



Diseño de la muestra censal

Censo de Población y Vivienda 2020

Instituto Nacional de Estadística y Geografía

Censo de Población y Vivienda 2020

Diseño de la muestra censal



Obras complementarias publicadas por el INEGI sobre el tema:

Diseño de la muestra censal 2010.

Catalogación en la fuente INEGI:

304.601072 Censo de Población y Vivienda (2020).
Censo de Población y Vivienda 2020 : diseño de la muestra censal / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2021.

vii, 14 p.

1. México - Población - Censos, 2020 - Metodología. 2. México - Población - Censos, 2020 - 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

2021, **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)** en el ejercicio de las atribuciones que la Ley del Sistema Nacional de Información Estadística y Geográfica le confiere, realizó el Censo de Población y Vivienda 2020.

Como parte de la documentación metodológica que acompaña a los resultados definitivos de este proyecto estadístico, se presenta el *Diseño de la muestra censal 2020* con la finalidad de poner a disposición de los usuarios información que les permita tener un panorama de las bases metodológicas a partir de las cuales se realizó el diseño de la muestra censal.

De esta forma, el **Instituto**, como parte del Sistema Nacional de Información Estadística y Geográfica, cumple con la función de suministrar a la sociedad y al Estado información de calidad, pertinente, veraz y oportuna para contribuir al desarrollo de México.

ÍNDICE

INTRODUCCIÓN	VII
I. OBJETIVOS	1
II. DISEÑO DE LA MUESTRA	2
1. POBLACION OBJETO DE ESTUDIO	2
2. DOMINIOS DE ESTUDIO	2
3. MARCO MUESTRAL	3
4. ESQUEMA DE MUESTREO	4
4.1 Unidades de muestreo	4
4.2 Estratificación	5
5. TAMAÑO DE MUESTRA	5
5.1 Afijación de la muestra dentro de los estratos	6
5.2 Factores de expansión	7
5.3 Cálculo de estimadores	7
5.4 Estimadores de la varianza	8
5.5 Error estándar	8
5.6 Coeficiente de variación (CV)	8
5.7 Intervalo de confianza	8
5.8 Estimadores de totales	9
5.9 Estimadores de promedios	11
5.10 Estimadores de proporciones	12
5.11 Estimadores de razón	13
5.12 Estimación de la varianza por el método de Jackknife	14
5.13 Consideración al método Jackknife	14

INTRODUCCIÓN

A partir del XII Censo General de Población y Vivienda 2000, se realizó el levantamiento de información utilizando dos tipos de cuestionarios: básico y ampliado. De la misma manera, durante el Censo de Población y Vivienda 2020 se aplicaron estos dos tipos de cuestionarios. Con el Cuestionario Ampliado se censaron alrededor de 4 millones de viviendas en el territorio nacional.

El Cuestionario Ampliado se aplicó en viviendas seleccionadas mediante criterios probabilísticos e incluye las preguntas del Cuestionario Básico y preguntas adicionales que ahondan sobre algunos aspectos de las características de las viviendas y sus habitantes, por lo que todas las viviendas fueron censadas con el Cuestionario Básico y algunas, las seleccionadas, respondieron preguntas adicionales.

En este documento se describen todos los aspectos técnicos relacionados con la muestra censal 2020, el marco de muestreo, el tamaño de muestra, el esquema de muestreo, así como la obtención de estimadores.

Es de vital importancia que las y los usuarios de la información de la muestra censal consideren el diseño de la muestra para la obtención de los estimadores que requieran.

El diseño de la muestra permite generar estimadores con precisión y confianza aceptables para la gran mayoría de las variables adicionales incluidas en el Cuestionario Ampliado y para niveles de desagregación municipal, sin embargo, siempre que se obtenga un estimador se debe considerar la precisión y confianza, antes de tomar una decisión basada en este.

I. OBJETIVOS

La muestra censal 2020 tiene por objetivo general proporcionar información sobre características de las viviendas y sus habitantes, con precisión y confianza medibles para tasas, promedios y proporciones, a nivel nacional, por entidad federativa, por municipio o demarcación territorial de la Ciudad de México y para cada una de las localidades de 50 mil o más habitantes.

