIREH 2021.
- Q_{ehij} = número de mujeres de 15 años o más en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

7.2 En complemento urbano

1. De las n_{eh} UPM que integran la muestra maestra en el h-ésimo estrato, en la e-ésima entidad se eligieron n_{eh}^* UPM con igual probabilidad para la ENDIREH 2021.
2. En cada UPM se seleccionaron 20 viviendas con igual probabilidad para la ENDIREH 2021.
3. En cada vivienda se seleccionó una mujer de 15 o más años de manera aleatoria.

La probabilidad de seleccionar la i-ésima UPM con probabilidad proporcional al tamaño para la muestra maestra del h-ésimo estrato, de la e-ésima entidad, está dada por:

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

La probabilidad de seleccionar una UPM de la muestra maestra del h-ésimo estrato, de la e-ésima entidad para la ENDIREH 2021 está dada por:

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

La probabilidad de seleccionar una vivienda de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad está dada por:

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

La probabilidad de selección de una mujer de 15 o más años de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad está dada por:

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

Por lo tanto, la probabilidad total de selección de las viviendas de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es el producto de las probabilidades de selección, definida por:

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

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

Su factor de expansión⁴ está dado por:

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

Donde:

- n_{eh} = número de UPM seleccionadas para la muestra maestra en el h-ésimo estrato, en la e-ésima entidad.
- n_{eh}^* = número de UPM seleccionadas para la ENDIREH 2021, en el h-ésimo estrato, en la e-ésima entidad.
- 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.
- m_{ehi}^* = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento de la actualización del listado de viviendas, previo al levantamiento de la ENDIREH 2021.
- Q_{ehij} = número de mujeres de 15 años o más en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

7.3 En rural

1. De las n_{eh} UPM que integran la muestra maestra en el h-ésimo estrato, en la e-ésima entidad se eligieron n_{eh}^* UPM con igual probabilidad para la ENDIREH 2021.
2. En cada UPM se seleccionaron cuatro segmentos de cinco viviendas con igual probabilidad para la ENDIREH 2021.
3. En cada vivienda se seleccionó una mujer de 15 o más años de manera aleatoria.

La probabilidad de seleccionar la i-ésima UPM con probabilidad proporcional al tamaño para la muestra maestra del h-ésimo estrato, de la e-ésima entidad, está dada por:

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

La probabilidad de seleccionar una UPM de la muestra maestra del h-ésimo estrato, de la e-ésima entidad para la ENDIREH 2021 está dada por:

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

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

La probabilidad de seleccionar una vivienda de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad está dada por:

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

La probabilidad de selección de una mujer de 15 o más años de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad está dada por:

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

Por lo tanto, la probabilidad total de selección de las viviendas de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es el producto de las probabilidades de selección, definida por:

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

$$= \frac{n_{eh} m_{ehi}}{m_{eh}} \frac{n_{eh}^*}{n_{eh}} \frac{4 * 5}{m_{ehi}^*} \frac{1}{Q_{ehij}} = \frac{20 n_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^* Q_{ehij}}$$

Su factor de expansión⁵ está dado por:

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

Donde:

- n_{eh} = número de UPM seleccionadas para la muestra maestra en el h-ésimo estrato, en la e-ésima entidad.
- n_{eh}^* = número de UPM seleccionadas para la ENDIREH 2021, en el h-ésimo estrato, en la e-ésima entidad.
- 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.
- m_{ehi}^* = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento de la actualización del listado de viviendas, previo al levantamiento de la ENDIREH 2021.
- Q_{ehij} = número de mujeres de 15 años o más en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

8. Ajuste a los factores de expansión

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

⁵ El factor de expansión se define como el inverso de la probabilidad de selección.

