

Estimación de población a través del marco de muestreo



Instituto Nacional de Estadística y Geografía

**Estimación de población
a través del marco de muestreo**



Catalogación en la fuente INEGI:

310.01 Instituto Nacional de Estadística y Geografía (México).
Estimación de población a través del marco de muestreo / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2021

67 p.

1. Estadística - Metodología.

Conociendo México

800 111 4634
www.inegi.org.mx
atencion.usuarios@inegi.org.mx

 **INEGI Informa**  **@INEGI_INFORMA**

Registro en trámite

DR © 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.

Contenido

Presentación.....	1
Introducción.....	2
1. Metodología de la estimación de población.....	4
1.1 Resumen metodológico.....	4
1.2 Diseño de la muestra.....	6
1.3 Suavizamiento de estimaciones.....	16
2. Resultados de la estimación de población a nivel nacional y por entidad.....	20
ANEXOS.....	21
1. Diseño del Marco Maestro de Muestreo (MMM) y de la Muestra Maestra (MM).....	21
2. Diseño del Marco de Áreas de Nuevo Crecimiento (MANC) y de la Muestra en Áreas de Nuevo Crecimiento (MM).....	28
3. Intervalos de confianza de las estimaciones de población.....	32
4. Gráficos de estimaciones de población.....	34
5. Diagnóstico del modelo.....	46
6. Estimador de población a nivel nacional y por entidad federativa.....	53
7. Estimación por enfoque de muestreo y por enfoque asistido por modelos.....	55
8. Prácticas Nacionales e Internacionales sobre Estimación de Población.....	60
Bibliografía.....	66

Presentación

El **Instituto Nacional de Estadística y Geografía (INEGI)** es un organismo público autónomo responsable de normar y coordinar el Sistema Nacional de Información Estadística y Geográfica (SNIEG), así como de captar y difundir información de México en cuanto al territorio, los recursos, la población y economía, que permita dar a conocer las características de nuestro país y ayudar a la toma de decisiones.¹

El objetivo prioritario del INEGI es lograr que el SNIEG suministre a la sociedad y al Estado información de calidad, pertinente, veraz y oportuna, a efecto de coadyuvar al desarrollo nacional, bajo los principios de accesibilidad, transparencia, objetividad e independencia. Dentro de sus propósitos se encuentra el producir información estadística, la promoción del conocimiento y el uso de la información.²

El Instituto genera estadística básica, a partir de tres fuentes de información: censos, encuestas y registros administrativos, además de generar estadística derivada por medio de otras técnicas estadísticas. Con esta información, produce indicadores demográficos, sociales y económicos, además de la contabilidad nacional.

Dentro de la estadística básica en materia sociodemográfica que produce el INEGI, se encuentra la cuenta poblacional de México. Por excelencia, la estadística se produce por medio del Censo de Población y Vivienda (CPV) que se realiza cada diez años. Este ejercicio estadístico es facultad exclusiva del INEGI (Ley del SNIEG, artículo 59). Sin embargo, en periodos intercensales, otra fuente de información que proporciona una estimación de la población es a través de un muestreo de viviendas representativo de México. Con el conteo de ocupantes en la muestra de viviendas se puede generar la **estimación de la población del país y sus entidades federativas, a través de un marco de muestreo de viviendas**.

El INEGI, comprometido con el principio de transparencia, pone a disposición del SNIEG el documento metodológico en el cual se describen los procesos que le permiten actualizar la estimación de la población de México.

1 INEGI, https://www.inegi.org.mx/inegi/quienes_somos.html

2 INEGI, <https://www.inegi.org.mx/inegi/contenido/instituto.html>

Introducción

Entre las finalidades del Instituto se encuentra el producir y difundir información estadística al SNIEG, en cuanto al territorio, los recursos, la población y economía, que permita dar a conocer las características de nuestro país y ayudar a la toma de decisiones. Para el caso de la información demográfica, el INEGI debe profundizar en temas de población y dinámica demográfica, salud, educación, empleo, vivienda, distribución de ingreso y pobreza.

La cuenta poblacional de un país representa una de las estadísticas más importantes para llevar a cabo la función del Estado, además de un sinfín de usos y decisiones de actores tanto públicos como privados, en función de dicha estadística.

La cuenta poblacional realizada por el CPV marca el punto cero de la demografía de un país para los próximos diez años, cuya dinámica irá modificándose en los años subsecuentes, debido a los nacimientos, la mortalidad y la migración.

Como alternativa a la información censal, se tienen las encuestas, o, dicho de otra manera, lo que se obtiene a través de la teoría del muestreo, donde a partir de la cuenta de un subconjunto de viviendas (muestra) representativas y aleatorias del país, se pueden realizar inferencias respecto a la población, lo que se denominará en lo subsecuente como **estimación de la población**. Es importante señalar que no es el resultado de predicciones futuras con base a supuestos del comportamiento demográfico en cualquiera de sus componentes (natalidad, mortalidad o migración); sino el **resultado inferencial del conteo directo de viviendas y sus residentes**, que genera una estimación del total poblacional de México.

Esta metodología se dio a conocer a través de la consulta pública “Actualización de la población a través de la muestra maestra de viviendas y su validación por estimaciones de población a corto plazo”, realizada del 11 de febrero al 8 de marzo de 2019, donde se convocó a usuarios interesados en la materia a proporcionar comentarios y recomendaciones. La consulta pública es un mecanismo conveniente para que el INEGI reciba una retroalimentación en términos de lo que piensa adecuar metodológicamente hablando.

La consulta se llevó a cabo de acuerdo con la siguiente estrategia:

1. Publicación vía web (11 de febrero al 08 de marzo 2019).
2. Reunión nacional con expertos en materia de población (05 de marzo 2019).
3. Reunión técnica del INEGI y otras instituciones para la revisión de la metodología (23 de abril 2019).
4. A través de los usuarios del foro del SNIEG (11 de febrero al 08 de marzo 2019).

El informe de la Consulta Pública se presentó a la Junta de Gobierno del INEGI en la Quinta sesión del 11 de junio de 2019, dando cumplimiento con la correspondiente normatividad, previo a su implementación en las encuestas en hogares, incluidas las que cumplen el carácter de Información de Interés Nacional.

Derivado del proceso de consulta pública, el INEGI determinó llevar a cabo la estimación de población como se describe a lo largo del presente documento, descartando la validación de estimaciones a corto plazo a través de modelos demográficos.

La estimación de población se ha venido desarrollando desde el primer trimestre de 2018 hasta el segundo trimestre de 2021. Se cuenta con catorce estimaciones en el tiempo (una estimación por trimestre) lo que permite contar con una primera serie de estadísticas de población.

Adicionalmente, con los resultados del CPV 2020 fue posible validar la calidad de las estimaciones obtenidas a través del marco de muestreo, lo que da certeza y robustez a la metodología, además de que permitió identificar nuevas áreas geográficas al Marco Geoestadístico del INEGI que a su vez fueron incorporadas a la metodología de estimación de población.

La finalidad del presente documento es describir la metodología y los principales resultados de la estimación de la población para los periodos comprendidos entre el primer trimestre de 2018 y el segundo trimestre de 2021.

El contenido se estructura en tres apartados, el primero corresponde a la metodología de la estimación de la población, el cual inicia con algunas características generales sobre aspectos metodológicos y como subapartados se desglosan cada uno de los componentes que implica el proceso para llegar a la estimación de la población a través de la Muestra Maestra y las Áreas de Nuevo Crecimiento; en el segundo apartado se presenta el ejercicio de la estimación de la población, por medio de la tabla de datos y gráficos por entidad federativa; finalmente, un tercer apartado que incluye un anexo con diferentes desarrollos y materiales de referencia.

1. Metodología de la estimación de población

1.1 Resumen metodológico

La estimación de población tiene por objeto medir el total de la población del país a través del tiempo, en periodos censales o intercensales.

La metodología se basa en la teoría de muestreo. En primer lugar, se selecciona una muestra probabilística de viviendas particulares habitadas. Posteriormente, se recaba información sobre el número de ocupantes en dichas viviendas. Finalmente, se estima población puntualmente y por intervalos de confianza. Lo anterior se realiza trimestralmente, por entidad federativa y a nivel nacional.

La muestra de viviendas se obtiene del Marco de Muestreo del INEGI, el cual comprende todas las áreas geográficas del país, donde puede haber población en un tiempo t . Este Marco de Muestreo se conformó a partir de la base cartográfica³ del CPV 2010 y de su información estadística, en el tiempo t_0 . Posteriormente, el Marco de Muestreo mantiene un esquema de actualización de manera periódica, incorporando nuevas áreas geográficas como manzanas o localidades, donde puede haber personas.

Para ilustrar lo anterior, en la Imagen 1 se puede observar del lado izquierdo las áreas geográficas de una ciudad con base en la información del CPV 2010 (tiempo t_0). Del lado derecho se puede apreciar en azul, la creación de nuevas manzanas al tiempo t_i .

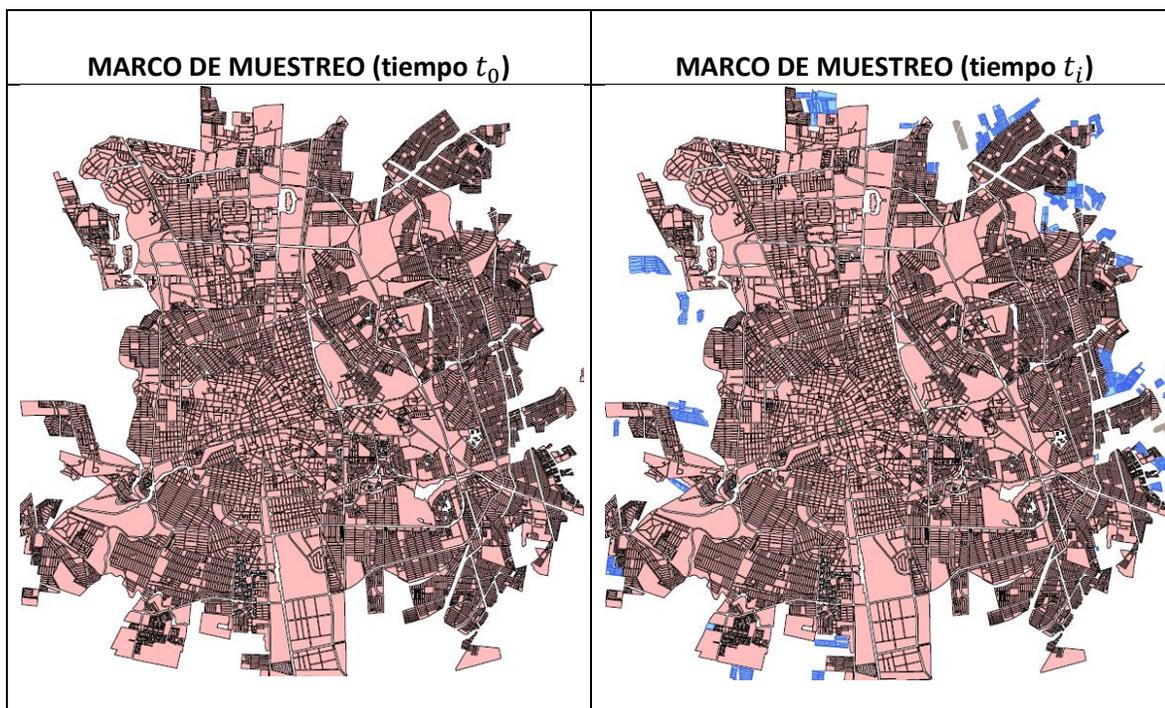


Imagen 1. Áreas geográficas de una ciudad en el tiempo t_0 (lado izquierdo) y áreas geográficas de la misma ciudad actualizadas al tiempo t_i .

³ Se tomó como base el Marco Geoestadístico 5.0 A, con la información cartográfica del CPV 2010 <https://www.inegi.org.mx/temas/mg/>

Por lo anterior, la estimación de población al tiempo t_i , denotada por $\hat{X}_{t_i}^{PT}$, se infiere por medio del conteo de personas en la muestra probabilística de viviendas (ver imagen 2, lado derecho) que se selecciona de la totalidad de las áreas geográficas (ver imagen 2, lado izquierdo) que existen al tiempo t_i .

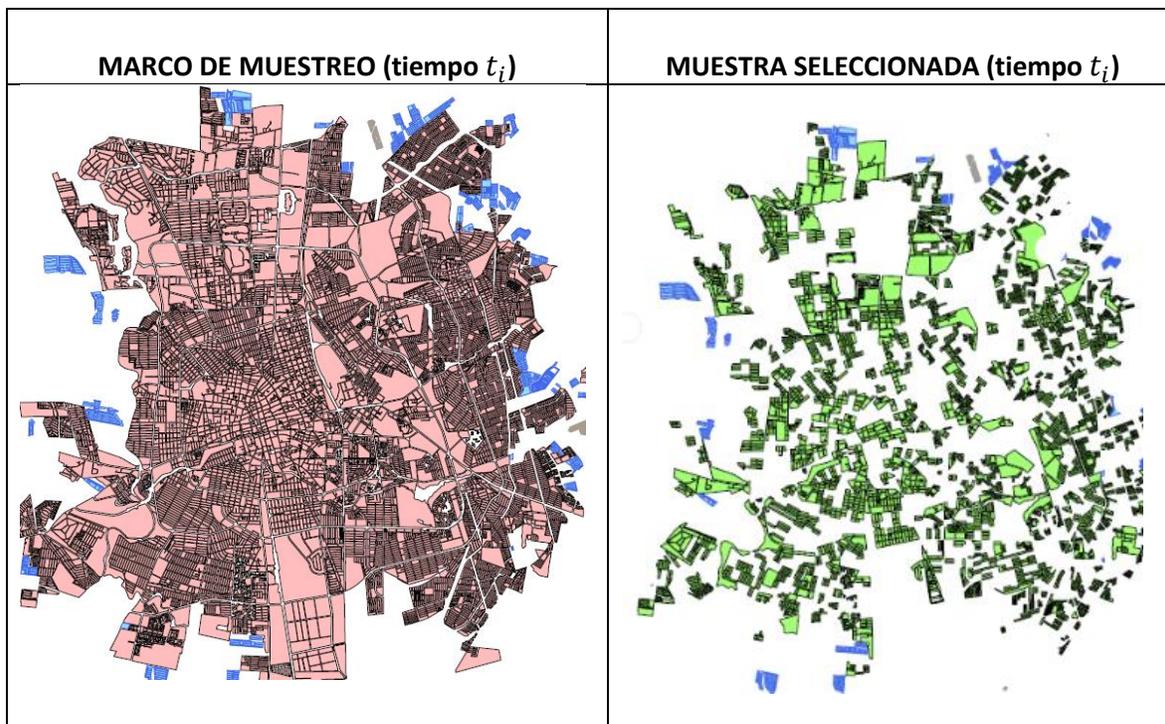


Imagen 2. Áreas geográficas de una ciudad actualizadas al tiempo t_i (lado izquierdo) y muestreo probabilístico de áreas donde se cuentan el número de personas, por su edad y sexo (lado derecho).

La muestra de n viviendas para el periodo t_i se puede denotar por $s_{t_i} = \{v_{t_i,1}, v_{t_i,2}, \dots, v_{t_i,n}\}$, donde cada vivienda tiene asociada una probabilidad de inclusión $w_{t_i,j}$ con $j = 1, \dots, n$, de acuerdo con el diseño muestral. A su vez, para cada vivienda se tiene el conteo de personas denotado por $x_{t_i,j}$ con $j = 1, \dots, n$.

Por lo tanto, la estimación de población en el tiempo t_i es el resultado de expandir los valores muestrales $x_{t_i,j}$ por el inverso de la probabilidad de inclusión $f_{t_i,j}$.

$$\hat{X}_{t_i}^{PT} = \sum_{\forall j} f_{t_i,j} * x_{t_i,j}$$

Finalmente, las estimaciones obtenidas para los diferentes periodos t_i , están sujetas a variaciones debidas al propio muestreo, por lo que es natural que oscilen a través del tiempo, lo que posibilita que, en ocasiones, el periodo k respecto al periodo $k-1$, presente decrementos o viceversa. Por lo tanto, para mitigar estas variaciones, el conjunto de estimaciones en el tiempo $\{(t_i, \hat{X}_{t_i}^{PT}), \forall i\}$

se ajusta a un modelo de regresión polinomial de grado m , de manera que la estimación de población será el resultado del ajuste de mínimos cuadrados de la regresión para el tiempo t_i .

En las próximas secciones se describen las características generales, el esquema de muestreo, el marco de muestreo y aspectos operativos.

1.2 Diseño de la muestra

Anteriormente se describió de manera general la metodología que se emplea para generar la estimación de la población. A partir de esta sección, se dan mayores elementos sobre cuáles son las variables estratégicas, las características del diseño de muestreo, la conformación del marco del cual se seleccionan las viviendas, así como los estimadores que se emplea en el proceso.

1.2.1 Objetivo general

Generar estimaciones trimestrales oportunas de población, a nivel nacional y por entidad federativa, por medio de un muestreo probabilístico.

1.2.2 Variables de interés

- *Residente habitual*: Persona que habita normalmente en la vivienda donde, por lo general, duerme, prepara sus alimentos, come y se protege del medio ambiente. Incluye a las personas que en el momento de la entrevista se alojan en la vivienda porque no tienen otro lugar para vivir.
- *Sexo*: Condición biológica que distingue a las personas en mujeres y hombres.
- *Edad*: Años cumplidos que tiene la persona desde la fecha de nacimiento hasta el momento de la entrevista.

1.2.3 Desagregación geográfica y periodicidad

Nacional y entidad federativa por trimestre

1.2.4 Informante adecuado en cada vivienda

El informante adecuado es un residente de 15 años y más de edad, que conozca las características de la vivienda y sus residentes.

1.2.5 Esquema de muestreo

La estimación de población se realiza con un esquema de muestreo probabilístico, estratificado, por conglomerados y en dos etapas de selección. En una primera etapa se seleccionan unidades primarias de muestreo (UPM) que corresponden a agrupaciones de manzanas o localidades puntuales. En una segunda etapa, se realiza una selección de viviendas o en el caso de UPM de áreas de nueva construcción se visita la totalidad.

1.2.6 Marco de Muestreo

El Marco de Muestreo es una lista de conglomerados o UPM estratificadas, donde cada UPM se construye por la unión de cierto número de manzanas o localidades, cuya conformación buscó que todas las UPM tuvieran tamaños similares en términos de viviendas. Es un marco actualizado, puesto que, de manera trimestral, incorpora nuevas manzanas o localidades a partir de la información más reciente del Marco Geoestadístico del INEGI, así como de los propios recorridos del personal operativo del marco.

El Marco de Muestreo a su interior contiene dos listas que no se traslapan y que se describen a continuación:

1. Marco Mastreo de Muestreo (MMM): Se refiere a una lista de UPM estratificadas, construidas en 2012 a partir del “Marco geoestadístico 2010 versión 5.0.A (CPV 2010)” simultáneamente con la información estadística del CPV 2010.

De este marco se selecciona una Muestra Maestra (MM) de UPM, que corresponde a la primera etapa de selección de los esquemas de muestreo bietápicos. La MM cuenta con un esquema de actualización que permite:

- a. Identificación de la ruta de acceso y elaboración del croquis (de zona, de UPM, de viviendas).
 - b. Llenado de datos de identificación de la UPM.
 - c. Recorrido del área geográfica, identificación y registro de la situación actual de los elementos geográficos.
 - d. Registro del domicilio geográfico de cada vivienda.
 - e. Determinación y registro de la condición de ocupación de la vivienda.
 - f. Registro del número de integrantes de la vivienda por edad y sexo.
2. Marco de Áreas de Nuevo Crecimiento (MANC): Se refiere a una segunda lista de UPM que se van creando conforme se actualiza el Marco Geoestadístico, en cuanto a nuevas manzanas o localidades. Estas manzanas y localidades son cotejadas con el MMM, a efecto de garantizar que su identificación no forma parte del propio esquema de actualización del MMM y garantizar que los marcos MMM y MANC sean mutuamente excluyentes.

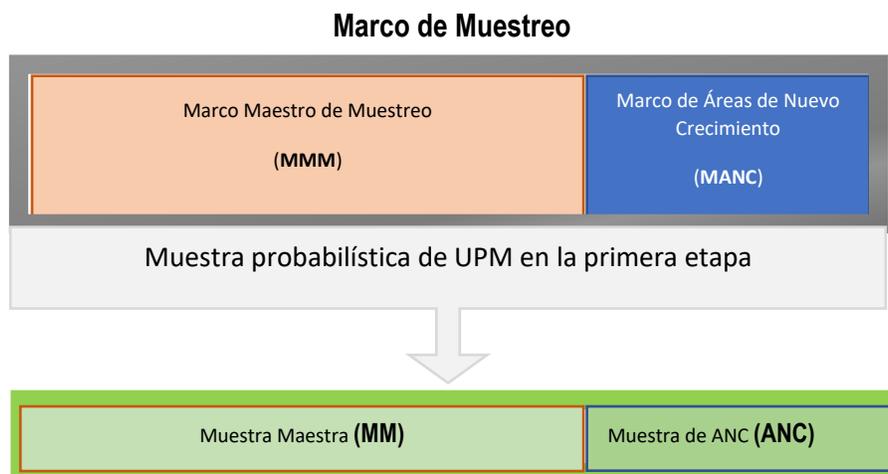
El Marco de Muestreo, bajo la unión de las listas de UPM del MMM y del MANC, contiene la base cartográfica más actualizada en cada trimestre, por lo que es un marco completo.

El MMM y el MANC son mutuamente excluyentes, lo que significa:

$$\text{MMM} \cap \text{MANC} = \emptyset \leftrightarrow (\forall a \in \text{MANC} \rightarrow a \notin \text{MMM}) \wedge (\forall b \in \text{MMM} \rightarrow b \notin \text{ANC})$$

Donde, a representa cualquier manzana o localidad del MANC que no puede ser parte de la actualización del MMM; por su parte, b representa cualquier manzana o localidad de la conformación inicial del MMM o de alguna etapa de actualización, que no puede ser parte de las ANC.

Para ilustrar lo anterior, a continuación se presenta el siguiente diagrama:



En el ANEXO se dan mayores elementos de la conformación del Marco de Muestreo.

1.2.7 Tamaño de muestra para la estimación de población

Con la finalidad de conocer el total de la población, el tamaño de muestra se obtuvo utilizando la expresión correspondiente para el total poblacional:

$$n = \frac{z^2 s^2 deff}{r^2 \bar{X}^2}$$

Donde:

- n = tamaño de la muestra de viviendas.
- z = valor asentado en las tablas estadísticas de la distribución normal estándar para una confianza prefijada.
- s^2 = estimación de la varianza poblacional de la variable de interés.
- \bar{X} = promedio de habitantes por vivienda.
- $deff$ = efecto de diseño definido como el cociente de la varianza en la estimación del diseño utilizado, entre la varianza obtenida considerando un muestreo aleatorio simple para un mismo tamaño de muestra.
- r = margen de error relativo.

Con un tamaño de muestra efectivo a nivel nacional de 245,866 viviendas particulares habitadas por trimestre, se obtienen estimaciones de población con un margen de error relativo de 0.46%, considerando un nivel de confianza de 90%, un efecto de diseño de 4.7, una tasa de no respuesta de 15%, una varianza poblacional de 4.4, para un promedio de habitantes por vivienda de 3.6 personas de acuerdo con el Censo de Población y Vivienda 2020.

En el caso de las entidades federativas, el tamaño de muestra permite estimar el total de la población, con márgenes de error entre 1.9 y 3.3 por ciento.

Dominio	Muestra total	r (%) error relativo
Nacional	245,866	0.46
01 Aguascalientes	6,581	2.7
02 Baja california	10,395	2.4
03 Baja california Sur	5,446	3.3
04 Campeche	8,214	2.5
05 Coahuila de Zaragoza	10,720	2.2
06 Colima	6,503	3.1
07 Chiapas	10,696	1.9
08 Chihuahua	9,962	2.5
09 Ciudad de México	7,661	2.8
10 Durango	7,526	2.5
11 Guanajuato	8,873	2.2
12 Guerrero	6,600	2.7
13 Hidalgo	6,429	2.8
14 Jalisco	7,654	2.6
15 México	9,907	2.2
16 Michoacán de Ocampo	6,323	2.8
17 Morelos	6,042	3.0
18 Nayarit	6,395	3.0
19 Nuevo León	8,273	2.6
20 Oaxaca	6,638	2.7
21 Puebla	8,324	2.3
22 Querétaro	6,439	2.9
23 Quintana Roo	7,495	2.9
24 San Luis Potosí	6,206	2.8
25 Sinaloa	6,350	2.9
26 Sonora	6,542	3.0
27 Tabasco	6,377	2.8
28 Tamaulipas	11,413	2.3
29 Tlaxcala	5,640	2.7
30 Veracruz de Ignacio de la Llave	11,864	2.2
31 Yucatán	6,158	2.9
32 Zacatecas	6,220	2.8

Del tamaño de muestra nacional por trimestre, 19,697 viviendas permiten estimar el total poblacional en las áreas de nuevo crecimiento, para un nivel de confianza de 90%, un margen de error relativo de 5%, un efecto de diseño de 38, una tasa de no respuesta de 15%, una varianza poblacional de 4.20 y un promedio de habitantes por vivienda de 3.23 personas. Por entidades federativas, el margen de error máximo esperado promedio es de 20 por ciento.

