

Encuesta Nacional sobre Uso del Tiempo 2019

ENUT

Diseño muestral



Encuesta Nacional sobre Uso del Tiempo (ENUT) 2019

Diseño muestral



DR © 2020, **Instituto Nacional de Estadística y Geografía**

Edificio Sede

Av. 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.

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

Diseño muestral de la Encuesta Nacional sobre Uso del Tiempo (ENUT) 2019

Presentación

El **Instituto Nacional de Estadística y Geografía (INEGI)**, en el marco de las atribuciones que le confiere la Ley del Sistema Nacional de Información Estadística y Geográfica, desarrolla y aplica estrategias orientadas a consolidar el Sistema Nacional de Información Estadística y Geográfica (SNIEG), entre las cuales se encuentra la documentación de los diferentes proyectos que lleva a cabo.

En tal contexto, presenta el **Diseño muestral de la Encuesta Nacional sobre Uso del Tiempo (ENUT) 2019**, con el propósito de ofrecer un panorama general sobre la metodología empleada para el diseño y ejecución del proyecto.

De esta forma, el **INEGI** ofrece a los usuarios un documento en el que se muestran, en forma detallada, los diferentes aspectos del diseño estadístico del proyecto.

Asimismo, a través de este tipo de documentos es posible contribuir a transparentar el proceso de generación de información estadística y disponer de un testimonio que permita fortalecer el diseño y la operación de futuros proyectos en la materia.

Índice

Introducción	VII
1. Bases metodológicas	1
1.1 Objetivos	1
1.2 Referencias metodológicas	2
2. Diseño estadístico	3
2.1 Marco de la encuesta	3
2.2 Formación de las UPM	3
2.3 Estratificación	4
2.4 Selección de las UPM de la Muestra Maestra	4
2.5 Tamaño de la muestra	5
2.6 Asignación de la muestra	5
2.7 Selección de la muestra	6
2.7.1 En urbano alto	6
2.7.2 En complemento urbano	7
2.7.3 En rural	8
2.8 Ajuste a los factores de expansión	9
2.8.1 Ajuste por no respuesta	9
2.8.2 Ajuste por proyección	10
2.9 Estimadores	12
2.10 Estimación de errores de muestreo	13
2.11 Homologación de la semaforización para los umbrales de indicadores de precisión estadística	14
Anexo	15
A. Indicadores empleados en la estratificación de los conglomerados de la Muestra Maestra	16
B. Distribución de la muestra en viviendas por entidad y dominio	17
C. Distribución de la muestra adicional en viviendas por entidad y dominio	18

Introducción

La **Encuesta Nacional sobre Uso del Tiempo (ENUT) 2019** es un ejercicio estadístico desarrollado por el INEGI que proporcionará información para la medición de todas las formas de trabajo de los individuos, ya sea remunerado o no; además, contribuirá a la visibilización de la producción doméstica y su aporte a la economía; la distribución en la ocupación del tiempo entre hombres y mujeres, así como la percepción del bienestar propio en la población de 12 años y más de áreas urbanas, rurales e indígenas.

El diseño muestral que se presenta a continuación está organizado en dos capítulos; en el primero, se abordan los objetivos y las bases metodológicas, mientras que el segundo está dedicado al diseño estadístico.

De esta manera, el **INEGI** da a conocer la metodología empleada y contribuye así a la transparencia del proceso de generación de información estadística.

1. Bases metodológicas

En este apartado se presentan los objetivos del proyecto y las referencias metodológicas que lo distinguen.

1.1 Objetivos

Con base en la necesidad de profundizar en el tema, se establecieron los siguientes objetivos:

General

Proporcionar información estadística necesaria para la medición de todas las formas de trabajo de los individuos, tanto remunerado como no remunerado, y hacer visible la importancia de la producción doméstica y su contribución a la economía.

