

ENAPE

Encuesta Nacional
sobre Acceso y Permanencia
en la Educación

2021

Diseño muestral



 **INEGI**

Instituto Nacional de Estadística y Geografía

Encuesta Nacional sobre Acceso y Permanencia en la Educación 2021

ENAPE

Diseño muestral



Catalogación en la fuente INEGI:

371.227 Encuesta Nacional sobre Acceso y Permanencia en la Educación.
Encuesta Nacional sobre Acceso y Permanencia en la Educación 2021 : ENAPE
: diseño muestral / Instituto Nacional de Estadística y Geografía.-- México : INEGI,
c2022.

8 p.

1. Acceso a la educación - Encuestas - Metodología. I. Instituto Nacional de
Estadística y Geografía (México).

Conociendo México

800 111 4634

www.inegi.org.mx

atencion.usuarios@inegi.org.mx



INEGI Informa



@INEGI_INFORMA

Registro en trámite

2022, **Instituto Nacional de Estadística y Geografía**

Edificio Sede

Avenida Héroe de Nacozari Sur 2301

Fraccionamiento Jardines del Parque, 20276 Aguascalientes,

Aguascalientes, Aguascalientes, entre la calle INEGI,

Avenida del Lago y Avenida Paseo de las Garzas.

Presentación

El **Instituto Nacional de Estadística y Geografía (INEGI)** presenta el documento muestral en el que se exponen las principales características de la muestra para la **Encuesta Nacional Acceso y Permanencia en la Educación (ENAPE)**, 2021.

En este documento, correspondiente al Diseño muestral, se describen las actividades que constituyen el proceso de obtención y tratamiento estadístico de la muestra, que inicia con la determinación de la cobertura y dominios de interés, para los cuales se pueden obtener estimaciones estadísticamente confiables, en función de los cuales se calcula el tamaño de la muestra y se determina su distribución, seguida de la selección de las unidades de muestreo.

Como parte del tratamiento, se describe el cálculo de las probabilidades de selección y su empleo para asignar el peso que debe tener cada unidad seleccionada.

Finalmente se presentan las expresiones empleadas para obtener las estimaciones y sus medidas de calidad.

Índice

1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Marco de la encuesta	1
4.1 Estratificación	1
5. Tamaño de la muestra	1
6. Distribución de la muestra	2
7. Selección de la muestra	2
8. Ajuste a los factores de expansión	3
8.1 Ajuste por no respuesta	3
8.2 Ajuste por la cobertura de población con teléfono	3
8.3 Ajuste por estimación de población	4
8.4 Ajuste por calibración	4
9. Estimadores	5
10. Estimación de errores de muestreo	6
11. Homologación de la Semaforización para los Umrales de Indicadores de precisión estadística	7

Diseño muestral

1. Objetivo de la encuesta

Generar información estadística sobre el acceso y permanencia de la población de 0 a 29 años en el Sistema Educativo Nacional (SEN), a través de la inscripción en el ciclo escolar pasado 2020-2021 y ciclo escolar vigente 2021-2022; los motivos o razones de no conclusión y no inscripción; la consideración de retomar sus estudios para quienes no están inscritos en el ciclo escolar actual; además de dar información sobre algunos aspectos de la dinámica escolar, entre otras variables de las condiciones educativas de esta población.

2. Población objetivo

La población objeto de estudio fueron las personas de 3 a 29 años, que contaban con un teléfono y que residían dentro del territorio nacional en la fecha del levantamiento.

3. Cobertura geográfica

La encuesta está diseñada para dar resultados a nivel nacional.

4. Marco de la encuesta

El marco de muestreo de la ENAPE 2021 se conformó a partir del Plan Nacional de Numeración, publicado por el Instituto Federal de Telecomunicaciones (IFT) y actualizado a noviembre de 2021 y de los cuales se generaron de manera aleatoria una muestra de números telefónicos en cada una de las 32 entidades federativas. El diseño estadístico es probabilístico por lo que los resultados se pueden generalizar a nivel nacional, unietápico porque la selección de los números se realizó en una etapa, se dice que es estratificado debido a que cada entidad federativa es un estrato de diseño.