1.2.8 Estratificación

La estratificación de las UPM del MMM, inicia con una primera clasificación de las localidades de acuerdo con los siguientes dominios de estudio:

- Urbano alto (UA): localidades mayores de 100 mil habitantes,
- Complemento urbano (CU): localidades entre 2,500 habitantes y 99,999 habitantes y,
- Rural (R): localidades menores de 2,500 habitantes.

Al interior de cada dominio, las UPM nuevamente son clasificadas mediante una estratificación sociodemográfica realizada con métodos estadísticos, con base a un conjunto de variables del CPV 2010, formando cuatro estratos: bajo (1), medio bajo (2), medio alto (3) y alto (4). Esta segunda clasificación tiene el objetivo de reducir la varianza de los estimadores de las encuestas en hogares, al agrupar las UPM de acuerdo con variables poblacionales sobre niveles de educación, acceso a salud, empleo, así como variables relacionadas con el equipamiento y servicios básicos de las viviendas.

1.2.9 Distribución y selección de la muestra

Para la muestra correspondiente al MMM, la distribución de la muestra al interior de cada entidad federativa se realiza de manera proporcional de cada estrato.

La distribución de la muestra de cada periodo t_i y entidad federativa, se hace de manera proporcional con la siguiente relación:

$$\frac{n_{t_i e h}}{n_{t_i e}} = \frac{N_{t_i e h}}{N_{t_i e}}$$

Donde:

$N_{t_i e}$ = total de viviendas particulares de la MM para la e -ésima entidad.

$N_{t_i e h}$ = total de viviendas particulares de la MM en el h -ésimo estrato de la e -ésima entidad.

$n_{t_i e}$ = total de muestra en viviendas de la e -ésima entidad.

$n_{t_i e h}$ = muestra en viviendas en el h -ésimo estrato para la e -ésima entidad.

La selección de la muestra se realiza en dos etapas. En la primera etapa, se seleccionan las UPM de la MM con probabilidad proporcional al tamaño (PPT) de viviendas. La probabilidad de seleccionar la j -ésima UPM, del h -ésimo estrato, de la e -ésima entidad en el periodo t_i es:

$$p_{t_i e h j} = p(UPM_{t_i e h j} \in MM) = \frac{m_{t_i e h} v_{t_i e h j}}{v_{t_i e h}}$$

En una segunda etapa, se emplea submuestreo de viviendas particulares habitadas en cada una de las UPM de la MM. El tamaño de la submuestra en cada UPM es fijo dependiendo del dominio del que se trate; en el dominio UA se seleccionan 10 viviendas particulares habitadas en cada UPM; en el dominio CU y R se seleccionan 20 viviendas particulares habitadas.

La probabilidad de seleccionar la k -ésima vivienda en la j -ésima UPM, del h -ésimo estrato, de la e -ésima entidad, en el periodo t_i es:

$$p_{t_i e h j} = \frac{m_{t_i e h} v_{t_i e h j}}{v_{t_i e h}} \frac{v_{t_i}}{v_{t_i e h j}^*}$$

Su factor de expansión está dado por:

$$f_{t_i e h j} = 1/p_{t_i e h j}$$

Donde:

- $m_{t_i e h}$ = es el total de UPM seleccionadas en el estrato h -ésimo estrato, en la e -ésima entidad.
- $v_{t_i e h j}$ = es el número de viviendas en la j -ésima UPM, del h -ésimo, en la e -ésima entidad de la MM.
- $v_{t_i e h}$ = es el número de viviendas en el h -ésimo estrato, en la e -ésima entidad de la MM.
- $v_{t_i e h j}^*$ = es el número de viviendas en la j -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad al momento de la actualización del listado de viviendas.
- v_{t_i} = es el número de viviendas seleccionadas por UPM, 10 viviendas en el UA y 20 viviendas en CU y R.

La muestra correspondiente al MANC, se encuentra dividida en cinco grupos de rotación. En cada trimestre se actualiza un grupo de UPM que corresponde al 20% de la muestra. Así, cada cinco trimestres se termina de actualizar completamente la muestra. Entre dos trimestres consecutivos se conserva el 80% de la muestra y un 20% de nuevas UPM.

Para cada trimestre t_i , la selección de las UPM se realiza en forma independiente para cada entidad federativa mediante el siguiente procedimiento.

1. De las $N_{t_i e}^*$ UPM se seleccionan $n_{t_i e}^*$ UPM con igual probabilidad.
2. En cada una de las UPM se realiza el registro de cada vivienda con sus residentes habituales.

La probabilidad de selección de la k -ésima UPM al interior de la e -ésima entidad federativa es:

$$P_{t_i e k} = P(UPM_{t_i e k} \in MANC) = \frac{n_{t_i e}^*}{N_{t_i e}^*}$$

Y su factor de expansión está dado por $f_{t_i e k} = \frac{N_{t_i e}^*}{n_{t_i e}^*}$

Donde:

$N_{t_i e}^*$ = número de UPM en la e -ésima entidad del MANC.

$n_{t_i e}^*$ = número de UPM seleccionadas en la e -ésima entidad.

Los factores de expansión antes descritos, requieren de ajustes por no respuesta para compensar el peso de las viviendas habitadas con respuesta respecto a las viviendas habitadas en las que no se pudo captar información al momento de la visita. En el caso de este ejercicio estadístico, se mantienen las viviendas habitadas sin información, a las que se imputa el número de habitantes de alguna vivienda que sí cuenta con información dentro del mismo estrato.

1.2.10 Estimador de la población por diseño de muestreo

Dado que, la población está contenida en los dos marcos de muestreo excluyentes, el MMM y el de ANC; se seleccionan como últimas unidades de muestreo a las viviendas particulares habitadas y se registra el número de ocupantes habituales en estas, entonces la población del país se puede expresar como la suma del número de habitantes en cada marco.

Lo anterior se puede expresar como sigue. Primero, denotemos a la poblacional nacional por U , a la población en el MMM por U_{MMM} y a la población en las ANC por U_{ANC} . Así, por la construcción de ambos marcos, se tiene que $U_{MMM} \cap U_{ANC} = \emptyset$ y $U_{MMM} \cup U_{ANC} = U$. Consecuentemente, la población al tiempo t_i se puede expresar como la suma de la población en ambos marcos de muestreo:

$$\begin{aligned} X_{t_i}^{PT} &= \sum_{k \in U} z_{t_i k}^{PT} \\ &= \sum_{k \in U_{MMM}} z_{t_i k}^{PT_MMM} + \sum_{l \in U_{ANC}} z_{t_i l}^{PT_ANC} \\ &= X_{t_i}^{PT_MMM} + X_{t_i}^{PT_ANC}, \end{aligned}$$

Donde

$$X_{t_i}^{PT_MMM} = \sum_{k \in U_{MMM}} Z_{t_i k}^{PT_MMM}$$

Es el número de habitantes en el MMM y

$$X_{t_i}^{PT_ANC} = \sum_{l \in U_{ANC}} Z_{t_i l}^{PT_ANC},$$

Es el correspondiente en las ANC. Aquí, $Z_{t_i k}^{PT_MMM}$ es el número de ocupantes habituales de la k -ésima vivienda particular habitada en el MMM. Además, $Z_{t_i l}^{PT_ANC}$ representa a la misma variable, pero en las ANC.

De esta forma, si se tienen estimadores de población para ambos marcos de muestreo, entonces un estimador de $X_{t_i}^{PT}$ estará dado por la suma de los estimadores en cada uno de estos.

El estimador del total de la población se calcula mediante el estimador de Horvitz-Thompson, tanto de la muestra seleccionada del MMM, como de la muestra seleccionada mediante el MANC.

Como se mencionó anteriormente, el MMM y MANC son en realidad dos marcos que no se traslapan, por lo que el diseño de muestra en cada uno, permite estimar poblaciones $\hat{X}_{t_i}^{PT_MM}$ y $\hat{X}_{t_i}^{PT_ANC}$ que son complementarias.

$$\hat{X}_{t_i}^{PT_MM} = \sum_{\forall e, h, j} f_{t_i e h j} * x_{t_i e h j}^{PT}$$

$$\hat{X}_{t_i}^{PT_ANC} = \sum_{\forall e, k} f_{t_i e k} * x_{t_i e k}^{PT}$$

Por lo anterior, el estimador de población total $\hat{X}_{t_i}^{PT}$ del trimestre t_i se expresa por medio de:

$$\hat{X}_{t_i}^{PT} = \hat{X}_{t_i}^{PT_MM} + \hat{X}_{t_i}^{PT_ANC}$$

Donde:

- $\hat{X}_{t_i}^{PT_MM}$ = es la estimación de población del trimestre t_i que se obtiene con la MM.
- $\hat{X}_{t_i}^{PT_ANC}$ = es la estimación de población del trimestre t_i que se obtiene con la ANC.
- $x_{t_i e h j}^{PT}$ = es la población de las viviendas en muestra de la k -ésima UPM, en la e -ésima entidad de la MM.
- $f_{t_i e h j}$ = es el factor de expansión de la j -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad, en la MM.

$x_{t_i e k}^{PT}$ = es la población total de la k -ésima UPM, en la e -ésima entidad, en las ANC.

$f_{t_i e k}$ = es el factor de expansión de la k -ésima UPM, en la e -ésima entidad, en las ANC.

1.2.11 Estimadores de errores de muestreo

Para la evaluación trimestral de los errores de muestreo de la estimación de población nacional⁴, se usa el método de Conglomerados Últimos⁵, 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.

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

$$\hat{V}(\hat{X}_{t_i}^{PT}) = \sum_{\forall e, h} \frac{k_{eh}}{k_{eh} - 1} \sum_{j=1}^{k_{eh}} \left(\hat{X}_{ehj} - \frac{1}{k_{eh}} \hat{X}_{eh} \right)^2$$

Donde:

\hat{X}_{ehj} = total ponderado de la población en la j -ésima UPM, en el h -ésimo estrato, en la e -ésima entidad.

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

k_{eh} = número de UPM en el h -ésimo estrato, en la e -ésima entidad.

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

$$DE = \sqrt{\hat{V}(\hat{X}_{t_i}^{PT})} \quad deff = \frac{\hat{V}(\hat{X}_{t_i}^{PT})}{\hat{V}(\hat{X}_{t_i}^{PT})_{MAS}} \quad CV = \frac{\sqrt{\hat{V}(\hat{X}_{t_i}^{PT})}}{\hat{X}_{t_i}^{PT}}$$

Donde:

⁴ De manera análoga, se pueden generar las estimaciones de errores de muestreo para cada una de las entidades federativas. En las expresiones, el índice correspondiente al estrato incluye los estratos de la MM, así como el estrato de las ANC.

⁵ Hansen, M. H. Horwitz, W.N. y Madow, W.G., Sample Survey Methods and Theory, (1953) Vol. 1 pág. 242.

$\hat{X}_{t_i}^{PT}$ = estimador del total de la población.

$\hat{V}(\hat{X}_{t_i}^{PT})_{MAS}$ = estimador de la varianza bajo muestreo aleatorio simple.

$\hat{V}(\hat{X}_{t_i}^{PT})$ = estimador de la varianza bajo el diseño de muestreo descrito en este documento.

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

$$\left(\hat{X}_{t_i}^{PT} - z_{1-\alpha/2} \sqrt{\hat{V}(\hat{X}_{t_i}^{PT})}, \hat{X}_{t_i}^{PT} + z_{1-\alpha/2} \sqrt{\hat{V}(\hat{X}_{t_i}^{PT})} \right)$$

Donde α es el nivel de significancia.

1.3 Suavizamiento de estimaciones

Como se mencionó anteriormente, el método de estimación puntual de población está dividido en dos partes. En la primera parte y para un trimestre determinado, se estima población utilizando únicamente la información de la muestra. En la segunda parte, se utiliza la información histórica de las estimaciones trimestrales para producir un estimador suavizado, utilizando regresión polinomial.

Adicionalmente, se reitera que esta parte de suavizamiento se realiza para proporcionar un estimador puntual de población que no tenga variaciones abruptas entre valores trimestres consecutivos.

Dicho lo anterior, el proceso de estimación por regresión polinomial se puede expresar como sigue. Sean $A = \{(t_i, \hat{X}_{t_i}^{PT}), \forall i\}$, donde $\hat{X}_{t_i}^{PT}$ es la estimación de población para el tiempo t_i , entonces el modelo de regresión polinomial de grado m para el conjunto A , de los puntos $\hat{X}_{t_i}^{PT}$ en función de los periodos t_i , se expresa como:

$$\hat{X}_{t_i}^{PT} = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \dots + \beta_m t_i^m + \varepsilon_i$$

Donde ε_i representa la parte residual del modelo con media cero y varianza constante para toda i .

Los parámetros del anterior modelo se estimaron con el método de mínimos cuadrados. Así, la estimación de población a nivel nacional, $\hat{Y}_{t_i}^{PT}$, se expresa como:

$$\hat{Y}_{t_i}^{PT} = \hat{\beta}_0 + \hat{\beta}_1 t_i + \hat{\beta}_2 t_i^2 + \dots + \hat{\beta}_m t_i^m,$$

Donde $\hat{\beta}_0, \dots, \hat{\beta}_m$ son los valores estimados de los respectivos parámetros por mínimos cuadrados.

En forma general, el proceso consiste en calcular la regresión polinomial para cada una de las 32 entidades federativas adicional a la regresión del dato nacional. Específicamente, el modelo polinomial para la e -ésima entidad sería de la forma:

$$\hat{X}_{et_i}^{PT} = \beta_{e0} + \beta_{e1} t_{ei} + \beta_{e2} t_{ei}^2 + \dots + \beta_{em} t_{ei}^m + \varepsilon_{ei}$$

donde $e = 1, \dots, 32$. Consecuentemente, su respectiva estimación se expresa por:

$$\hat{Y}_{et_i}^{PT} = \hat{\beta}_{e0} + \hat{\beta}_{e1} t_i + \hat{\beta}_{e2} t_i^2 + \dots + \hat{\beta}_{em} t_i^m.$$

A manera de ejemplo, el agregado nacional tiene las siguientes observaciones:

$t_1 = 2018.00$	$\hat{X}_{t_1}^{PT} = 122196865$
$t_2 = 2018.25$	$\hat{X}_{t_2}^{PT} = 122577853$
$t_3 = 2018.50$	$\hat{X}_{t_3}^{PT} = 123097653$
$t_4 = 2018.75$	$\hat{X}_{t_4}^{PT} = 123463070$
$t_5 = 2019.00$	$\hat{X}_{t_5}^{PT} = 123878773$
$t_6 = 2019.25$	$\hat{X}_{t_6}^{PT} = 124303233$
$t_7 = 2019.50$	$\hat{X}_{t_7}^{PT} = 124730109$
$t_8 = 2019.75$	$\hat{X}_{t_8}^{PT} = 124959595$
$t_9 = 2020.00$	$\hat{X}_{t_9}^{PT} = 125368738$
$t_{10} = 2020.25$	$\hat{X}_{t_{10}}^{PT} = 125735095$
$t_{11} = 2020.50$	$\hat{X}_{t_{11}}^{PT} = 126048951$
$t_{12} = 2020.75$	$\hat{X}_{t_{12}}^{PT} = 126372353$
$t_{13} = 2021.00$	$\hat{X}_{t_{13}}^{PT} = 126700767$

La escala para las x_i es grande, lo cual implicará tener coeficientes de regresión $\hat{\beta}_i$ más grandes aún. Para evitar lo anterior se aprovecha que las estimaciones son tomadas en lapsos de tiempo iguales y se plantea un cambio de variable antes de obtener los coeficientes de regresión mediante la siguiente expresión:

$$t'_i = 4(t_i - 2018) + 1 = 4t_i - 8071$$

De tal manera que las estimaciones quedan ahora referenciadas a los puntos:

$$t'_1 = 1, t'_2 = 2, \dots, t'_{13} = 13$$

Así para el conjunto de puntos anterior, la estimación a nivel nacional, $\hat{Y}_{t'_i}^{PT}$, resultante es:

$$\begin{aligned} \hat{Y}_{t'_i}^{PT} = & 122240905.777649 + (448181.084331964)t'_i + \\ & + (-1762.51817969211)t_i'^2 + (-551.374852540513)t_i'^3 + \\ & + (22.5548680049984)t_i'^4 \end{aligned}$$

Ahora, de entre los diferentes grados de la regresión polinomial, se determinó que el modelo de grado 4, obtuvo en promedio coeficientes de determinación más altos respecto a grados inferiores a lo largo de las entidades federativas. Entidades federativas como Colima, Michoacán, Zacatecas, Morelos y Nayarit, incrementan el coeficiente en al menos 2.6 puntos porcentuales bajo la consideración de grado 4. La consideración de grado 6, en la mitad de las entidades federativas no genera una ganancia adicional, pero sí complica la simplicidad que los modelos deben tener por el principio de parsimonia. En la sección 2 se presenta los resultados de la estimación de población bajo el ajuste de la regresión polinomial de grado 4, para cada entidad federativa, nacional y periodo de interés.

Independientemente del grado del polinomio elegido, se determinó que dicho grado fuera el mismo para todas las entidades federativas y a nivel nacional para que **la estimación de población**

nacional obtenida por el modelo polinomial fuera la suma de las estimaciones de las entidades obtenidas por el mismo tipo de modelos. La demostración de esta propiedad se puede ver en el apartado 6 del anexo. En segundo lugar y como se mencionó en el párrafo precedente, los resultados de las estimaciones de población se muestran en la Sección 2. Ahí se podrá verificar la propiedad antes mencionada, salvo efectos de redondeo en cada entidad federativa. Por último, se estimó el modelo polinomial a nivel nacional para poder determinar la precisión de las estimaciones a dicho nivel de agregación, como se describe más adelante.

En lo relacionado con la validación del modelo, en el apartado 5 del anexo se pueden consultar los resultados de las pruebas estadísticas correspondientes, así como visualizar gráficamente los supuestos para cada entidad federativa.

En este mismo orden de ideas, la estimación de los intervalos del 90% de confianza se calcula con la fórmula siguiente:

$$\left(\hat{X}_{et_i}^{PT} - 1.645 \sqrt{\widehat{Var}(\hat{X}_{et_i}^{PT})}, \hat{X}_{et_i}^{PT} + 1.645 \sqrt{\widehat{Var}(\hat{X}_{et_i}^{PT})} \right),$$

Donde $\widehat{Var}(\hat{X}_{et_i}^{PT})$ es la varianza muestral estimada del estimador de población.

El método anterior de estimación proporciona los intervalos de menor longitud y con un 90% de confianza, como se puede ver en la Sección 9.3 de Casella y Berger (1990). Estos intervalos así contruidos contienen el valor real de población con una probabilidad del 90%, por lo que cualquier valor dentro de dichos intervalos, una estimación puntual, tiene la misma posibilidad de concordar con dicho valor real. Convencionalmente, se proporciona el valor central del intervalo, aquí $\hat{X}_{et_i}^{PT}$, como estimación puntual del parámetro de interés, $X_{et_i}^{PT}$. Sin embargo, como se había mencionado anteriormente, **para evitar variaciones pronunciadas entre trimestres en las estimaciones de población, se propone a $\hat{Y}_{et_i}^{PT}$ como estimación puntual de $X_{et_i}^{PT}$.** Aunado a lo anterior, la estimación suavizada $\hat{Y}_{et_i}^{PT}$ está dentro del intervalo de confianza arriba descrito, para cada entidad federativa y tiempos de observación.

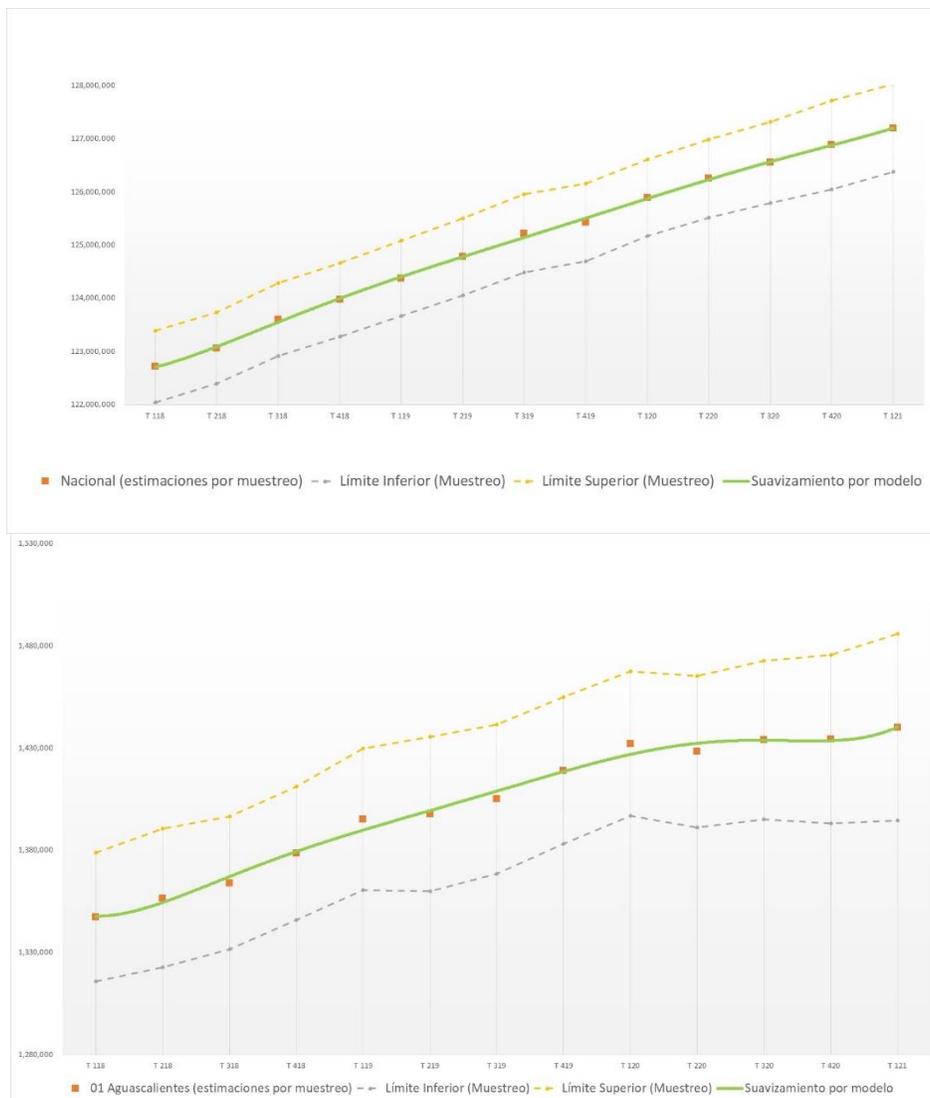
Es conveniente hacer notar que para cada nuevo trimestre, se estima nuevamente el modelo polinomial y se obtiene la estimación de población correspondiente para dicho trimestre, pero sin modificar las estimaciones de períodos anteriores.

Cabe destacar que, **para el presente proyecto, la estimación por intervalo se toma solo bajo el enfoque basado en diseños de muestreo.** Para mayor detalle sobre estos enfoques se puede consultar las Secciones 6.4 y 6.7 de Särndal et al. (1992). Aquí el modelo polinomial solo se utiliza como instrumento para calcular una estimación suavizada de las respectivas estimaciones basadas únicamente en los datos de la muestra.

Si se utilizara la varianza de $\hat{Y}_{et_i}^{PT}$ para calcular los intervalos de confianza, entonces estos intervalos podrían ser diferentes a los intervalos basados en el muestreo, como se muestra en el apartado 7 de los Anexos. Por lo tanto, usar el intervalo de confianza bajo el enfoque asistido por modelos, generaría un cambio en la cobertura $(1 - \alpha)\%$ proporcionada por el diseño de muestreo.

Para ilustrar la metodología hasta aquí descrita se muestra lo siguiente. En el gráfico 1 se presentan las estimaciones trimestrales de población a nivel nacional y para Aguascalientes, así como los intervalos de confianza basados en el diseño de muestreo. La línea continua representa la estimación de población $\hat{Y}_{et_i}^{PT}$ asistida por el modelo, cuyos valores forman parte de los valores plausibles de la estimación por intervalo del diseño de muestreo.

Gráfico 1. Estimación de población basada en diseño de muestreo para el nacional y la entidad de Aguascalientes del primer trimestre de 2018 al primer trimestre de 2021, así como estimación asistida por modelos.



2. Resultados de la estimación de población a nivel nacional y por entidad

A continuación, se presentan las estimaciones de población ajustadas por medio de la regresión polinomial, del primer trimestre de 2018 al segundo trimestre de 2021.

