

Encuesta Nacional de Ingresos y Gastos de los Hogares 2020

ENIGH
Nueva serie

Diseño muestral



 **INEGI**

Instituto Nacional de Estadística y Geografía

Encuesta Nacional de Ingresos y Gastos de los Hogares 2020

ENIGH

**Nueva serie
Diseño muestral**



Obras complementarias publicadas por el INEGI sobre el tema:

Encuesta Nacional de Ingresos y Gastos de los Hogares 2018 ENIGH, Nueva serie, Diseño muestral.

Catalogación en la fuente INEGI:

640.4201 Encuesta Nacional de Ingresos y Gastos de los Hogares (2020).
Encuesta Nacional de Ingresos y Gastos de los Hogares 2020 : ENIGH : nueva serie : diseño muestral / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2021.

18 p.

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

Conociendo México

01 800 111 4634

www.inegi.org.mx

atencion.usuarios@inegi.org.mx



INEGI Informa



@INEGI_INFORMA

DR © 2021, **Instituto Nacional de Estadística y Geografía**

Edificio Sede

Avenida Héroe de Nacozari Sur 2301

Fraccionamiento Jardines del Parque, 20276 Aguascalientes,

Aguascalientes, Aguascalientes, entre la calle INEGI,

Avenida del Lago y Avenida Paseo de las Garzas.

Presentación

El **Instituto Nacional de Estadística y Geografía (INEGI)** presenta diversos documentos metodológicos en los que se exponen las principales características de la **Encuesta Nacional de Ingresos y Gastos de los Hogares 2020 (ENIGH 2020)**.

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
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	5
7. Selección de la muestra	5
7.1 En urbano alto	5
7.2 En complemento urbano	7
7.3 En rural	8
8. Ajuste a los factores de expansión	9
8.1 Ajuste por no respuesta	9
8.2 Ajuste por estimación de población	9
9. Estimadores	10
10. Estimaciones de errores de muestreo	10
11. Homologación de la Semaforización para los Umbrales de Indicadores de precisión estadística	11
Anexo	13
A. Indicadores empleados en la estratificación de UPM de la muestra maestra.	15
B. Relación de las zonas dentro del ámbito urbano alto.	16
C. Distribución de la muestra en viviendas por entidad según dominio de estudio para la Encuesta Nacional de Ingresos y Gastos de los Hogares 2020 (ENIGH-2020)	18

Diseño muestral

1. Objetivo de la encuesta

El objetivo de la ENIGH-2020 es obtener información estadística sobre los ingresos y gastos de los hogares, así como las actividades económicas desarrolladas por los miembros de los hogares.

2. Población objetivo

La encuesta está dirigida a los hogares del territorio nacional.

3. Cobertura geográfica

La encuesta está diseñada para dar resultados a nivel nacional y entidad federativa, con cortes urbano y rural.

4. Marco de la encuesta

El diseño de la submuestra para la ENIGH-2020 se caracteriza por ser probabilístico; en consecuencia, los resultados obtenidos de la encuesta se generalizan a toda la población de dominio estudio, a su vez es bietápico, estratificado y por conglomerados, donde la unidad última de selección es la vivienda y la unidad de observación es el hogar.

La submuestra de ENIGH-2020 se seleccionó a partir de la muestra maestra 2012 del INEGI, esta muestra maestra se diseñó y seleccionó del Marco Maestro de Muestreo 2012 (MMM) el cual se conformó de conglomerados de viviendas llamados Unidades Primarias de Muestreo (UPM), construidos a partir de la información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2010. La muestra maestra permite la selección de submuestras para todas las encuestas en viviendas que realiza el INEGI; su diseño es probabilístico, estratificado, unietápico y por conglomerados, pues es en ellos donde se seleccionaron, en una segunda etapa, las viviendas que integran las submuestras de las diferentes encuestas. En diseño del MMM se construyó 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.

4.2 Estratificación

Una vez construido el conjunto de UPM, estas se agrupan considerando sus características similares, o sea se estratifican.

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	Localidades de 50 000 a 99 999 habitantes
	35	Localidades de 15 000 a 49 999 habitantes
	45	Localidades de 5 000 a 14 999 habitantes
	55	Localidades de 2 500 a 4 999 habitantes Rural
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 zonas del ámbito urbano alto se muestra en el Anexo B.

