

2024

DISEÑO MUESTRAL

Encuesta Nacional sobre Uso del Tiempo 2024

ENUT

Diseño muestral



Obras complementarias publicadas por el INEGI sobre el tema:
Diseño muestral de la Encuesta Nacional sobre Uso del Tiempo ENUT 2019.

Catalogación en la fuente INEGI:

306.8501 Encuesta Nacional sobre Uso del Tiempo (2024).
 Encuesta Nacional sobre Uso del Tiempo 2024 : ENUT : diseño muestral /
 Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2025.

23 p.

1. Uso del tiempo - México - Encuestas - Metodología. I. Instituto Nacional de
Estadística y Geografía (México).

Conociendo México

800 111 4634
www.inegi.org.mx
atencion.usuarios@inegi.org.mx

 **INEGI Informa**  **@INEGI_INFORMA**

Registro en trámite

2025, **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

La Encuesta Nacional sobre Uso del Tiempo (*ENUT*) representa un instrumento estadístico fundamental para México. Su propósito central es generar información exhaustiva sobre la distribución del tiempo de la población en una amplia gama de actividades diarias. Esta información es crucial para cuantificar el valor económico y social del trabajo no remunerado, analizar las brechas de género en las cargas de trabajo y fundamentar el diseño, monitoreo y evaluación de políticas públicas dirigidas a promover la igualdad, el bienestar social y el desarrollo sostenible. La edición de la *ENUT* 2024 simboliza una renovación y continuidad de los ejercicios previos, al integrar las mejores prácticas internacionales y adaptar su metodología para satisfacer las demandas informativas actuales del país.

El objetivo primordial de este documento es detallar el diseño de muestreo implementado para la *ENUT* 2024. Aquí se exponen los elementos metodológicos que garantizan la representatividad estadística de la muestra, la precisión de las estimaciones y la confiabilidad de los resultados a los niveles de desagregación establecidos. Este documento sirve como una referencia metodológica esencial para usuarios, investigadores y el personal involucrado en la producción y análisis de la información generada por la *ENUT* 2024.

Índice

1. Bases metodológicas	1
1.1 Objetivo general	1
1.2 Población objetivo	1
1.3 Cobertura geográfica	1
1.4 Estimación en subpoblaciones	1
2. Marco de muestreo	3
2.1 Descripción del marco de muestreo	3
2.2 Estratificación del marco de muestreo	3
2.3 Muestra Maestra y actualización del marco	4
3. Diseño estadístico de la muestra	5
3.1 Descripción del diseño de muestreo	5
3.2 Tamaño de la muestra	5
3.3 Distribución de la muestra en los estratos	6
3.4 Determinación del número de viviendas por UPM en la tercera etapa de muestreo	7
4. Ponderación	9
4.1 Ponderadores de diseño	9
4.2 Factores de corrección por no respuesta	10
4.2.1 Ajuste por no respuesta de UPM (segunda etapa de muestreo)	10
4.2.2 Ajuste por no respuesta de vivienda (tercera etapa de muestreo)	11
4.2.3 Ajuste por no respuesta de hogar	11
4.2.4 Ajuste por no respuesta de persona	12
4.3 Calibración de ponderadores	12
4.4 Ponderadores para el proceso de estimación	13
5. Estimación	15
5.1 Estimación de totales y razones	15
5.2 Estimación de la varianza	15
5.3 Intervalos de confianza	17
6. Anexos	19
A. Muestra de viviendas con hablantes de lenguas indígenas	19

B. Distribución de la muestra nacional de viviendas por entidad federativa y dominios 21

7. Referencias 23

1. Bases metodológicas

1.1 Objetivo general

La *ENUT* tiene como objetivo generar información estadística sobre las actividades realizadas por las personas de 12 años y más durante la semana de referencia, así como el tiempo que les dedican. Su finalidad es medir todas las formas de trabajo, tanto remunerado como no remunerado (incluyendo el trabajo doméstico, de cuidados, el trabajo comunitario y el voluntario). Además, de cuantificar el tiempo destinado a actividades personales, de estudio, de convivencia social, entre otras. Estos datos permiten analizar las diferencias en la distribución del tiempo entre mujeres y hombres, grupos de edad, ubicación geográfica, condición étnica, entre otros factores, para el diseño de políticas públicas y estudios especializados.

1.2 Población objetivo

La población objetivo de la *ENUT* 2024 está constituida por todas las personas de 12 años y más de edad, residentes habituales en viviendas particulares habitadas dentro del territorio nacional de México.

1.3 Cobertura geográfica

El proyecto cubre todo el territorio nacional, incluyendo áreas urbanas y rurales.

1.4 Estimación en subpoblaciones

La encuesta tiene como objetivo proporcionar estimaciones a nivel nacional, por entidad federativa y para localidades de menos de 10 000 habitantes y de 10 000 o más habitantes.

