

Instituto Nacional de Estadística y Geografía

**Encuesta sobre la percepción Pública
de la Ciencia y la Tecnología
en México 2009**

ENPECyT

Documento metodológico de diseño muestral

| ÍNDICE | Página |
|---|--------|
| 1. Objetivo de la encuesta | 1 |
| 2. Población objetivo | 1 |
| 3. Cobertura geográfica | 1 |
| 4. Diseño de la muestra | 1 |
| 5. Marco de la encuesta | 1 |
| 5.1 Formación de las unidades primarias de muestreo (UPM) | 1 |
| En urbano alto | 1 |
| 5.2 Estratificación | 2 |
| 6. Esquema de muestreo | 2 |
| a) Probabilístico | 2 |
| b) Estratificado | 2 |
| c) Bietápico | 2 |
| d) Por conglomerados | 2 |
| 7. Tamaño de la muestra | 2 |
| 8. Afijación de la muestra | 3 |
| 9. Selección de la muestra | 3 |
| 10. Ajuste a los factores de expansión | 4 |
| 10.1 Ajuste por no Respuesta a nivel vivienda | 4 |
| 10.2 Ajuste por no Respuesta a nivel hogar | 5 |
| 10.3 Ajuste por proyección | 5 |
| 11. Estimadores | 5 |
| 12. Estimación de las precisiones | 6 |
| Anexos | |
| - Indicadores empleados en la estratificación del marco de la muestra maestra por ámbito de estudio (Cuadro 1) | |
| - Distribución de la muestra en UPM y viviendas por área urbana para la Encuesta sobre la Percepción Pública de la Ciencia y la Tecnología en México (ENPECyT) 2009 (Cuadro 2). | |

1. Objetivo de la encuesta

Obtener información estadística que permita generar múltiples indicadores que midan el conocimiento, entendimiento y actitud de las personas hacia las actividades científicas y tecnológicas.

2. Población objetivo

La población objetivo son las personas de 18 años cumplidos y más que residen en viviendas particulares ubicadas en las áreas urbanas de 100,000 y más habitantes.

3. Cobertura geográfica

La ENPECyT 2009 proporciona resultados al agregado de las áreas urbanas de 100,000 y más habitantes.

4. Diseño de la muestra

El diseño de la muestra para la ENPECyT 2009 se caracteriza por ser probabilístico, en consecuencia, los resultados obtenidos de la encuesta se generalizan a toda la población. A su vez, el diseño es bietápico, estratificado y por conglomerados, donde la unidad última de selección es la vivienda y la unidad de observación es el hogar.

5. Marco de la encuesta

El marco de muestreo que se empleó para la ENPECyT 2009 es el Marco Nacional de Viviendas 2002 del INEGI, construido a partir de información cartográfica y demográfica que se obtuvo del XII Censo General de Población y Vivienda 2000. Este marco es en realidad una muestra maestra, a partir de la cual se seleccionan las muestras para todas las encuestas en viviendas que realiza el INEGI; como tal, su diseño es probabilístico, estratificado, unietápico y por conglomerados, a los que se denominó unidades primarias de muestreo (UPM), pues es en ellos donde se seleccionan en una segunda etapa, las viviendas que integran las muestras de las diferentes encuestas.

5.1 Formación de las unidades primarias de muestreo (UPM)

Las unidades primarias de muestreo están constituidas por agrupaciones de viviendas con características diferenciadas dependiendo del ámbito al que pertenecen, en el caso del urbano alto es de la siguiente manera:

En urbano alto

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

- Una manzana.
- La unión de dos o más manzanas contiguas de la misma AGEB.¹
- La unión de dos o más manzanas contiguas de diferentes AGEB de la misma localidad.
- La unión de dos o más manzanas contiguas de diferentes localidades, pero del mismo tamaño de localidad.

5.2 Estratificación

En el ámbito urbano alto, se forman cuatro estratos en los que se agrupan todas las UPM del país, esta estratificación considera las características sociodemográficas de los habitantes de las viviendas, así como, las características físicas y

¹ Área Geoestadística Básica.

el equipamiento de las mismas, expresadas por medio de 21 indicadores* contruidos con información del XII Censo General de Población y Vivienda 2000, para lo cual se emplearon métodos estadísticos multivariados.

