

Encuesta Nacional sobre Diversidad Sexual y de Género

ENDISEG 2021

Diseño muestral



Instituto Nacional de Estadística y Geografía

Encuesta Nacional sobre Diversidad Sexual y de Género 2021

ENDISEG

Diseño muestral



Catalogación en la fuente INEGI:

304.6021072 Encuesta Nacional sobre Diversidad Sexual y de Género (2021)
Encuesta Nacional sobre Diversidad Sexual y de Género 2021 : ENDISEG :
diseño muestral / Instituto Nacional de Estadística y Geografía.-- México : INEGI,
c2022.

15 p.

1. Comportamiento sexual - México - Estadísticas. 2. Rol sexual - México -
Encuestas. 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)** presenta el documento metodológico en el que se exponen las principales características de la **Encuesta Nacional sobre Diversidad Sexual y Género (ENDISEG) 2021**.

En este documento, correspondiente al **Diseño muestral**, se describen las actividades que constituyen el proceso de obtención y tratamiento estadístico de la muestra, que inicia con la determinación de la cobertura y dominios de interés, para los cuales se pueden obtener estimaciones estadísticamente confiables, en función de los cuales se calcula el tamaño de la muestra y se determina su distribución, seguida de la selección de las unidades de muestreo.

Como parte del tratamiento, se describe el cálculo de las probabilidades de selección y su empleo para asignar el peso que debe tener cada unidad seleccionada. Finalmente se presentan las expresiones empleadas para obtener las estimaciones y sus medidas de calidad.

Índice

Diseño muestral

1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Marco de la encuesta	1
4.1 Formación de las unidades primarias de muestreo (UPM)	1
4.1.1 En urbano alto	1
4.1.2 En complemento urbano	2
4.1.3 En rural	2
4.2 Estratificación	2
4.3 Selección de las UPM de la muestra maestra	3
5. Tamaño de la muestra	3
6. Distribución de la muestra	4
7. Selección de la muestra	4
7.1 En urbano alto	4
7.2 En complemento urbano	5
7.3 En rural	6
8. Ajuste de los factores de expansión	8
8.1 Ajuste por no respuesta	8
8.1.1 Ajuste por no respuesta a nivel vivienda	8
8.1.2 Ajuste por no respuesta a nivel persona	8
8.2 Ajuste por proyección	9
9. Estimadores	9
10. Estimación de errores de muestreo	9
11. Homologación de la Semaforización para los Umbrales de Indicadores de precisión estadística	11

Anexo

A. Indicadores empleados en la estratificación de la muestra maestra.	14
B. Distribución de la muestra de UPM y viviendas seleccionadas para la Encuesta Nacional Sobre Diversidad Sexual y Género (ENDISEG) 2021.	15

Diseño muestral

1. Objetivo de la encuesta

Identificar de la población de 15 años y más, aquella que se reconoce a sí misma con orientación sexual y/o identidad de género (OSIG) LGBTI+, esto es, no normativa o no convencional, así como conocer sus principales características, tales como: sexualidad, educación, empleo, servicios de salud, salud emocional, satisfacción personal, entre otros.

2. Población objetivo

La encuesta está dirigida a la población de 15 años y más de edad que reside permanentemente en viviendas particulares dentro del territorio nacional.

3. Cobertura geográfica

La encuesta está diseñada para proporcionar información con representatividad nacional.

4. Marco de la encuesta

El diseño de la muestra para la ENDISEG 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 es el hogar.

Para la selección de la muestra se utilizó el Marco Nacional de Viviendas 2012 del INEGI, construido a partir de la información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2010. Esta muestra es una muestra maestra a partir de la cual se seleccionan las submuestras para todas las encuestas en viviendas que realiza el INEGI; su diseño es probabilístico, estratificado, unietápico y por conglomerados; estos últimos también se consideran unidades primarias de muestreo, pues es en ellos donde se seleccionan, en una segunda etapa, las viviendas que integran las muestras de las diferentes encuestas. La muestra maestra se construye de la siguiente manera:

4.1 Formación de las unidades primarias de muestreo (UPM)

Primeramente, se construye el conjunto de UPM que cubrirá el territorio nacional.