Particularmente, la muestra busca proporcionar información, para los siguientes niveles de desagregación geográfica:

- Nacional.
- Para cada una de las 32 entidades federativas.
- Cada uno de los 2 469 municipios y demarcaciones territoriales en el país.
- Cada una de las localidades de 50 000 o más habitantes.

Adicionalmente, la muestra permite generar indicadores al mínimo nivel de desagregación geográfica permitido por el criterio de confidencialidad para los municipios que cumplen los siguientes criterios, teniendo como base los resultados de la Encuesta Intercensal 2015:

- Municipios con menos de 1 300 viviendas particulares habitadas.
- Municipios con muy alto grado de rezago social¹ y con 10 000 o menos viviendas particulares habitadas.
- Municipios con menor Índice de Desarrollo Humano² (IDH menor o igual al 0.55) y con 10 000 o menos viviendas particulares habitadas.
- Municipios con 40% o más de su población de 3 años y más que hable alguna lengua indígena y que no hable español.
- Municipios con 40% o más de su población que se considera afrodescendiente.

¹ https://www.coneval.org.mx/Medicion/IRS/Paginas/Indice_Rezago_Social_2015.aspx

² <https://www.mx.undp.org/content/mexico/es/home/library/poverty/informe-de-desarrollo-humano-municipal-2010-2015--transformando-.html>

II. DISEÑO DE LA MUESTRA

El diseño de la muestra censal 2020 es estratificado por conglomerados y se seleccionó en una sola etapa, es decir, se seleccionaron áreas completas, manzanas o localidades según sea el caso, en las cuales se aplicó el Cuestionario Ampliado en todas las viviendas particulares habitadas.

1. POBLACION OBJETO DE ESTUDIO

La población objeto de estudio está constituida por las viviendas particulares habitadas y sus residentes habituales dentro del territorio nacional.

2. DOMINIOS DE ESTUDIO

Se definieron lo siguientes dominios de estudio:

- Nacional.
- Estatal.
- Estatal con desagregación en cuatro tamaños de localidad:
 - Menos de 2 500 habitantes.
 - De 2 500 a 14 999 habitantes.
 - De 15 000 a 49 999 habitantes.
 - 50 000 y más habitantes.
- Cada uno de los 2 453 municipios de los 31 estados y de cada una de las 16 demarcaciones territoriales de la Ciudad de México.
- Las localidades de 50 mil o más habitantes³ del país existentes en 2019.

Es importante señalar que durante 2020 se consolidó la conformación de cuatro municipios (Honduras de la Sierra, Chiapas; Seybaplaya, Campeche; Hueyapan, Morelos y San Quintín, Baja California), por lo que no se fueron considerados en el diseño de la muestra. Asimismo, en el municipio de La Magdalena Tlaltelulco, Tlaxcala, no fue posible

³ Posterior al Censo 2020, se identificaron 18 localidades que superan 50 mil habitantes y que no fueron consideradas en el diseño debido a que antes del evento censal no superaban dicha población. Los resultados para estos municipios deben revisarse con cuidado antes de ser utilizados, considerando el error estándar y los intervalos de confianza.

realizar la captación de la información. Por lo anterior, solo se presenta la estimación de viviendas y población para los municipios de La Magdalena Tlaltelulco, Tlaxcala; Honduras de la Sierra, Chiapas; Seybaplaya, Campeche. Para los municipios de Hueyapan, Morelos y San Quintín, Baja California, los resultados deben revisarse con cuidado antes de ser utilizados, considerando el error estándar y los intervalos de confianza. A continuación, se presentan las características del diseño considerando los 2 465 municipios incluidos en el marco de muestreo utilizado para el diseño de la muestra censal.

Los 2 465 municipios existentes en 2019 se agruparon de la siguiente manera:

AGRUPACIÓN DE MUNICIPIOS	
Municipios que se censan con Cuestionario Ampliado (certeza)	Municipios con menos de 1 300 viviendas particulares.
	Municipios con poblaciones de interés (IDH, rezago social, lengua indígena y afrodescendencia).
Resto de los municipios	Sin localidades de más de 50 mil hab.
	Con localidades de más de 50 mil hab.

3. MARCO MUESTRAL

El marco de muestreo está conformado por los catálogos de AGEB, localidad y manzanas del Marco Geoestadístico 2019, así como las estadísticas asociadas a cada área.

