

Indicadores Laborales para los **Municipios de México**

ILMM

Metodología de estimación
en áreas pequeñas



Instituto Nacional de Estadística y Geografía

Indicadores Laborales para los Municipios de México

ILMM

**Metodología de estimación
en áreas pequeñas**



Obras complementarias publicadas por el INEGI sobre el tema:

Indicadores laborales para los municipios de México 2017. Estimación en áreas pequeñas; Indicadores laborales para los municipios de México 2018. Estimación en áreas pequeñas; Indicadores laborales para los municipios de México 2019. Estimación en áreas pequeñas; Indicadores laborales para los municipios de México 2020. Estimación en áreas pequeñas; Indicadores laborales para los municipios de México 2021. Estimación en áreas pequeñas.

Catalogación en la fuente INEGI:

331.2021 Instituto Nacional de Estadística y Geografía (México).
Indicadores Laborales para los Municipios de México : ILMM : metodología de estimación en áreas pequeñas / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2022.

viii, 34 p. : il.

1. Trabajo - Estadísticas - México.

Conociendo México

800 111 4634

www.inegi.org.mx

atencion.usuarios@inegi.org.mx



INEGI Informa



@INEGI_INFORMA

DR © 2023, **Instituto Nacional de Estadística y Geografía**

Edificio Sede

Avenida Héroe de Nacozari Sur 2301

Fraccionamiento Jardines del Parque, 20276, Aguascalientes,

Aguascalientes, Aguascalientes, entre la calle INEGI,

Avenida del Lago y Avenida Paseo de las Garzas.

Presentación

El **Instituto Nacional de Estadística y Geografía (INEGI)** ha tenido la necesidad de realizar diversas tareas encaminadas a la búsqueda de nuevas opciones técnicas y metodológicas para fortalecer la infraestructura estadística con información que facilite la toma de decisiones en torno a la planeación, el diseño y la evaluación de programas sociales. Por un lado, se tiene el propósito de responder de forma eficiente y eficaz a la creciente demanda de información estadística bajo exigencias de oportunidad, confiabilidad y comparabilidad, pero con presupuestos generalmente acotados, que obligan a cambiar los esquemas de trabajo previstos para poder costear los recursos materiales, humanos y las operaciones de los proyectos. Por otro lado, se debe responder a las tendencias de la dinámica mundial que implican una continua innovación de las técnicas y las formas para registrar la realidad.

La demanda creciente de la sociedad para tener respuestas satisfactorias a sus necesidades de información a través de las oficinas nacionales de estadística se ha convertido en una constante universal. Particularmente, los gobiernos locales requieren de información actualizada con mayor nivel de desagregación que los considerados en los proyectos de generación de información mediante encuestas nacionales por muestreo. Tales niveles de desagregación pueden ser subregiones geográficas, como municipios¹ o entidades federativas (en el caso de diseños para cifras nacionales), o bien, dominios temáticos que originalmente no se consideran en el diseño muestral.

Lograr estimaciones confiables, mediante encuestas, para niveles locales que permitan generar un análisis descriptivo y líneas base de indicadores, requiere ampliar las muestras, lo que conlleva el aumento en los costos de los proyectos, situación que difícilmente puede ser subsanada por los organismos nacionales de estadística.

Con este proyecto se busca generar **indicadores laborales** de la **población económicamente activa**, la **población ocupada** y la **población ocupada informal** a nivel de cada uno de los municipios del México, mediante técnicas de estimación en áreas pequeñas, para fortalecer el análisis estadístico subnacional y ampliar la oferta de información derivada de la **Encuesta Nacional de Ocupación y Empleo Nueva Edición (ENOE^N)**.

¹ A lo largo de este documento se usará la palabra municipio (y sus palabras derivadas, tal como municipales) para referirse tanto a los municipios de los 31 estados como a las 16 demarcaciones territoriales de la CDMX.

El propósito de este documento es dar a conocer la metodología de estimación de los ILMM y el proceso que se lleva a cabo para su obtención, además de describir la importancia que tienen estos indicadores al ser únicos en su tipo para la toma de decisiones. De esta manera, se da transparencia al quehacer del INEGI.

Índice

Introducción	VII
1. Los ILMM en el marco de la Dirección General de Estadísticas Sociodemográficas	1
2. La estimación en áreas pequeñas	3
2.1 Qué son las áreas pequeñas	3
2.2 Las metodologías de estimación en áreas pequeñas	3
3. Proceso para la generación de las estadísticas en áreas pequeñas	7
4. Fuentes de información, base conceptual y base metodológica	9
4.1 Fuentes de información	9
4.2 Clasificación de la población de 15 y más años de edad, según características laborales	10
4.3 Variables de los modelos	10
4.4 Predictores y modelos de estimación en áreas pequeñas	12
5. Preparación de información	19
5.1 Ajustes de los factores de expansión municipales	19
5.2 Matriz de distancias	19
6. Construcción de los modelos	25
6.1 <i>Software</i>	25
6.2 Procesamiento	25
7. Consecuencias	29
Bibliografía	31

Introducción

Se estiman las tasas de las siguientes características laborales: población económicamente activa (PEA), población ocupada (ocupados) y población ocupada informal (informales) para los municipios de México. Esto se realiza mediante técnicas de estimación en áreas pequeñas dado que, en la fuente especializada en la temática, la Encuesta Nacional de Ocupación y Empleo Nueva Edición (ENOE^N), no existe información con tal nivel de desagregación geográfica. Para ello, se utiliza información combinada de distintas fuentes, a partir de las cuales se construyen modelos estadísticos en los que se toman como variables dependientes a las tasas de la PEA, los ocupados y los informales, cuya información proviene del primer trimestre del año referido en dicha encuesta. Las variables auxiliares seleccionadas se obtienen de los registros administrativos del Instituto Mexicano del Seguro Social (IMSS), del Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE) y de la Secretaría de Educación Pública (SEP); así como de las bases de datos de los censos de población y vivienda.

Para cada variable dependiente, es decir, para cada variable objeto de estimación, se construyen sistemáticamente varios modelos, de los cuales es seleccionado el mejor (conforme a criterios estadísticos de calidad), y se aplican para obtener las estimaciones correspondientes. Se comprueban los supuestos del modelo y los resultados son validados al contrastarlos con las estimaciones directas de la propia encuesta para aquellos municipios con muestra suficiente, según su coeficiente de variación (CV), y con los resultados de aplicar la estructura porcentual de los censos (EPORC, para el caso de la PEA y los ocupados), a las cifras proporcionadas por la ENOE^N a nivel de entidad federativa.

A través de procedimientos de validación, se concluye que los resultados obtenidos por las técnicas de estimación en áreas pequeñas son plausibles con la realidad. De esta forma, los datos que se obtienen a nivel municipal para las variables de interés complementan la información de la fuente primaria, la ENOE^N, que proporciona cifras nacionales, por entidad federativa y para ciudades consideradas importantes por su dinamismo económico.

La razón por la cual se elige el primer trimestre de cada año es por el tiempo de referencia entre las diferentes fuentes de información y la oportunidad de liberación de los indicadores estratégicos. Hay registros administrativos (variables auxiliares) que son importantes, por su poder explicativo, para estimar los indicadores laborales de interés, pero que son liberados anualmente, es decir, al final del año (31 de diciembre). De manera que la fecha más próxima

de coincidencia con la ENOE^N es justamente el primer trimestre (a la mitad del periodo, 14 de febrero). Adicionalmente, es importante contar con información a principio de año para fines de planeación y diseño de la política social, de los tomadores de decisiones y del usuario en general.

Este documento se divide en siete capítulos. En el primero se explica el origen de los ILMM y su ubicación en el quehacer institucional. En el segundo se da a conocer qué son las técnicas de estimación en áreas pequeñas y los distintos métodos, así como los modelos matemáticos para su cálculo. Este capítulo permite adentrarse en el tema, de manera general, a cualquier usuario que desconozca estas técnicas.

En el apartado tres se explica, también de manera general, el proceso a través del cual se obtienen los ILMM; es resultado de aplicar la normativa del Instituto, pero al mismo tiempo, concuerda con las sugerencias que han formulado los expertos en técnicas de estimación en áreas pequeñas a nivel mundial. El cuarto presenta las fuentes de información que se utilizan para lograr las estimaciones de los ILMM y las variables involucradas en su cálculo, así como los modelos utilizados para generar las estimaciones. Mientras tanto, en el quinto se explica el tratamiento que se da a la información que sirve como insumo antes de ser utilizada.

El capítulo seis profundiza más la parte del proceso que implica la construcción de los distintos modelos y la selección del mejor, para cada variable de interés. Finalmente, en el apartado siete se da una reflexión respecto de la importancia que tiene este producto para la sociedad mexicana.

1. Los ILMM en el marco de la Dirección General de Estadísticas Sociodemográficas

El INEGI atiende la necesidad de información estadística para niveles de desagregación menores que los considerados en los proyectos de generación de información mediante encuestas nacionales por muestreo, bajo exigencias de oportunidad, confiabilidad y comparabilidad.

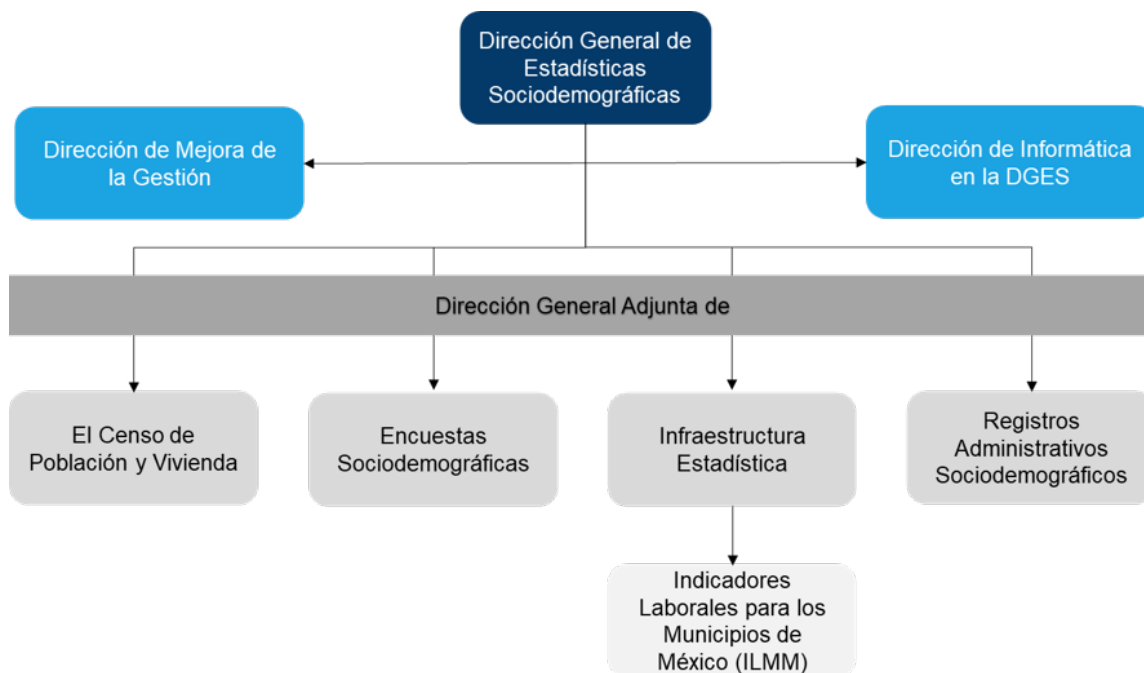
La ENOE^N es una de las encuestas más importantes levantadas por el Instituto; proporciona estimaciones laborales para el país, referidas a cada una de las entidades federativas y para algunas ciudades específicas; sin embargo, existe la necesidad y la demanda de proporcionar información a nivel municipal. Es así como surgen los Indicadores Laborales para los Municipios de México, que en un principio fueron enmarcados en el ámbito de Estadísticas Experimentales, pero luego de la publicación de los indicadores correspondientes a cinco años (2017- 2021), y una vez adquirido el grado de madurez en su producción, dejan de ser parte de este ámbito y a partir del año 2022 pasan a formar parte de los proyectos consolidados y regulares a cargo de la Dirección General de Estadísticas Sociodemográficas (DGES).

La DGES es una de las ocho direcciones generales del INEGI. Esta dirección se compone de cuatro direcciones generales adjuntas, siendo una de ellas la Dirección General Adjunta de Infraestructura Estadística (DGAIE). Los ILMM se encuentran dentro del quehacer de la DGAIE. Ver la ilustración 1.

El conjunto de productos regulares a cargo de la DGES se diversifica al incluir a los ILMM. Además de la estadística básica que se produce mediante censos, encuestas y registros administrativos, se abre paso a la publicación regular de estadística derivada con los tres indicadores laborales mencionados.