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, mediante la siguiente expresión:

$$F'_{ehi} = F_{ehi} \frac{\sum_{i \in h} F_{ehi} V_{ehi}}{\sum_{i \in h} F_{ehi} V_{ehi}^*}$$

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 con respuesta en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

8.2 Ajuste por estimación de población

Los factores de expansión ajustados por la No respuesta se corrigieron, con el fin de asegurar que en cada dominio de interés de la encuesta se obtenga el monto de la estimación generada por el INEGI⁶ 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 INEGI.
- $PEXP_D$ = población total a la que expande la encuesta en el dominio D.
- D = nivel de desagregación en que se efectúa el ajuste.

⁶ 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.

9. Estimadores

El estimador del total de la característica X es:

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

Donde:

F_{ehi}^{UA} = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio urbano alto.

$X_{ehij\ell}^{UA}$ = valor observado de la característica de interés X de la ℓ -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.

$X_{ehij\ell}^{CU}$ = valor observado de la característica de interés X de la ℓ -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.

$X_{ehij\ell}^R$ = valor observado de la característica de interés de la ℓ -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, \hat{Y} se define en forma análoga a \hat{X} .

10. Estimación de los errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones estatales y nacionales se usa el método de Conglomerados Últimos⁷, basado en que la mayor contribución a la varianza de un estimador, en un diseño bietápico es la que se presenta entre las Unidades Primarias de Muestreo (UPM). El término Conglomerados Últimos se utiliza para denotar el total de unidades en muestra de una unidad primaria de muestreo.

⁷ Véase Hasen, M. H. Horwitz, W.N. y Madow, W.G., *Sample Survey Methods and Theory*, (1953) Vol. 1 pág. 242.

¹⁰ INEGI. Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021. SNIEG. Información de Interés Nacional

Para obtener las precisiones de los estimadores de razón, conjuntamente al método de Conglomerados Últimos se aplica el método de Series de Taylor, obteniéndose la siguiente fórmula para estimar la precisión de \hat{R} :

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

Donde:

- \hat{X}_{ehi} = total ponderado de la variable de estudio X para 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} = total ponderado de la variable de estudio Y.

Estas definiciones son análogas para la variable de estudio Y.

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

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

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

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

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 nivel $100(1-\alpha)\%$, 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.

11. Homologación de la semaforización para los umbrales de indicadores de precisión estadística

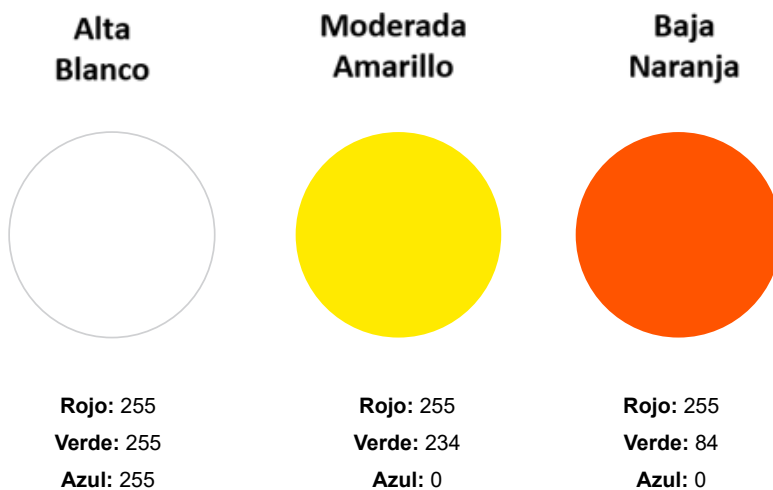
Para facilitar la interpretación de las precisiones estadísticas de la información pública en tabulados, el Comité de Aseguramiento de la Calidad, en la cuarta sesión celebrada el 1 de noviembre de 2018, aprobaron los siguientes umbrales y especificaciones para la publicación en los tabulados los coeficientes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

Umbrales aprobados para la cobertura del CV		
Interpretación	Semaforización	Viviendas/Hogares/Otras unidades diferentes a las económicas DGES/DGEGSPJ
Alta	Blanco	[0%, 15%)
Moderada	Amarillo	[15%, 30%)
Baja	Naranja oscuro	>=30%

Umbrales aprobados para el reporte de la precisión de acuerdo con el coeficiente de variación en los tabulados de resultados de los proyectos con muestreo probabilístico (acuerdo CAC-007/01/2018).

