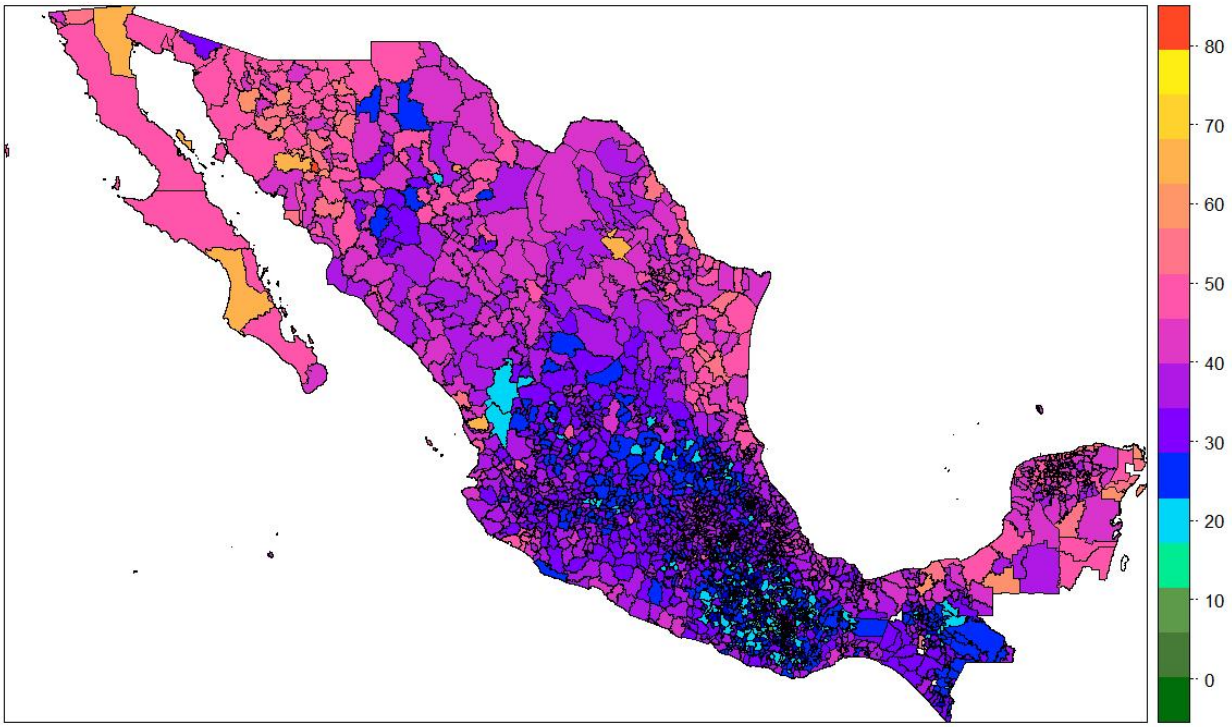
Nota metodológica



Prevalencia de Obesidad, Hipertensión y Diabetes para los Municipios de México 2018

Estimación para Áreas Pequeñas

Julio, 2020

Contenido

1 Resumen

Se estimó la prevalencia de la Obesidad, Hipertensión y Diabetes para los municipios y alcaldías del país para el año de 2018. Esto se realizó mediante técnicas de Estimación para Áreas Pequeñas (las siglas en inglés son SAE, en este documento se abreviará usando las siglas en español EAP), dado que no existe fuente de información reciente con tal nivel de desagregación geográfica. Para tal efecto, se utilizó información combinada de distintas fuentes existentes, a partir de la cual se construyeron modelos estadísticos en los que se tomaron como variables dependientes el porcentaje de población de 20 años y más con obesidad, el porcentaje de población de 20 años y más con diagnóstico previo de hipertensión y el porcentaje de población de 20 años y más con diagnóstico previo de diabetes, cuya información proviene de la Encuesta Nacional de Salud y Nutrición (ENSANUT, 2018). Las variables auxiliares fueron seleccionadas de las Estadísticas Vitales de defunciones registradas, la Estadística de Salud en Establecimientos Particulares, la Encuesta Intercensal 2015, así como de las proyecciones de población.

De esta forma se complementa la información de las fuentes de información primarias que proporcionan cifras nacionales, por entidad federativa y dominios mayores al municipio.