Organigrama de la DGES

Ilustración 1



Fuente: INEGI.

El proyecto de estimación de áreas pequeñas surge a partir de la demanda creciente observada en el sitio de internet², que el Instituto pone a disposición de la población usuaria para que formule por escrito sus requerimientos de información. De ahí se desprende la necesidad de información a nivel subnacional de áreas locales para tomar decisiones de diseño y formulación de políticas de manera focalizada. Esto —aunado a la fuerte tendencia mundial por el uso de las técnicas de estimación de áreas pequeñas, para la obtención de indicadores estratégicos de alta precisión mediante la combinación de fuentes estadísticas básicas, sin tener que generar fuentes adicionales de información que implican el consumo de enormes recursos para su operación— fue lo que motivó a generar estadística derivada.

² La siguiente liga te dirige al sitio de solicitud de información estadística y geográfica del INEGI: <https://www.inegi.org.mx/inegi/contacto.html>

2. La estimación en áreas pequeñas

2.1 Qué son las áreas pequeñas

Las áreas pequeñas son subconjuntos poblacionales de tamaño inferior al considerado en el diseño original de una encuesta por muestreo probabilístico, pueden ser áreas geográficas o dominios temáticos no considerados explícitamente. De esta manera, si una encuesta no fue diseñada para proporcionar estimaciones a nivel municipal, como es el caso de la ENOE^N, entonces los municipios constituyen áreas pequeñas; o bien, dado que esta encuesta proporciona datos sobre las características laborales para la población de 15 y más años de edad, al realizar estimaciones sobre grupos específicos de edad, se estaría también adentrando en el campo de las áreas pequeñas.

Para el caso de la encuesta de referencia, las entidades federativas no pueden ser clasificadas como áreas pequeñas, pues su diseño se realiza para obtener estimaciones directas confiables para cada una de ellas (lado izquierdo de la ilustración 2). En cambio, como la encuesta no es diseñada para generar indicadores laborales para los municipios, estos sí son áreas pequeñas; a manera de ejemplo y sin pérdida de generalidad, los municipios del estado de Aguascalientes (lado derecho de la ilustración 2) constituyen áreas pequeñas en dicha encuesta.

Ejemplo de áreas geográficas pequeñas en la ENOE^N

Ilustración 2



Fuente: INEGI.

2.2 Las metodologías de estimación en áreas pequeñas

Las técnicas de estimación en áreas pequeñas son herramientas estadísticas relativamente novedosas que permiten obtener estimadores confiables, sin necesidad de desarrollar ninguna encuesta adicional, tan solo mediante el uso de fuentes de información combinadas e integradas de propósitos múltiples: encuestas, censos, registros administrativos y otras (como pueden ser las imágenes satelitales y las redes sociales).

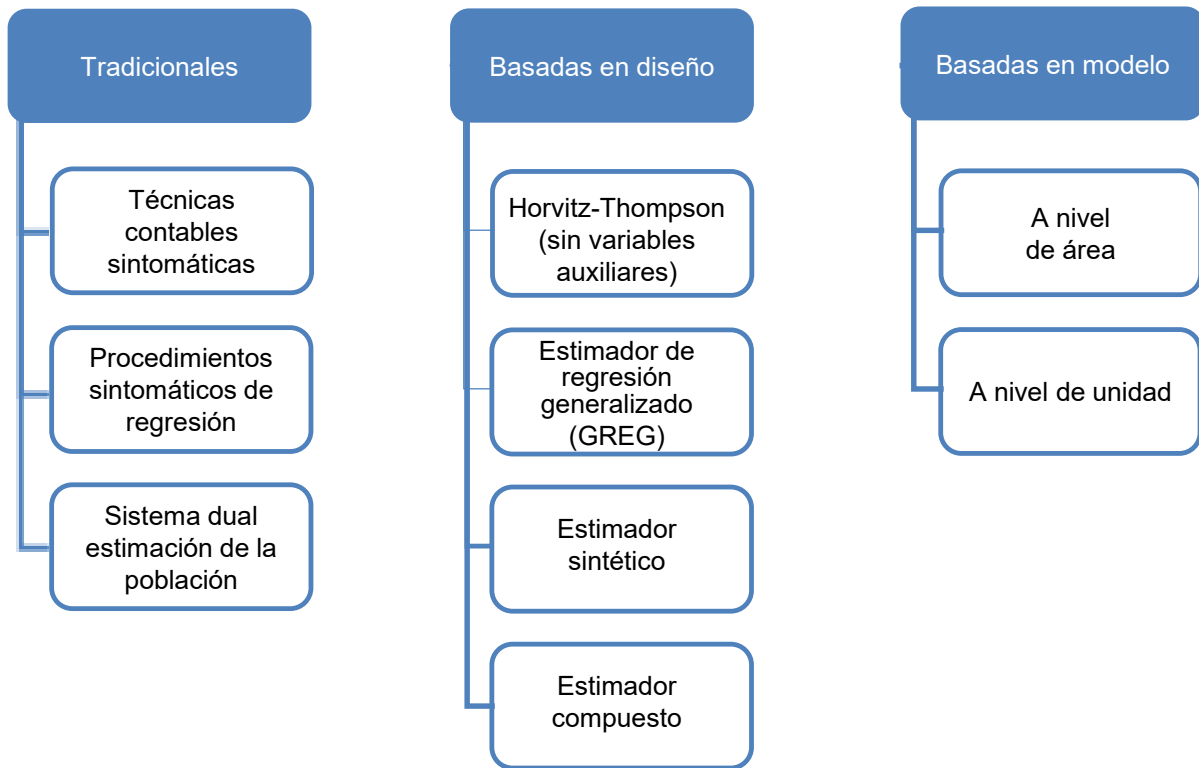
Diferentes métodos para la estimación en áreas pequeñas fueron recopilados por Harter (1983), en cuyo documento se explican de manera general los primeros métodos conocidos como “técnicas contables sintomáticas” basadas en registros de nacimientos y defunciones, aplicando métodos mejorados

de componentes demográficos, métodos de regresión, análisis de datos categóricos y estimadores de predicción (asumiendo que las áreas pequeñas mantienen un comportamiento similar a las de mayor agregación, en términos de las variables de interés), con el objetivo de conocer cifras de poblaciones intercensales o postcensales de ciudades y condados de la Unión Americana.

Recientemente Rao (2003) y Molina & Rao (2015) hacen una compilación de los métodos de estimación de áreas pequeñas en el libro *Small Area Estimation*, primera y segunda edición (ver la ilustración 3). Ambos textos describen varios métodos de estimación en áreas pequeñas basados en el diseño muestral. En ausencia de variables auxiliares se encuentra el estimador Horvitz-Thompson, el cual únicamente requiere del valor de la variable de la encuesta y la probabilidad de inclusión (factor de expansión). Otros métodos, en cambio, requieren para su operación variables auxiliares, como el estimador de regresión generalizado (GREG), el estimador sintético y el estimador compuesto.

Diferentes técnicas en las estimaciones en áreas pequeñas

Ilustración 3



Fuente: INEGI.

Las estimaciones basadas en diseño muestral se refieren a que, si mediante las encuestas tradicionales se obtienen resultados confiables solamente hasta cierto nivel geográfico, y el interés fuera contar con estimaciones para niveles geográficos más desagregados (o dominios de estudio), entonces se tendría que diseñar una encuesta con tamaño de muestra notablemente mayor, a fin de alcanzar la precisión deseada. Desde luego, esto conlleva a costos de operación de alta consideración.

Un ejemplo de un proyecto con estimaciones que no requieren variables auxiliares, es la Encuesta Intercensal 2015, que fue diseñada para obtener resultados confiables del total de la población por sexo y grupo de edad, así como del total de viviendas y algunas características sociodemográficas para los dominios geográficos: nacional, entidad federativa, municipal y zonas metropolitanas. La encuesta tuvo un tamaño de muestra de 7 853 702 viviendas visitadas.

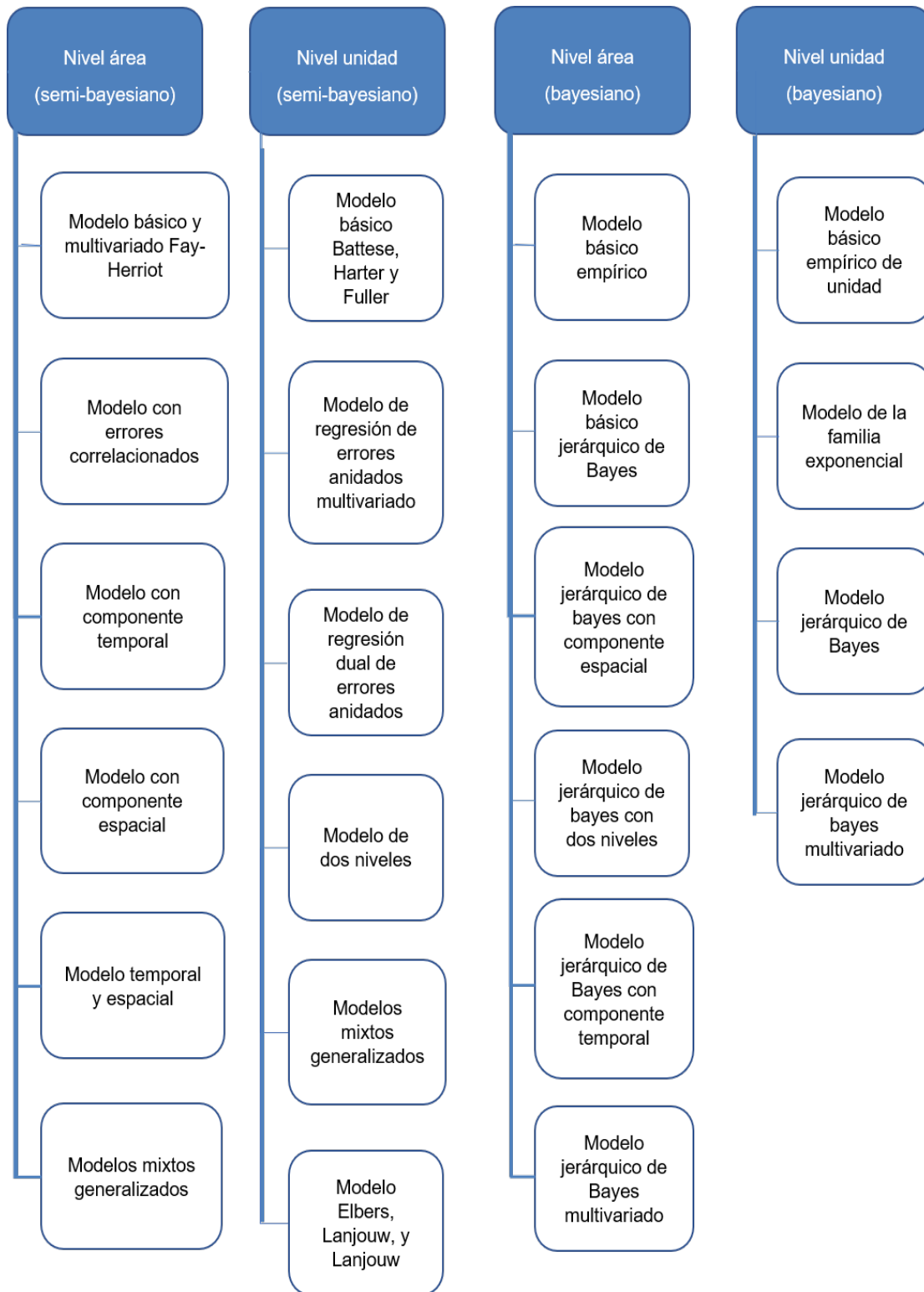
Las estimaciones en áreas pequeñas basadas en modelos estadísticos son derivadas de los modelos mixtos y los modelos mixtos generalizados. Ambos involucran efectos aleatorios específicos del área

pequeña con el fin de incorporar en el modelo la variación existente entre ellas, más allá de la explicada por las variables auxiliares incluidas en el modelo.

La compilación de los métodos de estimación en áreas pequeñas basados en modelos estadísticos es muy diversa, se dividen en dos vertientes: los modelos de área (Fay y Herriot, 1979) y los modelos de unidad (Battese, Harter y Fuller, 1988). Ver la ilustración 4.

Diferentes modelos empleados en las estimaciones en áreas pequeñas

Ilustración 4



Fuente: INEGI.

Los modelos de área incorporan, tanto en las variables de interés como en las variables auxiliares, información agregada a nivel de área. Por su parte, los modelos a nivel de unidad relacionan los valores unitarios de la variable de interés con los valores unitarios de las variables auxiliares, lo cual complica la construcción de las bases de datos, ya que se tiene que relacionar, por ejemplo, la vivienda visitada en la encuesta con el registro censal o registro administrativo de la misma vivienda; para este tipo de modelos, los censos constituyen la fuente de información principal en cuanto obtención de las variables auxiliares se trata.