Las unidades primarias de muestreo están constituidas por agrupaciones de viviendas con características diferenciadas dependiendo del ámbito al que pertenecen, como se especifica a continuación:

4.1.1 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.

4.1.2 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.

4.1.3 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 está conformada con un total de 240 912 UPM a nivel nacional.

4.2 Estratificación

Una vez construido el conjunto de UPM, se agrupan 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 estratificación geográfica. En cada entidad federativa se distinguen tres ámbitos, divididos a su vez en zonas, como se indica en el siguiente cuadro:

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 indicadores* construidos con información del Censo de Población y Vivienda 2010, para lo cual se emplearon métodos estadísticos multivariados.

¹ Área Geoestadística Básica.

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

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 = la muestra maestra.

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

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

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

Las UPM seleccionadas forman la muestra maestra que permite seleccionar las submuestras de las encuestas en hogares en las siguientes etapas del diseño estadístico.

5. Tamaño de la muestra

El tamaño de la muestra se calculó empleando la siguiente expresión:

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

Donde:

n = tamaño de la muestra.

p = estimación de la proporción de interés.

q = $1 - p$.

r = error relativo máximo aceptable.

z = valor asentado en las tablas estadísticas de la distribución normal estándar para una confianza prefijada.

DEFF = efecto de diseño definido como el cociente de la varianza en la estimación del diseño utilizado, entre la varianza obtenida considerando un muestreo aleatorio simple para un mismo tamaño de muestra.

tnr = tasa de no respuesta máxima esperada.

Considerando una confianza del 90%, un efecto de diseño de 4.67, un error relativo máximo esperado del 7.43%, una tasa de no respuesta máxima esperada del 15%, para una proporción de población de 15 años y más de edad que se autoidentifica como LGBTI+ de 5.1%, se obtuvo un tamaño de muestra de 50 113 viviendas, mismo que se ajustó a 50 258 viviendas a nivel nacional.

6. Distribución de la muestra

La distribución de la muestra se realizó dentro de cada entidad federativa, por tamaño de localidad y estrato de manera proporcional a su tamaño, para lo cual se empleó la siguiente expresión:

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

Donde:

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

n_e = número total de viviendas en muestra en la e-ésima entidad.

N_{eh} = número total de viviendas en el h-ésimo estrato, en la e-ésima entidad.

N_e = número total de viviendas en la e-ésima entidad.

En el anexo B, se presenta la distribución de la muestra por UPM y viviendas seleccionadas.

7. Selección de la muestra

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

7.1 En urbano alto

1. De las k_{eh} UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad para la muestra maestra, se seleccionaron k_{eh}^* con igual probabilidad para la ENDISEG 2021.
2. En cada UPM seleccionada, se seleccionaron cinco viviendas con igual probabilidad.
3. En cada vivienda seleccionada se seleccionó una persona de 15 años y más de edad con igual probabilidad.

La probabilidad de selección de las viviendas y personas se calculó de la siguiente manera:

- a) La probabilidad de selección de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra maestra está definida por:

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

- b) La probabilidad de seleccionar en la muestra maestra a la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra de la ENDISEG 2021 está definida por:

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

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

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

d) La probabilidad de seleccionar una persona de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es:

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

Por lo tanto, la probabilidad total de selección de una persona en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad es el producto de las probabilidades de selección:

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

Su factor de expansión está dado por:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad para la muestra maestra.
- m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según 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 ENDISEG 2021 con igual probabilidad de selección, en el h-ésimo estrato, en la e-ésima entidad.
- Q_{ehij} = número de personas de 15 años y más de edad, 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 k_{eh} UPM que integran la muestra maestra se seleccionaron k_{eh}^* UPM con igual probabilidad para la ENDISEG 2021.
2. En cada UPM seleccionada, se seleccionaron 20 viviendas con igual probabilidad.
3. En cada vivienda seleccionada se seleccionó una persona de 15 años y más de edad con igual probabilidad.