2 Introducción

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 con respecto a la planeación, diseño y evaluación de programas sociales, con 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, y para responder a las tendencias de la dinámica mundial que implican una continua innovación de los métodos y formas para registrar la realidad.

La demanda creciente de la sociedad para tener respuestas satisfactorias a sus necesidades de información por parte de las oficinas nacionales de estadística, se ha convertido en el transcurso del tiempo en una constante universal. Particularmente, los gobiernos locales requieren contar con información actualizada y desagregada para niveles geográficos pequeños, 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 los municipios, en el caso de diseños para cifras nacionales o por entidad federativa, o bien, dominios temáticos explícitos que originalmente no se consideran en el diseño muestral.

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

Específicamente, el trabajo estadístico que se reporta en este documento, responde al requerimiento de contar con información referida a los municipios de México de la prevalencia de enfermedades de alto impacto individual y social, ya que no es posible obtenerla en forma directa con las fuentes de información actuales; por lo que se utilizaron técnicas estadísticas para la estimación de las variables de interés correspondientes: Obesidad, Hipertensión y Diabetes. Y de esta forma complementar las fuentes de información primarias que proporcionan cifras nacionales y por entidad federativa.

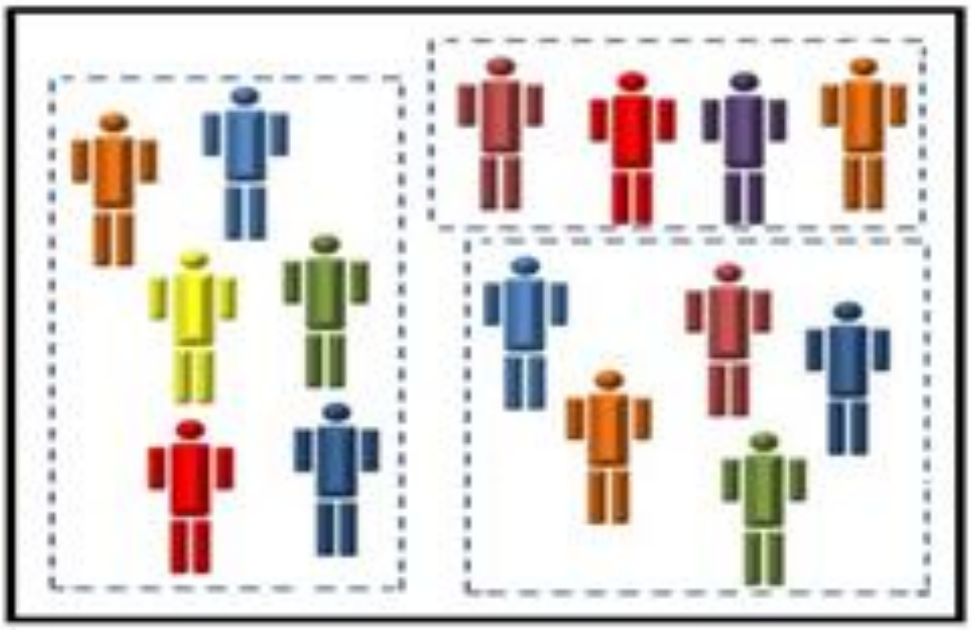
3 Objetivo

Estimar la proporción de la población de 20 años y más que padece enfermedades de Obesidad, Hipertensión y Diabetes para los municipios de México, mediante técnicas de Estimación para Áreas Pequeñas (EAP), a fin de ampliar la oferta de información derivada de la Encuesta Nacional de Salud y Nutrición (ENSANUT, 2018) y apoyar la toma de decisiones.

