

Encuesta Nacional de Adicciones 2002



Diseño muestral



www.inegi.gob.mx



www.ssa.gob.mx



www.conadic.gob.mx



www.inprf.org.mx

DIRECCIÓN
GENERAL DE
EPIDEMIOLOGÍA

DR © 2004, **Instituto Nacional de Estadística,**
Geografía e Informática
Edificio Sede
Av. Héroe de Nacozari Sur Núm. 2301
Fracc. Jardines del Parque, CP 20270
Aguascalientes, Ags.

www.inegi.gob.mx
atencion.usuarios@inegi.gob.mx

Encuesta Nacional de Adicciones. Diseño muestral
ENA-2002

Impreso en México

DISEÑO MUESTRAL

**ENCUESTA NACIONAL DE ADICCIONES
(ENA)2002**

ÍNDICE	Página
1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Diseño de la muestra	1
4.1 Marco de la encuesta	1
4.2 Formación unidades de muestreo	2
a) Unidades primarias de muestreo (UPM)	2
b) Unidades secundarias de muestreo (USM)	2
c) Unidad terciaria de muestreo (UTM)	2
4.3 Estratificación	3
5. Esquema de muestreo	3
a) Probabilístico	3
b) Estratificado	3
c) Biétapico	3
d) Por conglomerados	3
6. Tamaño de la muestra	3
7. Afijación de la muestra	4
8. Selección de la muestra	4
8.1 Urbano alto (ciudades ENEU)	4
8.2 Urbano alto (ciudades no ENEU) y complemento urbano	5
8.3 Rural	6
9. Ajuste a los factores de expansión	6
9.1 Ajuste por no Respuesta	6
9.2 Ajuste por proyección	7
10. Estimadores	7
11. Estimación de las precisiones	8
 Anexos	
- Distribución de la muestra en viviendas por entidad federativa para la Encuesta Nacional de Adicciones	

1. Objetivo de la encuesta

El objetivo de la Encuesta Nacional de Adicciones 2002 (ENA-2002) es proporcionar información que permita determinar la prevalencia del consumo de drogas en México así como las características sociodemográficas de la población consumidora.

2. Población objetivo

La población objeto de estudio son las personas de 12 a 65 años residentes de las viviendas particulares habitadas ubicadas en el territorio nacional.

3. Cobertura geográfica

La encuesta tiene cobertura nacional.

4. Diseño de la muestra

El diseño es probabilístico, polietápico, estratificado y por conglomerados.

4.1 Marco de la encuesta

El marco muestral de la ENA-2002 está constituido con información demográfica y cartográfica obtenida a partir del levantamiento del Censo de Población y Vivienda de 1995, cuya organización es la siguiente:

Al interior de cada entidad federativa se conforman zonas que agrupan a las localidades como se muestra en el siguiente cuadro.

Zona	Descripción
Urbano alto.	<ul style="list-style-type: none">• Ciudades y áreas metropolitanas objeto de estudio de la Encuesta Nacional de Empleo Urbano (ENEU).• Resto de las ciudades de 100 000 y más habitantes y/o capitales de estado (No ENEU).
Complemento urbano de alta densidad.	<ul style="list-style-type: none">• Localidades de 15 000 a 99 999 habitantes.
Complemento urbano de baja densidad.	<ul style="list-style-type: none">• Localidades de 2 500 a 14 999 habitantes.
Rural.	<ul style="list-style-type: none">• Localidades con menos de 2 500 habitantes.

4.2 Formación de Unidades de Muestreo

a) Unidad primaria de muestreo (UPM)

Las Unidades Primarias de Muestreo están constituidas por una o parte de un Área Geoestadística Básica (AGEB)¹, o de varias AGEB colindantes, esto es, cuando no se completa el mínimo de viviendas requeridas para formar la UPM. Dependiendo de la zona de referencia su constitución es la siguiente:

UPM en urbano alto

- Un AGEB con un mínimo de 480 viviendas.
- La unión de dos o más AGEB colindantes y del mismo estrato, con un mínimo de 480 viviendas en conjunto.

UPM en el resto de las zonas

Un AGEB o la unión de dos o más AGEB que contengan:

- De 100 a 1 000 viviendas en localidades urbanas y
- De 50 a 850 viviendas en localidades rurales.