Específicos

- I. Contar con información que permita asignar un valor social y económico al trabajo no remunerado que se realiza en los hogares, que permita contabilizar la producción doméstica y contribuir a la construcción de la cuenta satélite de los hogares.
- II. Evidenciar la permanencia de roles de género, en función de las diferencias en la participación de mujeres y hombres en el trabajo remunerado y no remunerado, y en otras actividades cotidianas.
- III. Disponer de información sobre las actividades que realizan cotidianamente las mujeres y los hombres, con base en sus características demográficas y socioeconómicas.
- IV. Analizar algunos aspectos de la vida cotidiana de las familias en relación con sus formas de organización, dinámicas y ciclo de vida.
- V. Conocer la participación y el tiempo asignado al cuidado y atención de personas dependientes, niñas y niños, personas adultas mayores, enfermas o con algún tipo de discapacidad, para hacer visibles las responsabilidades que han asumido las familias y la necesidad de participación del Estado y de las empresas para satisfacer las demandas de cuidado de las personas.
- VI. Tener un acercamiento a los niveles de participación social y política en actividades para el bien de la comunidad y cuidado del medio ambiente.
- VII. Informar sobre la participación y el tiempo que destinan las personas a actividades recreativas, deportivas, culturales y de esparcimiento, como un acercamiento a su calidad de vida.
- VIII. Estimar los tiempos que invierte la población en traslados para hacer visibles los problemas de transporte y vías de comunicación, además de la ubicación centralizada de fuentes de empleo y centros educativos.
- IX. Estimar los tiempos que invierte la población en realizar trámites diversos para contribuir a hacer visible la necesidad de simplificarlos.

- X. Identificar el tiempo destinado a actividades educativas y otras actividades relacionadas con la educación formal o capacitación para el trabajo.
- XI. Identificar el tiempo que invierten las personas en realizar actividades de apoyo a otros hogares relacionadas con trabajo no remunerado y cuidado de personas.
- XII. Identificar el tiempo que destinan las personas a la utilización de distintos medios de comunicación como una forma de evidenciar la necesidad de utilizarlos para informar a la población sobre aspectos importantes para su salud, educación y sensibilización en género, entre otros.
- XIII. Estimar el tiempo invertido por las personas en el autocuidado: dormir, comer, ir al médico, recuperarse de una enfermedad y al arreglo personal.
- XIV. Identificar la distribución del tiempo de la población indígena en diversas actividades, tanto entre las específicas de sus comunidades como entre las actividades que han adoptado del contexto nacional.
- XV. Identificar las actividades que muestran mayores brechas de género en las poblaciones indígenas con relación a la población nacional.
- XVI. Identificar la percepción de bienestar de la población con el tiempo que dedica a diversos grupos de actividades y con su vida en general.

1.2 Referencias metodológicas

Unidad de análisis. Fueron consideradas las unidades básicas de observación y análisis:

- Las viviendas
- Los hogares que residen en viviendas particulares.
- Los integrantes del hogar de 12 años y más.

Población objeto de entrevista. Todos los integrantes del hogar de 12 años y más.

Periodo de referencia. La semana anterior a la fecha de la entrevista, en dos ciclos: de lunes a viernes y de sábado a domingo.

Unidad de medida. El tiempo dedicado a las actividades se solicitó en horas y minutos, por ser las medidas de tiempo más prácticas para su registro.

Cobertura geográfica. El proyecto cubre todo el territorio nacional, incluyendo áreas urbanas, rurales y poblaciones indígenas seleccionadas.

Periodo de levantamiento. El trabajo de campo se realizó del 21 de octubre al 01 de diciembre de 2019.

Método de recolección. La captación de los datos se realizó por entrevista directa, mediante un cuestionario electrónico operado en un dispositivo móvil (mini *laptop*), cuya aplicación fue realizada directamente por los entrevistadores. En aquellos casos en que no se pudo acceder al sistema de captura o utilizar la mini *laptop*, se levantó la información en el cuestionario impreso, para posteriormente capturarlo en el electrónico.