4 La estimación de 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, como se muestra en la Ilustración 1.

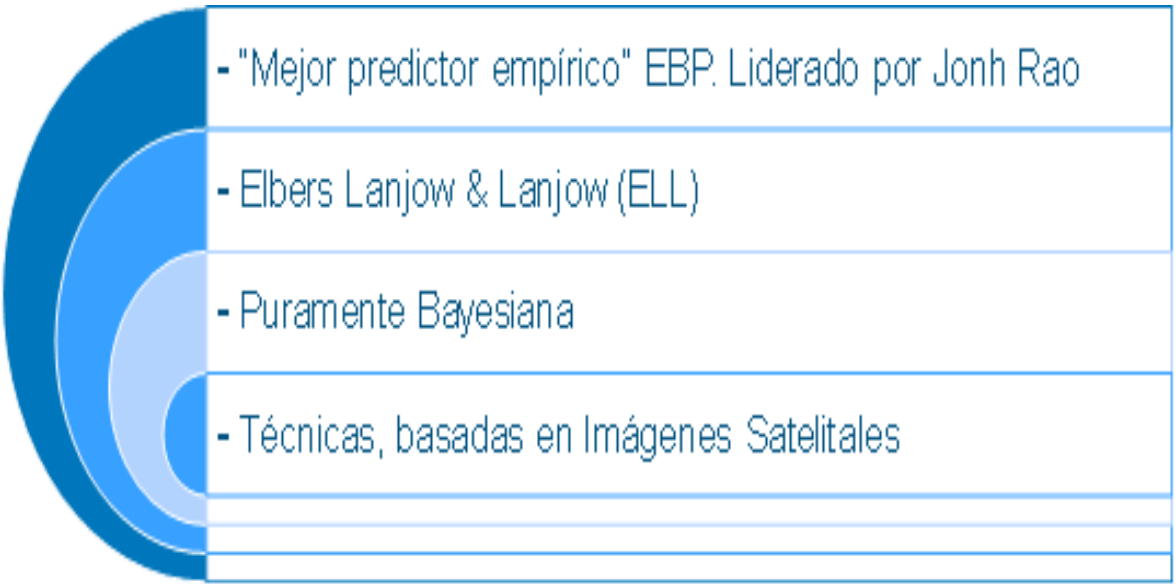
Ilustración 1. Subconjunto de poblaciones.



Fuente: Elaboración a partir de las fuentes consultadas.

Las técnicas de EAP son herramientas estadísticas relativamente novedosas que permiten estimar parámetros, 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. Algunas de estas técnicas se muestran en la Ilustración 2.

Ilustración 2. Técnicas de estimación de áreas pequeñas.



Fuente: Elaboración a partir de las fuentes consultadas.

Los principales métodos para estimar los parámetros generales de áreas pequeñas son: el Mejor Predictor Empírico Lineal Insesgado (EBLUP) basado en el conocido modelo de nivel de área de Fay-Herriot Fay y Herriot (1979); el método de Elbers (2003), llamado método ELL y utilizado por el Banco Mundial; mejor método empírico o empírico Bayes (EB) de Molina y Rao (2010); método jerárquico Bayes (HB) de Molina (2014); y otras variantes del método EB para tratar el muestreo en dos etapas o el muestreo informativo de unidades, en Molina, Isabel & Nandram, Balgobin & Rao, J. (2014).

5 Proceso para la generación de las estadísticas de áreas pequeñas

El INEGI cuenta con un proceso estándar para la generación de estadística básica y derivada plasmado en el instrumento normativo "Norma Técnica del Proceso de Producción de Información Estadística y Geográfica" (INEGI, 2018), que a su vez toma como premisa al Modelo Genérico del Proceso Estadístico (GSBPM por sus siglas en inglés, Generic Statistical Business Process Model) de la Comisión Económica de las Naciones Unidas para Europa (UNECE).

Por lo anterior, el proceso para la generación de estadísticas de EAP fue alineado a dicha Norma, como se ilustra en la Tabla 1.