Cabe aclarar que en la zona rural, por lo regular, un AGEB está conformada por varias localidades con población menor a 2 500 habitantes, por lo tanto la dispersión de la población en el área geográfica puede ser muy grande, en este caso se clasifican en UPM con población no dispersa o de fácil acceso y UPM con población dispersa o de difícil acceso.

Dependiendo del tipo de situación que tenga la UPM, se le aplicó el tratamiento correspondiente en el momento de la selección de la última etapa.

b) Unidad secundaria de muestreo (USM)

USM en urbano alto

La formación de la USM o área de listado se realiza únicamente en las ciudades ENEU y está conformada por la agrupación de viviendas bajo las siguientes condiciones:

- Puede estar formada por una manzana que tenga un mínimo de 40 viviendas habitadas.
- Puede estar formada por dos o más manzanas contiguas con al menos 40 viviendas habitadas.

USM en el resto de las zonas

En las zonas definidas como no ENEU, la unidad secundaria de muestreo está constituida por las viviendas particulares habitadas permanentemente o aptas para habitarse.

c) Unidad terciaria de muestreo (UTM)

Las unidades terciarias de muestreo se definen solamente en la zona denominada ciudades ENEU, donde se conforman por las viviendas particulares habitadas permanentemente o aptas para habitarse.

¹ Áreas Geoestadísticas Básicas definidas por el INEGI de acuerdo al tipo de uso del suelo.

4.3 Estratificación

Las UPM se sometieron a una estratificación socioeconómica al interior de cada entidad y zona, utilizando las siguientes variables.

- Porcentaje de población de 6 a 14 años de edad que saben leer y escribir.
- Porcentaje de personas de 15 años y más de edad analfabeta.
- Porcentaje de viviendas con drenaje conectado a la red pública.
- Porcentaje de viviendas con agua entubada dentro de la vivienda.

Para la zona definida como rural, adicionalmente se añadió la siguiente variable:

- Porcentaje de viviendas con electricidad.

5. Esquema de muestreo

La ENA-2002 fue diseñada bajo un esquema de muestreo probabilístico, polietápico, estratificado y por conglomerados:

a) Probabilístico

Esto significa que todas las unidades de muestreo tienen una probabilidad conocida y distinta de cero de ser seleccionadas lo cual, permite conocer la precisión de los resultados obtenidos.

b) Estratificado

Porque las unidades de muestreo con características socioeconómicas similares, que pertenecen a un mismo tamaño de localidad se agrupan para formar estratos.

c) Polietápico

Porque la unidad última de selección (vivienda) es seleccionada después de varias etapas.

d) Por Conglomerados

Porque previamente se conforman conjuntos de unidades muestrales de los cuales se obtiene la muestra.

6. Tamaño de muestra

El tamaño de muestra fue calculado para dar estimaciones a nivel nacional tomando como variable pivote la proporción de personas que, según el levantamiento de 1998 manifestaron haber consumido alguna vez drogas ilegales, para lo cual se empleó la siguiente expresión.

$$n = \frac{z^2 q DEFF}{r^2 p (1 - tnr) (PHV)}$$

Donde:

n = es el tamaño de la muestra.

p = estimación de la proporción.

q = $(1-p)$.

r = error relativo máximo aceptable.

z = valor asentado en las tablas estadísticas de la distribución normal para una confianza prefijada.

DEFF = efecto de diseño definido como el cociente de la varianza obteniendo en la estimación con el diseño utilizado, entre la varianza del muestreo aleatorio simple para un mismo tamaño de muestra.

tnr = tasa de no Respuesta esperada.

PHV = promedio de habitantes por vivienda.

Fijando un nivel de confianza del 90%, un efecto de diseño de 2.183, un error relativo máximo esperado de 5.45%, una tasa de no Respuesta máxima esperada del 15%, un promedio de habitantes por vivienda de 3, para una proporción del 5.27%, se obtuvo un tamaño de muestra de 14 019 el cual se ajusto a 14 020 viviendas a nivel nacional.

7. Afijación de la muestra

La distribución de la muestra se hizo en una primera etapa de manera proporcional al número de viviendas, primero de la entidad y al interior de cada entidad al número de viviendas de sus estratos.

En una segunda etapa se hicieron ajustes a la distribución proporcional atendiendo a los siguientes aspectos:

- Garantizar una muestra mínima por entidad y por estrato.
- Prevee la no Respuesta en base a la experiencia obtenida en otras muestras para algunos dominios o estratos.
- La capacidad del marco en virtud de su distribución y la suficiencia de viviendas en algunos estratos o dominios en el momento del levantamiento.
- Los costos de levantamiento por dominio de estudio.

