

Diseño de la muestra

Las unidades de análisis de la muestra del XII Censo General de Población y Vivienda 2000 se seleccionan con criterios probabilísticos para poder realizar estimaciones de los indicadores de interés de la población bajo estudio y al mismo tiempo conocer la calidad de las mismas. Los indicadores a estimar son fundamentalmente tasas, promedios y proporciones.

El diseño de la muestra permite generar información con desagregación municipal en la mayor parte de los indicadores obtenidos a partir del cuestionario ampliado. Adicionalmente, será posible producir información para cada una de las localidades de 50 000 y más habitantes.

Esquema de muestreo

El esquema de muestreo es por conglomerados y en una sola etapa, es decir, se seleccionan áreas geográficas completas, ya sean Ageb⁶ (urbanas o rurales), manzanas (urbanas o rurales) o localidades rurales.

Entraron con certeza a la muestra todos los municipios del país, así como aquellas localidades que de acuerdo con la cifra obtenida durante el Censo de Población y Vivienda 1995 tenían 2 000 y más habitantes.

Marco muestral

Para definir el marco muestral se tomó como base la información generada por el Censo de Población y Vivienda 1995 para la parte rural, y por la Enumeración Integral 1998 tanto para la parte urbana como para las localidades rurales amezanadas.

Las unidades de muestreo varían de acuerdo con el tipo de área y el diseño muestral que se aplica en cada una de ellas.

- **Área urbana y localidades amezanadas de 2 000 a 2 499 habitantes.** La Unidad Primaria de

Muestreo (UPM) está constituida generalmente por **manzanas**, aunque en algunas localidades con baja densidad de población, la UPM está conformada por un **Ageb urbana completa**.

- **Área rural.** Las UPM pueden ser **localidades o Ageb rurales**.

El marco de muestreo lo constituyen los catálogos de Ageb, localidad y manzanas, así como la información estadística asociada a cada área y el material cartográfico que permite su identificación y ubicación en campo.

La selección de espacios geográficos completos hace más fácil aplicar un solo tipo de cuestionario en cada área: el ampliado en las seleccionadas y el básico en el resto. De esta manera, el total de población es resultante de sumar la información proveniente del cuestionario básico y del ampliado.

Tamaño de muestra y precisión

Para garantizar que las estimaciones que se obtengan a partir de la muestra del XII Censo General de Población y Vivienda tengan calidad aceptable, es necesario, entre otros elementos, que el tamaño de muestra que se defina para cada parámetro de interés sea suficiente.

Para la determinación de dicho tamaño muestral se seleccionó, dentro de las variables incluidas en el cuestionario ampliado, el indicador “hogares con perceptores de ingresos por otras fuentes distintas al salario”.

El tamaño de la muestra se obtiene mediante la siguiente expresión:

$$n = \frac{k^2}{r^2} * \frac{Q}{P} * \frac{DEFT}{1-TNR}$$

donde:

P = Proporción a estimar.

Q = $1-P$

⁶ Dependiendo de sus características, las Ageb se clasifican en dos tipos:

Ageb urbana. Área geográfica ocupada por un conjunto de manzanas (generalmente de 1 a 50) perfectamente delimitadas por calles, avenidas, andadores o cualquier otro rasgo de fácil identificación en el terreno cuyo uso del suelo es principalmente habitacional, industrial, de servicios, comercial, etcétera.

Ageb rural. Área geográfica delimitada por rasgos naturales (ríos, arroyos, barrancas, etc.) y culturales (vía de ferrocarril, líneas de conducción eléctrica, carreteras, brechas, etc.) que se caracteriza por el uso del suelo de tipo agropecuario o forestal. Contiene localidades, y su extensión territorial es variable.

Diseño de la muestra

r = Error relativo máximo aceptable.

k = Valor asentado en tablas estadísticas que garantiza realizar las estimaciones a una confianza prefijada.

$DEFT$ = Efecto de diseño, que es la pérdida o ganancia en la eficiencia del diseño por el efecto de conglomerar elementos de la población para formar unidades muestrales.