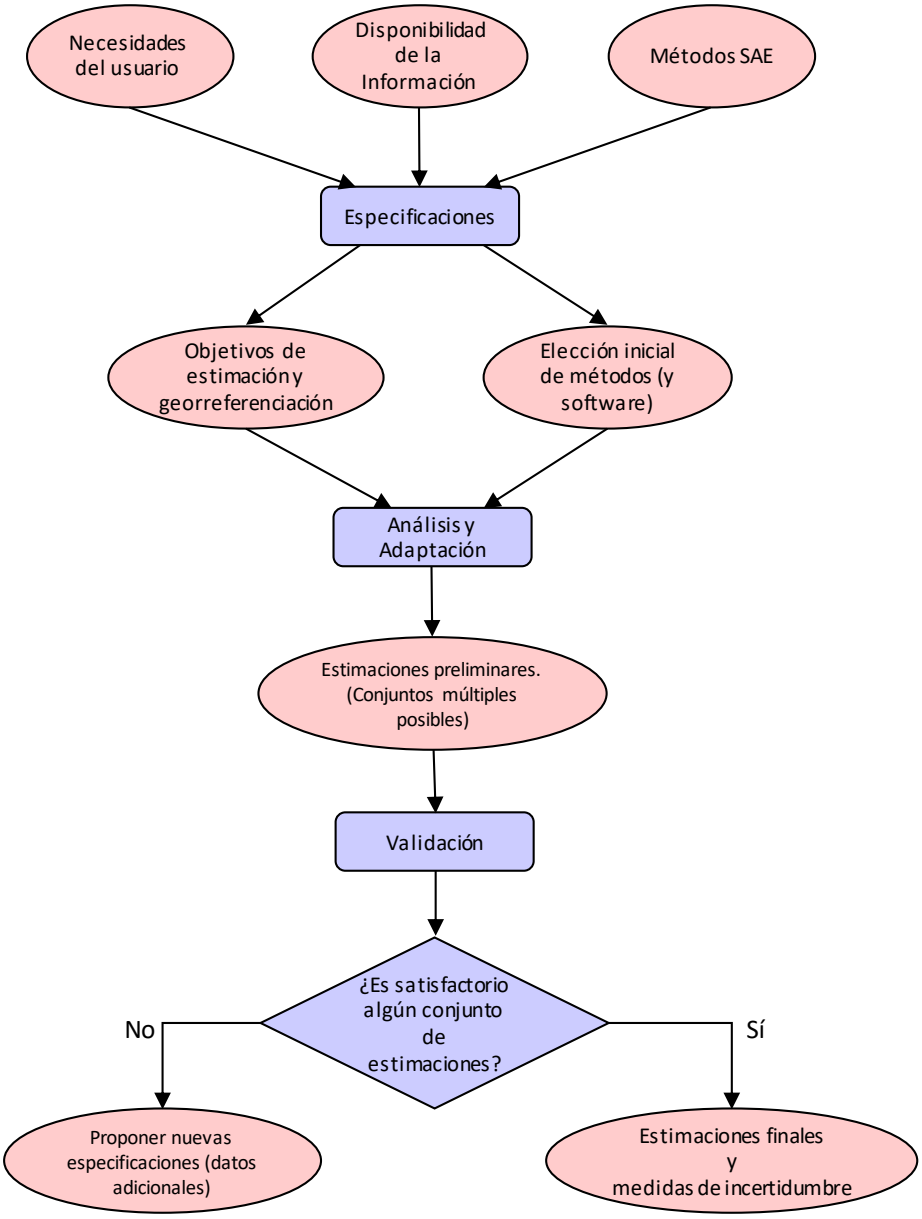
Tabla 1. Proceso de generación para las estadísticas de áreas pequeñas.

FASES	SUBPROCESOS
Especificación de necesidades	Determinación de las variables objeto de estimación conforme a las necesidades de información no contenidas en las estadísticas tradicionales.
Diseño	Especificación del modelo estadístico EAP de acuerdo a la disponibilidad de información.
Construcción	Preparación de la información fuente de la ENSANUT 2018, de los registros administrativos y de variables de la Encuesta Intercensal disponible a nivel municipal correlacionada con las variables de interés. Identificación y elaboración de software para EAP.
Procesamiento	Selección de variables auxiliares. Incorporación de variables geográficas. Obtención de parámetros del modelo. Comprobación de los supuestos del modelo estadístico de EAP. Detección de valores extremos. Nuevos parámetros del modelo. Diagnóstico de resultados. Procesamiento de datos municipales. Obtención de tabulados.
Análisis	Validación de resultados por comparación con otras fuentes de información. Validación de resultados con expertos.
Difusión	Preparación de información. Liberar la información a 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: Elaboración a partir de las fuentes consultadas.

En la revisión de literatura, Tzavidis, Zhang, Luna, Schmid y Rojas-Perilla (2018) en su artículo "*From start to finish: a framework for the production of small area official statistics*", publicado en *Journal of the Royal Statistical Society, Statistics and Society Series A, p. 927-979,* sugieren un esquema general para la producción de estadísticas de áreas pequeñas, que en términos generales coincide con el proceso estándar del INEGI (ver la Ilustración 3).

Ilustración 3. Diagrama general para la producción de estadísticas de áreas pequeñas.



Fuente: Tzavidis, Zhang, Luna, Schmid y Rojas-Perilla (2018) en su artículo "*From start to finish: a framework for the production of small area official statistics*".