La probabilidad de selección de las viviendas y personas se calculó de la siguiente manera:

- a) La probabilidad de selección de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra maestra está definida por:

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

b) La probabilidad de seleccionar en la muestra maestra a la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la muestra de la ENDISEG 2021 está definida por:

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

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

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

d) La probabilidad de seleccionar una persona de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad es:

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

Por lo tanto, la probabilidad total de selección de una persona en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad es el producto de las probabilidades de selección:

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

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

Su factor de expansión está dado por:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad para la muestra maestra.
- m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según 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 ENDISEG 2021 con igual probabilidad de selección, en el h-ésimo estrato, en la e-ésima entidad.
- Q_{ehij} = número de personas de 15 años y más de edad, 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 k_{eh} UPM se seleccionaron k_{eh}^* UPM para la ENDISEG 2021 con igual probabilidad.
2. En cada UPM seleccionada, se seleccionaron cuatro segmentos de cinco viviendas aproximadamente con igual probabilidad.
3. En cada vivienda seleccionada se seleccionó una persona de 15 años y más de edad con igual probabilidad.

La probabilidad de selección de las viviendas y personas se calculó de la siguiente manera:

- a) La probabilidad de selección de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad para la muestra maestra está definida por:

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

- b) La probabilidad de seleccionar en la muestra maestra a la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad para la muestra de la ENDISEG 2021 está definida por:

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

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

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

- d) La probabilidad de seleccionar una persona de la j -ésima vivienda, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad es:

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

Por lo tanto, la probabilidad total de selección de una persona en la j -ésima vivienda, en la i -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad es el producto de las probabilidades de selección:

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

Su factor de expansión² está dado por:

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

Donde:

- k_{eh} = número de UPM seleccionadas en el h -ésimo estrato, en la e -ésima entidad, para la muestra maestra.
- m_{eh} = número de viviendas en el h -ésimo estrato, en la e -ésima entidad.
- m_{ehi} = número de viviendas en la i -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad, según Censo de Población y Vivienda 2010.
- m_{ehi}^* = número de viviendas en la i -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad, según listado de viviendas actualizado.
- k_{eh}^* = número de UPM seleccionadas para la ENDISEG 2021, en el h -ésimo estrato, en la e -ésima entidad.

² El factor de expansión está definido como el inverso de la probabilidad de selección.

Q_{ehij} = número de personas de 15 años y más de edad, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

8. Ajuste de los factores de expansión

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

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 UPM, en cada uno de los dominios a nivel vivienda y a nivel persona:

8.1.1 Ajuste por no respuesta 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, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.

F_{ehij} = factor de expansión de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.

I_{ehij} = función indicadora con valor 1 o 0; esta función toma el valor 1 si una vivienda seleccionada en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad tiene respuesta completa, y toma el valor de cero 0 en caso contrario.

8.1.2 Ajuste por no respuesta a nivel persona

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

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

Donde:

F''_{ehijk} = factor de expansión corregido por no respuesta de la k-ésima vivienda seleccionada, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.

F'_{ehijk} = factor de expansión para la k-ésima vivienda seleccionada, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.

I_{ehijk} = función indicadora con valor 1 o 0; esta función toma el valor 1 si una vivienda seleccionada en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad tiene respuesta completa, y toma el valor de cero 0 en caso contrario.

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 = es el nivel de desagregación de la subpoblación en que se hace el ajuste por estimación.

9. Estimadores

El estimador del total de la característica X es:

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

Donde:

- F_{ehis}^{15+} = factor de expansión final de la s-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, en las localidades de 15,000 y más habitantes.
- $X_{ehis\ell}^{15+}$ = valor observado de la característica de interés X en la ℓ -ésima persona, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en las localidades de 15,000 y más habitantes.
- F_{ehis}^{15-} = factor de expansión final de la s-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, en las localidades menores de 15,000 habitantes.
- $X_{ehis\ell}^{15-}$ = valor observado de la característica de interés X de la ℓ -ésima persona, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en las localidades menores de 15,000 habitantes.

Para la estimación de proporciones, tasas y promedios se utiliza el estimador de razón:

$$\hat{R} = \frac{\hat{X}}{\hat{Y}}$$

Donde, \hat{Y} se define en forma análoga a \hat{X} .

10. Estimación de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones, se usó el método de Conglomerados Últimos³, basado en que la mayor contribución a la varianza de un estimador, en un diseño multietápico, 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.

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

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

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

Donde:

\hat{X}_{ehi} = total ponderado de la variable de estudio X en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

\hat{X}_{eh} = total ponderado de la variable de estudio X en el h-ésimo estrato, en la e-ésima entidad.