** 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:

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

Donde:

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

S_m = la muestra maestra.

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

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

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

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

5. Tamaño de la muestra

Como resultado del ejercicio de consulta pública, se dio revisión a los criterios para el reforzamiento de la muestra de ENIGH 2020, estableciendo los siguientes:

1. Se partió del tamaño de muestra de la ENIGH 2018, como cota mínima, para cada entidad federativa por ámbito urbano y rural.
2. Se aplicó una Tasa de no respuesta diferenciada por entidad y ámbito (urbano, rural). Partiendo de la tasa de no respuesta de la ENIGH 2018, se consideró una tasa de no respuesta mínima del 15%. En aquellos casos con tasa de no respuesta superior al 15%, se consideraron los redondeos (hacia arriba o hacia abajo) a 15, 20 y 25%.
 - a) Tasa de no respuesta del 20%:
Ámbito urbano: Campeche, Coahuila de Zaragoza, Colima, Chihuahua, Durango, Guerrero, Hidalgo, México, Puebla, Quintana Roo, Tabasco, Tamaulipas, Tlaxcala y Veracruz.
Ámbito rural: Chihuahua.
 - b) Tasa de no respuesta del 25%:
Ámbito urbano: Jalisco, Nuevo León, Querétaro, San Luis Potosí y Yucatán.
Ámbito rural: Ciudad de México.
 - c) El resto de las entidades y ámbitos permanece en 15 por ciento.

Además de la tasa de no respuesta, el tamaño de muestra se incrementó por ámbito (urbano, rural), considerando el tamaño de muestra mínimo requerido para estimar el promedio del ingreso corriente trimestral por hogar, el promedio del ingreso por trabajo trimestral por hogar, así como el promedio del gasto corriente trimestral por hogar, con base en la ENIGH 2018, una confianza del 90%, un margen de error relativo del 15% y una Tasa de no respuesta de acuerdo al punto 2.

3. Las entidades que requirieron incrementos de muestra fueron:

Ámbito urbano: Zacatecas.

Ámbito rural: Baja California Sur, Chihuahua, Ciudad de México, Durango, Guanajuato, Guerrero, Oaxaca, Querétaro, San Luis Potosí, Tamaulipas y Veracruz.

El ámbito rural de Oaxaca permite estimar el promedio del ingreso corriente trimestral por hogar y el promedio del gasto corriente trimestral por hogar. Sin embargo, en el caso de la variable promedio del ingreso por trabajo por hogar, se requiere una muestra superior a 3 000 viviendas, las cuales no es posible cubrir debido a la insuficiencia del marco de muestreo.

El cálculo del tamaño de muestra de la ENIGH-2020 se calculó para todas las variables de interés de la encuesta, pero para describir el procedimiento del cálculo solo se presenta para el promedio del ingreso corriente total por hogar. La expresión utilizada fue la siguiente:

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

Donde:

n = tamaño de la muestra.

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

s^2 = estimación de la varianza poblacional de la variable de interés.

\bar{x} = estimación del promedio de la variable de interés.

$DEFF$ = 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.

r = error relativo máximo aceptable.

tnr = tasa de no respuesta máxima esperada.

PHV = promedio de hogares por vivienda.

El tamaño de muestra para la ENIGH 2020 se calculó a nivel Entidad con ámbito urbano y rural considerando las variables y las tasas de no respuesta mencionadas anteriormente.

A nivel Entidad con ámbito urbano para el caso de la variable Ingreso Corriente promedio trimestral se tiene una variación entre 31 554.58 y 91 003.53, con una varianza que oscila entre 609 706 543.70 y 570 110 356 234.59, y un efecto de diseño que fluctúa entre 1.09 y 4.24.

A nivel Entidad con ámbito rural la variable Ingreso Corriente promedio trimestral varía entre 14 115.33 y 44 778.03, con una varianza que oscila entre 156 406 519.01 y 12 108 216 477.28, y un efecto de diseño que fluctúa entre 1.00 y 9.43. En el anexo C se presentan los tamaños de muestra para ámbito Entidad urbano y rural.

Integrando los tamaños de muestra a nivel nacional se tiene un tamaño de muestra de 87 826 viviendas, el cual garantiza un error de 4.485% a nivel nacional para la variable ingreso corriente promedio trimestral.

