



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**ESTIMACIONES DE POBLACIÓN EN FECHAS
ESPECÍFICAS PARA LA EXPANSIÓN
DE ENCUESTAS EN HOGARES**

REPORTE DE TRABAJO PROFESIONAL

QUE PARA OBTENER EL TÍTULO DE:

**A C T U A R I A
P R E S E N T A:**

SARA IVETH MERA CEBALLOS



**DIRECTOR DE TESIS:
DRA. MARÍA EDITH PACHECO
GÓMEZ MUÑOZ
2012**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de datos del jurado

1. Datos del alumno
Mera
Ceballos
Sara Iveth
51 11 41 83
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
097236858
2. Datos del tutor
Dra.
Ma. Edith
Pacheco
Gómez Muñoz
3. Datos del sinodal 1
Dr.
Víctor Alfredo
Bustos
Y de la Tijera
4. Datos del sinodal 2
M. en D.
Alejandro
Mina
Valdés
5. Datos del sinodal 3
M. en D.
Víctor
García
Vilchis
6. Datos del sinodal 4
M. en D.
Elsa
Pérez
Paredes
7. Datos del trabajo escrito
Estimaciones de población en fechas específicas para la expansión de encuestas en hogares
41 p
2012

Dedico este trabajo a mi familia amada
Benjamín, Rosa Elsa, Sandy, Gris y Benji.

Agradecimientos:

A mis padres Benjamín Mera y Rosa Elsa Ceballos, por ser ejemplo y guía en mi vida, por todo su apoyo y confianza, además de haber compartido conmigo su amor, paciencia y consejos.

A mis hermanos Sandy, Gris y Benji, por su cariño, compañía y comprensión, por lo que hemos aprendido juntos y todo lo que hemos compartido.

A Eloina Meneses, mi gran amiga, por ser mi consejera y la compañía constante en los momentos cruciales de mi vida durante los últimos años.

A Elsa Pérez Paredes, mi amiga y maestra, por darme la oportunidad de aprender y por la confianza de tantos años.

A la Universidad Nacional Autónoma de México, por ser una gran institución y porque me ofreció la oportunidad única de adquirir conocimientos que van más allá de la técnica y la metodología. Por el orgullo de pertenecer a su comunidad y de llevar sus colores.

A mis queridos amigos de toda la vida, porque han sido mi segunda familia, incluso a la distancia.

A mis amigos y compañeros Alejandro Mozo y Felipe Pérez, por su colaboración en el desarrollo del procedimiento que aquí se documenta.

En general, agradezco a todas las personas que durante este trayecto de vida han compartido conmigo un poco de su cariño, conocimiento y mucho de su valioso tiempo, porque de cada uno conservo un recuerdo, una lección y ejemplo.

A la Dra. Edith Pacheco por todo el apoyo, tiempo y disposición dedicados a la revisión de este trabajo, por sus valiosas recomendaciones para llevarlo a término.

Al Mtro. Alejandro Mina, al Mtro. Víctor García, al Dr. Víctor Alfredo Bustos, a la Mtra. Elsa Pérez y a la Act. Eloina Meneses, por tomarse el tiempo para dar lectura a este documento y por los comentarios que ayudaron a consolidar su contenido.

Empieza por hacer lo necesario, luego lo que es posible y de pronto te encontrarás haciendo lo imposible.

San Francisco de Asís

*Aprender es descubrir lo que ya sabes.
Hacer es demostrar que lo sabes.
Enseñar es recordar a otros que lo saben tanto como tú. Todos somos aprendices, hacedores, maestros.*

Richard Bach

Experiencia no es lo que le sucede a un hombre; es lo que un hombre hace con lo que le sucede.

Aldous Huxley

Índice

Introducción	3
I. Antecedentes y contexto	5
1.1 Captación de información sociodemográfica, relevancia y métodos de recolección	5
1.2 Algunos elementos técnicos de las encuestas en hogares	7
1.3 Ponderación y poblaciones de referencia	7
1.4 Cálculo de las poblaciones de referencia	8
II. Definición del problema	10
III. Algunos métodos para la estimación de poblaciones en fechas específicas	11
3.1 Generalidades sobre la interpolación	11
3.2 Interpolación lineal	12
3.3 Tasa de crecimiento medio constante	13
3.4 Los multiplicadores de Karup-King	15
3.5 Interpolación con spline cúbico	17
IV. Implementación y resultados	21
4.1 Argumentos para la selección del mejor método para interpolar	21
4.2 Enumeración de las etapas para la estimación completa de montos de población en fechas específicas para la ponderación de encuestas	22
4.3 Cálculo de los totales mensuales nacionales y por entidad federativa	23

4.4	Ajuste por tamaño de localidad	26
4.5	Adecuación para la estimación por ciudad	27
4.6	Resultados	29
V.	Conclusiones	35
	Glosario	39
	Bibliografía	40

Introducción

La aplicación de la técnica matemática tiene cabida en todas las áreas donde se requiere realizar análisis cuantitativo o cualitativo de la información disponible a través de distintas fuentes. El tratamiento de la información y la interpretación de los resultados son los que deben adecuarse al contexto en el que se enmarca el problema y la consecuente solución.

En actuaría se hace uso de técnicas y modelos matemáticos para la resolución de problemas de naturaleza diversa, con propósitos encaminados tanto a la evaluación de la eficiencia y la eficacia de sistemas o proyectos en marcha, como al desarrollo óptimo de nuevos productos, en el sentido más general del término *producto*.

Áreas como seguros, pensiones, finanzas, estadística, demografía e informática son en donde se desempeña con mayor frecuencia el actuario, sin que ello signifique exclusividad o restricción en los campos para la aplicación del conocimiento y las habilidades que posee el profesional de la actuaría.

La aplicación que se describe en el presente documento obedece a la provisión de insumos para la solución de un problema enmarcado directamente dentro de un procedimiento de la estadística muestral, que por la temática y los resultados que proporciona considera elementos básicos de la demografía: la interpolación de poblaciones de referencia en distintos dominios geográficos para la ponderación de las encuestas en hogares que realiza el Instituto Nacional de Estadística y Geografía (INEGI).

Los insumos para el desarrollo del procedimiento fueron los datos de población obtenidos del Censo 2000 y del Conteo 2005, desagregados por entidad federativa y localidad, además de la serie anual de proyecciones de población por entidad federativa del periodo 2000-2012 que incorpora la información del ejercicio de Conciliación Demográfica Conapo-INEGI-Colmex 2006 y que fueron publicadas por el Consejo Nacional de Población (CONAPO).

En la primera etapa se revisó la factibilidad de la aplicación de otros métodos, entre ellos el uso de los factores de Karup-King para la desagregación de las tasas de crecimiento por periodo en mitades, quintas y décimas partes, sin embargo, por sus características y flexibilidad se optó por la implementación final de una variante de la interpolación cúbica de

Hermite (interpolación tipo spline) sobre la tasa de crecimiento exponencial de la población anual proyectada, con ello se estimaron montos de población nacionales a mitad de cada mes para el total nacional y por cada entidad federativa, y por último, cada una de las series se ajustó proporcionalmente a las estructuras proyectadas en cada entidad federativa por tamaño de localidad. El procedimiento también se adecuó para obtener las series por ciudades.

El producto final del procedimiento es una serie mensual de montos de población total, por entidad federativa y ciudad, desagregada por cuatro tamaños de localidad, para el periodo 2005-2010. Además de una plantilla de trabajo para realizar interpolaciones de la misma naturaleza y en donde las fechas se pueden modificar a fin de ajustarla a las necesidades del ejercicio.

Vale la pena señalar que la estimación completa de los montos involucra otros procedimientos adicionales a la interpolación y que aún cuando la descripción detallada en este documento menciona la totalidad de las etapas, la mayor parte del contenido se enfocará primordialmente en la definición del método para interpolación de poblaciones.

La metodología se desarrolló en el año 2007 y sustituyó a los tradicionalmente utilizados: interpolación lineal de poblaciones o estimación por tasa de crecimiento geométrico (bajo el supuesto de tasa de crecimiento medio anual constante por periodo), y continúa vigente hasta el presente año, con la salvedad de que algunas etapas han tenido ajustes para optimizar la implementación y los insumos incorporan información nueva conforme a su disponibilidad.

El contenido del documento se organizó en cinco apartados principales. El primer capítulo tiene como propósito describir la importancia que tiene la información estadística básica y las necesidades a las que atiende; también describe a grandes rasgos los tipos de fuentes, en particular, explica algunas características de las encuestas en hogares y de manera más específica detalla la relevancia del proceso de ponderación.

La segunda parte es la más breve pero fundamental, ya que corresponde a la descripción detallada de la problemática a resolver.

En el tercer capítulo se refiere a la descripción básica de la técnica para la interpolación de poblaciones y de cuatro métodos: lineal, por tasa de crecimiento medio constante, utilizando multiplicadores de Karup-King y el spline cúbico.

La implementación del spline cúbico para la interpolación a nivel nacional y por entidad federativa, los procedimientos adicionales para el ajuste en áreas geográficas más pequeñas o respecto de su nivel de urbanización, así como los resultados definitivos en cada caso, conforman el cuarto apartado.

Finalmente, el quinto capítulo corresponde al compendio de conclusiones y comentarios a la solución implementada, a los resultados y a la importancia que tiene la vinculación entre la formación profesional del actuario y el conocimiento de las problemáticas reales en los campos de aplicación.

I. Antecedentes y contexto

1.1 Captación de información sociodemográfica, relevancia y métodos de recolección

El Instituto Nacional de Estadística y Geografía (INEGI), es el órgano responsable de integrar y coordinar el Sistema Nacional de Información Estadística y Geográfica (SNIEG), que es el conjunto de Unidades organizadas a través de los Subsistemas, articuladas mediante la Red Nacional de Información, con el propósito de producir y difundir la Información de Interés Nacional.

La ejecución de los proyectos de generación de información estadística como son censos, encuestas, así como la integración de las cuentas satélite de los distintos sectores de actividad económica; de los registros administrativos de hechos vitales, sociales, judiciales, entre otros; son responsabilidad del INEGI y además de alimentar el SNIEG, permiten al instituto dar cumplimiento a su función de generador de información que sirva al sector público y a la sociedad civil, ya que proporcionan los elementos necesarios para el monitoreo periódico de las componentes demográficas y de las variables económicas de relevancia, para que a partir de ello exista una toma oportuna de decisiones que garantice la provisión de servicios adecuados, así como el diseño y evaluación de políticas públicas encaminadas a procurar el bienestar de la población.

En lo que se refiere a la información sociodemográfica, los censos de población por su característica de universalidad son el método principal para captar información de este tipo, se levantan en la totalidad del territorio nacional y por lo tanto recaban la información básica de la población residente en nuestro país en una fecha fija, de ahí que sea común encontrar la descripción de un censo como *una fotografía instantánea de la población en un momento determinado*.

Si bien es cierto que en nuestro país la información que provee un censo es exhaustiva en cuanto a la cobertura territorial, es necesario considerar algunas desventajas: tiene un alto costo, lo que a su vez propicia un espaciamiento temporal considerable entre levantamientos (10 años), y el hecho de que su propósito principal es dar a conocer en términos descriptivos

a la totalidad de la población, hace que se priorice la calidad de la información por encima de la cantidad.

Así pues, los datos que capta son relativamente limitados pues se refieren a la caracterización demográfica básica (edad y sexo), el conocimiento del nivel de algunos de los fenómenos demográficos principales que afectan a la población (tasa de crecimiento, fecundidad y migración), así como de algunas variables socioeconómicas con alto impacto en la dinámica del cambio en su volumen y estructura (escolaridad, condición de actividad, habla de lengua indígena, entre otras).

En este contexto, los registros administrativos y las encuestas surgen como opciones adecuadas para la captación de información más detallada respecto de eventos particulares o fenómenos de interés, que complementan los datos disponibles a partir de un censo, además de que por sus características brindan mayor oportunidad con costos financieramente viables.