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

L_e = número de estratos en la e-ésima entidad.

\hat{Y}^2 = el cuadrado del total ponderado de la característica Y.

Estas definiciones son análogas para la variable de estudio Y.

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

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

Las estimaciones del error estándar (EE), efecto de diseño (DEFF) y coeficiente de variación (CV) se calculan mediante las siguientes expresiones:

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

Donde:

$\hat{\theta}$ = estimador del parámetro poblacional θ .

$\hat{V}(\hat{\theta})_{MAS}$ = estimador de la varianza bajo muestreo aleatorio simple.

$\hat{V}(\hat{\theta})$ = estimador de la varianza bajo el diseño de muestreo descrito en este documento.

Finalmente, el intervalo de confianza $I_{1-\alpha}$ al 100 (1- α)%, se calcula 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 estadística⁴

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

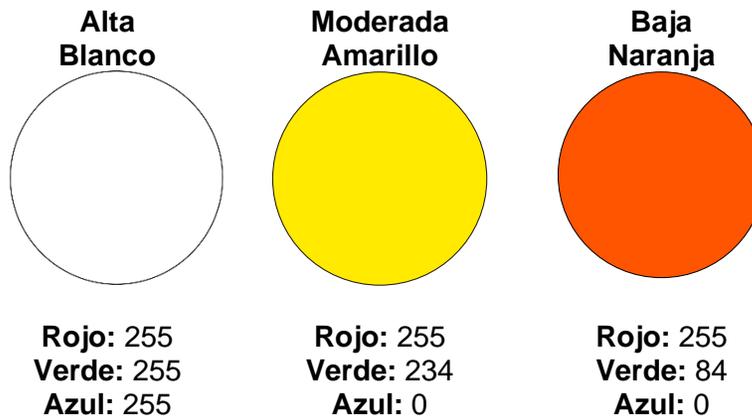
Umbrales aprobados para la cobertura del CV

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, 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 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

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

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 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 y más de edad que aprobaron algún grado de escolaridad diferente al nivel preescolar.
PP15PRI_CO	De 15 años y más de edad que tienen como máxima escolaridad 6 grados aprobados en primaria.
PP15SEC_CO	De 15 años y 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.
PVPH3YMASC	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.
PVDREDED	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 para la Encuesta Nacional Sobre Diversidad Sexual y Género (ENDISEG) 2021

Entidad		Muestra	
Clave	Nombre	UPM	Viviendas
Estados Unidos Mexicanos		6 374	50 258
01	Aguascalientes	239	1 566
02	Baja California	257	1 569
03	Baja California Sur	161	1 569
04	Campeche	200	1 570
05	Coahuila de Zaragoza	251	1 573
06	Colima	221	1 567
07	Chiapas	134	1 570
08	Chihuahua	248	1 569
09	Ciudad de México	312	1 569
10	Durango	206	1 574
11	Guanajuato	185	1 569
12	Guerrero	152	1 573
13	Hidalgo	125	1 575
14	Jalisco	206	1 570
15	México	257	1 569
16	Michoacán de Ocampo	140	1 572
17	Morelos	206	1 575
18	Nayarit	161	1 572
19	Nuevo León	269	1 571
20	Oaxaca	119	1 575
21	Puebla	173	1 572
22	Querétaro	206	1 571
23	Quintana Roo	254	1 572
24	San Luis Potosí	185	1 571
25	Sinaloa	197	1 570
26	Sonora	224	1 570
27	Tabasco	125	1 568
28	Tamaulipas	248	1 569
29	Tlaxcala	215	1 568
30	Veracruz de Ignacio de la Llave	158	1 575
31	Yucatán	200	1 572
32	Zacatecas	140	1 563

Título: Encuesta Nacional sobre Diversidad Sexual y de Género 2021.ENDISEG. Diseño muestral.

FE DE ERRATAS

Página	Apartado	Dice	Debe decir	Fecha de modificación
Página 1	4. Marco de la Encuesta Primer párrafo al final	...donde la unidad última de selección es la vivienda y la unidad de observación es el hogar.	...donde la unidad última de selección es la persona.	05 de julio del 2022.