

DISEÑO MUESTRAL

**MÓDULO DE HOGARES Y MEDIO AMBIENTE 2017
(MOHOMA-2017)**

Índice	Página
1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Marco de la encuesta	1
5. Tamaño de la muestra	1
6. Selección de la muestra	1
7. Ajuste a los factores de expansión	2
7.1 Ajuste por no respuesta	2
7.2 Ajuste por proyección	2
8. Estimadores	3
9. Estimación de errores de muestreo	3
Anexo	
Distribución de la muestra en UPM y viviendas por entidad para el Módulo de Hogares y Medio Ambiente MOHOMA-2017.	5

1. Objetivo de la encuesta

El Módulo de Hogares y Medio Ambiente 2017 (MOHOMA-2017) tiene por objetivo obtener información estadística de los hogares sobre temas como el cuidado del agua, ahorro de energía eléctrica, manejo de residuos y basura, transporte y movilidad, así como percepción del medio ambiente.

2. Población objetivo

La encuesta está dirigida a los hogares de las viviendas particulares del territorio nacional.

3. Cobertura geográfica

La encuesta está diseñada para dar resultados a nivel nacional.

4. Marco de la encuesta

El diseño muestral del MOHOMA-2017 es el empleado en la Encuesta Nacional de Hogares (ENH) 2017 ya que el MOHOMA se levanta en las mismas viviendas correspondiente al levantamiento del segundo trimestre de la ENH 2017¹. Este módulo se caracteriza por ser probabilístico, por tanto, los resultados obtenidos de la encuesta se generalizan a toda la población; a su vez el diseño muestral es bietápico, estratificado y por conglomerados, donde la unidad última de selección es la vivienda y la unidad de observación es el hogar.

5. Tamaño de la muestra

El tamaño de la muestra fue de 15 880 viviendas, el cual se determinó con base en las viviendas con respuesta en la ENH-2017 del segundo trimestre, de 2017.

6. Selección de la muestra

Para conformar la muestra del MOHOMA-2017 se consideraron las viviendas de la ENH-2017 correspondientes al segundo trimestre, por lo tanto, la probabilidad de seleccionar a la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad se define de la siguiente manera:

$$P_{ehij} = P_{ehij}^{ENH}$$

Su factor de expansión es:

$$F_{ehij} = F_{ehij}^{ENH}$$

Donde:

P_{ehij} = probabilidad de seleccionar de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, para el MOHOMA- 2017.

P_{ehij}^{ENH} = probabilidad de seleccionar de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, para la ENH-2017.

¹ Ver el documento del Diseño Muestral de la Encuesta Nacional de Hogares (ENH-2017).

F_{ehij} = factor de expansión de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, para el MOHOMA-2017.

F_{ehij}^{ENH} = factor de expansión de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad para la ENH-2017.

En el anexo, se presenta la distribución de la muestra de UPM y viviendas por entidad.

7. Ajuste a los factores de expansión

Los factores de expansión elaborados conforme al procedimiento antes descrito se ajustaron para los siguientes conceptos:

7.1 Ajuste por no respuesta

El ajuste por no respuesta atribuida al informante se realizó a nivel estrato, mediante la siguiente expresión:

$$F'_{ehi} = F_{ehi} \frac{\sum_{i \in h} F_{ehi} V_{ehi}}{\sum_{i \in h} F_{ehi} V_{ehi}^*}$$

Donde:

F'_{ehi} = factor de expansión corregido por no respuesta 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.

V_{ehi} = número de viviendas seleccionadas en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

V_{ehi}^* = número de viviendas seleccionadas con respuesta en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad.

7.2 Ajuste por proyección

Los factores de expansión ajustados por la no respuesta se corrigieron, 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 CONAPO 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 proyección.

$PEXP_D$ = población total a la que expande la encuesta en el dominio D.

D = dominio de interés para la encuesta.