TNR = Tasa de no respuesta.

La fórmula anterior se aplicó para el indicador mencionado.

$P = 0.2386$ Proporción de los hogares con perceptores de ingresos por otras fuentes distintas al salario⁷.

$Q = 1 - P$

$r = 0.1427$ Error relativo máximo en la estimación de 14.27 por ciento.

$k = 1.645$ Valor en tablas para una confianza de 90 por ciento.

$DEFT = 1.44$ Efecto del diseño esperado⁷.

$TNR = 0.15$ Tasa de no respuesta máxima de 15 por ciento.

Así, se obtuvo que el tamaño de muestra era de 718 hogares. Dividiendo este total entre el promedio de hogares por vivienda, se obtuvo un tamaño de muestra en términos de viviendas igual a 700, lo cual significa que para obtener estimaciones a nivel municipal con calidad aceptable, se debe tener un mínimo de 700 viviendas en muestra por cada municipio, lo mismo que en localidades de 50 000 y más habitantes.

Debido a que varios municipios tienen un número de viviendas aproximado e incluso menor al tamaño de muestra descrito, se realizó un ajuste por "población finita" al tamaño de muestra mínimo requerido. Sobre todo en

entidades con gran cantidad de municipios y en su mayoría pequeños, ya que un tamaño de 700 viviendas por municipio provocaría una muestra demasiado grande en la entidad.

Así, para Chiapas, Hidalgo, Jalisco y Veracruz, el tamaño de muestra mínimo por municipio se fijó en 500 viviendas, mientras que en Oaxaca, Yucatán, Puebla y Tlaxcala la muestra mínima por municipio varía entre 150 y 250 viviendas, que representan entre 20 y 50% del total de viviendas del municipio. En estas entidades, los municipios con menos de 1 000 habitantes entraron completos a la muestra, es decir, fueron censados a través del cuestionario ampliado.

El ajuste por población finita se calculó como sigue:

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

donde:

n' = Tamaño de muestra ajustado por población finita.

n = Tamaño de muestra original.

N = Total de viviendas.

Aplicando el promedio de habitantes por vivienda⁸ a municipios con 1 000 habitantes:

$$N = \frac{1\,000_{\text{habs.}}}{4.9_{\text{habs./viv.}}} = 204_{\text{viv.}}$$

Sustituyendo este resultado en la fórmula anterior:

$$n' = \frac{700}{1 + \frac{700}{204}} = 158$$

Que para efectos prácticos se redondea a 150 viviendas.

En resumen, el tamaño de muestra de cada entidad depende de:

⁷ Valor derivado de la Encuesta del Censo de Población y Vivienda 1995.

⁸ Para el estado de Oaxaca el promedio fue 4.9, de acuerdo con los resultados del Censo de Población y Vivienda 1995.

Diseño de la muestra

- El total de municipios (ya que se fija un mínimo de 700 viviendas para cada municipio).
- El número de localidades de 50 000 y más habitantes.
- El total de Ageb urbanas.
- El promedio de viviendas por manzana (urbano) y localidad (rural).

A nivel nacional, el tamaño de muestra obtenido fue de aproximadamente 2.2 millones de viviendas.

Con base en los tamaños de muestra obtenidos, se calcularon las precisiones que se tendrían para algunos indicadores, así como sus posibilidades de explotación en los ámbitos municipal, estatal y nacional.

Precisión de los indicadores

<i>Variable</i>	<i>Nivel geográfico</i>	<i>Valor del indicador (porcentaje)</i>	<i>Error relativo máximo</i>
Causa de discapacidad			
Población discapacitada por accidente	Estatad	15.94	14.82
Bienes de la vivienda			
Viviendas con radio	Estatad	82.20	11.35
Viviendas con teléfono	Estatad	21.78	12.23
Educación			
Población de 15 y más años con primaria terminada	Municipal	18.84	9.70
Población analfabeta	Municipal	10.60	13.57
Causa de abandono escolar (7-24 años)			
Porque no quiso estudiar	Municipal	34.22	10.20
No tenía recursos	Municipal	22.87	13.51
Empleo			
Población económicamente activa	Municipal	54.94	4.00
Ocupados que trabajaron más de 48 horas	Municipal	31.77	8.75
Ocupados que trabajaron menos de 8 horas	Estatad	4.32	7.45
Hogares			
Con perceptores de riesgo por otras fuentes distintas al salario	Municipal	23.86	14.27
Con jefatura femenina	Estatad	14.10	5.22

Cabe aclarar que las precisiones para las localidades de 50 000 y más habitantes, serán similares a las presentadas en el ámbito municipal.