Informante adecuado. Para las secciones *I Características de la vivienda*, *II Identificación y características del hogar* y *III Características sociodemográficas*, se entrevistó al jefe o la jefa del hogar, el cónyuge o algún integrante (de 15 o más años) del hogar que conociera la información. A partir de la sección *IV Características personales*, las preguntas se realizaron a cada integrante del hogar de 12 años y más.

2. Diseño estadístico

Comprende el conjunto de actividades relacionadas con la selección de la muestra, el tamaño suficiente que permita realizar estimaciones para la población objeto de estudio, el marco de muestreo, y la construcción y evaluación de las estimaciones planteadas para la expansión de la información, a partir de los datos obtenidos en campo.

2.1 Marco de la encuesta

El diseño de la muestra para la ENUT 2019 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 bietápico, estratificado y por conglomerados, donde las unidades de observación son los hogares y las personas de 12 años o más de edad.

La submuestra para la ENUT 2019 se seleccionó a partir de la Muestra Maestra 2012 del INEGI, la cual se diseñó y seleccionó del Marco Maestro de Muestreo (MMM) 2012, conformado por 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.

2.2 Formación de las UPM

En principio, se construye el conjunto de UPM que cubren el territorio nacional.

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

a) En urbano alto

El tamaño mínimo de una UPM es de 80 viviendas habitadas y el máximo es de 160. Pueden estar formadas por:

- Una manzana.
- La unión de dos o más manzanas contiguas de la misma AGEB¹.
- 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.

¹ Área Geoestadística Básica.

c) En rural

El tamaño mínimo de una UPM es de 160 viviendas habitadas y el máximo es de 300. Pueden estar formadas por:

- Una AGEB.
- Parte de una AGEB.
- La unión de dos o más AGEB colindantes del mismo municipio.
- La unión de una AGEB con una parte de otra AGEB colindante del mismo municipio.

2.3 Estratificación

Una vez construido el conjunto de UPM, se agruparon aquellas con características similares, de acuerdo a lo siguiente:

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

Estratificación por tamaño de localidad según ámbito geográfico

Ámbito	Zona	Descripción
Urbano alto	01 a 09	Ciudades con 100 000 o más habitantes.
Complemento urbano	25	De 50 000 a 99 999 habitantes.
	35	De 15 000 a 49 999 habitantes.
	45	De 5 000 a 14 999 habitantes.
	55	De 2 500 a 4 999 habitantes.
Rural	60	Localidades menores de 2 500 habitantes.

De manera paralela, se formaron cuatro estratos sociodemográficos en los que se agruparon todas las UPM del país; esta estratificación considera las características sociodemográficas de los habitantes de las viviendas, así como las características físicas y el equipamiento de las mismas, expresadas por medio de 34 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 obtuvo un total de 683 estratos en todo el territorio nacional.

2.4 Selección de las UPM de la Muestra Maestra

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

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

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.

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

2.5 Tamaño de la muestra

La expresión para calcular el tamaño de muestra fue:

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

Donde:

- n = tamaño de la muestra.
- z = valor asentado en las tablas estadísticas de la distribución normal estándar para una confianza prefijada.
- p = estimación de la proporción de interés.
- r = error relativo máximo aceptable.
- q = 1-p.
- tnr = tasa de no respuesta máxima esperada.
- 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.

Considerando una confianza de 90%, un efecto de diseño de 2.12, un error relativo máximo esperado de 15%, una tasa de no respuesta máxima esperada de 15%, y una proporción de 1%, se obtuvo un tamaño de muestra de 29 686 viviendas, que se ajustó a 30 009 viviendas a nivel nacional. Este tamaño de muestra alcanza para cubrir proporciones iguales o mayores de 1%, obteniendo errores esperados menores de 15%; para proporciones menores de 1% se obtendrán errores esperados mayores de 15%. En el anexo B se presenta la distribución de la muestra.