El marco para la muestra del Censo de Población y Vivienda 2020 tomó como base la información obtenida en el Censo de Población y Vivienda 2010, la Encuesta Intercensal 2015, el Registro Único de Viviendas 2012 (RUV 2012) y los Censos Económicos 2014 y 2019.

No se consideraron dentro del marco de muestreo algunas áreas del catálogo cartográfico por no incluir población objeto de estudio:

- Localidades sin población residente habitual en los últimos tres eventos (Censos y Encuesta Intercensal 2015).
- Áreas en las que únicamente hay viviendas colectivas.
- Áreas por su tipo (bajo puente, camellón, económica, glorieta, parque o jardín, etc.).
- Áreas que únicamente tienen servicios (escuelas, edificaciones, cuerpos de agua, etcétera).

4. ESQUEMA DE MUESTREO

El esquema de muestreo para la muestra censal es estratificado, por conglomerados y en una sola etapa de selección, las unidades primarias de muestreo corresponden a manzanas o localidades completas, es decir, en cada una de las áreas seleccionadas todas las viviendas particulares habitadas son censadas utilizando el Cuestionario Ampliado. Para la selección de las áreas se utilizó un muestreo aleatorio simple sin reemplazo para cada estrato definido.

Debido a que se requiere más detallada de algunos sectores de la población, se determinó censar con Cuestionario Ampliado a todos los municipios que cumplieran con al menos uno de los siguientes criterios:

MUNICIPIOS A CENSAR CON CUESTIONARIO AMPLIADO ⁴		786
Estimación de la Encuesta Intercensal 2015	Con menos de 1 300 viviendas particulares habitadas.	683
	Municipios con 40% o más de su población de 3 años y más que hablan alguna lengua indígena y que no hablan español.	22
	Municipios con 40% o más de su población que se considera afrodescendiente.	12
Municipios con muy alto grado de rezago social en 2015 con 10 000 o menos viviendas particulares habitadas.		169
Municipios con Índice de Desarrollo Humano (IDH) menor a 0.55 y con 10 000 o menos viviendas particulares habitadas.		131

4.1 Unidades de muestreo

Las unidades primarias de muestreo son áreas geográficas completas, ya sean manzanas o localidades. En este mismo sentido, las unidades de muestreo a seleccionar (manzanas o localidades) varían de acuerdo al diseño muestral que se aplique en cada una de ellas.

- Localidades de 250 o más viviendas particulares habitadas. La Unidad Primaria de Muestreo (UPM) está constituida generalmente por manzanas.
- Las localidades de menos de 250 viviendas, las UPM son las propias localidades.

⁴ Algunos municipios cumplen con más de un criterio de inclusión, por lo que la suma de los mismos no corresponde al total de municipios censados con cuestionario ampliado.

4.2 Estratificación

Los 2 453 municipios de los 31 estados y las 16 demarcaciones territoriales de la Ciudad de México se clasificaron en dos grupos: “municipios seleccionados con certeza” y resto de los municipios. En el caso de los municipios seleccionados con certeza todas sus áreas se incluyeron en la muestra con probabilidad uno, por lo cual, la totalidad de viviendas de estas áreas se censaron aplicando el Cuestionario Ampliado.

En el resto de los municipios se seleccionaron áreas completas (manzanas o localidades menores de 250 viviendas particulares habitadas). Dentro de cada municipio muestreado se definieron estratos por tamaño de localidad de la siguiente manera.

- Menos de 50 000 habitantes:
 - Menos 2 500 habitantes.
 - 2 500 a 4 999 habitantes.
 - 5 000 a 14 999 habitantes.
 - 15 000 a 49 999 habitantes.
- Para las localidades de 50 mil y más habitantes, cada una de ellas conformó un estrato.

5. TAMAÑO DE MUESTRA

Con el objetivo de garantizar que las estimaciones que se obtengan a partir de la muestra del Censo de Población y Vivienda 2020 tengan precisión y confianza aceptables, es necesario que el tamaño de muestra que se defina sea suficiente para cada parámetro de interés.