Entidad	T 118	T 218	T 318	T 418	T 119	T 219	T 319	T 419	T 120	T 220	T 320	T 420	T 121	T 221
Estados Unidos Mexicanos	122,686,862	123,126,089	123,556,455	123,975,914	124,382,988	124,776,764	125,156,899	125,523,613	125,877,695	126,220,500	126,553,949	126,880,530	127,203,298	127,489,041
01 Aguascalientes	1,347,408	1,355,783	1,366,343	1,377,941	1,389,630	1,400,664	1,410,496	1,418,783	1,425,379	1,430,339	1,433,921	1,436,580	1,438,975	1,449,082
02 Baja california	3,733,400	3,738,974	3,744,086	3,748,840	3,753,338	3,757,676	3,761,950	3,766,250	3,770,662	3,775,273	3,780,160	3,785,402	3,791,071	3,793,353
03 Baja california Sur	741,590	749,688	756,907	763,487	769,641	775,554	781,383	787,258	793,281	799,526	806,040	812,841	819,921	825,243
04 Campeche	864,861	868,423	874,991	883,378	892,582	901,788	910,369	917,881	924,072	928,871	932,397	934,955	937,036	939,217
05 Coahuila de Zaragoza	3,001,647	3,018,759	3,037,433	3,056,553	3,075,259	3,092,955	3,109,306	3,124,238	3,137,937	3,150,851	3,163,690	3,177,422	3,193,280	3,211,339
06 Colima	709,968	716,854	720,651	722,507	723,369	723,985	724,899	726,457	728,805	731,887	735,445	739,024	741,964	743,516
07 Chiapas	5,266,026	5,298,107	5,325,082	5,350,482	5,376,824	5,405,611	5,437,332	5,471,464	5,506,467	5,539,789	5,567,865	5,586,114	5,588,943	5,609,726
08 Chihuahua	3,636,661	3,660,055	3,674,915	3,684,880	3,692,801	3,700,738	3,709,967	3,720,972	3,733,452	3,746,316	3,757,687	3,764,897	3,764,493	3,769,759
09 Ciudad de México	9,178,198	9,188,553	9,194,164	9,197,236	9,199,506	9,202,250	9,206,277	9,211,931	9,219,090	9,227,169	9,235,117	9,241,418	9,244,090	9,248,530
10 Durango	1,776,242	1,782,123	1,787,137	1,792,116	1,797,646	1,804,068	1,811,478	1,819,729	1,828,426	1,836,931	1,844,360	1,849,586	1,851,234	1,854,349
11 Guanajuato	6,045,089	6,051,455	6,063,499	6,079,131	6,096,628	6,114,635	6,132,162	6,148,585	6,163,648	6,177,460	6,190,499	6,203,607	6,217,994	6,227,057
12 Guerrero	3,486,345	3,492,843	3,499,652	3,506,703	3,513,921	3,521,224	3,528,522	3,535,721	3,542,717	3,549,400	3,555,655	3,561,358	3,566,378	3,571,523
13 Hidalgo	3,006,361	3,023,783	3,035,861	3,044,453	3,051,069	3,056,870	3,062,666	3,068,918	3,075,738	3,082,890	3,089,786	3,095,490	3,098,716	3,108,961
14 Jalisco	8,099,269	8,138,719	8,186,357	8,235,041	8,279,413	8,315,898	8,342,710	8,359,843	8,369,078	8,373,981	8,379,903	8,393,978	8,425,126	8,434,682
15 México	16,791,211	16,820,278	16,840,479	16,857,036	16,873,963	16,894,061	16,918,923	16,948,927	16,983,243	17,019,827	17,055,427	17,085,577	17,104,602	17,139,604
16 Michoacán de Ocampo	4,513,692	4,582,047	4,632,308	4,667,948	4,692,346	4,708,785	4,720,451	4,730,437	4,741,737	4,757,250	4,779,729	4,812,034	4,856,624	4,875,692
17 Morelos	1,946,605	1,952,615	1,958,584	1,963,958	1,968,359	1,971,584	1,973,609	1,974,585	1,974,840	1,974,880	1,975,386	1,977,215	1,981,402	1,986,204
18 Nayarit	1,202,081	1,210,415	1,217,581	1,223,438	1,227,947	1,231,169	1,233,269	1,234,516	1,235,278	1,236,026	1,237,335	1,239,880	1,244,440	1,247,392
19 Nuevo León	5,442,781	5,443,595	5,473,877	5,520,795	5,574,337	5,627,309	5,675,333	5,716,852	5,753,125	5,788,229	5,829,060	5,885,330	5,969,572	5,995,779
20 Oaxaca	3,938,300	3,964,516	3,988,135	4,010,518	4,032,600	4,054,895	4,077,490	4,100,048	4,121,810	4,141,588	4,157,776	4,168,337	4,170,815	4,184,523
21 Puebla	6,441,706	6,454,283	6,466,893	6,480,509	6,495,760	6,512,937	6,531,987	6,552,520	6,573,801	6,594,758	6,613,976	6,629,700	6,639,833	6,651,117
22 Querétaro	2,232,638	2,252,948	2,276,058	2,299,610	2,321,790	2,341,323	2,357,476	2,370,061	2,379,428	2,386,473	2,392,630	2,399,877	2,410,736	2,424,186
23 Quintana Roo	1,790,524	1,796,181	1,803,859	1,812,918	1,822,791	1,832,981	1,843,068	1,852,700	1,861,602	1,869,567	1,876,465	1,882,236	1,886,892	1,891,162
24 San Luis Potosí	2,797,703	2,799,050	2,801,584	2,804,870	2,808,561	2,812,396	2,816,199	2,819,882	2,823,441	2,826,959	2,830,605	2,834,634	2,839,388	2,844,673
25 Sinaloa	2,978,502	2,985,049	2,991,088	2,996,891	3,002,656	3,008,503	3,014,478	3,020,550	3,026,613	3,032,483	3,037,903	3,042,537	3,045,976	3,051,040
26 Sonora	2,883,412	2,888,840	2,897,215	2,907,032	2,917,098	2,926,535	2,934,778	2,941,575	2,946,986	2,951,386	2,955,460	2,960,211	2,966,949	2,971,302
27 Tabasco	2,353,030	2,359,365	2,363,601	2,367,084	2,370,820	2,375,470	2,381,357	2,388,461	2,396,418	2,404,527	2,411,741	2,416,673	2,417,595	2,421,170
28 Tamaulipas	3,458,628	3,469,030	3,480,974	3,492,854	3,503,497	3,512,163	3,518,544	3,522,767	3,525,391	3,527,410	3,530,248	3,535,767	3,546,257	3,553,459
29 Tlaxcala	1,294,205	1,297,846	1,301,782	1,306,225	1,311,282	1,316,954	1,323,140	1,329,635	1,336,129	1,342,208	1,347,356	1,350,951	1,352,266	1,355,230
30 Veracruz de Ignacio de la Llave	7,885,544	7,912,901	7,931,227	7,945,106	7,958,097	7,972,740	7,990,550	8,012,018	8,036,617	8,062,792	8,087,969	8,108,551	8,119,916	8,132,837
31 Yucatán	2,226,132	2,234,855	2,245,732	2,257,989	2,270,932	2,283,948	2,296,502	2,308,141	2,318,490	2,327,254	2,334,219	2,339,248	2,342,288	2,348,418
32 Zacatecas	1,617,102	1,618,154	1,618,397	1,618,384	1,618,527	1,619,097	1,620,227	1,621,909	1,623,995	1,626,198	1,628,090	1,629,102	1,628,528	1,628,916

ANEXOS

1. Diseño del Marco Maestro de Muestreo (MMM) y de la Muestra Maestra (MM)

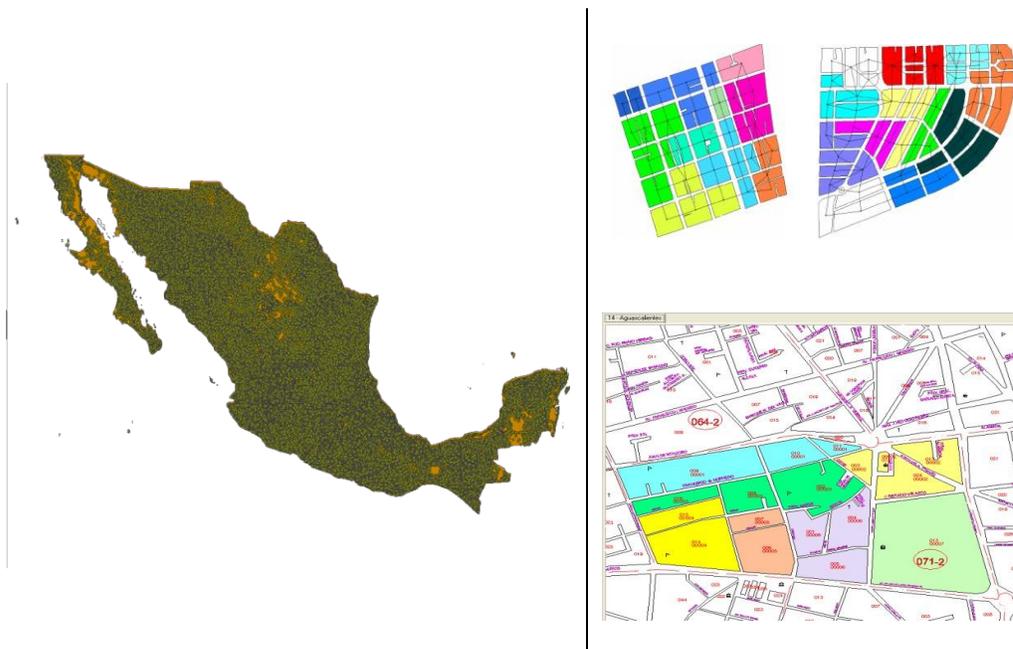
Con la información estadística obtenida del CPV 2010, se elaboró el MMM 2012. Se consideraron las viviendas particulares captadas por el CPV 2010 y con base en ellas se formaron los conglomerados⁶ de viviendas adyacentes (solamente manzanas completas), los cuales se basan en un peso óptimo en viviendas y con un intervalo de tolerancia para tales pesos. Se establecieron dos tamaños de UPM, en las localidades mayores de 100 000 habitantes el tamaño va de un mínimo de 80 viviendas particulares habitadas, hasta un máximo de 160 viviendas habitadas y en el resto se estableció un tamaño mínimo de 160 viviendas hasta un máximo de 300 viviendas habitadas.

El diseño muestral de la MM es probabilístico, por lo cual los resultados obtenidos de la encuesta se generalizan a toda la población. Tiene un diseño muestral en una etapa, de conglomerados estratificados, donde las unidades de observación son las viviendas y sus residentes habituales.

1.1 Formación de UPM

El número de UPM que conforman el MMM 2012 es de 245,279 a nivel nacional. En ellas se cubrió la totalidad de viviendas y de la población captada por el CPV 2010.

Imagen A1. Al lado izquierdo se presentan los 245,279 UPM a nivel nacional, mientras que al lado derecho se hace un acercamiento a ciertas UPM que se distinguen por tener el mismo color en su conformación de manzanas.



⁶ La formación de los conglomerados se basa en el trabajo de "Enfoque de optimización combinatoria para la construcción de marcos de muestreo en hogares"; (Romero & all), a estos conglomerados se les denomina Unidades primarias de Muestreo (UPM).

En el cuadro 1 se presenta el total de UPM formadas para el MMM 2012 agrupadas por entidad federativa, así como tamaño de localidad, de acuerdo con las definiciones:

- Urbano alto (UA): localidades mayores de 100 mil habitantes,
- Complemento urbano (CU): localidades entre 2,500 habitantes y hasta 99,999 habitantes y,
- Rural (R): localidades menores de 2,500 habitantes.

Cuadro A1. Distribución de las UPM por entidad según dominio de estudio

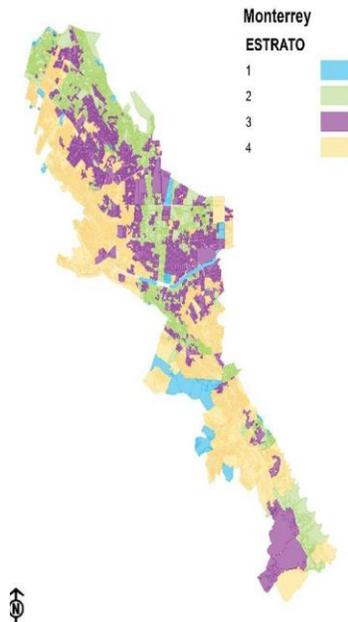
Clave	Nombre	Dominio de estudio			Total
		Urbano Alto	Complemento Urbano	Rural	
01	Aguascalientes	2130	241	261	2 632
02	Baja California	7366	934	417	8 717
03	Baja California Sur	721	707	204	1 632
04	Campeche	1106	416	275	1 797
05	Coahuila de Zaragoza	5970	815	451	7 236
06	Colima	1260	296	124	1 680
07	Chiapas	2584	2067	2581	7 232
08	Chihuahua	8030	1057	951	10 038
09	Ciudad de México	22424	3	55	22 482
10	Durango	2579	963	661	4 203
11	Guanajuato	6291	2870	1864	11 025
12	Guerrero	3004	1906	1684	6 594
13	Hidalgo	1528	1473	1580	4 581
14	Jalisco	10468	4457	1431	16 356
15	México	26403	2946	2271	31 620
16	Michoacán de Ocampo	3085	3079	1791	7 955
17	Morelos	2720	873	438	4 031
18	Nayarit	1137	703	474	2 314
19	Nuevo León	10295	1266	532	12 093
20	Oaxaca	1735	1812	2880	6 427
21	Puebla	5925	2649	1978	10 552
22	Querétaro	2621	455	649	3 725
23	Quintana Roo	3040	436	220	3 696
24	San Luis Potosí	3166	1096	1114	5 376
25	Sinaloa	4098	1247	1005	6 350
26	Sonora	5268	1544	750	7 562
27	Tabasco	1031	1199	1040	3 270
28	Tamaulipas	7522	982	624	9 128
29	Tlaxcala	1711	407	283	2 401
30	Veracruz de Ignacio de la Llave	7212	3443	3748	14 403
31	Yucatán	2971	1183	515	4 669
32	Zacatecas	1188	1511	803	3 502
Nacional		166589	45036	33654	245 279

1.2. Estratificación de UPM

Adicionalmente, se llevó a cabo un proceso de estratificación de las UPM, a efecto de mejorar la eficiencia del diseño y disminuir la variación de las estimaciones de las encuestas. Para lograr esto se

realizó una estratificación geográfica y otra sociodemográfica a partir de un grupo de indicadores relacionados con aspectos de la población como educación, salud y empleo, además de servicios y equipamiento de las viviendas. Para la estratificación se emplearon métodos estadísticos multivariados.

Imagen A2. UPM de Monterrey, según estrato bajo (1), medio bajo (2), medio alto (3) y alto (4).



Vale la pena mencionar que la Muestra Maestra mantiene un esquema de actualización, de manera que cada año y tres meses, todas las UPM de la Muestra Maestra son recorridas y actualizadas. Con la Muestra Maestra, se implementa el diseño estadístico de todas las encuestas en hogares que produce el INEGI.

1.3 Diseño de la Muestra Maestra (MM)

A partir de la construcción del Marco Maestro de Muestreo 2012 se diseñó la Muestra Maestra (MM) para lograr mantener actualizada de forma continua la información de las viviendas particulares dentro de esta muestra. El diseño de la MM consideró y respetó las UPM formadas y la estratificación con que fue construido el MMM.

El tamaño de la MM se calculó en función de las encuestas continuas, además de las encuestas periódicas y especiales que realiza el INEGI año tras año. El número de UPM y sus viviendas particulares son suficientes para obtener las submuestras de cada encuesta en viviendas. Por tanto, la MM tiene una doble función, por un lado, proporcionar las submuestras de las encuestas en viviendas, así como la generación de submuestras para estimar la población del país en cada uno de los trimestres del año.

En la siguiente tabla se presenta la distribución de la MM de UPM actualizadas y el máximo de viviendas particulares observadas por dominio y nacional.

Cuadro A2. Distribución de las UPM y viviendas en la MM por dominio

Dominio	Número de UPM	Número de viviendas particulares máximo
Estados Unidos Mexicanos	22 477	3 127 902
UA	19 159	2 282 220
CU	1 801	445 529
R	1 517	400 153

Imagen A3. Representación del MMM y la selección de UPM que forman la MM.



1.4 Esquema de actualización de la Muestra Maestra

La Muestra Maestra está diseñada con traslapes parciales y paneles de rotación. Las UPM tienen un peso suficiente en viviendas particulares habitadas que permiten un largo periodo de permanencia en la muestra, con el propósito de que se puedan seleccionar viviendas para distintas encuestas en hogares. A esta muestra se le da mantenimiento durante el periodo intercensal, al actualizar cada UPM seleccionada, una vez cada cinco trimestres.

Con el objetivo de registrar las variaciones en cada UPM, cada 5 trimestres se realiza una actualización del listado de viviendas de la UPM, la cual consiste en dar de alta nuevas construcciones de vivienda, se dan de baja aquellas que dejaron de ser vivienda, se actualiza la condición de habitación: viviendas habitadas, deshabitadas, habitadas de uso temporal y no vivienda. De igual manera se actualiza la información de los habitantes de las viviendas habitadas como son el sexo y la edad de los habitantes. Además, durante la actualización se registran los cambios cartográficos dentro de la UPM, como son, fusión de manzanas o localidades, partición de manzanas y variación de los límites de la UPM.

Estas actualizaciones las realiza personal capacitado en cada una de las entidades federativas y su importancia radica en que permiten la actualización de las probabilidades de inclusión y con ellas realizar las estimaciones de la población objetivo de manera confiable.

Para el operativo de campo, se involucra la colaboración de 475 personas que organizan, validan y visitan viviendas a lo largo del territorio nacional. La distribución por cargo a nivel nacional se muestra en el Cuadro A3.

Cuadro A3. Personal operativo de INEGI encargado de la actualización del Marco de Muestreo.

Personal INEGI	Responsable del Marco	Crítico - Verificador	Listador	Total
Total	88	183	204	475

La actualización de una UPM, conlleva el llenado del formato MNV-08 que se presenta en las imágenes A5a y A5b.

Imagen A5b. Interior del Listado de viviendas MNV-08.

PROG. DE VIV.	AGEB	TIPO Y NOMBRE ASENT. H.	MANZANA	TIPO Y NOM. DE VIALIDAD	KM / TRAMO	NÚMERO		DESCRIPCIÓN: Elementos perdurables que identifiquen a la vivienda del resto. NTDG: Datos complementarios de la Norma Técnica sobre Domicilios Geográficos. Nombre: Del(la) Jefe(a) de familia. Hora: Factible de entrevista.	TIPO DE VIVIENDA
						EXT.	INT.		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1		FRACCIONAMIENTO PARAÍSO MAYA I I		CERRADA LAGUNA CHANCANBACAB		3	-	Descripción: CFE 67X9B1 NTDG: Núm. de habitantes: Nombre: Hora: 20:00 H M	2 4
2	S/N					-	Descripción: CFE 67X8B8 NTDG: Núm. de habitantes: 2 Nombre: Juan Antonio Caamal Hora: H 46 M 51	1	
3	7					-	Descripción: NTDG: Núm. de habitantes: Nombre: Placa Fam. Miranda Mendez Hora: 20:00 H M	2 4	
4	9					-	Descripción: NTDG: Núm. de habitantes: Nombre: Hora: 20:00 H M	2 4	
5	11					-	Descripción: NTDG: Núm. de habitantes: Nombre: Hora: H M	2 5	
6	13					-	Descripción: NTDG: Núm. de habitantes: 4 Nombre: Celia de Jesús Olguín López Hora: H 24 6 00 M 23	1	
7	15					-	Descripción: NTDG: Núm. de habitantes: 1 Nombre: Blanca Sánchez Espinoza Hora: H M 48	1	
8	17					-	Descripción: NTDG: Núm. de habitantes: 3 Nombre: Verónica Sanchez Hora: H 99 M 99 99	1	
9	19					-	Descripción: NTDG: Núm. de habitantes: Nombre: Hora: 20:00 H M	2 4	
10	21					-	Descripción: NTDG: Núm. de habitantes: 5 Nombre: Eva Carrasco Hora: H 22 20 17 M 37 17	1	

2

Posterior al trabajo en campo, se lleva a cabo la captura en oficinas estatales, registrando la información levantada en campo de cada UPM de la MM. Se hace uso de equipo de cómputo de escritorio o Laptop y un sistema informático desarrollado en WEB que tiene interacción con otros sistemas informáticos para el mantenimiento permanente, denominado Sistema Integral para la Muestra Maestra (SIMM). Esta plataforma permite la captura, crítica, validación, seguimiento y entrega de la información captada en campo. La captura es enviada para su integración a la Base de Datos de MM.

Otro sistema complementario en el flujo de trabajo es el Módulo Cartográfico (MC-MM) para la digitalización de los croquis de viviendas. Este utiliza como plataforma el Mapa Digital de México que es un Sistema de Información Geográfica (SIG), desarrollado por el INEGI, y que integra información de los elementos geológicos naturales y culturales que conforman el entorno geográfico del país y permite relacionarlos con información estadística. La aplicación permite la inserción del croquis de viviendas del listado en la cartografía digital, mediante “el dibujo” de cada una de las viviendas que integran el croquis, en la manzana o localidad correspondiente, asentada en la cartografía oficial. Para tal efecto, este sistema tiene habilitadas una serie de funciones digitales que permiten captar las particularidades del croquis en todos sus detalles, desde la visualización del área comprendida por la UPM, hasta la inserción de los detalles de interés en las viviendas, como la localización del acceso a las mismas, la distribución de las viviendas dentro de vecindades y su ubicación en edificaciones de más de una planta.

2. Diseño del Marco de Áreas de Nuevo Crecimiento (MANC) y de la Muestra en Áreas de Nuevo Crecimiento (MM)

2.1 Diseño del Marco de Áreas de Nuevo Crecimiento (MANC)

Con el paso del tiempo, el esquema de actualización del MMM es insuficiente para estimar los cambios⁷ en el total de la población, debido a la existencia de nuevas áreas geográficas que no se integran a la propia actualización de la MM, particularmente en espacios geográficos que están fuera de las localidades conurbadas. Para tal efecto, se construye un segundo marco de muestreo, donde se incorporan dichas áreas, que en conjunto, son mutuamente excluyentes respecto al MMM.

El MANC se actualiza de manera trimestral a partir de las áreas geográficas que se incorporan al Marco Geoestadístico (MG) del INEGI,⁸ el cual es un instrumento único, de carácter nacional, donde la cartografía geoestadística, garantiza la cobertura y referencia geográfica de la información estadística. La representación gráfica del Marco Geoestadístico y de la integración territorial del país se concreta en la Cartografía Geoestadística, conformada por mapas, croquis, planos y catálogos en distintos niveles geográficos de representación, en ella se apoyan las actividades de planeación, ejecución, procesamiento y divulgación de resultados de los censos y encuestas que desarrolla el INEGI.

⁷ En el documento del programa para desarrollar la capacidad nacional de efectuar encuestas en hogares, de Naciones Unidas 1987, sección IV, D5; se presentan las estrategias de actualización de los marcos de muestreo. El INEGI sigue la estrategia de “revisión de la muestra”, estableciendo un “estrato de nueva construcción” especial, es decir, un conjunto de áreas donde se sabe que se han producido en gran cantidad nuevas unidades habitacionales posteriores al desarrollo del Marco de Muestreo. La composición del estrato de nueva construcción podrá actualizarse periódicamente, sobre la base de información procedente de los registros oficiales o mediante una inspección directa del listador sobre el terreno.

⁸ <https://www.inegi.org.mx/temas/mg/>

Para la incorporación de nuevas áreas al MANC, se hace un cotejo respecto a la cartografía del CPV 2010, para corroborar que se trata de un área que no fue considerada en la construcción del MMM.

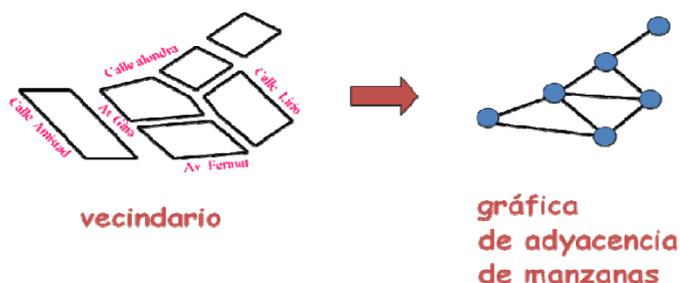
Adicionalmente, el personal operativo del Marco de Muestreo del INEGI mantiene estrategias de campo que le permite hacer recorridos de áreas geográficas, con el objetivo de identificar nuevos puntos geográficos donde pudiera existir población.

Cabe mencionar que el MANC incluye las nuevas áreas geográficas detectadas por el CPV 2020, así como otras que se han generado posterior a la realización de dicho censo y que se van registrando en el Marco Geoestadístico del INEGI. La actualización incluye nuevas manzanas y localidades detectadas al final del primer semestre de 2021, pero mantendrá su actualización en trimestres posteriores.

2.2 Formación de UPM del MANC

Con las manzanas o localidades de las áreas de nuevo crecimiento, se lleva a cabo una nueva conformación de UPM, con base en la cartografía digitalizada de dichas áreas. Al igual que el MMM 2012, el método de conformación de UPM fue el “Enfoque de optimización combinatoria para la construcción de marcos de muestreo en hogares”; (Romero & all), bajo el siguiente procedimiento:

1. En la cartografía más actual, se identifican las áreas de nuevo crecimiento periféricas de las localidades urbanas y las nuevas localidades dispersas en AGEB rural.
2. Se identifican las nuevas manzanas creadas después del 2010 y localidades rurales dispersas en las AGEB rurales.
3. En el caso de áreas amanzanadas, debido a que en la cartografía solo se cuenta con la referencia de manzanas y no se cuenta con el número de viviendas particulares, se formaron las UPM agrupando dos o más manzanas vecinas. Esto último se hizo con la intención de formar UPM con tamaños similares de viviendas. El punto de partida consiste en definir la adyacencia entre manzanas, la cual se puede representar mediante una gráfica o grafo, donde los centros de las manzanas son los nodos del grafo, y las aristas o enlaces se trazarán cuando dos manzanas son adyacentes, como se muestra en las siguientes figuras.



