

Diseño de la muestra censal 2010



Presentación

El **Instituto Nacional de Estadística y Geografía (INEGI)** realizó el Censo de Población y Vivienda del 31 de mayo al 25 de junio de 2010. El presente documento **Diseño de la Muestra Censal 2010** forma parte de los resultados definitivos de este proyecto estadístico y tiene la finalidad de poner a disposición de los usuarios información que les permita tener un panorama general de las bases metodológicas que rigieron el diseño y ejecución del censo, las etapas del proceso operativo y del tratamiento de la información, así como los principales componentes conceptuales que sustentaron la generación e integración de la información estadística que se proporciona.

De esta forma, el **INEGI**, como parte del Sistema Nacional de Información Estadística y Geográfica, cumple con la función de suministrar a la sociedad y al Estado información de calidad, pertinente, veraz y oportuna para contribuir al desarrollo de México.

Índice

Introducción	VII
Objetivos	1
Diseño de la muestra	1
Población objeto de estudio	1
Unidades de muestreo	2
Marco de muestreo	2
Dominios de estudio	2
Estratificación	3
Tamaño de muestra	3
Selección de la muestra	5
Factores de expansión	6
Estimadores	6
Error estándar (EE)	7
Microdatos de la muestra censal	8

Introducción

Para el levantamiento de la muestra censal 2010 se diseñó un cuestionario ampliado, con el que fueron censadas alrededor de 2.9 millones de viviendas en el país, seleccionadas con criterios probabilísticos. Al igual que en el Censo 2000, el cuestionario ampliado contiene todas las preguntas del cuestionario básico y preguntas adicionales sobre las características de las viviendas y sus habitantes. El Censo de Población y Vivienda 2010 considera un empadronamiento universal de las viviendas habitadas del país; así, cada vivienda es censada con un solo tipo de cuestionario y el total de la población se obtiene al sumar la información de ambos instrumentos de captación.

En este documento se describen todos los aspectos técnicos relacionados con la muestra censal 2010, el marco de muestreo, el tamaño de muestra, el esquema de muestreo, así como la obtención de estimadores. Es de vital importancia que el usuario de la información de la muestra censal considere el diseño de muestra para la obtención de los estimadores que requiera, ya que como se detallará más adelante, no se trata de una muestra autoponderada, esto es, las viviendas que conformaron la muestra no tienen igual probabilidad de selección. Esencialmente, por medio de la muestra censal se estiman tasas, promedios y proporciones; el usuario deberá tener en cuenta que cada uno de estos estimadores tiene asociadas una precisión y confianza, que indicarán la calidad del estimador.

El diseño de la muestra garantiza estimadores con precisión y confianza aceptables para la gran mayoría de variables a un nivel de desagregación de municipio o mayor; sin embargo, siempre que se obtenga un estimador se debe considerar la precisión y confianza, antes de tomar una decisión basada en este estimador.

Objetivos

El objetivo general de la muestra censal es:

- Proporcionar información a nivel municipal con precisión y confianza medibles para tasas, promedios y proporciones.

Entre los objetivos específicos de la muestra se tiene:

- Proporcionar información de las 75 variables con base en estimaciones que tengan precisión y confianza medibles para cada uno de los 2 456 municipios del país.
- Proporcionar información de las 75 variables con base en estimaciones que tengan precisión y confianza medibles para cada una de las localidades de 50 000 y más habitantes.
- Proporcionar información de las 75 variables al menor nivel de desagregación geográfico permitido por el criterio de confidencialidad de la información, para:
 - Todos los municipios de menos de 1 100 viviendas habitadas del país.
 - Los 125 municipios identificados como los de menor Índice de Desarrollo Humano (IDH) del país¹.

Diseño de la muestra

La muestra fue diseñada para garantizar el cumplimiento de los objetivos planteados. El diseño de la muestra censal 2010 es estratificado por conglomerados y se realizó en una sola etapa de selección, es decir, se seleccionan áreas geográficas completas, ya sean manzanas o localidades. Dentro de estas áreas se aplica el cuestionario ampliado a todas las viviendas particulares habitadas.