Posterior a esta clasificación, los modelos estadísticos se bifurcan en dos ramas según el enfoque: frecuentistas y bayesianos. Los modelos estadísticos bajo el enfoque frecuentista se caracterizan por establecer que el parámetro poblacional de interés sea fijo pero desconocido, y su forma general de estimación se fundamenta en elegir el valor del parámetro que maximiza la probabilidad con base en una muestra observada. Mientras que los modelos estadísticos con enfoque bayesiano establecen que el parámetro poblacional de interés es desconocido, pero no es fijo, es variable, y su forma general de estimación es contar con una distribución subjetiva preliminar de los parámetros antes de analizar una muestra; al incorporar en el análisis la información de la muestra se obtendrá una nueva distribución.

Para algunas técnicas de estimación en áreas pequeñas, la forma de estimación consiste en obtener el mejor predictor empírico lineal insesgado; la parte fija de este predictor se obtiene mediante métodos como la Máxima Verosimilitud o Máxima Verosimilitud Restringida, o bien, el método de ajuste de constantes de Henderson (1953). Estos son métodos frecuentistas. Para efectos de exposición, a los modelos que incluyan métodos frecuentistas se les denominará modelos semi-bayesianos y a los modelos que incorporen funciones *a priori* y funciones *a posteriori* se les designará como modelos bayesianos.

El mejor predictor empírico lineal insesgado está basado en el renombrado modelo a nivel de área de *Fay-Herriot*, propuesto por Fay y Herriot (1979). Este predictor es de los más utilizados para estimar los parámetros generales en áreas pequeñas. El INEGI se basa en él para obtener los indicadores laborales municipales, pero adicionándole un componente espacial.

3. Proceso para la generación de las estadísticas en áreas pequeñas

El INEGI cuenta con un proceso estándar para la producción de información estadística y geográfica. Este proceso se refleja en el instrumento *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica*. Tal instrumento toma como premisa al *Modelo Genérico del Proceso Estadístico* de la Comisión Económica de las Naciones Unidas para Europa.

Por lo anterior, el proceso para la generación de técnicas de estimación en áreas pequeñas fue alineado a dicho instrumento normativo, como se ilustra en el cuadro 1.

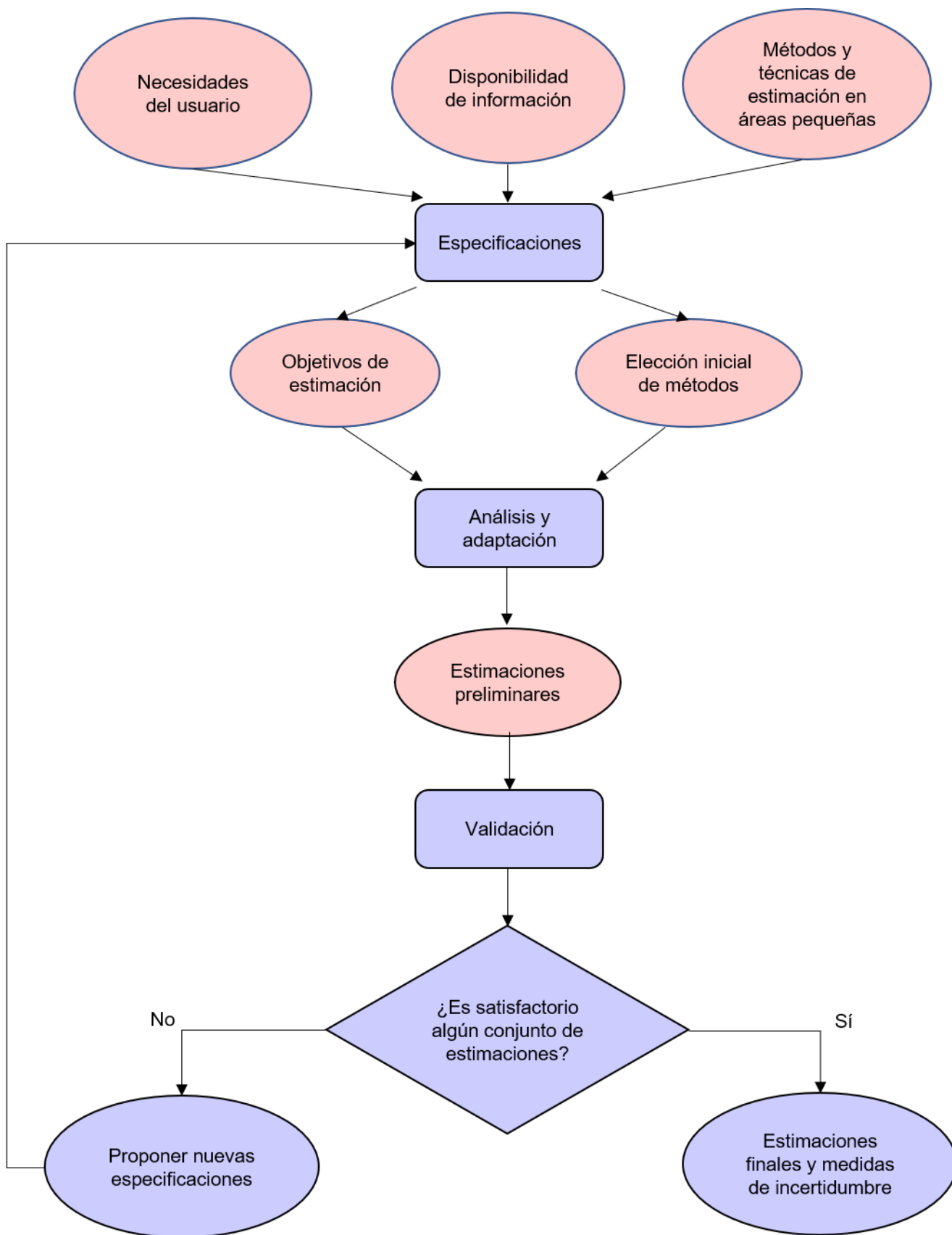
Proceso de generación para las estadísticas en áreas pequeñas

Cuadro 1

Fases	Subprocesos
Especificación de necesidades	Determinación conforme a las necesidades de información de las variables de interés no contenidas en las estadísticas tradicionales para los dominios requeridos.
Diseño	Especificación del modelo y predictor estadístico por utilizar para la estimación en áreas pequeñas, de acuerdo con la disponibilidad de información y las estimaciones de interés.
Construcción	Preparación de información fuente de la ENOE ^N y de los registros administrativos disponibles a nivel municipal correlacionados con las variables de interés. Identificación y elaboración de <i>software</i> para el modelo y predictor estadístico propio de las estimaciones en áreas pequeñas.
Procesamiento	Determinación de variables auxiliares seleccionadas. Incorporación del componente espacial. Estimación de parámetros del modelo. Comprobación de los supuestos del modelo obtenido. Procesamiento de datos municipales. Elaboración de tabulados.
Análisis	Validación de resultados por comparación con otras fuentes de información.
Difusión	Preparación de información. Liberación de la información en la plataforma institucional. Difusión de resultados.
Evaluación	Evaluación del acceso a la información. Evaluación del uso de la información. Detección de áreas de oportunidad.

Fuente: INEGI.

En el año 2018 Tzavidis *et al.* publicaron un artículo donde sugieren un esquema general para la producción de estadísticas en áreas pequeñas. El bosquejo de este esquema se puede apreciar en la ilustración 5. En términos generales, el contenido de esta ilustración coincide con el proceso estándar del INEGI (explicado en el cuadro anterior).



Fuente: Tzavidis (2018).

4. Fuentes de información, base conceptual y base metodológica

En las siguientes secciones se describen las fuentes de información para estimar las variables de interés, el predictor utilizado en el marco de las técnicas de estimación en áreas pequeñas y sus consideraciones para la construcción de los modelos.

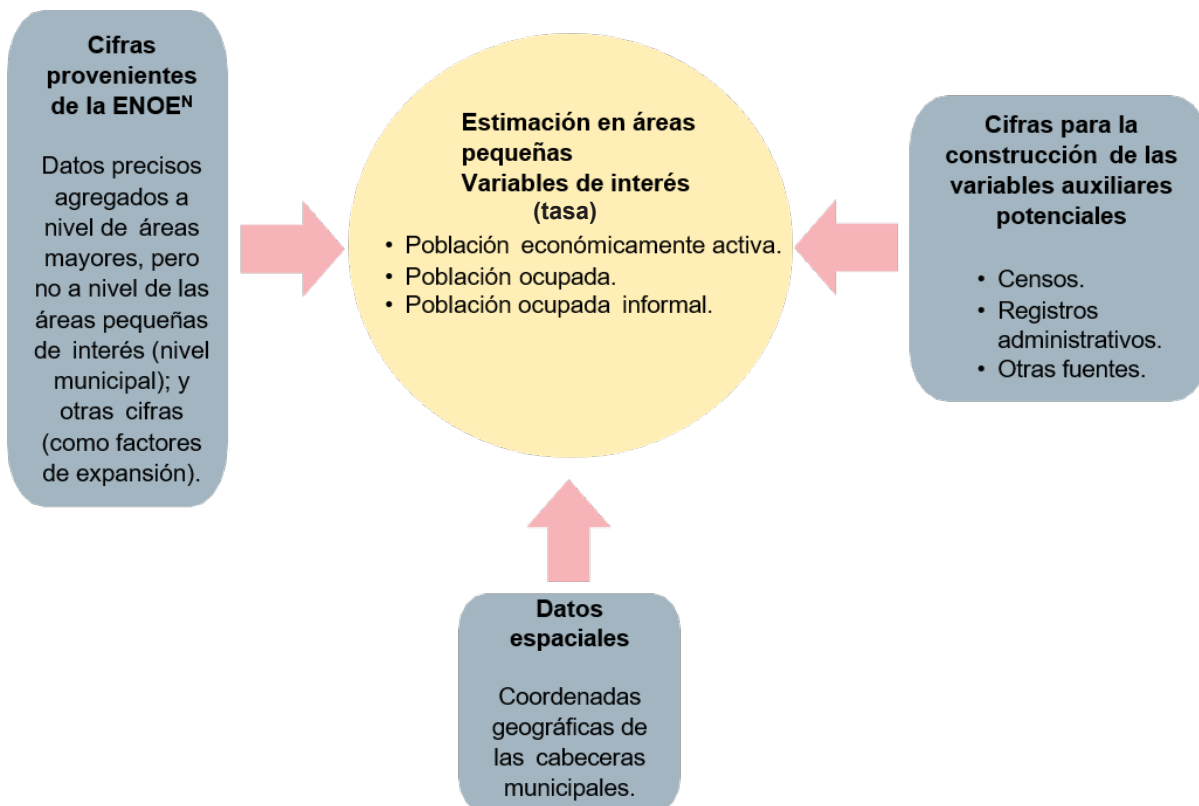
4.1 Fuentes de información

En la actualidad, en México no existe ninguna fuente de datos que anualmente ofrezca resultados directos de cifras laborales a nivel municipal (salvo los censos y conteos intercensales que son realizados, uno u otro, cada cinco años), de manera que se ha optado por utilizar fuentes de información combinadas y aplicar técnicas de estimación en áreas pequeñas. Por un lado, se utiliza la ENOE^N, de donde se toma la información de las variables objeto de estimación; mientras que, por el otro, se utilizan los censos y los registros administrativos, a partir de los cuales se obtienen las variables auxiliares para explicar el comportamiento de las primeras.

En la ilustración 6 se muestra la información requerida para obtener las estimaciones de las variables de interés. Esta información se compone de cifras provenientes de la ENOE^N, de cifras para la construcción de las variables auxiliares potenciales (datos que provienen de censos, registros administrativos y otras fuentes) y datos espaciales (coordenadas geográficas de cabeceras municipales).

Vínculo entre las variables de interés y la información necesaria para sus estimaciones

Ilustración 6

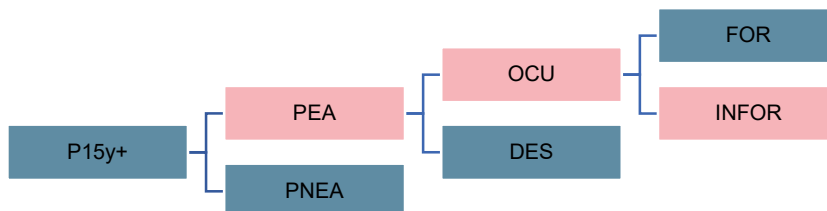


4.2 Clasificación de la población de 15 y más años de edad, según características laborales

Para el caso específico de la producción de los ILMM, en el marco conceptual de la ENOE^N, la variable correspondiente a la población de 15 y más años de edad se ramifica de acuerdo con la ilustración 7. Con las técnicas de estimación en áreas pequeñas se calculan tasas para las poblaciones en rosa.