2. Marco de muestreo

2.1 Descripción del marco de muestreo

La ENUT 2024 utiliza el Marco Maestro de Muestreo (MMM) del INEGI como marco de muestreo.

El *MMM* está compuesto por Unidades Primarias de Muestreo (*UPM*). Las *UPM* se delimitan geográficamente para agrupar un conjunto de viviendas particulares habitadas. En una zona urbana, las *UPM* están conformadas por una o varias manzanas que contienen viviendas particulares habitadas. Por su parte, en áreas rurales, las *UPM* pueden estar constituidas por una o más localidades pequeñas, o por agrupaciones de viviendas dispersas en un territorio delimitado, garantizando la contigüidad geográfica.

2.2 Estratificación del marco de muestreo

El marco de muestreo está estratificado exhaustivamente para mejorar la eficiencia del diseño y asegurar la representatividad de la muestra en los principales dominios de estudio.

Las variables de estratificación son

- **Entidad federativa:** Se establecen 32 conjuntos geográficos principales, correspondiendo a cada una de las entidades federativas de la República Mexicana
- **Tamaño de localidad:** Se consideran cuatro tamaños de localidad
 1. 100 000 habitantes o más
 2. De 15 000 a 99 999 habitantes
 3. De 2 500 a 14 999 habitantes
 4. De 1 a 2 499 habitantes

Estos cuatro rangos de población se agrupan en tres dominios geográficos principales para fines de muestreo: Urbano Alto (100 000 habitantes o más); Complemento Urbano (que abarca las localidades de 2 500 a 99 999 habitantes, correspondientes a los rangos 2 y 3) y Rural (localidades de 1 a 2 499 habitantes, correspondiente al rango 4).

- **Nivel socioeconómico:** Para determinar el nivel socioeconómico de las *UPM*, se utiliza información del último censo de población y vivienda, que incluye datos sobre empleo, ocupación, sexo, edad, educación, inserción laboral, características de las viviendas y acceso a servicios de salud. A partir de esta información, se establecieron cuatro niveles socioeconómicos para las *UPM*: Bajo, Medio Bajo, Medio Alto y Alto.

2.3 Muestra Maestra y actualización del marco

Para la gestión eficiente y la actualización continua del *MMM*, se implementó la selección de una Muestra Maestra (*MM*). Esta *MM* está constituida por una muestra de *UPM* seleccionadas del *MMM*, diseñada para ser lo suficientemente amplia como para mantener la representatividad estadística, pero con un tamaño que facilita su mantenimiento y actualización periódica.

La selección de las *UPM* que componen la *MM* se realizó en cada estrato mediante un muestreo balanceado con probabilidades proporcionales al tamaño, ver el Capítulo 6 de (Tillé, 2019). Este método garantiza que las *UPM* de un mismo estrato con mayor número de viviendas particulares habitadas tengan una mayor probabilidad de formar parte de la *MM*, lo que contribuye a la eficiencia y representatividad de las encuestas posteriores.

El *INEGI* mantiene la *MM* vigente a través de un operativo de campo continuo. Este proceso implica la verificación y actualización periódica de la información de las viviendas particulares habitadas dentro de las *UPM* que conforman la *MM*. Estas visitas garantizan que los listados de viviendas estén actualizados y que el marco de muestreo refleje la realidad demográfica del momento.

La *MM* sirve como la primera etapa de selección para las encuestas en hogares, incluida la *ENUT 2024*. A partir de las *UPM* seleccionadas en la *MM*, se procede a la selección de las muestras finales de viviendas particulares habitadas para cada encuesta específica. Este enfoque centralizado en la *MM* optimiza los recursos, asegura la consistencia metodológica entre distintas encuestas y facilita la comparabilidad de sus resultados.

3. Diseño estadístico de la muestra

3.1 Descripción del diseño de muestreo

El diseño de muestreo de la *ENUT* 2024 es **estratificado** aprovechando la estratificación preexistente del *MMM*. Dentro de cada estrato, se implementa un muestreo en tres etapas, diseñado para asegurar la representatividad y eficiencia de la muestra:

1. Primera etapa: selección de *UPM* para la *MM*

Como se mencionó en la Sección 2.3, en esta etapa inicial, se seleccionan las *UPM* que conforman la *MM* a partir del universo del *MMM*. Este proceso se lleva cabo mediante un muestreo balanceado con probabilidades proporcionales al tamaño. Este método garantiza que las *UPM* con un mayor número de viviendas tengan una mayor probabilidad de ser incluidas en la *MM*, lo que optimiza su representatividad inicial y su capacidad para ser la base de múltiples encuestas.

2. Segunda etapa: selección de *UPM* para la *ENUT* 2024

Para el levantamiento específico de la *ENUT* 2024, dentro de cada estrato previamente definido, se selecciona un conjunto de *UPM* de la *MM*. La selección en esta etapa se realiza mediante un muestreo aleatorio simple sin reemplazo. Este método garantiza que cada *UPM* disponible en la *MM* dentro de un estrato tenga una probabilidad igual de ser seleccionada para participar en la encuesta actual.