Los registros administrativos cuentan entre sus ventajas el hecho de ser permanentes, por lo que su actualización es constante y el costo de su levantamiento no requiere de disponibilidad de recursos adicionales a los programados; sin embargo, a pesar de ser útiles en términos de la estadística sociodemográfica, por lo general su diseño original no está pensado para los mismos propósitos que un proyecto de generación estadística como censo, conteo o encuesta; de ahí que en ocasiones presentan problemas de homogeneidad o no respuesta alta en la captación de algunas variables y categorías relevantes.

Por su parte, las encuestas manejan temáticas específicas, por lo tanto permiten profundizar en aspectos relevantes que por su extensión no son tratados en un cuestionario censal básico. Generalmente su diseño se apega tanto a las necesidades de información como a marcos conceptuales consolidados a propósito del tema. No obstante entre sus desventajas se encuentra que el alcance de los resultados es limitado a determinados grupos de población y dominios geográficos, además de que al ser ejercicios estadísticos, dependen de un buen diseño muestral para garantizar que la información que se colecta en sólo una porción de la población se infiera como válida para el total.

Ante el panorama ya descrito, es evidente la relevancia que tiene la disponibilidad de información estadística básica oportuna, confiable y con cobertura amplia. De la misma manera, también se advierte que aunque se puede obtener a partir de distintas fuentes, cada una con características y propósitos particulares, ninguno de los proyectos de generación estadística está exento de imprecisiones o desventajas, sin embargo, dado que todos tienen en común el objetivo de proveer información de calidad, es posible complementarlos entre sí para conocer las características sociodemográficas de la población mexicana, entender los fenómenos que la afectan y monitorear los cambios que en ella se presentan a lo largo del tiempo.

1.2 Algunos elementos técnicos de las encuestas en hogares

Las encuestas en hogares son una herramienta utilizada recurrentemente en nuestro país para recolectar la información sobre características y fenómenos muy específicos que se presentan en la población, su viabilidad es alta tanto en términos de oportunidad como de costo, de ahí que en la actualidad existen proyectos de este tipo que tienen carácter de permanentes. Así pues, son muchos los motivos para garantizar la calidad de la información que ofrecen y para ello, antes de poner la información a disposición de los usuarios, se implementan mecanismos de evaluación, procesamiento y corrección de datos.

Desde el diseño de la muestra se incorporan elementos técnicos que aseguran la cobertura, por ejemplo, las características descriptivas sobre el volumen total de la población objetivo (monto, proporción respecto del total, etc.), indicadores base a estimar, la tasa de no respuesta esperada, entre otros. Por su parte, en el procesamiento de la información existen fases como la validación de campo y la automática, que están encaminadas a la detección de errores en la captación y a la verificación de la congruencia de la información, tanto en el momento del levantamiento como en etapas posteriores; y por último, se aplican procedimientos que corrigen el efecto de la diferencia entre lo programado y lo efectivamente recabado, tal es el caso de la reponderación de los factores de expansión.

En el caso de las encuestas que diseña y levanta el INEGI, se trata de ejercicios probabilísticos, lo que garantiza que los resultados obtenidos a través de ellas tienen validez para el total de la población nacional e incluso para algunas desagregaciones geográficas o categóricas específicas, siempre y cuando éstas hayan sido consideradas en el diseño de la muestra, que usualmente es de tipo complejo¹. La unidad de selección muestral es la vivienda particular, mientras que los hogares al interior y las personas que los conforman y que residen habitualmente constituyen las categorías de análisis.

1.3 Ponderación y poblaciones de referencia

El ponderador de una muestra se define como el inverso multiplicativo de la probabilidad de selección de cada uno de los elementos en la población, en términos prácticos es equivalente al número de individuos a los que representa la unidad en muestra dentro de la población de la cual se extrajo, de ahí que también sea conocido como factor de expansión, ya que las características que se captan de dicha unidad se hacen extensivas a tantos elementos como lo indica su factor.

Por lo general, su uso obedece a la búsqueda de lograr que las estimaciones sean consistentes con la evolución de la población a lo largo del tiempo y con otras encuestas, así como para corregir las observaciones que faltan debido a la no respuesta, aumentando el peso de las

¹ Se refiere al conjunto de unidades seleccionadas a través de un esquema polietápico y que combina distintos tipos de muestreo.

unidades que sí proporcionan información; adicionalmente al hecho de que una muestra ponderada ofrece la opción de generar estimaciones insesgadas de los totales poblacionales para distintos propósitos.

La ponderación se realiza tomando como base la información disponible sobre el conjunto de población objetivo del ejercicio. Dado que en las encuestas en hogares que realiza el INEGI la unidad de análisis suele ser la población o algún subconjunto de ella, las cifras de control para el cálculo de factores de expansión están dadas por los montos de población que proyecta el CONAPO, de tal forma que las estimaciones de totales resultantes de las encuestas las replican.

Cabe señalar que el cálculo de los ponderadores considera únicamente las cifras de control para ámbitos geográficos, por lo que no interviene de ninguna manera con las variaciones muestrales relacionadas con características de la población (edad y sexo). Los dominios geográficos que consideran son: total nacional, entidad federativa y cuatro tamaños de localidad; adicionalmente algunos ejercicios requieren información para calibrar los factores de expansión en dominios más concretos como ciudades.

1.4 Cálculo de las poblaciones de referencia

Los operativos de levantamiento, tanto de los proyectos continuos como de los especiales², se programan a lo largo del año y toman en cuenta factores tales como la cantidad de recursos humanos destinados, el tamaño total de la muestra, la extensión de cada cuestionario y el tiempo promedio de entrevista. Por lo general esta fase se cuenta en semanas, puede ser menor a un mes o extenderse hasta un trimestre completo.

Debido a que los resultados de la encuesta deben tener una referencia temporal bien definida para la ponderación de los resultados y que ésta no puede ser el periodo completo debido a la dinámica de la población y las variaciones constantes (aunque mínimas) en su tamaño a lo largo del mismo; se utiliza la fecha media del operativo como el momento en el que debe conocerse el monto de población de cada ámbito geográfico de interés y al que debe representar la muestra expandida del ejercicio.

Por su parte, las series de población proyectadas son anuales y por lo general tienen referencia solamente a dos momentos: al inicio y la mitad de cada año (1ro. de enero y 30 de junio), esta situación explica el porqué no pueden utilizarse directamente para el cálculo de factores de expansión, sin embargo, sí se pueden utilizar como base para la estimación de los montos de población en cada momento que se requiera.

² Los proyectos continuos o también llamados tradicionales, corresponden a encuestas que por la temática que atienden y su relevancia, tienen carácter de permanente y su levantamiento se encuentra programado a lo largo del año, con variaciones mínimas. Por otra parte, los proyectos especiales corresponden a encuestas que se levantan de manera ocasional a solicitud de un sector específico o sobre una temática especial y que su continuidad no está asegurada. En esta segunda clasificación también se encuentran las encuestas cuya temática está por demás justificada, sin embargo, su levantamiento no es periódico y depende de la disponibilidad de recursos extraordinarios al ejercicio fiscal de cada año.

Existe una gran variedad de métodos para interpolar montos de población para fechas específicas, las variantes se originan por los supuestos que subyacen en el procedimiento, la técnica específica para su cálculo e incluso las características de los resultados finales. En el caso de la estimación para la expansión de encuestas se distinguen dos métodos que por la sencillez de su implementación son los más utilizados, el primero es la interpolación lineal directa de los totales de población, y el segundo es la interpolación utilizando la tasa de crecimiento bajo el supuesto de que el crecimiento medio anual es constante entre los dos momentos base.

En términos netamente prácticos, ambos métodos cumplen con su principal cometido que es brindar la información en la fecha necesaria para la expansión, sin embargo, tienen limitaciones ya que la tasa de crecimiento de la población es variable en el tiempo, por lo tanto el supuesto de linealidad y de constancia entre momentos base no se cumple, así que es de esperarse que los montos resultantes por interpolación directa puedan observar inconsistencias en la tendencia de la intensidad de crecimiento al interior del periodo y los que se estiman a partir de la tasa constante observen un comportamiento atípico, por ejemplo tasas escalonadas entre estimaciones.

II. Definición del problema

El desarrollo, adecuación y aplicación de modelos matemáticos hace posible la resolución de problemas en distintos ámbitos, o bien, aporta mejoras en los procedimientos ya establecidos.

En el caso de la estimación de poblaciones para la expansión de encuestas probabilísticas, aunque el planteamiento del problema nos remite directamente a la obtención de totales de población en momentos determinados, es necesario tener en cuenta una serie de elementos adicionales como el comportamiento del crecimiento poblacional a través del tiempo, la distribución de la población en el territorio nacional, su distribución espacial de acuerdo con el grado de urbanización de la localidad en la que reside, entre otras; todo ello para que los resultados obtenidos no solamente sean consistentes en términos numéricos, sino que garanticen un comportamiento coherente en términos demográficos.

De manera específica, el problema se enfocó en determinar para periodos intercensales y postcensales, los totales de población del país, de cada entidad federativa, para 32 ciudades y por cuatro tamaños de localidad³, cifras que habrían de servir como base para el cálculo de los factores de expansión de las distintas encuestas en hogares que realiza el INEGI, tanto a proyectos continuos como especiales. Los datos se requirieron a mitad de cada mes en el periodo 2000-2010, teniendo como base e insumo la serie anual de la población proyectada por el CONAPO.

Además de ser consistente con los montos estimados, la serie total, de cada entidad federativa y ciudad debía observar tasas de crecimiento regulares y sin perturbaciones o cambios drásticos, mientras que cada una de las series por tamaño de localidad debía respetar la tendencia de la tasa de crecimiento correspondiente y al mismo tiempo conservar la proporción respecto del total proyectado para el periodo.

³La clasificación se realiza de acuerdo con el número de habitantes y se agrupan en: menos de 2 500 habitantes (rurales), de 2 500 a 14 999 habitantes (urbano medio bajo), de 15 000 a 99 999 habitantes (urbano medio alto) y de 100 000 o más (urbano alto).

III. Algunos métodos para la estimación de poblaciones en fechas específicas

La disponibilidad de las cifras de población desagregadas en categorías muy específicas y con temporalidad prácticamente continua es una solicitud recurrente de los usuarios de la información sociodemográfica, de ahí que a través de la historia han surgido distintas alternativas para atender a dicha necesidad.

Los métodos desarrollados e implementados se enfocan en dos grandes propósitos: proveer cifras por edad desplegada, es decir, que sirven para obtener totales de población en cada edad partiendo de la información agrupada (generalmente en grupos quinquenales o decenales); y el segundo en atención a la necesidad de estimar los montos en fechas distintas a las de levantamiento, las cifras anuales en periodos intercensales son un ejemplo de este resultado.

De acuerdo con el problema a resolver, los métodos que se revisaron y propusieron son los que están diseñados para la desagregación temporal y aunque existen muchas posibilidades, en este apartado se describen *grosso modo* solamente los más utilizados y aquellos que al implementarse ofrecieron mejores resultados en términos de las características deseadas.

3.1 Generalidades sobre la interpolación

La interpolación se puede definir de una manera sencilla como la acción de estimar valores intermedios en una serie de datos conocidos utilizando fórmulas matemáticas o procedimientos gráficos. Generalmente el éxito de un procedimiento de interpolación se demuestra por el cambio mínimo en las diferencias sucesivas entre los datos conocidos y los estimados. Es una cuestión de juicio sobre si el método seleccionado es adecuado y si la aproximación es o no de buena calidad y de utilidad.⁴

⁴ Shriock, Henry S, Siegel Jacob S. and Associate. "Interpolation of point data", en: *The Methods and Materials of Demography*, Volume 2, second printing. Washington, D.C, Bureau of the Census, 1973, p. 681-682.

