

DISEÑO MUESTRAL

**ENCUESTA NACIONAL SOBRE NIVELES
DE VIDA DE LOS HOGARES**

Línea Basal
(ENNViH 2002)

Dirección de Diseño y Marcos Estadísticos
Instituto Nacional de Estadística, Geografía e Informática (INEGI), 2004.

ENCUESTA NACIONAL SOBRE NIVELES DE VIDA EN HOGARES Línea Basal (ENNViH-1)

1. OBJETIVO DE LA ENCUESTA.

Obtener información estadística que permita generar múltiples indicadores sobre la dinámica sociodemográfica y económica de la población que conforman los hogares mexicanos.

2. POBLACIÓN OBJETIVO.

La población objeto de estudio son todas las personas que habitan en las viviendas particulares ubicadas dentro del territorio nacional.

3. COBERTURA GEOGRÁFICA.

La ENNViH-1 permitió dar resultados para tres dominios de estudio: nacional, urbano, rural y cinco regiones de interés. Las regiones son las mismas que se consideran en el plan Nacional de Desarrollo.

4. DISEÑO MUESTRAL.

El diseño de la muestra para la ENNViH-1 se caracteriza por ser probabilístico, en consecuencia, los resultados obtenidos de la encuesta se generalizan a toda la población. A su vez el diseño es polietápico, estratificado y por conglomerados, donde la unidad última de selección es la vivienda y la unidad de observación es el hogar.

4.1 MARCO MUESTRAL

El marco de muestreo que se empleó para la ENNViH-1 es el marco muestral de la Encuesta Nacional de Empleo (ENE)¹.

¹Ver documento metodológico de la ENE.

4.1.2 ESTRATIFICACIÓN DE LAS UNIDADES DE MUESTREO.

Las UPM's se clasifican en 3 estratos; alto, medio y bajo, con base a las siguientes variables captadas durante el levantamiento de la Encuesta Nacional de Empleo (ENE) del 2do. trimestre del 2001:

- Proporción de la población económicamente activa.
- Proporción de la población con más de 2 salarios mínimos de ingreso mensual.
- Proporción de la población ocupada en el sector secundario.
- Proporción de la población que son funcionarios y personal directivo.
- Proporción de la población con secundaria terminada.
- Proporción de la población con educación media superior.
- Proporción de viviendas con piso diferente de tierra.
- Proporción de viviendas con agua.
- Proporción de viviendas con drenaje.
- Nivel de no-hacinamiento.
- Proporción de la población de mujeres en edad reproductiva con embarazos ocurridos con atención en las etapas del ciclo reproductivo.
- Proporción de la población de mujeres en edad reproductiva con embarazos ocurridos con atención en el puerperio.
- Proporción de hijos nacidos vivos de la población de mujeres en edad reproductiva.
- Proporción de la población de 15 años y más derechohabiente.

Las UPM's estratificadas se agrupan en las 3 diferentes zonas como son:

- 1.- La zona ENEU está constituida por 48 ciudades y áreas metropolitanas de ENEU.
- 2.- La zona de Complemento Urbano está constituida por las ciudades con 100,000 y más habitantes que no están en la ENEU y por un complemento de localidades urbanas de 2,500 a 99,999 habitantes.
- 3.- La zona Rural está constituida por localidades con menos de 2,499 habitantes.

5. ESQUEMA DE MUESTREO

El esquema de muestreo es probabilístico, estratificado, polietápico y por conglomerados:

- Probabilístico. Porque todas las unidades de muestreo tienen una probabilidad conocida y distinta de cero de ser seleccionadas.
- Estratificado. Porque las unidades de muestreo con características geográficas y socioeconómicas similares se agrupan para formar estratos.
- Polietápico. Porque la unidad última de selección (vivienda) es elegida después de varias etapas.
- Por Conglomerados. Porque las unidades de selección son grupos de unidades muestrales.