3. Tercera etapa: selección de viviendas particulares habitadas

Una vez seleccionadas las *UPM* en la segunda etapa, en cada una de estas *UPM* se procede a la selección de una muestra de viviendas particulares habitadas. Esta selección se realiza mediante un muestreo aleatorio simple sin reemplazo a partir del listado actualizado de viviendas en la *UPM*. Este es el último nivel de selección antes de la identificación de hogares y personas.

Este diseño multietápico permite al *INEGI* gestionar eficientemente los recursos, al concentrar las operaciones de campo en un conjunto limitado de *UPM* seleccionadas, mientras se mantiene la validez estadística y la representatividad de las estimaciones a los niveles requeridos.

Además, uno de los objetivos específicos de la *ENUT* 2024 es recopilar información sobre la población indígena. Dada su dispersión y su proporción relativamente menor en el conjunto de la población del país, el *MMM* por sí solo no es suficiente para identificarla eficazmente y garantizar un número adecuado de casos en la muestra que permita realizar estimaciones precisas. Por consiguiente, se ha diseñado e implementado una muestra específica para esta subpoblación. En el anexo A, se ofrece más información sobre esta muestra.

3.2 Tamaño de la muestra

El tamaño de la muestra de la *ENUT* 2024 se calculó para garantizar la representatividad estadística a nivel nacional, sobre todo en lo que respecta a la estimación de proporciones clave de la población objetivo. Para este cálculo, se utilizó la siguiente fórmula para estimar el tamaño de muestra necesario en términos de personas, ver la Sección 7.5 de (Lohr, 2010):

$$n = \frac{z_{1-\alpha/2}^2 \times (1 - p) \times DEFF}{\varepsilon^2 \times p \times (1 - tnr)}, \quad (1)$$

donde:

- p es la proporción por estimar
- ε es el error relativo máximo de estimación
- $z_{1-\alpha/2}$ es el percentil del $100 \times (1 - \alpha/2) \%$ de una normal estándar
- $(1 - \alpha)$ es la probabilidad mínima de observar un error relativo máximo de valor ε
- $DEFF$ es el efecto de diseño
- tnr es la tasa máxima de no respuesta

Para este proyecto en particular, se consideraron los siguientes valores de los parámetros, basados en la experiencia de la *ENUT* 2019 y en los objetivos de precisión establecidos:

- **Proporción por estimar (p):** Se tomó como referencia la tasa de participación de la población objetivo en los cuidados especiales a integrantes del hogar por enfermedad crónica, temporal o discapacidad, la cual fue del 5.87 % (o 0.0587 en proporción)
- **Error relativo máximo de estimación (ε):** Se estableció un error relativo máximo del 3.7 % (o 0.037)
- **Percentil de la normal estándar (z):** Para una probabilidad mínima de $(1 - \alpha) = 0.9$ (90 % y $\alpha = 0.1$) de observar un error relativo máximo de $\varepsilon = 0.037$, el percentil correspondiente de una distribución normal estándar es $z=1.6448$
- **Efecto de diseño ($DEFF$):** Se estableció un efecto de diseño de 2.5
- **Tasa máxima de no respuesta (tnr):** Se consideró una tasa de no respuesta máxima del 15 % (o 0.15)

Aplicando estos valores a la fórmula, se obtuvo un tamaño de muestra inicial de 93 210 personas de la población objetivo.

Considerando que, en promedio, se estima que hay tres personas de la población objetivo por vivienda, este tamaño de muestra en personas se traduce a 31 070 viviendas ($93\,210/3=31\,070$).

Cabe aclarar que, debido a la forma de distribuir la muestra de viviendas por estratos, la cual se describirá en el apartado 3.3, el tamaño de muestra final fue de 32 048 viviendas a nivel nacional.

3.3 Distribución de la muestra en los estratos

Para optimizar el uso de los recursos humanos y económicos, se decidió asignar un tamaño de muestra aproximadamente igual en cada entidad federativa. Al interior de cada entidad federativa, la asignación del tamaño de muestra por estrato se realizó de manera proporcional al tamaño de dicho estrato.

De esta forma, el tamaño de muestra asignado a un estrato específico h dentro de una entidad federativa e , denotado como n_{eh}^V , se determinó mediante la siguiente fórmula:

$$n_{eh}^V = n_e^V \frac{V_{eh}}{V_e}, \quad (2)$$

donde:

- n_e^V es el tamaño de muestra total de viviendas asignado a la entidad federativa e
- V_e es el número total de viviendas particulares habitadas en la entidad federativa e , según el marco de muestreo
- V_{eh} es el número de viviendas particulares habitadas en el estrato h de la entidad federativa e , según el marco de muestreo

Esta asignación proporcional dentro de cada entidad federativa, combinada con una asignación casi igual entre entidades, busca optimizar la eficiencia del diseño.