8. Selección de la muestra

La muestra de la ENA-2002 es una submuestra de la muestra maestra, la selección de ésta se realizó de forma independiente en cada entidad y estrato, bajo el siguiente procedimiento.

8.1 Urbano Alto (ciudades ENEU)

De las k áreas de listado seleccionadas por UPM en el dominio ENEU², se seleccionaron k^* áreas de listado para la ENA-2002, con igual probabilidad.

De éstas áreas de listado se seleccionaron cinco viviendas con igual probabilidad, en consecuencia la probabilidad de seleccionar una vivienda es:

² Del total de UPM formadas para integrar el marco, se seleccionaron n_{eh} UPM con probabilidad proporcional al tamaño para formar la muestra maestra en el dominio ENEU, en cada UPM seleccionada, se seleccionaron k USM o áreas de listado con probabilidad proporcional a su tamaño.

$$P\{V_{ehij}\} = \frac{n_{eh} m_{ehi}}{m_{eh}} \frac{k m_{ehij}}{m_{ehi}} \frac{k^*}{k} \frac{5}{m_{ehij}^*} = \frac{5 k^* n_{eh} m_{ehij}}{m_{eh} m_{ehij}^*}$$

Donde:

- n_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato en la e-ésima entidad.
- m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehij} = número total de viviendas en la j-ésima USM, de la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.
- k^* = número total de USM seleccionadas para la ENA-2002.
- m_{ehij}^* = número total de viviendas en la j-ésima USM, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento del levantamiento de la ENA-2002.
- K = áreas de listado seleccionadas por UPM.

Su factor de expansión está dado por.

$$F_{ehij} = \frac{m_{eh} m_{ehij}^*}{5 k^* n_{eh} m_{ehij}}$$

8.2 Urbano Alto (ciudades no ENEU) y Complemento Urbano

1. Entre las n_{eh} UPM seleccionadas para la muestra maestra³, se seleccionaron n_{eh}^* UPM con igual probabilidad.
2. En cada UPM seleccionada, se seleccionaron 20 viviendas con igual probabilidad.

Por lo tanto, la probabilidad de seleccionar una vivienda es.

$$P\{V_{ehi}\} = \frac{n_{eh} m_{ehi}}{m_{eh}} \frac{n_{eh}^*}{n_{eh}} \frac{20}{m_{ehi}^*} = \frac{20 n_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^*}$$

Donde:

- n_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi} = número de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.
- m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi}^* = número total de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento del levantamiento de la ENA-2002.
- n_{eh}^* = número de UPM seleccionadas para la ENA-2002 en el h-ésimo estrato, en la e-ésima entidad.

Su factor de expansión está dado por:

³ Del total de UPM que integran el marco en este dominio se seleccionaron n_{eh} UPM con probabilidad proporcional a su tamaño para integrar la muestra maestra.

$$F_{ehi} = \frac{m_{eh} m_{ehi}^*}{20 n_{eh}^* m_{ehi}}$$

8.3 Rural

1. Entre las n_{eh} UPM seleccionadas para la muestra maestra⁴ se seleccionaron n_{eh}^* UPM con igual probabilidad.
2. En cada UPM seleccionada se seleccionaron dos segmentos de 10 viviendas aproximadamente, con igual probabilidad.

Por lo tanto, la probabilidad de seleccionar una vivienda está dada por.

$$P \{ V_{ehi} \} = \frac{n_{eh} m_{ehi}}{m_{eh}} \frac{n_{eh}^*}{n_{eh}} \frac{2}{m_{ehi}^*} \frac{10}{m_{ehi}} = \frac{20 n_{eh}^* m_{ehi}}{m_{eh} m_{ehi}^*}$$

Donde:

- n_{eh} = número de UPM seleccionadas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi} = número de viviendas de la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.
- m_{eh} = número de viviendas en el h-ésimo estrato, en la e-ésima entidad.
- m_{ehi}^* = número total de viviendas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad al momento del levantamiento de la ENA-2002.
- n_{eh}^* = número de UPM seleccionadas para la ENA-2002 en el h-ésimo estrato, en la e-ésima entidad.