Se pueden identificar dos tipos de metodologías para la interpolación de datos: la primera por aproximación a los puntos observados y la segunda por ajuste o réplica exacta de los datos de origen, estos últimos son conocidos como de suavizamiento. En el primer grupo se encuentran los métodos que pretenden generar datos intermedios que sean consistentes con la información original en términos de la tendencia, el patrón y el nivel, aunque no necesariamente están diseñados para replicar los puntos base de manera precisa y en algunos casos la consistencia con información observada se limita a una localidad de la serie completa.

Los del segundo tipo, también tienen como propósito proporcionar datos intermedios congruentes con la serie de origen, sin embargo, la diferencia radica en que pretenden la réplica exacta de los datos base, además de que consideran el comportamiento de la serie completa para determinar la consistencia de las estimaciones. En ambos casos hay variantes que suavizan la tendencia; en el caso de los de aproximación, incluso corrigen irregularidades derivadas de datos atípicos observados.

3.2 Interpolación lineal

Este método es netamente aritmético y está clasificado como de aproximación, ya que aunque proporciona valores congruentes con el comportamiento de los puntos extremos observados y adicionalmente permite replicarlos, la tendencia de los puntos intermedios estimados puede no ser congruente con la de la serie completa observada.

El método consiste en obtener estimaciones intermedias utilizando una recta entre cada par de puntos, de modo que la representación general de la recta de interpolación entre los puntos base (x_i, y_i) , (x_{i+1}, y_{i+1}) , se puede escribir como sigue:

$$F(k) = y_i + \frac{y_{i+1} - y_i}{x_{i+1} - x_i} (k - x_i)$$

$$\text{donde, } |x_i| \leq |k| \leq |x_{i+1}|$$

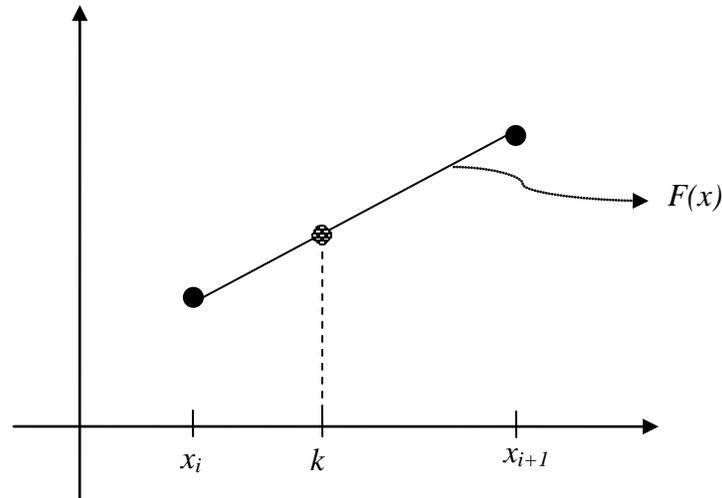
Una vez definida la recta, se puede calcular el valor de $P(x)$ para tantos puntos como se requiera sin necesidad de cubrir condiciones adicionales, teniendo como garantía que los valores estimados cumplirán con:

$$|F(x_i)| \leq |F(k)| \leq |F(x_{i+1})|$$

En términos más sencillos esta condición implica que los valores interpolados cumplirán con la tendencia creciente o decreciente que se observe entre cada par de puntos base.

Figura 1

Interpolación lineal



3.3 Tasa de crecimiento medio constante

Este método es una variante de la interpolación lineal, ya que también supone que el crecimiento en el intervalo $[x_t, x_{t+h}]$ es uniforme y sostenido, aunque difiere en el supuesto de la forma en que ocurre, pues considera que el incremento de personas en la población a través del tiempo no es lineal. En cuestiones relacionadas con la población de nuestro país la tasa geométrica es la más comúnmente utilizada, aunque también se utiliza la tasa exponencial.

Tomando en consideración que la tasa de crecimiento es la velocidad o ritmo de cambio de los efectivos de la población en un periodo determinado⁵ y que la tasa geométrica se refiere al incremento acumulado en un periodo de longitud conocida en función de la población al inicio del mismo, la expresión analítica para estimar el total de individuos en cualquier momento intermedio del intervalo se encuentra a continuación:

$$P_{t+h} = P_t(1 + r)^h$$

Donde,

r Tasa media de crecimiento del periodo

P_t Población al inicio del periodo

P_{t+h} Población al final del periodo

h Duración del periodo

⁵ Haupt, Arthur y Thomas T. Kane (2003). *Guía rápida de población*, 4ª edición. Washington, D.C, Population Reference Bureau.

Dado que r es la tasa media de crecimiento en el periodo $[t, t+h]$ la expresión para su cálculo se obtiene como un despeje de la expresión anterior:

$$r = \left[\left(\frac{P_{t+h}}{P_t} \right)^{1/h} - 1 \right] \times 100$$

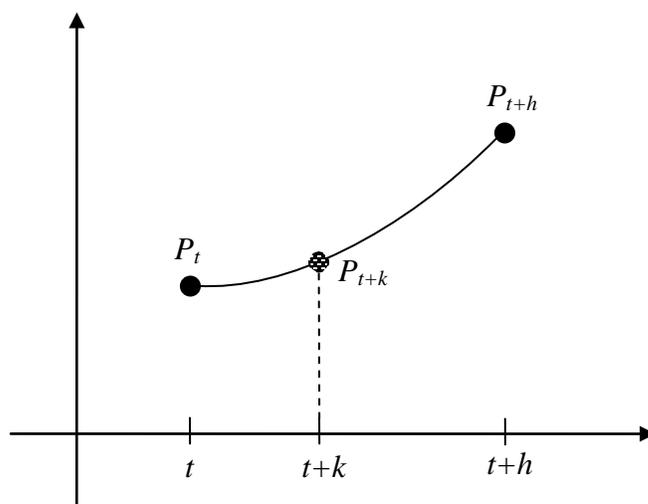
Así pues, bajo el supuesto de que el crecimiento de la población mantiene el mismo ritmo durante todo el periodo, cualquier valor intermedio se puede obtener en función de la población inicial multiplicada por el efecto de la tasa a lo largo del tiempo transcurrido entre la fecha de la observación inicial y la fecha de interés, como se ilustra en la figura 2 y cuya expresión analítica se expresa de la siguiente forma:

$$P_{t+k} = P_t(1+r)^k \quad \text{donde } k < h$$

Vale la pena señalar que el valor estimado de P_{t+k} puede tener imprecisiones derivadas del hecho de que aún cuando el crecimiento en el intervalo efectivamente ocurriera con una intensidad r , no necesariamente es uniforme en el periodo, lo que en términos gráficos significaría que la forma real de la curva de crecimiento entre P_t y P_{t+h} puede ser distinta a la que supone el método.

Figura 2

Interpolación por tasa de crecimiento medio constante



3.4 Los multiplicadores de Karup-King

La fórmula de Karup-King es uno de los métodos más sencillos desarrollados y utilizados para la aplicación de interpolación por tramos con polinomios sociables, o lo que también se conoce como interpolación osculatoria. Su uso obedece a dos propósitos básicos, el primero es la estimación de poblaciones entre fechas dadas (generalmente periodos intercensales) y el segundo es para la desagregación de información que se presenta agrupada.

Es un método que utiliza cuatro puntos dados *equidistantes* (o los límites de tres grupos de un total de cuatro). La fórmula se asemeja a la de un polinomio de segundo grado ordinario (expresada en diferencias) obtenido a partir de los tres primeros puntos, con un ajuste que requiere del cuarto punto.⁶ La popularidad de su uso radica en que existe un conjunto de coeficientes fijos (multiplicadores) generados a partir de la fórmula básica y que se pueden utilizar de forma genérica, sin necesidad de hacer el cálculo para una aplicación en particular.

Cuadro 1

Multiplicadores de Karup-King para interpolar datos entre puntos dados (en quintas partes)

Puntos a interpolar	Coeficientes a ser aplicados para los puntos base:			
	N _{1,0}	N _{2,0}	N _{3,0}	N _{4,0}
Primer intervalo				
N _{1,0}	1.000	0.000	0.000	0.000
N _{1,2}	0.656	0.552	-0.272	0.064
N _{1,4}	0.408	0.856	-0.336	0.072
N _{1,6}	0.232	0.984	-0.264	0.048
N _{1,8}	0.104	1.008	-0.128	0.016
Segundo intervalo (medio)				
N _{2,0}	0.000	1.000	0.000	0.000
N _{2,2}	-0.064	0.912	0.680	-0.016
N _{2,4}	-0.072	0.696	0.424	-0.048
N _{2,6}	-0.048	0.424	0.696	-0.072
N _{2,8}	-0.016	0.168	0.912	-0.064
Tercer intervalo				
N _{3,0}	0.000	0.000	1.000	0.000
N _{3,2}	0.016	-0.128	1.008	0.104
N _{3,4}	0.048	-0.264	0.984	0.232
N _{3,6}	0.072	-0.336	0.856	0.408
N _{3,8}	0.064	-0.272	0.552	0.656
N _{4,0}	0.000	0.000	0.000	1.000

Nota: Los coeficientes de interpolación fueron calculados a partir de las formulas básicas por Wilson H. Grabill del Buró de Censos de Estados Unidos.

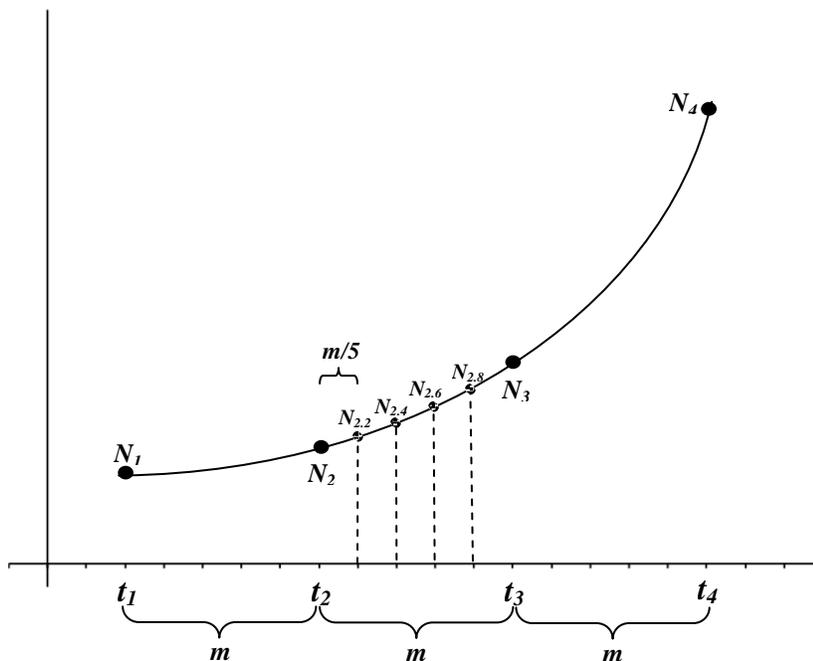
Fuente: Anexo C de Tablas de coeficientes de interpolación del libro *The Methods and materials of demography*.

⁶ Op.cit p. 700

Por la forma de construcción de la técnica, aunque permite interpolar montos para los tres intervalos entre los cuatro puntos de base, al llevar a cabo pruebas sobre datos conocidos se ha observado que los resultados son de mejor calidad para el intervalo central, ya que el error en los extremos se incrementa, sobre todo si la población observó algún cambio en su ritmo de crecimiento.

Figura 3

Interpolación de datos en el intervalo intermedio utilizando multiplicadores de Karup-King



Otra de las características de este método, es que los datos interpolados también son equidistantes entre sí y distribuidos uniformemente dentro de cada intervalo, por lo tanto la fecha de estimación es fija y depende directamente de la fecha de referencia de los puntos de base; en otras palabras, si m es la distancia entre cada uno de los puntos originales, las estimaciones tendrán una distancia de $m/5$ entre cada una; en términos prácticos es una muy buena opción para desagregar datos, no obstante, debe tenerse en cuenta que esta situación también significa una limitante en la posibilidad de obtener cifras para cualquier momento deseado.