8. Estimadores

El estimador del total de la característica X es:

$$\hat{X} = \sum_e \sum_h \sum_i F_{ehi}^{UA} \left(\sum_j \sum_k X_{ehijk}^{UA} \right) + \sum_e \sum_h \sum_i F_{ehi}^{CU} \left(\sum_j \sum_k X_{ehijk}^{CU} \right) + \sum_e \sum_h \sum_i F_{ehi}^R \left(\sum_j \sum_k X_{ehijk}^R \right)$$

Donde:

- F_{ehi}^{UA} = factor de expansión final para el MOHOMA-2017 de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad en el dominio urbano alto.
- X_{ehijk}^{UA} = valor observado de la característica de interés X en la k-ésima persona, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el dominio urbano alto.
- 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_{ehijk}^{CU} = valor observado de la característica de interés X en la k-ésima persona, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el dominio complemento urbano.
- F_{ehi}^R = factor de expansión final de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad del dominio rural.
- X_{ehijk}^R = valor observado de la característica de interés X en la k-ésima persona, en la j-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad del dominio rural.

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

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

Donde la variable \hat{Y} se define en forma análoga a \hat{X} .

9. Estimación de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones nacionales se usó el método de "Conglomerados Últimos",² basado en que la mayor contribución a la varianza de un estimador, en un diseño bietápico, es la que se presenta entre las unidades primarias de muestreo (UPM). El término "Conglomerados Últimos" 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 aplicó el método de Conglomerados Últimos conjuntamente con el método de series de Taylor, obteniéndose la siguiente fórmula para estimar la precisión de \hat{R} .

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

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.

² Vease Hansen, M.H. Horwitz, W.N. y Madow, W.G, *Sample Survey Methods and Theory*, (1953), Vol. 1 página 242.

- \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.
- L_e = número de estratos en la e-ésima entidad.
- \hat{Y}^2 = el cuadrado del total estimado de la variable de estudio Y.

Estas definiciones son análogas para la variable de estudio Y.

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

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

Las estimaciones del error estándar (EE), coeficiente de variación (CV) o error relativo del estimador y el efecto de diseño (DEFF) se calculan mediante las siguientes expresiones:

$$EE = \sqrt{\hat{V}(\hat{\theta})} \qquad CV = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}} \qquad DEFF = \frac{\hat{V}(\hat{\theta})}{\hat{V}(\hat{\theta})_{mas}}$$

Donde:

- $\hat{\theta}$ = estimador del parámetro poblacional θ .
- $\hat{V}(\hat{\theta})_{MAS}$ = estimador de la varianza, bajo muestreo aleatorio simple.
- $\hat{V}(\hat{\theta})$ = estimador de la varianza bajo el diseño de muestreo descrito en este documento.

Finalmente, el intervalo de confianza $I_{1-\alpha}$ al $100(1-\alpha)\%$, se construye de la siguiente forma:

$$I_{1-\alpha} = \left(\hat{\theta} - z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\theta})} \right)$$

Donde α es el nivel de significancia.

**Distribución de la muestra en UPM y viviendas por entidad
para el Módulo de Hogares y Medio Ambiente MOHOMA-2017**

Entidad		Muestra	
Clave	Nombre	UPM	Viviendas
01	Aguascalientes	80	564
02	Baja California	83	505
03	Baja California Sur	52	515
04	Campeche	62	480
05	Coahuila de Zaragoza	78	495
06	Colima	72	509
07	Chiapas	42	481
08	Chihuahua	78	481
09	Ciudad de México	101	505
10	Durango	67	515
11	Guanajuato	58	504
12	Guerrero	46	471
13	Hidalgo	40	488
14	Jalisco	66	496
15	México	81	480
16	Michoacán de Ocampo	43	486
17	Morelos	66	496
18	Nayarit	49	471
19	Nuevo León	87	525
20	Oaxaca	37	485
21	Puebla	53	465
22	Querétaro	64	503
23	Quintana Roo	80	505
24	San Luis Potosí	58	487
25	Sinaloa	64	500
26	Sonora	70	487
27	Tabasco	38	477
28	Tamaulipas	76	470
29	Tlaxcala	69	525
30	Veracruz de Ignacio de la Llave	49	469
31	Yucatán	65	521
32	Zacatecas	46	519
Total		2 020	15 880