

# DISEÑO MUESTRAL

TERCERA ENCUESTA DE CULTURA CONSTITUCIONAL 2017

267



## INTRODUCCIÓN

Con la finalidad de replicar la Segunda Encuesta Nacional de Cultura Constitucional, se diseñó una muestra a nivel nacional, que sigue un diseño muestral similar en cuanto a la estratificación de la población y al esquema de selección realizados en el año 2010. Para recopilar la información se propuso un muestreo con un esquema polietápico, por conglomerados, como se detalla a continuación:

## 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 localidad<sup>1</sup> y de Área Geoestadística Básica (AGEB)<sup>2</sup> urbana.

## PERIODO DE LEVANTAMIENTO

El periodo de levantamiento de esta encuesta fue del 20 al 31 de octubre de 2016.

## POBLACIÓN OBJETIVO

Para esta investigación la población objetivo la constituyen todas las personas de 15 años o más que habitan en hogares distribuidos en todo el país.

## 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
Norte	Baja California, Baja California Sur, Chihuahua, Coahuila, Durango, Sinaloa, Sonora, Nuevo León, Tamaulipas.
Centro-occidente	Aguascalientes, Colima, Guanajuato, Jalisco, Michoacán de Ocampo, Nayarit, Querétaro, San Luis Potosí, Zacatecas.
Centro	Distrito Federal, Hidalgo, Estado de México, Morelos, Puebla, Tlaxcala.
Sur-sureste	Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz, Yucatán.

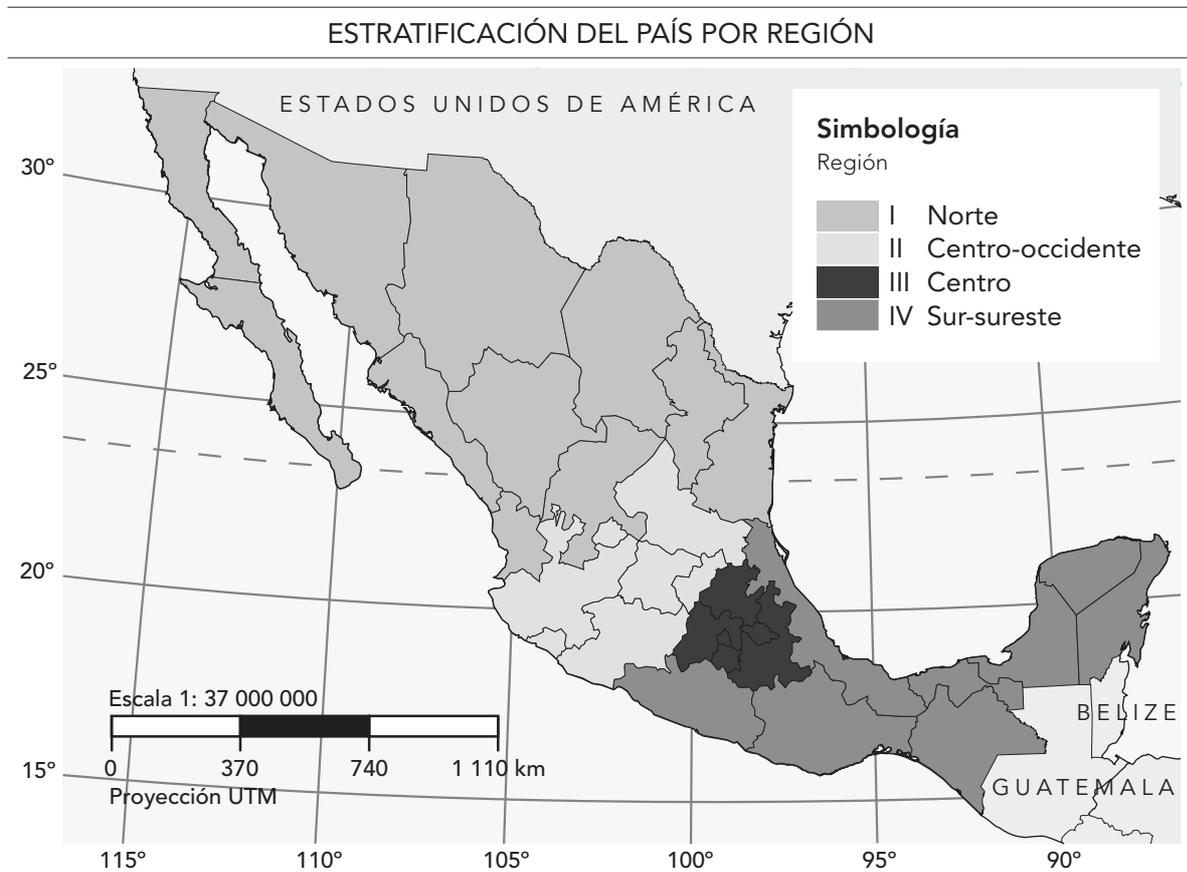
Fuente: Elaboración propia.