Población objeto de estudio

La población objeto de estudio está constituida por las viviendas particulares habitadas en el territorio nacional y por sus residentes habituales.

¹ Incluidos en la Estrategia 100x100, del Gobierno Federal.

Unidades de muestreo

Las Unidades Primarias de Muestreo (UPM) son áreas geográficas completas, ya sean manzanas o localidades. Por razones de tipo operativo y de control muestral, para el conjunto de localidades de menos de 250 viviendas habitadas, la unidad primaria de muestreo fueron las localidades completas, aun cuando estuvieran amanzandas o no.

Marco de muestreo

Para la conformación del marco de muestreo del Censo de Población y Vivienda 2010 se tomó como base la información generada por el Conteo de Población y Vivienda 2005 complementada con la generada por el Censo Agropecuario 2007 y los Censos Económicos 2009. Toda esta información se actualizó cartográficamente y se integró en un catálogo cartográfico por medio del cual se llevó a cabo la planeación operativa y la selección de la muestra censal 2010.

No se consideraron dentro del marco de muestreo algunas áreas del catálogo cartográfico, por no incluir población objeto de estudio. En particular las áreas no consideradas para la selección son:

- Ageb o localidades sin población residente habitual, en los últimos tres eventos (Conteo 2005, Censo Agropecuario 2007 y Censos Económicos 2009).
- Áreas en las que únicamente hay viviendas colectivas.
- Áreas con población pero sin viviendas, como aquellas en las que se tiene referenciada la población de indigentes.

Dominios de estudio

Los dominios de estudio definidos para la muestra censal son:

- Nacional
- Estatal
- Estatal con desagregación en cuatro tamaños de localidad:
 - Menos de 2 500 habitantes
 - De 2 500 a 14 999 habitantes
 - De 15 000 a 49 999 habitantes
 - 50 000 y más habitantes
- Municipal
- Localidades de 50 mil y más habitantes del país

Estratificación

Los 2 456 municipios del país se clasificaron en tres grupos de acuerdo a su tamaño, medido como el número de viviendas habitadas registradas. Los puntos de corte para definir estos grupos fueron:

- Menos de 1 100 viviendas habitadas
- 1 100 a 4 000 viviendas habitadas
- Más de 4 000 viviendas habitadas

Dados los requerimientos de información, se decidió que las viviendas habitadas de los municipios con menos de 1 100 viviendas habitadas, así como las viviendas de los 125 municipios con menor IDH, independientemente de su tamaño, se incluyeran con certeza en la muestra, esto es, con probabilidad igual a uno, por lo que la totalidad de viviendas de estos municipios se censaron con el cuestionario ampliado.

En el resto de los municipios se seleccionaron áreas completas (manzanas o localidades menores de 250 viviendas particulares habitadas). Al interior de cada municipio muestreado, se conformaron los siguientes estratos:

- Localidades de menos de 250 viviendas habitadas.
- Localidades de 250 viviendas a menos de 50 000 habitantes. Dentro de estas localidades, se conformaron estratos de acuerdo al número de habitantes de la localidad.
 - Menos de 50 habitantes
 - 50 a 499 habitantes
 - 500 a 999 habitantes
 - 1 000 a 1 499 habitantes
 - 1 500 a 1 999 habitantes
 - 2 000 a 2 499 habitantes
 - 2 500 a 4 999 habitantes
 - 5 000 a 14 999 habitantes
 - 15 000 a 49 999 habitantes
- Al interior de cada una de las localidades de 50 mil y más habitantes, cada Ageb constituyó un estrato.

Es conveniente aclarar que el tamaño de la localidad se definió con base en los resultados del II Censo de Población y Vivienda 2005.