6 Base teórica y conceptual

En este apartado se describen las fuentes de datos para estimar las variables de interés, los métodos utilizados de estimación para áreas pequeñas y sus consideraciones para la construcción del modelo.

6.1 Fuentes de datos

En la actualidad no existe fuente de datos para ofrecer cifras sobre prevalencia de enfermedad con desglose geográfico por municipio, de manera que se ha optado por utilizar fuentes de información distintas y aplicar técnicas de EAP: por un lado, se utiliza la Encuesta Nacional de Salud y Nutrición 2018 (ENSANUT), de donde se toman las variables objeto de estimación; su diseño estadístico garantiza resultados precisos a nivel nacional, para cada entidad federativa y para otros niveles geográficos superiores a los de municipio. En tanto que, por otro lado, se utiliza la Encuesta Intercensal 2015, registros administrativos de defunciones y de infraestructura hospitalaria, fuentes de las cuales se obtienen las variables auxiliares.

En la Ilustración 4 se muestra el vínculo entre las variables de interés que se desean estimar, a partir de la encuesta base de salud y nutrición; y las variables auxiliares potenciales provenientes de censos, encuestas mayores y registros administrativos, disponibles a niveles de desagregación municipal. Fueron obtenidas de un análisis teórico-conceptual al interior del Instituto, y de la revisión documental de ejercicios desarrollados en organismos nacionales de estadística de otros países.

Ilustración 4. Vínculo entre variables de interés.

Fuente: Elaboración a partir de las fuentes consultadas.

6.2 Variables del modelo

En este apartado se describen las variables objeto de estimación a nivel municipal y las variables auxiliares para la construcción del modelo.

6.2.1 Variables por estimar a nivel municipal

En este trabajo, las variables objeto a EAP son:

- Porcentaje de población de 20 años y más con obesidad (Obesidad).
- Porcentaje de población de 20 años y más con diagnóstico previo de hipertensión (Hipertensión).
- Porcentaje de población de 20 años y más con diagnóstico previo de diabetes (Diabetes).

6.2.2 Variables auxiliares

Para la estimación de las variables, es necesario recurrir a otras fuentes de información, las cuales deben estar relacionadas conceptual y estadisticamente con la ENSANUT.

Por lo anterior, se hizo la búsqueda y procesamiento de las fuentes, con fines de tener un conjunto de variables que permitan una estimación estadísticamente adecuada. De esta forma, se consideró un conjunto inicial de 67 variables, clasificadas en alguna de las siguientes 7 temáticas: estructura etaria, servicios de salud, características educativas, características económicas, vivienda, defunciones e infraestructura hospitalaria.

6.3 Métodos de estimación de áreas pequeñas

Durante el desarrollo del proyecto se han considerado diversos métodos de estimación a nivel de áreas pequeñas, después de realizar el análisis y ajustes diversos, primero se decidió utilizar el Mejor Predictor Empírico Lineal Insesgado (EBLUP, por sus siglas en inglés), que es una combinación de la estimación directa y la sintética (producto de un modelo lineal mixto).

A continuación se describe, en forma concisa, la técnica de EAP utilizada para la estimación de prevalencia de Obesidad, Hipertensión y Diabetes para los municipios de México y los supuestos estadísticos del modelo.

6.3.1 Modelo a nivel área EBLUP

- 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): $\theta_a^M = X_a\beta + Zv_a + e_a$
- Predictor BLUP (B): $\theta_a^B = X_a\beta + v_a = X_a\beta + \gamma_a\left(\hat{\theta}_a^D - X_a\beta\right)$
- Predictor empírico EBLUP(EB): $\hat{\theta}_a^{(EB)} = x_a^T\hat{\beta}^{FH} + \hat{v}_a = x_a^T\hat{\beta}^{FH} + \hat{\gamma}_a\left(\hat{\theta}_a^D - x_a^T\hat{\beta}^{FH}\right)$

  $= \hat{\theta}_a^S + \hat{\gamma}_a\left(\hat{\theta}_a^D - \hat{\theta}_a^S\right) = \hat{\gamma}_a\hat{\theta}_a^{(D)} + (1-\hat{\gamma}_a)\,\hat{\theta}_a^S$
- Estimador sintético (S): $\hat{\theta}_a^S = x_a^T\hat{\beta}^{FH}$
- Función log-verosímil (LV): $l(\beta, \sigma_v^2, \psi) = -\frac{m}{2}ln(2\pi) - \frac{1}{2}\sum_{a=1}^{m} ln(\sigma_v^2 + \psi_a)$

