



MODECULT

Módulo sobre
Eventos Culturales
Seleccionados

2025

DISEÑO MUESTRAL



Módulo sobre Eventos Culturales Seleccionados 2025

MODECULT

Diseño muestral

Obras complementarias publicadas por el INEGI sobre el tema:

Módulo sobre Eventos Culturales seleccionados MODECULT 2016 Diseño conceptual.

Catalogación en la fuente INEGI:

306.97201 Instituto Nacional de Estadística y Geografía (México).
Módulo sobre Eventos Culturales seleccionados 2025 : MODECULT : diseño muestral / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2025.

19 p.

1. México - Cultura - Metodología. 2. México - Condiciones sociales - Metodología.

Conociendo México

800 111 4634

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

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

El presente documento detalla el diseño de muestreo para el Módulo sobre Eventos Culturales Seleccionados (Modecult). Este módulo es un esfuerzo del INEGI para generar información estadística sobre la asistencia y participación de la población de 12 años en adelante en eventos culturales.

El Modecult genera datos valiosos para comprender los hábitos de la población de México, tanto si asisten a eventos culturales en su localidad como en otras. Entre los temas que abarca se encuentran los eventos culturales a los que se asiste, los motivos para no asistir, la influencia de los estímulos culturales recibidos en la infancia, el tipo de participación o colaboración en actividades culturales y el papel en la organización o patrocinio de eventos. El módulo también explora las actividades culturales que las personas realizan cuando viajan a otras localidades y el consumo de eventos a través de internet, lo que ofrece una visión completa del panorama cultural del país.

La metodología de muestreo desarrollada para el Modecult está concebida para cumplir objetivos primordiales, como garantizar la representatividad de las estimaciones en los ámbitos de estudio, optimizar la precisión de los indicadores clave y asegurar la coherencia y comparabilidad de los datos. Esta estructura metodológica permitirá generar estadísticas sólidas y oportunas que contribuirán al conocimiento y análisis de la dinámica cultural del país y proporcionarán información valiosa para la toma de decisiones en materia de política cultural.

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	3
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	10
4.2.2 Ajuste por no respuesta de vivienda	10
4.2.3 Ajuste por no respuesta de persona	11
4.3 Calibración de ponderadores	11
4.4 Ponderadores para el proceso de estimación	11
5. Estimación	13
5.1 Estimación de totales y razones	13
5.2 Estimación de la varianza	13
5.3 Intervalos de confianza	14
6. Anexo	15
Distribución de la muestra nacional por entidad federativa y dominios	17
7. Referencias	19

1. Bases metodológicas

1.1 Objetivo general

El objetivo es generar información estadística sobre la asistencia a eventos culturales y la participación en actividades artísticas de la población de 12 años en adelante en México, así como ofrecer información sobre los estímulos culturales recibidos durante la infancia, el consumo de eventos culturales a través de Internet y las actividades culturales durante las visitas a otras localidades. De este modo, se proporcionarán datos clave a investigadores, expertos y cualquier persona interesada en el tema, y se contribuirá al diseño de políticas públicas.

1.2 Población objetivo

La población objetivo del Modocult está constituida por todas las personas de 12 años en adelante, residentes habituales en viviendas particulares habitadas dentro del territorio nacional.

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 para los dominios urbano y rural, y para el agregado urbano conformado por 32 ciudades, una por cada entidad federativa.

2. Marco de muestreo

2.1 Descripción del marco de muestreo

El Modecult utiliza el Marco Maestro de Muestreo (MMM) del INEGI como marco de muestreo. Dicho marco se construyó a partir de la información del último Censo de Población y Vivienda.

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, correspondiente al rango 1); 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

La conformación final de los estratos se realiza al interior de cada entidad federativa, combinando la clasificación por tamaño de localidad y nivel socioeconómico.

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 el Modecult. 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 del Modecult es estratificado, aprovechando la estratificación preexistente del MMM. Dentro de cada estrato, se implementa un muestreo en cuatro 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 a 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 el Modecult
Para el levantamiento específico del Modecult, 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
4. Cuarta etapa: selección de personas.
Dentro de cada vivienda seleccionada en la etapa anterior, se elige aleatoriamente a una persona de 12 años o más que resida habitualmente en la vivienda. Esta persona funge como informante para la aplicación del cuestionario del módulo Modecult. La selección aleatoria garantiza la representatividad de la población residente en los hogares encuestados

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.

3.2 Tamaño de la muestra

El tamaño de la muestra del Modecult 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 parámetros, basados en la experiencia del Modcult 2024 y en los objetivos de precisión establecidos:

- **Proporción por estimar (p):** se estableció una proporción del 5.62 % (o 0.0562 en proporción a la población objetivo)
- **Error relativo máximo de estimación (ε):** se estableció un error relativo máximo del 10 % (o 0.10)
- **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 $\varepsilon = 0.1$, el percentil correspondiente de una distribución normal estándar es $z = 1.6448$
- **Efecto de diseño ($DEFF$):** se estableció un efecto de diseño de 2.7
- **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 14 433 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 4 811 viviendas ($14\ 433/3=4\ 811$).