6. Esquema de muestreo

El tipo de muestreo empleado es probabilístico, estratificado, bietápico y por conglomerados. A continuación, se mencionan las definiciones de cada uno de estos métodos:

a) Probabilístico

Las unidades de muestreo tienen una probabilidad conocida y distinta de cero de ser seleccionadas, lo que significa que se puede hacer inferencia estadística de la población.

b) Estratificado

Las unidades primarias de muestreo con características socioeconómicas similares que pertenecen a un mismo tamaño de localidad, se agrupan para formar estratos.

c) Bietápico

La unidad última de muestreo (vivienda) es seleccionada en dos etapas.

d) Por conglomerados

Previamente a la primera selección, se forman conjuntos de unidades muestrales llamadas unidades primarias de muestreo (UPM), de las cuales se obtiene la muestra.

7. Tamaño de la muestra

Para el cálculo del tamaño de la muestra se analizaron algunos parámetros relacionados con los indicadores de interés, de los cuales se tomó como variable de referencia a la proporción de la población con nivel de escolaridad hasta primaria con interés moderado en el conocimiento de nuevos descubrimientos científicos, lo cual permite que las variables que se den en proporciones mayores queden automáticamente cubiertas. La expresión para el cálculo de la muestra es:

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

Donde:

- n = tamaño de la muestra
- p = estimación de la proporción de interés
- q = (1-p)
- z = valor en tablas estadísticas de la distribución normal para una confianza prefijada
- r = error relativo máximo esperado
- 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

* La descripción de estos indicadores se presenta en el cuadro 1.

Considerando una confianza de 90%, un error relativo máximo esperado de 15%, un efecto de diseño de 2.60, una tasa de no Respuesta de 15% y una proporción de 10.43%, se obtuvo una muestra de 3,159 viviendas, la cual se ajustó a 3,200 viviendas.

8. Afijación de la muestra

La afijación de la muestra se realiza dentro de cada ciudad entre los diferentes estratos de manera proporcional a su tamaño, para lo cual se emplea la siguiente expresión:

$$n_{ch} = \frac{N_{ch}}{N_c} n_c$$

Donde:

n_{ch} = número de viviendas en muestra en el h-ésimo estrato, en la c-ésima ciudad

n_c = número total de viviendas en muestra, en la c-ésima ciudad

N_{ch} = número total de viviendas en el h-ésimo estrato, en la c-ésima ciudad

N_c = número total de viviendas, en la c-ésima ciudad

En el cuadro 2 se presenta la distribución de la muestra por área urbana.

9. Selección de la muestra

La selección de la muestra para la ENPECyT 2009 se realizó en forma independiente para cada área urbana, el procedimiento es como sigue:

1. De las k_{ch} UPM que integran la muestra maestra² se eligieron k_{ch}^* UPM con igual probabilidad para la ENPECyT 2009
2. En cada UPM se seleccionaron cinco viviendas con igual probabilidad
3. En cada vivienda se seleccionó una persona de 18 años cumplidos y más

a) La probabilidad de seleccionar una vivienda de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad es:

$$P\{V_{chi}\} = \frac{k_{ch} m_{chi}}{m_{ch}} \frac{k_{ch}^*}{k_{ch}} \frac{5}{m_{chi}^*} = \frac{5 k_{ch}^* m_{chi}}{m_{ch} m_{chi}^*}$$

b) La probabilidad de seleccionar una persona de 18 años cumplidos y más de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad, H_{chij} , es:

$$P\{H_{chij}\} = \frac{k_{ch} \cdot m_{chi}}{m_{ch}} \frac{k_{ch}^*}{k_{ch}} \frac{5}{m_{chi}^*} \frac{1}{H} = \frac{5 k_{ch}^* m_{chi}}{m_{ch} m_{chi}^* H}$$

² Del total de UPM que integran el marco de propósitos múltiples se seleccionaron con probabilidad proporcional a su tamaño k_{ch} UPM para localidades 100,000 y más habitantes.

