

Encuesta Nacional de Ocupación y Empleo

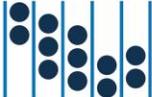
Nueva edición

ENOE ^N

Diseño muestral



Segunda edición

 **INEGI**

Instituto Nacional de Estadística y Geografía

Encuesta Nacional de Ocupación y Empleo

Nueva edición

ENOE^N

Diseño muestral

Segunda edición

Noviembre de 2021



Índice	Página
1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Diseño muestral	1
4.1 Submuestra	1
5. Tamaño de la submuestra	1
6. Ajuste a los factores de expansión	1
6.1 Colapsamiento de estratos	2
6.2 Ajuste por no respuesta	2
6.3 Ajuste por proyecciones de población	3
7. Estimadores	3
8. Estimación de errores de muestreo	4
9. Homologación de la Semaforización para los Umbrales de Indicadores de precisión estadística	5
Anexo	
- Indicadores empleados en la estratificación de los conglomerados de la muestra maestra (Cuadro 1).	7

1. Objetivo de la encuesta

Obtener información estadística sobre las características ocupacionales de la población, así como de variables demográficas y económicas para el análisis de la fuerza de trabajo, la toma de decisiones, el diseño y seguimiento de las políticas laborales.

2. Población objetivo

La población objetivo de la encuesta son todos los integrantes de la vivienda, así como las personas de 12 y más años de edad.

3. Cobertura geográfica

El tamaño de muestra de la ENOEN^N permite dar resultados de los principales indicadores estratégicos a nivel nacional, por entidad federativa y ciudad auto representada.

4. Diseño muestral

El diseño muestral de la ENOEN^N se caracteriza por ser probabilístico, por lo cual los resultados obtenidos de la encuesta se generalizan a toda la población, a su vez es bietápico, estratificado y por conglomerados, donde la unidad última de observación es la persona.

4.1 Submuestra

Para el mes de julio de 2020, la ENOEN^N toma como referencia 16 948 viviendas de la ENOE, cuyas entrevistas se realizaron cara a cara, adicionalmente 6 607 viviendas de la misma ENOE del primer trimestre de 2020, la cual está conformada por todas las viviendas de las cuales se dispone del número de contacto telefónico

En julio de 2021, se deja de usar la submuestra del primer trimestre de 2020, y la muestra de la ENOEN^N se conforman de 41,719 viviendas cuyas entrevistas se realizaron cara a cara y se adicionan 6,197 nuevas viviendas independientes a las cuales se les realiza una entrevista telefónica.

5. Tamaño de la submuestra

El tamaño de muestra de la ENOEN^N en julio de 2020 fue de 16 948 viviendas cara a cara y 6 607 viviendas con número telefónico, el cual da un tamaño de muestra total de 23 555 viviendas, este tamaño de muestra alcanza para cubrir valores de la tasa de desocupación de 2.5% y mayores, bajo una confianza del 90%, con un error relativo máximo esperado de 7.8%, una tasa neta de participación de 60%, un promedio de habitantes de 15 y más años de edad por vivienda de 2.47, así como un valor de 2.00 para el efecto de diseño.

6. Ajuste de los factores de expansión

Debido a que la muestra de hogares de la ENOEN^N es una submuestra de la ENOE, conservan sus factores de expansión, los cuales se ajustan debido al colapsamiento de estratos, a la no respuesta y a la estimación de población del INEGI, el procedimiento de ajuste se describe a continuación:

6.1 Colapsamiento de estratos

El colapso de estratos genera nuevos estratos a los cuales se les conoce como pseudo-estratos y son construidos con el objetivo de corregir la subestimación de la población de interés y demás variables debido a la existencia de estratos sin muestra o que cuenta con una sola UPM.

Los pseudo-estratos son construidos a través de la agrupación de dos o más estratos originales, de acuerdo con un patrón u ordenamiento jerárquico de variables geográficas o de tamaño, de modo que estos contengan al menos dos UPM, los que a su vez deberán contener suficientes viviendas en su interior para mantener los factores de expansión sin valores atípicos.