Ramificación de la variable población de 15 años y más, según características laborales

Ilustración 7



Fuente: INEGI.

Donde:

P15y+ es la población de 15 y más años de edad.

PEA es la población económicamente activa.

PNEA es la población no económicamente activa.

OCU es la población ocupada.

DESOCU es la población desocupada.

FOR es la población ocupada formal.

INFOR es la población ocupada informal.

Población económicamente activa. A esta variable la integran las personas de 15 y más años de edad que tuvieron vínculo con una actividad económica o que lo buscaron en la semana de referencia, por lo que se encontraban ocupadas o desocupadas.

Población ocupada. Comprende a las personas de 15 y más años de edad que en la semana de referencia realizaron alguna actividad económica durante al menos una hora. Incluye a quienes tenían trabajo, pero no lo desempeñaron temporalmente por alguna razón, sin que por ello perdieran el vínculo laboral con este; así como a quienes ayudaron en alguna actividad económica sin recibir un sueldo o salario.

Población ocupada informal. Comprende la suma de la población ocupada, sin duplicar, de los ocupados que son laboralmente vulnerables por la naturaleza de la unidad económica para la que trabajan, con aquellos cuyo vínculo o dependencia laboral no es reconocido por su fuente de trabajo.

4.3 Variables de los modelos

En este apartado se describen las variables objeto de estimación a nivel municipal, así como las variables auxiliares potenciales y las seleccionadas para la construcción del modelo mediante técnicas de estimación en áreas pequeñas.

4.3.1 Variables objeto de estimación

Las variables de interés que se requieren estimar a nivel municipal son las siguientes:

- Tasa de la población económicamente activa respecto a la población de 15 y más años de edad.
- Tasa de la población ocupada respecto a la población económicamente activa.
- Tasa de la población ocupada informal respecto a la población ocupada.

4.3.2 Variables auxiliares potenciales y seleccionadas

Para la estimación de las variables de interés, es necesario recurrir a distintas fuentes de información, las cuales deben ser consistentes con las fechas de referencia, la unidad de observación y la unidad de análisis de la ENOE^N.

En primer lugar, se realiza un análisis de las experiencias internacionales, bajo la premisa de indagar cuáles variables explicativas se han utilizado en ejercicios similares. Posteriormente, se hace la búsqueda y el procesamiento de fuentes de información para la obtención y construcción de las que se denominan variables auxiliares potenciales. Estas variables deben ser adecuadas temporal y geográficamente para hacerlas compatibles con la información agregada de la ENOE^N. En el cuadro 2 se muestran algunas de las variables auxiliares potenciales que se han utilizado para los ILMM.

Ejemplo de variables auxiliares potenciales

Cuadro 2

1. Proporción de la población en edad de trabajar.
Población de 15 a 64 años respecto a la de 15 y más años de edad.
2. Proporción de la población masculina.
Población de hombres entre 15 y 44 años respecto a la de 15 y más años de edad.
3. Proporción de la población en edad madura para trabajar.
Población de 45 y más años respecto a la de 15 y más años de edad.
4. Proporción de la población femenina joven.
Población de mujeres de 18 a 29 años respecto a la población total de 18 a 29 años de edad (la suma de hombres y mujeres).
5. Índice de masculinidad.
Población de hombres de 15 y más años respecto a la de mujeres de la misma edad.
6. Relación de dependencia económica.
Población de menores de 15 años y de 65 y más respecto a la población de 15 a 64 años de edad.
7. Proporción de la población afiliada al IMSS o al ISSSTE.
Población afiliada al IMSS y al ISSSTE (la suma de estas dos) respecto a la población de 15 y más años de edad.
8. Proporción de población beneficiaria del Seguro Popular.
Población afiliada al Seguro Popular respecto a la población de 15 y más años de edad.
9. Matrícula escolar básica.
Población matriculada en educación preescolar, primaria y secundaria (la suma de estas tres) respecto a la población de 15 y más años de edad.
10. Proporción de la población con conflictos individuales de jurisdicción local.
Población en condiciones de conflicto de jurisdicción local respecto a la población de 15 y más años de edad.
11. Beneficiados del programa Jóvenes Construyendo el Futuro.
Población beneficiada con el programa Jóvenes Construyendo el Futuro respecto a la población de 18 a 29 años de edad.
12. Beneficiados con el programa Producción para el Bienestar de Pequeños y Medianos Productores.
Población beneficiada con el programa Producción para el Bienestar de Pequeños y Medianos Productores respecto a la población de 18 y más años de edad.

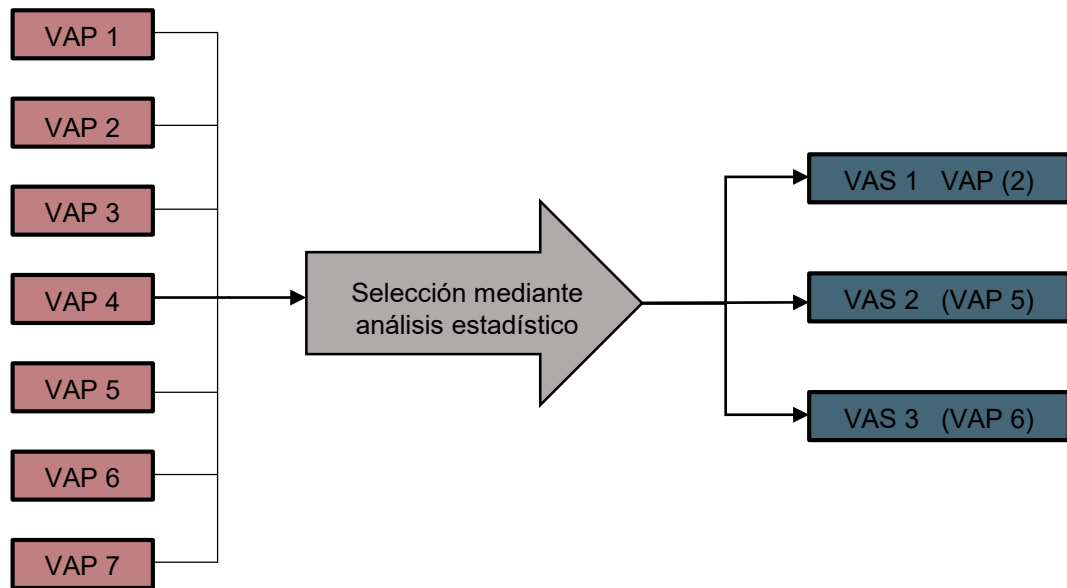
Fuente: INEGI.

Las variables auxiliares potenciales son analizadas estadísticamente para determinar cuáles de ellas se relacionan con las variables objeto de estimación. Aquellas que son estadísticamente significativas forman el conjunto de variables auxiliares seleccionadas. Ver la ilustración 8.

Las variables auxiliares seleccionadas son recopiladas con la finalidad de obtener el conjunto de variables explicativas que permitan la estimación estadísticamente adecuada para cada una de las variables objeto de estimación. Es importante mencionar que el conjunto de variables auxiliares potenciales y seleccionadas es específico para cada variable de estimación; así mismo, no necesariamente son las mismas para cada publicación anual (ver las presentaciones de resultados).

Variabes auxiliares

Ilustración 8



VAP: Variables auxiliares potenciales

VAS: Variables auxiliares seleccionadas

Fuente: INEGI.

4.4 Predictores y modelos de estimación en áreas pequeñas

Durante el desarrollo del proyecto se han considerado varios modelos estadísticos y predictores de estimación de áreas pequeñas, realizándose diversos análisis y ajustes. Principalmente se compararon dos predictores: el mejor predictor empírico lineal insesgado y el mejor predictor espacial empírico lineal insesgado³. El primero es una combinación lineal de la estimación directa y la estimación sintética⁴. Al incorporar un componente espacial a este, se obtiene el predictor espacial; tal componente considera las distancias euclidianas entre cabeceras municipales.

Al realizar la comparación, se observa una mayor verosimilitud y mejores criterios de información bayesiana y de *Akaike* con el predictor espacial, de modo que es este el utilizado para realizar las estimaciones de las tasas laborales.

A continuación, y en forma concisa, se describen los modelos estadísticos que utilizan ambos predictores, los cuales se aplican en el proceso de estimación de los Indicadores Laborales para los Municipios de México.

³ A lo largo de este documento se abrevia al mejor predictor espacial empírico lineal insesgado como predictor espacial.

⁴ La estimación directa se refiere a los valores obtenidos a través de la muestra de la encuesta. La estimación sintética resulta de un modelo estadístico alimentado con las variables auxiliares seleccionadas.

4.4.1 Modelo a nivel de área con el mejor predictor empírico lineal insesgado

Se consideran las abreviaturas de la siguiente lista:

Estimador directo (*D*): $\hat{\theta}_a^D = \theta_a + e_a$.

Modelo lineal (*L*): $\theta_a^L = X_a\beta + Zv_a$.

Modelo lineal mixto (*M*): $\theta_a^M = X_a\beta + Zv_a + e_a$.

Mejor predictor lineal insesgado (*B*): $\theta_a^B = X_a\beta + v_a = X_a\beta + \gamma_a(\hat{\theta}_a^D - X_a\beta)$.

Mejor predictor empírico lineal insesgado (*EB*): $\hat{\theta}_a^{EB} = x_a^T \hat{\beta}^{FH} + \hat{v}_a = x_a^T \hat{\beta}^{FH} + \hat{\gamma}_a(\hat{\theta}_a^D - x_a^T \hat{\beta}^{FH})$
 $= \hat{\theta}_a^S + \hat{\gamma}_a(\hat{\theta}_a^D - \hat{\theta}_a^S) = \hat{\gamma}_a \hat{\theta}_a^D + (1 - \hat{\gamma}_a) \hat{\theta}_a^S$.

Estimador sintético (*S*): $\hat{\theta}_a^S = x_a^T \hat{\beta}^{FH}$.

Función log-verosímil (*LV*): $l(\beta, \sigma_v^2, \psi) = -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \sum_{a=1}^m \ln(\sigma_v^2 + \psi_a)$
 $-\frac{1}{2} \sum_{a=1}^m (\hat{\theta}_a^D - x_a^T \hat{\beta})^2 / (\sigma_v^2 + \psi_a)$.

Donde,

$$v_a = \gamma_a(\hat{\theta}_a^D - X_a\beta),$$

$$\gamma_a = \frac{\sigma_v^2}{\sigma_v^2 + \hat{\psi}_a},$$

$$\hat{\beta}^{FH} = \left[\sum_{a=1}^m \frac{x_a x_a^T}{\hat{\sigma}_v^2 + \hat{\psi}_a} \right]^{-1} \left[\sum_{a=1}^m \frac{x_a \hat{\theta}_a^D}{\hat{\sigma}_v^2 + \hat{\psi}_a} \right], m \text{ es el número de municipios dentro del modelo seleccionado.}$$

$$\hat{\gamma}_a = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\psi}_a}.$$

El estimador directo (*D*) es el valor obtenido por la encuesta. Puede ser un valor como el total estimado de la variable a calcular para cada municipio (por ejemplo: el número total de personas de 15 y más años de edad). Se expresa como la suma del parámetro de interés más un error aleatorio debido al muestreo.

El modelo lineal (*L*) es una alternativa para estimar la variable objeto de estimación, siempre y cuando las variables auxiliares seleccionadas cumplan los supuestos del modelo (estos supuestos se explican en el apartado 6.2).

El modelo lineal mixto (M) surge de sustituir el modelo lineal (L) en el estimador directo (D). Se desglosa en tres partes: al primer sumando se le conoce como efecto fijo, al segundo se le conoce como efecto aleatorio y la última parte es el error aleatorio. En este modelo, X es la matriz de diseño de los efectos fijos, y a Z se le conoce como matriz de diseño de los efectos aleatorios.

El mejor predictor lineal insesgado (B) es la expresión matemática para el estimador del modelo lineal mixto a nivel de área. Las estimaciones son obtenidas en forma óptima por este predictor y son adecuadas para datos que no son de tipo panel o que no son longitudinales; sin embargo, el modelo asume que el vector de parámetros de la regresión y la varianza de los efectos aleatorios necesitan ser estimados.

Como las variables aleatorias no se pueden estimar sino solamente predecir, se propone el mejor predictor empírico lineal insesgado sustituyendo σ_v^2 por $\hat{\sigma}_v^2$. Este predictor es el que se aplica como punto de partida para obtener las predicciones de las cifras laborales cuando sí existe muestra de la ENOE^N en el municipio. Sin embargo, estas estimaciones se recalculan con el predictor espacial.