<sup>1</sup> 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.

<sup>2</sup> AGEB (Área Geoestadística Básica). Extensión territorial delimitada por el INEGI cuyos habitantes comparten características socioeconómicas parecidas.

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

**MAPA 1**



Fuente: Elaboración propia.

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 y más habitantes
II	De 15 000 a 99 999 habitantes
III	De 2 500 a 14 999 habitantes
IV	Menos de 2 500 habitantes

Fuente: Elaboración propia.





Con los criterios anteriores se obtuvieron 16 estratos para todo el país, tal y como se describe 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	Norte	I	11 826 157
2	Norte	II	2 188 477
3	Norte	III	1 357 198
4	Norte	IV	2 369 083
5	Centro-occidente	I	7 664 080
6	Centro-occidente	II	3 331 538
7	Centro-occidente	III	2 632 124
8	Centro-occidente	IV	4 514 182
9	Centro	I	15 154 860
10	Centro	II	3 092 997
11	Centro	III	3 708 779
12	Centro	IV	3 606 778
13	Sur-sureste	I	4 792 845
14	Sur-sureste	II	2 954 728
15	Sur-sureste	III	3 374 837
16	Sur-sureste	IV	6 863 440

Fuente. Censo de Población y Vivienda 2010, INEGI.

## ESQUEMA DE SELECCIÓN

Se propone un muestreo que seguirá un esquema polietápico, estratificado, por conglomerados. El país será dividido en estratos geográficos y poblacionales. Posteriormente la selección de las unidades de muestreo se realizará a través de etapas sucesivas y de manera independiente para cada uno de los estratos generados:

- *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.* La selección de AGEBS dentro de cada localidad se realizó con probabilidad proporcional al tamaño (PPT) de la AGEB. En el caso de las localidades de tamaño poblacional menor a 2 500, no se consideró esta etapa, por lo que pasó directamente a la selección de manzanas.
- *De manzanas.* Se seleccionaron dos manzanas para cada AGEB urbana ó localidad menor a 2 500 habitantes y dicha selección se hizo de forma sistemática, con arranque aleatorio (MSA).
- *De viviendas.* Se seleccionaron cuatro viviendas 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.
- 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.