Tamaño de muestra

Los tamaños de muestra se fijaron como sigue:

- 800 viviendas habitadas para los municipios de 1 100 a 4 000 viviendas habitadas.
- 1 100 viviendas habitadas para los municipios de 4 000 o más viviendas habitadas y que no cuentan con localidades de 50 mil y más habitantes.
- 800 viviendas habitadas para los municipios de 4 000 o más viviendas habitadas y que cuentan con localidades de 50 mil y más habitantes, es decir, este tamaño contempla al resto del municipio sin incluir a la localidad de 50 000 y más habitantes.
- En cada una de las localidades de 50 mil y más habitantes se estableció un tamaño de muestra de al menos 2 000 viviendas habitadas.

Al interior de cada municipio, la afijación de la muestra para cada estrato fue proporcional al número de viviendas habitadas por estrato. Para los municipios en el que todas sus viviendas entraron con certeza a la muestra, no fue necesario hacer la afijación.

En la siguiente tabla se puede ver un resumen de los tamaños de muestra considerados para cada municipio, según su clasificación:

Clasificación de municipios		Viviendas en muestra	Total de municipios	% de los municipios
Municipios con menos de 1 100 viviendas o 125 municipios con IDH más bajo		Censo	766	31.2
Municipios con 1 101 a 4 000 viviendas habitadas		800	794	32.3
Municipios con más de 4 000 viviendas habitadas	Sin localidades de 50 mil y más habitantes	1 100	703	28.6
	Con localidades de 50 mil y más habitantes (Sin considerar estas localidades)	800	193	7.9
	Cada localidad de 50 mil y más habitantes	Al menos 2 000		
Nacional			2 456	100

Estos tamaños de muestra se obtuvieron mediante la siguiente expresión:

$$n = \frac{z^2}{r^2} \times \frac{q}{p} \times \frac{DEFF}{1 - TNR}$$

donde:

p = Proporción que se desea estimar

$q = 1-p$

r = Error relativo máximo aceptable

z = Nivel de confianza

$DEFF$ = Efecto de diseño

TNR = Tasa de no respuesta

También se consideró un ajuste por población finita que se calculó como sigue:

$$n' = \frac{n}{1 + \frac{n}{N}}$$

donde:

n' = Tamaño de muestra ajustado por población finita

n = Tamaño de muestra estimado

N = Total de viviendas

Para el cálculo de los tamaños de muestra se consideró una confianza del 90%, un error relativo máximo de 0.2, un efecto de diseño de 1.44 y una tasa de respuesta de 90%.

Los tamaños de muestra fijados, garantizan estimaciones municipales aceptables, de acuerdo con los parámetros fijados, para proporciones cercanas a 0.01 o mayores.

Selección de la muestra

El esquema de selección es el mismo para cada municipio muestreado, sin embargo, las UPM varían de acuerdo con el estrato de cada municipio, como se menciona a continuación:

- En las localidades de menos de 250 viviendas habitadas se seleccionaron localidades completas.
- En las localidades de más de 250 viviendas habitadas y menos de 50 000 habitantes, se seleccionaron manzanas o localidades completas dentro de cada estrato de tamaño de localidad.
- Para las localidades de 50 mil y más habitantes, se seleccionaron en promedio 1.3 manzanas en cada una de las Ageb que conforman la localidad.

En todos los casos la selección se realizó mediante muestreo aleatorio simple.

Factores de expansión

De acuerdo con el esquema de muestreo, la probabilidad de selección de cada una de las UPM (manzanas o localidades) en el interior de cada estrato, está dada por la siguiente fórmula:

$$P_{hj} = \frac{n_h}{N_h}$$

donde:

P_{hj} = Probabilidad de selección de la j-ésima UPM en h-ésimo estrato

n_h = Número de UPM (manzanas o localidades) a seleccionar en el h-ésimo estrato

N_h = Total de UPM (manzanas o localidades) en el h-ésimo estrato

De esta manera, el factor de expansión para cada una de las viviendas (y cada uno de sus residentes) de la j-ésima UPM del h-ésimo estrato está dado por:

$$F_{hj} = \frac{1}{P_{hj}} = \frac{N_h}{n_h}$$

Para cada UPM, los factores de expansión se multiplicaron por un factor adicional que recibe el nombre de ajuste por no respuesta.