4. Para el caso de localidades dispersas, el procedimiento de formación de UPM se realizó uniendo localidades no amanzanadas vecinas dentro del mismo AGEB y dispersas en la parte rural, el número mínimo que conformarán cada UPM es de 30 localidades.

2.3 Estratificación de UPM del MANC

Las UPM del MANC se integran en un **“estrato de nueva construcción”** debido a que no se tiene información sociodemográfica que permita realizar una estratificación estadística, en cada una de las entidades federativas. Para fines de explotación y a efecto de contar con estadísticas más precisas, se realiza una post estratificación por el número de viviendas particulares habitadas encontradas en el proceso de captación a efecto de generar estadísticas más precisas.

Cuadro A4. Distribución de las UPM del MANC por entidad según el tipo de amanzamiento de estudio.

Clave	Nombre	Tipo		Total
		Amanzanado	Disperso	
1	Aguascalientes	389	357	746
2	Baja California	1,123	1,798	2,921
3	Baja California Sur	1,800	286	2,086
4	Campeche	182	362	544
5	Coahuila de Zaragoza	1,128	955	2,083
6	Colima	566	109	675
7	Chiapas	1,183	2,248	3,431
8	Chihuahua	1,504	1,046	2,550
9	Ciudad de México	31	81	112
10	Durango	1,209	1,137	2,346
11	Guanajuato	2,049	1,827	3,876
12	Guerrero	2,034	500	2,534
13	Hidalgo	866	1,849	2,715
14	Jalisco	3,626	2,099	5,725
15	México	1,236	1,831	3,067
16	Michoacán de Ocampo	1,720	805	2,525
17	Morelos	477	777	1,254
18	Nayarit	777	389	1,166
19	Nuevo León	3,352	1,819	5,171
20	Oaxaca	938	1,835	2,773
21	Puebla	546	1,301	1,847
22	Querétaro	2,006	1,167	3,173
23	Quintana Roo	2,391	823	3,214
24	San Luis Potosí	1,064	1,338	2,402
25	Sinaloa	1,875	777	2,652
26	Sonora	2,678	1,006	3,684
27	Tabasco	337	302	639
28	Tamaulipas	1,726	559	2,285
29	Tlaxcala	208	156	364
30	Veracruz de Ignacio de la Llave	1,599	2,660	4,259
31	Yucatán	1,656	502	2,158
32	Zacatecas	1,117	449	1,566
Nacional		43,393	33,150	76,543

3. Intervalos de confianza de las estimaciones de población

Límite inferior (90%) de las estimaciones de población obtenidas del diseño de muestreo, del primer trimestre de 2018 al segundo trimestre de 2021.

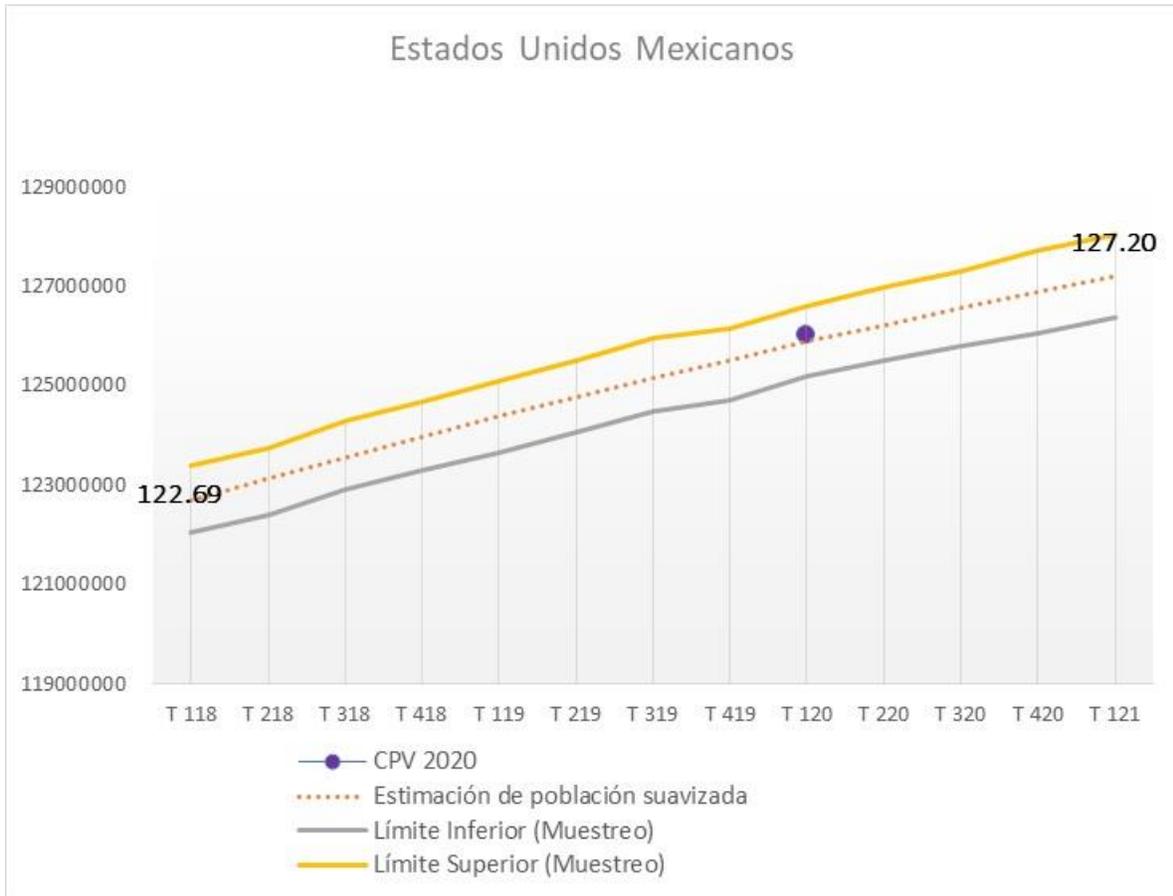
Entidad	T 118	T 218	T 318	T 418	T 119	T 219	T 319	T 419	T 120	T 220	T 320	T 420	T 121	T 221
Estados Unidos Mexicanos	122,036,931	122,391,424	122,919,780	123,284,703	123,664,613	124,051,038	124,487,221	124,691,341	125,174,547	125,510,916	125,786,035	126,051,152	126,380,756	126,664,425
01 Aguascalientes	1,315,828	1,322,603	1,331,540	1,345,787	1,360,631	1,359,924	1,368,622	1,383,029	1,396,823	1,391,186	1,395,025	1,393,064	1,404,850	1,412,330
02 Baja california	3,661,657	3,658,068	3,659,814	3,656,671	3,665,241	3,655,516	3,657,194	3,662,361	3,676,390	3,680,802	3,679,349	3,681,062	3,675,467	3,693,070
03 Baja california Sur	715,396	725,704	729,260	733,020	742,335	744,268	756,284	751,859	759,093	765,968	772,934	766,301	778,455	783,918
04 Campeche	847,888	853,063	857,863	866,019	872,305	882,129	889,922	898,203	906,857	910,133	910,580	911,200	909,991	906,391
05 Coahuila de Zaragoza	2,948,891	2,967,025	2,981,084	2,994,909	3,016,808	3,032,901	3,040,642	3,064,339	3,069,462	3,078,683	3,075,435	3,106,726	3,091,860	3,113,333
06 Colima	696,726	699,561	706,098	709,097	704,570	709,652	703,479	706,189	707,823	707,988	708,628	705,657	710,834	711,676
07 Chiapas	5,127,432	5,138,471	5,196,715	5,194,299	5,218,246	5,201,721	5,256,911	5,246,246	5,320,203	5,365,422	5,383,691	5,390,361	5,385,755	5,405,690
08 Chihuahua	3,576,143	3,583,303	3,612,567	3,620,272	3,625,193	3,624,697	3,628,598	3,644,716	3,664,213	3,674,461	3,679,534	3,693,406	3,696,557	3,702,248
09 Ciudad de México	8,998,422	9,002,104	9,008,107	9,011,403	9,012,077	9,021,913	9,024,327	9,030,338	9,038,513	9,046,690	9,055,129	9,038,810	9,094,409	9,110,591
10 Durango	1,740,903	1,740,458	1,744,165	1,757,392	1,755,529	1,764,311	1,771,437	1,771,186	1,786,857	1,790,891	1,799,563	1,803,357	1,807,673	1,812,428
11 Guanajuato	5,873,176	5,900,601	5,896,107	5,928,060	5,941,562	5,949,829	5,966,688	5,984,235	6,012,668	6,008,103	6,005,246	6,012,014	6,056,950	6,040,065
12 Guerrero	3,384,628	3,391,800	3,396,847	3,402,986	3,410,672	3,418,562	3,423,927	3,443,065	3,454,520	3,448,813	3,449,701	3,450,720	3,431,874	3,452,903
13 Hidalgo	2,906,960	2,941,808	2,936,497	2,937,748	2,943,788	2,938,595	2,964,694	2,965,025	2,960,881	2,985,259	2,988,214	2,986,501	2,959,419	2,988,007
14 Jalisco	7,893,081	7,928,700	7,973,839	8,041,760	8,084,072	8,073,987	8,076,661	8,093,410	8,119,387	8,066,721	8,044,472	7,997,978	7,963,142	8,018,482
15 México	16,438,459	16,489,002	16,540,768	16,488,188	16,509,714	16,567,611	16,605,274	16,602,650	16,697,401	16,700,545	16,708,982	16,729,438	16,799,627	16,802,168
16 Michoacán de Ocampo	4,399,086	4,444,499	4,509,046	4,551,854	4,569,237	4,588,539	4,616,753	4,582,240	4,580,358	4,623,605	4,675,039	4,648,634	4,647,158	4,650,974
17 Morelos	1,903,039	1,911,922	1,902,997	1,909,165	1,910,714	1,915,595	1,922,623	1,927,137	1,933,155	1,930,606	1,928,763	1,926,803	1,914,369	1,929,453
18 Nayarit	1,172,315	1,175,886	1,182,222	1,194,768	1,194,785	1,200,513	1,202,401	1,200,418	1,197,385	1,204,955	1,205,521	1,205,822	1,206,384	1,211,376
19 Nuevo León	5,260,597	5,264,056	5,299,458	5,316,027	5,399,320	5,429,708	5,515,863	5,553,210	5,572,475	5,624,189	5,631,755	5,716,740	5,835,123	5,826,092
20 Oaxaca	3,827,491	3,827,505	3,845,443	3,856,652	3,861,885	3,888,675	3,912,354	3,931,738	3,978,215	3,991,795	4,014,742	4,009,161	4,019,833	4,049,519
21 Puebla	6,239,790	6,255,408	6,249,580	6,291,143	6,293,161	6,306,715	6,323,861	6,324,210	6,338,079	6,384,269	6,426,412	6,435,482	6,476,853	6,478,155
22 Querétaro	2,169,577	2,181,459	2,219,109	2,233,316	2,236,285	2,264,112	2,275,822	2,294,965	2,308,124	2,284,311	2,295,515	2,307,495	2,317,450	2,348,765
23 Quintana Roo	1,730,182	1,730,048	1,737,782	1,752,514	1,761,395	1,761,963	1,763,530	1,765,929	1,778,402	1,778,820	1,801,716	1,807,351	1,811,815	1,816,150
24 San Luis Potosí	2,725,422	2,719,377	2,724,800	2,739,426	2,744,052	2,750,959	2,756,461	2,760,929	2,764,706	2,765,960	2,765,884	2,765,318	2,768,626	2,771,356
25 Sinaloa	2,909,764	2,919,772	2,923,825	2,924,710	2,937,565	2,946,227	2,946,897	2,949,736	2,953,429	2,955,627	2,956,176	2,953,202	2,953,155	2,967,247
26 Sonora	2,806,089	2,815,042	2,823,492	2,823,909	2,843,217	2,842,255	2,855,606	2,849,375	2,863,343	2,861,540	2,865,433	2,842,394	2,881,707	2,878,998
27 Tabasco	2,292,084	2,302,280	2,302,603	2,309,300	2,317,279	2,315,811	2,319,203	2,326,403	2,336,068	2,341,854	2,343,398	2,347,101	2,358,489	2,355,223
28 Tamaulipas	3,388,960	3,407,057	3,412,135	3,426,551	3,436,893	3,447,972	3,450,833	3,453,776	3,451,701	3,461,573	3,447,504	3,454,945	3,480,009	3,474,567
29 Tlaxcala	1,270,030	1,274,317	1,277,674	1,280,930	1,290,476	1,293,640	1,297,144	1,299,295	1,312,253	1,320,804	1,323,762	1,316,221	1,322,448	1,322,150
30 Veracruz de Ignacio de la Llave	7,718,209	7,711,474	7,745,592	7,758,987	7,730,687	7,762,975	7,752,989	7,776,474	7,822,294	7,870,754	7,880,408	7,883,162	7,908,989	7,928,351
31 Yucatán	2,173,638	2,190,810	2,207,078	2,205,591	2,218,122	2,226,402	2,251,222	2,261,088	2,266,715	2,260,910	2,256,129	2,262,132	2,270,937	2,281,912
32 Zacatecas	1,575,679	1,575,885	1,579,875	1,575,513	1,571,941	1,573,180	1,578,819	1,578,300	1,580,196	1,578,026	1,578,416	1,566,571	1,575,566	1,570,681

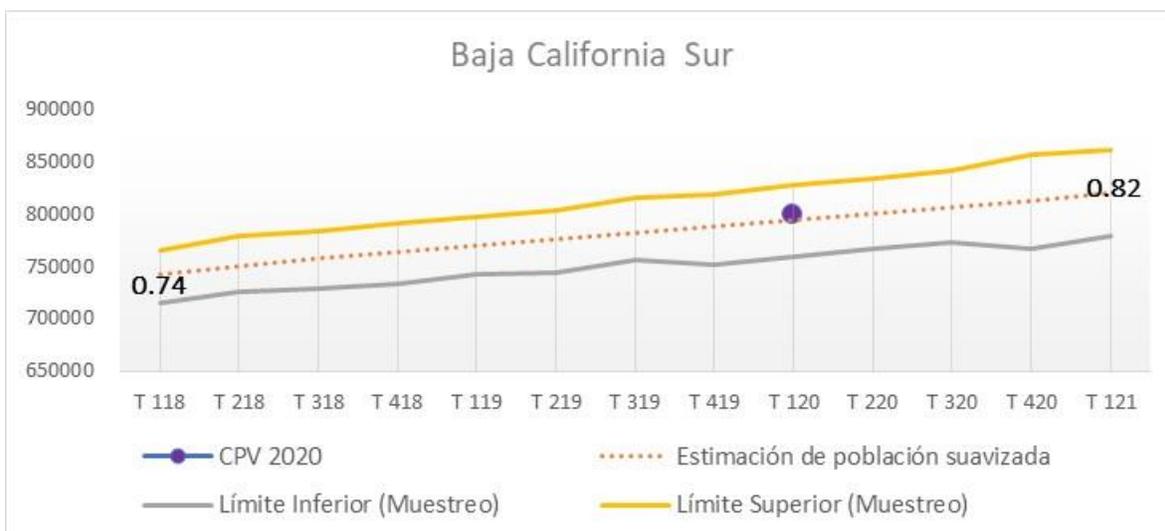
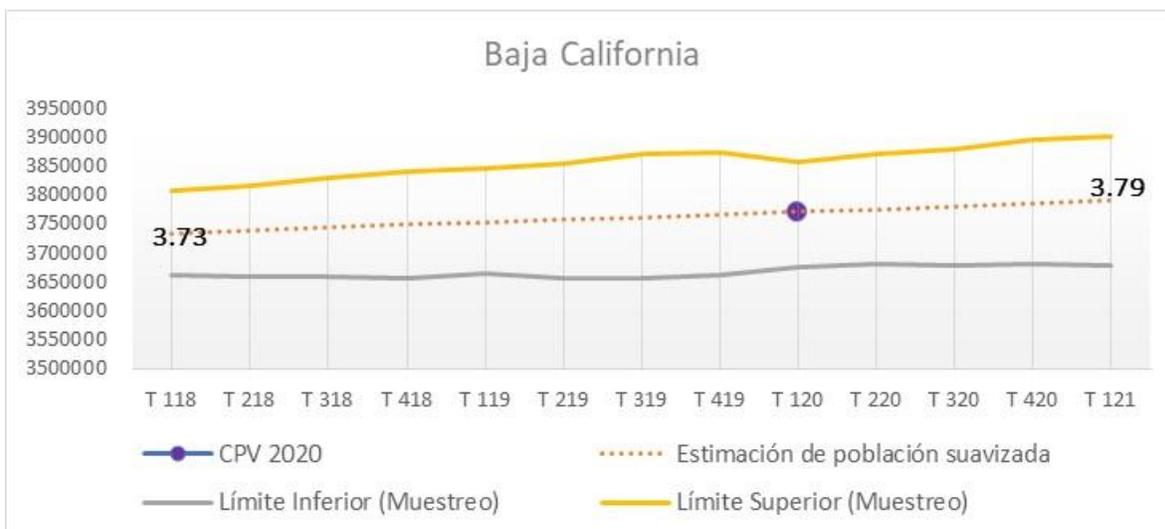
Límite superior (90%) de las estimaciones de población obtenidas del diseño de muestreo, del primer trimestre de 2018 al segundo trimestre de 2021.

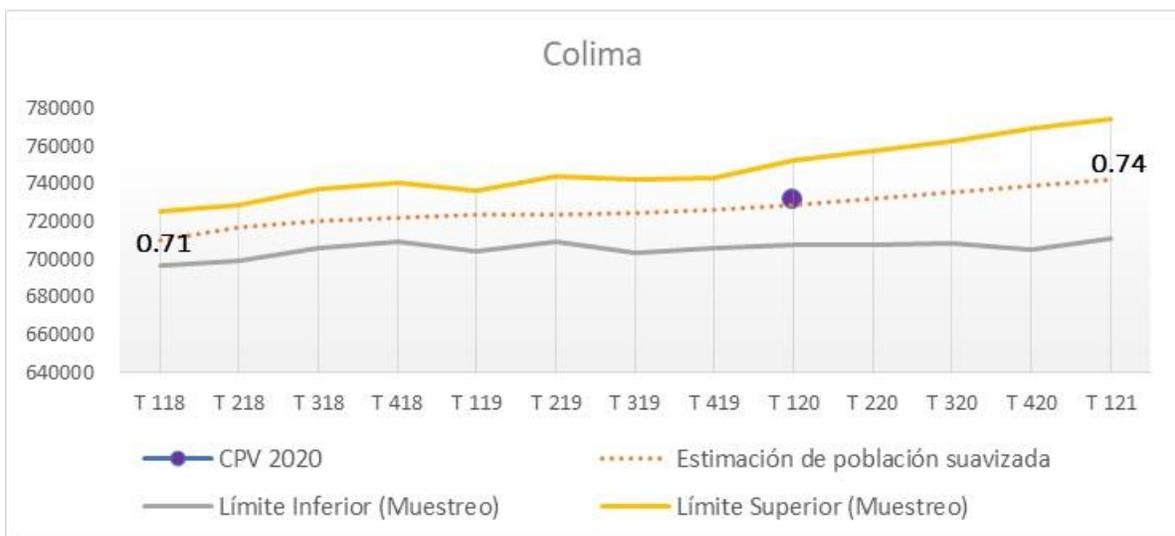
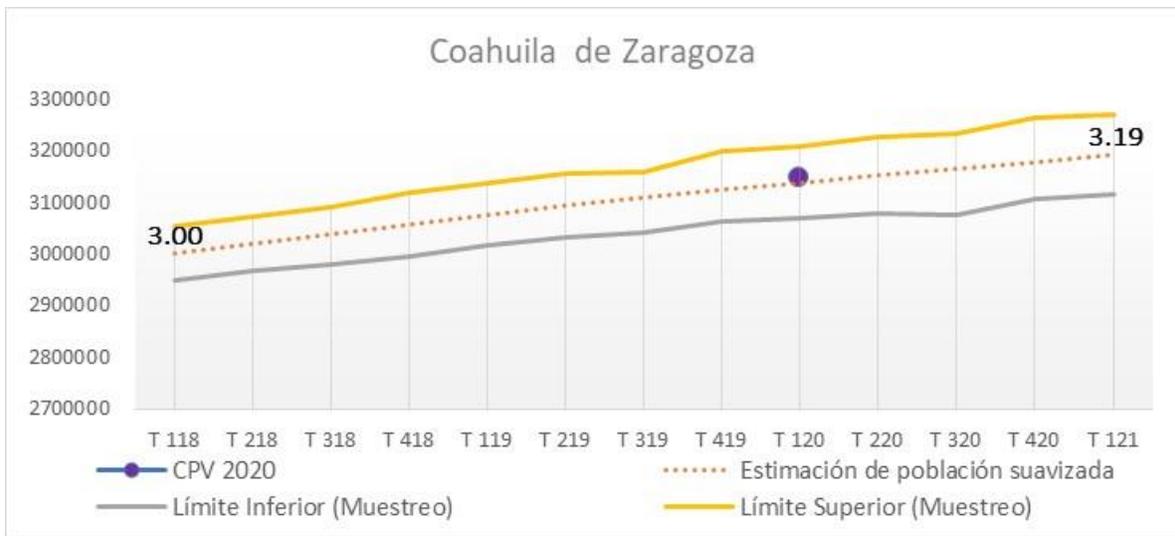
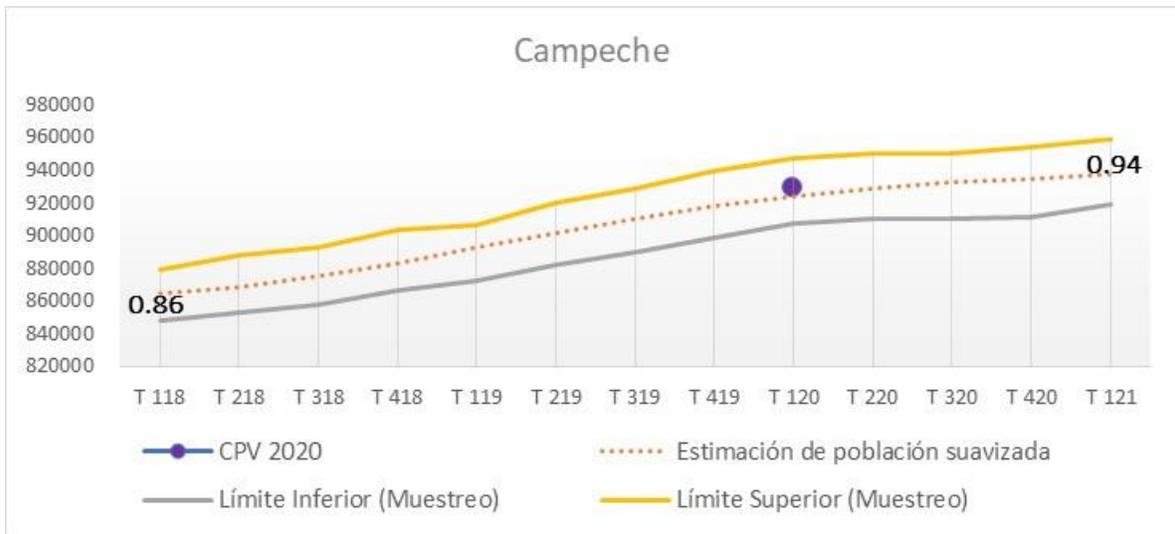
Entidad	T 118	T 218	T 318	T 418	T 119	T 219	T 319	T 419	T 120	T 220	T 320	T 420	T 121	T 221
Estados Unidos Mexicanos	123,385,963	123,729,644	124,283,250	124,667,779	125,084,243	125,507,546	125,953,267	126,158,781	126,607,021	126,990,790	127,313,437	127,711,962	128,050,640	128,293,137
01 Aguascalientes	1,378,730	1,390,597	1,396,528	1,411,189	1,429,751	1,435,364	1,441,528	1,455,037	1,467,419	1,465,391	1,472,845	1,475,712	1,477,744	1,487,190
02 Baja california	3,807,783	3,814,546	3,829,346	3,840,877	3,847,233	3,854,078	3,870,004	3,873,679	3,856,424	3,871,351	3,880,429	3,894,800	3,870,533	3,892,154
03 Baja california Sur	765,780	778,198	783,940	790,900	797,541	802,818	815,608	817,761	826,877	833,920	841,062	856,435	864,437	864,928
04 Campeche	879,182	887,843	892,517	903,847	906,591	919,977	929,120	939,053	947,175	950,091	950,180	954,114	972,229	971,847
05 Coahuila de Zaragoza	3,053,765	3,072,491	3,091,424	3,117,991	3,138,048	3,155,235	3,158,970	3,199,671	3,209,106	3,226,697	3,231,833	3,263,664	3,280,610	3,307,785
06 Colima	725,090	729,109	736,992	740,757	736,732	744,152	742,095	743,593	752,595	757,542	762,968	769,579	774,126	775,394
07 Chiapas	5,409,648	5,436,245	5,478,991	5,498,483	5,551,884	5,576,583	5,649,951	5,649,942	5,714,465	5,743,082	5,744,779	5,756,115	5,812,003	5,827,434
08 Chihuahua	3,705,135	3,714,615	3,747,661	3,761,828	3,760,117	3,771,895	3,779,284	3,799,808	3,808,807	3,823,334	3,832,700	3,832,252	3,826,731	3,842,856
09 Ciudad de México	9,361,522	9,368,758	9,379,479	9,384,615	9,388,595	9,386,773	9,389,885	9,391,118	9,389,757	9,407,763	9,426,061	9,440,530	9,371,829	9,389,385
10 Durango	1,814,281	1,820,684	1,821,811	1,839,756	1,836,129	1,845,507	1,854,737	1,857,318	1,875,129	1,881,282	1,892,081	1,895,733	1,898,599	1,899,606
11 Guanajuato	6,209,314	6,221,757	6,214,087	6,239,974	6,253,890	6,270,083	6,286,534	6,313,037	6,340,042	6,348,952	6,359,602	6,391,904	6,413,790	6,410,861
12 Guerrero	3,584,890	3,597,946	3,607,195	3,607,024	3,614,344	3,618,550	3,634,997	3,630,649	3,639,308	3,645,862	3,659,031	3,669,868	3,699,000	3,689,009
13 Hidalgo	3,099,878	3,120,592	3,130,983	3,140,968	3,161,148	3,165,915	3,183,344	3,173,277	3,167,657	3,192,028	3,194,996	3,199,675	3,247,723	3,234,677
14 Jalisco	8,318,149	8,333,160	8,367,809	8,462,810	8,501,148	8,541,121	8,592,705	8,626,440	8,632,567	8,660,917	8,719,734	8,807,466	8,830,028	8,834,384
15 México	17,122,629	17,165,742	17,215,612	17,165,304	17,199,392	17,220,641	17,268,330	17,251,304	17,337,729	17,348,231	17,364,082	17,420,898	17,511,285	17,486,800
16 Michoacán de Ocampo	4,651,174	4,681,309	4,749,272	4,790,994	4,824,097	4,850,487	4,871,285	4,828,664	4,847,438	4,891,930	4,944,631	4,969,564	5,042,704	5,088,072
17 Morelos	1,988,111	1,997,332	2,014,723	2,017,245	2,020,764	2,029,297	2,031,599	2,014,805	2,020,111	2,019,837	2,020,277	2,027,175	2,045,701	2,041,353
18 Nayarit	1,235,249	1,240,678	1,246,584	1,260,172	1,259,561	1,265,143	1,268,083	1,265,864	1,262,261	1,270,585	1,271,911	1,277,906	1,276,508	1,282,112
19 Nuevo León	5,629,615	5,620,772	5,656,040	5,679,891	5,767,700	5,855,592	5,857,351	5,873,794	5,877,427	5,966,465	6,011,385	6,107,150	6,132,855	6,132,132
20 Oaxaca	4,061,139	4,075,017	4,139,313	4,166,944	4,210,309	4,228,365	4,246,774	4,244,686	4,266,865	4,288,024	4,318,576	4,322,335	4,306,623	4,326,289
21 Puebla	6,642,926	6,660,864	6,665,694	6,684,669	6,695,017	6,719,789	6,743,557	6,779,620	6,801,301	6,807,071	6,808,830	6,820,044	6,828,105	6,827,713
22 Querétaro	2,296,505	2,310,217	2,360,583	2,371,062	2,374,425	2,418,072	2,446,336	2,453,207	2,472,216	2,461,538	2,485,893	2,502,591	2,503,140	2,500,053
23 Quintana Roo	1,854,236	1,856,512	1,866,618	1,879,628	1,888,119	1,902,725	1,914,892	1,945,029	1,950,706	1,944,143	1,960,072	1,960,071	1,957,869	1,964,162
24 San Luis Potosí	2,872,168	2,874,445	2,878,828	2,870,722	2,875,076	2,875,355	2,876,507	2,875,399	2,880,238	2,889,175	2,896,794	2,904,758	2,910,564	2,917,026
25 Sinaloa	3,046,778	3,051,502	3,058,329	3,067,656	3,067,669	3,070,983	3,084,175	3,089,268	3,101,081	3,109,531	3,116,344	3,135,306	3,137,753	3,136,003
26 Sonora	2,959,253	2,963,014	2,978,730	2,981,983	2,991,083	3,009,129	3,019,644	3,025,771	3,039,837	3,039,177	3,044,227	3,075,988	3,086,241	3,061,768
27 Tabasco	2,414,148	2,415,850	2,423,871	2,424,928	2,431,399	2,430,865	2,438,691	2,446,247	2,464,854	2,472,388	2,475,690	2,482,213	2,501,893	2,489,177
28 Tamaulipas	3,524,466	3,541,209	3,543,957	3,558,369	3,565,745	3,579,380	3,590,021	3,593,806	3,588,003	3,607,174	3,602,416	3,620,431	3,618,323	3,626,909
29 Tlaxcala	1,318,758	1,321,803	1,324,322	1,327,652	1,338,342	1,342,954	1,349,866	1,348,853	1,361,285	1,370,210	1,373,546	1,379,213	1,387,920	1,390,486
30 Veracruz de Ignacio de la Llave	8,075,545	8,062,782	8,127,222	8,165,213	8,163,827	8,218,257	8,208,681	8,229,168	8,232,280	8,284,771	8,298,492	8,330,060	8,325,133	8,342,923
31 Yucatán	2,271,616	2,286,214	2,300,576	2,299,665	2,311,810	2,330,216	2,353,986	2,365,230	2,374,171	2,389,630	2,406,123	2,414,284	2,416,799	2,415,082
32 Zacatecas	1,657,889	1,660,197	1,660,011	1,661,399	1,661,613	1,662,426	1,664,907	1,666,950	1,668,448	1,672,353	1,678,826	1,690,151	1,674,296	1,688,079