En general, para obtener el tamaño de muestra se consideró una confianza del 90%, un error relativo del 0.2, un efecto de diseño de 1.44 y una tasa de respuesta de 90%, según fuera el caso. Con estos parámetros, el tamaño de muestra distribuyó de la siguiente manera:

- Municipios de 1 300 viviendas o menos o certeza fueron censados.
- En los municipios de más de 1 300 viviendas al menos 2057 viviendas y si el municipio incluye localidades de más de 200 mil habitantes, al menos 4 470 viviendas.

Los tamaños de muestra se obtuvieron mediante la siguiente expresión:

$$n = \frac{Z^2}{r^2} \times \frac{q}{p} \times \frac{DEFF}{1 - TNR}$$

Donde:

p = Proporción que se desea estimar

$q = 1 - p$

r = Error relativo máximo

z = Nivel de confianza al 90%

TNR = Tasa de no respuesta

$DEFF$ = Efecto de diseño

Adicionalmente, se realizó un ajuste por población finita al tamaño de muestra requerido. La expresión es la siguiente:

$$n' = \frac{n}{1 + \frac{n}{N}}$$

Donde:

n' = Tamaño de muestra ajustado por población finita

n = Tamaño de muestra estimado

N = Total de viviendas

Los tamaños de muestra fijados garantizan estimaciones municipales con precisión y confianza aceptables (de acuerdo con los parámetros fijados) para proporciones cercanas a 0.01 o mayores.

5.1 Afijación de la muestra dentro de los estratos

Para el caso de los municipios seleccionados con certeza no es necesario realizar algún tipo de afijación de la muestra, pero para el resto de los municipios la afijación del tamaño de muestra se realizó proporcional al total de viviendas particulares habitadas en el municipio.

$$n_h = \frac{N_h}{N_{mun}} n_{mun}$$

Donde:

N_h = Total de viviendas en el estrato h

N_{mun} = Total de viviendas en el municipio

n_{mun} = Tamaño de muestra por municipio

n_h = Tamaño de muestra a seleccionar por estrato

5.2 Factores de expansión

De acuerdo con el esquema de muestreo, la probabilidad de selección de cada una de las UPM (manzanas o localidades) en el interior de cada estrato, está dada por la siguiente fórmula:

$$P_{hj} = \frac{n_h}{N_h}$$

n_h = Número de UPM en muestra en el h –ésimo estrato

N_h = Total de UPM en el h –ésimo estrato

De esta manera, el factor de expansión para cada una de las viviendas (y cada uno de sus residentes) de la j –ésima UPM del h –ésimo estrato está dado por:

$$F_{hj} = \frac{1}{P_{hj}} = \frac{N_h}{n_h}$$

Para cada UPM, los factores de expansión se multiplicaron por un factor adicional que recibe el nombre de ajuste por no respuesta.

El diseño de la muestra no es autoponderado, es decir, que cada vivienda en muestra representa a un número diferente de viviendas. Los factores de expansión son diferentes entre las áreas y también la tasa de no respuesta es diferencial, y por ello, cualquier indicador o tabulación a partir de la muestra censal requiere del uso de los factores de expansión.

5.3 Cálculo de estimadores

El estimador es un estadístico obtenido a partir de los datos captados de sólo una parte de la población (muestra), se utiliza como aproximación de un parámetro desconocido de la población.

Para la muestra censal 2020, el estimador del total se calculó mediante el método propuesto por Horvitz-Thompson, del cual se derivan los estimadores de los promedios, proporciones y razones.

Existen diversos paquetes estadísticos con los cuales se pueden obtener los estimadores, para el caso de la muestra censal 2020 se utilizó el *software* SAS 9.4.

5.4 Estimadores de la varianza

La varianza de los estimadores lineales que se presentan en los tabulados de la muestra censal 2020 se obtuvo mediante el método de series de Taylor. Al igual que para el Censo 2010 y la Encuesta Intercensal 2015 este método se seleccionó de manera que los usuarios que realicen la explotación de la muestra censal obtengan los mismos resultados haciendo uso de algún paquete estadístico de los más utilizados.

Asimismo, se decidió que la estimación de la varianza de los indicadores no lineales se aproximara por técnicas de remuestreo. Tanto para la muestra censal 2010, la Encuesta Intercensal 2015 y para la muestra censal 2020 se decidió utilizar la técnica conocida como Jackknife.