Con el objeto de disminuir los errores en la estimación de las variables de uso del tiempo entre la población de 12 años o más de edad hablante de lengua indígena, se seleccionó una muestra de 2 000 viviendas en localidades con 90% o más de población hablante de lengua indígena. La distribución de la muestra se presenta en el anexo C. A la muestra anterior se le agregaron las observaciones de población hablante de lengua indígena obtenidas en la muestra nacional, a fin de generar resultados sobre el uso del tiempo para esta subpoblación en particular.

2.6 Asignación de la muestra

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

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

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

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

Donde:

- n_{eh} = número de viviendas en muestra en el h-ésimo estrato, en la e-ésima entidad.
- n_e = número total de viviendas en muestra en la e-ésima entidad.
- N_{eh} = número total de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- N_e = número total de viviendas en la e-ésima entidad.
- k_{eh}^* = número de UPM en muestra en el h-ésimo estrato, en la e-ésima entidad.
- b = número de viviendas a seleccionar por UPM.

2.7 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:

2.7.1 En urbano alto

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.
2. En cada UPM se eligieron cinco viviendas con igual probabilidad.

La probabilidad de seleccionar la i-ésima UPM, en el 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 ENUT 2019 es:

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

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

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

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:

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

Su factor de expansión³ está dado por:

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

³ El factor de expansión se define como el inverso de la probabilidad de selección. En la *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, el factor de expansión se nombra Ponderador.

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 a seleccionadas, 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 el Censo de Población y Vivienda 2010.
- m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, según el Censo de Población y Vivienda 2010.
- m_{ehi}^* = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento de la actualización del listado de viviendas, previo al levantamiento de la ENUT 2019.

2.7.2 En complemento urbano

1. De las k_{eh} UPM que integran la Muestra Maestra, en el h-ésimo estrato, de la e-ésima entidad, se eligieron k_{eh}^* UPM con igual probabilidad.
2. En cada UPM seleccionada, se eligieron 20 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 ENUT 2019 es:

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

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

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

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:

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

Su factor de expansión está dado por:

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

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 el Censo de Población y Vivienda 2010.
- k_{eh}^* = número de UPM seleccionadas, 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 ENUT 2019.

2.7.3 En rural

1. De las k_{eh} UPM que integran la Muestra Maestra en el h-ésimo estrato, de la e-ésima entidad, se eligieron k_{eh}^* UPM con igual probabilidad.
2. En cada UPM seleccionada, se eligieron cuatro segmentos de cinco viviendas cada uno, 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 ENUT 2019 es:

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

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

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

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:

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

En consecuencia, su factor de expansión está dado por:

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

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 el Censo de Población y Vivienda 2010.
- m_{eh} = número total de viviendas en el h-ésimo estrato, en la e-ésima entidad, según el Censo de Población y Vivienda 2010.
- k_{eh}^* = número de UPM seleccionadas para la ENUT 2019, 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 ENUT 2019..

2.8 Ajuste a los factores de expansión

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

2.8.1 Ajuste por no respuesta

Se realizó a nivel de las viviendas, los hogares y las personas seleccionadas a nivel estrato, en cada uno de los dominios de estudio mediante las siguientes expresiones:

A NIVEL VIVIENDA

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

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

Donde:

- F'_{ehij} = factor de expansión corregido por no respuesta de la j-ésima vivienda seleccionada, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- F_{ehij} = factor de expansión de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- I_{ehij} = función indicadora con valor 1 o 0; esta función toma el valor 1 si la j-ésima vivienda seleccionada, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad proporcionó respuesta completa, y toma el valor de 0 en caso contrario.