3.4 Determinación del número de viviendas por UPM en la tercera etapa de muestreo

Con la finalidad de optimizar nuevamente los recursos humanos y económicos, se estableció un número fijo de viviendas a seleccionar en la tercera etapa de muestreo, dentro de cada *UPM* que forma parte de la muestra de la *ENUT 2024*. Esta determinación se realizó diferenciando por el tamaño de localidad al que pertenece la *UPM*, acorde con el marco de muestreo:

- Se seleccionan 5 viviendas en cada *UPM* ubicada en localidades del dominio Urbano Alto
- Se seleccionan 20 viviendas en cada *UPM* ubicada en las localidades de los dominios Complemento Urbano y Rural

Esta diferenciación en el número de viviendas por *UPM* permite balancear la carga de trabajo de campo con la densidad de población y la variabilidad esperada de las características de interés en los distintos tipos de localidades, contribuyendo a la eficiencia operativa sin comprometer la representatividad.

Estos tamaños de muestra de viviendas por *UPM* determinan el número de estas necesarias en la segunda etapa de muestreo.

El resultado final de la distribución de la muestra de viviendas por entidad federativa y tamaño de localidad se muestra en la tabla que aparece en el Anexo B.

4. Ponderación

Los ponderadores, que se utilizan para estimar los parámetros de interés de la encuesta, se componen de tres partes: los ponderadores de diseño, los factores de corrección por no respuesta y el factor de calibración. A continuación, se describe cada uno de estos componentes.

4.1 Ponderadores de diseño

Como se mencionó en el apartado 3.1, el diseño de muestreo se realizó en tres etapas.

1. Primera etapa. Esta etapa corresponde a la selección de *UPM* que conforman la *MM*, como se mencionó en el apartado 2.3. En cada estrato se utilizó un diseño con probabilidades proporcionales al tamaño de cada *UPM*. Por lo tanto, para la *i*-ésima *UPM* del estrato *h* de la entidad *e*, el ponderador de diseño de esta etapa, denotado por f_{ehi}^{MM} , está dado por

$$f_{ehi}^{MM} = \frac{V_{eh}}{n_{eh}^{MM} \times V_{ehi}}, \quad (3)$$

donde:

- n_{eh}^{MM} es el tamaño de muestra total de *UPM* asignado al estrato *h* de la entidad federativa *e* en la *MM*
- V_{eh} es el número total de viviendas particulares habitadas en el estrato *h* de la entidad federativa *e*, según el marco de muestreo
- V_{ehi} , tamaño de la *UPM*, es el número de viviendas particulares habitadas en la *i*-ésima *UPM* del estrato *h* de la entidad federativa *e*, según el marco de muestreo

Cabe mencionar que estos ponderadores de diseño son los inversos de las probabilidades de inclusión de primer orden de las unidades de muestreo, ver la Sección 2.5 de (Tillé, 2019). Para esta etapa, esta probabilidad de inclusión está dada por

$$\frac{n_{eh}^{MM} \times V_{ehi}}{V_{eh}}, \quad (4)$$

cuyo valor inverso es precisamente el factor de la expresión (3).

En ocasiones, el tamaño de una *UPM* (V_{ehi}) puede ser tan grande en relación con el tamaño total del estrato (V_{eh}) y el número de *UPM* a seleccionar (n_{eh}^{MM}), que la probabilidad de inclusión calculada resultaría mayor que uno. Dado que una probabilidad no puede exceder la unidad, en estos casos, a la respectiva *UPM* se le asigna una probabilidad de inclusión igual a uno. Consecuentemente, su ponderador de diseño para esta etapa es también uno ($f_{ehi}^{MM} = 1$).

Para el resto de las *UPM* en ese estrato, el cálculo de probabilidades, expresión (4), se ajusta: el tamaño de muestra a seleccionar se disminuye en uno y el denominador de la fórmula de probabilidad se sustituye por $V_{eh} -$

V_{ehi} . Este procedimiento se repite iterativamente si, después del ajuste, alguna otra *UPM* aún presentara una probabilidad de inclusión mayor que uno, hasta que todas las *UPM* restantes tengan probabilidades menores o iguales a uno.

2. Segunda etapa. Esta etapa corresponde a la selección de *UPM* para el proyecto *ENUT* 2024. En cada estrato se utilizó un muestreo aleatorio simple sin reemplazo de *UPM* de la *MM*. Por lo tanto, en el estrato h de la entidad e , el ponderador de diseño de esta etapa para la j -ésima *UPM*, denotado por f_{ehj}^{ENUT} , está dado por

$$f_{ehj}^{ENUT} = \frac{n_{eh}^{MM}}{n_{eh}^{ENUT}}, \quad (5)$$

donde n_{eh}^{ENUT} es el tamaño de muestra de *UPM* asignado al estrato h de la entidad federativa e para la *ENUT* 2024.