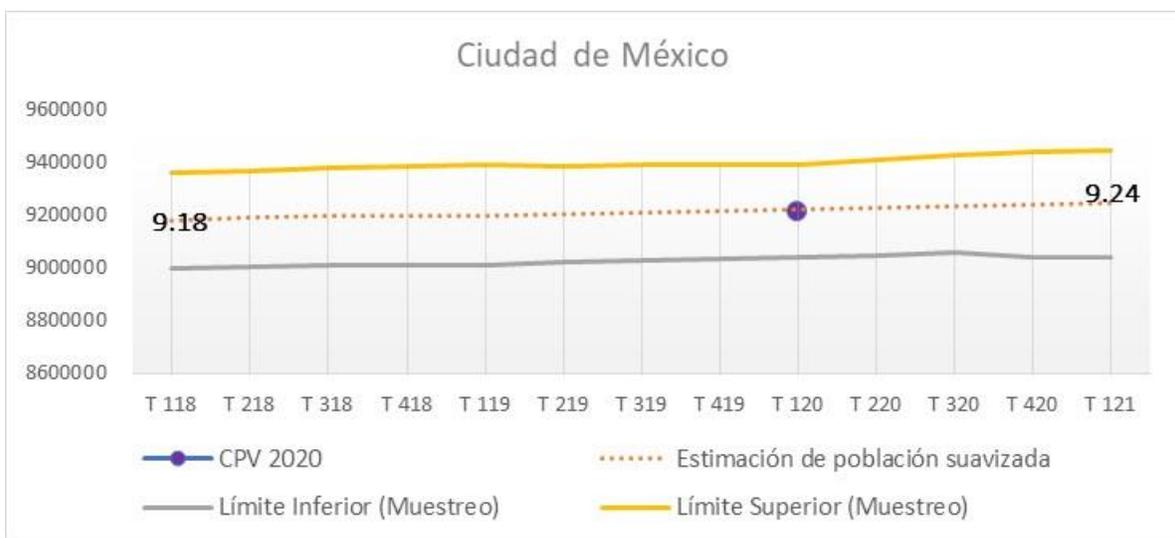
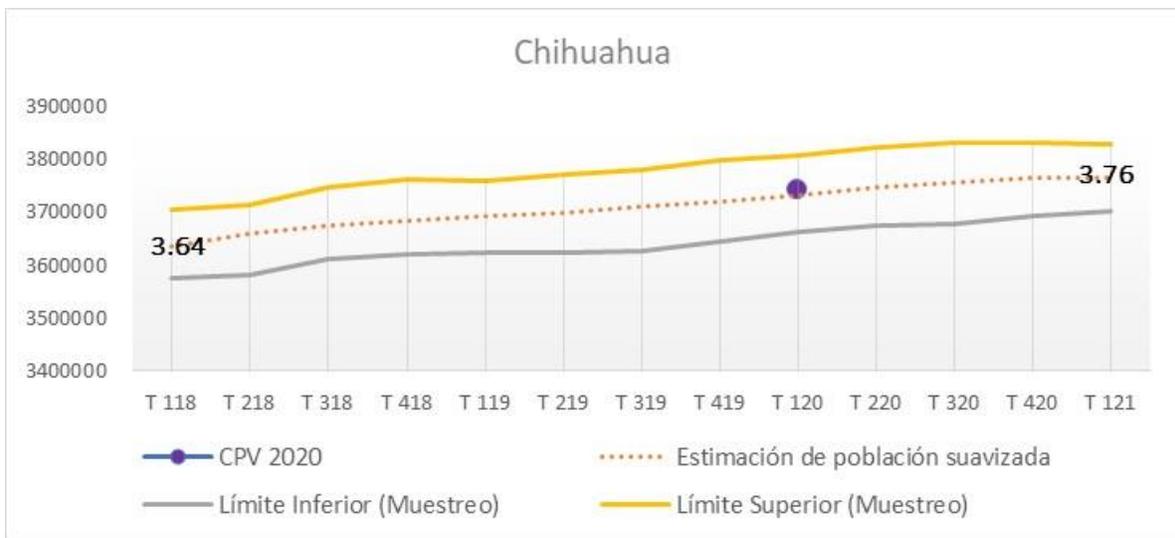
4. Gráficos de estimaciones de población

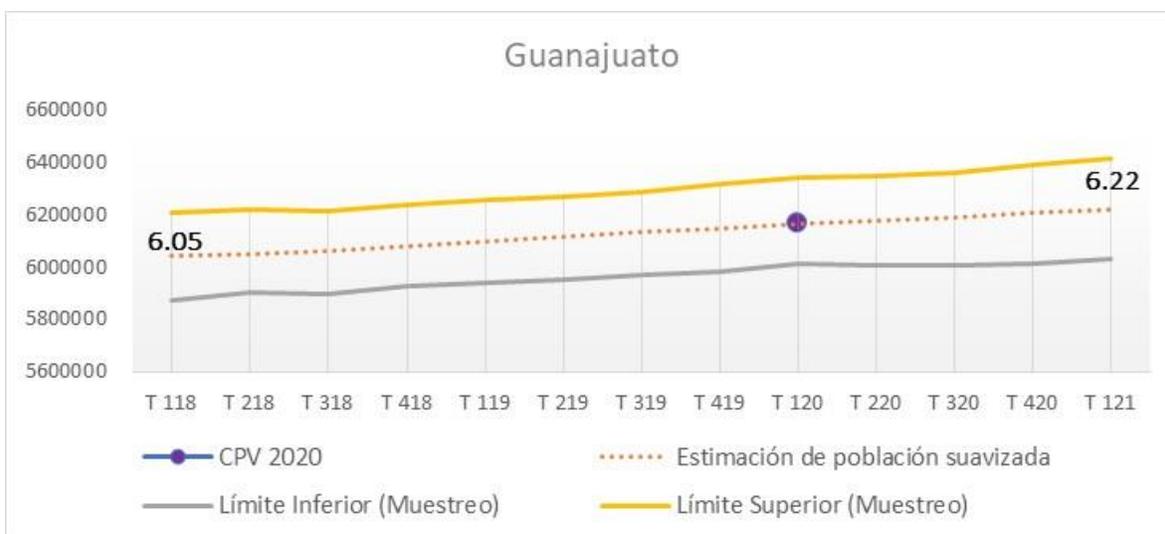
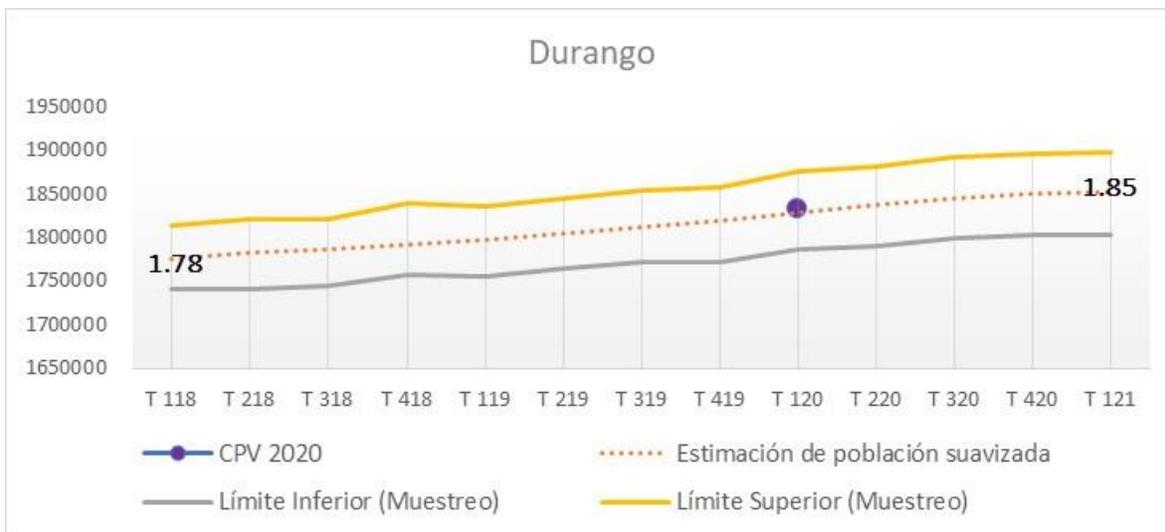
En los siguientes gráficos se presentan las estimaciones de población ajustadas por medio de la regresión polinomial, del primer trimestre de 2018 al primer trimestre de 2021. Los gráficos contienen la estimación por intervalo de confianza a 90 por ciento con base en el diseño de muestreo, información que se puede revisar en el anexo anterior.

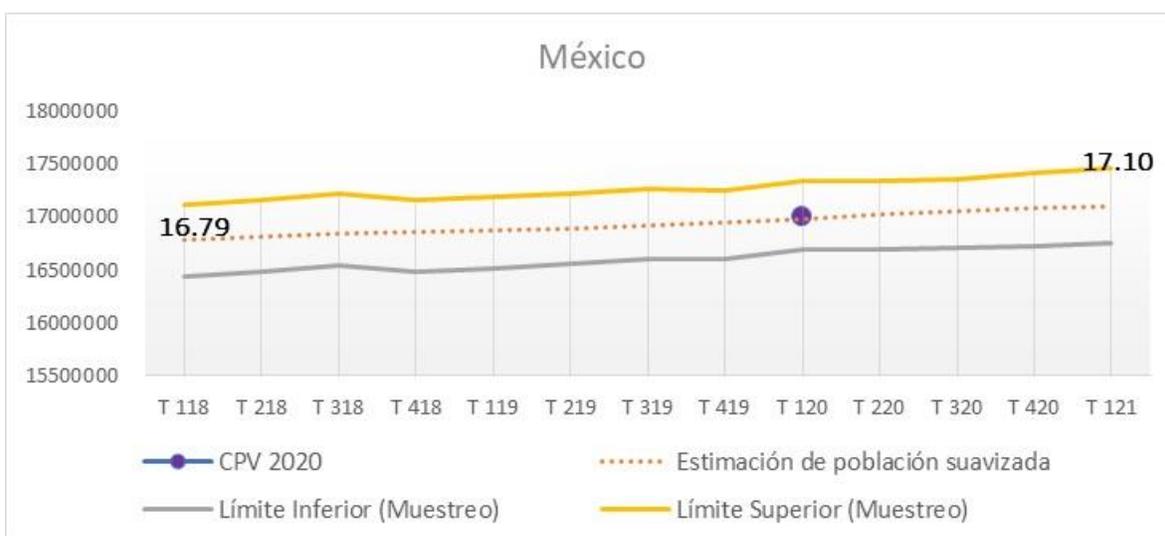
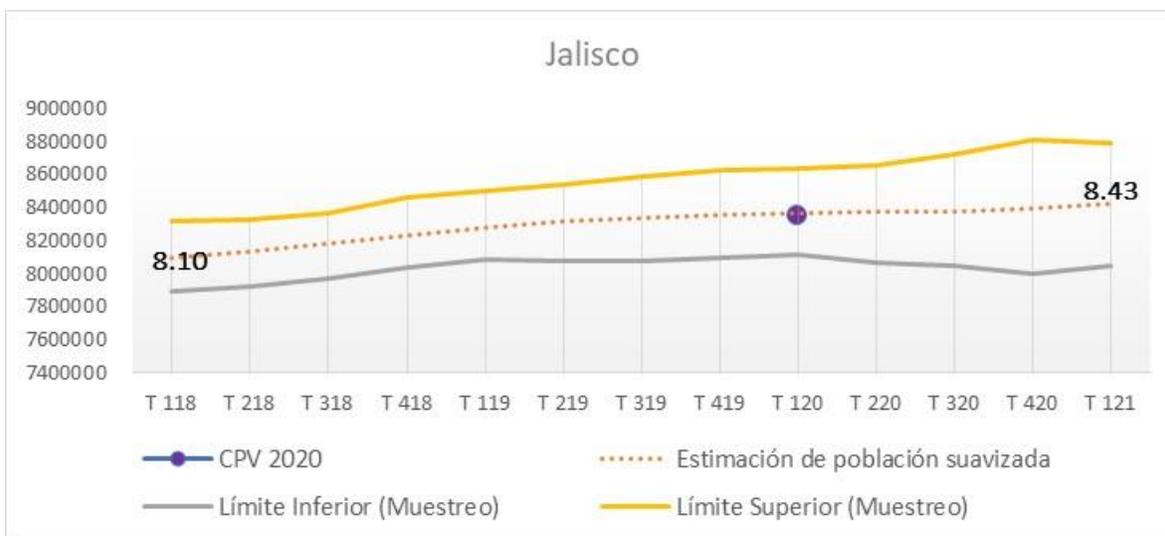


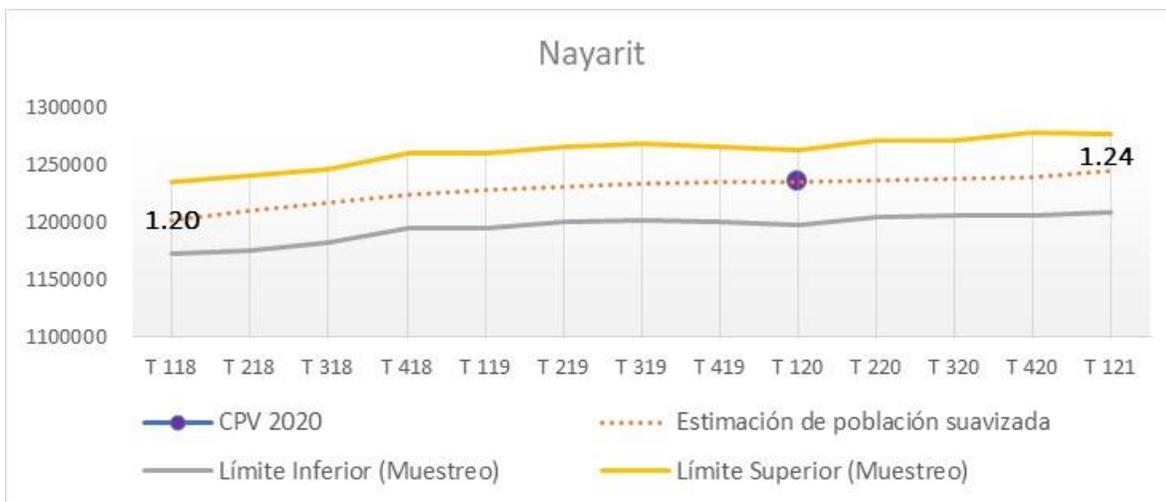
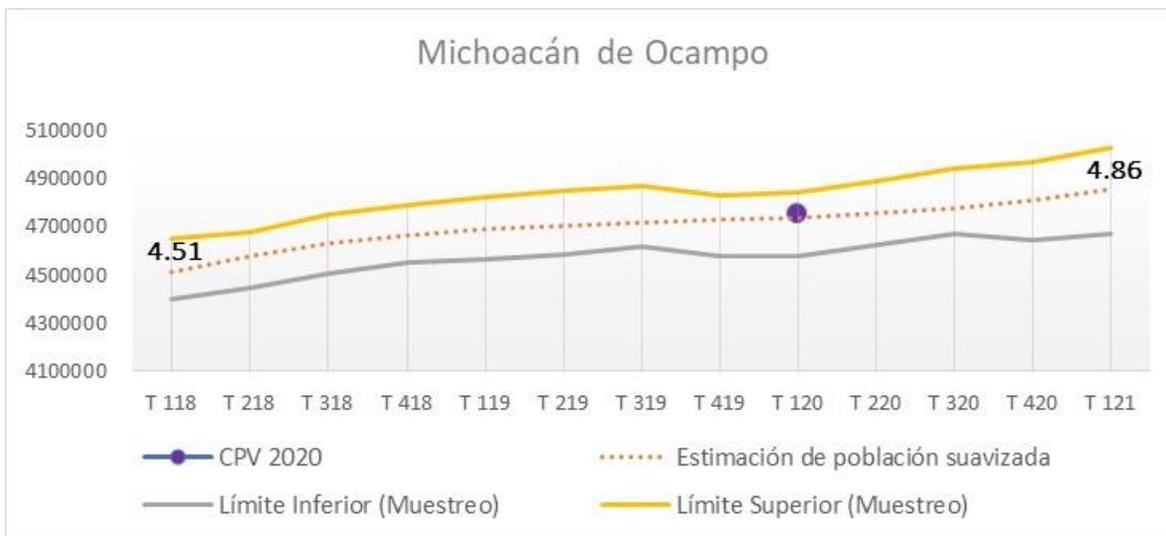


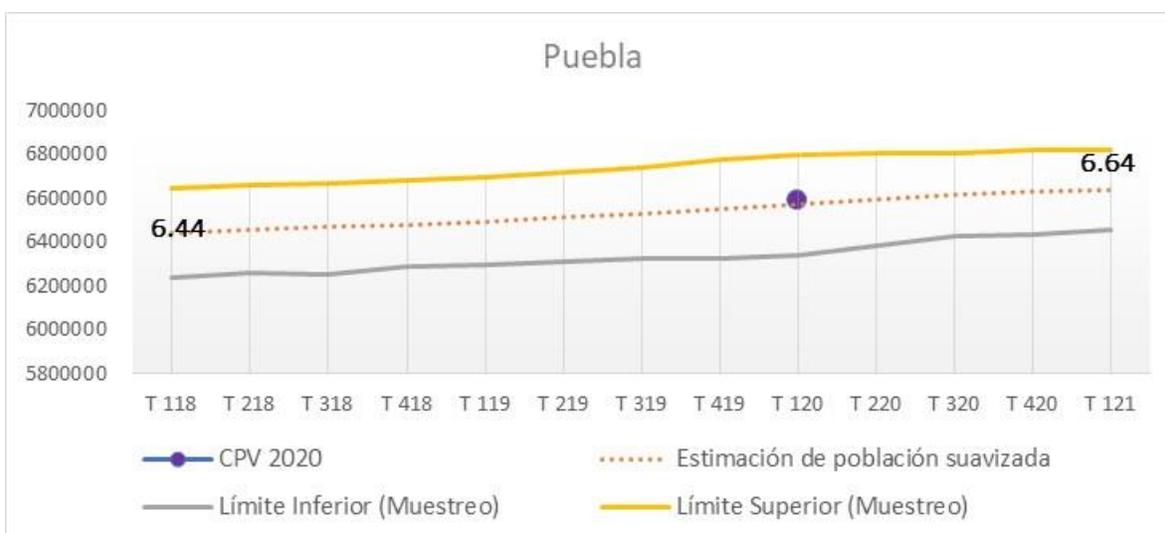
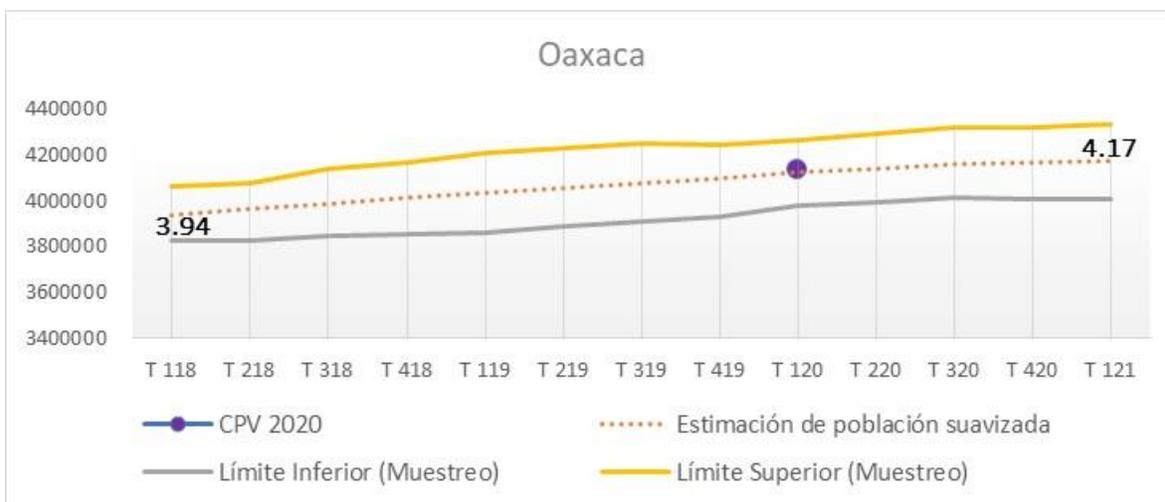