  $-\frac{1}{2}\sum_{a=1}^{m}\left(\hat{\theta}_a^D - x_a^T\hat{\beta}\right)^2 / (\sigma_v^2 + \psi_a)$

Donde,

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

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

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

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

El estimador directo (*D*) es el valor obtenido por el operativo de la encuesta; puede ser un valor como la proporción estimada de la variable de interés en el municipio (por ejemplo: el porcentaje de población de 20 años y más con obesidad) o un total del mismo (por ejemplo: el número total de personas de 20 años y más con diágnostico previo de diabetes) y puede ser expresado 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 de interés, siempre y cuando las variables regresoras tengan una relación lineal con la misma y se cumplan los supuestos del modelo.

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

El predictor BLUP (Best Linear Umbiased Prediction) o el Mejor Predictor Lineal Insesgado, es la expresión matemática para la estimación del modelo lineal mixto para áreas pequeñas, es decir, para combinaciones lineales de párametros de regresión y realizaciones de los efectos aleatorios y es propia para datos que no son del tipo panel o datos que no son longitudinales, sin embargo, el modelo sigue contemplando la variable aleatoria de los efectos aleatorios $\boldsymbol{\nu_a}$.

Como las variables aleatorias no se pueden estimar sino solamente predecir, Harville (1991) propone el predictor empírico EBLUP (Empirical Best Linear Umbiased Prediction) o el Mejor Predictor Empírico Lineal Insesgado sustituyendo σ_v^2 por $\hat{\sigma}_v^2$, el modelo EBLUP es el que se aplica para obtener las estimaciones de la prevalencia de enfermedad cuando existe muestra de la ENSANUT en el municipio.

La fórmula matemática del predictor empírico EBLUP ($\hat{\theta}_a^{(EB)} = \hat{\gamma}_a \hat{\theta}_a^{(D)} + (1 - \hat{\gamma}_a)\, \hat{\theta}_a^{S}$) expresa que las estimaciones obtenidas bajo este modelo (en términos de Inferencia Bayesiana la infomación posteriori) son una combinación entre lo que se observa de la ENSANUT (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 ($\boldsymbol{\hat{\sigma}_v^2}$) entre la suma de la varianza de los efectos aleatorios más la varianza observada del estimador directo de la ENSANUT ($\boldsymbol{\hat{\psi}_a}$)**,** por lo 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 ENSANUT tendrá más peso en el predictor EBLUP, 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 el predictor EBLUP.

Consultar la sección 6.1.2 de Rao & Molina (2015), como referencia de cómo obtener la varianza de los efectos aleatorios.

La estimación del error cuadrático medio del predictor empírico EBLUP 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,

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

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

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

g_1 se asocia al error de muestreo de la estimación directa de la ENSANUT, g_2 se asocia al error del estimador sintético y 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).

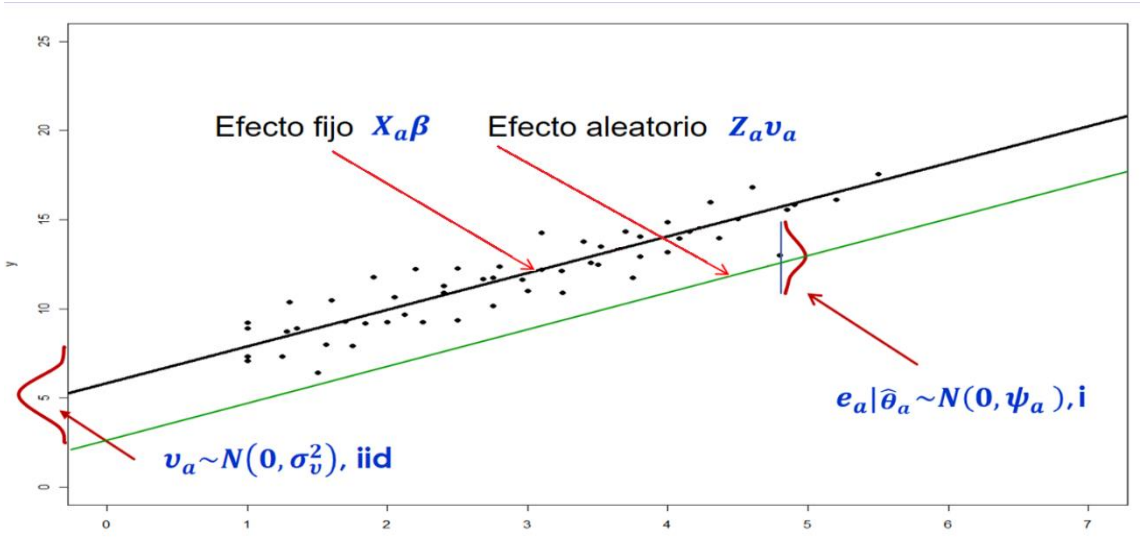
Por otra parte, el estimador sintético (*S*) o efecto fijo del modelo mixto, es la parte del modelo de áreas pequeñas que se aplica para obtener las estimaciones de prevalencia de enfermedad cuando no existe muestra de la ENSANUT en el municipio o bien porque el municipio fue excluido como parte del modelo por ser considerado como un valor extremo.