3. Tercera etapa. En esta etapa se seleccionan viviendas por muestreo aleatorio simple sin reemplazo. Por lo tanto, en el estrato h de la entidad e , el ponderador de diseño de esta etapa para la k -ésima vivienda de la j -ésima *UPM*, denotado por f_{ehjk}^V , está dado por

$$f_{ehjk}^V = \frac{V_{ehj}^*}{n^V}, \quad (6)$$

donde:

$$n^V = \begin{cases} 5 & \text{para el dominio Urbano Alto} \\ 20 & \text{para los dominios Complemento Urbano y Rural,} \end{cases}$$

esto de acuerdo con el apartado 3.4.

Además, V_{ehj}^* es el número actualizado de viviendas particulares habitadas en la j -ésima *UPM* de la *MM* del estrato h de la entidad federativa e .

4.2 Factores de corrección por no respuesta

La no respuesta es un fenómeno inherente a las encuestas por muestreo y puede ocurrir en diferentes etapas de la captación de información. Para mitigar el sesgo potencial que esto introduce y asegurar la representatividad de las estimaciones, se calculan factores de corrección por no respuesta en cada etapa relevante del muestreo, ver la Sección 3.4 de (Valliant & Dever, 2018). Estos factores se aplican multiplicativamente a los ponderadores de diseño.

4.2.1 Ajuste por no respuesta de *UPM* (segunda etapa de muestreo)

Es posible que, por diversas causas (e.g., inaccesibilidad, rechazo colectivo), no se obtenga respuesta de ninguna vivienda en muestra de alguna *UPM* seleccionada para este proyecto. Previendo esta situación, se calcula un factor

de corrección por no respuesta de *UPM*. Este factor se aplica al ponderador de la segunda etapa de muestreo y se calcula dentro de cada estrato, ver la Sección 15.6.2 de (Särndal, Swensson, & Wretman, 1992).

Para un estrato *eh* (correspondiente a la entidad federativa *e* y el estrato *h* de tamaño de localidad y nivel socioeconómico), el factor de corrección por no respuesta, FNR_{eh}^{ENUT} , está dado por

$$FNR_{eh}^{ENUT} = \frac{\sum_{s_{eh}^{ENUT}} f_{ehj}^{MM}}{\sum_{r_{eh}^{ENUT}} f_{ehl}^{MM}}, \quad (7)$$

donde s_{eh}^{ENUT} es la muestra de *UPM* de la *ENUT* y r_{eh}^{ENUT} es el conjunto de *UPM* con respuesta en esta misma muestra.

4.2.2 Ajuste por no respuesta de vivienda (tercera etapa de muestreo)

Dentro de las *UPM* que tuvieron respuesta, es posible que no se obtenga información de todas las viviendas seleccionadas. Para estos casos, se calcula un factor de corrección por no respuesta de vivienda. Este factor se aplica al ponderador de la tercera etapa de muestreo.

Por lo tanto, para el estrato *eh*, el factor de corrección, FNR_{eh}^V , está dado por

$$FNR_{eh}^V = \frac{n^V \times \sum_{r_{eh}^{ENUT}} \{f_{ehj}^{MM} \times V_{ehj}^*\}}{\sum_{r_{eh}^{ENUT}} \{f_{ehl}^{MM} \times V_{ehl}^* \times R_{ehl}^V\}}, \quad (8)$$

donde R_{ehl}^V es el número de viviendas en muestra con respuesta para la *l*-ésima *UPM*.

4.2.3 Ajuste por no respuesta de hogar

El proceso de captación de información en cada vivienda en muestra requiere recabar información de las personas por hogar. Por esto, hay que prever la omisión de esta información a nivel de hogar.

Por lo tanto, para el estrato *eh*, el factor de corrección por no respuesta de hogar, denotado por FNR_{eh}^H , está dado por

$$FNR_{eh}^H = \frac{\sum_{r_{eh}^{ENUT}} \{f_{ehj}^{MM} \times V_{ehj}^* \times [\sum_{r_{ehj}^V} \eta_{ehjk}]\}}{\sum_{r_{eh}^{ENUT}} \{f_{ehl}^{MM} \times V_{ehl}^* \times [\sum_{r_{ehl}^V} R_{ehlm}^H]\}}, \quad (9)$$

donde,

- r_{ehj}^V es el conjunto de viviendas en muestra con respuesta en la *j*-ésima *UPM* de esta encuesta
- η_{ehjk} es el total de hogares en la *k*-ésima vivienda en muestra
- R_{ehlm}^H es el número de hogares con respuesta en la *m*-ésima vivienda

4.2.4 Ajuste por no respuesta de persona

Finalmente, en cada hogar que tuvo respuesta, se recaba información de las personas de la población objetivo. Si no es posible captar información de alguna de estas personas seleccionadas (e.g., ausencia prolongada, rechazo individual), es necesario calcular un factor de corrección por no respuesta de persona.