Debido a la apertura para salir a campo para realizar entrevistas cara a cara en las viviendas habitadas seleccionadas durante la época de la pandemia COVID-19, se tuvo indicios de tener un mayor rechazo de los hogares a aceptar las entrevistas, se decidió tomar medidas precautorias de incrementar la tasa de no respuesta hasta tener un 20% adicional, debido a estos ajustes la muestra resultante fue de 105 483 viviendas a nivel nacional.

6. Distribución de la muestra

La distribución de la muestra se realizó dentro de cada entidad federativa, de manera proporcional al número de viviendas de los estratos, 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 promedio de viviendas a seleccionar por UPM.

En el anexo C, se presenta la distribución de la muestra en viviendas, por entidad según dominio de estudio interés.

7. Selección de la muestra

La selección de la muestra fue bietápica, donde las unidades de selección de segunda etapa fueron las viviendas y la unidad de observación el hogar. Se hizo una selección independiente para cada entidad y estrato, el procedimiento varió dependiendo del ámbito, como se especifica a continuación:

7.1 En urbano alto

En el ámbito urbano alto la selección de la muestra se realizó en forma independiente por cada entidad y estrato mediante el siguiente procedimiento:

1. De las k_{eh} UPM que integran la muestra maestra en el h-ésimo estrato, en la e-ésima entidad se eligieron k_{eh}^* UPM con igual probabilidad para la ENIGH-2020.

2. En cada UPM se seleccionaron seis viviendas con igual probabilidad para la ENIGH-2020.

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

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

La probabilidad de selección de la i -ésima UPM_{ehi} , del h -ésimo estrato, de la e -ésima entidad para la submuestra de la ENIGH 2020 es:

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

La probabilidad de seleccionar a cada vivienda de la i -ésima, UPM_{ehi} del h -ésimo estrato, de la e -ésima entidad es:

$$P_{3ehi} = \frac{6}{m_{ehi}^*}$$

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

$$\begin{aligned} P_{ehi} &= P_{1ehi} * P_{2ehi} * P_{3ehi} \\ &= \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{k_{eh}^*}{k_{eh}} \frac{6}{m_{ehi}^*} = \frac{6 k_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^*} \end{aligned}$$

Y su factor de expansión² está dado por:

$$F_{ehi} = \frac{m_{eh} m_{ehi}^*}{6 k_{eh}^* m_{ehi}}$$

Donde:

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

k_{eh}^* = número de UPM seleccionadas para la ENIGH-2020, 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, 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 Censo de Población y Vivienda 2010.

² El factor de expansión se define como el inverso de la probabilidad de selección.

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 ENIGH-2020.

7.2 En complemento urbano

1. De las k_{eh} UPM que integran la muestra maestra en el h-ésimo estrato, en la e-ésima entidad, se eligieron k_{eh}^* UPM con igual probabilidad para la ENIGH-2020.
2. En cada UPM seleccionada, se eligieron 24 viviendas con igual probabilidad.

La probabilidad de seleccionar la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad con probabilidad proporcional al tamaño para la muestra maestra es:

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

La probabilidad de selección de la i-ésima UPM_{ehi} , del h-ésimo estrato, de la e-ésima entidad para la submuestra de la ENIGH 2020 es:

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

La probabilidad de seleccionar a cada vivienda de la i-ésima UPM_{ehi} , del h-ésimo estrato, de la e-ésima entidad es:

$$P_{3ehi} = \frac{24}{m_{ehi}^*}$$

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

$$\begin{aligned} P_{ehi} &= P_{1ehi} * P_{2ehi} * P_{3ehi} \\ &= \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{k_{eh}^*}{k_{eh}} \frac{24}{m_{ehi}^*} = \frac{24 k_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^*} \end{aligned}$$

Y su factor de expansión está dado por:

$$F_{ehi} = \frac{m_{eh} m_{ehi}^*}{24 k_{eh}^* m_{ehi}}$$

Donde:

k_{eh} = número de UPM seleccionadas para 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.

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

k_{eh}^* = número de UPM seleccionadas para la ENIGH-2020, en el h-ésimo estrato, en la e-ésima entidad.

m_{ehi}^* = número total de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento de la actualización de los listados, previo al levantamiento de la ENIGH-2020.