A continuación, se detalla el procedimiento de construcción de los pseudo-estratos.

- Primero se contabiliza, al interior de cada estrato original, el total de unidades que participa en la encuesta. Si el estrato no contiene unidades, entonces deberá ser colapsado con otro.
- Se ordenan todos los estratos, geográficamente, de acuerdo a los dominios de la encuesta en áreas autorrepresentadas, en complemento urbano y rural, al interior de cada entidad federativa.
- Finalmente, al interior de la misma área geográfica y entidad se colapsan aquellos estratos sin unidades, lo más cercano posible geográficamente.

El procedimiento de ajuste es el siguiente:

$$F_{eh'i}^* = F_{ehi}'' \frac{\sum_{eh} F_{ehi}'' V_{eh}}{\sum_{eh'} F_{ehi}'' V_{eh'}}$$

Donde:

- $F_{eh'i}^*$ = factor de expansión corregido por el colapsamiento de estratos de la i-ésima UPM, del h'-ésimo pseudo-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 de la ENOE.
- V_{eh} = número de viviendas seleccionadas en el h-ésimo estrato, en la e-ésima entidad.
- $V_{eh'}$ = número de viviendas seleccionadas en el h'-ésimo pseudo-estrato, en la e-ésima entidad.
- h' = representa los nuevos pseudo-estratos para la ENOE^N en cada entidad.

6.2 Ajuste por no respuesta

El ajuste por no respuesta atribuida al informante se realiza a nivel de estrato, en cada uno de los ámbitos, mediante la siguiente expresión:

$$F_{eh'i}^{**} = F_{ehi}'' \frac{\sum_{eh'} F_{eh'i}'' V_{eh'}}{\sum_{eh'} F_{eh'}'' V_{eh'}^*}$$

Donde:

- $F_{eh'i}^{**}$ = factor de expansión corregido por no respuesta de la i-ésima UPM, en el h'-ésimo pseudo-estrato, de la e-ésima entidad.
- F_{ehi}'' = factor de expansión de la i-ésima UPM, en el h-ésimo pseudo-estrato, de la e-ésima entidad.
- $V_{eh'}$ = número de viviendas seleccionadas en el h'-ésimo pseudo-estrato, de la e-ésima entidad.
- $V_{eh'}^*$ = número de viviendas con respuesta en el h'-ésimo pseudo-estrato, de la e-ésima entidad.

6.3 Ajuste por proyecciones de población

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 estimación de población generada por el INEGI referida al punto medio del levantamiento, mediante la siguiente expresión:

$$F_D'' = F_D' \frac{PRO_y}{PEXP_D}$$

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.
- PRO_y = población en el dominio D, según estimación de población INEGI.
- $PEXP_D$ = población total a la que expande la encuesta en el dominio D.
- D = nivel de desagregación en que se efectúa el ajuste.

7. Estimadores

El estimador del total de la característica X es:

$$\hat{X} = \sum_{e=1}^{32} \sum_{h=1}^{L_e} \sum_{i=1}^{k_{eh}^{UA}} F_{ehij}^{UA} \left(\sum_s \sum_{\ell} X_{ehis\ell}^{UA} \right) + \sum_{e=1}^{32} \sum_{h=1}^{L_e} \sum_{i=1}^{k_{eh}^{CU}} F_{ehij}^{CU} \left(\sum_s \sum_{\ell} X_{ehis\ell}^{CU} \right) + \sum_{e=1}^{32} \sum_{h=1}^{L_e} \sum_{i=1}^{k_{eh}^R} F_{ehij}^R \left(\sum_s \sum_{\ell} X_{ehis\ell}^R \right)$$

Donde:

- F_{ehij}^{UA} = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad del ámbito urbano alto.
- $X_{ehis\ell}^{UA}$ = valor observado de la característica de interés X en la ℓ -ésima persona, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el ámbito urbano alto.
- F_{ehij}^{CU} = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad, del ámbito complemento urbano.
- $X_{ehis\ell}^{CU}$ = valor observado de la característica de interés X en la ℓ -ésima persona, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, en el ámbito complemento urbano.
- F_{ehij}^R = factor de expansión final de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la e-ésima entidad del ámbito rural.
- $X_{ehis\ell}^R$ = valor observado de la característica de interés en la ℓ -ésima persona, en la s-ésima vivienda, en la i-ésima UPM, en el h-ésimo estrato, en la e-ésima entidad, del ámbito rural.