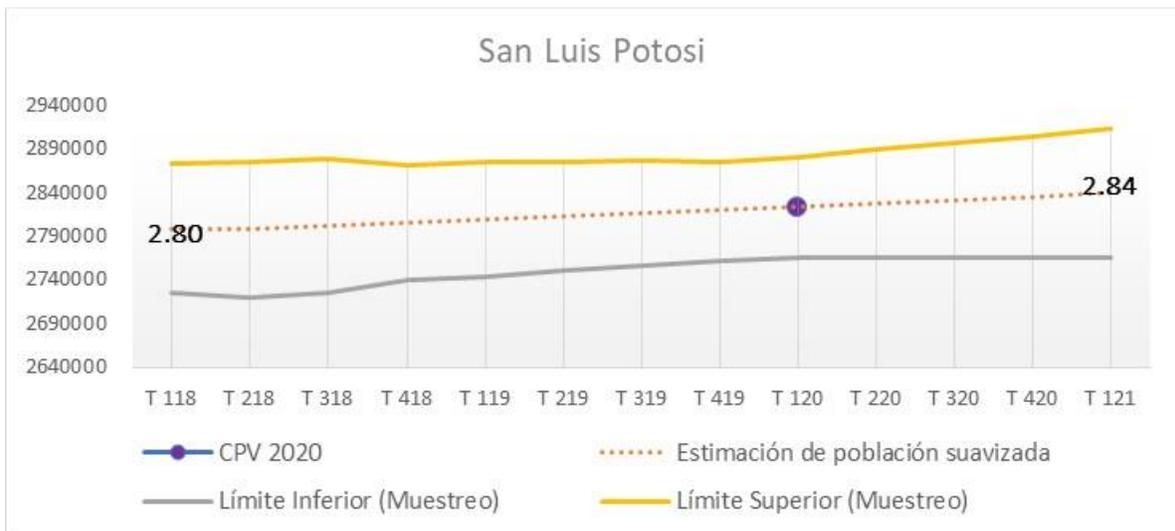
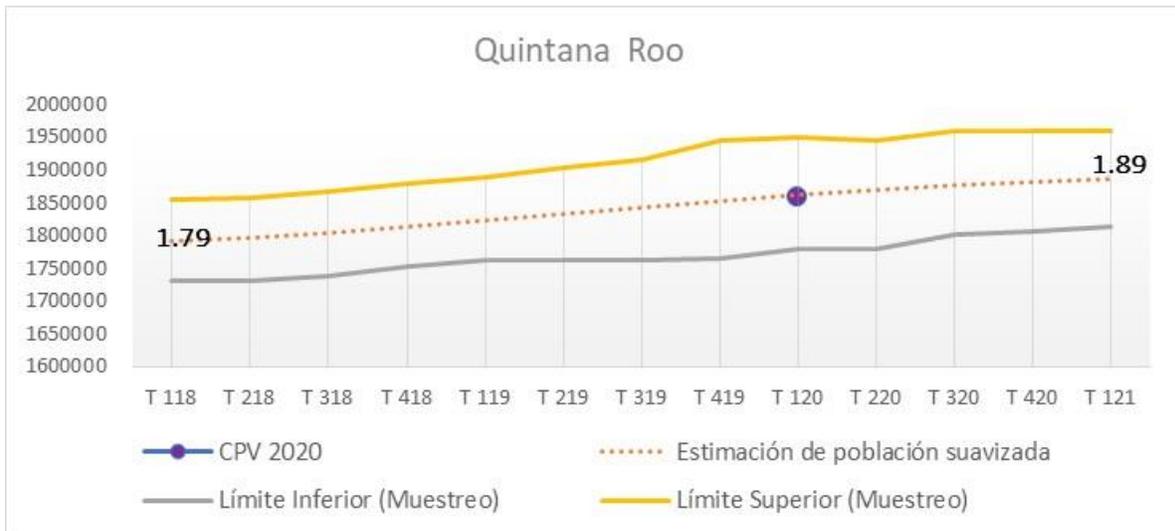
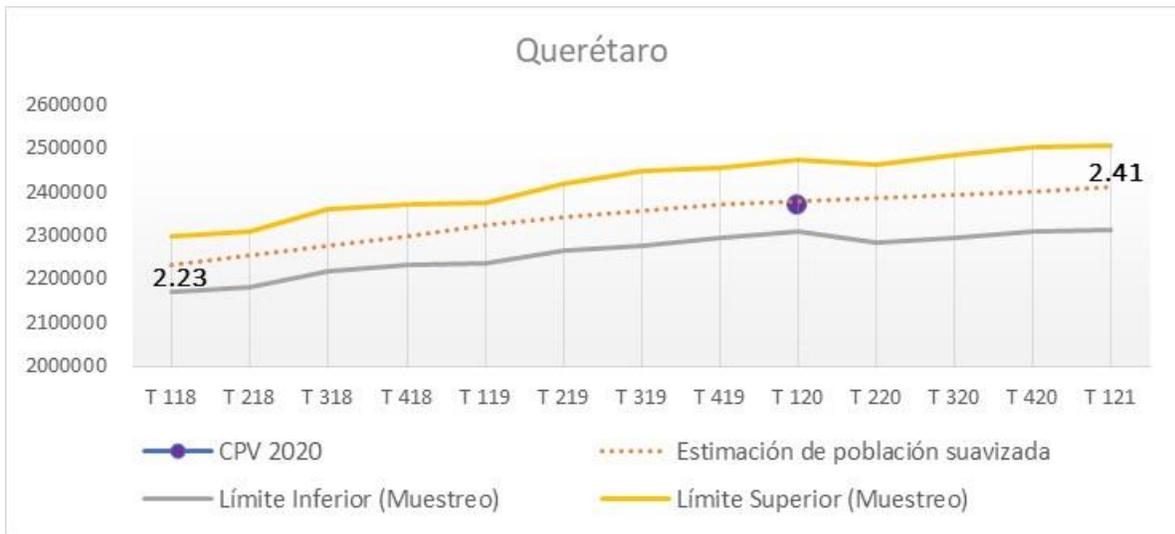


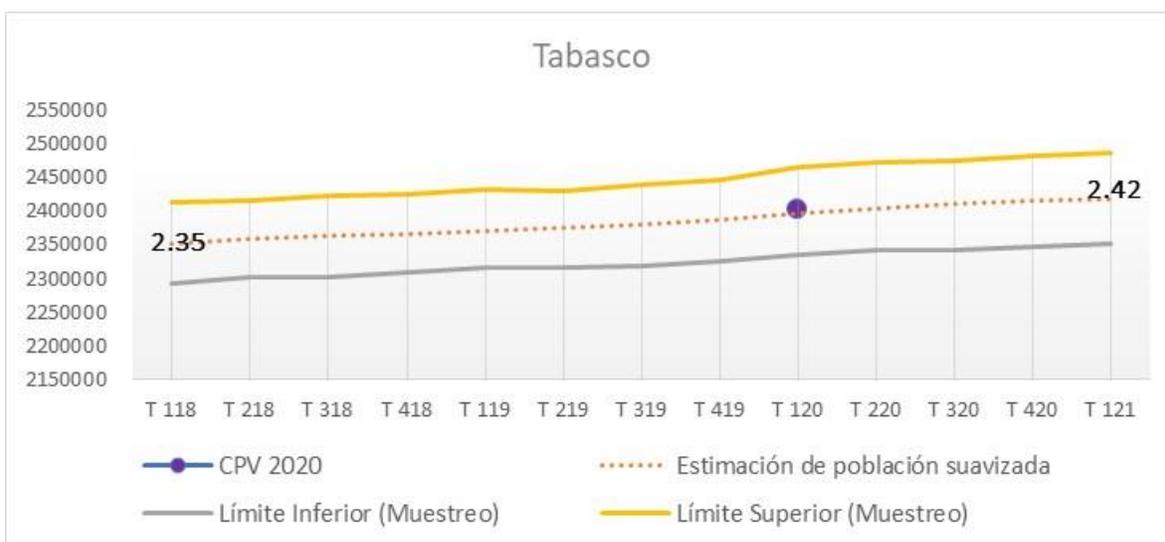
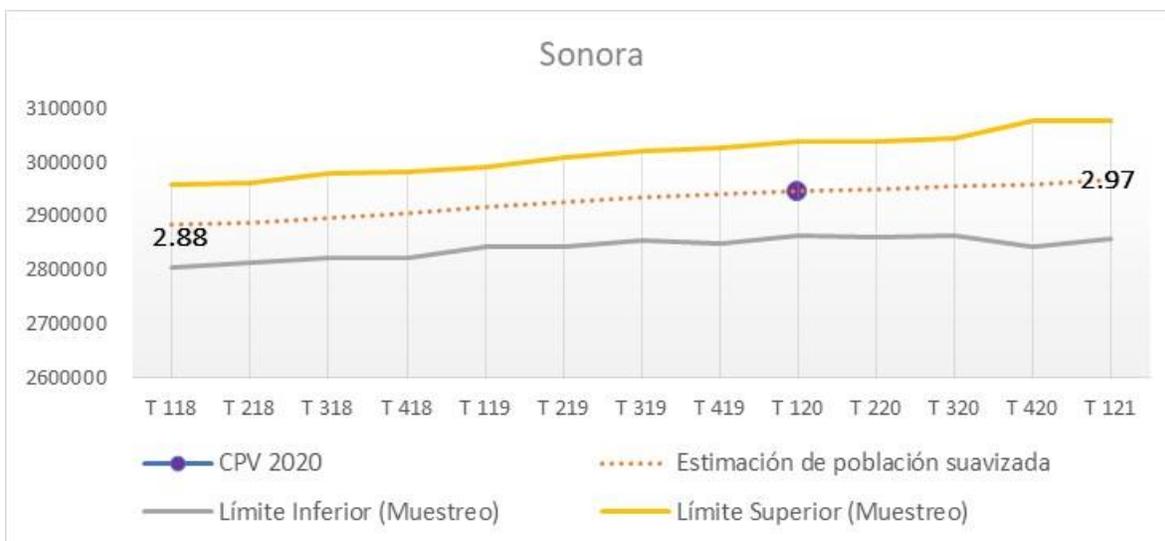
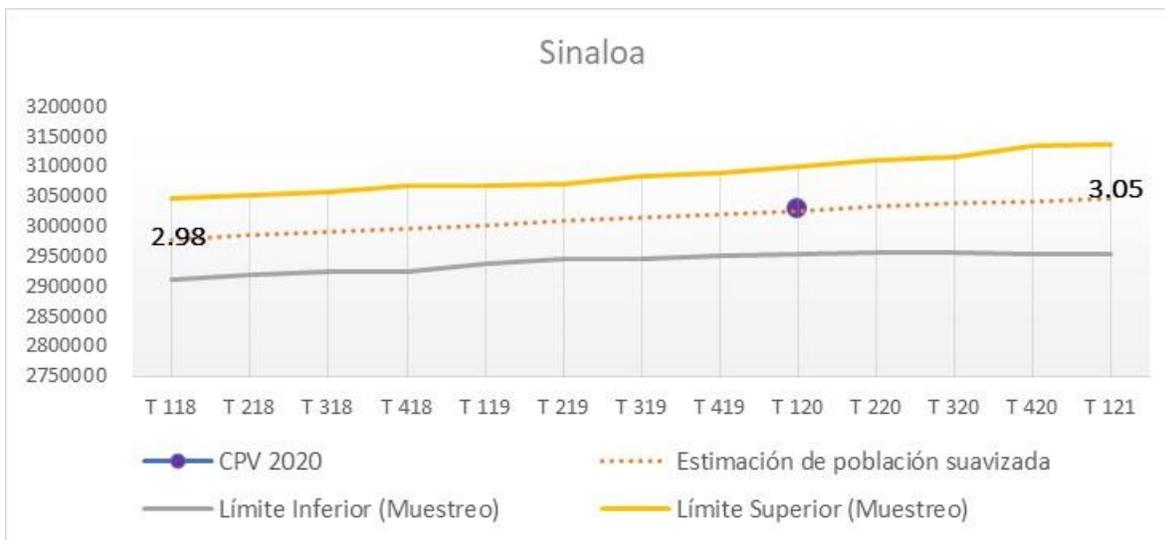


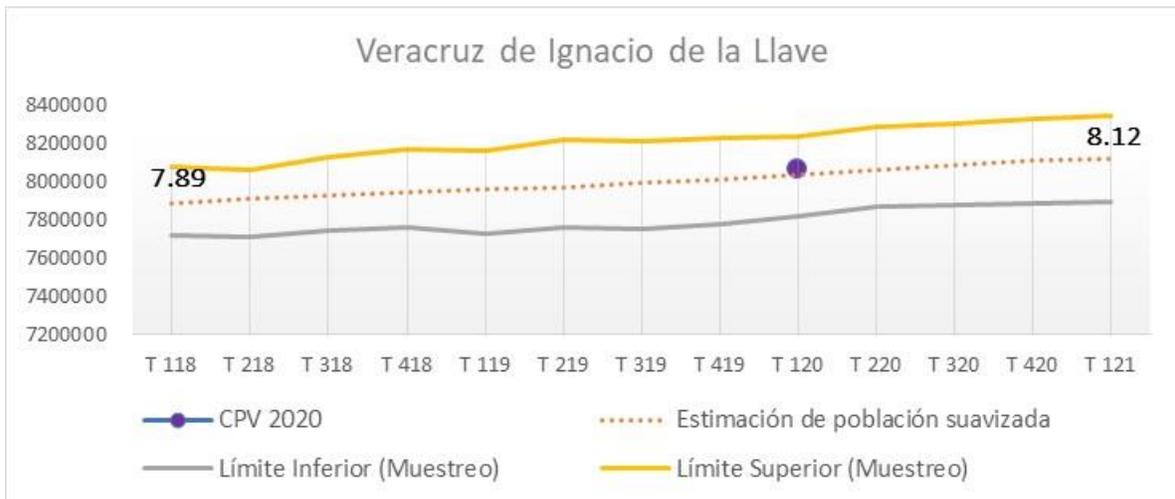
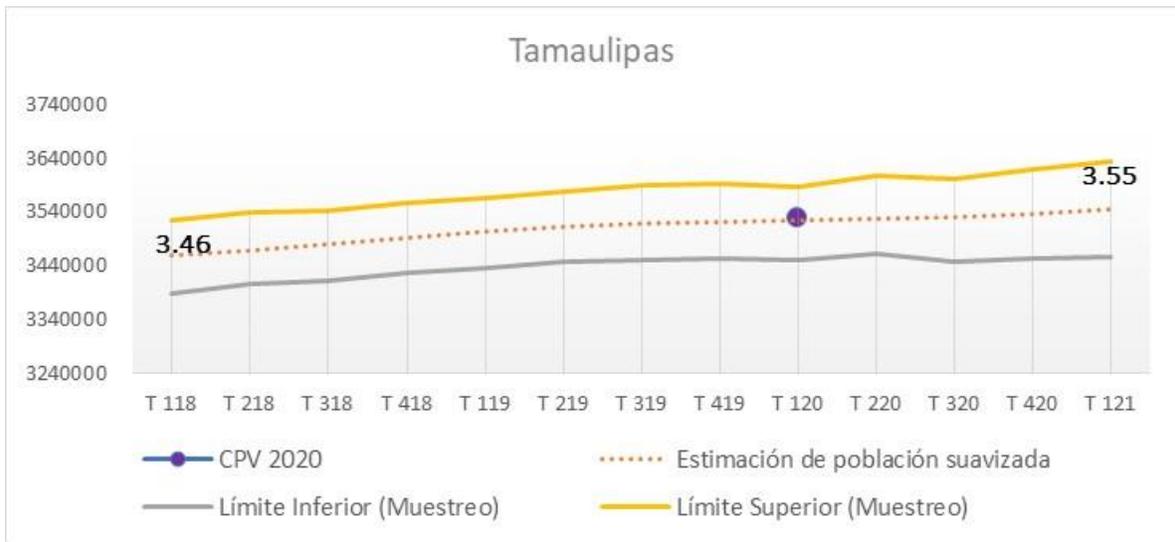


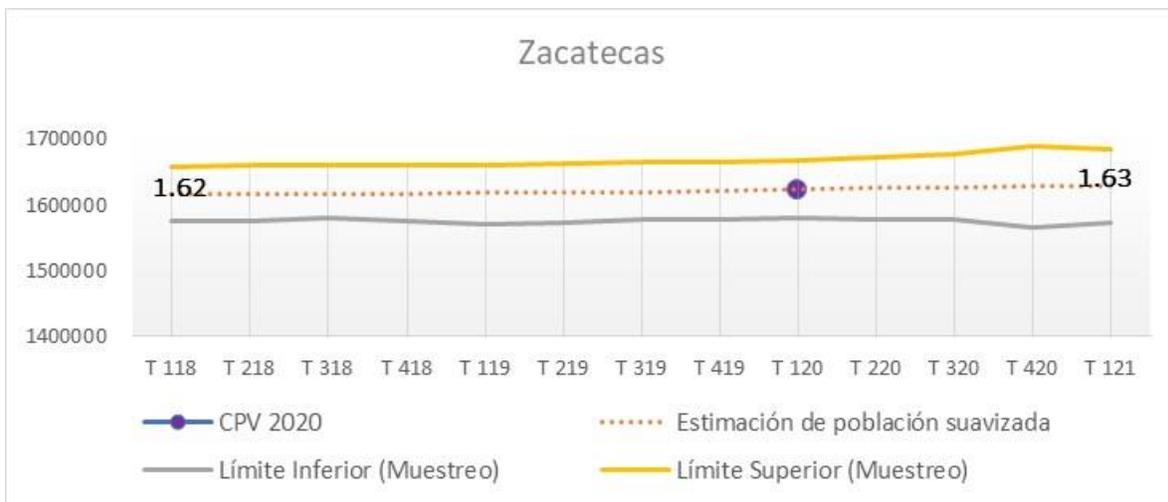
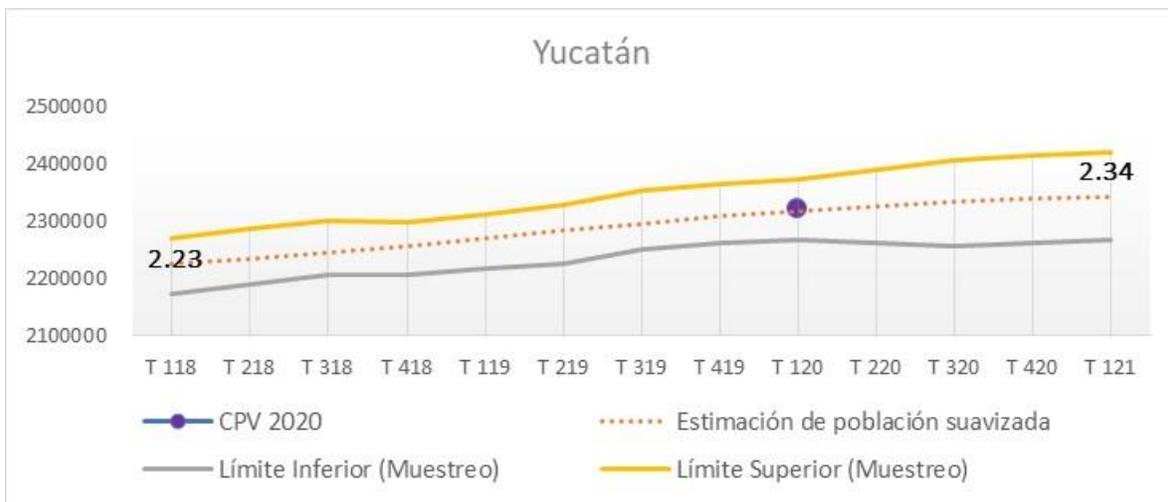












Durante el periodo de la pandemia por la COVID-19, los operativos de campo que captan información cara a cara en las viviendas, han estado sujetos a la disponibilidad del personal, de acuerdo a su condición de vulnerabilidad, o resultado de contagios. Lo anterior se traduce en posibles reducciones de muestra que derivan en incrementos de varianza en algunas entidades federativas.

5. Diagnóstico del modelo

Estadísticos de Coeficiente de Determinación (R^2) y Coeficiente de Determinación Ajustado (R^2 ajustado) de la regresión polinomial de las estimaciones de población, según entidad federativa y dominio nacional.

De entre diferentes grados de la regresión polinomial, se determinó que el modelo de grado 4, obtuvo en promedio coeficientes de determinación más altos respecto a grados inferiores a lo largo de las entidades federativas. Entidades federativas como Colima, Michoacán, Zacatecas, Morelos y Nayarit, incrementan el coeficiente en al menos 2.6 puntos porcentuales bajo la consideración de grado 4. La consideración de grado 6, en la mitad de las entidades federativas no genera una ganancia adicional, pero sí complica la simplicidad que los modelos deben tener por el principio de parsimonia.

ENTIDAD	Grado 2		Grado 4		Grado 6	
	R ²	R ² ajustado	R ²	R ² ajustado	R ²	R ² ajustado
Estados Unidos Mexicanos	0.999	0.999	0.999	0.999	0.999	0.999
01 Aguascalientes	0.986	0.983	0.989	0.984	0.992	0.983
02 Baja California	0.984	0.981	0.985	0.978	0.989	0.978
03 Baja California Sur	0.993	0.992	0.994	0.991	0.995	0.990
04 Campeche	0.985	0.982	0.995	0.992	0.999	0.998
05 Coahuila de Zaragoza	0.992	0.991	0.993	0.989	0.993	0.986
06 Colima	0.909	0.890	0.958	0.936	0.971	0.942
07 Chiapas	0.981	0.977	0.987	0.981	0.990	0.981
08 Chihuahua	0.976	0.971	0.987	0.980	0.995	0.991
09 Ciudad de México	0.970	0.964	0.985	0.978	0.991	0.983
10 Durango	0.980	0.976	0.987	0.981	0.990	0.980
11 Guanajuato	0.985	0.982	0.988	0.983	0.991	0.983
12 Guerrero	0.993	0.992	0.994	0.991	0.998	0.996
13 Hidalgo	0.949	0.939	0.958	0.937	0.964	0.929
14 Jalisco	0.979	0.975	0.991	0.986	0.994	0.988
15 México	0.956	0.947	0.960	0.940	0.974	0.948
16 Michoacán de Ocampo	0.917	0.901	0.967	0.950	0.981	0.963
17 Morelos	0.935	0.922	0.968	0.952	0.974	0.948
18 Nayarit	0.923	0.907	0.955	0.933	0.977	0.953
19 Nuevo León	0.985	0.982	0.992	0.989	0.997	0.993
20 Oaxaca	0.991	0.989	0.994	0.990	0.997	0.994
21 Puebla	0.993	0.991	0.997	0.995	0.997	0.993
22 Querétaro	0.977	0.972	0.979	0.969	0.981	0.961
23 Quintana Roo	0.985	0.982	0.989	0.984	0.991	0.981
24 San Luis Potosí	0.993	0.991	0.995	0.992	0.999	0.998
25 Sinaloa	0.998	0.997	0.998	0.998	0.999	0.998
26 Sonora	0.987	0.985	0.991	0.987	0.992	0.985
27 Tabasco	0.980	0.976	0.990	0.985	0.995	0.990
28 Tamaulipas	0.967	0.960	0.982	0.973	0.984	0.969
29 Tlaxcala	0.978	0.974	0.986	0.980	0.989	0.977
30 Veracruz de Ignacio de la Llave	0.967	0.960	0.972	0.958	0.985	0.971
31 Yucatán	0.982	0.978	0.987	0.980	0.996	0.991
32 Zacatecas	0.910	0.892	0.950	0.925	0.963	0.927

A continuación, se muestran los resultados de realizar la prueba estadística de homocedasticidad Breusch-Pagan, Breusch y Pagan (1972), de los modelos polinomiales. Específicamente se muestra el p-valor que se obtuvo para cada entidad federativa. Cabe destacar primero que la hipótesis nula en esta prueba es la homocedasticidad. En segundo lugar, a menor p-valor se puede interpretar como mayor evidencia en los datos a favor de rechazar dicha hipótesis.