La fórmula matemática del mejor predictor empírico lineal insesgado ($\hat{\theta}_a^{EB} = \hat{\gamma}_a \hat{\theta}_a^D + (1 - \hat{\gamma}_a) \hat{\theta}_a^S$) expresa que las estimaciones obtenidas bajo este modelo (la información a *posteriori* en términos de inferencia bayesiana) son una combinación entre lo que se observa de la ENOE^N (la información a *priori* en términos de inferencia bayesiana) y lo que se obtiene con el modelo sintético basado en la verosimilitud.

El ponderador gamma $\hat{\gamma}_a$ resulta de dividir la varianza de los efectos aleatorios ($\hat{\sigma}_v^2$) entre la suma de la varianza de los efectos aleatorios más la varianza observada del estimador directo de la ENOE^N ($\hat{\psi}_a$). Por tanto, si la varianza de la encuesta es pequeña comparada con la varianza de los efectos aleatorios, el ponderador gamma tendrá un valor cercano a uno y, en consecuencia, la estimación de la ENOE^N tendrá más peso en el mejor predictor empírico lineal insesgado. Por el contrario, si la varianza de la encuesta es grande comparada con la varianza de los efectos aleatorios, el ponderador gamma tendrá un valor cercano a cero y, por ende, la estimación sintética tendrá más peso en este predictor.

Para una referencia de cómo obtener la varianza de los efectos aleatorios, se recomienda consultar la sección 6.1.2 de Rao & Molina (2015). La estimación del error cuadrático medio del mejor predictor empírico lineal insesgado se obtiene de la siguiente forma:

$$\widehat{ECM}(\hat{\theta}_a^{EB}) = g_1(\hat{\sigma}_v^2) + g_2(\hat{\sigma}_v^2) + 2g_3(\hat{\sigma}_v^2).$$

Donde:

$$g_1(\hat{\sigma}_v^2) = \hat{\gamma}_a \hat{\psi}_a,$$

$$g_2(\hat{\sigma}_v^2) = (1 - \hat{\gamma}_a)^2 x_a^T \left[\sum_{a=1}^m \frac{x_a x_a^T}{\hat{\sigma}_v^2 + \hat{\psi}_a} \right]^{-1} x_a,$$

$$g_3(\hat{\sigma}_v^2) = (1 - \hat{\gamma}_a)^2 \hat{\gamma}_a (\hat{\sigma}_v^2)^{-1} \bar{V}(\hat{\sigma}_v^2).$$

El valor de g_1 se asocia al error de muestreo de la estimación directa de la ENOE^N, el de g_2 se asocia al error del estimador sintético y el de g_3 al error de los efectos aleatorios; $\bar{V}(\hat{\sigma}_v^2)$ es la varianza asintótica de $\hat{\sigma}_v^2$. Para más detalle de esta expresión, consultar la sección 6.2.1 de Rao & Molina (2015).

Por otra parte, el estimador sintético (S), o efecto fijo del modelo mixto, es la parte del modelo que se aplica para obtener las predicciones de las cifras laborales cuando no existe muestra de la ENOE^N en el municipio, o bien, porque fue excluido como parte del modelo por ser considerado como un valor extremo (estas exclusiones se explican más en el apartado 6.2).

La estimación del error cuadrático medio del estimador sintético (S) para municipios que no tienen muestra, $l = m + 1, \dots, M$, tiene la expresión:

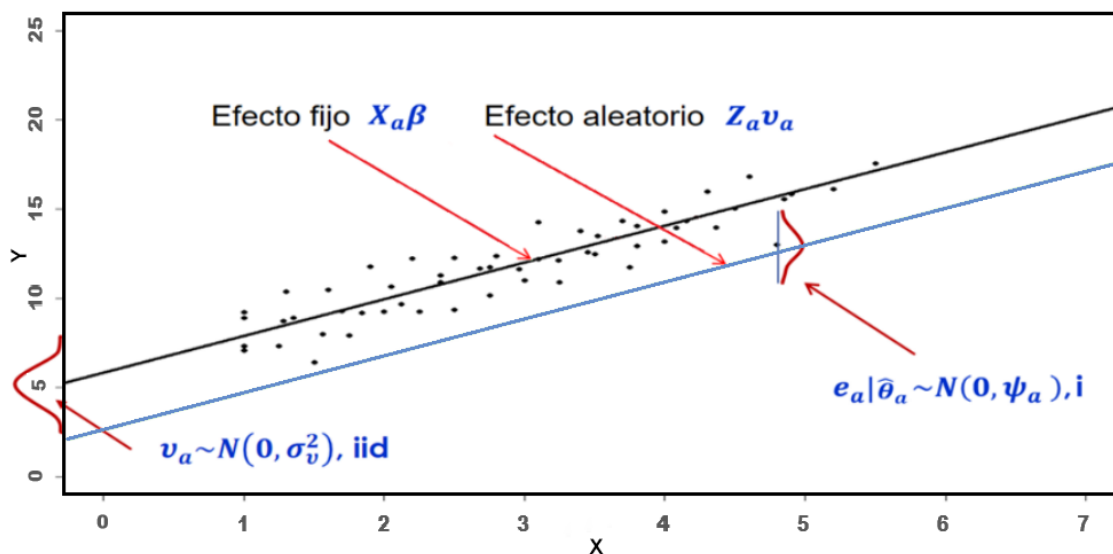
$$\widehat{ECM}(\hat{\theta}_l^S) = \hat{\sigma}_v^2 + x_l^T \left[\sum_{a=1}^m \frac{x_a x_a^T}{\hat{\sigma}_v^2 + \hat{\psi}_a} \right]^{-1} x_l.$$

Por último, la función log-verosímil (LV) sirve para estimar los componentes de varianza y el estimador de β . Esto se hace a través del método de máxima verosimilitud o máxima verosimilitud restringida, utilizando derivadas parciales para obtener la estimación de β y algoritmos iterativos cuando no se tenga una solución analítica para estimar $\hat{\sigma}_v^2$; en lo que respecta a ψ , esta se considera como una constante conocida que ofrece la ENOE^N.

Para ilustrar en forma gráfica el modelo lineal mixto, se realiza un diagrama de dispersión (ver la ilustración 9) en donde cada punto en el plano cartesiano representa un municipio. En esta gráfica, el eje vertical representa los valores del parámetro de interés, el estimador directo (D); en tanto que el eje horizontal representa los valores de una variable auxiliar.

Modelo lineal mixto en forma gráfica

Ilustración 9



Fuente: INEGI.

Dado que se necesitan al menos dos puntos para obtener un modelo lineal por municipio y la información solamente cuenta con uno, se calcula una sola recta de regresión para todos los municipios con muestra. Posteriormente, para cada municipio se predice un valor aleatorio que, sumado al efecto fijo, da como resultado una serie de líneas rectas paralelas al efecto fijo, tantas como municipios haya. Estas líneas paralelas son las que se utilizan para obtener un predictor de la variable de interés para cada municipio.

Por otro lado, los efectos aleatorios utilizados deberán tener una distribución normal y ser independientes e idénticamente distribuidos. La varianza de los efectos aleatorios se puede asociar como la varianza que hay entre los municipios una vez aplicado el ponderador gamma descrito anteriormente (varianza *between*).

Adicionalmente, los errores derivados de utilizar el mejor predictor empírico lineal insesgado en el modelo deberán tener también una distribución normal y ser independientes, mas no idénticamente distribuidos, en virtud de que dentro del modelo se incorpora la varianza obtenida por el operativo de la encuesta, la cual es diferente para cada municipio (varianza *within*). Dentro del proceso de estimación, los errores de muestreo se mantendrán como información fija y conocida.

El modelo utilizado para obtener las estimaciones está basado en el modelo a nivel de área de *Fay-Herriot*. Se asume disponible el dato auxiliar a nivel de área X_a (nivel municipio) y que una muestra s_a de tamaño n_a proviene de las N_a unidades en cada área pequeña a ($a = 1, \dots, A$).

4.4.2 Modelo a nivel de área con el mejor predictor espacial empírico lineal insesgado

El predictor que considera dependencia espacial, postula que el efecto aleatorio se expresa como un modelo autorregresivo espacial:

$$v = \rho Wv + u = (I_m - \rho W)^{-1}u.$$

$$v = (I_m - \rho W)^{-1}u.$$

El modelo lineal mixto espacial para estimación en áreas pequeñas se expresa como:

$$\theta^{SM} = X\beta + Z(I_m - \rho W)^{-1}u + e.$$

De esta forma, el modelo da cuenta de la importancia de la influencia de las demás áreas sobre el área específica a . En este modelo ρ es el coeficiente de correlación espacial, W es la matriz de porcentajes de los valores inversos de las distancias entre las cabeceras municipales y u es un vector de términos de error con media cero y varianza constante τ_u^2 .

En el predictor espacial, v se distribuye como una función normal de media cero y matriz de varianzas y covarianzas:

$$G(\tau_u^2, \rho) = \tau_u^2 [(I_m - \rho W)^T (I_m - \rho W)]^{-1} = (\widehat{\sigma}_v^2)'$$

Utilizando la notación sigma y términos algebraicos se tiene:

$$v_a = \rho \sum_{a \langle j} w_{aj} v_j + u_a, \quad \gamma'_a = \frac{(\widehat{\sigma}_v^2)'}{(\widehat{\sigma}_v^2)' + \widehat{\psi}_a}$$

Tomando la última igualdad, se tiene la expresión para el predictor espacial:

$$\hat{\theta}_a^{SEB} = x_a^T \hat{\beta}^{FH'} + \hat{\gamma}'_a (\hat{\theta}_a^D - x_a^T \hat{\beta}^{FH'}) = \hat{\theta}_a^{S'} + \hat{\gamma}'_a (\hat{\theta}_a^D - \hat{\theta}_a^{S'}) = \hat{\gamma}'_a \hat{\theta}_a^D + (1 - \hat{\gamma}'_a) \hat{\theta}_a^{S'}$$

En notación matricial, el predictor espacial se define como:

$$\hat{\theta}_a^{SEB} = x_a^T \hat{\beta}^{FH'} + z_m^T [\hat{\sigma}_v^2 (I_D - \hat{\rho}W)^{-1} (I_D - \hat{\rho}W^T)^{-1}] z^T * \\ \left[[diag(\widehat{\psi}_a) + \hat{\sigma}_v^2 z [(I_D - \hat{\rho}W)^{-1} (I_D - \hat{\rho}W^T)^{-1}] z^T] \right]^{-1} (\hat{\theta}_a^D - x_a^T \hat{\beta}^{FH'})$$

Donde,

$$\hat{\beta}^{FH'} = (X^T V^{-1} X)^{-1} X^t V^{-1} \hat{\theta}_a^D,$$

$$V = \text{diag}(\hat{\psi}_a) + \hat{\sigma}_v^2 z [(I_D - \hat{\rho} W)^{-1} (I_D - \hat{\rho} W^T)^{-1}] z^T.$$

Se tiene que z_m es un vector de tamaño $1 * m = (0,0,1,0, \dots)$ con 1 en la m -ésima posición, y z se compone por los elementos de la matriz de diseño de los efectos aleatorios (generalmente matrices diagonales).

5. Preparación de información

Antes de realizar las estimaciones estadísticas, es necesario preparar la información que servirá como insumo principal para la construcción de los modelos y su aplicación. En este capítulo se describen dos de los procedimientos en la preparación de información más importantes: la reponderación de los factores de expansión y la conformación de la matriz de distancias.

5.1 Ajustes de los factores de expansión municipales

La ENOE^N estima la población nacional y estatal a partir de una muestra cuya asignación se realiza dentro de cada entidad federativa entre los diferentes estratos de manera proporcional a su tamaño, en que por azar cae muestra en algunos municipios, alrededor de 1 000, por lo que en ellos se asigna toda la población, incluyendo la que le correspondería al resto de los municipios.

Con el propósito de que las poblaciones estimadas a nivel municipal reflejen un aproximado de su valor real, se debe ajustar los factores de expansión de la encuesta.

La expresión para el cálculo de este ajuste es:

$$w'_{ai} = w_{ai} \frac{\hat{N}_{aAJUSTADA}}{\hat{N}_{aENOEN}}.$$

Donde,

w'_{ai} = factor de expansión de la vivienda i ajustado a la población en el municipio a .

w_{ai} = factor de expansión original de la población de la ENOE^N para la vivienda i en el municipio a .

$\hat{N}_{aAJUSTADA}$ = número de personas calculadas mediante el ajuste (al aplicar proyecciones, o bien, la estructura porcentual del censo a las cifras estatales de la ENOE^N).

\hat{N}_{aENOEN} = número de personas estimadas por la ENOE^N en el municipio a .