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

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

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

8. Estimación de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones estatales y nacionales se usó el método de Conglomerados Últimos¹, 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 la muestra total de unidades primarias de muestreo.

Para obtener las precisiones de los estimadores de razón, conjuntamente al método de Conglomerados Últimos se aplicó el método de *Series de Taylor*, obteniéndose la siguiente fórmula para estimar la precisión de \hat{R} a nivel nacional:

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

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.

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

k_{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 ponderado de la característica Y.

Estas definiciones de X son análogas para la variable de estudio Y.

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

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

Las estimaciones de la desviación estándar (DE), efecto de diseño (DEFF) y coeficiente de variación (CV) se calculan mediante las siguientes expresiones:

$$DE = \sqrt{\hat{V}(\hat{\theta})}$$

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

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

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.

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

Finalmente, el intervalo de confianza $I_{1-\alpha}$ al $100(1-\alpha)\%$, se calcula 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.

9. Homologación de la Semaforización para los Umbrales de Indicadores de precisión estadística²

Para facilitar la interpretación de las precisiones estadísticas de la información pública en tabulados, el Comité de Aseguramiento de la Calidad, en la cuarta sesión celebrada el 1 de noviembre de 2018, aprobaron los siguientes umbrales y especificaciones para la publicación en los tabulados los CV, así como su semaforización de estos.

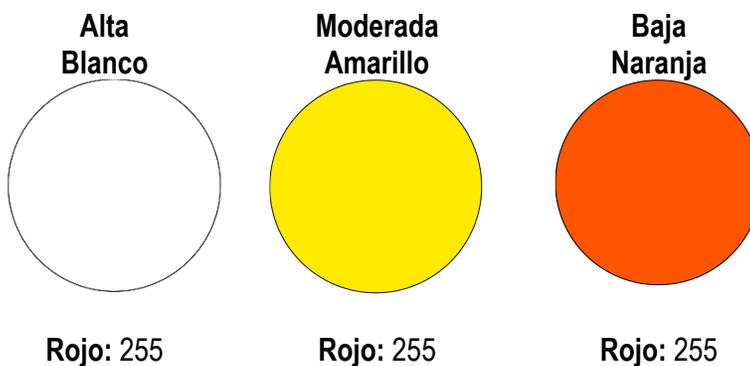
Umbrales aprobados para la cobertura del CV

Interpretación	Semaforización	Viviendas/Hogares/Otras unidades diferentes a las económicas DGES/DGEGSPJ
Alta	Blanco	[0%, 15%)
Moderada	Amarillo	[15%, 30%)
Baja	Naranja oscuro	>=30%

Umbrales aprobados para el reporte de la precisión de acuerdo con el coeficiente de variación en los tabulados de resultados de los proyectos con muestreo probabilístico (acuerdo CAC-007/01/2018).

A partir del segundo trimestre de 2018, se publican los siguientes indicadores de precisión estadística en la presentación de resultados en tabulados de todas las encuestas con muestreo probabilístico del INEGI: error estándar, intervalo de confianza y coeficiente de variación (CV). Adicionalmente, se estandariza la coloración en los tabulados para indicar el nivel de precisión de las estimaciones con base en el CV. A continuación, se presenta el código RGB de los colores utilizados en la semaforización:

Parámetros RGB para la semaforización del coeficiente de variación.