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, coeficiente de variación (CV) e intervalo de confianza. 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 de los colores utilizados en la semaforización:

Parámetros RGB para la semaforización del coeficiente de variación.



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

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

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

Anexo

A. Indicadores empleados en la estratificación de conglomerados 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 años o más de edad que aprobaron algún grado de escolaridad diferente al nivel preescolar.
PP15PRI_CO	De 15 años o más de edad que tienen como máxima escolaridad 6 grados aprobados en primaria.
PP15SEC_CO	De 15 años o más de edad que tienen como máxima escolaridad 3 grados aprobados en secundaria.
PGDO_ESC	Grado promedio de escolaridad.
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.
PTASAOcupa	Tasa de ocupación.
TOCU12A17	No ocupada de 12 a 17 años de edad entre la población de 12 a 17 años de edad.
PPOMAYED	Ocupada de 18 y más años de edad entre la población ocupada.
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 la vivienda 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.
PVDRENERED	Que disponen de drenaje conectado a la red pública.
PVEXCAGU	Que disponen de excusado con descarga directa de agua.
PVPH_CSERV	Que disponen de luz eléctrica, agua entubada dentro o fuera de la vivienda, pero dentro del terreno, así como drenaje.
PSIN_HASIN	Que no se encuentran en situación de hacinamiento a nivel manzana.
Proporción de viviendas particulares habitadas que disponen de:	
PVPH_TV	Televisor.
PVPH_AUTOM	Automóvil o camioneta.
PVPH_CEL	Teléfono celular.
PVCELFIJ	Teléfono celular y teléfono fijo.
PV4ELEC	Radio, televisor, refrigerador y lavadora.
PVRADTEL	Radio y televisor.
PVPHCBIEN	Todos los bienes.

B. Distribución de la muestra en viviendas por entidad federativa según ámbito de estudio para la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021

Clave	Entidad	Distribución de viviendas			
		UA	CU	Rural	Total
1	Aguascalientes	2 980	640	781	4 401
2	Baja California	3 340	720	335	4 395
3	Baja California Sur	1 520	2 280	592	4 392
4	Campeche	2 240	1 120	1 022	4 382
5	Coahuila de Zaragoza	3 220	740	442	4 402
6	Colima	2 680	1 240	489	4 409
7	Chiapas	1 000	1 340	2 031	4 371
8	Chihuahua	3 180	560	654	4 394
9	Ciudad de México	4 360	0	40	4 400
10	Durango	2 360	720	1 311	4 391
11	Guanajuato	1 960	1 140	1 323	4 423
12	Guerrero	1 380	1 280	1 766	4 426
13	Hidalgo	860	1 560	1 980	4 400
14	Jalisco	2 400	1 400	592	4 392
15	México	3 280	600	514	4 394
16	Michoacán de Ocampo	1 140	1 900	1 386	4 426
17	Morelos	2 380	1 340	675	4 395
18	Nayarit	1 500	1 580	1 321	4 401
19	Nuevo León	3 520	640	234	4 394
20	Oaxaca	740	1 400	2 239	4 379
21	Puebla	1 760	1 440	1 197	4 397
22	Querétaro	2 380	800	1 206	4 386
23	Quintana Roo	3 240	700	469	4 409
24	San Luis Potosí	1 960	920	1 529	4 409
25	Sinaloa	2 180	1 080	1 159	4 419
26	Sonora	2 720	1 080	613	4 413
27	Tabasco	820	1 780	1 780	4 380
28	Tamaulipas	3 200	660	560	4 420
29	Tlaxcala	2 520	1 080	817	4 417
30	Veracruz de Ignacio de la Llave	1 480	1 280	1 630	4 390
31	Yucatán	2 280	1 460	662	4 402
32	Zacatecas	1 120	1 500	1 755	4 375
Total		71 700	35 980	33 104	140 784