A NIVEL HOGAR

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

$$F''_{ehijk} = F'_{ehij} \frac{\sum_h \sum_i \sum_j F'_{ehij}}{\sum_h \sum_i \sum_j F'_{ehij} I_{ehijk}}$$

Donde:

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

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:

$$F'''_{ehijk\ell} = F''_{ehijk} \frac{\sum_h \sum_i \sum_j F''_{ehijk}}{\sum_h \sum_i \sum_j F''_{ehij} I_{ehijk\ell}}$$

Donde:

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

2.8.2 Ajuste por proyección

Los factores de expansión ajustados por la no respuesta se corrigieron, con el propósito de asegurar que en cada dominio de interés de la encuesta se obtenga la población total determinada por la proyección generada por el Consejo Nacional de Población (CONAPO) referida al punto medio del levantamiento, mediante la siguiente expresión:

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

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.
- $PROy_D$ = población en el dominio D, según proyección.
- $PEXP_D$ = población total a la que expande la encuesta en el dominio D.
- D = nivel de desagregación de la población al que se realiza el ajuste por proyección.

AJUSTE POR NO RESPUESTA A NIVEL PERSONA PARA INFORMANTES DE 12 AÑOS O MÁS DE EDAD

El factor persona se ajusta para los informantes de 12 años o más de edad a nivel estrato, mediante la siguiente expresión:

$$F''_{ehi} = F'_D \frac{Q_{eh} + Q'_{eh}}{Q_{eh}}$$

Donde:

F'_D = factor vivienda ajustado por proyección de población.

Q_{eh} = número de informantes de 12 años o más de edad con respuesta en el h-ésimo estrato, en la e-ésima entidad.

Q'_{eh} = número de informantes de 12 años o más de edad sin respuesta en el h-ésimo estrato, en la e-ésima entidad.

AJUSTE PARA LA MUESTRA DE POBLACIÓN HABLANTE DE LENGUA INDÍGENA

El cálculo de los factores de expansión para la población de 12 años o más de edad hablante de lengua indígena se elaboró conforme al procedimiento descrito antes, con el fin de dar resultados del uso del tiempo para esa población.

Los factores de expansión se ajustaron de acuerdo con el siguiente procedimiento:

$$F^i_{ehi} = \alpha F^*_{ehi}$$

$$F^{ii}_{ehi} = (1 - \alpha) F^{**}_{ehi}$$

$$\alpha = \frac{n_1}{n_1 + n_2}, \quad \alpha + (1 - \alpha) = 1$$

Donde:

F^*_{ehi} = factor de expansión de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad a nivel persona de 12 años o más de edad de la población hablante de lengua indígena, en la muestra tradicional.

F^{**}_{ehi} = factor de expansión de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad a nivel persona de 12 años o más de edad de la población hablante de lengua indígena, en la muestra adicional.

F^i_{ehi} = factor de expansión ajustado por el ponderador muestral de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad a nivel persona de 12 años o más de edad de la población hablante de lengua indígena, en la muestra tradicional.

F^{ii}_{ehi} = factor de expansión ajustado por el ponderador muestral de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad a nivel persona de 12 años o más de edad de la población hablante de lengua indígena, en la muestra adicional.

n_1 = número de personas de 12 años o más de edad que hablan lengua indígena en la muestra tradicional.

n_2 = número de personas de 12 años o más de edad que hablan lengua indígena en la muestra adicional.

α = factor de ponderación muestral para unir las dos muestras de población de 12 o más años de edad que hablan lengua indígena.

2.9 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 dominio 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 dominio urbano alto.
- F_{ehi}^{CU} = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, del dominio 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 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.
- $X_{ehis\ell}^R$ = 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 dominio rural.

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

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

La construcción de los estimadores para muestra de la población de 12 años o más de edad hablante de lengua indígena, se hace de forma análoga al caso nacional:

$$\hat{X} = \sum_e \sum_h \sum_i F_{ehi}^i \left(\sum_j X_{ehij}^i \right) + \sum_e \sum_h \sum_i F_{ehi}^{ii} \left(\sum_j X_{ehij}^{ii} \right)$$

Donde:

- F_{ehi}^i = factor de expansión final, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad de la muestra tradicional.
- X_{ehij}^i = valor observado de la característica de interés X en la j-ésima persona de 12 años y más hablante de lengua indígena, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad de la muestra tradicional.
- F_{ehi}^{ii} = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, de la muestra adicional.
- X_{ehij}^{ii} = valor observado de la característica X en la j-ésima persona de 12 años y más hablante de lengua indígena, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad de la muestra adicional.