6. TAMAÑO DE MUESTRA.

Debido a la importancia de cada uno de los temas que conformaran la ENNViH-1 se realizaron y analizaron una gran cantidad de cálculos del tamaño de muestra para cada uno de los parámetros relacionados con los indicadores de interés, entre ellos podemos mencionar algunos como: empleo, migración, educación, ingresos, salud y características de la vivienda.

Entre los indicadores que tienen proporciones más pequeñas están, la tasa de desempleo abierto que tiene una importante variación entre el 1.6% y el 2.0%, el de población ocupada con puestos de funcionarios superiores y personal directivo 2.27%, otros indicadores importantes son el de la renta de la propiedad 2.0% y el indicador de la proporción de migrantes al extranjero 2.23%. Para calcular el tamaño de muestra se tomó como variable de referencia a la proporción de migrantes al extranjero, el cual por ser un indicador de baja frecuencia permite que las demás variables con proporciones mayores queden automáticamente cubiertas.

La expresión empleada para el cálculo fue:

$$n = \frac{z^2 q deff}{r^2 p (1-TNR)PHIV}$$

Donde:

Z = Valor asentado en las tablas estadísticas de la distribución normal para una confianza prefijada.

N = Es el tamaño de muestra en viviendas.

P = Proporción de la variable de referencia.

Q = 1-P

$Deff$ = Es el cociente de dividir la varianza del diseño utilizado entre la varianza del muestreo aleatorio simple, para un mismo tamaño de muestra.

R = Es el error relativo máximo esperado.

TNR = Es la tasa de no respuesta máxima esperada.

PIB = Promedio de habitantes de la categoría de interés por vivienda.

Considerando una confianza del 90%, un error relativo máximo esperado del 13.75%, un efecto de diseño de 6.010, un promedio de habitantes por vivienda de 4.5%, una tasa de no respuesta esperada del 15%; el tamaño de muestra calculado fue de 9,860 el cual se ajustó a 10,000 viviendas.

7. AFIJACIÓN DE LA MUESTRA.

El procedimiento para afijar la muestra en forma proporcional al tamaño es el siguiente:

1. Se afijó la muestra en forma igual entre las cinco regiones.
2. Al interior de cada región se afijó la muestra proporcionalmente entre las zonas.
3. Al interior de cada zona la muestra se afijó proporcionalmente entre los estratos.

$$n_{rgh} = \frac{N_{rh} * n_r}{N_r}$$

Donde:

n_{rgh} = Es el número de viviendas a seleccionar en el h-ésimo estrato de la g-ésima entidad de la r-ésima región.

N_{rh} = Es el total de viviendas del h-ésimo estrato de la g-ésima entidad en la r-ésima región.

N_r = Es el número de viviendas de la r-ésima región.

n_r = Es el número de viviendas a seleccionar en la r-ésima región.

A continuación en el cuadro 1 se presenta la distribución de la muestra en viviendas por dominio y por región; en el cuadro 2 se presenta la distribución de la muestra en listados por entidad y región; en el cuadro 3 se presenta la muestra en viviendas seleccionadas y región.

CUADRO 1

DISTRIBUCIÓN DE LA MUESTRA POR DOMINIO Y REGIÓN

<i>REGIÓN</i>		<i>UPM SELECCIONADOS</i>				<i>VIVIENDAS SELECCIONADAS</i>			
<i>No</i>	<i>NOMBRE</i>	<i>ENEU</i>	<i>CU</i>	<i>RURAL</i>	<i>TOTAL</i>	<i>ENEU</i>	<i>CU</i>	<i>RURAL</i>	<i>TOTAL</i>
1	<i>SUR-SURESTE</i>	13	8	15	36	800	400	800	2,000
2	<i>CENTRO</i>	13	8	15	36	800	400	800	2,000
3	<i>CENTRO-OCCIDENTE</i>	13	8	15	36	800	400	800	2,000
4	<i>NOROESTE</i>	13	8	15	36	800	400	800	2,000
5	<i>NORESTE</i>	13	8	15	36	800	400	800	2,000
<i>NACIONAL</i>		<i>65</i>	<i>40</i>	<i>75</i>	<i>180</i>	<i>4,000</i>	<i>2,000</i>	<i>4,000</i>	<i>10,000</i>