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 4 818 viviendas a nivel nacional.

Para estar en condiciones de proporcionar resultados a nivel nacional, el tamaño de muestra del Modcult está constituido por dos componentes. El primer componente corresponde a la muestra con 2 336 viviendas que tienen una cobertura en el agregado de las 32 ciudades del país de mayor importancia. El segundo componente corresponde a una muestra adicional de 2 482 viviendas, distribuidas en los ámbitos urbano alto, complemento urbano y rural de cada entidad federativa. Cabe mencionar que el tamaño de muestra no es fijo debido a que en el ámbito nacional la muestra de las viviendas se hace seleccionando segmentos geográficos de aproximadamente cuatro viviendas en promedio cada uno, lo que hace que tengamos tamaños de muestra variable en cada mes de levantamiento de la encuesta.

3.3 Distribución de la muestra en los estratos

Para dar continuidad a la serie histórica del Modcult se asignaron 2 336 viviendas al agregado de las 32 ciudades. El resto de las viviendas se distribuyó aproximadamente igual en cada entidad federativa y, al interior de estas, la asignación del tamaño de muestra por estrato se realizó de manera a su tamaño.

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, y
- 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 seleccionaron cuatro viviendas en cada UPM en muestra.

Este tamaño de muestra por UPM determina 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.

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 cuatro 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 y
- 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 Modecult. En cada estrato se utilizó un diseño aleatorio simple sin reposición. 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 $f_{ehj}^{MODECULT}$, está dado por

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

donde $n_{eh}^{MODECULT}$ es el tamaño de muestra de UPM asignado al estrato h de la entidad federativa e para el Modecult.

3. Tercera etapa. En esta etapa se seleccionan viviendas por muestreo aleatorio simple sin reposición. 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

$$f_{ehjk}^V = \frac{V_{ehj}^*}{4}, \quad (6)$$

donde 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. Cuarta etapa. En esta etapa se selecciona de manera aleatoria a una persona de 12 o más años dentro de cada vivienda. Por lo tanto, en el estrato h de la entidad e , el ponderador de diseño correspondiente a esta etapa para la l -ésima persona de la k -ésima vivienda de la j -ésima UPM, denotado por f_{ehjkl}^P , está dado por

$$f_{ehjkl}^P = \frac{Q_{ehjk}}{1}, \quad (7)$$

donde Q_{ehjk} es el total de personas de 12 años o más en la k -ésima vivienda de la j -ésima UPM, perteneciente a la MM del estrato h en 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

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 socio-económico), el factor de corrección por no respuesta, $FNR_{eh}^{MODECULT}$, está dado por

$$FNR_{eh}^{MODECULT} = \frac{\sum_{MODECULT} f_{ehj}^{MM}}{\sum_{r_{eh}^{MODECULT}} f_{ehl}^{MM}}, \quad (8)$$

donde $s_{eh}^{MODECULT}$ es la muestra de UPM del Modecult y $r_{eh}^{MODECULT}$ es el conjunto de UPM con respuesta en esta misma muestra.

4.2.2 Ajuste por no respuesta de vivienda

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 la 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

$$FNR_{eh}^V = \frac{4 \times \sum_{r_{eh}}^{MODECULT} \{f_{ehj}^{MM} \times V_{ehj}^*\}}{\sum_{r_{eh}}^{MODECULT} \{f_{ehl}^{MM} \times V_{ehl}^* \times R_{ehl}^V\}}, \quad (9)$$

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 persona

Finalmente, en cada vivienda que tuvo respuesta, se selecciona aleatoriamente una persona de la población objetivo. Si no es posible captar información de esta (e.g., ausencia prolongada o rechazo individual), es necesario calcular un factor de corrección por no respuesta de la persona.

Por lo tanto, para el estrato eh , el factor de corrección por no respuesta de la persona, denotado por FNR_{eh}^P , está dado por

$$FNR_{eh}^P = \frac{\sum_{r_{eh}}^{MODECULT} \{f_{ehj}^{MM} \times V_{ehj}^* \times [\sum_{r_{ehj}}^V \eta_{ehjk}]\}}{\sum_{r_{eh}}^{MODECULT} \{f_{ehl}^{MM} \times V_{ehl}^* \times [\sum_{r_{ehl}}^V \eta_{ehjk} R_{ehl}^P]\}}, \quad (10)$$

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 personas de la población objetivo del hogar principal en la k -ésima vivienda en muestra.
- R_{ehl}^P es uno si hubo respuesta de la persona seleccionada y cero en otro caso.

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 la población estimada por el INEGI dividida en tres grupos de edad: de 0 a 11 años, de 12 a 17 y de 18 y más años.

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 *ranking* con estas variables es que la muestra de este proyecto estime, con la mayor precisión posible, la población total de los tres grupos de edad antes mencionados.