Los valores de los indicadores se obtuvieron de la Encuesta del Conteo de Población y Vivienda 1995 y del Censo Piloto del XII Censo General de Población y Vivienda 2000.

Como se observa en el cuadro, entre menor es el valor del indicador, las posibilidades de explotación se reducen en el ámbito municipal, no así en el estatal; sin embargo, se podrá ofrecer información confiable al presentar la característica más importante de cada distribución.

Diseño de la muestra

Procedimiento de selección

El procedimiento de selección aplicado en cada municipio para la muestra del XII Censo General de Población y Vivienda 2000 es diferente de acuerdo con el tipo de área (urbana o rural).

- **Área urbana.** En general, aunque todas las Ageb tienen muestra, sólo en algunas localidades urbanas con menos de 5 000 habitantes y baja densidad de población, esta regla no se cumple.

En este sentido, se tienen dos tipos de selección en el interior de las localidades urbanas:

- Selección de manzanas en el interior de cada Ageb.

Se hizo una selección de al menos dos manzanas en cada Ageb. La elección se realizó con probabilidad igual mediante un muestreo aleatorio simple.

Las Ageb con dos o menos manzanas, entraron completas a la muestra.

En los casos en donde la muestra no alcanzó el mínimo requerido en el municipio o localidad, se seleccionaron más manzanas por Ageb.

- Selección de Ageb completas en el interior de la localidad.

En algunas localidades urbanas menores de 5 000 habitantes y con un promedio de viviendas por manzana inferior a cinco, se optó por la selección de áreas más claramente definidas en campo, como lo son las Ageb, ya que en ocasiones la traza irregular o poco definida de las manzanas impidió identificar las áreas seleccionadas.

La selección de Ageb se realizó con probabilidad igual mediante un muestreo aleatorio simple.

- **Área rural.** Todas las localidades de 2 000 a 2 499 habitantes entraron con certeza a la muestra.

Si éstas contaban con ameznamiento regular, entonces se les daba tratamiento de localidad urbana, es decir, se seleccionaron en su interior dos o más manzanas; en cambio, cuando no contaban con plano, entraban completas a la muestra.

En el resto del área rural (localidades menores de 2 000 habitantes), el tipo de selección dependió de las características de cada municipio, por lo que se aplicaron tres esquemas de selección, que se presentan a continuación:

- Selección de localidades por cada Ageb del municipio.

Esta selección se caracteriza, al igual que en el área urbana, por una gran dispersión de la muestra, lo cual se traduce en una mejora en las precisiones de los estimadores.

Se escogieron dos o más localidades de cada Ageb (dependiendo del mínimo de viviendas requerido para el municipio). La selección fue aleatoria y con probabilidad igual.

- Selección de localidades en el interior del municipio.

Al tratarse de un muestreo por áreas, se pierde el control sobre el tamaño de muestra obtenido, ya que éste depende del número de viviendas de las áreas seleccionadas.

Con el objeto de controlar y reducir el tamaño de muestra cuando éste sobrepasa con mucho al requerido, se aplicó otra selección que consistió en estratificar las localidades rurales en el interior del municipio, según el total de habitantes; en el interior de cada estrato conformado se hizo una selección de algunas localidades, de acuerdo con el siguiente cuadro:

<i>Estrato</i>	<i>Rangos de tamaño de localidad (habitantes)</i>	<i>Número de localidades a seleccionar</i>
1	Con menos de 50	4
2	De 50 a 499	2
3	De 500 a 1 999	1

La selección en el interior de cada estrato se realizó con probabilidad igual y mediante muestreo sistemático.