7.3 En rural

1. De las k_{eh} UPM que integran la muestra maestra en el h-ésimo estrato, en la e-ésima entidad, se eligieron k_{eh}^* UPM con igual probabilidad para la ENIGH-2020.
2. En cada UPM seleccionada, se eligieron, 24 viviendas con igual probabilidad.

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

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

La probabilidad de selección de la i-ésima UPM_{ehi}, del h-ésimo estrato, de la e-ésima entidad para la submuestra de la ENIGH 2020 es:

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

La probabilidad de seleccionar a cada vivienda de la i-ésima UPM_{ehi}, del h-ésimo estrato, de la e-ésima entidad es:

$$P_{3ehi} = \frac{24}{m_{ehi}^*}$$

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

$$\begin{aligned} P_{ehi} &= P_{1ehi} * P_{2ehi} * P_{3ehi} \\ &= \frac{k_{eh} m_{ehi}}{m_{eh}} \frac{k_{eh}^*}{k_{eh}} \frac{24}{m_{ehi}^*} = \frac{24 k_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^*} \end{aligned}$$

En consecuencia, su factor de expansión está dado por:

$$F_{ehi} = \frac{m_{eh} m_{ehi}^*}{24 k_{eh}^* m_{ehi}}$$

Donde:

k_{eh} = número de UPM seleccionadas para 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, según Censo de Población y Vivienda 2010.

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

k_{eh}^* = número de UPM seleccionadas para la ENIGH-2020, en el h-ésimo estrato, en la e-ésima entidad.

m_{ehi}^* = número total de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento de la actualización de los listados, previo al levantamiento de la ENIGH-2020.

8. Ajuste a los factores de expansión

Los factores de expansión elaborados conforme al procedimiento antes descrito se ajustan para los siguientes conceptos:

8.1 Ajuste por no respuesta

El ajuste por no respuesta atribuida al informante se realiza 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 ajustados por no respuesta se corrigen, a fin de asegurar que en cada dominio de interés de la encuesta se obtenga la población total determinada por la estimación de población determinada por el INEGI³, la cual está referida al punto medio del levantamiento, mediante la siguiente expresión:

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

³ 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 y derivó en un reprocesamiento de la información publicada en la ENIGH.

Donde:

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

9. Estimadores

El estimador del total de la característica X, a nivel nacional es:

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

Donde:

- F_{ehi}^{UA} = factor de expansión final, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el ámbito urbano alto.
- $X_{ehis\ell}^{UA}$ = valor observado de la característica de interés X 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 ámbito urbano alto.
- F_{ehi}^{CU} = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, del ámbito complemento urbano.
- $X_{ehis\ell}^{CU}$ = valor observado de la característica X 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 ámbito complemento urbano.
- F_{ehi}^R = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad del ámbito rural.
- $X_{ehis\ell}^R$ = valor observado de la característica X 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 ámbito 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} .

10. Estimación de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones se usa el método de “Conglomerados Últimos”,⁴ basado en que la mayor contribución a la varianza de un estimador, en un diseño polietápico, 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.

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

⁴ Véase Hansen, M.H. Horwitz, W.N. y Madow, W.G. *Sample Survey Methods and Theory*, (1953), Vol. 1 página 242.

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

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 = el cuadrado del estimador del total 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:

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

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

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

Donde:

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

$\hat{V}(\hat{\theta})_{mas}$ = estimador de la varianza, bajo un 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.

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 sección 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.

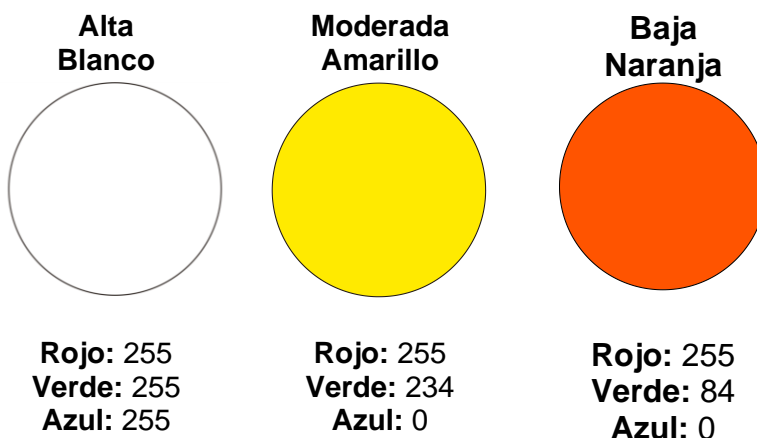
Umbrales aprobados para la cobertura de la CV

