

DISEÑO MUESTRAL

LOS MEXICANOS VISTOS POR SÍ MISMOS
LOS GRANDES TEMAS NACIONALES

211



INTRODUCCIÓN

La colección *Los mexicanos vistos por sí mismos. Los grandes temas nacionales* dibuja un panorama extenso del país, en textos en los que convergen la teoría, el contexto actual y la voz de mujeres y hombres. Para ello se recurrió a la aplicación de 25 encuestas en vivienda de 1 200 casos cada una, a personas de 15 años y más distribuidas en todo el país. Las encuestas levantadas para la colección fueron las siguientes:

1. Encuesta Nacional de Corrupción y Cultura de la Legalidad
2. Encuesta Nacional de Cultura, Lectura y Deporte
3. Encuesta Nacional de Derechos Humanos, Discriminación y Grupos Vulnerables
4. Encuesta Nacional de Familia
5. Encuesta Nacional de Salud
6. Encuesta Nacional de Seguridad Pública
7. Encuesta Nacional de Movilidad y Transporte



8. Encuesta Nacional de Pobreza
9. Encuesta Nacional de Migración
10. Encuesta Nacional de Género
11. Encuesta Nacional de Globalización
12. Encuesta Nacional de Niños, Adolescentes y Jóvenes
13. Encuesta Nacional sobre las Condiciones de Habitabilidad de la Vivienda
14. Encuesta Nacional de Envejecimiento
15. Encuesta Nacional de Religión, Secularización y Laicidad
16. Encuesta Nacional de Ciencia y Tecnología
17. Encuesta Nacional de Educación
18. Encuesta Nacional de Economía y Empleo
19. Encuesta Nacional de Indígenas
20. Encuesta Nacional de Justicia
21. Encuesta Nacional de Sociedad de la Información
22. Encuesta Nacional de Medio Ambiente
23. Encuesta Nacional de Federalismo
24. Encuesta Nacional de Identidad y Valores
25. Encuesta Nacional de Cultura Política

Cada encuesta contiene un diseño muestral que permite la sistematización de la información y la inferencia de los resultados en el ámbito nacional.

El objetivo de este apartado es dar a conocer la metodología asociada al diseño muestral general de la investigación. El primer apartado describe las fuentes de información utilizadas al seleccionar a los informantes para cada una de las etapas del muestreo. Posteriormente, se describe el diseño general que se implementó para las 25 encuestas. Por último, se presenta el procedimiento para obtener el tamaño de la muestra, así como el cálculo de las probabilidades de selección, los factores de expansión y su ajuste por variables sociodemográficas.

LA MUESTRA Y SUS ELEMENTOS

Con el fin de realizar cada una de las encuestas, se propuso la elaboración de distintas muestras en el país, una por cada tema de la investigación, siguiendo un diseño muestral similar en cuanto a la estratificación de la población y al esquema de selección. La selección de individuos dentro de una muestra es independiente de la selección de individuos en cualquier otra muestra de la investigación.

MARCO MUESTRAL

Se utilizó el programa Mapa Digital de México versión 6.0.1 y SCINCE 2010 versión 1.0.2 del Instituto Nacional de Estadística y Geografía (INEGI) para obtener la georreferenciación de todas las entidades del país, así como los datos de población desagregados en el ámbito de localidad¹ y de Área Geoestadística Básica (AGEB)² urbana.

PERIODO DE LEVANTAMIENTO

El periodo de levantamiento de las encuestas fue de octubre a noviembre de 2014.

POBLACIÓN OBJETIVO

Para esta investigación, la población objetivo la constituyeron todos los habitantes de 15 años o más. A cada habitante seleccionado en la muestra se le aplicó un cuestionario de opinión que recopila diversas opiniones y actitudes sobre uno de los temas dentro de la investigación.

¹ Localidad. El INEGI lo define como todo lugar ocupado con una o más viviendas y reconocido por un nombre dado por la ley o la costumbre.

² Área Geoestadística Básica: extensión territorial delimitada por el INEGI cuyos habitantes comparten características socioeconómicas parecidas.