CUADRO 2

DISTRIBUCIÓN DE LA MUESTRA POR DOMINIO Y REGIÓN

REGIÓN		ENTIDAD		UPM SELECCIONADAS			
CLAVE	NOMBRE	CLAVE	NOMBRE	ENEU	CU	RURAL	TOTAL
1	Sur-Sureste	20	Oaxaca	3	2	7	12
		30	Veracruz	6	4	6	16
		31	Yucatán	4	2	2	8
<i>TOTAL</i>				<i>13</i>	<i>8</i>	<i>15</i>	<i>36</i>
2	Centro	09	Distrito Federal	4	0	1	5
		15	México	4	2	7	13
		17	Morelos	2	2	4	8
		21	Puebla	3	4	3	10
<i>TOTAL</i>				<i>13</i>	<i>8</i>	<i>15</i>	<i>36</i>
3	Centro-Occidente	16	Michoacán	6	3	4	13
		14	Jalisco	3	3	7	13
		11	Guanajuato	4	2	4	10
<i>TOTAL</i>				<i>13</i>	<i>8</i>	<i>15</i>	<i>36</i>
4	Noroeste	03	Baja California Sur	3	2	3	8
		25	Sinaloa	5	3	8	16
		26	Sonora	5	3	4	12
<i>TOTAL</i>				<i>13</i>	<i>8</i>	<i>15</i>	<i>36</i>
5	Noreste	05	Coahuila	3	4	4	11
		10	Durango	3	2	5	10
		19	Nuevo León	7	2	6	15
<i>TOTAL</i>				<i>13</i>	<i>8</i>	<i>15</i>	<i>36</i>
<i>TOTAL NACIONAL</i>				<i>65</i>	<i>40</i>	<i>75</i>	<i>180</i>

CUADRO 3

MUESTRA DE VIVIENDAS SELECCIONADAS POR ENTIDAD Y REGIÓN.

REGIÓN		ENTIDAD		VIVIENDAS SELECCIONADAS			
CLAVE	NOMBRE	CLAVE	NOMBRE	ENEU	CU	RURAL	TOTAL
1	Sur-Sureste	20	Oaxaca	189	108	368	665
		30	Veracruz	378	216	324	918
		31	Yucatán	233	76	108	417
<i>TOTAL</i>				800	400	800	2000
2	Centro	09	Distrito Federal	233	0	54	287
		15	México	252	108	378	738
		17	Morelos	126	108	206	440
		21	Puebla	189	184	162	535
<i>TOTAL</i>				800	400	800	2000
3	Centro-Occidente	16	Michoacán	378	130	206	714
		14	Jalisco	170	162	378	710
		11	Guanajuato	252	108	216	576
<i>TOTAL</i>				800	400	800	2000
4	Noroeste	03	Baja California Sur	170	108	152	430
		25	Sinaloa	315	162	432	909
		26	Sonora	315	130	216	661
<i>TOTAL</i>				800	400	800	2000
5	Noreste	05	Coahuila	189	184	216	589
		10	Durango	189	108	270	567
		19	Nuevo León	422	108	314	844
<i>TOTAL</i>				800	400	800	2000
<i>TOTAL NACIONAL</i>				4000	2000	4000	10000

8. SELECCIÓN DE LA MUESTRA.

La selección de la muestra de la ENNViH-1, se realizó en forma independiente para cada región y estrato, el procedimiento de selección, se aplicó de acuerdo a la zona.

8.1 CIUDADES ENEU

En la zona ENEU la selección de la muestra se realizó en forma independiente para cada estrato y región el procedimiento es el siguiente:

1. Dentro de cada región se seleccionaron t_r entidades con igual probabilidad.
2. Dentro de cada entidad y estrato se seleccionaron n_{rh} UPM con probabilidad proporcional al tamaño.
3. En cada UPM² seleccionada se eligieron 6 USM con probabilidad proporcional al tamaño.
4. En cada USM elegida se seleccionaron 'v' viviendas para la ENNViH-1.