Interpretación	Semaforización	Intervalo del CV
Alta	Blanco	[0%, 15%)
Moderada	Amarillo	[15%, 30%)
Baja	Naranja oscuro	$\geq 30\%$

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 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 los conglomerados de la Muestra Maestra

Mnemónico	Descripción
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.
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.
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.
Viviendas Particulares Habitadas que disponen de:	
PVPH_TV	Televisor.
PVPH_AUTOM	Automóvil o camioneta.
PVPH_CEL	Teléfono celular.
PVCELFIJ	Teléfono celular y teléfono fijo.
PV4ELEC	Radio, televisor, refrigerador y lavadora.
PVRADTEL	Radio y televisor.
PVPHCBIE	Todos los bienes.

B. Relación de las zonas dentro del ámbito urbano alto

Clave	Entidad	Zona	Nombre de la zona
01	Aguascalientes	01	Aguascalientes
		01	Tijuana
02	Baja California	02	Mexicali
		03	Ensenada
		01	La Paz
03	Baja California Sur	01	Campeche
04	Campeche	02	Ciudad del Carmen
05	Coahuila de Zaragoza	01	Saltillo
		02	Torreón
		03	Monclova
		04	Piedras Negras
		05	Ciudad Acuña
		01	Colima
06	Colima	02	Manzanillo
07	Chiapas	01	Tuxtla Gutiérrez
		02	Tapachula
		03	San Cristóbal de las Casas
08	Chihuahua	01	Ciudad Juárez
		02	Chihuahua
		03	Delicias
		04	Hidalgo del Parral
		05	Cuauhtémoc
		01	Ciudad de México
09	Ciudad de México	01	Durango
10	Durango	02	Torreón
11	Guanajuato	01	León
		02	Irapuato
		03	Celaya
		04	Salamanca
		05	Guanajuato
		01	Acapulco
12	Guerrero	02	Chilpancingo
		03	Iguala
		01	Pachuca
13	Hidalgo	02	Tulancingo
		01	Guadalajara
14	Jalisco	02	Puerto Vallarta
		01	Ciudad de México
15	México	02	Toluca
		01	Morelia
16	Michoacán de Ocampo	02	Uruapan
		03	Zamora
		01	Cuernavaca
17	Morelos	02	Cuautla
18	Nayarit	01	Tepic
19	Nuevo León	01	Monterrey
20	Oaxaca	01	Oaxaca
		05	San Juan Bautista Tuxtepec
21	Puebla	01	Puebla
		02	Tehuacán
22	Querétaro	01	Querétaro
		02	San Juan del Río

B. Relación de las zonas dentro del ámbito urbano alto

23	Quintana Roo	01	Cancún
		02	Chetumal
		03	Playa del Carmen
24	San Luis Potosí	01	San Luis Potosí
		02	Ciudad Valles
25	Sinaloa	01	Culiacán Rosales
		02	Mazatlán
		03	Los Mochis
26	Sonora	01	Hermosillo
		02	Ciudad Obregón
		03	Heroica Nogales
		04	San Luis Río Colorado
		05	Guaymas
		06	Navojoa
27	Tabasco	01	Villahermosa
28	Tamaulipas	01	Tampico
		02	Reynosa
		03	Matamoros
		04	Nuevo Laredo
		05	Ciudad Victoria
29	Tlaxcala	01	Tlaxcala
30	Veracruz de Ignacio de la Llave	01	Veracruz
		02	Xalapa
		03	Orizaba
		04	Coatzacoalcos
		05	Poza Rica
		06	Córdoba
		07	Minatitlán
		08	Tuxpan
		09	Tampico
31	Yucatán	01	Mérida
32	Zacatecas	01	Zacatecas
		02	Fresnillo

C. Distribución de la muestra en viviendas por entidad según dominio de estudio para la Encuesta Nacional de Ingresos y Gastos de los Hogares 2020 (ENIGH-2020)