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

$$w_{ehjk}^V = g_{ehj} \times f_{ehj}^{MM} \times [f_{ehj}^{MODECULT} \times FNR_{eh}^{MODECULT}] \times [f_{ehjk}^V \times FNR_{eh}^V], \quad (12)$$

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

$$w_{ehjkl}^P = w_{ehjk}^V \times f_{ehjkl}^P \times FNR_{eh}^P. \quad (13)$$

5. Estimación

5.1 Estimación de totales y razones

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

$$X = \sum_{e=1}^{32} \sum_{h=1}^{H_e} \sum_{i=1}^{N_{he}} \sum_{k=1}^{V_{ehi}^*} x_{ehik}, \quad (14)$$

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 y
- 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 4 y la información de la muestra, la estimación del parámetro X se puede expresar por

$$\hat{X} = \sum_{e=1}^{32} \sum_{h=1}^{H_e} \sum_{s_{eh}^{MODECULT}} \sum_{s_{ehj}^V} w_{ehjl}^V \times x_{ehjl}, \quad (15)$$

donde $s_{eh}^{MODECULT}$ 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

$$\hat{R} = \frac{\hat{X}}{\hat{Y}}, \quad (16)$$

donde \hat{X} y \hat{Y} están dados por la expresión (15).

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.

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

$$\widehat{Var}(\hat{X}) = \sum_{e=1}^{32} \sum_{h=1}^{H_e} \left\{ \frac{n_{eh}^{MODECULT}}{(n_{eh}^{MODECULT} - 1)} \times \sum_{s_{eh}^{ENUT}} \left[w_{ehj}^{MODECULT} \times \hat{X}_{ehj} - \frac{\hat{X}_{eh}}{n_{eh}^{MODECULT}} \right]^2 \right\}, \quad (17)$$

donde:

- $w_{ehj}^{MODECULT} = g_{ehj} \times f_{ehj}^{MM} \times f_{ehj}^{MODECULT} \times FNR_{eh}^{MODECULT}$, 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

$$\hat{X}_{ehj} = \sum_{s_{ehj}^V} f_{ehjk}^V \times FNR_{eh}^V \times x_{ehjl}. \quad (18)$$

- La estimación del parámetro de interés en el estrato eh es

$$\hat{X}_{eh} = \sum_{s_{eh}^{MODECULT}} w_{ehl}^{MODECULT} \times \hat{X}_{ehl}. \quad (19)$$

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):

$$\begin{aligned} \widehat{Var}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_{e=1}^{32} \sum_{h=1}^{H_e} \left\{ \frac{n_{eh}^{MODECULT}}{(n_{eh}^{MODECULT} - 1)} \right. \\ \times \sum_{s_{eh}^{MODECULT}} \left[\left(w_{ehj}^{MODECULT} \times \hat{X}_{ehj} - \frac{\hat{X}_{eh}}{n_{eh}^{MODECULT}} \right) \right. \\ \left. \left. - \hat{R} \times \left(w_{ehj}^{MODECULT} \times \hat{Y}_{ehj} - \frac{\hat{Y}_{eh}}{n_{eh}^{MODECULT}} \right) \right]^2 \right\} \end{aligned} \quad (20)$$

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.

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

$$\left(\hat{X} - z_{1-\alpha/2} \times \sqrt{\widehat{Var}(\hat{X})}, \hat{X} + z_{1-\alpha/2} \times \sqrt{\widehat{Var}(\hat{X})} \right). \quad (21)$$

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.

Anexo

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

Entidad	Ciudad no autorrepresentada				Ciudad autorrepresentada	Total
	Complemento urbano	Rural	Urbano alto	Total		
Estados Unidos Mexicanos	1 012	866	604	2 482	2 336	4 818
Aguascalientes	44	36	0	80	48	128
Baja California	16	8	48	72	56	128
Baja California Sur	20	12	44	76	40	116
Campeche	28	28	24	80	40	120
Coahuila de Zaragoza	20	8	52	80	56	136
Colima	28	12	28	68	48	116
Chiapas	28	49	12	89	48	137
Chihuahua	12	12	52	76	48	124
Ciudad de México	0	64	0	64	200	264
Durango	20	32	28	80	48	128
Guanajuato	28	32	24	84	80	164
Guerrero	32	40	12	84	48	132
Hidalgo	40	40	4	84	48	132
Jalisco	52	20	8	80	256	336
México	28	16	0	44	232	276
Michoacán de Ocampo	36	28	12	76	48	124
Morelos	44	20	16	80	56	136
Nayarit	44	28	0	72	48	120
Nuevo León	52	12	12	76	256	332
Oaxaca	32	44	4	80	48	128
Puebla	44	32	4	80	64	144
Querétaro	40	28	12	80	48	128
Quintana Roo	24	12	44	80	48	128
San Luis Potosí	28	45	8	81	48	129
Sinaloa	28	28	28	84	48	132
Sonora	28	12	40	80	48	128
Tabasco	40	44	0	84	40	124
Tamaulipas	12	8	60	80	48	128
Tlaxcala	48	24	0	72	40	112
Veracruz de Ignacio de la Llave	28	32	20	80	64	144
Yucatán	56	24	0	80	48	128
Zacatecas	32	36	8	76	40	116

Referencias

- Lohr, S. (2010). *Sampling: Design and Analysis* (segunda edicion). 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.