5.5 Error estándar

El error estándar se calcula con los datos de la muestra y se define como la raíz cuadrada de la varianza estadística del estimador. Mide el grado de precisión con que el estimador se aproxima al resultado que se habría obtenido si se hubiera entrevistado a toda la población objeto de estudio bajo las mismas condiciones. También se le conoce como error de muestreo o error muestral. Por sí solo es difícil de interpretar, por lo que se utiliza el coeficiente de variación y el intervalo de confianza.

5.6 Coeficiente de variación (CV)

Para efectos de presentación de resultados de la muestra censal 2020, el coeficiente de variación se define como la variación porcentual del error estándar a la estimación central. Es el cociente entre el error estándar y el estimador multiplicado por 100. También se le conoce como error relativo. Constituye un criterio uniforme para determinar la precisión de un estimador.

5.7 Intervalo de confianza

El intervalo de confianza es una de las formas más efectivas de la inferencia estadística. A partir de éste se puede conocer qué tan acertada es la estimación realizada, pues indica que el valor verdadero se encuentra entre el límite inferior y el límite superior del intervalo,

con una probabilidad o nivel de confianza α prefijado. Para la muestra censal se utilizó 90% de confianza.

5.8 Estimadores de totales

El estimador del total de la población para un dominio de estudio dado d , es la suma de los estimadores de los totales de los estratos:

$$\hat{Y}_d = \sum_{h=1}^H \hat{Y}_h$$

$$\hat{Y}_d = \sum_{h=1}^H \sum_{j=1}^{n_h} F_{hj} \sum_{k=1}^{M_{hj}} y_{hjk}$$

Donde:

\hat{Y}_d = Total estimado de la característica de interés para el dominio d

\hat{Y}_h = Total estimado de la característica de interés en el estrato h del dominio d

F_{hj} = Factor de expansión de la j -ésima UPM en el h -ésimo estrato

y_{hjk} = Valor de la característica de interés en la k -ésima vivienda de la j -ésima UPM en el h -ésimo estrato

H = Número de estratos en el dominio d

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

M_{hj} = Número de unidades en muestra dentro de la j -ésima UPM en el h -ésimo estrato

Al contar con muestras independientes por estrato, la varianza del estimador del total es igual a la suma de las varianzas de los estimadores de los totales de estratos.

$$var(\hat{Y}_d) = \sum_{h=1}^H var(\hat{Y}_h) = \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

Donde $S_h^2 = \sum_{j=1}^{n_h} \frac{(Y_{hj} - \bar{Y}_h)^2}{N_h - 1}$ es la varianza del estrato h .

Y el estimador de la varianza del estimador del total, para un dominio dado, está dada por:

$$\widehat{var}(\hat{Y}_d) = \sum_{h=1}^H \widehat{var}(\hat{Y}_h) = \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{S}_h^2}{n_h}$$

Donde:

$$\hat{S}_h^2 = \sum_{j=1}^{n_h} \frac{(y_{hj} - \hat{y}_h)^2}{n_h - 1}$$

Nótese que \hat{S}_h^2 es la misma expresión que S_h^2 , aunque la primera se refiere a los valores de la muestra y la segunda al valor poblacional de cada estrato.

La varianza $var(\hat{Y}_d)$ mide la variabilidad de los estimadores en el dominio y $\widehat{var}(\hat{Y}_h)$ en los diferentes estratos.

El factor $(1 - \frac{n_h}{N_h})$ llamado Factor de corrección de población finita (FCP) en general tiende a 1 para los diversos dominios de estudio de la muestra censal 2020, por lo que no se consideró para el cálculo de los estimadores de varianza; además el no involucrar el FCP simplificó los cálculos.

Entonces el error estándar asociado a \hat{Y}_d está dado por:

$$EE_{\hat{Y}_d} = \sqrt{\widehat{var}(\hat{Y}_d)}$$

Y la estimación de su coeficiente de variación (CV) es:

$$CV = \frac{EE_{\hat{Y}_d}}{\hat{Y}_d} \times 100$$

A este coeficiente también se le conoce como error relativo.