La probabilidad de seleccionar la vivienda s esta dado por la siguiente expresión:

$$P\{V_{rehijs}\} = \frac{t_r}{T_r} P_{UPM}^{ENE} \frac{n_{reh}}{N_{reh}} \frac{6 \times V_{rehijs}}{V_{rehi}} \frac{v}{V_{rehijs}^*}$$

Su factor de expansión esta dado por:

$$F_{rehijs} = \frac{T_r}{t_r} F_{UPM}^{ENE} \frac{N_{reh}}{n_{reh}} \frac{V_{rehi}}{6 \times V_{rehijs}} \frac{V_{rehijs}^*}{V}$$

Donde:

² Dentro de las n_{gh} UPM seleccionadas para la zona urbano alto, se seleccionó k USM o áreas de listado con probabilidad proporcional al tamaño.

$P\{V_{rehijs}\}$	=	Es la probabilidad de seleccionar la vivienda s , de la j -ésima USM, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, de la r -ésima región.
t_r	=	Es el número de entidades seleccionadas en la r -ésima región.
T_r	=	Es el total de entidades en la r -ésima región.
P_i^{ENE}	=	Es la probabilidad con la que fue seleccionada la UPM para el marco de propósitos múltiples de la ENE.
n_{reh}	=	Es el número de UPM seleccionadas en el h -ésimo estrato, de la e -ésima entidad, de la r -ésima región para la ENNViH-1.
N_{reh}	=	Es el número total de UPM en el h -ésimo estrato, de la e -ésima entidad, de la r -ésima región.
V_{rehij}	=	Es el total de viviendas en la j -ésima USM, de la i -ésima UPM del h -ésimo estrato, de la e -ésima entidad de la r -ésima región, según conteo de población y viviendas de 1995.
V_{rehi}	=	Es el total de viviendas de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, de la r -ésima región, según conteo de población y vivienda 1995.
V	=	Es el número de viviendas seleccionadas en cada USM.
V_{rehij}^*	=	Es el total de viviendas en la j -ésima USM, de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, de la r -ésima región, según el listado de viviendas.
F_{rehijs}	=	Es el factor de expansión para la vivienda s , de la j -ésima USM de la i -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, de la r -ésima región.
F_{UPM}^{ENE}	=	Es el factor de expansión de la UPM seleccionada para el Marco de Propósitos Múltiples de la ENE.

8.2 COMPLEMENTO URBANO Y RURAL

1. Al interior de cada región se seleccionó t_r entidades con igual probabilidad.
2. Al interior de cada entidad se seleccionó n_{reh} UPM con igual probabilidad empleando muestreo sistemático con arranque aleatorio.
3. Al interior de cada UPM seleccionada se eligió v viviendas con igual probabilidad y arranque aleatorio.

La probabilidad de seleccionar la vivienda está dada por la siguiente expresión:

$$P(V_{rehijs}) = \frac{t_r}{T_r} P_i^{ENE} \frac{n_{reh}}{N_{reh}} \frac{V}{V_{rehij}^*}$$

Su factor de expansión está dado por:

$$F_{rehijs} = \frac{T_r}{t_r} P_i^{ENE} \frac{N_{reh}}{n_{reh}} \frac{V_{rehij}^*}{V}$$

Donde:

$P(V_{rehijs})$ = Es la probabilidad de seleccionar la vivienda s, de la j-ésima USM, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, de la r-ésima región.

t_r = Es el número de entidades seleccionadas en la r-ésima región.

T_r = Es el total de entidades en la r-ésima región.

P_i^{ENE} = Es la probabilidad con la que fue seleccionada la UPM para el marco de propósitos múltiples de la ENE .

n_{reh} = Es el número de UPM seleccionadas en el h-ésimo estrato, de e-ésima entidad, de la r-ésima región para la ENNViH-1.