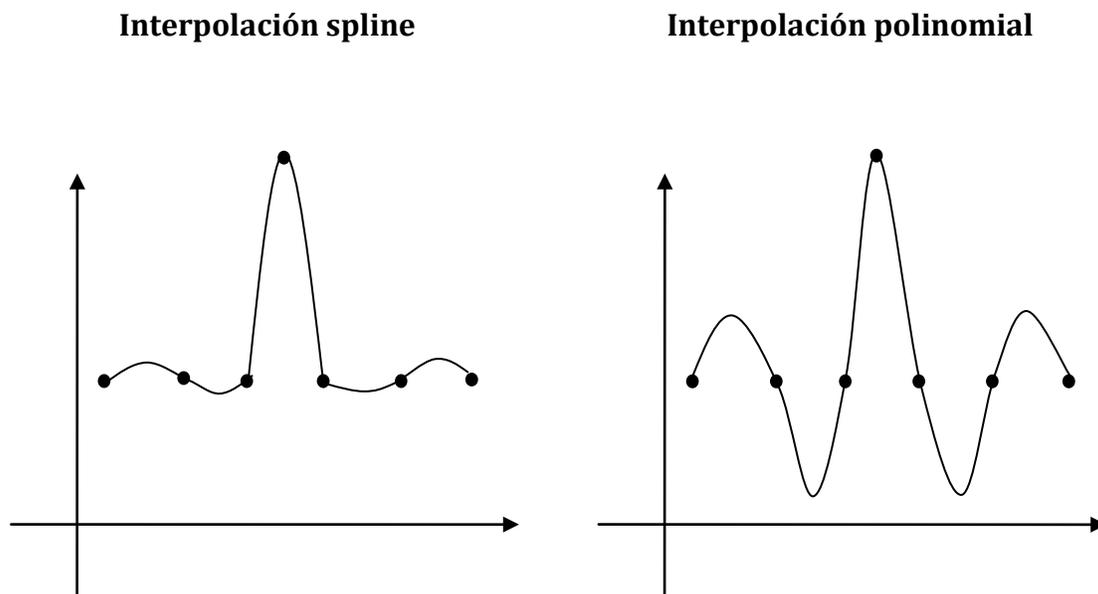
Cabe señalar que existen variantes de este método desarrolladas para lograr desagregaciones distintas, por ejemplo, en mitades, décimas partes, doceavos o aquellas que no requieren de datos equidistantes, sin embargo, la bibliografía para su implementación no necesariamente es inmediata y de alguna manera aunque significan una mejora respecto a la información que ofrecen, conservan la rigidez de las fechas de estimación, motivo principal para omitir su explicación en este documento. La referencia bibliográfica de una de las variantes se encuentra en el apartado respectivo al final de este reporte.

3.5 Interpolación con spline cúbico

Esta técnica al igual que los multiplicadores de Karup-King, consta de un grupo de funciones polinomiales para interpolar un conjunto de puntos conocidos y relacionadas entre sí de forma suave. En términos nada rigurosos, pero sí técnicos, se puede decir que un spline es una función definida por una familia de polinomios sociables, donde el término sociable indica que los polinomios que constituyen una función de este tipo están estrechamente vinculados entre sí.⁷ Aunque no existe alguna restricción respecto del grado de los polinomios a utilizar, los de grado tres o cúbicos suelen ser los más populares en el uso puesto que sus características ofrecen una buena combinación entre estabilidad, flexibilidad y suavidad para el ajuste.

Debe tenerse muy claro que la interpolación cúbica tipo spline es distinta de una interpolación común, ya que en la primera se construye un polinomio de grado n entre cada par de puntos base y se garantiza que la unión de cada uno es continua y sin cambios abruptos en la tendencia; mientras que en la segunda se pretende que solamente una función ajuste al total de puntos originales, de esa manera el grado del polinomio queda en función del número de ellos y aunque en general el comportamiento de los polinomios es suave, también es cierto que conforme se incrementa el grado de la función también aumentan las oscilaciones de la misma, lo que tiene como consecuencia un comportamiento probablemente inconsistente con el del fenómeno que origina los puntos base. En la figura cuatro se ilustra el efecto descrito.

Figura 4



⁷ Barrera, Pablo, Victoria Hernández y Claudia Duran. *El ABC de los splines*. Serie textos. México, Sociedad Matemática Mexicana, 1995. p.3

En términos matemáticos lo que hace un spline es definir un conjunto de funciones continuas y de tendencia suave sujetas a dos condiciones básicas: pasan por puntos conocidos sin importar la distancia entre ellos y la unión entre cada tramo de interpolación es continua.

Hasta el momento se ha utilizado el término *suave* haciendo alusión al hecho de que la función no presenta quiebres, esta propiedad es relevante cuando se trata de interpolación por tramos debido a que cada par de puntos tiene asociada una función distinta, y para tener una solución conjunta la unión de cada uno de esas funciones debe tener garantía de continuidad; matemáticamente hablando esa característica se logra imponiendo condiciones sobre la primera y la segunda derivada.

A continuación se describen algebraicamente el interpolante cúbico y las condiciones que cumple.

Se desea construir una función $P(y)$ que pase por los puntos conocidos $[x_i, f(x_i)]$, donde utilizaremos $g_i = f(x_i)$ para simplificar la notación.

\mathbf{x}	x_1	x_2	x_3	x_4	\dots	x_n
\mathbf{g}	g_1	g_2	g_3	g_4	\dots	g_n

Entonces en principio consideraremos la construcción de una función $P_i(y)$ que pasará por los puntos conocidos $[x_i, g_i]$ y $[x_{i+1}, g_{i+1}]$, por lo tanto deberá cumplir con:

$$\begin{aligned} P_i(x_i) &= g_i \\ P_i(x_{i+1}) &= g_{i+1} \end{aligned} \tag{1}$$

Por otro lado, para garantizar que el comportamiento sea suave, la derivada de la función en cada punto debe estar definida, aunque no sea un valor conocido a priori, en este caso se supondrá un valor s .

$$\begin{aligned} P'_i(x_i) &= s_i \\ P'_i(x_{i+1}) &= s_{i+1} \end{aligned} \tag{2}$$

Así pues, para cada intervalo se deberá construir una función del tipo

$$P_i(x) = g_i q_1(y) + g_{i+1} q_2(y) + s_i \Delta x_i q_3(y) + s_{i+1} \Delta x_i q_4(y)$$

donde $y = \frac{x - x_i}{\Delta x_i},$

$$\Delta x_i = x_{i+1} - x_i,$$

$$x_i \leq x \leq x_{i+1}$$

Entonces q_1, q_2, q_3, q_4 son polinomios de tercer grado, con coeficientes que satisfacen las condiciones (1) y (2) y están definidos como.

$$q_1 = \frac{(x - x_{i+1})^2 [2(x - x_i) + \Delta x_i]}{\Delta x_i^3} \quad q_2 = \frac{(x - x_i)^2 [2(x_{i+1} - x) + \Delta x_i]}{\Delta x_i^3}$$

$$q_3 = \frac{(x_{i+1} - x)^2 (x - x_i)}{\Delta x_i^2} \quad q_4 = \frac{(x - x_i)^2 (x_{i+1} - x)}{\Delta x_i^2}$$

Si se tiene un polinomio de esta forma para cada uno de los $n-1$ segmentos a interpolar, entonces el resultado es una familia de polinomios $\{P_i(x), i=1 \dots n-1\}$

Hasta este punto se ha supuesto que el valor de la derivada es conocido de alguna manera, sin embargo, en la mayor parte de los problemas donde son aplicables los splines, esto no sucede para cada punto, de ahí que su cálculo sea parte de la solución. En este caso, para determinar los valores idóneos lo primero es imponer la condición de que la derivada exista, y para ello se establecen condiciones de continuidad sobre la segunda derivada en cada punto base.

$$P_i''(x) = P_{i+1}''(x) \quad i = 2 \dots n - 1$$

Mientras que en los extremos del spline, que son los únicos puntos donde no coinciden dos polinomios, se establece la pendiente igual a cero, es decir:

$$P_i''(x_1) = P_i''(x_n) = 0$$

Sustituyendo los valores de los polinomios q_i en la expresión de $P(y)$ y derivando dos veces se obtiene la ecuación:

$$P_i''(x) = 6 \frac{g_{i+1} - g_i}{\Delta x_i^3} (x_{i+1} + x_i - x) - 2s_i \frac{2x_{i+1} + x_i - 3x}{\Delta x_i^2} - 2s_{i+1} \frac{2x_i + x_{i+1} - 3x}{\Delta x_i^2}$$

Entonces, la expresión para determinar los valores de la derivada en los puntos desde 2 hasta $n-1$ resulta ser:

$$a_i s_{i-1} + d_i s_i + (1 - a_i) s_{i+1} = b_i$$

donde

$$d_i = 2$$

$$a_i = \frac{\Delta x_i}{(\Delta x_{i-1} + \Delta x_i)}$$

$$b_i = 3 \left[\frac{\Delta x_i}{(\Delta x_{i-1})} \frac{g_i - g_{i-1}}{(\Delta x_{i-1} + \Delta x_i)} + \frac{\Delta x_{i-1}}{(\Delta x_i)} \frac{g_{i+1} - g_i}{(\Delta x_{i-1} + \Delta x_i)} \right]$$

IV. Implementación y resultados

4.1 Argumentos para la selección del mejor método para interpolar

Los criterios que se tuvieron en consideración para decidir el método más adecuado para la estimación de las poblaciones base de las encuestas en hogares que realiza el INEGI se listan a continuación en orden de prioridad:

- Flexibilidad del método en cuanto a la cantidad y características de los insumos requeridos.
- Estabilidad de las estimaciones en términos del comportamiento continuo de las tasas de crecimiento a lo largo del periodo.
- Consistencia de las estimaciones con la información de origen.
- Facilidad de implementación.

Con base en los criterios anteriores se tomó la decisión de utilizar el método de interpolación cúbica como el mejor calificado para la estimación de los totales de población a nivel nacional y para cada área geográfica de interés (entidad federativa, tamaño de localidad y ciudad).

Aunque la interpolación spline no resulta ser el método más sencillo de implementar en comparación con los métodos tradicionalmente utilizados para la estimación de poblaciones, su aplicación se puede llevar a cabo utilizando equipos de cómputo con paquetería básica de oficina, ello facilita en gran medida los procedimientos de cálculo más elaborados (inversión de matrices), esto elimina la complejidad aparente de la metodología. Adicionalmente, se observa que sus resultados muestran mayor consistencia en el comportamiento de las tasas de crecimiento en cada periodo, así como al interior de cada dominio geográfico.

De manera adicional, el hecho de que los insumos para la aplicación observan pocos requisitos iniciales respecto a la cantidad de información y la temporalidad de los mismos, se impuso como una condición deseable y de peso considerable para la selección final del método.

4.2 Enumeración de las etapas para la estimación completa de montos de población en fechas específicas para la ponderación de encuestas

La expansión de encuestas, es un procedimiento que involucra múltiples consideraciones, tanto de cuestiones estadísticas como demográficas,⁸ con la intención de proveer certeza y garantizar la fiabilidad de los resultados del proyecto completo. A continuación se describen las características demográficas primordiales a vigilar durante la estimación de cifras absolutas de población para la ponderación de encuestas.⁹

- En lo que se refiere a las características demográficas de vigilancia, se encuentra el comportamiento regular de la tasa de crecimiento de las cifras de población y apegado a la realidad que se observa a través de información de las fuentes que sirven como insumo, en este caso, la serie anual de proyecciones de población.
- La consistencia entre los resultados del total nacional, por entidad federativa, por tamaño de localidad en cada caso y por ciudad, de tal modo que se conserven las proporciones estimadas en cada caso y que al mismo tiempo las sumas de absolutos se respeten para todos los ámbitos.
- Es importante mencionar que la ponderación de las encuestas se lleva a cabo con base en estimaciones de población total sin distinguir entre características demográficas específicas como edad o sexo, por lo tanto, las estructuras para los rasgos demográficos respetan íntegramente la información colectada en campo.