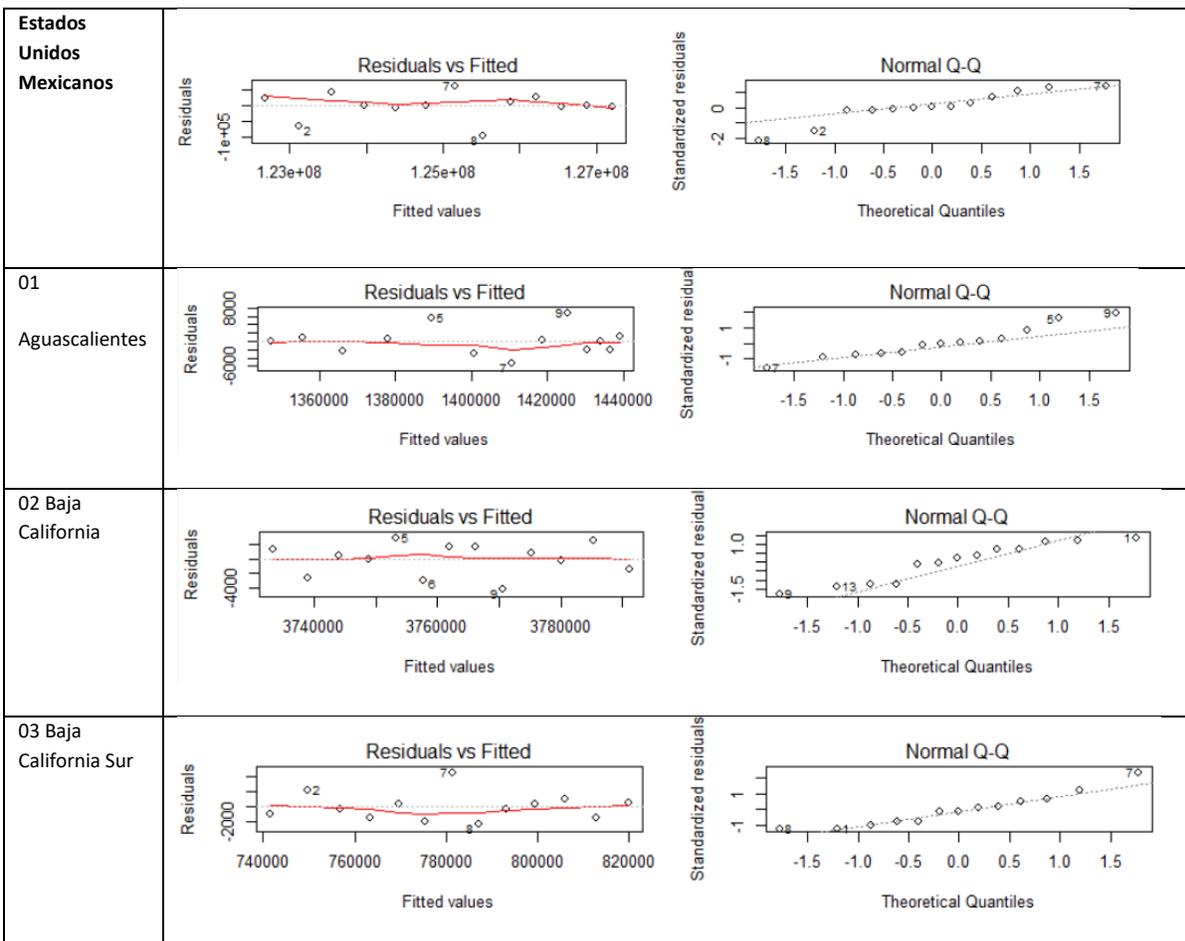
Entidad	p-valor
01 Aguascalientes	0.45
02 Baja california	0.95
03 Baja california Sur	0.92
04 Campeche	0.64
05 Coahuila de Zaragoza	0.15
06 Colima	0.16
07 Chiapas	0.78
08 Chihuahua	0.02
09 Ciudad de México	0.35
10 Durango	0.35
11 Guanajuato	0.77
12 Guerrero	0.97
13 Hidalgo	0.96
14 Jalisco	0.22
15 México	0.76
16 Michoacán de Ocampo	0.74
17 Morelos	0.83
18 Nayarit	0.67
19 Nuevo León	0.25
20 Oaxaca	0.24
21 Puebla	0.13
22 Querétaro	0.60
23 Quintana Roo	0.44
24 San Luis Potosí	0.13
25 Sinaloa	0.05
26 Sonora	0.86
27 Tabasco	0.20
28 Tamaulipas	0.69
29 Tlaxcala	0.57
30 Veracruz de Ignacio de la Llave	0.03
31 Yucatán	0.10
32 Zacatecas	0.45

De la anterior tabla, se puede observar que los residuos predichos por los modelos estimados para las entidades Chihuahua, p-valor=0.02, y Veracruz, p-valor=0.03, se encuentran en el límite de la decisión sobre la hipótesis de homocedasticidad, donde para un nivel de significancia de la prueba de 0.01, no existe suficiente evidencia para rechazar. Para los fines de este proyecto, se mantuvieron las estimaciones de los modelos para todas las entidades federativas. Sin embargo, casos como las

entidades de Chihuahua y Veracruz se seguirán evaluando en posteriores aplicaciones de la metodología aquí descrita.

Adicionalmente, las siguientes gráficas nos permite revisar visualmente el supuesto de homocedasticidad, donde se espera que los residuos y los valores predichos no presenten patrones claros de asociación. La segunda es la gráfica de cuantil-cuantil normal, donde se espera que los puntos puedan alinearse a la diagonal siempre que los residuos se distribuyan como una normal.

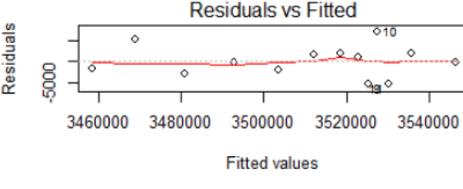
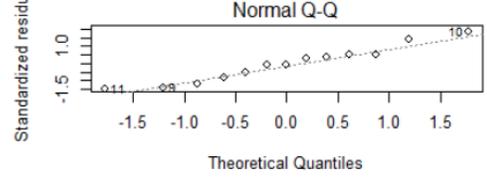
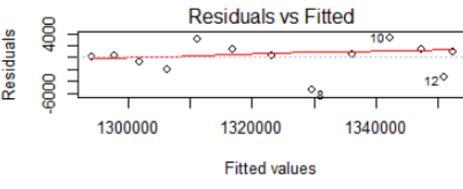
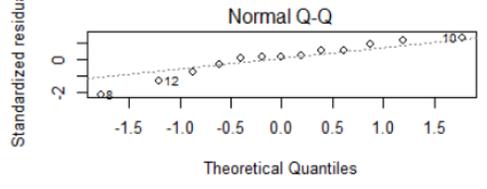
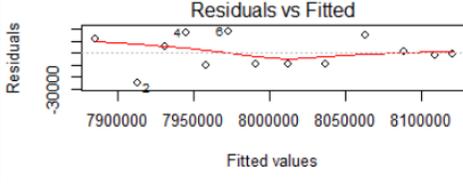
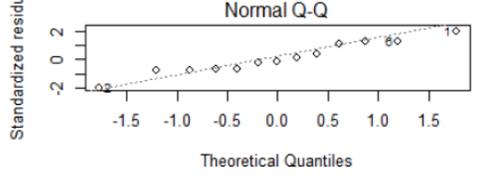
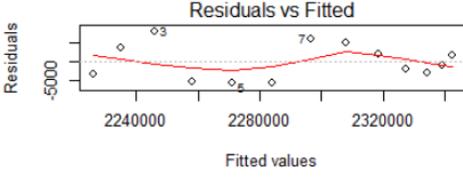
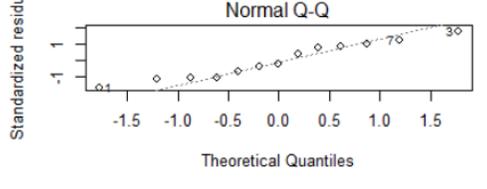
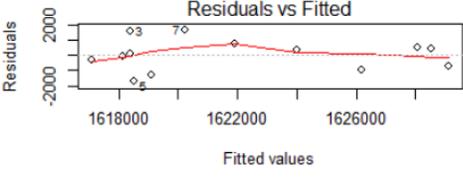
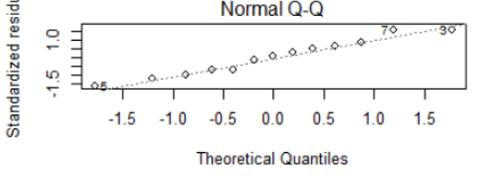
En términos generales, a lo largo de los gráficos de las entidades federativas, los supuestos de homocedasticidad y normalidad son soportados por las observaciones y residuales predichos. En el caso de los gráficos de residuales contra observaciones, los puntos mantienen una aleatoriedad o dispersión como era de esperarse bajo el principio de homocedasticidad, salvo los casos ya mencionados. En los gráficos de Normal Q-Q, en general también se aprecia que los residuos tienden a una distribución normal, comportamiento que podrá analizarse de mejor manera conforme se tengan mayores estimaciones.



04 Campeche		
05 Coahuila de Zaragoza		
06 Colima		
07 Chiapas		
08 Chihuahua		
09 Ciudad de México		
10 Durango		
11 Guanajuato		

12 Guerrero		
13 Hidalgo		
14 Jalisco		
15 México		
16 Michoacán de Ocampo		
17 Morelos		
18 Nayarit		
19 Nuevo León		

20 Oaxaca		
21 Puebla		
22 Querétaro		
23 Quintana Roo		
24 San Luis Potosí		
25 Sinaloa		
26 Sonora		
27 Tabasco		

28 Tamaulipas	 <p>Residuals vs Fitted plot for Tamaulipas. The y-axis is labeled 'Residuals' and ranges from -5000 to 5000. The x-axis is labeled 'Fitted values' and ranges from 3460000 to 3540000. A red smoothing line is shown. Several data points are labeled with their IDs: 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>	 <p>Normal Q-Q plot for Tamaulipas. The y-axis is labeled 'Standardized residuo' and ranges from -1.5 to 1.0. The x-axis is labeled 'Theoretical Quantiles' and ranges from -1.5 to 1.5. A dashed diagonal line represents the normal distribution. Data points are labeled with their IDs: 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>
29 Tlaxcala	 <p>Residuals vs Fitted plot for Tlaxcala. The y-axis is labeled 'Residuals' and ranges from -6000 to 4000. The x-axis is labeled 'Fitted values' and ranges from 1300000 to 1340000. A red smoothing line is shown. Data points are labeled with their IDs: 10, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>	 <p>Normal Q-Q plot for Tlaxcala. The y-axis is labeled 'Standardized residuo' and ranges from -2 to 0. The x-axis is labeled 'Theoretical Quantiles' and ranges from -1.5 to 1.5. A dashed diagonal line represents the normal distribution. Data points are labeled with their IDs: 10, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>
30 Veracruz de Ignacio de la Llave	 <p>Residuals vs Fitted plot for Veracruz de Ignacio de la Llave. The y-axis is labeled 'Residuals' and ranges from -30000 to 30000. The x-axis is labeled 'Fitted values' and ranges from 7900000 to 8100000. A red smoothing line is shown. Data points are labeled with their IDs: 4, 3, 2, 1, 0, -1, -2, -3, -4, -5, -6, -7, -8, -9, -10, -11, -12, -13, -14, -15, -16, -17, -18, -19, -20, -21, -22, -23, -24, -25, -26, -27, -28, -29, -30, -31, -32, -33, -34, -35, -36, -37, -38, -39, -40, -41, -42, -43, -44, -45, -46, -47, -48, -49, -50, -51, -52, -53, -54, -55, -56, -57, -58, -59, -60, -61, -62, -63, -64, -65, -66, -67, -68, -69, -70, -71, -72, -73, -74, -75, -76, -77, -78, -79, -80, -81, -82, -83, -84, -85, -86, -87, -88, -89, -90, -91, -92, -93, -94, -95, -96, -97, -98, -99, -100.</p>	 <p>Normal Q-Q plot for Veracruz de Ignacio de la Llave. The y-axis is labeled 'Standardized residuo' and ranges from -2 to 2. The x-axis is labeled 'Theoretical Quantiles' and ranges from -1.5 to 1.5. A dashed diagonal line represents the normal distribution. Data points are labeled with their IDs: 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>
31 Yucatán	 <p>Residuals vs Fitted plot for Yucatán. The y-axis is labeled 'Residuals' and ranges from -5000 to 5000. The x-axis is labeled 'Fitted values' and ranges from 2240000 to 2320000. A red smoothing line is shown. Data points are labeled with their IDs: 3, 7, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>	 <p>Normal Q-Q plot for Yucatán. The y-axis is labeled 'Standardized residuo' and ranges from -1 to 1. The x-axis is labeled 'Theoretical Quantiles' and ranges from -1.5 to 1.5. A dashed diagonal line represents the normal distribution. Data points are labeled with their IDs: 3, 7, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>
32 Zacatecas	 <p>Residuals vs Fitted plot for Zacatecas. The y-axis is labeled 'Residuals' and ranges from -2000 to 2000. The x-axis is labeled 'Fitted values' and ranges from 1618000 to 1626000. A red smoothing line is shown. Data points are labeled with their IDs: 3, 7, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>	 <p>Normal Q-Q plot for Zacatecas. The y-axis is labeled 'Standardized residuo' and ranges from -1.5 to 1.0. The x-axis is labeled 'Theoretical Quantiles' and ranges from -1.5 to 1.5. A dashed diagonal line represents the normal distribution. Data points are labeled with their IDs: 3, 7, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.</p>

6. Estimador de población a nivel nacional y por entidad federativa

El propósito del presente anexo es mostrar que la estimación de población a nivel nacional es la suma de las estimaciones por entidad federativa, utilizando para ambos casos un modelo polinomial. Lo anterior se hará usando el método de mínimos cuadrados para estimar los parámetros de los modelos involucrados.

Para efectos del anterior propósito y como se estableció en la Sección 1.4, se supondrá que las estimaciones de población por muestreo se pueden modelar por la siguiente expresión polinomial de grado m :

$$\hat{X}_{et} = \beta_{e0} + \beta_{e1}t + \dots + \beta_{em}t^m + \varepsilon_{et}, \quad (1)$$

Donde \hat{X}_{et} es la estimación de población por muestreo para la entidad e al tiempo t , $\beta_{e0}, \dots, \beta_{em}$ son los coeficientes del polinomio y ε_{et} es la parte residual del modelo, donde $e = 1, \dots, 32$ y $t = 1, \dots, \tau$.

Ahora, el modelo (1) se puede expresar en forma matricial como:

$$\hat{X}_e = T\beta_e + \varepsilon_e, \quad (2)$$

Donde:

$$\hat{X}_e = \begin{pmatrix} \hat{X}_{e1} \\ \hat{X}_{e2} \\ \vdots \\ \hat{X}_{e\tau} \end{pmatrix}, T = \begin{pmatrix} 1 & 1 & \dots & (1)^m \\ 1 & 2 & \dots & (2)^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \tau & \dots & (\tau)^m \end{pmatrix}, \beta_e = \begin{pmatrix} \beta_{e0} \\ \beta_{e1} \\ \vdots \\ \beta_{em} \end{pmatrix} \text{ y } \varepsilon_e = \begin{pmatrix} \varepsilon_{e1} \\ \varepsilon_{e2} \\ \vdots \\ \varepsilon_{e\tau} \end{pmatrix}. \quad (3)$$

Es conveniente observar que la matriz T no depende de la entidad federativa que se trate.

De esta forma, el estimador por mínimos cuadrados de β_e se puede expresar como:

$$\hat{\beta}_e = (T'T)^{-1}T'\hat{X}_e. \quad (4)$$

Aquí T' denota la matriz transpuesta de T y $(T'T)^{-1}$ la matriz inversa de $T'T$.

Así, las predicciones de población, \hat{Y}_e , resultantes del modelo (2), utilizando el estimador en la expresión (4) de los coeficientes de este, se pueden expresar como:

$$\begin{aligned} \hat{Y}_e &= T\hat{\beta}_e \\ &= T(T'T)^{-1}T'\hat{X}_e. \end{aligned} \quad (5)$$

Con lo anterior, para mostrar que las predicciones de población a nivel país se pueden expresar como la suma de predicciones de cada entidad, se procede como sigue.

Primero sea $\hat{X} = \hat{X}_1 + \dots + \hat{X}_{32}$ la estimación de población por muestreo a nivel país. En segundo lugar, si \hat{X} se puede modelar por una expresión similar a la de (2), entonces:

$$\hat{X} = T\beta + \epsilon \quad (6)$$

y

$$\hat{\beta} = (T'T)^{-1}T'\hat{X}, \quad (7)$$

Donde:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_\tau \end{pmatrix} \quad (8)$$

Y $\hat{\beta}$ es el estimador de mínimos cuadrados de β . Así, la predicción de población a nivel país, \hat{Y} , con el modelo polinomial (6) estimado resulta ser:

$$\hat{Y} = T\hat{\beta}. \quad (9)$$

Ahora, utilizando las expresiones (5), (7) y (9), obsérvese que la suma de predicciones de población de cada entidad federativa se puede expresar como:

$$\begin{aligned} \hat{Y}_1 + \dots + \hat{Y}_{32} &= T(T'T)^{-1}T'\hat{X}_1 + \dots + T(T'T)^{-1}T'\hat{X}_{32} \\ &= [T(T'T)^{-1}T'][\hat{X}_1 + \dots + \hat{X}_{32}] \\ &= [T(T'T)^{-1}T']\hat{X} \\ &= T\hat{\beta} \\ &= \hat{Y}. \end{aligned} \quad (10)$$

Lo anterior muestra que la suma de las predicciones de población para cada entidad federativa, utilizando el modelo polinomial (2) estimado, resulta ser la predicción de población a nivel nacional, utilizando el modelo polinomial (6) estimado.

7. Estimación por enfoque de muestreo y por enfoque asistido por modelos

Para el presente método de suavizamiento se supone que las estimaciones puntuales de población por muestreo se pueden modelar por la siguiente expresión polinomial de grado m :

$$\hat{X}_{et} = \beta_{e0} + \beta_{e1}t + \dots + \beta_{em}t^m + \varepsilon_{et}, \quad (11)$$

Donde \hat{X}_{et} es la estimación de población por muestreo para la entidad e al tiempo t , $\beta_{e0}, \dots, \beta_{em}$ son los coeficientes del polinomio y ε_{et} es la parte residual o residuo del modelo, donde $e = 1, \dots, 32$ y $t = 1, \dots, \tau$. Dicha parte residual absorbe la parte de la estimación puntual que no es explicada por el modelo polinomial. En este sentido, el residuo es función de los coeficientes:

$$\varepsilon_{et} = \varepsilon_{et}(\beta_{e0}, \beta_{e1}, \dots, \beta_{em}). \quad (12)$$

Es conveniente hacer notar que, la elección de dichos coeficientes influirá en la varianza del estimador suavizado por el modelo.

Ahora, el modelo (1) al tiempo t se puede expresar en forma matricial como:

$$\hat{X}_e(t) = T(t)\beta_e(t) + \varepsilon_e(t), \quad (13)$$

Donde

$$\hat{X}_e(t) = \begin{pmatrix} \hat{X}_{e1} \\ \hat{X}_{e2} \\ \vdots \\ \hat{X}_{et} \end{pmatrix}, T(t) = \begin{pmatrix} 1 & 1 & \dots & (1)^m \\ 1 & 2 & \dots & (2)^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t & \dots & (t)^m \end{pmatrix}, \beta_e(t) = \begin{pmatrix} \beta_{e0} \\ \beta_{e1} \\ \vdots \\ \beta_{em} \end{pmatrix} \text{ y} \quad (14)$$

$$\varepsilon_e(t) = \begin{pmatrix} \varepsilon_{e1} \\ \varepsilon_{e2} \\ \vdots \\ \varepsilon_{et} \end{pmatrix}.$$

Obsérvese que los coeficientes $\beta_{e0}, \beta_{e1}, \dots, \beta_{em}$ son función también de t , pero se omite en la notación por simplicidad de esta.

La estrategia para elegir los valores de los coeficientes del modelo es elegir aquellos valores tales que la estimación suavizada esté lo más cercano al valor observado de la estimación puntual por muestreo, en términos de su distancia euclidiana. Por esto, se usarán los coeficientes que minimicen la suma de residuos al cuadrado. En otros términos, se utilizarán los parámetros $\beta_e(t)$ tales que minimicen:

$$\varepsilon_e'(t)\varepsilon_e(t) = (\hat{X}_e'(t) - \beta_e'(t)T'(t))(\hat{X}_e(t) - T(t)\beta_e(t)), \quad (15)$$

En la notación utilizada aquí, A' denotará la transpuesta de alguna matriz A .

Es posible mostrar que los coeficientes que minimizan la expresión en (15) son:

$$\hat{\beta}_e(t) = (T'(t)T(t))^{-1}T'(t)\hat{X}_e(t). \quad (16)$$

Así, la estimación suavizada de población al tiempo t resultante del modelo (11), utilizando los coeficientes de la expresión (16), se puede escribir como:

$$\begin{aligned} \hat{Y}_e(t) &= T(t)\hat{\beta}_e(t) \\ &= T(t)(T'(t)T(t))^{-1}T'(t)\hat{X}_e(t). \end{aligned} \quad (17)$$

Obsérvese que, la estimación suavizada al tiempo t es una combinación lineal de las estimaciones de población por muestreo hasta el tiempo t . Específicamente, la estimación suavizada al tiempo t se puede expresar como:

$$\hat{Y}_{et} = h_{t1}\hat{X}_{e1} + \dots + h_{tt}\hat{X}_{et}, \quad (18)$$

Donde (h_{t1}, \dots, h_{tt}) es el t –ésimo, el último, renglón de la matriz:

$$H(t) = T(t)(T'(t)T(t))^{-1}T'(t). \quad (19)$$

Esta última matriz se le conoce como matriz Sombrero (Hat en inglés) o matriz de proyección en la literatura referente a regresión lineal, ver por ejemplo la Sección 8.1 de (Draper, N.R. & H. Smith (1998). *Applied Regression Analysis*. New York: John Wiley & Sons).

Adicionalmente, si se utiliza la expresión (18), la varianza del estimador en (18) está dada por:

$$\begin{aligned} Var(\hat{Y}_{et}) &= \sum_{i=1}^t h_{t(t-i+1)}^2 Var(\hat{X}_{e(t-i+1)}) \\ &+ 2 \sum_{i=1}^t \sum_{j>i}^t h_{t(t-i+1)} h_{t(t-j+1)} Cov(\hat{X}_{e(t-i+1)}, \hat{X}_{e(t-j+1)}) \end{aligned} \quad (20)$$

Obsérvese que, la varianza muestral del estimador de población, $Var(\hat{X}_{et})$, es parte de la varianza anterior. Por lo tanto, es de esperar que los intervalos de confianza para la población de la entidad e calculados con las respectivas estimaciones de $Var(\hat{Y}_{et})$ o $Var(\hat{X}_{et})$ serán diferentes.

De lo anterior, supóngase que $Var(\hat{Y}_{et}) < Var(\hat{X}_{et})$, entonces los intervalos de confianza teóricos para X_{et} y centrados en \hat{X}_{et}^{PT} tendrán la siguiente relación de contención:

$$\begin{aligned} &\left(\hat{X}_{et}^{PT} - z_{1-\alpha/2} \sqrt{Var(\hat{Y}_{et})}, \hat{X}_{et}^{PT} + z_{1-\alpha/2} \sqrt{Var(\hat{Y}_{et})} \right) \\ &\subset \left(\hat{X}_{et}^{PT} - z_{1-\alpha/2} \sqrt{Var(\hat{X}_{et})}, \hat{X}_{et}^{PT} \right. \\ &\quad \left. + z_{1-\alpha/2} \sqrt{Var(\hat{X}_{et})} \right), \end{aligned} \quad (21)$$

Con probabilidad 1.

Ahora, como se mencionó en los apartados 1.2.11 y 1.3, si la inferencia es basada en diseños de muestreo, entonces el intervalo de confianza teórico:

$$\left(\hat{X}_{et}^{PT} - z_{1-\alpha/2} \sqrt{Var(\hat{X}_{et})}, \hat{X}_{et}^{PT} + z_{1-\alpha/2} \sqrt{Var(\hat{X}_{et})} \right) \quad (22)$$

Es el de menor longitud y tiene una cobertura de $(1 - \alpha)\%$, esto es:

$$P \left(\left(\hat{X}_{et}^{PT} - z_{1-\alpha/2} \sqrt{\text{Var}(\hat{X}_{et})}, \hat{X}_{et}^{PT} + z_{1-\alpha/2} \sqrt{\text{Var}(\hat{X}_{et})} \right) \ni X_{et} \right) = 1 - \alpha. \quad (23)$$

Por lo tanto, usando la contención en la expresión (21), resultará que:

$$P \left(\left(\hat{X}_{et}^{PT} - z_{1-\alpha/2} \sqrt{\text{Var}(\hat{Y}_{et})}, \hat{X}_{et}^{PT} + z_{1-\alpha/2} \sqrt{\text{Var}(\hat{Y}_{et})} \right) \ni X_{et} \right) < 1 - \alpha. \quad (24)$$

Lo que significa que si $\text{Var}(\hat{Y}_{et}) < \text{Var}(\hat{X}_{et})$, entonces los intervalos de confianza, que involucran la $\text{Var}(\hat{Y}_{et})$, tienen menor cobertura que la esperada.

Por otro lado, siguiendo una argumentación similar a la anterior, si $\text{Var}(\hat{Y}_{et}) > \text{Var}(\hat{X}_{et})$, entonces los intervalos de confianza teóricos, que involucran la $\text{Var}(\hat{Y}_{et})$, tendrán una cobertura de al menos $(1 - \alpha)\%$, pero serán de mayor longitud, con probabilidad 1, que el intervalo de la expresión (22).