N_{reh} = Es el número total de UPM en el h-ésimo estrato, de la e-ésima entidad, de la r-ésima región.

V = Es el número de viviendas seleccionadas en cada USM.

V_{rehij}^* = Es el total de viviendas en la j-ésima USM, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, de la r-ésima región, según el listado de viviendas.

F_{rehijs} = Es el factor de expansión para la vivienda s, de la j-ésima USM de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, de la r-ésima región.

1. Entre las n_{gh} UPM que se seleccionó en el marco³, se eligió n_{gh}^* UPM con igual probabilidad.
2. De cada UPM elegida, se seleccionaron 20 viviendas con igual probabilidad utilizando muestreo sistemático.

³ Del total de UPM que integran el marco de propósitos múltiples se seleccionó n_{gh} UPM con probabilidad proporcional a su tamaño para la zona complemento urbano.

Por lo tanto, la probabilidad de seleccionar la vivienda s en la i -ésima UPM en el h -ésimo estrato de la g -ésima entidad en la r -ésima región es:

$$P\{V_{rghis}\} = \frac{n_{rgh} m_{rghi}}{m_{rgh}} \frac{n_{rgh}^*}{n_{rgh}} \frac{20}{m_{rgh}^*} = \frac{20 n_{rgh}^* m_{rghi}}{m_{rgh} m_{rgh}^*}$$

Su factor de expansión está dado por:

$$F_{rghis} = \frac{m_{rgh} m_{rghi}^*}{20 n_{rgh}^* m_{rghi}}$$

Donde:

- n_{rgh} = Número de UPM seleccionadas en el marco en el h -ésimo estrato de g -ésima entidad, en la r -ésima región.
- m_{rghi} = Número de viviendas en la i -ésima UPM en el h -ésimo estrato de la g -ésima entidad, en la r -ésima región.
- m_{rgh} = Número de viviendas del h -ésimo estrato para la g -ésima entidad, en la r -ésima región.
- n_{rgh}^* = Número de UPM seleccionadas para la ENNViH-1 en el h -ésimo estrato para la g -ésima entidad, en la r -ésima región.
- m_{rghi}^* = Número total de viviendas en la i -ésima UPM en el h -ésimo estrato para la g -ésima entidad en la r -ésima región. Al momento del levantamiento de la ENNViH-1.

8.3 Rural.

1. Entre las n_{rgh} UPM que se seleccionó en el marco ⁴, se tomaron en muestra n_{rgh}^* UPM con igual probabilidad.

⁴ Del total de UPM que integran el marco de propósitos múltiples, se seleccionó n_{gh} UPM con la probabilidad proporcional a su tamaño para la zona rural.

2. En cada UPM seleccionada, se eligieron 2 segmentos de 10 viviendas aproximadamente, con igual probabilidad.

Por lo tanto, la probabilidad de seleccionar la vivienda s de la i -ésima UPM en el estrato h de la g -ésima entidad en la r -ésima región es:

$$P \left\{ V_{rghis} \right\} = \frac{n_{rgh} m_{rghi}}{m_{rgh}} \frac{n_{rgh}^*}{n_{rgh}} \frac{2 * 10}{m_{rghi}^*} = \frac{20 n_{rgh}^* m_{rghi}}{m_{rgh} m_{rghi}^*}$$

En consecuencia su factor de expansión está dado por:

$$F_{rghi} = \frac{m_{rgh} m_{rghi}^*}{20 n_{rgh}^* m_{rghi}}$$

Donde:

n_{rgh} = Número de UPM seleccionadas del h -ésimo estrato de la g -ésima entidad, en la r -ésima región.

m_{rghi} = Número de viviendas de la i -ésima UPM en el h -ésimo estrato para la g -ésima entidad en la r -ésima región.

m_{rgh} = Número total de viviendas en el h -ésimo estrato para la g -ésima entidad, en la r -ésima región.

n_{rgh}^* = Número de UPM seleccionadas para la ENNViH-1 en el h -ésimo estrato de la g -ésima entidad, en la r -ésima región.

m_{rghi}^* = Número total de viviendas en la i -ésima UPM del h -ésimo estrato para la g -ésima entidad, en la r -ésima región al momento del levantamiento de la ENNViH-1.