La estimación del error cuadrático medio del estimador sintético (*S*) para municipios que no tienen muestra $l = m + 1 \ldots M$, tiene la siguiente 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 $(\boldsymbol{\beta}, \boldsymbol{\sigma_v^2})$ por máxima verosimilitud o máxima verosimilitud restringida, utilizando derivadas parciales para obtener la estimación de $\boldsymbol{\beta}$ y algoritmos iterativos cuando no se tenga una solución analítica para estimar $\boldsymbol{\sigma_v^2}$ **,** por lo que respecta a ψ se considera como constante conocida que ofrece la ENSANUT.

Gráfica 1. El modelo en forma gráfica.



Fuente: Elaboración propia.

Para ilustrar en forma gráfica el modelo lineal mixto, se realiza un diagrama de dispersión (ver la Gráfica 1) en el cual el eje *Y* contiene los valores de la encuesta del parámetro de interés a nivel de los municipios muestreados (aquí hay que recordar que no todos los municipios de México se visitan en el operativo de la encuesta) y, por otra parte, en el eje de las *X* se presenta a nivel

municipal el valor de la variable auxiliar (como se comentó, puede ser un registro administrativo, o bien, un valor de la Encuesta Intercensal).

La problemática que se tiene con la información, es que se necesitan al menos dos puntos para obtener un modelo lineal por municipio y la información solamente cuenta con uno. Para solventar esta situación, se calcula una sola recta de regresión para todos los puntos (el efecto fijo); posteriormente, para cada punto, 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 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 del 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 muy diferente para cada municipio (varianza *within*), por otra parte, dentro del proceso de estimación, estos errores debidos al muestreo, se mantendrán como información fija y conocida.

Los supuestos del modelo son:

- Se asume disponible el dato auxiliar a nivel de área X_a (nivel municipio).
- Se asume que una muestra s_a de tamaño n_a, proviene de las N_a unidades en el área *a* (*a* = 1,… , *A*).
- Linealidad, la variable de interés Y_a se asume relacionada con X_a mediante un modelo de regresión lineal.
- Normalidad de los efectos aleatorios y de los residuales.
- Homocedasticidad, igualdad de varianzas en los residuales.
- No-colinealidad, inexistencia de colinealidad o multicolinealidad.

6.3.2 Modelo a nivel área SEBLUP. Se incorpora el componente espacial

El modelo de dependencia espacial postula que el afecto aleatorio del área *a* se expresa como un modelo autoregresivo espacial (SAR por sus siglas en inglés):

$$\upsilon = \rho W \upsilon + u$$

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

El modelo mixto espacial para áreas pequeñas se expresa:

$$\theta_a^{SM} \;\; = \; X_a\beta + Z(I_m - \rho W)^{-1}u_a + e_a$$

De esta forma, el nuevo modelo da cuenta de la importancia de la influencia de las demás áreas sobre el área a, donde ρ es el coeficiente de correlación espacial, W es la matriz del inverso de las distancias entre las cabeceras municipales de los municipios (en el caso de la Ciudad de México al no tener cabeceras municipales oficiales se consideraron las localidades con clave 001), u es un vector de términos de error con media cero y varianza constante τ_u^2.