- Selección de Ageb (rurales) en el interior del municipio.

La selección de Ageb rurales completas se aplica cuando el municipio tiene un gran número de localidades rurales y el promedio de viviendas por localidad es menor a 15. Dicha selección se realiza con probabilidad igual y con muestreo aleatorio.

En los tres esquemas anteriores, las Ageb que tienen menos de X viviendas⁹ entran completas a la muestra.

Procedimiento de estimación

De acuerdo con el proceso de selección, se tienen dos métodos para el cálculo de las probabilidades de selección de las viviendas.

El primero considera la selección de manzanas en el interior de cada Ageb urbana, selección de Ageb completas en el interior de la localidad urbana, selección de localidades por cada Ageb rural del municipio, así como selección de Ageb rurales en el interior del municipio. La fórmula que se emplea es la siguiente:

$$P_{ij} = \frac{n_i}{N_i}$$

donde:

P_{ij} = Probabilidad de selección de la j -ésima UPM en el ámbito de selección correspondiente i .

n_i = Número de UPM (manzanas, localidades o Ageb) a seleccionar. Con $n_i \geq 2$.

N_i = Total de UPM en el ámbito de selección (Ageb urbana o rural, municipio).

n_i y N_i varían según el tipo de selección que se realice, de acuerdo con el siguiente cuadro:

<i>Tipo de selección</i>	<i>n_i</i>	<i>N_i</i>
Area urbana		
Selección de manzanas	Número de manzanas a seleccionar en la i -ésima Ageb.	Total de manzanas en la i -ésima Ageb.
Selección de Ageb completas en el interior de la localidad urbana.	Número de Ageb a seleccionar en la i -ésima localidad urbana.	Total de Ageb en la i -ésima localidad urbana.
Area rural		
Selección de manzanas por localidad amezanada de 2 000 a 2 499 habitantes.	Número de manzanas a seleccionar en la i -ésima localidad rural.	Total de manzanas en la i -ésima localidad rural.
Selección de localidades por cada Ageb del municipio.	Número de localidades a seleccionar en la i -ésima Ageb.	Total de localidades en la i -ésima Ageb.
Selección de Ageb en el interior del municipio.	Número de Ageb a seleccionar en el i -ésimo municipio.	Total de Ageb en el i -ésimo municipio.

⁹ X va de 50 a 150 viviendas, dependiendo de la dispersión en la entidad.

Diseño de la muestra

El segundo caso es cuando en el área rural se realiza una selección de localidades en el interior del municipio, la expresión que se utiliza es:

$$P_{hi} = \frac{l_h}{L_h}$$

donde:

P_{hi} = Probabilidad de selección de la i -ésima localidad del estrato h en el municipio; todas las viviendas que conforman las localidades seleccionadas tienen la misma probabilidad de selección.

l_h = Número de localidades a seleccionar en el estrato h del municipio.

L_h = Total de localidades en el estrato h del municipio.

Estimadores

La construcción de totales de población se hizo en dos etapas:

- Inicialmente se obtuvieron estimaciones de totales a partir de los ponderadores (inverso de la probabilidad de selección).
- Posteriormente se utilizó un estimador separado de razón, el cual empleó como variable auxiliar a la población en viviendas particulares habitadas resultantes de la integración de las cifras preliminares del XII Censo General de Población y Vivienda 2000.

En general:

- El estimador del total municipal obtenido directamente de la muestra es:

$$\hat{X}_m = \sum_i \left(F_i \sum_j \sum_k x_{ijk} \right)$$

donde:

F_i = Es el inverso de la probabilidad de selección de las UPM correspondientes al ámbito de selección i .

x_{ijk} = Es el valor de la característica de interés en la k -ésima entrevista de la j -ésima UPM, en el ámbito de selección i .

- Para definir el estimador separado de razón se utilizó como variable auxiliar la población residente en viviendas particulares habitadas, obtenida durante la integración de las cifras preliminares del XII Censo General de Población y Vivienda 2000:

$$X_m^* = \sum \frac{\hat{X}_m}{\hat{Y}_m} Y_m$$

donde:

X_m^* = Estimador separado de razón del total de la característica X para el municipio m .