Así pues, se establecieron las siguientes etapas de estimación y el orden de las mismas, teniendo en cuenta los insumos en cada caso, las coherencias a vigilar en cada nivel y los procedimientos de comprobación de las mismas.

1. Interpolación para cada mitad de mes de los totales de población nacional y por entidad federativa apegada a proyecciones de población utilizando spline cúbico.
2. Ajuste de montos nacionales para la coincidencia de cifra interpolada con la suma de las entidades.
3. Cálculo de la distribución porcentual de la población por tamaño de localidad para el total nacional (cálculo de estructuras por tamaño de localidad a nivel nacional).
4. Cálculo de la distribución porcentual de la población por tamaño de localidad para cada entidad federativa (cálculo de estructuras por tamaño de localidad a nivel entidad).

⁸ En el caso de las encuestas en hogares se involucran supuestos y consideraciones demográficas debido a que la población objetivo son personas con determinadas características relacionadas con la temática del ejercicio estadístico. En el caso de encuestas económicas o de otra naturaleza, los supuestos involucrados están en términos de las características de las unidades de estudio.

⁹ Las características y consideraciones de corte estadístico no son tratadas en este documento debido a que corresponden a una etapa posterior en la que no se tuvo participación directa y cuya descripción rebasa los objetivos de este documento.

5. Estimación de la evolución mensual de las proporciones de los pasos 2 y 3, utilizando el método de estimación de poblaciones con tasa de crecimiento medio constante.
6. Obtención de cifras absolutas por tamaño de localidad para total nacional y por entidad, aplicando las distribuciones porcentuales a los totales por entidad interpolados por spline cúbico.
7. Ajuste de la suma de totales por tamaño de localidad y entidad federativa utilizando tabla cuadrada para lograr coincidencia entre las cifras interpoladas con spline y las estimadas por proporción.
8. Obtención de total nacional a mitad de cada mes por suma de las entidades federativas.
9. La estimación para ciudades retoma los pasos 4 al 8, pero referidos a la población de ciudades seleccionadas.

4.3 Cálculo de los totales mensuales nacionales y por entidad federativa

El objetivo de esta etapa es generar series de montos de población estimada para la mitad de cada mes, para el total nacional y para cada entidad federativa, tomando como base las cifras anuales de las Proyecciones de Población publicadas por el CONAPO y con la condición de que los resultados deberán ser consistentes en términos de la tendencia de la información base, así como tener una tasa de crecimiento sin sobresaltos. A continuación se describen los insumos necesarios para esta etapa:

- Serie de población anual de proyecciones, estimada a mitad de cada año para el total nacional y por entidad federativa, 2000-2012.
- Serie anual de la tasa de crecimiento exponencial acumulada, es decir, calculada para cada punto base tomando como momento de referencia común la primera observación de proyecciones (primero de julio de 2000).

Como ya se explicó en el apartado 3.5 de este documento, la interpolación cúbica tipo spline se puede aplicar sobre casi cualquier conjunto de datos que se desee representar a partir de una curva suave. En el caso particular de la interpolación de poblaciones a mitad de mes, la flexibilidad del método permite la opción de calcular interpolaciones tomando directamente las cifras de población como puntos base, sin embargo, esta forma de cálculo no asegura que la tasa de crecimiento resultante entre los montos estimados tenga el comportamiento deseado, de ahí que se optara por tomar la serie de tasas de crecimiento exponencial acumuladas como base para el ejercicio.

Si bien es cierto que la tasa de crecimiento más comúnmente utilizada para la población de nuestro país es la que asume ritmo geométrico, en este caso se utilizó la tasa exponencial, con la intención de aprovechar la propiedad aditiva de los exponentes, lo cual permite desagregar la tasa de un periodo como la suma de las tasas parciales de subperiodos más pequeños.

En otras palabras, cuando r es la tasa de crecimiento exponencial de la población en el periodo $[t, t+k]$, entonces, si la tasa se puede descomponer en m partes se cumple que:

$$P_{t+k} = P_t * e^{rk} \tag{a}$$

$$= P_t * e^{r_1+r_2+\dots+r_m}$$

donde, $r = r_1 + r_2 + \dots + r_m$

Para el contexto de nuestro problema, se tomaron las tasas de crecimiento acumuladas anualmente como base para la interpolación, por lo tanto, la estimación entre cada par de observaciones proporcionó el crecimiento acumulado para la fecha específica, a partir de ella y en conjunto con la cifra de población en el momento inicial, se pudo calcular la población en el momento deseado.

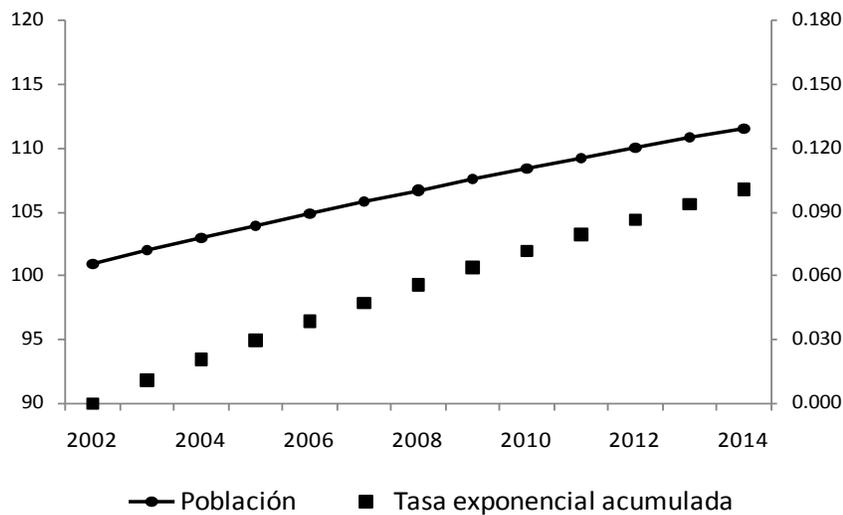
En términos analíticos, considerando que nuestro punto de partida es la población de 2000 y que contamos con una cifra absoluta para cada mitad de año, la fórmula para el cálculo de las tasas acumuladas se obtiene despejando r de la expresión (a)

$$r_k = \frac{1}{k} \ln \left(\frac{P_{2000+k}}{P_{2000}} \right), \quad k = 0 \dots 12$$

Por lo tanto, se generó una serie de puntos base con 13 tasas de crecimiento acumuladas (incluyendo la tasa inicial igual a cero), que sirvió como insumo para la construcción del interpolante cúbico. De acuerdo con la forma de construcción y por las características de la serie, el spline se conformó como una función de 12 tramos y consecuentemente constó del mismo número de conjuntos de coeficientes para los polinomios en cada sección.

Figura 6

Población base y tasa de crecimiento exponencial acumulada

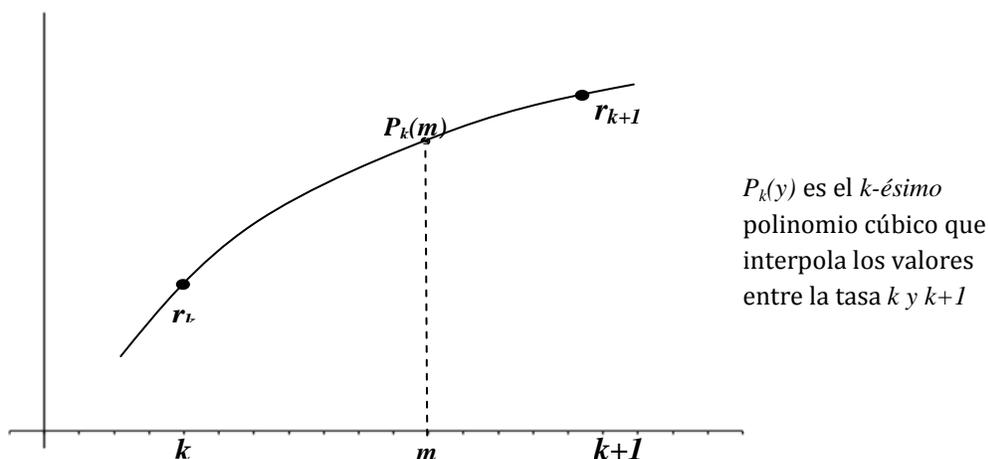


Como ya se expuso en el apartado 3.5, el procedimiento de interpolación es sencillo una vez que se generan los coeficientes del polinomio para cada par de puntos, ya que basta con identificar en qué intervalo se encuentra el punto a estimar y evaluar la función correspondiente en el mismo.

Debido a que los datos se originan a partir de información anual (del 01 de julio de un año a la misma fecha del siguiente), y que el propósito es estimar valores a la mitad de cada mes, entonces, entre cada par de puntos se estiman conjuntos de 12 valores, correspondientes al día 15 de cada uno de los meses desde julio del año t a junio del año $t+1$.

Figura 7

Interpolación de tasas acumuladas con spline cúbico



El resultado del proceso de interpolación es una serie de tasas de crecimiento acumuladas, $P_k(m)$ de acuerdo con la figura 7, no obstante se trata de indicadores de periodo y no anualizados, por lo tanto, para obtener el monto de población en el momento de interés, es necesario anualizar las tasas de crecimiento y posteriormente multiplicar cada una de ellas por el valor de población base. Así, retomando la expresión (a) se tiene:

$$\widehat{P_{k+m}^i} = P_k^i * e^{P(m) * \left\{ \frac{m-k}{365} \right\}}$$

donde $i = 0 \dots 32$

la entidad cero representa al total nacional

Así pues, con este procedimiento es posible estimar de manera independiente los montos de población a mitad de cada mes tanto para el total nacional, como para cada entidad federativa; y dado que la interpolación se hace directamente sobre las tasas, existe garantía de que el comportamiento es suave y regular.

De acuerdo con el orden de las etapas para el cálculo, una vez que se obtuvieron las 33 series mensuales, es necesario realizar un ajuste para lograr la coincidencia de los montos interpolados para el total nacional y los que resultan de la suma de la población por entidad,

ya que aunque la suma era cercana, por el efecto de redondeo en algunos casos mostró diferencias en algunas unidades. En términos de notación, se buscó la congruencia:

$$\widehat{P}_m^0 = \sum_{i=1}^{32} \widehat{P}_m^i$$

Este procedimiento se llevó a cabo con la función buscar objetivo de Excel y vale la pena señalar que las diferencias máximas en este caso fueron de 20 unidades a nivel nacional, por lo tanto, el ajuste no significó cambios en el comportamiento de las tasas de crecimiento.

4.4 Ajuste por tamaño de localidad

La obtención de montos por tamaño de localidad tuvo una mecánica distinta, debido a que al momento en que se llevó a cabo la estimación no se encontraban disponibles proyecciones de población por tamaño de localidad, entonces la información censal fue el único insumo disponible para el cálculo, por lo tanto, solamente se tenían dos puntos de apoyo que además no cubrían el periodo de interés completo, esta razón fue la que hizo inviable la interpolación cúbica para este caso.

A partir de la información del Censo 2000 y del II Conteo 2005, se conformaron los montos de población agrupada en cuatro tamaños de localidad y se calculó la tasa de crecimiento geométrico medio anual para el periodo intercensal, posteriormente se utilizaron dichas tasas (por cada tamaño y entidad) para estimar el monto de población a mitad de cada mes, tanto del periodo intercensal (2000-2005), como para el post-censal (2006-2010).

