

---

**ENCUESTA PARA LA DETERMINACIÓN DE  
UMBRALES MULTIDIMENSIONALES DE  
POBREZA 2007  
DISEÑO DE LA MUESTRA<sup>1</sup>**

---

**2007**

---

<sup>1</sup> Documento elaborado por la empresa Consulta Mitofsky, empresa ganadora de la Licitación Pública Nacional No. 20167001-001-07 para la elaboración del proyecto “ENCUESTA PARA LA DETERMINACIÓN DE UMBRALES MULTIDIMENSIONALES DE POBREZA 2007”.

# DISEÑO MUESTRAL

## NOTA METODOLÓGICA

### POBLACIÓN OBJETIVO

Consulta S.A. de C.V. toma en cuenta que la población objeto de este estudio está determinada por el total de hogares en viviendas particulares ubicadas en el territorio nacional. Una segunda población de estudio está constituida por la población mexicana de 18 años o más residente en el país.

### MARCO MUESTRAL

Para cubrir todos los estratos socioeconómicos del país, el marco muestral utilizado corresponde al Censos de Población y Vivienda del 2000 (*Hasta el día de hoy el INEGI no ha liberado al mercado datos del Censo de Población y Vivienda del 2005 a nivel de Ageb*). Por cuestiones de diseño el marco muestral se reagrupó en dos componentes y/o estratos los cuales se explican a continuación:

- a) Listado de localidades rurales del país (localidades menores a 15,000 habitantes y que no son cabeceras municipales) para obtener la muestra en las zonas rurales.
- b) Listado de localidades con al menos 15,000 habitantes y de las AGEB'S (Áreas Geoestadísticas Básicas) que las conforman, para seleccionar la muestra respectiva de la parte urbana del país.

## DISEÑO MUESTRAL

---

Se contempla un diseño probabilístico estratificado y polietápico donde cada elemento de la población o bien cada vivienda tiene una probabilidad de selección conocida y ésta es distinta de cero. Los dos marcos muestrales se prepararon para proceder a la 1ª etapa de selección, en la cual se eligen municipios como Unidades Primarias de Muestreo (UPM); posteriormente y abordando la 2ª etapa de selección se clasificaron, dentro de cada municipio, las AGEBS y Localidades Rurales, como Unidades Secundarias de Muestreo (USM) mediante un parámetro indicativo del estrato socio-económico a nivel nacional (denominado SES), el cual se estructuró de la siguiente manera:

- ✍ Primero se promediaron el porcentaje de población que recibe más de 5 salarios mínimos, el porcentaje de población con instrucción universitaria o mayor, el porcentaje de analfabetismo, el porcentaje de viviendas con drenaje y el porcentaje de población que recibe 2 salarios mínimos o menos en cada UPM.
  
- ✍ Con lo anterior, se produjo enseguida una variable continua (índice socioeconómico) que al acumularse a través de las UPM, previamente ordenadas descendientemente en función de la misma, se pudo categorizar el marco en los estratos que a continuación se describen:

- 1) Estrato Bajo
- 2) Estrato Medio-Bajo
- 3) Estrato Medio
- 4) Estrato Medio-Alto
- 5) Estrato Alto

La 3ª etapa de selección se realiza aleatoriamente sobre cartografía, en la cual se eligen dos manzanas o caseríos de los AGEBS o localidades rurales en muestra. Dentro de cada manzana o caserío se escogen 5 viviendas y dentro de cada una de ellas a un informante mayor de 18 años.

## TAMAÑO DE MUESTRA

El cálculo del tamaño de muestra se realiza mediante la siguiente expresión:

$$n = \frac{z^2 q \text{ deff}}{d^2 p(1 - \text{tnr})} \approx 2800 \text{ cuestionarios}$$

- z** es el valor en tablas estadísticas *z de una normal* necesario para lograr el nivel de confianza deseada. En este caso se utilizó una distribución normal para el 90% de confianza, resultando un valor de  $t=1.64$
- p** es el rango de las proporciones va de 0.10 a 0.90
- P** (1-p)
- d** es el error relativo máximo que se está dispuesto a aceptar
- Deff** Es el cociente de la varianza del diseño complejo entre la varianza de un muestreo simple aleatorio; en nuestro caso se consideró un efecto de diseño de 2.183 dato tomado de la encuesta sobre adicciones 2002
- tnr** es la tasa de rechazo del 12%

## AFIJACIÓN MUESTRAL

Se realiza con base en las proporciones poblacionales del CENSO de población y vivienda 2000, en el que se señala que el 61% de la población del país habita en localidades urbanas de más de 15,000 habitantes, mientras que el resto lo hace en localidades rurales de 15,000 o menos habitantes. Por lo anterior el tamaño de muestra de 2,800 ya calculado permite dar estimaciones para los siguientes niveles de desagregación:

- a) Nacional (2,800)
- b) Localidades rurales de menos de 15, 000 habitantes (1,100) – *Estrato Rural*
- c) Localidades urbanas con al menos 15, 000 habitantes (1,700) – *Estrato Urbano*

## PROCEDIMIENTO DE SELECCIÓN.

- a) Primera etapa de selección: Estrato urbano: Se ordenan las UPM de mayor a menor población urbana anidadamente dentro de los estratos SES, los cuales a su vez se ordenaron de mayor a menor denominación, y de ahí se seleccionan 85 con probabilidad proporcional al tamaño de su población. Estrato rural: Se repite el proceso anterior, ordenando las UPM de mayor a menor población rural, seleccionando 55 con probabilidad proporcional al tamaño de su población.
- b) Segunda etapa de selección: Dentro de cada municipio en muestra se ordenan las USM (AGEB) de mayor a menor tamaño de población; después se seleccionan 2 AGEB o localidades rurales, según el estrato que se trate (urbano o rural), 10 viviendas, 2 manzanas en cada una, 10 viviendas por manzana y un informante en cada vivienda de acuerdo al procedimiento descrito en el diseño muestral. Una vez seleccionado el hogar, se realizarán hasta 4 visitas hasta encontrar al informante adecuado para la entrevista.