Su factor de expansión³ está dado por:

$$F_{chj} = \frac{H m_{ch} m_{chi}^*}{5 k_{ch}^* m_{chi}}$$

Donde:

- k_{ch} = número de UPM seleccionadas para el marco de la muestra maestra en el h-ésimo estrato, en la c-ésima ciudad
- m_{chi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad
- m_{ch} = total de viviendas en el h-ésimo estrato, en la c-ésima ciudad
- m_{chi}^* = total de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad al momento del levantamiento de la ENPECyT 2009
- k_{ch}^* = número de UPM seleccionadas, en el h-ésimo estrato, en la c-ésima ciudad
- H = número de personas de 18 años cumplidos y más en la vivienda seleccionada

10. Ajuste a los factores de expansión

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

10.1. Ajuste por no Respuesta a nivel vivienda

El ajuste por no Respuesta atribuida al informante se realiza a nivel UPM en cada uno de los dominios mediante la siguiente expresión:

$$F'_{chi} = F_{chi} \frac{V_{chi}}{V_{chi}^*}$$

Donde:

- F'_{chi} = factor de expansión corregido por no Respuesta para las viviendas de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad
- F_{chi} = factor de expansión de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad
- V_{chi} = número de viviendas habitadas seleccionadas en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad
- V_{chi}^* = número de viviendas habitadas seleccionadas en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad para las cuales se obtuvo respuesta

10.2. Ajuste por no Respuesta a nivel persona

Los factores de expansión “persona” se corrigieron por no Respuesta a nivel UPM, debido a que algunas de las personas de 18 años cumplidos y más que fueron seleccionadas, no contestaron.

$$F''_{chij} = F'_{chij} \frac{H}{H^*}$$

³ El factor de expansión se define como el inverso de la probabilidad de selección.

Donde:

- F_{chij}'' = factor de expansión corregido por no Respuesta a nivel persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad
- F_{chij}' = factor de expansión a nivel persona seleccionada en la UPM, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad
- H = número de personas de 18 años cumplidos y más, seleccionadas por UPM
- H^* = número de personas de 18 años cumplidos y más, seleccionadas por UPM, para las que se obtuvo respuesta

10.3. Ajuste por proyección

Los factores de expansión ajustados anteriormente se corrigen, a fin de asegurar que se obtenga la población determinada por la proyección de población generada por 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

11. Estimadores

El estimador del total de la característica X es:

$$\hat{X} = \sum_c \sum_h \sum_i \sum_j F_{chij} X_{chij}$$

Donde:

- F_{chij} = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad
- X_{chij} = valor observado de la característica de interés X de personas seleccionadas, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad

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

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

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

12. Estimación de las precisiones

Para la evaluación de los errores de muestreo de las principales estimaciones nacionales, se usó el método de Conglomerados Últimos,⁴ basado en que la mayor contribución a la varianza de un estimador en un diseño bietápico, es la que se presenta entre las unidades primarias de muestreo (UPM). El término “Conglomerados Últimos” se utiliza para denotar el total de unidades en muestra de una unidad primaria de muestreo.

Para obtener las precisiones de los estimadores de razón, se 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} :

$$\hat{V}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_c \left\{ \sum_h \frac{k_{ch}}{k_{ch}-1} \sum_i^{k_{ch}} \left[\left(\hat{X}_{chi} - \frac{1}{k_{ch}} \hat{X}_{ch} \right) - \hat{R} \left(\hat{Y}_{chi} - \frac{1}{k_{ch}} \hat{Y}_{ch} \right) \right]^2 \right\}$$

Donde:

\hat{X}_{chi} = total ponderado de la variable de estudio X en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad

\hat{X}_{ch} = total ponderado de la variable de estudio X en el h-ésimo estrato, en la c-ésima ciudad

k_{ch} = número de UPM en el h-ésimo estrato, en la c-ésima ciudad

Estas definiciones son análogas para la variable de estudio Y.

El cálculo de la varianza del estimador de un total, se obtiene con la siguiente expresión:

$$\hat{V}(\hat{X}_{NAL}) = \sum_c \sum_{h=1}^{L_c} \frac{k_{ch}}{k_{ch}-1} \sum_{i=1}^{k_{ch}} \left(\hat{X}_{chi} - \frac{1}{k_{ch}} \hat{X}_{ch} \right)^2$$

Las estimaciones de la desviación estándar (D.E.), coeficiente de variación (C.V.) y efecto de diseño (DEFF) se calculan mediante las siguientes expresiones:

$$D.E. = \sqrt{\hat{V}(\hat{\theta})}$$

$$C.V. = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}}$$

$$DEFF = \frac{\hat{V}(\hat{\theta})}{\hat{V}(\hat{\theta})_{MAS}}$$

Donde:

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

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

Finalmente, el intervalo de confianza a 90%, se calcula de la siguiente forma:

$$I_{1-\alpha} = \left(\hat{\theta} - 1.645 \sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + 1.645 \sqrt{\hat{V}(\hat{\theta})} \right)$$

Donde el nivel de significancia es de $\alpha = 0.10$.