ESTRATIFICACIÓN DEL PAÍS

Para fines de selección de la muestra, el país fue dividido tanto por regiones geográficas con características comunes como por tamaño de localidad. Las regiones se describen en el cuadro 1.

CUADRO 1

REGIONALIZACIÓN DEL PAÍS	
REGIÓN	ESTADOS DE LA REPÚBLICA
Centro	Aguascalientes, Colima, Guanajuato, Hidalgo, Jalisco, Michoacán, Morelos, Puebla, Querétaro, San Luis Potosí y Tlaxcala
DF-Estado de México	Distrito Federal y Estado de México
Norte	Baja California, Baja California Sur, Coahuila, Chihuahua, Durango, Nayarit, Nuevo León, Sinaloa, Sonora, Tamaulipas y Zacatecas
Sur	Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz y Yucatán

Diseño muestral *Los mexicanos vistos por sí mismos. Los grandes temas nacionales, 2014.*

De igual manera, el mapa 1 muestra la división del país por región.

MAPA 1



Dentro de cada región, de manera independiente, se estratificaron las localidades según su número de habitantes. Derivado de ello se establecieron cuatro estratos, los cuales se describen en el cuadro 2.

CUADRO 2

ESTRATIFICACIÓN POR TAMAÑO DE LOCALIDAD	
TIPO DE LOCALIDAD	TAMAÑO DE LA LOCALIDAD
I	100 000 habitantes o más
II	De 15 000 a 99 999 habitantes
III	De 2 500 a 14 999 habitantes
IV	2 499 habitantes o menos

Diseño muestral *Los mexicanos vistos por sí mismos. Los grandes temas nacionales*, 2014.

Con los criterios anteriores se obtuvieron 16 estratos para todo el país, tal y como se describen en el cuadro 3.

CUADRO 3

ESTRATOS GENERADOS POR REGIÓN Y TIPO DE LOCALIDAD			
ESTRATO	REGIÓN	TIPO DE LOCALIDAD	POBLACIÓN DE 15 AÑOS Y MÁS*
1	Centro	I	9 145 749
2	Centro	II	4 683 034
3	Centro	III	4 279 203
4	Centro	IV	6 160 059
5	DF-Estado de México	I	13 165 701
6	DF-Estado de México	II	1 461 226
7	DF-Estado de México	III	1 698 166
8	DF-Estado de México	IV	1 313 940
9	Norte	I	12 333 646
10	Norte	II	2 468 751
11	Norte	III	1 720 733
12	Norte	IV	3 016 044
13	Sur	I	4 792 845
14	Sur	II	2 954 728
15	Sur	III	3 374 837
16	Sur	IV	6 863 440

Fuente: Censo de Población y Vivienda 2010.

Diseño muestral *Los mexicanos vistos por sí mismos. Los grandes temas nacionales*, 2014.



ESQUEMA DE SELECCIÓN

La selección de unidades de muestreo se realizó a través de etapas sucesivas y de manera independiente para cada estrato previamente definido.

- *De localidades.* La selección de localidades que pertenecen a una zona se realizó con probabilidad proporcional al tamaño (PPT) de la población de 15 años o más.
- *De AGEBS.* Se seleccionaron dos AGEBS dentro de cada localidad de tamaños I, II y III, con probabilidad proporcional al tamaño de la población de la AGEB con remplazo. En el caso de las localidades con tamaño poblacional menor a 2500, no se consideró esta etapa, por lo que se pasó directamente a la selección de manzanas.
- *De manzanas.* Se seleccionaron tres manzanas³ para cada localidad y dicha selección se hizo de manera sistemática con arranque aleatorio (MSA).
- *De viviendas.* Se seleccionaron cuatro viviendas⁴ para cada manzana y dicha selección se hizo de forma sistemática, con arranque aleatorio.
- *De individuos.* Se seleccionó de manera aleatoria (MAS) a un individuo de 15 años o más dentro de la vivienda.