Lo anterior ratifica que el diseño de la muestra no es autoponderado, es decir, que cada vivienda en muestra representa a un número diferente de viviendas (los factores de expansión son diferentes entre las áreas y también la tasa de no respuesta es diferencial), y por ello cualquier indicador o tabulación a partir de la muestra censal requiere el uso de los factores de expansión.

Estimadores

El estimador del total para un dominio de estudio dado (por ejemplo, un municipio) es:

$$\hat{Y}_m = \sum_{h=1}^H \hat{Y}_h$$
$$\hat{Y}_m = \sum_{h=1}^H \sum_{j=1}^{n_h} F_{hj} \sum_{k=1}^{M_j} y_{hjk}$$

donde:

\hat{Y}_m = Es el total estimado de la característica de interés para el dominio m

\hat{Y}_h^m = Es el total estimado de la característica de interés en el estrato h del dominio m

F_{hj} = Es el factor de expansión de la j-ésima UPM en el h-ésimo estrato

y_{hjk} = Es el valor de la característica de interés en la k-ésima vivienda de la j-ésima UPM en el h-ésimo estrato

H = Número de estratos en el dominio m

M_j = Número de viviendas en muestra dentro de la j -ésima UPM en el h -ésimo estrato

El estimador de la media es:

$$\hat{Y}_m = \frac{\hat{Y}_m}{M_m}$$
$$= \frac{\hat{Y}_m}{\sum_{h=1}^H \sum_{j=1}^{n_h} F_{hj}}$$

donde:

M_m = Estimador del total de viviendas en el dominio m

Error estándar (EE)

Existen expresiones específicas para el cálculo del error estándar acorde con cada tipo de estimador (total, media, razón, proporción, etcétera).

Para la estimación de un total, el estimador de la varianza del estimador, para un dominio dado, está dada por:

$$\hat{V}(\hat{Y}_m) = \sum_{h=1}^H \hat{V}(\hat{Y}_h) = \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{S}_h^2}{n_h}$$

donde:

$$\hat{S}_h^2 = \sum_{j=1}^{n_h} \frac{(y_{hj} - \bar{y}_h)^2}{n_h - 1}$$

Por ende, el error estándar asociado está dado por:

$$EE_{\hat{Y}_m} = \sqrt{\hat{V}(\hat{Y}_m)}$$

Si lo que se requiere es el estimador de la varianza del estimador de la media de una característica en particular, se usa la fórmula:

$$\hat{V}(\hat{Y}_m) = \sum_{h=1}^H \hat{V}\left(\frac{\hat{Y}_m}{M_m}\right) = \sum_{h=1}^H \frac{1}{M_m^2} \hat{V}(\hat{Y}_m)$$

Y el error estándar para la media, está dado por:

$$EE_{\hat{Y}_m} = \sqrt{\hat{V}(\hat{Y}_m)}$$

Micro-datos de la muestra censal

Los micro-datos de la muestra censal 2010 están disponibles de manera gratuita en el portal electrónico del INEGI. Estos datos pueden ser usados en el paquete estadístico de preferencia del usuario.

El conjunto de datos de cada entidad federativa está compuesto por tres tablas: Viviendas, Personas y Migrantes.

Convencionalmente los paquetes estadísticos requieren para el cálculo de estimadores y sus varianzas que se especifiquen los campos donde se incluye la información sobre:

- Unidades Primarias de Muestreo
- Estratos considerados en el diseño
- Factores de expansión o ponderadores

Estos campos se han colocado al final de cada una de las tablas con los nombres: UPM, Estrato y Factor respectivamente.