Su factor de expansión está dado por:

$$F_{ehi} = \frac{m_{eh} m_{ehi}^*}{20 n_{eh}^* m_{ehi}}$$

9. Ajustes a los factores de expansión

Los factores elaborados conforme al procedimiento antes descrito se ajustan para los siguientes conceptos:

9.1 Ajuste por no Respuesta

El ajuste por no Respuesta atribuida al informante se realiza a nivel USM o área de listado en el dominio ENEU, y a nivel UPM en el resto, mediante las siguientes expresiones.

En el dominio ENEU

$$F'_{ehij} = F_{ehij} \frac{nv_{ehij}}{nv_{ehij} CR_{ehij}}$$

⁴ Del total de UPM que integran el marco en este dominio se seleccionaron n_{eh} UPM con probabilidad proporcional a su tamaño para integrar la muestra maestra.

Donde:

- F'_{ehij} = factor de expansión corregido por no Respuesta para las viviendas de la j-ésima USM, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.
- F_{ehij} = factor de expansión de la j-ésima USM, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.
- nvh_{ehij} = número de viviendas habitadas de la j-ésima USM, en la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- $nvhCR_{ehij}$ = número de viviendas habitadas con respuesta del la j-ésima USM, de la i-ésima UPM, del h-ésimo estrato, del e-ésima entidad.

En el resto de los dominios

$$F'_{ehi} = F_{ehi} \frac{nvh_{ehi}}{nvhCR_{ehi}}$$

Donde:

- F'_{ehi} = factor de expansión corregido por no Respuesta para las viviendas de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- F_{ehi} = factor de expansión de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- nvh_{ehi} = número de viviendas habitadas de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.
- $nvhCR_{ehi}$ = número de viviendas habitadas con respuesta de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad.

9.2 Ajuste por proyección

Los factores ajustados por no Respuesta se corrigen, a fin de asegurar que en cada dominio de interés de la Encuesta se obtenga la población total determinada por la proyección de población generada por INEGI referida al punto medio del levantamiento, mediante la siguiente expresión:

$$F''_D = F'_D \frac{PROY_D}{PEXP_D}$$

Donde:

- F''_D = factor de expansión corregido por proyección en el dominio D.
- F'_D = factor de expansión corregido por no Respuesta en el dominio D.
- $PROY_D$ = población en el dominio D, según la proyección.
- $PEXP_D$ = población total a la que expande la encuesta en el dominio D.

10. Estimadores

El estimador del total de la característica X:

A nivel nacional es.

$$\hat{X}_{NAL} = \sum_h \sum_i \sum_j F_{ehij}^{ENEU} \left(\sum_s \sum_l X_{ehijsl}^{ENEU} \right) + \sum_h \sum_i F_{ehi}^R \left(\sum_s \sum_l X_{ehisl}^R \right) + \sum_h \sum_i F_{ehi}^{CU} \left(\sum_s \sum_l X_{ehisl}^{CU} \right)$$

Donde:

- F_{ehij}^{ENEU} = factor de expansión final de la j-ésima USM, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio ENEU.
- X_{ehijsl}^{ENEU} = valor observado de la característica de interés X de la l-ésima persona, de la s-ésima vivienda, de la j-ésima USM, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio ENEU.
- F_{ehi}^R = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio rural.
- X_{ehisl}^R = valor observado de la característica X en la l-ésima persona, de la s-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio rural.
- F_{ehi}^{CU} = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, en el dominio complemento urbano.
- X_{ehisl}^{CU} = valor observado de la característica X en la l-ésima persona, de la s-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio complemento urbano.

Para la estimación de proporciones, tasas y promedios, se utiliza el estimador de razón.

$$\hat{R}_{NAL} = \frac{\hat{X}_{NAL}}{\hat{Y}_{NAL}}$$

Donde, la variable \hat{Y}_{NAL} es definida en forma análoga a \hat{X}_{NAL} .

11. Estimación de las precisiones

Para la evaluación de los errores de muestreo de las principales estimaciones estatales y nacionales se usa el método de conglomerados últimos⁵, basado en que la mayor contribución a la varianza de un estimador, en un diseño polietápico, es la que se presenta entre las unidades primarias de muestreo (UPM), el término conglomerado último se utiliza para denotar el total de unidades en muestra de una unidad primaria de muestreo.

