

Encuesta sobre la Percepción Pública de la Ciencia y la Tecnología en México 2017

ENPECYT

Diseño muestral



**INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA**

DISEÑO MUESTRAL

**ENCUESTA SOBRE LA PERCEPCIÓN PÚBLICA DE LA
CIENCIA Y LA TECNOLOGÍA EN MÉXICO (ENPECyT-2017)**

Índice	Página
1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Diseño de la encuesta	1
4.1. Marco de la encuesta	1
4.2. Formación de las unidades primarias de muestreo (UPM)	1
4.3. Estratificación	2
5. Tamaño de la muestra	2
6. Afijación de la muestra	2
7. Selección de la muestra	3
8. Ajuste a los factores de expansión	4
8.1. Ajuste por No respuesta a nivel vivienda	4
8.2. Ajuste por No respuesta a nivel persona	4
8.3. Ajuste por proyección	4
9. Estimadores	5
10. Estimaciones de errores de muestreo	5
Anexos	
- Indicadores empleados en la estratificación del marco de la muestra maestra (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-2017) (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 áreas urbanas de 100 000 y más habitantes.

3. Cobertura geográfica

La ENPCyT-2017 proporciona resultados al agregado de 32 áreas urbanas de interés de 100 000 y más habitantes.

4. Diseño de la encuesta

El diseño de la muestra para la ENPCyT-2017 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 trietápico, estratificado y por conglomerados, donde la unidad última de selección es la persona que al momento de la entrevista tenga 18 años cumplidos o más.

4.1. Marco de la encuesta

El marco de muestreo que se empleó para la ENPCyT-2017 es el Marco Nacional de Viviendas 2012 del INEGI, construido a partir de información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2010. 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, pues es en ellos donde se seleccionan en una segunda etapa, las viviendas que integran las muestras de las diferentes encuestas.

4.2. 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.

¹ Área Geoestadística Básica.

4.3. Estratificación

En el ámbito urbano alto, se forman cuatro estratos en los que se agruparon todas las UPM del país, esta estratificación considera las características sociodemográficas de los habitantes de las viviendas, así como, las características físicas y el equipamiento de las mismas, expresadas por medio de 34 indicadores* construidos con información del Censo de Población y Vivienda 2010, para lo cual se emplearon métodos estadísticos multivariados.

5. 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 la proporción de la población masculina con nivel de escolaridad hasta primaria con un 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 estándar 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.

Considerando una confianza de 90%, un error relativo máximo esperado de 15%, un efecto de diseño de 3.35, una tasa de No respuesta de 15% y una proporción de 12.93%, se obtuvo una muestra de 3 159 viviendas, la cual se ajustó a 3 200 viviendas.

6. 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.

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

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

En el cuadro 2, se presenta la distribución de la muestra en UPM y viviendas por área urbana.

7. Selección de la muestra

La selección de la muestra para la ENPECyT-2017 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-2017.
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 es:

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

Su factor de expansión³ está dado por:

$$F_{chij} = \frac{H_{chij} m_{ch} m_{chi}^*}{5 k_{ch}^* m_{chi}}$$

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, según Censo de Población y Vivienda 2010.

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-2017.

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

H_{chij} = número de personas de 18 años cumplidos y más en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad.

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 de 100 000 y más habitantes.

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

8. Ajuste a los factores de expansión

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

8.1 Ajuste por No respuesta 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.

8.2 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.

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

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_{chij} = número de personas de 18 años cumplidos y más en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad.
- H^*_{chij} = número de personas de 18 años cumplidos y más, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad para las que se obtuvo respuesta.

8.3 Ajuste por proyección

Los factores de expansión ajustados por No respuesta se corrigen, a fin de asegurar que se obtenga la población determinada por la proyección de población generada por el INEGI 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 en que se efectúa el ajuste.

9. 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 la persona seleccionada, 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} .

10. Estimaciones de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones 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} :

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

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

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:

$$\hat{V}(\hat{X}_{NAL}) = \sum_{c=1}^{32} \sum_{h=1}^{L_c} \frac{k_{ch}}{k_{ch} - 1} \sum_{i=1}^{k_{ch}} \left(\hat{X}_{chi} - \frac{1}{k_{ch}} \hat{X}_{ch} \right)^2$$

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.

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

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

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

MNEMÓNICO	DESCRIPCIÓN
Población	
PPSSNOSP	Que tiene derecho a recibir servicios médicos en alguna institución de salud pública o privada excepto seguro popular.
PPDER_SS	Derechohabiente a servicios de salud.
PDP3A14A	De 3 a 14 años de edad que asiste a la escuela.
PDP15A24A	De 15 a 24 años de edad que asiste a la escuela.
PDP8A14ALF	De 8 a 14 años de edad que saben leer o escribir.
PDP15YM_SE	De 15 años o más de edad que aprobaron algún grado de escolaridad diferente al nivel preescolar.
PP15PRI_CO	De 15 años o más de edad que tienen como máxima escolaridad 6 grados aprobados en primaria.
PP15SEC_CO	De 15 años o más de edad que tienen como máxima escolaridad 3 grados aprobados en secundaria.
PGDO_ESC	Grado promedio de escolaridad.
PPEA	De 12 años y más que trabajaron; tenían trabajo pero no trabajaron o; buscaron trabajo en la semana de referencia.
PPEA_F	Femenina de 12 años y más que trabajaron; tenían trabajo pero no trabajaron o; buscaron trabajo en la semana de referencia.
PTASAOcupa	Tasa de ocupación.
TOCU12A17	No ocupada de 12 a 17 años de edad entre la población de 12 a 17 años de edad.
PPOMAYED	Ocupada de 18 y más años de edad entre la población ocupada.
Viviendas Particulares Habitadas	
PVIVSINH	Que no tienen hacinamiento.
PVPH_PISDT	Que tienen piso de cemento o firme, madera, mosaico u otro material.
PVPH2YMASD	Que usan para dormir entre 2 y 25 cuartos.
PVPH_2MASC	Que tienen más de un cuarto.
PVPH3YMASC	Que tienen entre 3 y 25 cuartos.
PVPH_C_ELE	Que disponen de luz eléctrica.
PVPHAGUADV	Que tienen disponibilidad de agua entubada dentro de la vivienda, o fuera de la vivienda pero dentro del terreno.
PVPH_EXCSA	Que tienen excusado, retrete, sanitario, letrina u hoyo negro.
PVPHDRENAJ	Que tienen drenaje conectado a la red pública, fosa séptica, barranca, grieta, río, lago o mar.
PVDRENERED	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	Total de manzanas sin hacinamiento.
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

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-2017) **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