4.1 Estratificación

El diseño de la muestra es estratificado debido a que cada una de las 32 entidades federativas conforman los estratos de diseño, al interior de cada entidad se ha seleccionado una muestra aleatoria de números telefónicos en forma independiente.

5. Tamaño de la muestra

El tamaño de la muestra se calculó para una proporción mínima de 1 %, la cual requiere el tamaño de muestra mayor, lo que garantiza que las estimaciones de proporciones mayores al 1 % queden cubiertas con este tamaño. La expresión empleada para el cálculo es la siguiente:

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

Donde:

- n = tamaño de la muestra.
- p = estimación de la proporción de interés.
- q = 1-p.
- r = valor asentado en las tablas estadísticas de la distribución normal estándar que garantiza obtener las estimaciones con una confianza prefijada.
- z = tasa de no respuesta máxima esperada.
- tnr = tasa de no respuesta máxima esperada.

Considerando una confianza de 90%, un error relativo esperado de 11.1 %, una proporción de 1 %, y la tasa de no respuesta esperada de 85 %, se determinó una muestra a nivel nacional de aproximadamente 144 928 números telefónicos, el cual se ajustó a 145 015.

6. Distribución de la muestra

La muestra se distribuyó en las 32 entidades federativas, en promedio se seleccionaron 4 532 números telefónicos por entidad.

En la siguiente tabla se presenta la distribución de la muestra a nivel nacional.

Resultado y código de la entrevista	ENAPE 2021	
	Total	Relativo
Muestra de números telefónicos	145 015	100.0
Viviendas con contacto	38 774	26.7
Viviendas con entrevista lograda (00)	22 719	58.6
Viviendas con población objetivo	14 456	63.6
(Información de personas de 0 a 29 años)	33 922	
Sin población objetivo	8 263	36.4
Entrevista No lograda	16 052	41.4
Sin Contacto (No respondió la llamada, no existe el número, otro)	106 244	73.3

7. Selección de la muestra

La selección de la muestra de números telefónicos para la ENAPE 2021, se realizó de manera independiente al interior de cada entidad federativa. El procedimiento de selección de números telefónicos se realizó de manera aleatoria.

La probabilidad de seleccionar el i -ésimo número telefónico en la e -ésima entidad, está dada por la siguiente expresión:

$$p_{ei} = \frac{k_e}{K_e}$$

Su factor de expansión¹ está dado por:

$$F_{ei} = \frac{K_e}{k_e}$$

¹ El Factor de Expansión se define como el inverso de la probabilidad de selección. En la Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía, el Factor de Expansión se nombra Ponderador.

Donde:

- k_e = números telefónicos por seleccionar en la e-ésima entidad.
- K_e = total de números telefónicos activos que conforman el marco muestral en la e-ésima entidad.
- F_{ei} = factor de expansión del i-ésimo número telefónico de la e-ésima entidad.

8. Ajuste a los factores de expansión

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

8.1 Ajuste por no respuesta

El ajuste por no respuesta atribuida al informante se efectuó a nivel de entidad, mediante la siguiente expresión:

$$F_{ei}^* = F_{ei} \frac{\sum_e F_{ei} Q_e}{\sum_e F_{ei} Q_e^*}$$

Donde:

- F_{ei}^* = factor de expansión corregido por no respuesta del i-ésimo número telefónico, de la e-ésima entidad.
- F_{ei} = factor de expansión del i-ésimo número telefónico, de la e-ésima entidad.
- Q_e = total de números telefónicos seleccionados en la e-ésima entidad.
- Q_e^* = total de números telefónicos seleccionados con respuesta en la e-ésima entidad.

8.2 Ajuste por la cobertura de población con teléfono

Los factores de expansión ajustados por no respuesta se corrigen por la cobertura de población que cuenta con teléfono por entidad federativa. Con información obtenida de ENDUTIH 2020 la cobertura a nivel nacional es aproximadamente de 95 % para las viviendas con teléfono, el procedimiento es el siguiente:

$$F_{ei}^{**} = F_{ei}^* \frac{\hat{Z}_e}{\hat{Y}_e}$$

Donde:

- \hat{Z}_e = estimación de la población que reside en viviendas con teléfono obtenida de la ENDUTIH 2020 en la e-ésima entidad.
- \hat{Y}_e = estimación de la población que reside en las viviendas con teléfono obtenida de la ENAPE 2021 en la e-ésima entidad.
- F_{ei}^* = factor de expansión ajustado por no respuesta, de la i-ésima vivienda con teléfono, de la e-ésima entidad.
- F_{ei}^{**} = factor de expansión ajustado por estimación de población, de la i-ésima vivienda con teléfono, de la e-ésima entidad.