Con base en las cifras estimadas, para cada mitad de mes se obtuvo la distribución porcentual de la población por tamaño de localidad para las 33 series y ésta se aplicó a la cifra interpolada por spline, el resultado fueron montos de población por tamaño de localidad para el total nacional y por cada entidad federativa.

Entonces, la población por tamaño de localidad estimada a nivel nacional y para cada entidad se obtuvo como:

$$\widehat{P}_m^{i,Tj} = \widehat{P}_m^i * p_m^{i,Tj}$$

$$\text{con } i = 0 \dots 32 ; j = 1 \dots 4$$

Donde

$\widehat{P}_m^{i,Tj}$ es la población estimada de la entidad i en el tamaño de localidad j en el momento m

\widehat{P}_m^i es la población total estimada por interpolación spline para la entidad i en el momento m

$p_m^{i,Tj}$ es la proporción de población de la entidad i en el tamaño de localidad j en el momento m , esta es la que se obtiene de la estimación por tasa geométrica.

De la misma manera que en el apartado 4.3, para cada tamaño de localidad es necesario lograr la coincidencia entre las cifras del total nacional estimadas por cálculo y los totales por suma, solamente que en este caso la coincidencia no es tan inmediata, ya que hay que conciliar dos cifras a la vez, tanto el total de las entidades como el de los tamaños de localidad y para ello se recurrió al método de la tabla cuadrada, también conocido como *ajuste bilineal convergente*.

En los términos de la notación, después de realizar el ajuste se cumplen las condiciones:

para cada tamaño de localidad j se cumple

$$\widehat{P}_m^{0,Tj} = \sum_{i=1}^{32} \widehat{P}_m^{i,Tj}$$

y al mismo tiempo para cada entidad i se cumple

$$\widehat{P}_m^i = \sum_{j=1}^4 \widehat{P}_m^{i,Tj}$$

Para esta etapa la verificación de resultados incluyó la revisión de la tendencia de las tasas de crecimiento geométricas para las poblaciones por tamaño de localidad.

4.5 Adecuación para la estimación por ciudad

La estimación por ciudad se llevó a cabo con la misma metodología descrita en el apartado 4.4 y respetando los montos de población obtenidos en dicha etapa, ello debido a que los insumos también se limitaron a la información censal de 2000 y 2005, no obstante, por la naturaleza de estas unidades la conformación de la población en cada una es distinta, ya que su delimitación obedece a criterios que no necesariamente coinciden con los límites político administrativos de las entidades, entonces, existen unidades que se conforman por localidades de más de una entidad.

Esta última condición significó la adecuación más importante a la forma de ajuste, puesto que el procedimiento para la estimación de crecimiento y distribución a través del tiempo debió considerar el efecto y vigilancia del comportamiento de más de una entidad de manera simultánea. Como resultado de la inclusión de dicha característica, en cada entidad la población se subclasificó de acuerdo con su condición de residencia en una ciudad, en los casos donde en una entidad se ubicó más de un área metropolitana el número de categorías se incrementó en la misma medida.

En este caso, para cada entidad federativa se construyó una tabla cuadrada que procuró la coincidencia entre los montos estimados para el total de la entidad, sus tamaños de localidad y la distribución al interior de cada uno por condición de residencia en la ciudad y fuera de ella.

Así, para cada entidad i se cumple

$$\widehat{P}_m^i = \widehat{P}_m^{i,C} + \widehat{P}_m^{i,FC}$$

para cada tamaño de localidad j en la entidad i se cumple

$$\widehat{P}_m^{i,Tj} = \widehat{P}_m^{i,Tj,C} + \widehat{P}_m^{i,Tj,FC}$$

y para verificar la consistencia para cada categoría de residencia o no dentro de cada ciudad en la entidad i se debe cumplir

para la población que reside en las ciudades

$$\widehat{P}_m^{i,C} = \sum_{j=1}^4 \widehat{P}_m^{i,Tj,C}$$

para la población que reside fuera de las ciudades

$$\widehat{P}_m^{i,FC} = \sum_{j=1}^4 \widehat{P}_m^{i,Tj,FC}$$

Donde:

\widehat{P}_m^i población total estimada en la entidad i

$\widehat{P}_m^{i,C}$ población estimada que reside en alguna ciudad de interés en la entidad i

$\widehat{P}_m^{i,FC}$ población estimada que reside fuera de las ciudades de interés en la entidad i

$\widehat{P}_m^{i,Tj,C}$ población estimada que reside en alguna ciudad de interés en la entidad i , en localidades de tamaño j

$\widehat{P}_m^{i,Tj,FC}$ población estimada que reside fuera de las ciudades de interés en la entidad i , en localidades de tamaño j

La metodología se mantuvo sin mayores cambios (en términos técnicos), no obstante, sí hubo adecuaciones en la conformación inicial de poblaciones para el cálculo de tasas de crecimiento, pues ésta incluyó un paso adicional en el que para las ciudades que involucraron a más de una entidad se calculó la proporción que representó en cada una y su evolución se estimó de manera fragmentada, de tal manera que se respetara el comportamiento de la entidad en su conjunto.

4.6 Resultados

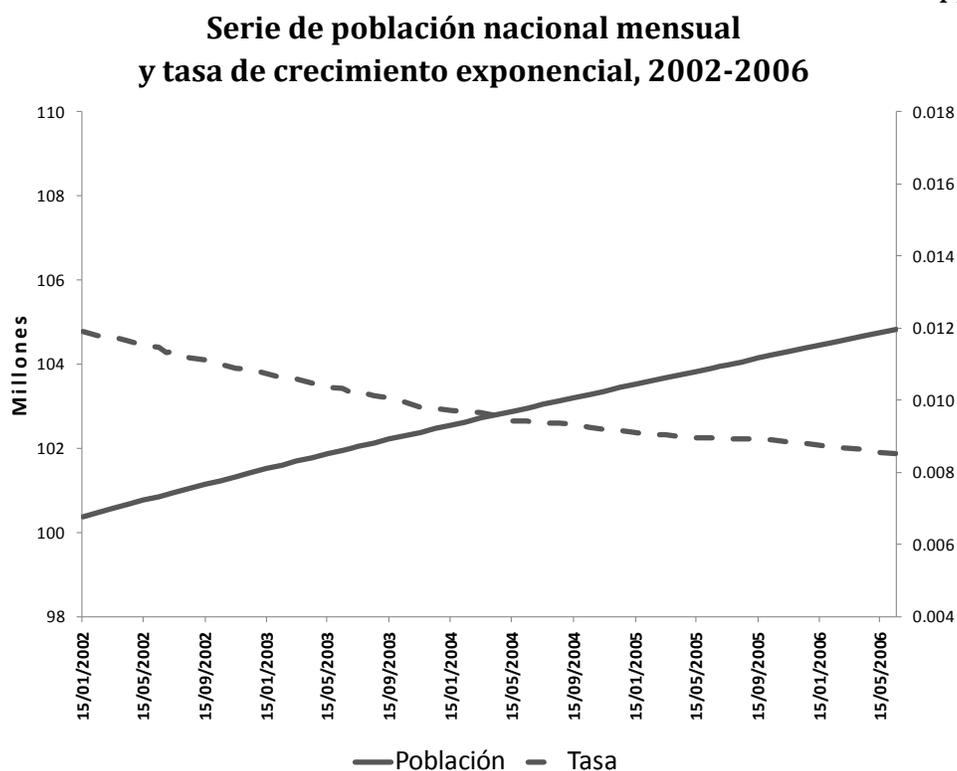
Los resultados finales se resumen en series de población estimada a mitad de mes para el total nacional, para cada entidad federativa, ambas por cuatro cortes de tamaño de localidad y un resultado adicional, donde se encuentra la información por entidad federativa dividida en población residente en la ciudad seleccionada para esa entidad y resto.

Aún cuando la interpolación consideró el periodo completo 2000-2012, es decir, que se generó una cifra estimada para cada mitad de mes en el decenio, no se utilizó la información de los dos primeros y los dos últimos años (2000, 2001, 2011 y 2012), debido a que el spline es una interpolación de tipo osculatoria, entonces, el ajuste es mejor al centro y las desviaciones se incrementan hacía los extremos.

Por cuestiones de espacio y de confidencialidad de la información, las series completas estimadas para cada mes no se incluyen en este documento, sin embargo, sí es posible presentar algunos de los gráficos que evidencian la tendencia y comportamiento de la población estimada y su tasa de crecimiento.

En la figura número 8, se ilustra la serie de valores de población interpolados para el total nacional, así como la tasa de crecimiento exponencial mensual, ambas series correspondientes al periodo 2002-2006 referentes a la primera serie solicitada para reponderación de encuestas con resultados ya publicados.

Figura 8



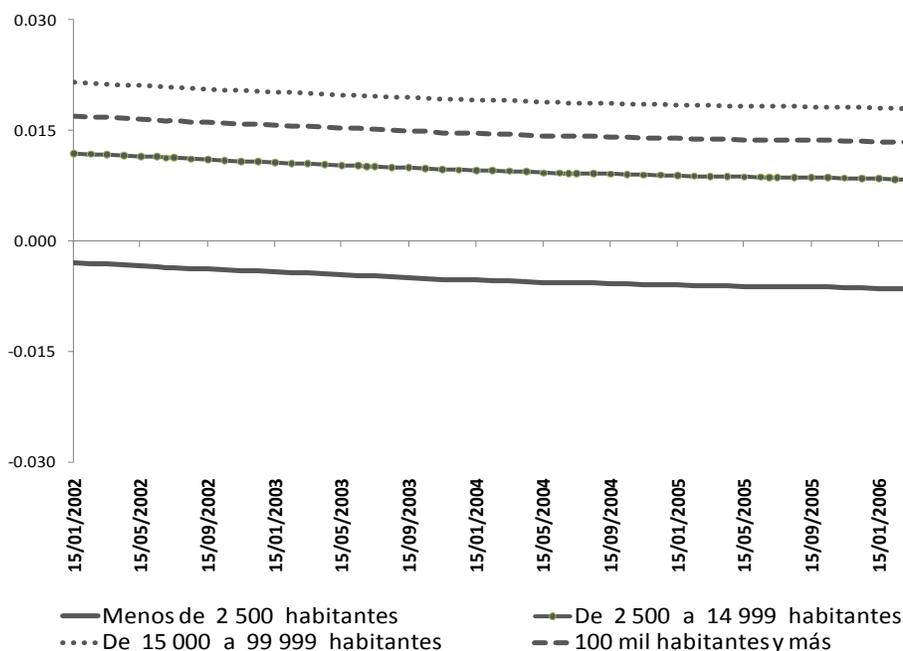
En cuanto a las cifras de población, es evidente que las estimaciones son estables y respetan la tendencia monótona creciente de la serie de proyecciones que sirvió de base, en este sentido, la interpolación reprodujo de manera fiel los montos de población a la mitad de cada año y los valores intermedios no tuvieron sobresaltos o cambios bruscos.

Por su parte, la tasa de crecimiento se observa aceptablemente suave y sin cambios abruptos, aunque también debe señalarse que muestra oscilaciones pequeñas en algunos puntos, ya que aún cuando al aplicar la interpolación spline la gráfica de las tasas fue completamente suave, las variaciones son efecto del redondeo aplicado al momento de obtener poblaciones a partir de los valores interpolados, pues fue necesario redondear a enteros por la naturaleza de las cifras, ya que no tiene sentido presentar fracciones de persona.

Las tasas por tamaño de localidad (figura 9), consistentemente replican la tendencia de la serie total, con la diferencia de que ocurren con distinta intensidad en cada caso. En este sentido, solamente la tasa de localidades rurales observa tasa negativa durante el periodo.