Derivado de dicho esquema se consideraron las siguientes unidades de muestreo:

- a) *Unidades primarias de muestreo (UPM):* en todos los estratos generados fueron las localidades.
- b) *Unidades secundarias de muestreo (USM):* en los estratos compuestos por los tipos de localidad I, II y III fueron las AGEB, mientras que en los estratos compuestos por el tipo de localidad IV fueron las manzanas.
- c) *Unidades terciarias de muestreo (UTM):* en los estratos compuestos por los tipos de localidad I, II y III fueron las manzanas, y en los estratos compuestos por el tipo de localidad IV fueron las viviendas.

³ Para el estrato sur-I se seleccionaron cuatro manzanas por AGEB.

⁴ Para el estrato centro-IV se seleccionaron cinco viviendas por manzana.



- d) *Unidades cuaternarias de muestreo (UCM)*: de igual manera, en los estratos compuestos por los tipos de localidad I, II y III fueron las viviendas y para los de tipo de localidad IV fueron los individuos que para estos estratos también se consideran como las unidades últimas de muestreo.
- e) *Unidades últimas de muestreo (UUM)*: dentro de los estratos por tipos de localidad I, II y III fueron los individuos de cada vivienda seleccionada.

El cuadro 4 indica el tamaño de muestra para cada estrato generado por región y tipo de localidad.

CUADRO 4

TAMAÑO DE MUESTRA PARA CADA ESTRATO

REGIÓN	TIPO DE LOCALIDAD	TAMAÑO DE MUESTRA
Centro	I	144
Centro	II	72
Centro	III	72
Centro	IV	60
DF-Estado de México	I	192
DF-Estado de México	II	48
DF-Estado de México	III	48
DF-Estado de México	IV	24
Norte	I	192
Norte	II	48
Norte	III	24
Norte	IV	48
Sur	I	96
Sur	II	48
Sur	III	48
Sur	IV	36

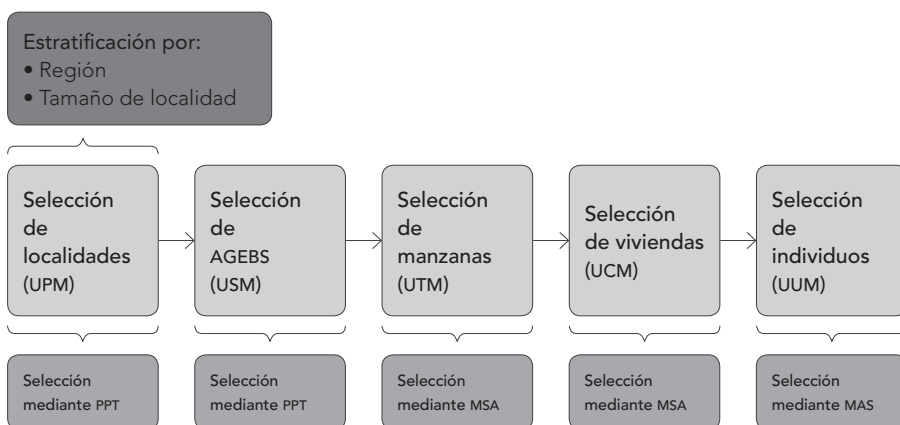
Diseño muestral *Los mexicanos vistos por sí mismos. Los grandes temas nacionales*, 2014.



Con fines prácticos, la gráfica 1 resume el esquema de selección descrito en los párrafos anteriores.

GRÁFICA 1

ESQUEMA DE SELECCIÓN UTILIZADO EN LAS 25 MUESTRAS



TAMAÑO DE MUESTRA

El tamaño de muestra obtenido para cada encuesta fue de 1 200 casos y fue distribuido entre los diferentes estratos descritos anteriormente con el fin de generar una disminución de los márgenes de error estadísticos. Para determinarlo se consideró la siguiente fórmula:

$$n = \frac{(z^2)(p)(1-p)(Deff)}{d^2(1-TNR)}$$

donde:

n = tamaño de la muestra;

z = valor en tablas de valores de probabilidad acumulada para la distribución normal estándar; este valor depende del nivel de confianza asignado $(1 - \alpha) * 100$ por ciento;

p = probabilidad de éxito del evento; se refiere a la probabilidad de éxito esperada. El más conservador de los valores se obtiene cuando $p = 0.5$;

d = diferencia entre el valor estimado y el valor poblacional;
TNR = tasa de no respuesta esperada;
Deff = efecto de diseño por utilizar un muestreo diferente del muestreo aleatorio simple.