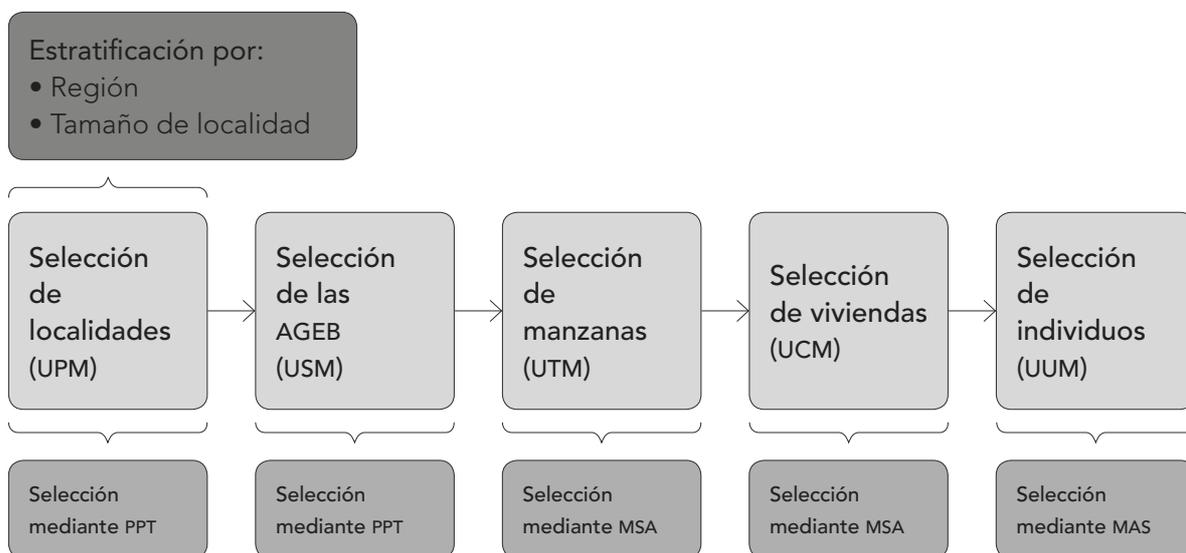




Con fines prácticos, la gráfica 1 resume el esquema de selección descrito en los párrafos anteriores.

## GRÁFICA 1

### ESQUEMA DE SELECCIÓN UTILIZADO EN LAS 25 MUESTRAS



Fuente: Elaboración propia.

## TAMAÑO DE MUESTRA

El tamaño de muestra fue de 1 200 casos, y fue distribuida entre los diferentes estratos, 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

$z^2$  = Es el valor en tablas de valores de probabilidad acumulada para la distribución normal estándar. Este valor depende del nivel de confianza asignado  $(1 - \alpha) * 100$ .

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

$d^2$  = es la diferencia entre el valor estimado y el valor poblacional (en este caso, elevado al cuadrado).

$TNR$  = Tasa de no respuesta esperada.

$Deff$  = Efecto de diseño por utilizar un muestreo diferente al muestreo aleatorio simple.

Considerando un efecto de diseño ( $Deff$ ) de 2.00, una tasa de no respuesta ( $TNR$ ) inferior a 5% y un margen de error ( $d$ ) de 4.1 puntos porcentuales con un nivel de confianza de 95%, se calculó un tamaño de muestra aproximado de 1 200 casos.

## CÁLCULO DE LOS PONDERADORES

Para obtener los ponderadores utilizados dentro del cálculo de los estimadores de los parámetros se utiliza la probabilidad de selección de cada una de las etapas 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:

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

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  $i$ .

$m_{i,j}$  es el número de AGEB seleccionadas para la localidad  $j$  del 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 estrato  $i$ .

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

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

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

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

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

## 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 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 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 estrato  $i$ .

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

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

## AJUSTE DE LOS FACTORES DE EXPANSIÓN

Se calibraron los factores de expansión de acuerdo al censo de población y vivienda de 2010 con base en la edad<sup>3</sup> por decenios, sexo y región, 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.

<sup>3</sup> El número de individuos, cuya edad no fue especificada dentro del censo, se distribuyó uniformemente dentro de los demás rangos de edades.





Sea ahora  $g(x_{r,s,e}) = N_{r,s,e}$  la población total de acuerdo al censo de población y vivienda de 2010 perteneciente a la región del país  $r$ , grupo decenal  $e$  y sexo  $s$ . Con ello es posible calibrar cada uno de los factores 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  $X_{r,s,e}$ :

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