Para un estrato eh , el factor de corrección por no respuesta de persona, FNR_{eh}^P , está dado por

$$FNR_{eh}^P = \frac{\sum_{r_{eh}^{ENUT}} \{ f_{ehj}^{MM} \times V_{ehj}^* \times [\sum_{r_{ehj}^V} \sum_{r_{ehjk}^H} n_{ehjkl}^P] \}}{\sum_{r_{eh}^{ENUT}} \{ f_{ehl}^{MM} \times V_{ehl}^* \times [\sum_{r_{ehl}^V} \sum_{r_{ehlm}^H} R_{ehlm\chi}^P] \}} \quad (10)$$

donde:

- r_{ehjk}^H es el conjunto de hogares con respuesta en la k -ésima vivienda
- n_{ehjkl}^P es el número de personas de la población objetivo en el l -ésimo hogar de la k -ésima vivienda
- $R_{ehlm\chi}^P$ es el número de personas de la población objetivo con respuesta en el χ -ésimo hogar de la m -ésima vivienda

4.3 Calibración de ponderadores

El proceso de calibración de los ponderadores es una etapa fundamental para asegurar la coherencia de las estimaciones obtenidas de este proyecto. Esta coherencia es vital para la integración de datos y la producción de estadísticas oficiales armonizadas.

Los ponderadores ajustados por no respuesta se someterán a un proceso de calibración para que las estimaciones de ciertas variables auxiliares coincidan con los totales poblacionales conocidos y fiables. Al incorporar información auxiliar en este proceso de calibración, se mejora la precisión de las estimaciones y se reduce el posible sesgo inducido por la corrección de la no respuesta.

Las variables auxiliares utilizadas en este proyecto para la calibración son

- El número de residentes habituales de la población objetivo en las viviendas
- El número de personas hablantes de una lengua indígena en las viviendas

Para lograr esta calibración, se utiliza el método *Raking*, ver la Sección 12.2.2 de (Tillé, 2019). Este método ajusta los ponderadores de forma iterativa para que las distribuciones marginales de las variables auxiliares en la muestra se alineen con las distribuciones marginales poblacionales conocidas. El objetivo específico de aplicar *Raking* con estas variables es que la muestra de este proyecto estime, con la mayor precisión posible, la población total objetivo y la población de hablantes de lengua indígena a los niveles de desagregación relevantes.

De este proceso iterativo de calibración se obtiene el factor de calibración, denotado por g_{ehj} , el cual se aplica multiplicativamente al ponderador ajustado por no respuesta para cada *UPM* en la muestra de esta encuesta.

4.4 Ponderadores para el proceso de estimación

Conjuntando lo descrito en las secciones anteriores, el ponderador para el proceso de estimación que se aplica a cada vivienda en muestra de la j -ésima UPM del estrato eh está dado por:

$$w_{ehjk}^V = g_{ehj} \times f_{ehj}^{MM} \times [f_{ehj}^{ENUT} \times FNR_{eh}^{ENUT}] \times [f_{ehjk}^V \times FNR_{eh}^V], \quad (11)$$

para la k -ésima vivienda en muestra.

Consecuentemente, el ponderador utilizado en proceso de estimación que se aplica a la l -ésima persona de la población objetivo de la k -ésima vivienda en muestra es

$$w_{ehjkl}^P = w_{ehjk}^V \times FNR_{eh}^H \times FNR_{eh}^P. \quad (12)$$

5. Estimación

5.1 Estimación de totales y razones

Supóngase que se desea estimar el total nacional del parámetro X , dado por

$$X = \sum_{e=1}^{32} \sum_{h=1}^{H_e} \sum_{i=1}^{N_{he}} \sum_{k=1}^{V_{ehi}^*} x_{ehik}, \quad (13)$$

donde:

- H_e es el número de estratos (por tamaño de localidad y nivel socioeconómico) en la entidad e
- N_{he} es el número de *UPM* en el *MMM* para el estrato eh
- V_{ehi}^* es el número, actualizado, de viviendas en la ehi -ésima *UPM*
- x_{ehik} es el valor de la característica de interés para k -ésima vivienda de la ehi -ésima *UPM*

Entonces, con los ponderadores dados en la Sección cuatro y la información de la muestra, la estimación del parámetro X se puede expresar por:

$$\hat{X} = \sum_{e=1}^{32} \sum_{h=1}^{H_e} \sum_{s_{eh}^{ENUT}} \sum_{s_{ehj}^V} w_{ehjl}^V \times x_{ehjl}, \quad (14)$$

donde s_{eh}^{ENUT} es la muestra de *UPM* para este proyecto en el estrato eh y s_{ehj}^V es la muestra de viviendas para la j -ésima *UPM* en muestra.