Figura 9

Tasa de crecimiento exponencial nacional mensual según tamaño de localidad, 2002-2006



En las figuras 10 y 11, se ilustra el porcentaje que cada tamaño de localidad representa del total y su evolución en el periodo. Para los cuatro tamaños se observa como regular y consistente con el comportamiento de las observaciones censales que le dieron origen. Cabe señalar que la separación obedece a la mejora en la visualización de las tendencias y no a criterios relacionados con el grado de urbanización.

Figura 10

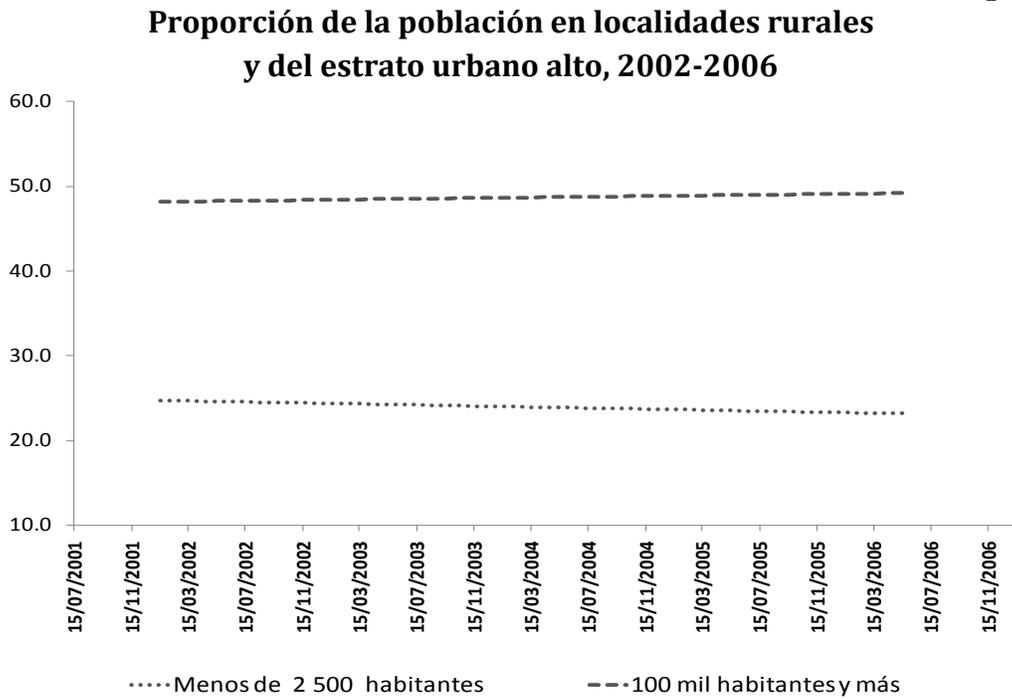
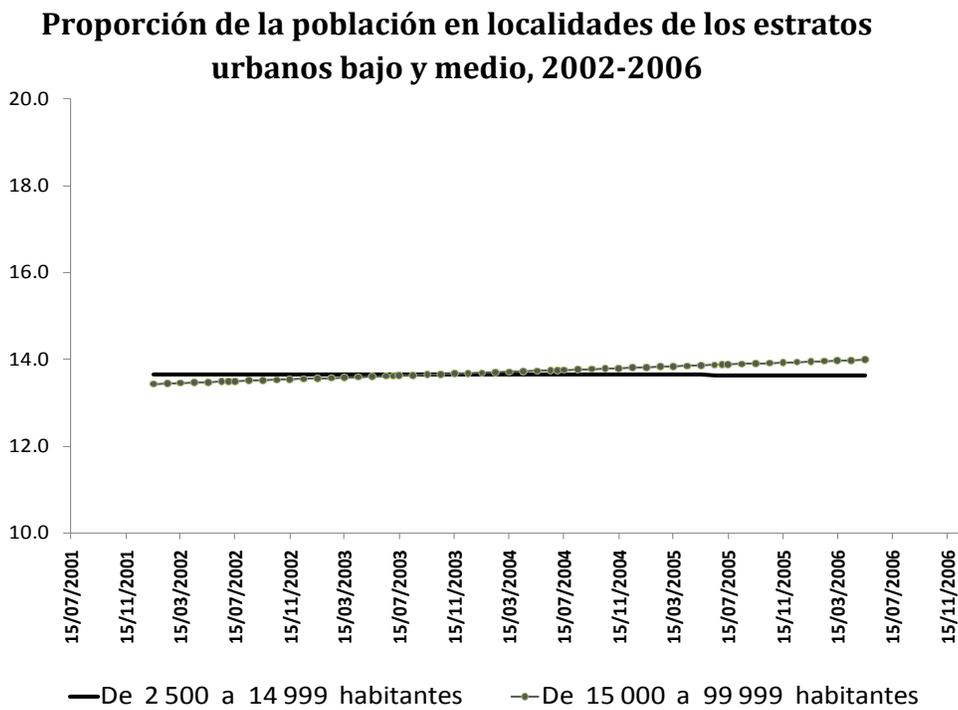


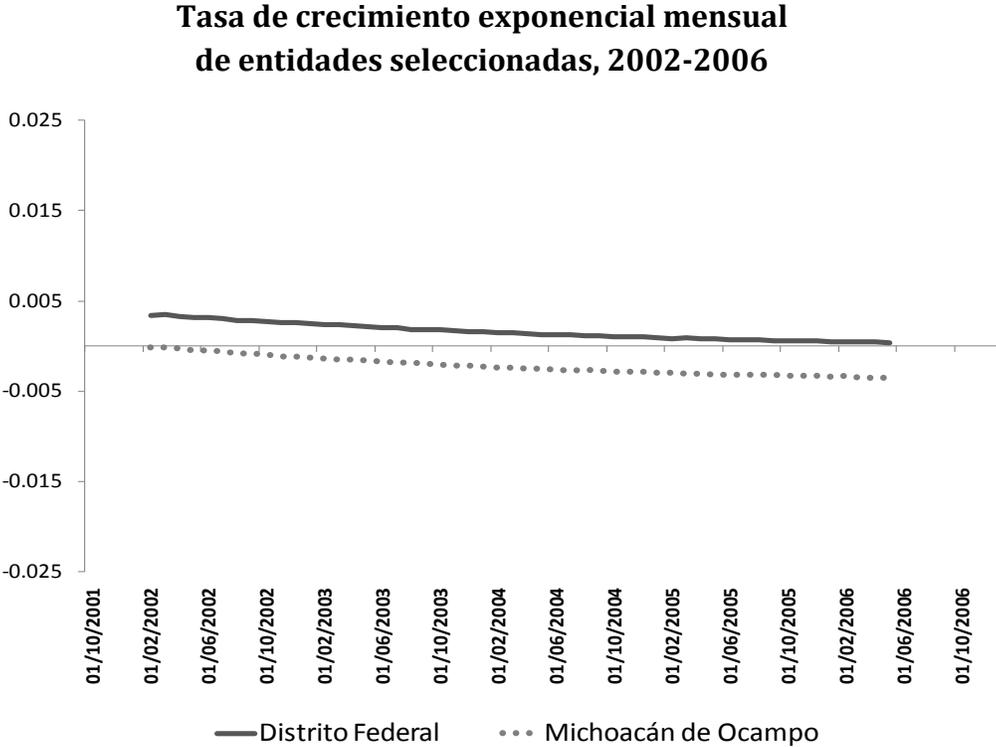
Figura 11



La revisión de las estimaciones por entidad federativa se enfocó prácticamente en la coincidencia con los montos de proyecciones del CONAPO para cada mitad de año y en que la tasa de crecimiento no observara comportamientos irregulares en el periodo de referencia. Así pues, algunos de los resultados por entidad también presentaron pequeñas fluctuaciones en la tasa de crecimiento debidas al redondeo y fue mucho más evidente en el caso de las entidades con poblaciones más pequeñas.

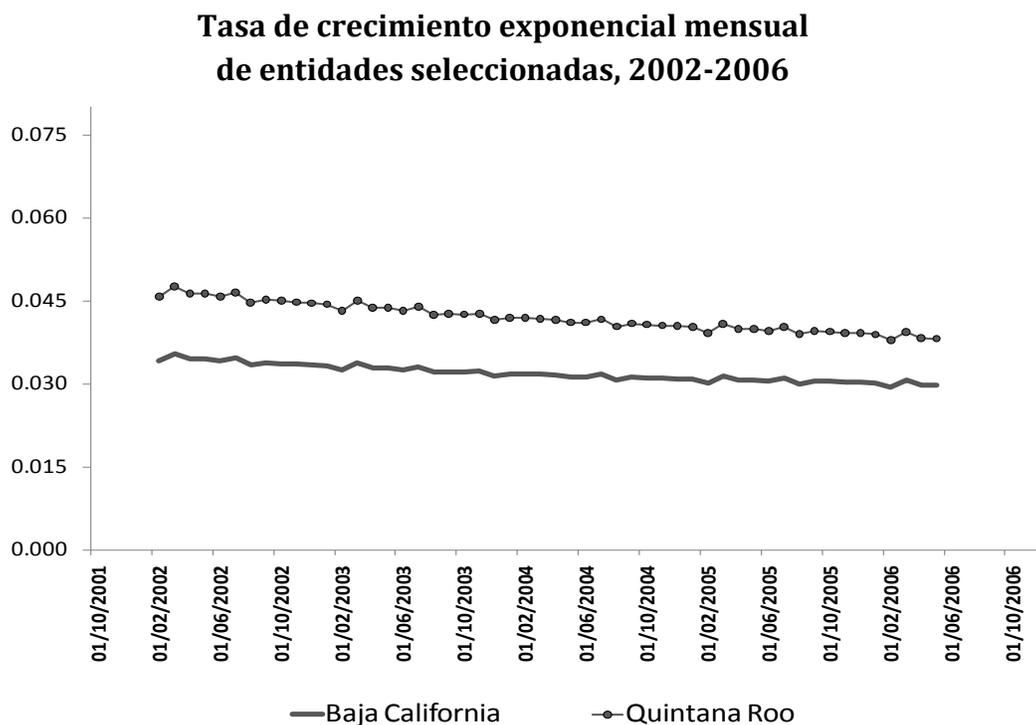
Los gráficos 12 y 13 ilustran la situación mencionada, en el caso de entidades con poblaciones superiores a 4 millones como el caso de Michoacán y el Distrito Federal, las tasas de periodo resultantes tienen un comportamiento bastante suave y sin sobresaltos, coincidente con la tendencia de los valores base provenientes de proyecciones.

Figura 12



Conforme el monto de población se reduce, entonces, la tasa de crecimiento se vuelve más sensible a ganancias o pérdidas de población, de ahí que entidades como Baja California y Quintana Roo presenten una serie con oscilaciones visibles en algunas fechas, ligeramente más evidentes en el caso de la entidad del sureste de nuestro país, ya que su población no rebasa el millón de habitantes.

Figura 13



Por otro lado, vale la pena señalar que los controles de tendencia en la información a nivel de área metropolitana se aplicaron únicamente sobre los montos de la población que forma parte de la ciudad en cuestión, no así para la población residente en localidades que no forman parte de alguna de las ciudades de interés, que a esta porción en el ejercicio se denominó como *resto*, por lo tanto, es altamente probable que la tasa de crecimiento de esta categoría observe irregularidades mayores a las que se suscitan por efecto del redondeo, aún cuando los montos de cada ciudad estén acotados por los estimados para la entidad que las contiene.

Como se puede observar en la figura 14, las tasas en ciudades tienen tendencias estables, independientemente si se trata de áreas metropolitanas contenidas completamente en una entidad (Guadalajara y Monterrey) o si involucran a dos o más entidades (Distrito Federal), sin embargo, en ambos casos los montos en este nivel de desagregación son mucho más susceptibles de oscilaciones.