El intervalo de confianza que se presenta en los tabulados se construyó con una confianza de 90% y está dado por:

$$P[\hat{Y}_d - 1.645 \sqrt{\widehat{var}(\hat{Y}_d)} < Y_d < \hat{Y}_d + 1.645 \sqrt{\widehat{var}(\hat{Y}_d)}] \approx 0.90$$

5.9 Estimadores de promedios

$$\hat{Y}_d = \frac{\hat{Y}_d}{\hat{M}_d} = \frac{\hat{Y}_d}{\sum_{h=1}^H \sum_{j=1}^{n_h} F_{hj}}$$

\hat{M}_d = Estimación del total en el dominio d .

Y la varianza de \hat{Y}_d es:

$$var(\hat{Y}_d) = \sum_{h=1}^H var\left(\frac{\hat{Y}_d}{\hat{M}_d}\right) = \sum_{h=1}^H \frac{1}{\hat{M}_d^2} var(\hat{Y}_d) = \sum_{h=1}^H \frac{N_h^2}{\hat{M}_d^2} \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

Si lo que se requiere es estimar la varianza de la media muestral de una característica en particular, se usa la fórmula:

$$var(\hat{Y}_d) = \sum_{h=1}^H var\left(\frac{\hat{Y}_d}{\hat{M}_d}\right) = \sum_{h=1}^H \frac{1}{\hat{M}_d^2} var(\hat{Y}_d) = \sum_{h=1}^H \frac{N_h^2}{\hat{M}_d^2} \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{S}_h^2}{n_h}$$

Donde el error estándar para la media está dado por:

$$EE_{\hat{Y}_d} = \sqrt{var(\hat{Y}_d)}$$

Y la estimación del coeficiente de variación (CV) es:

$$CV = \frac{EE_{\hat{Y}_d}}{\hat{Y}_d} \times 100$$

De manera similar, para la estimación de la media en el dominio d , se puede construir el intervalo de confianza aproximado a 90 por ciento:

$$P[\hat{Y}_d - 1.645 \sqrt{var(\hat{Y}_d)} < \bar{Y}_d < \hat{Y}_d + 1.645 \sqrt{var(\hat{Y}_d)}] \approx 0.90$$

5.10 Estimadores de proporciones

Si lo que se requiere estimar es la proporción P de elementos de la población que tienen una característica determinada, se usan las equivalencias dadas por:

$$\bar{Y}_d = P_d, \bar{y}_d = p_d$$

Entonces el estimador para la proporción es:

$$\hat{P}_d = \frac{\hat{Y}_d}{\hat{M}_d} = \frac{\hat{Y}_d}{\sum_{h=1}^H \sum_{j=1}^{n_h} F_{hj}}$$

Con una varianza estimada dada por la siguiente expresión:

$$\widehat{var}(\hat{Y}_d) = \widehat{var}(\hat{P}_d) = \sum_{h=1}^H \frac{N_h^2}{\hat{M}_d^2} \left(1 - \frac{n_h}{N_h}\right) \frac{p_d(1-p_d)}{n_h}$$

Donde el error estándar de la proporción se obtiene con:

$$EE_{\hat{P}_d} = \sqrt{\widehat{var}(\hat{P}_d)}$$

Y la estimación de su coeficiente de variación (CV) está dada por:

$$CV = \frac{EE_{\hat{P}_d}}{\hat{P}_d} \times 100$$

Para el cálculo del intervalo de confianza para un porcentaje o proporción se usan, a diferencia de los anteriores, intervalos de confianza *logit*.

Los límites de confianza *logit* P_L y P_U se calculan como:

$$\text{Limite inferior } P_L = \frac{e^{\lambda_L}}{(1 + e^{\lambda_L})}$$

$$\text{Limite superior } P_U = \frac{e^{\lambda_U}}{(1 + e^{\lambda_U})}$$

Donde:

$$\lambda_L = \log\left(\frac{\hat{p}}{1 - \hat{p}}\right) - z_{df, \frac{\alpha}{2}} \times \frac{EE_{\hat{p}}}{\left(\frac{\hat{p}}{1 - \hat{p}}\right)}$$

$$\lambda_U = \log\left(\frac{\hat{p}}{1 - \hat{p}}\right) + z_{df, \frac{\alpha}{2}} \times \frac{EE_{\hat{p}}}{\left(\frac{\hat{p}}{1 - \hat{p}}\right)}$$