Ahora, si se desea estimar el cociente de dos totales, denotado por $R = X/Y$, entonces la estimación con la información de la muestra se expresa por

$$\hat{R} = \frac{\hat{X}}{\hat{Y}}, \quad (15)$$

donde \hat{X} y \hat{Y} están dados por la expresión (14).

5.2 Estimación de la varianza

Dada la complejidad del diseño de muestreo (estratificado y multietápico) de este proyecto, las varianzas de los estimadores se calculan utilizando el método de Conglomerados Últimos, ver la Sección 4.6 de (Särndal, Swensson, & Wretman, 1992).

El método de conglomerados últimos es una técnica robusta y ampliamente utilizada para la estimación de la varianza en diseños de muestreo complejos. Este método simplifica el cálculo de la varianza al considerar únicamente la variabilidad de los estimadores entre las *UPM*.

Este enfoque es particularmente adecuado para este proyecto, donde las *UPM* de la segunda etapa actúan como los conglomerados últimos para el cálculo de la varianza. Al emplear este método, se logra una estimación viable y robusta de la varianza, se puede calcular el efecto de diseño, *DEFF*, y se pueden construir intervalos de confianza.

Especificamente, la estimación de la varianza del estimador \hat{X} está dada por

$$\widehat{Var}(\hat{X}) = \sum_{e=1}^{32} \sum_{h=1}^{H_e} \left\{ \frac{n_{eh}^{ENUT}}{(n_{eh}^{ENUT} - 1)} \times \sum_{s_{eh}^{ENUT}} \left[w_{ehj}^{ENUT} \times \hat{X}_{ehj} - \frac{\hat{X}_{eh}}{n_{eh}^{ENUT}} \right]^2 \right\}, \quad (16)$$

donde:

- $w_{ehj}^{ENUT} = g_{ehj} \times f_{ehj}^{MM} \times f_{ehj}^{ENUT} \times FNR_{eh}^{ENUT}$, resultado del producto del factor de calibración, por el ponderador de la primera etapa, por el ponderador de la segunda etapa y por el factor de corrección por no respuesta de *UPM*.
- la estimación del parámetro de interés a nivel de la j -ésima *UPM* en muestra es

$$\hat{X}_{ehj} = \sum_{s_{ehj}^V} f_{ehjl}^V \times FNR_{eh}^V \times x_{ehjl} \quad (17)$$

- la estimación del parámetro de interés en el estrato *eh* es

$$\hat{X}_{eh} = \sum_{s_{eh}^{ENUT}} w_{ehl}^{ENUT} \times \hat{X}_{ehl}. \quad (18)$$

Ahora, si se desea calcular la estimación de la varianza del cociente \hat{R} , esta estimación se obtiene de la siguiente fórmula, ver la Sección 15.4.1 de (Tillé, 2019):

$$\begin{aligned} \widehat{Var}(\hat{R}) = & \frac{1}{\hat{Y}^2} \sum_{e=1}^{32} \sum_{h=1}^{H_e} \left\{ \frac{n_{eh}^{ENUT}}{(n_{eh}^{ENUT} - 1)} \right. \\ & \times \left. \sum_{s_{eh}^{ENUT}} \left[\left(w_{ehj}^{ENUT} \times \hat{X}_{ehj} - \frac{\hat{X}_{eh}}{n_{eh}^{ENUT}} \right) - \hat{R} \times \left(w_{ehj}^{ENUT} \times \hat{Y}_{ehj} - \frac{\hat{Y}_{eh}}{n_{eh}^{ENUT}} \right) \right]^2 \right\} \end{aligned} \quad (19)$$

5.3 Intervalos de confianza

Un intervalo de confianza proporciona un rango de valores dentro del cual se espera que se encuentre el valor del parámetro poblacional (e.g., la proporción real de la población que realiza una actividad específica o el tiempo promedio real dedicado a una actividad), con una alta probabilidad o nivel de confianza.

La construcción de estos intervalos de confianza se basará en las estimaciones puntuales y las estimaciones de la respectiva varianza.

Especificamente, si se desea un intervalo del $100 \times (1 - \alpha)\%$ de confianza para el total X , entonces este se calcula por la siguiente expresión:

$$\left(\hat{X} - z_{1-\alpha/2} \times \sqrt{\widehat{Var}(\hat{X})}, \hat{X} + z_{1-\alpha/2} \times \sqrt{\widehat{Var}(\hat{X})} \right). \quad (20)$$

Aquí se recomienda que el valor de α (y, por tanto, el de $z_{1-\alpha/2}$) sea el mismo que se utilizó en el cálculo del tamaño de muestra, ver el apartado 3.2.