Considerando un efecto de diseño (*Deff*) de 2.00, una tasa de no respuesta inferior (*TNR*) a 10 por ciento y un margen de error (d) de 4.2 puntos porcentuales con un nivel de confianza de 95 por ciento, se calculó para cada encuesta un tamaño de muestra aproximado de 1 200 casos, resultando un total de 30 000 cuestionarios por las 25 encuestas.



CÁLCULO DE LOS PONDERADORES

Para obtener los ponderadores utilizados dentro del cálculo de los estimadores de los parámetros tanto en individuos como en viviendas se utiliza la probabilidad de selección de cada etapa del muestreo.

EL CASO DE LOS ESTRATOS COMPUESTOS POR EL TIPO DE LOCALIDAD I, II Y III

La probabilidad de selección del individuo n -ésimo que pertenece a la vivienda m , manzana l , AGEB k , localidad j y estrato i se calcula de la siguiente forma:

$$P\left[x_{j,k,l,m,n}^i\right] = \frac{m_i N_j^i}{N^i} \frac{2N_{j,k}^i}{N_j^i} \frac{m_{i,j,k}}{I_{i,j,k}} \frac{m_{i,j,k,l}}{I_{i,j,k,l}} \frac{1}{I_{i,j,k,l,m}}$$

donde:

$x_{j,k,l,m,n}^i$ es el individuo n -ésimo que pertenece a la vivienda m , manzana l , AGEB k , localidad j y estrato i .

m_i es el número de localidades seleccionadas para el estrato i .

$m_{i,j,k}$ es el número de manzanas seleccionadas en la AGEB k , localidad j y estrato i .

$m_{i,j,k,l}$ es el número de viviendas seleccionadas en la manzana l , AGEB k , localidad j y estrato i .



N^i es la población total de personas de 15 años o más que habitan en el estrato i .

N_j^i es la población total de personas de 15 años o más que habitan en la localidad j y el estrato i .

$N_{j,k}^i$ es la población total de personas de 15 años o más que habitan en la AGEB urbana k , la localidad j y el estrato i .

$I_{i,j,k}$ es el número de manzanas existentes en la AGEB k , la localidad j y el estrato i .

$I_{i,j,k,l}$ es el número de viviendas en la manzana l , AGEB k , la localidad j y el estrato i .

$I_{i,j,k,l,m}$ es el número de individuos de 15 años o más que habitan la vivienda m , en la manzana l , la AGEB k , la localidad j y el estrato i .

El recíproco de la probabilidad de selección del individuo da el factor de expansión correspondiente, es decir:

$$F_{j,k,l,m,n}^i = \frac{1}{P[x_{j,k,l,m,n}^i]}$$

La probabilidad de selección de la vivienda m -ésima que pertenece a la manzana l , AGEB k , localidad j y estrato i se calcula de la siguiente forma:

$$P[x_{j,k,l,m}^i] = \frac{m_i N_j^i}{N^i} \frac{2N_{j,k}^i}{N_j^i} \frac{m_{i,j,k}}{I_{i,j,k}} \frac{m_{i,j,k,l}}{I_{i,j,k,l}}$$

Respecto al factor de expansión correspondiente, se obtiene de igual manera que el anterior, es decir:

$$F_{j,k,l,m}^i = \frac{1}{P[x_{j,k,l,m}^i]}$$

EL CASO DE LOS ESTRATOS COMPUESTOS POR EL TIPO DE LOCALIDAD IV

La probabilidad de selección del individuo n -ésimo que pertenece a la vivienda m , manzana l , localidad j y estrato i se calcula de la siguiente forma:



$$P[x_{j,l,m,n}^i] = \frac{m_i N_j^i}{N^i} \frac{m_{i,j}}{I_{i,j}} \frac{m_{i,j,l}}{I_{i,j,l}} \frac{1}{I_{i,j,l,m}}$$

donde:

$x_{j,l,m,n}^i$ es el individuo n -ésimo que pertenece a la vivienda m manzana l , localidad j y estrato i .

m_i es el número de localidades seleccionadas en el estrato i .