Figura 14



V. Conclusiones

En el año 2010 se colectó la información demográfica básica de poco más de 112.3 millones de personas en nuestro país, eso significó que por su población, México ocupa el lugar número 11 en el mundo y el tercer lugar respecto de los países de América Latina. Una población de tal magnitud implica retos del mismo tamaño y el primero de ellos es llevar una contabilidad eficiente de la totalidad de sus elementos, identificar las situaciones que cada uno vive y el impacto que éstas tienen en los distintos ámbitos en los que se desarrollan.

Disponer de cifras de población actualizadas, continuas, oportunas, desagregadas para todos los ámbitos geográficos posibles y por características demográficas específicas, es un logro alcanzado solamente en algunos países desarrollados, en donde se ha podido crear, consolidar y mantener un sistema coordinado y eficiente de registros administrativos.

En el caso de México, existen esfuerzos importantes para cubrir las necesidades básicas de información estadística, así como para mejorar la calidad de los datos que se generan sistemática y tradicionalmente a partir de distintas fuentes; en este sentido, la búsqueda continua de la mejora en los procedimientos de origen constituye una parte importante de dichos esfuerzos.

Sin importar el tipo de proyecto de generación de información estadística que se implemente, es de vital importancia garantizar una cobertura adecuada de los distintos temas y ámbitos geográficos, la calidad de los datos obtenidos y la disponibilidad de los mismos con oportunidad, ya que su objetivo principal es atender a las necesidades de los usuarios, teniendo en consideración los distintos fines que cada uno tiene de acuerdo con el sector en el que se desempeña.

En el caso particular de la información de carácter sociodemográfico, es importante procurar la consistencia entre los resultados derivados de censos, encuestas en hogares y de los registros administrativos, de manera que los datos obtenidos de cada fuente se complementen entre sí para optimizar el análisis conjunto y que a su vez coadyuven a subsanar los vacíos de información, así como a la generación de indicadores necesarios para el estudio de la población y los fenómenos que la afectan.

Las encuestas en hogares son la herramienta recurrente para la obtención de información especializada en los temas de población y complementan a las otras fuentes de datos, sin embargo, por su naturaleza requieren de cifras de control para afinar su diseño, así como para la evaluación de sus resultados y para ampliar la oferta de información que proveen.

De la misma manera que el resto de los procedimientos intrínsecos en la realización de una encuesta, la calibración de los factores de expansión tiene gran relevancia debido a que es una parte importante del ajuste por no respuesta. Debido a que el insumo principal de esta fase son las cifras de población estimada en los momentos de levantamiento y para los ámbitos geográficos de representatividad en cada caso, es importante que dichos datos sean consistentes con la dinámica demográfica del país, de modo que la representatividad del ejercicio completo esté garantizada y sobre todo, que la información de la encuesta no se vea afectada.

Ante la imposibilidad de contar con cifras absolutas observadas de población continuas o con una periodicidad menor a la de un evento censal, la existencia de una variedad de metodologías para la estimación de montos en momentos específicos es una situación ventajosa en términos de elegir la opción que se ajuste mejor a las necesidades de cada proyecto, e incluso que sea la más adecuada conforme a las características de los insumos para el cálculo.

Con referencia específica en la metodología aplicada para la estimación de poblaciones para la expansión de encuestas de INEGI, la interpolación cúbica tipo spline es un método flexible, bonito y barato, no obstante, su uso no puede ser completamente indiscriminado, ya que deben tenerse algunas precauciones para conservar la estabilidad del método y la consistencia de los resultados.

En particular es recomendable no interpolar un gran número de puntos entre cada intervalo, ya que por su naturaleza la función cúbica spline puede curvarse demasiado, de tal modo que el comportamiento de los datos estimados ya no sea consistente con el fenómeno al que se ajustan; de la misma manera, aunque no existe una restricción en cuanto al espaciamiento entre puntos base, las estimaciones resultarán con menos oscilaciones cuando las distancias entre las observaciones originales sean cortas.

En términos generales, las series de población estimadas para cada desagregación geográfica, respecto de su tasa de crecimiento, presentan comportamiento estable, suave y apegado en lo más posible al conjunto de datos que sirven de base. Por entidad federativa, los resultados conservan el apego a la información observada y aún en este nivel se puede afirmar que no existen cambios abruptos en la tendencia que deriven en complicaciones para la aplicación del método.

Conforme se reduce el dominio geográfico, las poblaciones de referencia son más pequeñas, por lo tanto más susceptibles a presentar volatilidad en el comportamiento de su tasa de crecimiento, ya que el incremento o disminución de su población en pocas unidades significa un cambio significativo en la intensidad con la que ocurre, así pues, la condición primordial de

estabilidad en la tasa de crecimiento reduce los grados de libertad para el ajuste e incrementa el grado de dificultad para la obtención de estimaciones con las características deseadas.

Si bien es cierto que la consecución del resultado final de este ejercicio involucra la implementación de algunos procedimientos adicionales a la interpolación cúbica, debe tenerse en cuenta que los datos obtenidos a partir de este procedimiento son la base y el tope para el ajuste de las estructuras en dominios geográficos más pequeños, de ahí la relevancia de su buen cálculo.

Aunque las estimaciones finales se pueden calificar como *adecuadas*, definitivamente la metodología es susceptible de mejorar, principalmente en los mecanismos de ajuste en áreas geográficas pequeñas, particularmente el ajuste por tamaño de localidad y por ciudad, ya que en este apartado hay mucho trabajo por hacer para afinar la estimación, puesto que estos dominios tienen una dinámica de crecimiento muy distinta a la de las entidades o de la población nacional, situación que no se reproduce fielmente en las estimaciones actuales debido a su forma de construcción, la información disponible e incluso por la falta de un marco conceptual completamente delimitado sobre el tema.

El orden de estimación se estableció partiendo de dominios grandes y hacia dominios más específicos, debido a que no hay fuentes disponibles, oportunas y confiables para determinar la dinámica de cambio de las ciudades o más aún de las localidades en sí. En particular, las últimas muestran un ritmo de cambio hasta cierto punto inestable, ya que se crean, se fusionan o desaparecen sin un patrón de comportamiento predecible, en muchas ocasiones con dependencia completa de criterios arbitrarios, lo cual no provee de los elementos necesarios para comprometer su correcta estimación, por lo tanto, en este ejercicio fue preferible acotar sus tendencias a la de las entidades.

En nuestros días, México todavía no cuenta con un sistema de registros administrativos infalible, veraz y oportuno, que provea de toda la información que requieren los distintos sectores. Ante esta imposibilidad se recurre a la estimación con el propósito ideal de que los datos de población que se producen en forma retrospectiva y prospectiva, garanticen certeza absoluta en los comportamientos resultantes de los ejercicios; sin embargo, también debe tenerse presente que la tendencia de los factores involucrados depende, en la mayoría de los casos, de situaciones no previsibles y no controlables, lo que se convierte en la fuente potencial de la imprecisión de los pronósticos.

Así, aunque no tenemos forma de prever los fenómenos sociales o naturales de gran escala y su efecto en la dinámica y crecimiento de la población, lo que sí podemos hacer es conocer y revisar la información disponible e incorporarla en las metodologías, de tal manera que su inclusión provea elementos de la realidad de un fenómeno y que compense a los supuestos que por lo general se involucran en la resolución de distintos problemas.

Es importante destacar la importancia que tiene la vinculación entre las problemáticas reales en distintas áreas de aplicación y la técnica que se desarrolla en la academia, ya que en los dos ámbitos es muy poco el panorama que se tiene del trabajo que se realiza en el otro, aún cuando las necesidades son evidentes.

Por último y como una opinión personal, es deseable que la formación académica incluya talleres de aplicación, pláticas o coloquios que acerquen a los estudiantes con los problemas reales que se enfrentan en los distintos campos donde se podrán desempeñar, de tal manera que desde su formación el profesionista adquiera conciencia de las áreas donde puede participar y aportar soluciones (más allá de los campos tradicionales para un actuario), también que valore la relevancia que tiene la técnica que adquiere a lo largo de la licenciatura, que identifique un enfoque más específico de sus habilidades (sin que ello sea restrictivo en su formación posterior) y que su inserción laboral sea más orientada.

Glosario

Dominio geográfico. Delimitación territorial político administrativa a la que hace referencia un dato, un indicador o un conjunto de los mismos.

Interpolación. Es la acción de estimar valores intermedios en una serie de datos conocidos utilizando fórmulas matemáticas o procedimientos gráficos.

Interpolación cúbica. Es la acción de interpolar datos utilizando polinomios de tercer grado como fórmulas de aproximación.

Interpolación tipo spline. Es la acción de interpolar datos utilizando familias de polinomios sociables, donde el grado de los polinomios define el grado del spline.

Marco muestral. Lista de unidades o elementos susceptibles de muestreo.

Muestra. Es el conjunto de unidades o elementos de análisis seleccionadas de uno o varios marcos muestrales.

Población objetivo. Conjunto de individuos con características bien definidas de los cuales se requiere obtener información.

Ponderación de una encuesta. Es la definición del número de unidades de la población que están representadas por la unidad en muestra. Donde las estimaciones básicas se ajustan para compensar la ausencia de respuesta y de cobertura para hacer que las estimaciones muestrales ponderadas correspondan a los totales conocidos de la población.

Tasa de no respuesta. Representa el porcentaje de viviendas que por distintos motivos no es posible visitar y recabar información, ello garantiza un número suficiente de entrevistas de modo que no se vea afectado el cálculo de los indicadores base del ejercicio.

Unidades de análisis. Objeto o individuo sobre el que se capta la información de interés para el estudio.

Unidades de muestreo. Conjunto de elementos no traslapados de la población que cubren a la población completa. Todo miembro de la población corresponderá a una y sólo una unidad de selección.

Bibliografía

Barrera, Pablo, Victoria Hernández y Claudia Duran. *El ABC de los splines*. Serie textos. México, Sociedad Matemática Mexicana, 1995.

Crofts, Geoffrey. "A Karup-King Formula with Unequal Intervals" en: *Actuarial Research Clearing House*, Vol. 1.1998, pp. 299-303.

Haupt, Arthur y Thomas T. Kane (2003). *Guía rápida de población*, 4ª edición. Washington, D.C, Population Reference Bureau.

Instituto Nacional de Estadísticas y Censos (INDEC). *Ponderación de la muestra y tratamiento de valores faltantes en las variables de ingreso en la Encuesta Permanente de Hogares*. Metodología No 15. Argentina, INDEC.

Instituto Nacional de Estadística y Geografía (INEGI) (2007). *Cómo se hace la ENOE, Métodos y procedimientos*. Consulta 2011 en:
www.inegi.org.mx/est/contenidos/espanol/metodologias/enoe/ENOE_como_se_hace_la_ENO_E1.pdf

Instituto Nacional de Estadística, Geografía e Informática (INEGI). *ENOE. 50 Preguntas y Respuestas*. Consulta 2011 en:
www.inegi.org.mx/est/contenidos/espanol/metodologias/encuestas/hogares/50_preguntase_noe.pdf

Instituto Nacional de Estadística, Geografía e Informática (INEGI) (1997). *Manual de Medidas Sociodemográficas*. México, INEGI.

McNeil, Donald, James Trussell and John C. Turner. "Spline Interpolation of Demographic Data" en: *Demography*, Vol. 14, No. 2. Population Association of America, 1977, pp. 245-252.

Organización de las Naciones Unidas (ONU). *Encuestas de hogares en los países en desarrollo y transición*. Estudios de métodos, Serie F No. 96. Nueva York, ONU, 2007.

Richard L . Scheaffer, William Mendenhall, Lyman Ott. *Elementos de muestreo*, 6a edición. México, Thomson, 2007.

Shriock, Henry S, Siegel Jacob S. and Associate. "Osculatory Interpolation", en: *The Methods and Materials of Demography*, Volume 2, second printing. Washington, D.C, Bureau of the Census, 1973, pp. 687-724.

Welti, Carlos (ed.) (1997). *Demografía I*. México, Programa Latinoamericano de Actividades en Población-IISUNAM-CELADE.