Anexos

A. Muestra de viviendas con hablantes de lenguas indígenas

Uno de los objetivos específicos de la *ENUT 2024* es obtener información detallada sobre las actividades cotidianas de mujeres y hombres, con base en sus características demográficas y socioeconómicas, a fin de visibilizar brechas de género y entre las poblaciones indígenas. Dada la dispersión y la proporción relativamente menor de la población hablante de lengua indígena en el país, el *MMM* por sí solo no es suficiente para identificar eficientemente a esta población y garantizar un número adecuado de casos en muestra para generar estimaciones precisas.

Por ello, se diseñó e implementó una muestra específica para esta subpoblación de 2 000 viviendas con hablantes de lenguas indígenas. Para construir el marco de muestreo para esta muestra, se utilizó la información del Censo de Población y Vivienda de 2020, que identifica a las personas que hablan lenguas indígenas, lo que permitió centrarse en las localidades con una concentración significativa de esta subpoblación (85 % o más).

La selección de esta muestra se realizó mediante un diseño probabilístico de tres etapas de la siguiente forma:

- **Primera etapa:** en esta etapa se seleccionaron localidades mediante muestreo aleatorio simple sin reemplazo.
- **Segunda etapa:** dentro de cada localidad seleccionada en la primera etapa, se procedió a la selección de manzanas. La selección de estas unidades se realizó mediante muestreo con probabilidades proporcionales al tamaño, utilizando el número de viviendas particulares habitadas como medida de tamaño de la unidad.
- **Tercera etapa:** en cada manzana seleccionada en la segunda etapa, se realizó un listado *in situ* de todas las viviendas particulares habitadas. Posteriormente, se seleccionaron 10 viviendas por muestreo aleatorio simple sin reemplazo de este listado.

La tabla al final de este anexo detalla la distribución de la muestra por entidad federativa.

Los ponderadores utilizados para las estimaciones de los parámetros de interés obtenidos de esta muestra se calcularon de acuerdo con este diseño de muestreo específico, considerando las probabilidades de selección en cada etapa.

Finalmente, la información captada de esta muestra fue integrada a la muestra nacional de viviendas para efectos de estimar los parámetros de interés del proyecto. Esta integración se realiza durante el proceso de ponderación y calibración, permitiendo que los ponderadores finales reflejen adecuadamente la probabilidad total de selección de cada individuo y garantizando la representatividad para la población hablante de lengua indígena a los niveles de desagregación establecidos.

Distribución de la muestra por entidad federativa

Entidad	Localidades en el Marco de Muestreo	Localidades en Muestra	Viviendas en Muestra
Chiapas	1 027	34	340
Chihuahua	40	3	30
Guerrero	330	14	140
Hidalgo	81	13	135
México	2	2	145
Michoacán de Ocampo	35	4	50
Oaxaca	562	42	425
Puebla	201	20	210
San Luis Potosí	275	8	80
Veracruz de Ignacio de la Llave	217	25	250
Yucatán	31	19	195
Total	2 801	184	2 000

B. Distribución de la muestra nacional de viviendas por entidad federativa y dominios

Entidad	Urbano Alto	Complemento Urbano	Rural	Total
Estados Unidos Mexicanos	16 740	7 910	7 398	32 048
Aguascalientes	655	160	120	935
Baja California	780	100	60	940
Baja California Sur	695	160	80	935
Campeche	475	240	220	935
Coahuila de Zaragoza	740	140	60	940
Colima	615	240	80	935
Chiapas	220	360	708	1 288
Chihuahua	720	120	130	970
Ciudad de México	920	0	20	940
Durango	535	160	240	935
Guanajuato	460	240	240	940
Guerrero	300	320	464	1 084
Hidalgo	215	395	464	1 074
Jalisco	560	280	100	940
México	700	230	155	1 085
Michoacán de Ocampo	300	415	276	991
Morelos	555	260	124	939
Nayarit	335	360	240	935
Nuevo León	760	140	40	940
Oaxaca	180	470	719	1 369
Puebla	440	365	349	1 154
Querétaro	555	220	160	935
Quintana Roo	735	120	80	935
San Luis Potosí	455	210	350	1 015
Sinaloa	515	220	201	936
Sonora	615	220	100	935
Tabasco	180	360	404	944
Tamaulipas	760	100	80	940
Tlaxcala	595	240	100	935
Veracruz de Ignacio de la llave	340	350	508	1 198
Yucatán	535	415	184	1 134
Zacatecas	295	300	342	937

Referencias

- Lohr, S. (2010). *Sampling: Design and Analysis* (Second ed.). Austria: Cengage Learning.
- Särndal, C., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. Alemania: Springer.
- Tillé, Y. (2019). *Sampling and Estimation from Finite Populations*. Reino Unido: Wiley.
- Valliant, R., & Dever, J. (2018). *Survey Weights: A Step-by-step Guide to Calculation*. Estados Unidos: Stata Press.