Para el cálculo de las variables objeto de estimación, se conserva el diseño complejo de la encuesta, es decir, se respetan las unidades primarias de muestreo (UPM), los estratos y se incluye el factor de expansión ajustado.

Bajo esta misma consideración se calculan también las precisiones estadísticas correspondientes, como los intervalos de confianza y los coeficientes de variación, que constituyen elementos para medir la calidad de las estimaciones resultantes.

5.2 Matriz de distancias

La autocorrelación espacial se define en términos generales como la concentración o dispersión de los valores de una variable en un mapa geográfico. Dicho de otra manera, la correlación espacial refleja el grado en que objetos o actividades en una unidad geográfica son similares a otros objetos o actividades en unidades geográficas próximas (Goodchild, 1987). Este tipo de autocorrelación prueba la primera ley

geográfica de Tobler (1970) “Todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes”.

En general, si el valor de una variable de dos objetos o actividades cercanas se parecen mucho entre sí, se dice que existe una autocorrelación espacial positiva. Si, por el contrario, el valor de la variable de los objetos cercanos, por el hecho de estar juntos difieren mucho entre sí, la autocorrelación espacial es negativa; por ejemplo, la delincuencia suele ser menor en las cercanías de las estaciones de policía, entonces hay una correlación espacial negativa entre centros de operación delictiva y estaciones de policía.

Así, la autocorrelación espacial tiene que ver tanto con la localización geográfica como con los valores de la variable en estudio. Para determinar si el patrón de distribución espacial dista de ser meramente aleatorio, debe utilizarse un índice de comparación que responde a un mismo principio: poner en relación las diferencias del valor de la variable entre los objetos o actividades con la correspondiente distancia geométrica que les separa. Debido a lo anterior, existe un patrón común a todo índice de autocorrelación espacial del tipo:

$$\Gamma = \sum_a P_{aj} \sum_j C_{aj}.$$

Donde la matriz P_{aj} es formada por la distancia geográfica entre dos sucesos o casos, o bien, pueden ser valores que representen una forma de medición de la contigüidad en los datos originales. La matriz C_{aj} es la distancia en el espacio de datos o la diferencia entre los objetos o actividades, o una medida de la proximidad de los valores a y j , en otro tipo de distancia (euclídea, esférica, de Manhattan, etcétera).

Un método muy conocido para obtener una medida de autocorrelación espacial es el Índice de Moran (I de Moran), mismo que se utiliza para los ILMM. El atributo es cada una de las variables de estimación, y los objetos a los que se les mide tal atributo son los municipios que cuentan con muestra en la ENOE^N.

El índice se calcula a partir del valor medio y la varianza para el atributo laboral que se evalúa, posteriormente se resta el valor medio (\bar{y}) en cada valor del atributo para cada municipio con muestra en la ENOE^N (y_a o y_j), lo que crea una desviación del valor medio. Los valores de la desviación para todos los municipios vecinos (los municipios dentro del criterio de distancia especificada) se multiplican de forma conjunta para crear un producto cruzado. El numerador para la estadística I de Moran suma estos productos cruzados. Los índices a y j van de 1 a m , donde m es el número de municipios seleccionados. La finalidad del I de Moran es determinar si existe una dependencia espacial en la variable objeto de estimación. En caso de que sí, entonces se debe emplear el predictor espacial, en caso de que no, lo correcto es utilizar el mejor predictor empírico lineal insesgado.

Este índice resulta análogo al coeficiente de correlación convencional de *Pearson*, ya que su numerador se interpreta como la covarianza entre unidades contiguas, y sus valores oscilan entre +1 (significando fuerte correlación espacial positiva) y -1 (significando fuerte correlación espacial negativa).

A continuación, se muestran ambas expresiones de cálculo:

$$I \text{ de Moran} = \frac{1}{\sum_a \sum_j w_{aj}} \frac{\sum_a \sum_j w_{aj} (y_a - \bar{y})(y_j - \bar{y})}{\frac{\sum_a (y_a - \bar{y})^2}{m}}.$$

$$\text{Correlación de Pearson } (r) = \frac{1}{m} \frac{\sum_a (x_a - \bar{x})(y_a - \bar{y})}{\sqrt{\frac{\sum_a (x_a - \bar{x})^2}{m}} \sqrt{\frac{\sum_a (y_a - \bar{y})^2}{m}}}.$$

En la fórmula para el cálculo del I de Moran, w_{aj} es cada elemento de la matriz W , compuesta por los inversos de las distancias. Esta matriz considera la vecindad y distancias euclidianas entre las cabeceras municipales de los municipios que tienen muestra en la ENOE^N.

Las distancias euclidianas se pueden obtener en kilómetros, metros, o cualquier otra medida numérica de distancia. De estas, se toman los porcentajes respecto a la suma total de porcentajes recíprocos. Se toma el recíproco porque en este trabajo interesa darle más ponderación a los vecinos que geográficamente estén más cerca de cierto municipio.

Por ejemplo, el cuadro 3 muestra las distancias (en kilómetros) y sus respectivos porcentajes respecto a la suma de distancias que hay entre la cabecera municipal de un municipio A y las cabeceras municipales de tres municipios vecinos (B, C y D). En este ejemplo, la suma de distancias es de 100 km.

Distancias en kilómetros entre cabeceras municipales

Cuadro 3

Parámetro respecto al municipio A	Municipio vecino		
	B	C	D
Distancia	20 km	30 km	50 km
Porcentaje respecto a la suma de distancias	20.00	30.00	50.00

Fuente: INEGI.

En el cuadro 4 se observa el recíproco del porcentaje respecto a la suma de distancias para cada municipio vecino. Dividiendo cada uno de estos números por la suma de ellos, se obtiene el porcentaje de ponderación, mismo que cumple con el objetivo de dar más peso a los municipios más cercanos. En este ejemplo, la suma de los recíprocos del porcentaje respecto a la suma de distancias es de 10.33.

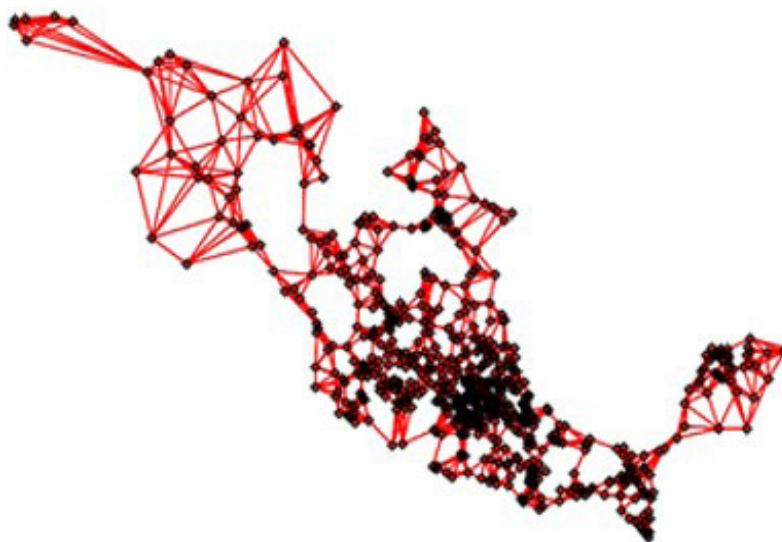
Distancias en porcentaje entre cabeceras municipales

Cuadro 4

Parámetro respecto al municipio A	Municipio vecino		
	B	C	D
Recíproco del porcentaje respecto a la suma de distancias	5.00	3.33	2.00
Porcentaje ponderador	48.39	32.26	19.35

Fuente: INEGI.

Como se explica en el capítulo seis, se prueban varios modelos para decidir con cuál aplicar el predictor espacial. Estos modelos se diferencian en considerar el número de municipios con muestra en la ENOE^N y el número de vecinos por considerar. En la ilustración 10 se muestra un ejemplo de los municipios con muestra en la ENOE^N considerados en algún modelo y los vecinos de cada uno de ellos.



Fuente: INEGI.

La matriz de los inversos de las distancias W , es una matriz cuadrada no simétrica, donde cada fila debe sumar 100 por ciento. A manera de ejemplo, considérese que un modelo particular que toma en cuenta a los seis vecinos más cercanos por cada municipio seleccionado de los que tienen muestra en la ENOE^N. Una parte de esta matriz se presenta en el cuadro 5, que corresponde a los municipios del estado de Aguascalientes que tuvieron muestra en la ENOE^N y que fueron seleccionados para ese modelo particular (los valores numéricos están dados en porcentajes).

Es importante mencionar que los municipios pueden tener entre sus vecinos más cercanos a los de otros estados (siempre y cuando tengan muestra en la ENOE^N y hayan sido seleccionados por el modelo respectivo). Como referencia de vecindades entre municipios, en la ilustración 11 se muestra un mapa con los municipios del estado de Aguascalientes (en azul) y los municipios vecinos a ellos (los de Jalisco en amarillo y los de Zacatecas en rosa).

Matriz de los inversos de las distancias W

Cuadro 5
(1a. parte)

Municipio	Vecinos							
	S. F. de los R.	Rincón de R.	San José de G.	Jesús María	Tepezalá	Aguascalientes	Pabellón de A.	Calvillo
S. F. de los R.		12.58	10.79	14.27	15.60	14.00	32.75	
Rincón de R.	10.57		12.88		19.14		16.25	
San José de G.	14.01	19.91		18.44			19.67	14.09
Jesús María	17.17	11.11	17.09			20.99	17.75	15.90
Tepezalá	14.98	21.87					18.52	
Aguascalientes	19.84		11.49	24.73			15.59	11.65
Pabellón de A.	28.76	16.98	13.30	12.95	16.95			
Calvillo			15.79	19.24		11.97		
Cosío		27.30	9.19		17.62		10.84	

Matriz de los inversos de las distancias W

Cuadro 5
(2a. y última parte)

Municipio	Vecinos							
	Cosío	Cuauhtémoc	Loreto	Encarnación de D.	Huanusco	Tabasco	Jalpa	Ojocaliente
S. F. de los R.								
Rincón de R.	26.66	14.50						
San José de G.	13.88							
Jesús María								
Tepezalá	19.66	11.04	13.93					
Aguascalientes				16.69				
Pabellón de A.	11.06							
Calvillo					19.19	21.77	12.04	
Cosío		22.47						12.58

Fuente: INEGI.

Vecindades entre municipios del estado de Aguascalientes⁵

Ilustración 11



Fuente: INEGI.

⁵ Los municipios cuyo nombre no aparece en el mapa no tuvieron muestra en la ENOE^N o no fueron seleccionados en el modelo plasmado en el ejemplo.

6. Construcción de los modelos

El principal *software* utilizado para calcular las variables objeto de estimación es la librería *sae* de R. En este capítulo se explica el porqué de la elección de este programa, así como el proceso para la obtención de las estimaciones a través de dicha librería.

6.1 *Software*

En los cálculos de los primeros ILMM se realizó un análisis para la selección del *software* de cómputo estadístico adecuado para el cálculo de las estimaciones a través de las técnicas de estimación en áreas pequeñas. Se estudiaron las herramientas que se describen a continuación.

Programas elaborados por el proyecto EURAREA⁶ (2004) en la plataforma SAS y para la estimación en áreas pequeñas de países europeos. En el INEGI se cuenta con licencias para este *software*, situación que fue aprovechada para replicar los programas disponibles al público, tomando como base las cifras de la ENOE^N y los registros administrativos disponibles. Se concluyó que lo desarrollado en esta plataforma ofrece resultados básicos y limitados para las necesidades propias del Instituto.