Clave	Entidad		Dominio	ENIGH nueva serie 2018	ENIGH nueva serie 2020
		Nombre			
Estados Unidos Mexicanos			Nacional	87 826	105 483
			Urbano	56 052	67 583
			Rural	31 774	37 900
01	Aguascalientes		Entidad	2 640	3 170
			Urbano	1 698	2 040
			Rural	942	1 130
02	Baja California		Entidad	4 035	4 854
			Urbano	1 860	2 304
			Rural	2 175	2 550
03	Baja California Sur		Entidad	2 737	3 262
			Urbano	1 770	2 124
			Rural	967	1 138
04	Campeche		Entidad	2 099	2 513
			Urbano	1 620	1 944
			Rural	479	569
05	Coahuila de Zaragoza		Entidad	4 146	4 987
			Urbano	3 378	4 056
			Rural	768	931
06	Colima		Entidad	3 270	3 927
			Urbano	1 980	2 544
			Rural	1 290	1 383
07	Chiapas		Entidad	2 002	2 411
			Urbano	1 068	1 284
			Rural	934	1 127
08	Chihuahua		Entidad	4 521	5 422
			Urbano	2 802	3 390
			Rural	1 719	2 032
09	Ciudad de México		Entidad	2 485	2 990
			Urbano	1 980	2 376
			Rural	505	614
10	Durango		Entidad	2 646	3 175
			Urbano	1 494	1 794
			Rural	1 152	1 381
11	Guanajuato		Entidad	2 856	3 450
			Urbano	1 404	1 686
			Rural	1 452	1 764
12	Guerrero		Entidad	2 420	2 918
			Urbano	1 290	1 548
			Rural	1 130	1 370
13	Hidalgo		Entidad	2 119	2 548
			Urbano	1 188	1 428
			Rural	931	1 120

C. Distribución de la muestra en viviendas por entidad según dominio de estudio para la Encuesta Nacional de Ingresos y Gastos de los Hogares 2020 (ENIGH-2020)

Clave	Entidad		Dominio	ENIGH nueva serie 2018	ENIGH nueva serie 2020
		Nombre			
14	Jalisco		Entidad	2 679	3 210
			Urbano	1 980	2 376
			Rural	699	834
15	Estado de México		Entidad	3 296	3 948
			Urbano	1 878	2 256
			Rural	1 418	1 692
16	Michoacán de Ocampo		Entidad	2 010	2 425
			Urbano	1 380	1 656
			Rural	630	769
17	Morelos		Entidad	2 467	2 966
			Urbano	1 692	2 034
			Rural	775	932
18	Nayarit		Entidad	2 005	2 407
			Urbano	1 404	1 686
			Rural	601	721
19	Nuevo León		Entidad	3 870	4 632
			Urbano	2 244	2 694
			Rural	1 626	1 938
20	Oaxaca		Entidad	2 448	2 930
			Urbano	972	1 170
			Rural	1 476	1 760
21	Puebla		Entidad	2 107	2 524
			Urbano	1 548	1 860
			Rural	559	664
22	Querétaro		Entidad	3 640	4 370
			Urbano	1 866	2 244
			Rural	1 774	2 126
23	Quintana Roo		Entidad	2 312	2 784
			Urbano	1 902	2 286
			Rural	410	498
24	San Luis Potosí		Entidad	2 568	3 074
			Urbano	1 488	1 788
			Rural	1 080	1 286
25	Sinaloa		Entidad	3 309	3 980
			Urbano	2 778	3 336
			Rural	531	644
26	Sonora		Entidad	2 446	2 956
			Urbano	1 740	2 088
			Rural	706	868
27	Tabasco		Entidad	2 084	2 493
			Urbano	1 266	1 523
			Rural	818	970
28	Tamaulipas		Entidad	2 377	2 870
			Urbano	1 878	2 256
			Rural	499	614

C. Distribución de la muestra en viviendas por entidad según dominio de estudio para la Encuesta Nacional de Ingresos y Gastos de los Hogares 2020 (ENIGH-2020)

Clave	Entidad Nombre	Dominio	ENIGH nueva serie 2018	ENIGH nueva serie 2020
29	Tlaxcala	Entidad	2 137	2 592
		Urbano	1 752	2 106
		Rural	385	486
30	Veracruz de Ignacio de la Llave	Entidad	2 715	3 265
		Urbano	1 344	1 614
		Rural	1 371	1 651
31	Yucatán	Entidad	2 976	3 545
		Urbano	1 920	2 304
		Rural	1 056	1 241
32	Zacatecas	Entidad	2 404	2 885
		Urbano	1 488	1 788
		Rural	916	1 097