2.10 Estimación de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones nacionales y por dominio se usó el método de Conglomerados Últimos⁴, basado en que la mayor contribución a la varianza de un estimador, en un diseño polietápico, es la que se presenta entre las UPM. El término Conglomerados Últimos se utiliza para denotar el total de unidades en muestra de una UPM.

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

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

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.

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:

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

En el caso de la estimación de varianza para las estimaciones de la muestra de la población de 12 años y más de edad hablante de lengua indígena, los cálculos se realizan utilizando la misma fórmula para estimar su precisión.

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

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

Donde:

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

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

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

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

Finalmente, el intervalo de confianza $I_{1-\alpha}$ al 100 (1- α)%, se construye de la siguiente forma:

$$I_{1-\alpha} = \left(\hat{\theta} - z_{\alpha/2} \sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + z_{\alpha/2} \sqrt{\hat{V}(\hat{\theta})} \right)$$

Donde α es el nivel de significancia.

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

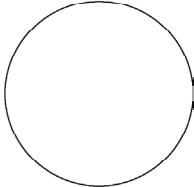
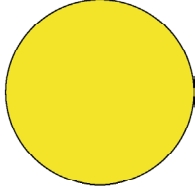
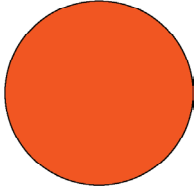
Para facilitar la interpretación de las precisiones estadísticas de la información pública en tabulados, el Comité de Aseguramiento de la Calidad, en la cuarta sección celebrada el 1 de noviembre de 2018, aprobaron los siguientes umbrales y especificaciones para los coeficientes de variación en la publicación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

Umbrales aprobados para la precisión de las estimaciones

Interpretación	Semaforización	Nivel de precisión
Alta	Blanco	[0%, 15%)
Moderada	Amarillo	[15%, 30%)
Baja	Naranja oscuro	>=30%

A partir del segundo trimestre de 2018, se publican los siguientes indicadores de precisión estadística en la presentación de resultados en tabulados de todas las encuestas con muestreo probabilístico del INEGI: error estándar, intervalo de confianza y coeficiente de variación. Adicionalmente, se estandariza la coloración en los tabulados para indicar el nivel de precisión de las estimaciones con base en el CV. A continuación, se presenta el código RGB (sigla en inglés de *red, green, blue*; en español: rojo, verde y azul) de los colores utilizados en la semaforización.

Parámetros RGB para la semaforización del CV

Código	Interpretación		
	Alta	Moderada	Baja
			
Rojo	255	255	255
Verde	255	234	84
Azul	255	0	0

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

Las estimaciones que aparecen en este cuadro están coloreadas, de acuerdo con su nivel de precisión, en alta, moderada y baja, tomando como referencia el 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.