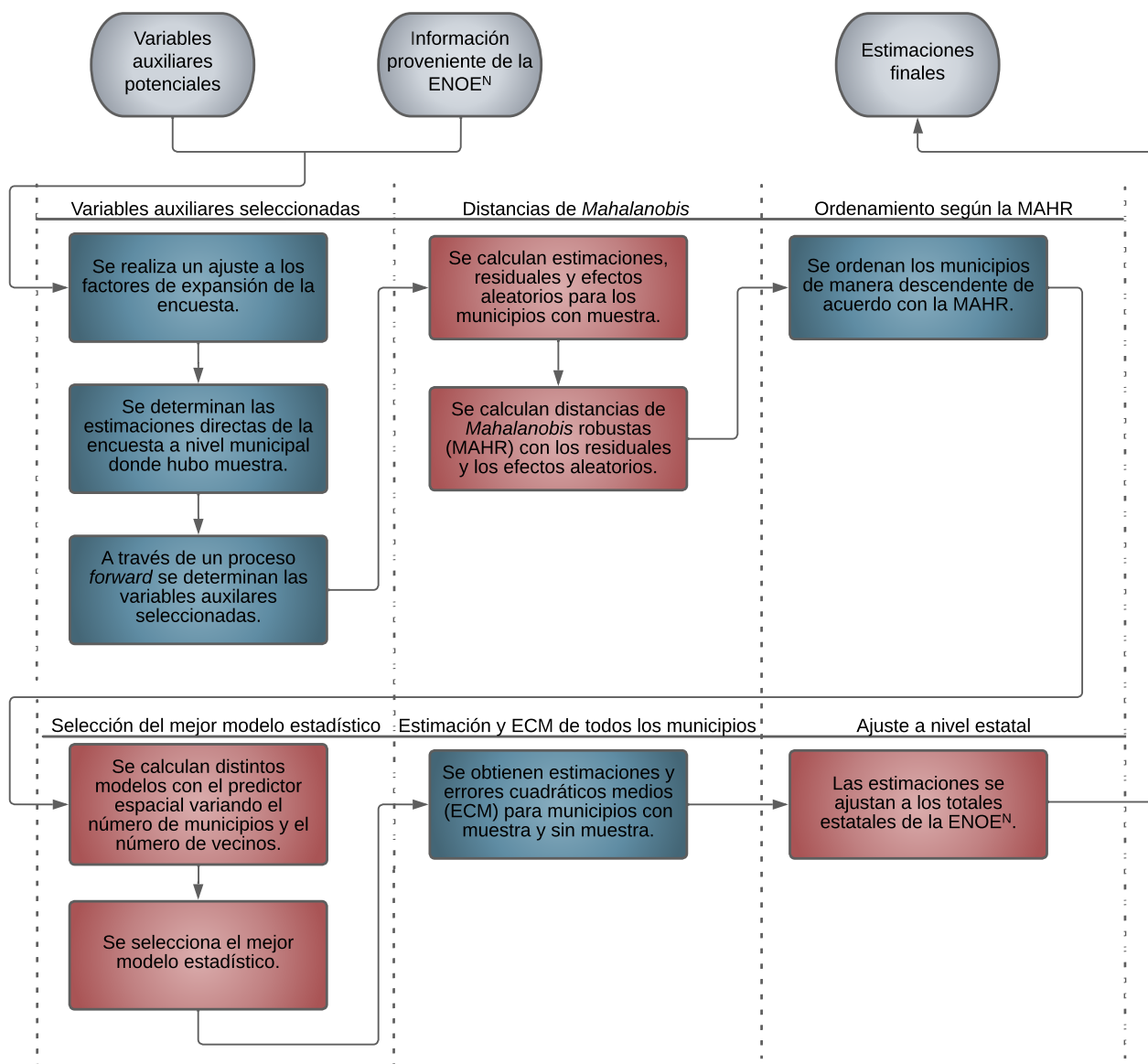
Librería *mme* del CRAN de R, desarrollada por E. López-Vizcaíno, M.J. Lombardía y D. Morales. Se replicaron las rutinas contenidas en esta librería; además, se desarrollaron nuevas rutinas con el fin de complementar la salida de productos como las estimaciones ajustadas y los errores estándar para todos los municipios. Al concluir esta actividad, se determinó que este *software* ofrecía una gran ventaja respecto a otros, ya que se pueden utilizar modelos multinomiales en donde se encuadran las cifras para dar un total por agregación. En los ejercicios se utilizaron hasta tres categorías, por ejemplo, la PEA que es una categoría conformada por otras dos, los ocupados y su complemento (los desocupados). Aun considerando sus ventajas, el *software* carece de la posibilidad de incluir el diseño complejo de la muestra, por tanto, para la obtención de resultados, se asume que la información proviene de un muestreo aleatorio simple. Adicionalmente, la librería no cuenta con la posibilidad de incorporar el componente de dependencia espacial (que estadísticamente es significativo para los ILMM).

Librería *sae* del CRAN de R, desarrollada por Isabel Molina y Yolanda Marhuenda. Esta librería es soporte básico para la estimación de los indicadores laborales de referencia; permite incorporar el diseño de la muestra, así como el componente espacial, y tiene la posibilidad de analizar datos tipo panel observados en el tiempo. Cabe mencionar que, aunque esta librería permite incluir un componente temporal, este no se incluye para obtener las estimaciones, en virtud de contar únicamente con información auxiliar de pocos años; además de que el panel solamente se puede construir con aproximadamente quinientos municipios.

6.2 Procesamiento

Una vez que se tienen los insumos para la construcción del modelo (la información agregada proveniente de la ENOE^N, los factores de expansión ajustados y las variables auxiliares potenciales), se realiza el procesamiento de la información que lleva a la obtención de las estimaciones de cifras laborales de los municipios de México. En la ilustración 12 se observa el proceso y, posterior a ella, se describen cada uno de los subprocesos.

⁶ EURAREA se refiere al proyecto *Enhancing Small Area Estimation Techniques to meet European needs*.



Fuente: INEGI.

Variables auxiliares seleccionadas. Se realiza un ajuste en el factor de expansión para cada uno de los municipios, alrededor de 1 000, donde cae la muestra de la ENOE^N; de esta forma es posible contar con estimaciones directas para ellos. Posteriormente, para cada variable objeto de estimación, se realiza un procedimiento *forward* a través del cual se determinan las variables auxiliares seleccionadas que mejor se encuentran correlacionadas con cada una de las variables de interés. Para las tasas de la PEA y los ocupados, por estar altamente correlacionadas entre sí, las variables auxiliares seleccionadas son las mismas y se determinan de manera conjunta.

Distancias de Mahalanobis. Para cada variable objeto de estimación se obtienen estimaciones con el mejor predictor empírico lineal insesgado tomando como insumo todos los municipios con muestra. Con el modelo lineal mixto y por cada variable de interés, se obtienen dos valores para cada uno de estos municipios: el efecto aleatorio y el residual. Para las tasas de la PEA y los ocupados, nuevamente por estar altamente correlacionadas, se calcula de manera conjunta una única distancia de Mahalanobis por municipio (se consideran cuatro valores por municipio: efecto aleatorio de la estimación para la tasa de la PEA,

efecto aleatorio de la estimación para la tasa de los ocupados, residual de la estimación para la tasa de la PEA y residual de la estimación para la tasa de los ocupados). Para la variable informales se calcula otra distancia de *Mahalanobis* para cada municipio (utilizando el residual y el efecto aleatorio de cada estimación para los informales).

Es importante mencionar que existen registros con estimaciones directas en la encuesta de la PEA, los ocupados y los informales donde no fue posible calcular su varianza muestral, debido a que solamente se contaba con una sola UPM. Para estos casos se desarrollan rutinas en R que permiten calcular una varianza por ajuste a una curva o a una línea recta según el comportamiento de los municipios que sí tienen varianza muestral y el monto de la población de 15 y más años de edad.

Ordenamiento según el p-valor. Posterior al cálculo de las distancias de *Mahalanobis* provenientes de un primer modelo a nivel de área de *Fay-Herriot* (utilizando el mejor predictor empírico lineal insesgado), se ordenan tales distancias de mayor a menor. Esto para cada variable objeto de estimación.

Selección del mejor modelo estadístico. Se construyen aproximadamente 1 200 modelos estadísticos para cada variable dependiente. Estos difieren entre sí en el número de municipios con muestra cuya información es insumo para el mismo modelo y en el número de vecinos que se considera para la formación de la matriz de distancias entre las unidades geográficas que forman parte de cada modelo espacial (la matriz *W* explicada en el apartado 5.2). El predictor espacial basado en el modelo a nivel de área de *Fay-Herriot* requiere de dicha matriz. En este caso se toma la distancia euclidiana entre las cabeceras municipales (utilizar otros puntos como los centroides sesgaría la correlación espacial pues pueden resultar ser lugares poco poblados o sin población). Para la obtención de esta matriz se requiere definir el número de municipios vecinos (los más cercanos y sin importar que pertenezcan a entidades federativas distintas).

Los tres primeros modelos se construyen considerando todos los municipios con muestra en la ENOE^N; entre ellos, el número de vecinos analizados para la matriz de distancias varía: cuatro, cinco o seis⁷. Se continúa de manera sucesiva: se va eliminado el primer municipio de la lista ordenada (aquel con la distancia de *Mahalanobis* más alta) y se construyen otros tres modelos estadísticos (considerando nuevamente cuatro, cinco y seis vecinos geográficos). La generación de modelos estadísticos continúa, eliminándose un municipio más en cada iteración. Esto se realiza mientras el número de municipios en los modelos representa, al menos, 30 % del total de municipios en el país, esto es recomendación de los expertos (INEGI Informa, 2017, 31m20s).

De todos los modelos se selecciona como el mejor aquel que pase el mayor número de pruebas estadísticas⁸, con mayor contundencia en el p-valor y con el mayor número de registros posible (municipios). El modelo seleccionado debe cumplir varios supuestos que dan soporte a la confiabilidad de las estimaciones. A través de métodos estadísticos, analíticos y gráficos, se verifica que los supuestos del modelo a nivel de área de *Fay-Herriot* seleccionado se cumplan. Los supuestos comprobados son:

- Linealidad, la variable de interés Y_a se asume relacionada con X_a mediante un modelo de regresión lineal (intrínsecamente).
- Normalidad de los efectos aleatorios y de los residuales.
- Homocedasticidad, misma varianza para los residuales.
- No-colinealidad, inexistencia de colinealidad o multicolinealidad.

Además, para utilizar el predictor espacial, es necesario que exista una autocorrelación espacial en las variables de interés, de lo contrario, lo mejor sería utilizar el mejor predictor empírico lineal insesgado.

⁷ Los números 4, 5 y 6 representan la media (redondeado), moda y mediana nacional del número de vecinos que tienen los municipios de México.

⁸ Hay más de 20 pruebas estadísticas. Estas pruebas evalúan los supuestos del mejor modelo estadístico.

La verificación de estos supuestos es muy importante, ya que si no se hace la verificación o no se aprueban, las estimaciones podrían carecer de confiabilidad. Los resultados de las pruebas a los supuestos se muestran en las presentaciones de resultados.

Además de la comprobación de los supuestos del modelo estadístico, los resultados son validados a través de diferentes procedimientos. Las validaciones realizadas también se encuentran en dicho documento complementario.

Estimación y ECM de todos los municipios. La librería *sae* proporciona las estimaciones y los errores cuadráticos medios por municipio del modelo espacial seleccionado. Para los municipios excluidos, aquellos con muestra en la ENOE^N que quedaron fuera del modelo espacial seleccionado y aquellos municipios sin muestra en la encuesta, se diseñaron rutinas en R con el fin de calcular sus estimaciones y errores cuadráticos mediante el modelo sintético explicado en la sección 4.4.1. De esta forma, se obtienen estimaciones que deben ser ajustadas para los todos los municipios.

Ajuste a nivel estatal. Los resultados obtenidos mediante el modelo espacial seleccionado para cada variable objeto de estimación, deben ser ajustados de forma tal que la suma de los valores municipales de una misma entidad federativa coincida con la cifra estatal proporcionada por la ENOE^N. Para el ajuste de las estimaciones se utiliza el método ajuste proporcional iterativo.

La replicación de los resultados puede ser hecha por el usuario. Se proporcionan dos *scripts* en R. En el primero ya se encuentra determinado el mejor modelo estadístico seleccionado y con él se realiza el cálculo de las estimaciones con el predictor espacial para todos los municipios. En el segundo fichero se realiza el ajuste IPF a nivel estatal.

7. Consecuencias

La información contenida en los ILMM es muy valiosa, pues anualmente se proporcionan datos laborales a nivel municipal sin necesidad de ampliar el tamaño de la muestra de la ENOE^N. Esto es muy importante, pues hasta la primera publicación de los ILMM no había antecedentes con tal periodicidad y a tal nivel de desagregación de este tipo de indicadores.

En el pasado, los análisis hechos en el contexto laboral se limitaban a conjuntos poblacionales agregados, por no haber estimaciones a nivel municipal; esta situación, en parte, se enmienda con los ILMM. De esta forma, los distintos niveles de gobierno, los tomadores de decisiones y el público en general cuentan con información relevante que pueden utilizar para fortalecer el análisis relacionado con el contexto laboral de la población municipal.

Se pretende que la publicación de los ILMM sea el punto de partida para extender el cálculo de otros indicadores derivados de encuestas que permitan generar resultados a niveles más específicos sin necesidad de aumentar los costos por cambios en el diseño de las encuestas.

Es importante mencionar que las técnicas de estimación en áreas pequeñas presentan alcances y limitaciones. Cualquier estimación lograda a través de estas técnicas depende de la calidad de las estimaciones directas obtenidas a través de la encuesta que se pretende desagregar, y de la información auxiliar (particularmente de los registros administrativos). Si la calidad de las estimaciones de la encuesta no es buena o la condición de los registros administrativos es incompleta o escasa, entonces no es seguro que se logren buenas estimaciones aplicando las técnicas de áreas pequeñas.