Para obtener las precisiones de los estimadores de razón, se aplica el método de series de Taylor conjuntamente con el método de conglomerados últimos, obteniéndose la siguiente fórmula para estimar la precisión de \hat{R}_{NAL} .

La expresión empleada para calcular la varianza de \hat{R}_{NAL} es la siguiente.

$$\hat{V}(\hat{R}_{NAL}) = \frac{1}{\hat{Y}_{NAL}^2} \sum_h \left\{ \sum_{eh} \frac{n_{eh}}{n_{eh} - 1} \sum_i^{n_{eh}} \left[\left(\hat{X}_{ehi} - \frac{1}{n_{eh}} \hat{X}_{eh} \right) - \hat{R}_{NAL} \left(\hat{Y}_{ehi} - \frac{1}{n_{eh}} \hat{Y}_{eh} \right) \right]^2 \right\}$$

⁵ Vease Hansen, M.H. Horwitz, W.N. y Madow, W.G, Sample Survey Methods and Theory, (1953), Vol. 1 página 242

Donde:

\hat{X}_{ehi} = total ponderado de la variable de estudio X en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

\hat{X}_{eh} = total ponderado de la variable de estudio X en el h-ésimo estrato, en la e-ésima entidad.

n_{eh} = número de UPM en el h-ésimo estrato, en la e-ésima entidad.

Estas definiciones son análogas para la variable de estudio Y.

La estimación de la varianza del estimador del total, se calcula con la siguiente expresión.

$$\hat{V}(\hat{X}_{NAL}) = \sum_{e=1}^{32} \sum_{h=1}^{L_e} \frac{n_{eh}}{n_{eh}-1} \sum_{i=1}^{n_{eh}} \left(\hat{X}_{ehi} - \frac{1}{n_{eh}} \hat{X}_{eh} \right)^2$$

Las estimaciones del error estándar (E.E.), coeficiente de variación (C.V.) o error relativo del estimador y el efecto de diseño (DEFF), se calculan mediante las siguientes expresiones.

$$E.E. = \sqrt{\hat{V}(\hat{\theta})}$$

$$C.V. = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}}$$

$$DEFF = \frac{\hat{V}(\hat{\theta})}{\hat{V}(\hat{\theta})_{mas}}$$

Donde:

$\hat{\theta}$ = estimador del parámetro poblacional θ .

$\hat{V}(\hat{\theta})_{mas}$ = estimador de la varianza, bajo muestreo aleatorio simple.

Finalmente, el Intervalo de Confianza al 90%, se construye de la siguiente forma.

$$I_{1-\alpha} = \left(\hat{\theta} - 1.645 \sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + 1.645 \sqrt{\hat{V}(\hat{\theta})} \right)$$

donde el nivel de significancia es de $\alpha = 0.10$.

Distribución de la muestra en viviendas por entidad federativa para la Encuesta Nacional de Adicciones 2002

Clave	Entidad	Viviendas			
		URBANO		RURAL	Total
		ENEU	Complemento URBANO		
01	Aguascalientes	170	80	100	350
02	Baja California	370	80	100	550
03	Baja California Sur	170	80	100	350
04	Campeche	210	40	100	350
05	Coahuila de Zaragoza	260	40	100	400
06	Colima	210	40	100	350
07	Chiapas	210	100	140	450
08	Chihuahua	380	80	140	600
09	Distrito Federal	560	0	40	600
10	Durango	240	60	100	400
11	Guanajuato	230	60	160	450
12	Guerrero	210	120	120	450
13	Hidalgo	200	80	120	400
14	Jalisco	450	80	120	650
15	México	340	60	100	500
16	Michoacán de Ocampo	270	160	120	550
17	Morelos	190	60	100	350
18	Nayarit	170	80	100	350
19	Nuevo León	490	60	100	650
20	Oaxaca	200	80	120	400
21	Puebla	210	100	140	450
22	Querétaro	210	40	100	350
23	Quintana Roo	170	80	100	350
24	San Luis Potosí	180	100	120	400
25	Sinaloa	150	80	120	350
26	Sonora	140	140	120	400
27	Tabasco	170	60	120	350
28	Tamaulipas	410	60	100	570
29	Tlaxcala	170	60	120	350
30	Veracruz de Ignacio de la Llave	300	80	120	500
31	Yucatán	170	80	100	350
32	Zacatecas	210	80	160	450
TOTAL		8 020	2 400	3 600	14 020