\hat{X}_m = Estimador del total de la característica X obtenido directamente de la muestra para el municipio m .

\hat{Y}_m = Estimador del total de población obtenido directamente de la muestra para el municipio m .

Y_m = Número de residentes generado de la cifra preliminar.

Para la estimación de proporciones, tasas y promedios se utilizaron estimadores de razón combinados, definiendo otra característica T de manera similar a X , quedando entonces:

$$R = \frac{X^*}{T^*}$$

Precisión y confianza

Debido a que los datos presentados se estimaron al incluir información de sólo una parte de la población (muestra), éstos pueden diferir de las cifras obtenidas a partir del Censo, en donde se realizó un recuento total de la población.

Diseño de la muestra

La magnitud de la diferencia entre el indicador muestral y el censal puede cuantificarse y estimarse a partir de la precisión y confianza de los estimadores, para los cuales se utilizaron las siguientes expresiones:

Error estándar (SE). El cálculo de error estándar para estimadores de razón de diseño de la muestra, se hizo con la siguiente fórmula:

$$SE = \sqrt{\text{Var}(\overset{*}{R})} = \sqrt{\frac{1}{T^2} \sum_m^M \sum_i^{I_m} \frac{n_{mi}}{n_{mi} - 1} (1 - f_{mi}) \left[\sum_j^{n_{mi}} d_{mij}^2 - \frac{d_{mi}^2}{n_{mi}} \right]}$$

donde:

M = Total de municipios

I_m = Ámbitos de selección en el municipio m (Ageb urbana, Ageb rural o parte rural del municipio).

n_{mi} = Total de UPM (manzanas, localidades o Ageb) seleccionadas en el ámbito de selección i del municipio m .

f_{mi} = Fracción de muestreo en cada ámbito de selección de cada municipio.

$$d_{mij} = \overset{*}{x'_{mij}} - R \overset{*}{t'_{mij}}$$

donde:

$\overset{*}{x'_{mij}}$ y $\overset{*}{t'_{mij}}$ son las estimaciones del total de población con característica X y característica T de la j -ésima UPM en el ámbito de selección i , municipio m , las cuales se calculan con la siguiente expresión:

$$\overset{*}{x'_{mij}} = \sum_k^{P_{mij}} (F_i x_{mijk})$$

para la j -ésima UPM, ámbito i , municipio m , donde P_{mij} es el total de unidades de análisis en la UPM.

$\overset{*}{t'_{mij}}$ se obtiene de manera análoga a $\overset{*}{x'_{mij}}$

$d_{mi} = \overset{*}{x'_{mi}} - R \overset{*}{t'_{mi}}$ es la suma para el i -ésimo ámbito de selección en el municipio m .

donde:

$\overset{*}{x'_{mi}}$ y $\overset{*}{t'_{mi}}$ son las estimaciones para los totales de la población con la característica x y t del ámbito de selección i del municipio m .

Coefficiente de variación (CV). Es una medida relativa, definida como el error estándar dividido entre el estimador.

$$CV = \frac{SE}{\overset{*}{R}}$$

Efecto del diseño (DEFT). Es la razón de comparación entre el error estándar de acuerdo con el diseño de la muestra y el error estándar bajo un diseño de muestreo aleatorio simple (SER).

$$DEFT = \frac{SE}{SER}$$

Intervalo de confianza al 90 por ciento:

$$\overset{*}{R} - 1.645 \overset{*}{SE} \text{ y } \overset{*}{R} + 1.645 \overset{*}{SE}$$

Para fines prácticos y simplificar su lectura, los resultados de la muestra del Censo se presentan con estimaciones puntuales, aunque su lectura e interpretación debe darse en términos de intervalos.

El intervalo de confianza indica que el valor real se encuentra entre el límite inferior y el límite superior del intervalo $\overset{*}{R} \pm 1.645 \overset{*}{SE}$ con una probabilidad del 90 por ciento.