En el Conjunto de herramientas SAE (*Small Area Estimation*, referencia en inglés de Estimación en Áreas Pequeñas) proyecto para orientación de países, División de Estadísticas de las Naciones Unidas (UNDS) se le dedica una sección a este proyecto. Con ello se busca transmitir la experiencia del INEGI en la selección de variables, construcción del modelo, validación y difusión de resultados a fin de orientar a los países interesados. Esta aportación fue publicada recientemente en el *UN Statistics Wiki*⁹.

El INEGI se encuentra bajo un monitoreo continuo de nuevos enfoques y desarrollos de técnicas de estimación en áreas pequeñas que puedan aplicarse para el caso de México, bajo tres criterios fundamentales:

1. Se apliquen a nivel de área (no de unidad).
2. Consideren el diseño complejo de muestreo de las encuestas.
3. Que exista el *software* correspondiente.

Por el momento se ha realizado un primer ejercicio de técnicas de estimación en áreas pequeñas, bajo el enfoque de modelos multinomiales para estimar características laborales. Los resultados muestran consistencia con los obtenidos mediante el modelo *Fay-Herriot* espacial.

⁹ La siguiente liga te dirige a la experiencia de México en el UN Statistics Wiki: <https://unstats.un.org/wiki/display/SAE4SDG/Mexico>

Bibliografía

- Anselin, L. (1988). *Spatial Econometrics. Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- Benedetti, R., Piersimoni, F. & Postiglione, P. (2015). *Sampling Spatial Units for Agricultural Surveys*. Berlin: Springer-Verlag.
- Battese, G. E., Harter, R. M. & Fuller, W. A. (1988). An Error-Components model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, 83 (401), 28-36.
- Bivand, R. S., Pebesma E. J. & Gómez-Rubio V. (2008). *Applied Spatial Data Analysis with R*. New York: Springer.
- Borrego, J. (2018). Modelos de Regresión para Datos Espaciales. Recuperado de <https://idus.us.es/bitstream/handle/11441/81660/Borrego%20Sánchez%20José%20Ángel%20TFG.pdf?sequence=1&isAllowed=y>
- Chandra, H., Salvati, N. & Chambers, R. (2009). Small Area Estimation for Spatially Correlated Populations - A Comparison of Direct and Indirect Model-Based Methods. *Statistics in Transition*, 8 (2), 331-350.
- Dorfman, A. (2018). Towards a Routine External Evaluation Protocol for Small Area Estimation. *International Statistical Review*, 86(2), 259-274. <https://doi.org/10.1111/insr.12248>
- Elbers, C., Lanjouw, J.O. & Lanjouw, P. (2003). Micro-Level Estimation of Poverty and Inequality. *Econometrica Journal of the Econometric Society*, 71(1), 355-364. <https://doi.org/10.1111/1468-0262.00399>
- Fay, R. E. & Herriot, R. A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, 74(366), 269-277. <https://doi.org/10.2307/2286322>
- Getis, A. & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189-206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Goodchild, M. (1987). Spatial Analytical Perspective on Geographical Information Systems. *International Journal of Geographical Information Systems*, 1(4), 327-334.
- Griffith, D. & Arbia, G. (2010). Detecting Negative Spatial Autocorrelation in Georeferenced Random Variables. *International Journal of Geographical Information Science*, 24(3), 417-437. <https://doi.org/10.1080/13658810902832591>
- Griffith, D. & Paelinck, J. (2018). *Morphisms for Quantitative. Spatial Analysis Advanced Studies in Theoretical and Applied Econometrics*. Cham: Springer.
- Harter, R. M. (1983). Small area estimation using nested-error models and auxiliary data. [A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of DOCTOR OF PHILOSOPHY, Major: Statistics]. Iowa State University.
- Harville, D. A. (1991). Comment (en "That BLUP Is a Good Thing: The Estimation of Random Effects" de Robinson, G.K.). *Statistical Science*, 6 (1), pp. 35-39.

Henderson, C. R. (1953). Estimation of Variance and Covariance Components. *Biometrics*, 9 (2), pp. 226-252.

Hidioglou, M. A. & Patak, Z. (2009). An Application of Small Area Estimation Techniques to the Canadian Labour Force Survey. *Statistical Society of Canada -Proceedings of the Survey Methods Section, 2009 Annual Meeting in Vancouver*.

Hunsinger, E. (2008). Iterative Proportional Fitting for a Two-Dimensional Table. Recuperado de <https://edyhsgr.github.io/IPFDescription/AKDOLWDIPFTWOD.pdf>

INEGI Informa. (09 de febrero de 2017). Estimación bajo el enfoque bayesiano [Archivo de Video]. Youtube. <https://www.youtube.com/watch?v=1CXBt2ffYHo>

Instituto Canario de Estadística (ISTAC). (2008). Estadísticas Laborales Encuesta de Población Activa Metodología para la estimación en pequeñas áreas de Canarias. Recuperado de <http://www.gobiernodecanarias.org/istac/estadisticas/empleo/empleo/actividadeconomica/E30308A.html>

Instituto de Seguridad y Servicios Sociales para los Trabajadores del Estado (ISSSTE). (2021). Anuarios Estadísticos. Recuperado de <https://www.gob.mx/issste/documentos/anuarios-estadisticos>

Instituto Nacional de Estadística y Geografía (INEGI). (2015). Principales resultados de la Encuesta Intercensal 2015: Estados Unidos Mexicanos. Recuperado de https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825078966.pdf

_____ (INEGI). (2018). Norma técnica del proceso de producción de información estadística y geográfica para el Instituto Nacional de Estadística y Geografía - comité de aseguramiento de la calidad. Recuperado de https://sc.inegi.org.mx/repositorioNormateca/O_05Sep18.pdf

_____ (INEGI). (2019). Cómo se hace la ENOE: métodos y procedimientos. Recuperado de https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825190613.pdf

_____ (INEGI). (2020). Censo de Población y Vivienda 2020. Recuperado de <https://www.inegi.org.mx/programas/ccpv/2020/>

Kordos, J. (2016). Development of Small Area Estimation in Official Statistics. *STATISTICS IN TRANSITION New Series*, 17(1), 105–32. <http://dx.doi.org/10.21307/stattrans-2016-008>

Krapavickait, D. & Rudys, T. (2015). Small Area Estimates for the Fraction of the Unemployed. *Lithuanian Mathematical Journal*, 55(2), 243–254. <https://doi.org/10.1007/s10986-015-9277-9>

López-Vizcaíno, E., Lombardía, M. J. & Morales, D. (2019). Package ‘mme’: Multinomial Mixed Effects Models. Versión 0.1-6. Recuperado de <https://cran.r-project.org/web/packages/mme/index.html>

Marhuenda, Y., Molina, I. & Morales, D. (2013). Small Area Estimation with Spatio-Temporal Fay-Herriot Models. *Computational Statistics and Data Analysis*, 58, 308-325.

Martini, A. & Loriga, S. (2016). Small Area Estimation of employment and unemployment for Local Labour. Recuperado de <https://www.dst.dk/Site/Dst/SingleFiles/GetArchiveFile.aspx?fi=4299453410&fo=0&ext=formid>

Molina, I. (2018). Desagregación de datos en encuestas en hogares. Metodología de estimación en áreas pequeñas. Santiago: Naciones Unidas-CEPAL.

- Molina, I. & Marhuenda, Y. (2015). sae: An R Package for Small Area Estimation. *The R Journal*, 7(1), 81–98. <https://doi.org.10.32614/RJ2015-007>
- Molina, I., Nandram, B. & Rao, J. (2014). Small Area Estimation of General Parameters with Application to Poverty Indicators: A hierarchical Bayes approach. *The Annals of Applied Statistics*, 8(2), 852-885.
- Molina, I. & Rao, J. N. K. (2010). Small Area Estimation of Poverty Indicators. *The Canadian Journal of Statistics*, 38(3), 369-385.
- Molina, I., Saei, A. & Lombardía, M. J. (2007). Small Area Estimates of Labour Force Participation under a Multinomial Logit Mixed Model. *Journal of the Royal Statistical Society Serie A*, 170(4), 975-1000.
- Montgomery, D. C., Peck, E. A. & Vining G. G. (2006). *Introducción al análisis de regresión lineal* (V. González Pozo, Trans.). 3rd ed. CDMX: Compañía Editorial Continental (Obra original publicada en 2001).
- Münnich, R., Burgard, J. P., Gabler, S., Ganninger, M. & Kolb, J. (2016). Small Area Estimation in the German Census 2011. *STATISTICS IN TRANSITION New Series*, 17(1), 25–40. <https://doi.org/10.21307/stattrans-2016-004>
- Paradis, E., et al. (2021). Moran's Autocorrelation Coefficient in Comparative Methods. Recuperado de <https://cran.r-project.org/web/packages/apec/index.html>
- Pratesi, M. & Salvati, N. (2007). Small Area Estimation in the Presence of Correlated Random Area Effects. *Journal of Official Statistics*, 25(1), 37-53.
- _____ (2007). Small Area Estimation: the EBLUP Estimator Based on Spatially Correlated Random Area Effects. *Statistical methods and applications*, 17(3), 113-141. <https://doi.org/10.1007/s10260-007-0061-9>
- Puspongoro, N. H. & Rachmawati, R. N. (2018). Spatial Empirical Best Linear Unbiased Prediction in Small Area Estimation of Poverty. *Procedia Computer Science*, 135, 712-718. <https://doi.org/10.1016/j.procs.2018.08.214>
- Ramos, R. (2017). Indicadores sociodemográficos a nivel área pequeña para la Encuesta Intercensal 2015: Incorporación de efectos espaciales y temporales. Recuperado de https://www.inegi.org.mx/eventos/2017/conacyt/doc/p_RogelioRamos.pdf
- Rao, J. N. K. (2003). *Small Area Estimation*. Hoboken: Wiley.
- Rao, J. N. K. & Mingyu, Y. (1994) Small Area Estimation by Combining Time Series and Cross-Sectional Data. *Canadian Journal of Statistics*, 22(4), 511-528. <https://doi.org/10.2307/3315407>
- Rao, J. N. K. & Molina, I. (2015). *Small Area Estimation*. Second Edition. Hoboken: Wiley.
- Rawlings, J. O., Pantula, S. G. & Dickey, D. A. (1998). *Applied Regression Analysis: A Research Tool*. 2nd ed. New York: Springer.
- Robinson, G. K. (1991). That BLUP Is a Good Thing: The Estimation of Random Effects. *Statistical Science*, 6 (1), 15-32.
- Rousseeuw, P. J. & Van Driessen K. (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41(3), 212-223. <https://doi.org/10.1080/00401706.1999.10485670>

Scealy, J. (2010). Small Area Estimation Using Multinomial Logit Mixed Model with Category Specific Random Effects. Australian Bureau of Statistics. Catalogue number 1351.0.55.029.

Singh, B., Shukla, G. & Kundu, D. (2005). Spatio-Temporal Models in Small Area Estimation. Statistics Canada, 31(2),183-195.

Song, S. (2011). Small area estimation of unemployment: From feasibility to implementation. Wellington: New Zealand Association of Economists Conference.

Suárez, C. M. A., Aguilar, M. G. & González M. R. (2015). Estimación del ingreso por trabajo en los municipios y las delegaciones de México utilizando técnicas de estimación para áreas pequeñas. Realidad Datos y Espacio Revista internacional de Estadística y Geografía, 6 (3), 44-61.

The EURAREA Consortium. (2004). Project reference Volume Vol 1. Enhancing Estimation Techniques to meet European Needs.

The EURAREA Consortium. (2004). Project reference Volume Vol 3. Enhancing Estimation Techniques to meet European Needs.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, 46(supplement 1), 234-240. <https://doi.org/10.2307/143141>

Torben, K. (2010). Spatial model selection and spatial knowledge spillovers: a regional view of Germany. ZEW Discussion Papers, 10-005. Recuperado de <https://www.zew.de/en/publications/zew-discussion-papers>

Tzavidis, N., Zhang, L., Rojas Perilla, N., Luna, A. & Schmid T. (2018). From Start to Finish: a Framework for the Production of Small Area Official Statistics. Journal of the Royal Statistical Society: Serie A (Statistics in Society), 181(4), 927-979.

United Nations Department of Economic and Social Affairs of the (UN DESA). (2017). Principles and Recommendations for Population and Housing Censuses. Rev. 3. New York: United Nations.

United States Census Bureau. (2018). Local Area Unemployment Statistics. Handbook of Methods. Recuperado de <https://www.bls.gov/opub/hom/lau/home.htm>

Vilalta y Perdomo, C. J. (2005). Cómo enseñar autocorrelación espacial. Economía, Sociedad y Territorio, V(18),323-333. Recuperado de <https://www.redalyc.org/articulo.oa?id=11101804>