En conclusión, si utilizamos la varianza del estimador por suavizamiento, $\text{Var}(\hat{Y}_{et})$, para expresar los intervalos de confianza y su valor es menor que la varianza del estimador por muestreo, $\text{Var}(\hat{X}_{et})$, entonces los intervalos serán de menor longitud y con menor cobertura nominal. Por otro lado, si dicha varianza por suavizamiento es mayor que la resultante del diseño de muestreo, entonces se tendrá una cobertura mayor o igual a la nominal, pero los intervalos resultantes serán de mayor longitud, lo que contradice la determinación referida en la ecuación (23). Por este motivo, **el enfoque será utilizar la varianza del estimador por muestreo para el cálculo de los intervalos de confianza.**

Antes de terminar este apartado, se hará una descripción general sobre la probabilidad de cobertura de un intervalo de confianza y los valores contenidos en una estimación de este. Posteriormente, dicha descripción se hará para el caso específico de estimación de población.

Considérese que el parámetro poblacional de interés es θ (un escalar), considérese también que un estimador de este parámetro es $\hat{\theta}$ y un estimador de la varianza de $\hat{\theta}$ es $\widehat{\text{Var}}(\hat{\theta})$. Adicionalmente, considérese que se tienen las condiciones suficientes para que el estadístico:

$$\frac{\hat{\theta} - \theta}{\sqrt{\widehat{\text{Var}}(\hat{\theta})}} \quad (22)$$

Tenga aproximadamente una distribución normal estándar. De esta forma, un intervalo de $100 \times (1 - \alpha)\%$, aproximadamente, de confianza está dado por:

$$\left(\hat{\theta} - z_{1-\alpha/2} \sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} + z_{1-\alpha/2} \sqrt{\text{Var}(\hat{\theta})} \right), \quad (23)$$

Donde $z_{1-\alpha/2}$ es el cuantil de orden $1 - \alpha/2$ de una normal estándar. Este estimador por intervalo es un objeto abstracto. Cuando se evalúa este estimador en una muestra específica, se obtiene la estimación por intervalos respectiva a dicha muestra (un objeto concreto). Finalmente, con dicha evaluación se obtiene el límite inferior y el superior que determinan la estimación del referido intervalo.

Dicho lo anterior, algo conveniente de aclarar es el significado de la probabilidad de cobertura. Esto es, la interpretación de que el intervalo de confianza en la expresión (23) contenga al parámetro θ con probabilidad $1 - \alpha$:

$$P \left(\left(\hat{\theta} - z_{1-\alpha/2} \sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} + z_{1-\alpha/2} \sqrt{\text{Var}(\hat{\theta})} \right) \ni \theta \right) = 1 - \alpha. \quad (24)$$

Dentro del contexto de muestreo de poblaciones finitas, dicha interpretación es como sigue. Si se tienen los recursos necesarios para seleccionar independientemente un número suficientemente grande de muestras de la población bajo estudio y con el mismo diseño de muestreo, entonces aproximadamente el $100 \times (1 - \alpha)\%$ de las muestras así seleccionadas producirán intervalos estimados que sí contienen el valor del parámetro θ . Es conveniente notar que, la probabilidad de cobertura es una propiedad del procedimiento de construcción del intervalo. En contraposición, dicha probabilidad no es una propiedad del intervalo de confianza estimado de una muestra particular. Para mayores detalles de la construcción de intervalos de confianza y su interpretación se puede consultar el Capítulo 2 de Meeker et al. (2017).

Sin embargo, es oportuno aclarar qué información proporcionan los valores contenidos en un intervalo estimado para una muestra específica. Para esto, se debe notar que el intervalo de la expresión (23), cumple las condiciones suficientes para denominar a los valores contenidos en la estimación de dicho intervalo como valores plausibles o verosímiles del parámetro de interés. Adicionalmente, estas mismas condiciones suficientes garantizan que los valores alrededor del centro del intervalo son más plausibles que los valores a los extremos de dicho intervalo. Los detalles técnicos sobre lo que se afirma aquí se pueden encontrar en el Capítulo 11 de Kalbfleisch (1985).

Con todo lo anterior, para la estimación de población X_{et} con el estimador puntual \hat{X}_{et} , se cumplen los supuestos suficientes para que:

$$\left(\hat{X}_{et} - 1.6449 \sqrt{\widehat{\text{Var}}(\hat{X}_{et})}, \hat{X}_{et} + 1.6449 \sqrt{\widehat{\text{Var}}(\hat{X}_{et})} \right) \quad (25)$$

Sea un intervalo de 90%, aproximadamente, de confianza. Para este objeto abstracto se aplica el criterio de cobertura. Esto es, si se tienen los recursos necesarios para seleccionar

independientemente un número suficientemente grande de muestras de la población bajo estudio y con el mismo diseño de muestreo, entonces aproximadamente el 90% de las muestras así seleccionadas producirán intervalos estimados que sí contienen el valor del parámetro X_{et} . Por otro lado, después de seleccionar una muestra, el intervalo estimado (objeto concreto) contendrá valores plausibles o verosímiles del parámetro de interés. Adicionalmente, los valores alrededor del centro del intervalo estimado son más plausibles que los cercanos a los extremos de dicho intervalo.

8. Prácticas Nacionales e Internacionales sobre Estimación de Población

Existen varias prácticas internacionales sobre la estimación de población con base en una muestra. Entre las experiencias más destacadas están lo realizado por la American Community Survey (ACS) de la Agencia de Censos de Estados Unidos y lo hecho en el French Rolling Census realizado por el Instituto Nacional de Estadística y de Estudios Económicos de Francia.

Con referencia a las prácticas nacionales, se encuentra la Encuesta Intercensal realizada por el Instituto.

A continuación, se hace una breve descripción de cada uno de estos casos donde se utiliza una muestra de viviendas para estimar el volumen de población. Lo anterior se hace con la finalidad de mostrar como algunas agencias nacionales de estadística utilizan dicha metodología.

8.1 American Community Survey (ACS)

La ACS es una encuesta nacional que todos los años recopila y produce información oportuna sobre las características sociales, económicas, de vivienda y demográficas de la población de EE. UU. Desde sus inicios, 2005, la ACS ha proporcionado un flujo continuo de información actualizada para los estados y las áreas locales del citado país.

Específicamente, la ACS recopila información anualmente en los 50 estados de EE. UU., el Distrito de Columbia y Puerto Rico. Sin embargo, no recopila información en los cuatro territorios principales de EE. UU. (Samoa Estadounidense, Guam, las Islas Marianas del Norte y las Islas Vírgenes de los EE. UU.). Para mayores detalles sobre esta encuesta se puede consultar el documento metodológico de la ACSO (2014).

Los datos que componen dicha información son utilizados por muchas partes interesadas del sector público, del sector privado y sin fines de lucro para asignar fondos, rastrear los cambios demográficos, planificar emergencias y aprender sobre las comunidades locales.

Un hecho importante de esta encuesta es que los formularios no se envían por correo postal a personas específicas, sino a direcciones específicas. La Oficina del Censo selecciona una muestra de direcciones para incluirlas en la ACS. Cada dirección tiene una probabilidad de alrededor de 1 en 480 de ser seleccionada en un mes, y ninguna dirección puede seleccionarse más de una vez cada cinco años. La Oficina del Censo envía cuestionarios a aproximadamente 295,000 direcciones al mes. Esta es una pequeña cantidad de hogares considerando que hay más de 140 millones de direcciones que cumplen con los requisitos en los Estados Unidos. Adicionalmente, una dirección que recibe las

instrucciones de la ACS probablemente no encontrará a un vecino o amistad que también las haya recibido.

La Oficina del Censo envía cartas por correo postal para informar a las personas que viven en las direcciones que han sido seleccionadas para participar en la ACS y para proporcionar instrucciones para completar la encuesta por internet. En Puerto Rico y en algunas áreas de difícil acceso en los Estados Unidos, solo se envía por correo postal un cuestionario impreso. A los hogares se les pide que completen la encuesta por internet o que envíen por correo postal el cuestionario impreso completo al Centro Nacional de Procesamiento de la Oficina del Censo. Si la Oficina del Censo no recibe la encuesta completa en unas pocas semanas, enviará otro cuestionario de encuesta impreso.

Luego de todos los contactos por correo postal, se toma una muestra de las direcciones que no hayan respondido por internet, por correo postal o de las direcciones con entregas en un apartado postal. Los representantes de campo de la Oficina del Censo visitan estas direcciones y realizan la entrevista en persona.

La ACS tiene la capacidad exclusiva de producir estimaciones sobre una gran cantidad de geografías, incluidos los niveles geográficos bajos tales como los sectores censales y los grupos de bloques censales. Las áreas geográficas están organizadas según su jerarquía geográfica. Las unidades más grandes, como los estados, incluyen a las unidades más pequeñas, como los condados y los sectores censales. Los grupos de bloque censales son el componente fundamental más pequeño de la ACS.

Los datos de la ACS son oportunos, ya que se publican en el año inmediatamente posterior al año en que fueron recopilados.

La ACS crea estimaciones periódicas, lo que significa que estas representan las características de la población y la vivienda durante un período específico de recopilación de datos. Estas son actualmente estimaciones de un año y de cinco años.

Las estimaciones a un año se realizan con la información recopilada en los 12 meses de año de referencia y produce datos de áreas con población de más de 65,000 habitantes. Las estimaciones de un año de la ACS del 2005 se publicaron por primera vez en el 2006.

Existen también las estimaciones suplementarias a un año. Estas estimaciones son versiones simplificadas de las tablas de la ACS. Produce datos de áreas con poblaciones de más de 20,000 habitantes. Las respectivas estimaciones de 2014 se publicaron por primera vez en el 2016.

Las estimaciones de cinco años se construyen con información de 60 meses y produce datos para todas las áreas. Las estimaciones de cinco años de la ACS del 2005 al 2009 se publicaron por primera vez en el año 2010.

Por otro lado, los productos de la ACS son utilizados, por ejemplo, en todo el gobierno federal. Las agencias que lo componen usan dichas estimaciones para informar a los legisladores de políticas públicas, distribuir fondos y evaluar programas. Por ejemplo, el Departamento de Justicia de los EE. UU., el Departamento de Trabajo de los EE. UU. y la Comisión para la Igualdad de Oportunidades en el Empleo de los EE. UU. usan las estimaciones de la ACS para hacer cumplir las leyes contra la discriminación en el empleo. El Departamento de Asuntos de Veteranos de los EE. UU. usa las estimaciones de la ACS para evaluar la necesidad de programas de cuidado de salud, educación y

empleo para aquellas personas que han prestado servicio en las fuerzas armadas. Adicionalmente, el Departamento de Educación de los EE. UU. utiliza las estimaciones de la ACS para desarrollar programas de educación y alfabetización para adultos.

La información de la ACS es también fundamental para las agencias locales y estatales. Los planificadores y legisladores usan las estimaciones actualizadas para evaluar la necesidad de construir nuevas carreteras, hospitales, escuelas, servicios para personas de la tercera edad y otros servicios básicos. Además, los datos de la ACS proporcionan a las comunidades locales información importante sobre sus ciudadanos, como por ejemplo el nivel de educación, los patrones de desplazamiento al trabajo y los idiomas que se hablan.

8.2 French Rolling Census

Desde 2004, Francia ha pasado de un censo tradicional a un censo continuo (rolling census en inglés). La legislación correspondiente actual establece los principios para la realización del censo y la difusión de la población oficial anual, estimada, de cada municipio. Se debe tomar en cuenta que existen alrededor de 35,000 municipios en Francia. Esta población estimada es referencia para alrededor de 350 textos administrativos o legislativos, en particular para el cálculo de la contribución financiera que el Estado otorga cada año a cada municipio. Para mayores detalles sobre este censo continuo se puede consultar INSEE (2021).

El censo continuo de población también se utiliza para conocer la población diversa y cambiante de Francia. Desde 2009, proporciona cada año datos sobre los habitantes y sus características a nivel municipal, pero también a un nivel inferior, áreas locales de alrededor de 2,000 habitantes: composición por edad y sexo, ocupación, condiciones de vivienda, medios de transporte, desplazamientos al trabajo o a la institución de educación.

Estos resultados son utilizados por las autoridades administrativas y locales para mejorar las instalaciones públicas como escuelas, instalaciones deportivas, transporte público, etc. Las empresas y asociaciones también los utilizan para mejorar el conocimiento de las personas que viven en un lugar determinado.

Desde 2004, una encuesta de población anual se realiza, por ley, en cooperación con cada municipio. El correspondiente proceso es gestionado y controlado por el Instituto Nacional de Estadística y Estudios Económicos de Francia (INSEE por sus siglas en francés), pero la recopilación de datos lo realizan los municipios que contratan a los empadronadores y gestionan la encuesta de campo. La recolección de información se realiza durante cuatro o cinco semanas, desde mediados de enero hasta mediados de febrero.

En este censo continuo, todos los municipios por debajo de 10,000 habitantes se censan una vez cada 5 años. Los municipios de 10,000 habitantes o más se realiza anualmente una encuesta con una muestra de viviendas que representa el 8 % de su población.

Así, cada año, la encuesta anual constituye una muestra del 14 % de la población total que vive en Francia. Durante un período de 5 años, se han tenido en cuenta en el censo todos los habitantes de los municipios con menos de 10,000 habitantes y el 40 % de la población de los municipios con 10,000 habitantes o más.

Cabe destacar que es obligatorio responder al censo y los ayuntamientos pueden imponer una multa de 38 euros a las personas que se nieguen a responder. Sin embargo, las sanciones no se aplican con frecuencia porque se hace hincapié en que responder al censo es un acto civil, lo cual es importante para que los municipios puedan tomar buenas decisiones de política pública. Además, no lleva mucho tiempo, alrededor de 5 minutos para un cuestionario. Gracias a una fuerte campaña de comunicación antes y durante el período censal, la tasa de respuesta se mantiene en un nivel alto, del 96.1 % en 2019 por ejemplo.

Luego, cada año, los datos recopilados durante los últimos cinco años se compilan para producir una estimación de la población actualizada de cada municipio de Francia y las características detalladas de la población y las viviendas para todos los niveles geográficos. La población calculada, por ejemplo, para el primero de enero de 2016, el año mediano del ciclo de 5 años, se basa en los datos recopilados de 2014 a 2018. Los resultados correspondientes se publicaron en enero de 2019. El método para producir estas estimaciones varía con el tamaño de los municipios.

Para los municipios de más de 10,000 habitantes, la población se estima multiplicando el número de viviendas presentes en el registro de edificios localizados (que es exhaustivo) al primero de enero de 2016 y el promedio de personas por vivienda calculado en las últimas 5 encuestas censales, para el ejemplo de 2014 a 2018.

Para los municipios menores de 10,000 habitantes, los cálculos utilizan información adicional como tendencias observadas en fuentes administrativas, como el archivo del impuesto a la vivienda que mantiene registros de todas las viviendas residenciales ocupadas o no.

Más precisamente, se crea una base de datos para todos los municipios pequeños con una fecha de referencia 2016. El 20 % es una enumeración directa; el 40 % se extrapola hacia delante de los dos años anteriores de recaudación utilizando el archivo del impuesto a la vivienda; el restante 40 % se interpola hacia atrás de los siguientes dos años de recaudación.

8.3 Encuesta Intercensal

La Encuesta Intercensal 2015 realizada por el Instituto Nacional de Estadística y Geografía, INEGI, de México tuvo como objetivo generar información estadística actualizada que proporcione estimaciones con calidad sobre el volumen, la composición y distribución de la población. Adicionalmente, otro objetivo fue estimar el número de viviendas del territorio nacional, cuya desagregación geográfica fue por municipios, entidad federativa, localidades de 50 mil o más habitantes y a nivel nacional. Un objetivo más fue mantener la comparabilidad histórica con los censos y encuestas nacionales y de otros países; así como obtener estimadores de proporciones, tasas y promedios de las variables de interés. Para mayores detalles sobre esta encuesta se puede consultar Encuesta Intercensal (2015).

En las últimas décadas, la demanda de los usuarios por disponer de información actualizada en períodos más cortos que los decenales ha sido creciente, al igual que la solicitud de que estos se desagreguen a nivel municipal; por esta razón, el INEGI en 1995 realizó el primer conteo de población y vivienda y en 2005 el segundo. En ambos casos, el objetivo central fue generar información socio demográfica básica para actualizar los datos sobre el tamaño, la composición y la distribución territorial de la población, así como de los hogares y las viviendas existentes en el país.

A partir de estas experiencias, y en las proximidades de 2015, el INEGI valoró que la opción para actualizar la información demográfica y socioeconómica de México en el período intercensal era realizar una encuesta de cobertura temática amplia. Esta actualización debería permitir proporcionar estimaciones del total de la población nacional, por entidad federativa y municipio, así como conocer el comportamiento de las características captadas para la población y las viviendas particulares habitadas y para las principales ciudades del país, en términos de proporciones, tasas y promedios.

La encuesta se desarrolló bajo las siguientes bases metodológicas:

- Para la integración del marco de muestreo, en 2014 se realizó un operativo denominado Recorrido de Actualización del Marco Geoestadístico Nacional, del Entorno Urbano y las Características de las Localidades, cuya finalidad fue actualizar la cartografía urbana y rural del país. Durante este proceso se aprovechó el recorrido en campo para poner al día la información sobre el entorno urbano y las características de las localidades, recabada en 2010.
- Las unidades de observación fueron las viviendas particulares habitadas y sus residentes habituales.
- El período de levantamiento fue del 2 al 27 de marzo de 2015 y el momento de referencia de la información es el 15 de marzo de 2015.
- En cada vivienda en muestra se aplicó un cuestionario impreso mediante una entrevista directa a un informante adecuado.
- El informante adecuado fue la jefa o el jefe de la vivienda o, en su ausencia, una persona de 18 años o más de edad que fuese residente habitual de la misma y conociera los datos de todos los residentes.
- Se realizaron hasta cuatro visitas a las viviendas particulares habitadas, en distinto horario o día, para conseguir la entrevista.
- El diseño estadístico consideró un tamaño de muestra esperado de 6.1 millones de viviendas a nivel nacional.
- Para asegurar la cobertura, del 27 de marzo al 19 de junio se realizó un operativo de verificación.

El diseño de la muestra para la Encuesta Intercensal 2015 permitió actualizar las estimaciones sobre el volumen, la composición y distribución de la población y de las viviendas particulares habitadas del territorio nacional, así como obtener estimadores de proporciones, tasas y promedios de las variables de interés.

En particular, la muestra permitió estimar los totales de viviendas particulares habitadas y la población que en ellas reside para los siguientes dominios de estudio:

- Estados Unidos Mexicanos.
- 32 entidades federativas.
- 2 457 municipios o delegaciones (número de municipios en 2015).
- Cada una de las localidades de 50 mil o más habitantes.

De igual manera posibilitó obtener estimadores de proporciones, tasas y promedios de las variables de interés, tanto para los niveles geográficos antes mencionados como para los siguientes dominios de estudio:

- Entidad federativa con cinco tamaños de localidad predefinidos.
- Cualquier agrupación geográfica del indicador deseado que por el tamaño de la muestra permita formar un nuevo dominio de estudio (zonas metropolitanas, regiones, etcétera).

La construcción del marco de muestreo tomó como base la cartografía actualizada al cierre del Recorrido de Actualización del Marco Geoestadístico Nacional en 2014. Así como las estadísticas obtenidas del Censo de Población y Vivienda 2010 (a nivel manzana y localidad rural), del Recorrido para la Actualización del Inventario Nacional de Viviendas 2012, de los Censos Económicos 2014, del Registro Único de Viviendas (RUV) y del Recorrido de Actualización del Marco Geoestadístico Nacional.

El tamaño de muestra mínimo por municipio para obtener estimaciones con precisión y confianza adecuada fue de aproximadamente 1300 viviendas particulares habitadas, por lo que se determinó censar a todos los municipios que en el 2010 contaban con igual o menor número de viviendas; también se censaron algunos municipios y localidades con población vulnerable, en atención a los requerimientos de información por parte de los usuarios. Entre estas poblaciones se encuentran principalmente:

- Los 100 primeros municipios con población en extrema pobreza.
- Municipios con rezago social muy alto.
- Algunas localidades con población afro mexicana.
- Algunas localidades con población hablante de lengua indígena y en particular donde se habla alguna lengua indígena en riesgo de desaparecer.

El tamaño de muestra esperado, considerando la selección de las áreas en los municipios muestreados y los municipios a censar, fue de 6.1 millones de viviendas y durante el levantamiento se obtuvieron 7.9 millones de las cuales 5.9 fueron habitadas.

El esquema de muestreo fue estratificado por conglomerados y en una sola etapa, es decir: para formar los estratos se clasificaron las localidades según tamaño y nivel socioeconómico, se seleccionaron áreas geográficas completas utilizando muestreo aleatorio simple, y en su interior, se visitaron todas las viviendas para captar sus características y las de sus residentes. En las localidades de 50 mil y más habitantes se estratificaron las áreas geoestadísticas básicas (AGEB) por nivel socioeconómico, y se seleccionaron en promedio dos manzanas de cada una de ellas.

Bibliografía

- American Community Survey Office, ACSO, (2014). American Community Survey Design and Methodology. United States Census Bureau.
- Breusch, T. S. y A. R. Pagan (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica*. 47 (5), págs. 1287–1294.
- Casella, R. y R. L. Berger (1990). *Statistical Inference*. Brooks/Cole Publishing Company.
- Consejo Nacional de Población. (2014). *Proyecciones de la Población 2010-2050*. CONAPO. México. Junio del 2015. <http://www.conapo.gob.mx/es/CONAPO/Proyecciones>
- Cochran, W. (1992). *Técnicas de Muestreo*. México. CECSA.
- Dalenius, T. & Hodges, JL. (1959). “Minimum variance stratification”. *Journal of American Statistics Association*. No. 54, pp. 88-101.
- Encuesta Intercensal (2015). Encuesta Intercensal 2015: síntesis metodológica y conceptual. México: Instituto Nacional de Estadística y Geografía.
- Forrester, JW. (1992). “Policies, decisions and information sources for modeling” *European Journal of Operational Research*. Volume 59. Issue 1, pp. 42-63.
- Hansen-Morris, H., Hurtwitz, WN., & Madow, WG. (1976). *Sample Survey: Methods and Theory*. New York, John Wiley and Sons.
- INEGI. (2012). *Marco Maestro de Muestreo 2012*. Dirección de Diseño y Marcos Estadísticos. México.
- INEGI. (2012). *Marco Nacional de Viviendas, muestras maestras 2012*. Dirección de Diseño y Marcos Estadísticos. México.
- INEGI. (2010). *Censo de Población y Vivienda 2010*. México.
- INEGI. (2010). *Encuesta Nacional de Ocupación y Empleo ENOE*. México.
- INEGI. (2010). *Sistema de Integración Territorial*. México.
- INEGI. *Censo de población 1990-2010*. INEGI. México. Junio del 2015. <http://www.inegi.org.mx/est/contenidos/proyectos/ccpv/default.aspx>
- INSEE. (2021, 14 octubre). Population Rolling Census. Institut National de La Statistique et Des Études Économiques. Recuperado 14 de octubre de 2021, de <https://www.insee.fr/en/metadonnees/source/serie/s1321/presentation>.

- Jarque, CM. (1981). "A solution to the problem of Optimum Stratification in Multivariate Sampling". JRSS, Series C. *Applied Statistics*. Vol. 30. No. 2, pp. 163-169.
- Kalbfleisch, J. G. (1985). *Probability and Statistical Inference: Statistical Inference* (Vol. 2). Springer Science & Business Media.
- Kish, L. (1972). *Muestreo de Encuestas*. México. Trillas.
- Levy, P., & Lemeshow, S. (1991). *Sampling Populations: Methods and Applications*. New York, John Wiley and Sons.
- Lynn, P. (2009). *Methodology of Longitudinal Surveys*. New York, John Wiley and Sons.
- Lohr, SL. (2000). *Muestreo: Diseño y Análisis*. México. International Thomson Editores.
- Meeker, W. Q., Hahn, G. J., & Escobar, L. A. (2017). *Statistical intervals: a guide for practitioners and researchers*. John Wiley & Sons.
- Naciones Unidas. (2009). *Diseño de Muestras para Encuestas de Hogares: directrices prácticas*. Nueva York. Estudios de Métodos. Departamento de Asuntos Económicos y Sociales. División de Estadística. Serie F. No. 98.
- Naciones Unidas. (1987). *Manual de Encuestas sobre hogares*. Nueva York. Departamento de Asuntos Económicos y Sociales Internacionales. Oficina de Estadística. Serie F. No. 31.
- Pettersson, H. (2007). "Diseño de marcos muestrales maestros y muestras maestras para las encuestas de hogares en países en desarrollo". Cap. V. *Encuestas de hogares en los países en desarrollo y en transición*. Estudios de Métodos. Serie F. No. 96. Nueva York. Naciones Unidas. Departamento de Asuntos Económicos y Sociales. División de Estadística.
- Raj, D. (1984). *Teoría del Muestreo*. México. Fondo de Cultura Económica.
- Romero, D., Burguete J., Velasco, JR. & Martínez, LE. (2006). "Un enfoque de optimización combinatoria para construcción de marcos de muestreo en hogares." *Boletín de los Sistemas Nacionales Estadístico y de Información Geográfica*. México. INEGI.
- Särndal, CE., Swensson., B. & Wretman J. (1992). *Model Assisted Survey Sampling*. New York. Springer.
- Scheaffer, RL., Mendenhall, W., & Ott, L. (1987). *Elementos de Muestreo*. México. Grupo Editorial Iberoamérica.
- Subdirección de Diseño Muestral en Vivienda (SDMV). (2016). *Glosario de términos de muestreo*. Dirección de Diseño y Marcos Estadísticos. México. INEGI.