9. ESTIMADORES.

El estimador del total de la característica X es:

A nivel nacional:

$$\hat{X}_{NAL} = \sum_{r=1}^R \sum_{g=1}^G \sum_{h=1}^H \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K F_{rghijk}^u \bar{x}_{rghijk}^u + \sum_{r=1}^R \sum_{g=1}^G \sum_{h=1}^H \sum_{i=1}^I \sum_{k=1}^K F_{rghik}^R \bar{x}_{rghik}^R$$

Donde:

F_{rghijk}^u = Factor de expansión de la s-ésima vivienda en la j-ésima USM de la i-ésima UPM en el h-ésimo estrato para la g-ésima entidad, en la r-ésima región de la zona Urbano Alto (Ciudades ENEU).

X_{rghijk}^u = Valor observado de la característica de interés x en la I-ésima entrevista para la s-ésima vivienda en la j-ésima USM de la i-ésima UPM en el h-ésimo estrato para la g-ésima entidad, de la r-ésima región de la zona Urbano Alto (Ciudades ENEU).

F_{rghik}^R = Factor de expansión de la s-ésima vivienda en la i-ésima UPM de el h-ésimo estrato para la g-ésima entidad en la r-ésima región del resto de las zonas.

X_{rghik}^R = Valor observado de la característica X en la i-ésima entrevista de la s-ésima vivienda en la i-ésima UPM del h-ésimo estrato para la g-ésima entidad, en la r-ésima región del resto de las zonas.

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} .

10. ESTIMACIÓN DE LAS PRECISIONES.

La estimación de los errores de muestreo para las estimaciones nacionales obtenidas de la ENNViH-1 se realizó mediante la aplicación del método de Conglomerados Últimos⁵, conjuntamente con el de Series de Taylor.

La ENNViH-1 emplea estimadores de promedios por hogar, que son considerados como estimadores de razón, donde la estimación de hogares a nivel nacional (en el denominador) se comporta como una variable aleatoria.

La fórmula empleada para calcular la precisión de \hat{R} es la siguiente:

$$\hat{V}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_h^L \frac{n_h}{n_h - 1} \sum_i^{n_h} \left[\left(\hat{X}_{hi} - \hat{X}_h \right) - \hat{R} \left(\hat{Y}_{hi} - \hat{Y}_h \right) \right]^2$$

Donde:

n_h = Número de UPM en el h-ésimo estrato.

\hat{X}_{hi} = Total ponderado de la variable de estudio X para la i-ésima UPM en el h-ésimo estrato.

\hat{X}_h = Promedio por UPM del total ponderado de la variable de estudio X para el h-ésimo estrato.

Estas definiciones son análogas para la variable de estudio Y.

La estimación de la precisión del estimador del total, se calcula con la fórmula siguiente:

⁵ El término conglomerado último se utiliza para denotar al conjunto de unidades seleccionadas en las Unidades Primarias de Muestreo, sin tomar en cuenta las etapas de selección y el método calcula la varianza entre estas unidades, debido a que representa la contribución más importante a la varianza total.

$$\hat{V}(\hat{X}) = \sum_{h=1}^{lg} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{1}{n_h} \hat{X}_h \right)^2$$

Las estimaciones de la desviación estándar (D.E.), coeficiente de variación (C.V.) y deff se calculan mediante las siguientes expresiones:

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

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

$$deff = \frac{\sqrt{\hat{V}(\hat{R})}}{\sqrt{\hat{V}(\hat{R})_{mas}}}$$

Donde:

$\hat{V}(\hat{R})_{mas}$ = Estimador de la varianza de \hat{R} bajo un muestreo aleatorio simple.

\hat{R} = Estimador del parámetro poblacional R.