5.11 Estimadores de razón

Supongamos que se desea calcular el estimador de la razón de las características Y y X , con $R = Y/X$, la cual está dada en la siguiente expresión:

$$\hat{R}_d = \frac{\hat{Y}_d}{\hat{X}_d} = \frac{\sum_{h=1}^H \sum_{j=1}^{n_h} F_{hj} \sum_{k=1}^{M_{hj}} y_{hjk}}{\sum_{h=1}^H \sum_{j=1}^{n_h} F_{hj} \sum_{k=1}^{M_{hj}} x_{hjk}}$$

Donde la estimación de su varianza se obtiene:

$$\widehat{var}(\hat{R}_d) = \sum_{h=1}^H \widehat{var}_h(\hat{R}_d)$$

Por tanto, el error estándar de la razón se estima por:

$$EE_{\hat{R}_d} = \sqrt{\widehat{var}(\hat{R}_d)}$$

Y la estimación de su coeficiente de variación (CV) se calcula:

$$CV = \frac{EE_{\hat{R}_d}}{\hat{R}_d} \times 100$$

Para el estimador de razón en el dominio d , se puede construir el intervalo de confianza aproximado a 90% con la siguiente expresión:

$$P[\hat{R}_d - 1.645 \sqrt{\widehat{var}(\hat{R}_d)} < R_d < \hat{R}_d + 1.645 \sqrt{\widehat{var}(\hat{R}_d)}] \approx 0.90$$

Con un efecto del diseño (DEFF) asociado a la proporción estará dado por:

$$DEFF_{\hat{R}_d} = \frac{\widehat{var}(\hat{R}_d)}{\hat{V}_{MAS}(\hat{R}_d)}$$

Donde $\hat{V}_{MAS}(\hat{R}_d)$ es la varianza estimada del estimador de la razón considerando un muestreo aleatorio simple (MAS).

5.12 Estimación de la varianza por el método de Jackknife

El método de estimación Jackknife forma parte de las técnicas de remuestreo, las cuales se usan principalmente cuando se requiere estimar la varianza de estimadores no lineales. El estimado calculado con este método para la presentación de resultados de la muestra censal 2020 fue la Tasa Global de Fecundidad.

Dicho método se sustenta en hacer estimaciones eliminando, cada vez, una UPM diferente de la muestra completa para crear réplicas; por lo tanto, se tiene la misma cantidad de estimaciones que de número de UPM en la muestra. El estimador de la varianza Jackknife para el estimador $\hat{\theta}$ se calcula mediante la expresión:

$$\hat{V}_{JK}(\hat{\theta}) = \sum_{r=1}^R \alpha_r (\hat{\theta} - \hat{\theta}_r)^2$$

Donde:

$\hat{\theta}$ = Estimador calculado a partir de la muestra completa

$\hat{\theta}_r$ = Estimador calculado en la r -ésima réplica usando los factores replicados $F_h^{(r)}$, dados por la expresión:

$$F_h^{(r)} \begin{cases} F_h & \text{Si la UPM eliminada en la } r\text{-ésima réplica no pertenece al estrato donante} \\ \frac{F_h}{\alpha_r} & \text{Si la UPM eliminada en la } r\text{-ésima réplica pertenece al estrato donante} \end{cases}$$

α_r = Coeficientes de Jackknife, $\alpha_r = \frac{n_{hr} - 1}{n_{hr}}$

h_r = Estrato donante o estrato del cual se elimina la UPM en la r -ésima réplica

n_{hr} = Número de UPM en estrato donante

La estimación de la varianza Jackknife tiene $R - H$ grados de libertad.

5.13 Consideración al método Jackknife

El método supone que se tiene más de una UPM por estrato en la muestra, de no ser así se debe dar un tratamiento especial. Éste consiste en no considerar para la eliminación de UPM a los estratos que tienen sólo una UPM (caso de áreas censadas, por ejemplo); sin embargo, sí se deben considerar para el cálculo de las $\hat{\theta}_r$, así como para $\hat{\theta}$. Si no se consideraran a las áreas censadas en los cálculos, para evitar el problema de los estratos con una única UPM en la muestra, se obtendrían estimaciones sesgadas.

Conociendo México

800 111 46 34

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

    **INEGI** Informa