8.3 Ajuste por estimación de población

Los factores corregidos se ajustan por estimación de población al mes de noviembre de 2021, el procedimiento es el siguiente:

$$F_{ei}^{***} = F_{ei}^{**} \frac{x_e}{\hat{Y}_e}$$

Donde:

- X_e = estimación de población de INEGI para la e-ésima entidad.
- \hat{Y}_e = estimación de la población obtenida de la ENAPE 2021 en la e-ésima entidad.
- F_{ei}^{**} = factor de expansión de la i-ésima vivienda con teléfono, de la e-ésima entidad, ajustado a la población que reside en viviendas con teléfono reportada por ENDUTIH 2020.
- F_{ei}^{***} = factor de expansión de la i-ésima vivienda con teléfono, de la e-ésima entidad, ajustado por estimación de población.

8.4 Ajuste por calibración

Debido a que, el marco de muestreo de la ENAPE 2021 puede presentar algunas diferencias respecto a las estructuras de los registros administrativos de la SEP, se procedió a ajustar sus factores de expansión mediante técnicas de calibración, a efecto de reducir estas diferencias.

La calibración consiste en minimizar una función que impone condiciones para generar nuevos factores de expansión (w_k), partiendo de los factores originales (d_k), tomando como referencia la información muestral de p variables de interés la muestra (X_j $j=1, \dots, p$), así como fuentes auxiliares de información (registros administrativos de la SEP) traducidas en totales t_j $j=1, \dots, p$. El procedimiento es el siguiente:

$$\min \sum_{k=1}^n d_k F\left(\frac{w_k}{d_k}\right)$$

Sujeta a:

$$\sum_{k=1}^n w_k X_{kj} = t_j; \quad j = 1, \dots, P$$

Para este ajuste se consideró la función distancia $F(z)=z \ln(z)-z+1$ conocida como *Raking Ratio* propuesta por Deville y Sarndal².

En una calibración, los factores de expansión de la ENAPE 2021 se ajustaron de manera que se mantuvieran la población objetivo y los niveles educativos de la SEP:

- Población de 3 a 29 años.
- Niveles educativos.

² Deville, Jean-Claude, Sarndal, Carl-Erik (1992). "Calibration Estimators in Survey Sampling". Journal of the American Statistical Association, Vol 87. Núm. 418, pp. 376-382.

Las variables de escolaridad utilizadas para calibrar los factores de expansión la ENAPE 2021 fueron las siguientes:

- Ninguno
- Preescolar
- Primaria
- Secundaria
- Bachillerato
- Superior

Cabe mencionar que la calibración tiene el efecto de minimizar las posibles desviaciones de las estimaciones de las variables más relevantes de la encuesta, que surgen, en este caso, al considerar una muestra aleatoria. Por lo anterior, a medida que las variables sean desglosadas en otras subcategorías conceptuales, las estimaciones se deberán tomar con reserva por la posible presencia de sesgos, debido al volumen reducido de muestra.

9. Estimadores

Para la ENAPE 2021 se pueden calcular diversos tipos de estimadores. A continuación, se presentan los siguientes:

- Estimador del total

El estimador del total de la característica X y Y están dados por:

$$\hat{X} = \sum_{e=1}^L \sum_{i=1}^{m_e} F_{ei} X_{ei}$$

$$\hat{Y} = \sum_{e=1}^L \sum_{i=1}^{m_e} F_{ei} Y_{ei}$$

Donde:

F_{ei} = factor de expansión final de la i-ésima vivienda con teléfono, de la e-ésima entidad.

X_{ei} = valor observado de la característica de interés X en la i-ésima vivienda con teléfono, en la e-ésima entidad.