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

| Indicadores empleados en la estratificación del marco de la muestra maestra por ámbito de estudio | | | | |
|---|-------------------|-------------|--------------------|-----------|
| Cuadro 1 | | | | |
| Descripción del indicador | Ámbito de estudio | | | |
| | Nacional | Urbano alto | Complemento urbano | Rural |
| % De vivienda | | | | |
| Que disponen de agua entubada dentro de la vivienda | | X | | |
| Con drenaje | X | X | | X |
| Con electricidad | | | | X |
| Que disponen de agua, luz y drenaje | X | X | X | X |
| Con piso diferente de tierra | X | | | X |
| Con paredes de material sólido | | | | X |
| Con cocina exclusiva | X | X | X | X |
| Sin hacinamiento | X | X | X | X |
| Con servicio sanitario exclusivo con conexión de agua | | X | | |
| Con servicio sanitario exclusivo con admisión de agua | X | | X | |
| Que utilizan gas para cocinar | X | | | X |
| Con radio o radiograbadora | X | | | X |
| Con televisión | X | | | |
| Con refrigerador | X | X | X | |
| Con licuadora | X | | | X |
| Con automóvil o camioneta propios | X | X | X | |
| Con videocasetera | | | X | |
| Con lavadora | X | | X | |
| Con teléfono | | X | X | |
| Con calentador de agua | | X | X | |
| Con cuatro bienes (teléfono, refrigerador, lavadora y boiler) | | X | | |
| Con cuatro bienes (radio, televisión, licuadora y refrigerador) | X | | X | |
| Con el mínimo equipamiento (radio o televisión y licuadora) | | | | X |
| % De población | | | | |
| Derechohabiente a servicio de salud | | X | X | |
| De 6 a 17 años que asiste a la escuela | X | X | X | |
| De 6 a 14 años que asiste a la escuela | | | | X |
| De 15 años y más alfabeto | X | | | X |
| De 15 años y más con postprimaria | X | X | X | X |
| Grado promedio de escolaridad | X | X | X | X |
| Ocupada que gana más de 2.5 salarios mínimos | X | X | X | X |
| Ocupada que gana más de 5 salarios mínimos | X | X | X | |
| Femenina de 12 años y más económicamente activa | X | X | X | |
| Económicamente activa de 20 a 49 años | X | X | X | |
| % De otros | | | | |
| Hogares de los deciles 8, 9 y 10 a nivel nacional | X | X | X | |
| Relación de dependencia económica | X | X | X | |
| Total de indicadores | 24 | 21 | 21 | 16 |

Distribución de la muestra en UPM y viviendas por área urbana

| Cuadro 2 | | | |
|----------|---------------------------------|-----|-----------|
| CVE | Nombre | UPM | Viviendas |
| 01 | Cd. México | 20 | 100 |
| 02 | Guadalajara | 20 | 100 |
| 03 | Monterrey | 20 | 100 |
| 04 | Puebla | 20 | 100 |
| 05 | León | 20 | 100 |
| 07 | San Luis Potosí | 20 | 100 |
| 08 | Mérida | 20 | 100 |
| 09 | Chihuahua | 20 | 100 |
| 10 | Tampico | 20 | 100 |
| 12 | Veracruz de Ignacio de la Llave | 20 | 100 |
| 13 | Acapulco | 20 | 100 |
| 14 | Aguascalientes | 20 | 100 |
| 15 | Morelia | 20 | 100 |
| 16 | Toluca | 20 | 100 |
| 17 | Saltillo | 20 | 100 |
| 18 | Villahermosa | 20 | 100 |
| 19 | Tuxtla Gutiérrez | 20 | 100 |
| 21 | Tijuana | 20 | 100 |
| 24 | Culiacán | 20 | 100 |
| 25 | Hermosillo | 20 | 100 |
| 26 | Durango | 20 | 100 |
| 27 | Tepic | 20 | 100 |
| 28 | Campeche | 20 | 100 |
| 29 | Cuernavaca | 20 | 100 |
| 31 | Oaxaca | 20 | 100 |
| 32 | Zacatecas | 20 | 100 |
| 33 | Colima | 20 | 100 |
| 36 | Querétaro | 20 | 100 |
| 39 | Tlaxcala | 20 | 100 |
| 40 | La Paz | 20 | 100 |
| 41 | Cancún | 20 | 100 |
| 43 | Pachuca | 20 | 100 |
| Total | | 640 | 3,200 |