Anexo

A. Indicadores empleados en la estratificación de los 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 o más años de edad que aprobaron algún grado de escolaridad diferente al nivel preescolar.
PP15PRI_CO	de 15 o más años de edad que tienen como máxima escolaridad 6 grados aprobados en primaria.
PP15SEC_CO	de 15 o más años de edad que tienen como máxima escolaridad 3 grados aprobados en secundaria.
PPEA	de 12 años y más que trabajaron; tenían trabajo, pero no trabajaron; o buscaron trabajo en la semana de referencia.
PPEA_F	femenina de 12 años y más que trabajaron; tenían trabajo pero no trabajaron; o buscaron trabajo en la semana de referencia.
TOCU12A17	no ocupada de 12 a 17 años entre la población total de este rango.
PPOMAYED	ocupada de 18 años y más entre la población total ocupada.
PGDO_ESC	Grado promedio de escolaridad.
PTASAOcupa	Tasa de ocupación.
Proporción de viviendas particulares habitadas	
PVIVSINH	que no tienen hacinamiento.
PVPH_PISDT	que tienen piso de cemento o firme, madera, mosaico u otro material.
PVPH2YMASD	que usan para dormir entre 2 y 25 cuartos.
PVPH_2MASC	que tienen más de un cuarto.
PVPH3YMASD	que tienen entre 3 y 25 cuartos.
PVPH_C_ELE	que disponen de luz eléctrica.
PVPHAGUADV	que tienen disponibilidad de agua entubada dentro de la vivienda, o fuera de ella, pero dentro del terreno.
PVPH_EXCSA	que tienen excusado, retrete, sanitario, letrina u hoyo negro.
PVPHDRENAJ	que tienen drenaje conectado a la red pública, fosa séptica, barranca, grieta, río, lago o mar.
PVDRENER	que disponen de drenaje conectado a la red pública.
PVEXCAGU	que disponen de excusado con descarga directa de agua.
PVPH_CSERV	que disponen de luz eléctrica, agua entubada dentro o fuera de la vivienda, pero dentro del terreno, así como drenaje.
PSIN_HASIN	que no se encuentran en situación de hacinamiento a nivel manzana.
Proporción de viviendas particulares habitadas que disponen de	
PVPH_TV	televisor.
PVPH_AUTOM	automóvil o camioneta.
PVPH_CEL	teléfono celular.
PVCELFIJ	teléfono celular y teléfono fijo.
PV4ELEC	radio, televisor, refrigerador y lavadora.
PVRADTEL	radio y televisor.
PVPHCBIE	todos los bienes.

B. Distribución de la muestra en viviendas, por entidad y dominio

Entidad federativa	Total	Urbano alto	Complemento urbano	Rural
Nacional	30 009	15 420	7 520	7 069
01 Aguascalientes	935	635	140	160
02 Baja California	943	700	160	83
03 Baja California Sur	933	335	480	118
04 Campeche	935	475	240	220
05 Coahuila de Zaragoza	941	680	160	101
06 Colima	935	575	260	100
07 Chiapas	935	220	280	435
08 Chihuahua	940	680	120	140
09 Ciudad de México	940	920	0	20
10 Durango	933	515	140	278
11 Guanajuato	938	420	240	278
12 Guerrero	938	300	260	378
13 Hidalgo	935	195	320	420
14 Jalisco	941	520	300	121
15 México	944	700	120	124
16 Michoacán de Ocampo	939	240	400	299
17 Morelos	934	515	280	139
18 Nayarit	937	335	320	282
19 Nuevo León	940	740	140	60
20 Oaxaca	935	160	300	475
21 Puebla	939	380	300	259
22 Querétaro	941	515	160	266
23 Quintana Roo	937	695	140	102
24 San Luis Potosí	937	415	200	322
25 Sinaloa	946	460	240	246
26 Sonora	934	595	220	119
27 Tabasco	935	195	360	380
28 Tamaulipas	941	680	140	121
29 Tlaxcala	936	555	220	161
30 Veracruz de Ignacio de la Llave	942	320	280	342
31 Yucatán	934	495	300	139
32 Zacatecas	936	255	300	381

C. Distribución de la muestra adicional en viviendas, por entidad y dominio

Entidad federativa	Total	Urbano alto	Complemento urbano	Rural
Nacional	2 000	300	450	1 250
07 Chiapas	350	0	75	275
08 Chihuahua	50	0	0	50
12 Guerrero	200	0	55	145
13 Hidalgo	100	0	20	80
15 México	50	50	0	0
16 Michoacán de Ocampo	50	0	35	15
20 Oaxaca	340	0	110	230
21 Puebla	200	100	0	100
24 San Luis Potosí	100	0	10	90
30 Veracruz de Ignacio de la Llave	350	150	0	200
31 Yucatán	210	0	145	65