² La fuente de esta información está basada en el documento del comité de Aseguramiento de la calidad depositado en el siguiente sitio http://intranet.inegi.org.mx/calidad/wp-content/uploads/2017/02/Homologacion_de_umbrales.pdf

Verde: 255
Azul: 255

Verde: 234
Azul: 0

Verde: 84
Azul: 0

El siguiente texto explicativo aparece en cada uno de los tabulados publicados de encuestas por muestreo probabilístico.

Las estimaciones que aparecen en este cuadro están coloreadas de acuerdo con su nivel de precisión, en *Alta*, *Moderada* y *Baja*, tomando como referencia el coeficiente de variación CV (%). Una precisión *Baja* requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.

Nivel de precisión de las estimaciones:

Alta, CV en el rango de (0,15)

Moderada, CV en el rango de [15, 30)

Baja, CV de 30% en adelante

Mnemónico	Descripción
Proporción de Población	
PPSSNOSP	Que tiene derecho a recibir servicios médicos en alguna institución de salud pública o privada excepto seguro popular.
PPDER_SS	Derechohabiente a servicios de salud.
PDP3A14A	De 3 a 14 años de edad que asiste a la escuela.
PDP15A24A	De 15 a 24 años de edad que asiste a la escuela.
PDP8A14ALF	De 8 a 14 años de edad que saben leer o escribir.
PDP15YM_SE	De 15 años o más de edad que aprobaron algún grado de escolaridad diferente al nivel preescolar.
PP15PRI_CO	De 15 años o más de edad que tienen como máxima escolaridad 6 grados aprobados en primaria.
PP15SEC_CO	De 15 años o más de edad que tienen como máxima escolaridad 3 grados aprobados en secundaria.
PGDO_ESC	Grado promedio de escolaridad.
PPEA	De 12 años y más que trabajaron; tenían trabajo pero no trabajaron o; buscaron trabajo en la semana de referencia.
PPEA_F	Femenina de 12 años y más que trabajaron; tenían trabajo pero no trabajaron o; buscaron trabajo en la semana de referencia.
PTASAOCUPA	Tasa de ocupación.
TOCU12A17	No ocupada de 12 a 17 años de edad entre la población de 12 a 17 años de edad.
PPOMAYED	Ocupada de 18 y más años de edad entre la población ocupada.
Proporción de Viviendas Particulares Habitadas	
PVIVSINH	Que no tienen hacinamiento.
PVPH_PISDT	Que tienen piso de cemento o firme, madera, mosaico u otro material.
PVPH2YMASD	Que usan para dormir entre 2 y 25 cuartos.
PVPH_2MASC	Que tienen más de un cuarto.
PVPH3YMASD	Que tienen entre 3 y 25 cuartos.
PVPH_C_ELE	Que disponen de luz eléctrica.
PVPHAGUADV	Que tienen disponibilidad de agua entubada dentro de la vivienda, o fuera de la vivienda pero dentro del terreno.
PVPH_EXCSA	Que tienen excusado, retrete, sanitario, letrina u hoyo negro.
PVPHDRENAJ	Que tienen drenaje conectado a la red pública, fosa séptica, barranca, grieta, río, lago o mar.
PVDREDED	Que disponen de drenaje conectado a la red pública.
PVEXCAGU	Que disponen de excusado con descarga directa de agua.
PVPH_CSERV	Que disponen de luz eléctrica, agua entubada dentro o fuera de la vivienda, pero dentro del terreno, así como drenaje.
PSIN_HASIN	Que no se encuentran en situación de hacinamiento a nivel manzana.
Proporción de Viviendas Particulares Habitadas que disponen de:	
PVPH_TV	Televisor.
PVPH_AUTOM	Automóvil o camioneta.
PVPH_CEL	Teléfono celular.
PVCELFIJ	Teléfono celular y teléfono fijo.
PV4ELEC	Radio, televisor, refrigerador y lavadora.
PVRADTEL	Radio y televisor.
PVPHCBIEN	Todos los bienes.