$m_{i,j}$ es el número de manzanas seleccionadas en la localidad j y el estrato i .

$m_{i,j,l}$ es el número de viviendas seleccionadas en la manzana l , localidad j y estrato i .

N_j^i es la población total de personas de 15 años o más que habitan en la localidad j y el estrato i .

N^i es la población total de personas de 15 años o más que habitan en el estrato i .

$I_{i,j}$ es el número de manzanas existentes en la localidad j y el estrato i .

$I_{i,j,l}$ es el número de viviendas en la manzana l , la localidad j y el estrato i .

$I_{i,j,l,m}$ es el número de individuos de 15 años o más que habitan la vivienda m , en la manzana l , localidad j y estrato i .

El recíproco de la probabilidad de selección del individuo da el factor de expansión correspondiente, es decir:

$$F_{j,l,m,n}^i = \frac{1}{P[x_{j,l,m,n}^i]}$$

La probabilidad de selección de la vivienda m -ésima que pertenece a la manzana l , localidad j y estrato i se calcula de la siguiente forma:

$$P[x_{j,l,m}^i] = \frac{m_i N_j^i}{N^i} \frac{m_{i,j}}{I_{i,j}} \frac{m_{i,j,l}}{I_{i,j,l}}$$

El recíproco de la probabilidad de selección de la vivienda da el factor de expansión correspondiente:

$$F_{j,l,m}^i = \frac{1}{P[x_{j,l,m}^i]}$$



AJUSTE DE LOS FACTORES DE EXPANSIÓN

Se calibraron los factores de expansión de acuerdo con el Censo de Población y Vivienda de 2010 con base en la edad⁵ por decenios, sexo y región. Esto con el fin de que los datos puedan expandir a la población en las proporciones que se presentan en el país. Este ajuste se logra multiplicando el factor de expansión original por una constante (componente de ajuste) que permita alcanzar tal objetivo.

Sea $X_{r,s,e}$ el conjunto de individuos x_i de la muestra que pertenecen a la región del país r , grupo decenal e y sexo s . Defínase $f_1(x_{r,s,e}) = \sum_{x_i \in X_{r,s,e}} \frac{1}{P[x_i]}$ como la expansión total del conjunto $X_{r,s,e}$, es decir, la suma de los factores de expansión correspondientes a los individuos que pertenecen a la misma región, grupo decenal y sexo.

Sea ahora $g(x_{r,s,e}) = N_{r,s,e}$ la población total de acuerdo con el Censo de Población y Vivienda de 2010 perteneciente a la región del r , grupo decenal e y sexo s . Con ello es posible calibrar cada factor de expansión de los individuos que componen la muestra. El factor de expansión ya corregido se calculó de la siguiente manera:

$$f_2(x_i) = \frac{g(X_{r,s,e})}{f_1(X_{r,s,e})} \frac{1}{P[x_i]}$$

$f_2(x_i)$ es el nuevo factor de expansión a utilizar del individuo i -ésimo de la muestra. Con ello se logra expandir los datos de la muestra al tamaño poblacional de cada uno de los nuevos estratos $X_{r,s,e}$:

$$\sum_{x_i \in X_{r,s,e}} f_2(x_i) = \sum_{x_i \in X_{r,s,e}} \frac{g(X_{r,s,e})}{f_1(X_{r,s,e})} \frac{1}{P[x_i]} = \frac{g(X_{r,s,e})}{f_1(X_{r,s,e})} \sum_{x_i \in X_{r,s,e}} \frac{1}{P[x_i]} = \frac{g(X_{r,s,e})}{f_1(X_{r,s,e})} f_1(X_{r,s,e}) = g(X_{r,s,e}) = N_{r,s,e}$$

⁵ El número de individuos cuya edad no fue especificada dentro del censo se distribuyó uniformemente dentro de los demás rangos de edades.