Por lo anteriormente expuesto la probabilidad de que resulte seleccionado un individuo  $i$  que vive en la USM  $jh$  del municipio  $m$  es:

$$P_{mghi} = \frac{N_1 P_{mh}}{P_h} \times \frac{2P_{jh}}{P_{mh}} \times \frac{10}{P_{jh}} = \frac{20N_1}{P_h}$$

Donde:

$P_{mghi}$  : Probabilidad de selección del individuo  $i$ , de la USM  $jh$ , del municipio  $m$

$N_1$  : Es el número de municipios a seleccionar en ese estrato, sabiendo que en cada uno tendremos 20 viviendas

$P_{mh}$  : Es la población perteneciente al estrato  $h$  en ese municipio

$P_{jh}$  : Es la población total de la USM  $jh$

$P_h$  : Es la población total del estrato  $h$

## FACTORES DE EXPANSIÓN

Los factores de expansión o ponderadores son parámetros destinados a representar la variable medida en función de la probabilidad de selección que el diseño muestral determinó para cada unidad final de muestreo (informante). Dichos factores de expansión son calculados como el inverso de la probabilidad de selección en la ya mencionada unidad final de muestreo; de tal suerte que, para este estudio en específico El factor de expansión del individuo  $i$  que vive la USM  $jh$ , del municipio  $m$  es:

$$Fe_{mji} = \frac{1}{P_{mji}} = \frac{P_h}{20N_1}$$

Donde :

$Fe_{mji}$  : Factor de expansión del individuo  $i$ , de la USM  $jh$ , del municipio  $m$

$P_{mji}$  : Probabilidad de selección del individuo  $i$ , del municipio  $m$ , USM  $j$  que pertenece al estrato  $h$

$N_1$  : Es el número de UPMs seleccionadas en ese estrato, sabiendo que en cada una se toman 20 viviendas

$P_h$  : Es la población que habita en el estrato  $h$

Es necesario puntualizar que al realizar el análisis global de los cuestionarios, aplicando a cada uno el factor de expansión del individuo correspondiente, los totales y proporciones obtenidos reflejarán el panorama poblacional, de acuerdo a la confianza y error que el estudio asegura.

## TRATAMIENTO DE LA NO RESPUESTA

En la práctica solemos tener distorsiones en la muestra, lo cual altera al factor de expansión inicial, debido a la no respuesta en algunas unidades de muestreo. Para ello se corrige el factor de expansión por no respuesta, dando origen a un nuevo factor de expansión, dicho procedimiento consiste en corregir a nivel de agebs modificando los factores de expansión de las respuestas obtenidas, es decir, si  $F_i$  es el factor de expansión original de una persona en una ageb determinada, entonces:

$$F_i^* = F_i \frac{\sum_{i=1}^{i=10} F_i}{\sum_{i=1}^{i=k} F_i}$$

Donde:

- )  $F_i^*$ : Nuevo factor de expansión
- ) El numerador es la suma de los factores de expansión de las 10 personas que se debieron haber enumerado
- ) El denominador es la suma del total de personas que efectivamente se enumeraron

Este procedimiento es conocido en el argot como Posestratificación para ello se requiere:

- a) Información acerca de las proporciones de la población en los diversos estratos.
- b) Información que sirva para clasificar los casos de la muestra en los mismo estratos.

Debe hacerse hincapié que los criterios de clasificación deben ser los mismos para (a) y (b) de no ocurrir esto tenemos un procedimiento sesgado.

Para este estudio el marco muestral se construyó a partir del Censo de Población y Vivienda del 2000, los que nos permitió tener las proporciones de individuos mayores a 18 años y de viviendas particulares habitadas, para ambos estratos. Posteriormente del Censo de Población y Viviendas del 2005 se obtuvo la información de cada una de las subclases mencionadas anteriormente por estrato y se aplicó el procedimiento descrito anteriormente.

## ESTIMACIÓN Y ERROR

Con los factores de expansión anteriores la fórmula para estimar la proporción de individuos con una característica es:

$$P = \frac{\sum_{\Omega} F_{mjhi} I_{mjhi}}{\sum_{\Omega} F_{mjhi}}$$
$$I_{mjhi} = \begin{cases} 1 \rightarrow \text{Si el individuo tiene la característica de interés} \\ 0 \rightarrow \text{Si no la tiene} \end{cases}$$

Donde:

$\Omega =$  Es la población para la que se desea estimar la proporción

De igual manera, con los factores de expansión la fórmula para estimar el total de individuos con una característica es:

$$\hat{T} = \sum_{\Omega} F_{mjhi} I_{mjhi}$$

Donde:

$$I_{mjhi} = \begin{cases} 1 \rightarrow \text{Si el individuo tiene la característica de interés} \\ 0 \rightarrow \text{Si no la tiene} \end{cases}$$

$\Omega =$  Es la población para el que se está obteniendo el total

El cálculo de la varianza para una proporción estimada por la encuesta se hace mediante la fórmula:

$$\hat{V}(P) = \frac{PQ}{n} \quad (4.65)$$

Y el intervalo de confianza al 95%:

$$I_{95\%} = \left( P - (1.96)\sqrt{\hat{V}(P)}, P + (1.96)\sqrt{\hat{V}(P)} \right)$$