Para la estimación de tasas se utiliza el estimador de razón:

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

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

- **Estimador de una proporción**

$$\hat{p}_e = \frac{\sum_{i=1}^{m_e} F_{ei} X_{ei}}{\sum_{i=1}^{m_e} F_{ei}}$$

$$\hat{q}_e = 1 - \hat{p}_e$$

$$\hat{p} = \frac{\sum_{e=1}^L \sum_{i=1}^{m_e} F_{ei} X_{ei}}{\sum_{e=1}^L \sum_{i=1}^{m_e} F_{ei}}$$

10. Estimación de errores de muestreo

Para la evaluación de los errores de muestreo de las principales estimaciones nacionales se emplearon las expresiones de estimación de varianza para los estimadores de razón, estimadores de totales y para proporciones empleando las siguientes expresiones:

- Totales

$$\hat{V}(\hat{X}) = \sum_{e=1}^L \hat{M}_e \sum_{i=1}^{m_e} \frac{F_{ei} (X_{ei} - \bar{X}_e)^2}{m_e - 1}$$

$$\hat{V}(\hat{Y}) = \sum_{e=1}^L \hat{M}_e \sum_{i=1}^{m_e} \frac{F_{ei} (Y_{ei} - \bar{Y}_e)^2}{m_e - 1}$$

Donde:

$$\bar{X}_e = \frac{\sum_{i=1}^{m_e} F_{ei} X_{ei}}{\sum_{i=1}^{m_e} F_{ei}}$$

$$\bar{Y}_e = \frac{\sum_{i=1}^{m_e} F_{ei} Y_{ei}}{\sum_{i=1}^{m_e} F_{ei}}$$

$$\hat{M}_e = \sum_{i=1}^{m_e} F_{ei} X_{ei}$$

$$m_e = \sum_{i=1}^{m_e} 1$$

- Proporciones

$$\hat{V}(\hat{p}) = \sum_{e=1}^L \frac{\hat{p}_e \hat{q}_e}{m_e - 1}$$

- Razones

$$\hat{V}(\hat{R}) = \frac{\hat{V}(\hat{X}) + \hat{R}^2 \hat{V}(\hat{Y}) - 2\hat{R}C\hat{O}V(\hat{X}, \hat{Y})}{\hat{Y}^2}$$

$$COV(\hat{X}, \hat{Y}) = \sum_{e=1}^L \hat{M}_e \frac{\sum_{i=1}^{m_e} (x_{ei} - \bar{X}_e)(y_{ei} - \bar{Y}_e)}{m_e - 1}$$

Donde:

- X_{ei} = la variable de estudio X de la i-ésima viviendas, en la e-ésima entidad.
- \bar{X}_e = promedio de la variable de estudio X en la e-ésima entidad.
- M_e = número de personas en la e-ésima entidad.
- L = número de estratos a nivel nacional.
- \hat{Y}^2 = el cuadrado del total ponderado de la característica Y.

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

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

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

Donde:

- $\hat{\theta}$ = estimador del parámetro poblacional θ .
- $\hat{V} \hat{\theta}$ = valor observado de la característica de interés X en la i-ésima vivienda con teléfono, en la e-ésima entidad.

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.

11. Homologación de la Sematorización para los Umbrales e Indicadores de presión estadística³

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.

³ 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_umbralles.pdf.

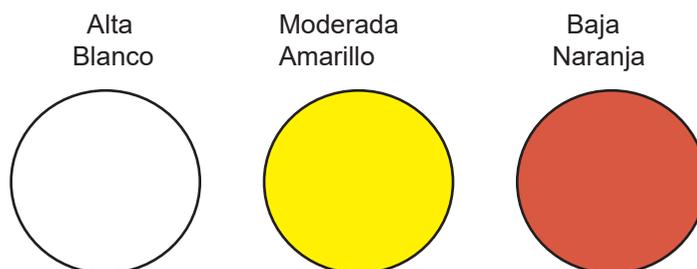
Umbrales aprobados para la cobertura de la CV

Interpretación	Semaforización	Viviendas/Hogares/Otras unidades diferentes a las económicas DGES/ DGEOSPJ
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.



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