Por otra parte, υ se distribuye como una función normal de media cero y matriz de varianzas y covarianzas:

$$G(\tau_u^2, \rho) = \tau_u^2[(I_m - \rho W)^T(I_m - \rho W)]^{-1} = \left(\widehat{\sigma_\upsilon^2}\right)'$$

Utilizando la notación sigma y términos algebraicos se tiene:

$$\upsilon_a = \rho \sum_{a<>j} w_{aj}\upsilon_j + u_a \qquad , \quad \gamma_a' = \frac{\left(\widehat{\sigma_\upsilon^2}\right)'}{\left(\widehat{\sigma_\upsilon^2}\right)' + \hat{\psi}_a}$$

Al tomar la última igualdad, se tiene la expresión para el predictor empírico espacial SEBLUP,

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

En notación matricial, el predictor espacial SEBLUP se define:

$$\hat{\theta}_a^{(SEB)} = x_a^T\hat{\beta}^{FH'} + z_m^T[\hat{\sigma}_\upsilon^2(I_D - \hat{\rho}W)^{-1}(I_D - \hat{\rho}W^T)^{-1}]z^T *$$

$$\left[\left[diag\left(\hat{\psi}_a\right) + \hat{\sigma}_\upsilon^2 z[(I_D - \hat{\rho}W)^{-1}(I_D - \hat{\rho}W^T)^{-1}]z^T\right]\right]^{-1}\left(\hat{\theta}_a^D - x_a^T\hat{\beta}^{FH'}\right)$$

Donde,

$\hat{\beta}^{FH'} = \; (X^TV^{-1}X)^{-1}\,X^tV^{-1}\hat{\theta}_a^D$

$V = diag\left(\hat{\psi}_a\right) + \hat{\sigma}_\upsilon^2 z[(I_D - \hat{\rho}W)^{-1}(I_D - \hat{\rho}W^T)^{-1}]z^T$

$z_m = es\ un\ vector\ de\ tamaño\ 1 * m = (0,0,1,0,\dots)\ con\ 1\ en\ la\ m - ésima\ posición$

$z \;\; = elementos\ de\ la\ matriz\ de\ diseño\ de\ los\ efectos\ aleatorios\ generalmente\ matrices\ diagonales.$

7 Preparación de Información

7.1 Las variables auxiliares seleccionadas

La Encuesta Intercensal 2015 se llevó a cabo con la finalidad de actualizar la información sociodemográfica a la mitad del periodo comprendido entre el Censo de 2010 y el de 2020. Aborda temas presentes en los últimos censos y guarda comparabilidad con ellos, pero también incorporó temas de reciente interés. Con un tamaño de muestra de 6.1 millones de viviendas, proporciona información a nivel nacional, entidad federativa, municipio y para cada una de las localidades con 50 mil o más habitantes. Con esta base se hizo la selección y procesamiento de las variables candidatas a ser auxiliares, con lo que se tuvo un conjunto de 48 variables de esta fuente estadística, clasificadas en las siguientes temáticas: servicios de salud, características educativas, características económicas y características de la vivienda.

Los registros administrativos de defunciones es otra fuente de información de donde se seleccionaron variables auxiliares. Su característica es generar información sobre el volumen de las defunciones registradas en el país, así como algunas características por edad y sexo de los fallecidos y las principales causas que originan los decesos. Para fines de la EAP, se estimó la proporción respecto a la población de 20 años y más para diez causas de muerte por enfermedad, seleccionadas por su mayor incidencia a nivel nacional.

A partir la Estadística de Salud en Establecimientos Particulares también se obtuvieron dos variables, referidas a la densidad de consultorios y de camas hospitalarias.

Finalmente, otro importante insumo fueron las estimaciones de población por grupos de edad, a partir de ellas se estimaron diez indicadores que describen la composición etaria y por sexos de la población.

De esta forma se tuvo un conjunto de 70 variables, a partir de las cuales se analizó su relación conceptual y estadística para cada una de las enfermedades objeto de estudio.

7.2 El Índice de Moran

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): "*Everything is related to everything else, but closer things more so*" ("Todo está relacionado con todo lo demás, pero las cosas más cercanas, lo están aún más").

En general, si los objetos o actividades se parecen mucho entre sí, se dice que existe una autocorrelación espacial positiva; si por el contrario, 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, hay una correlación espacial negativa entre casos de delincuencia y la presencia policial).

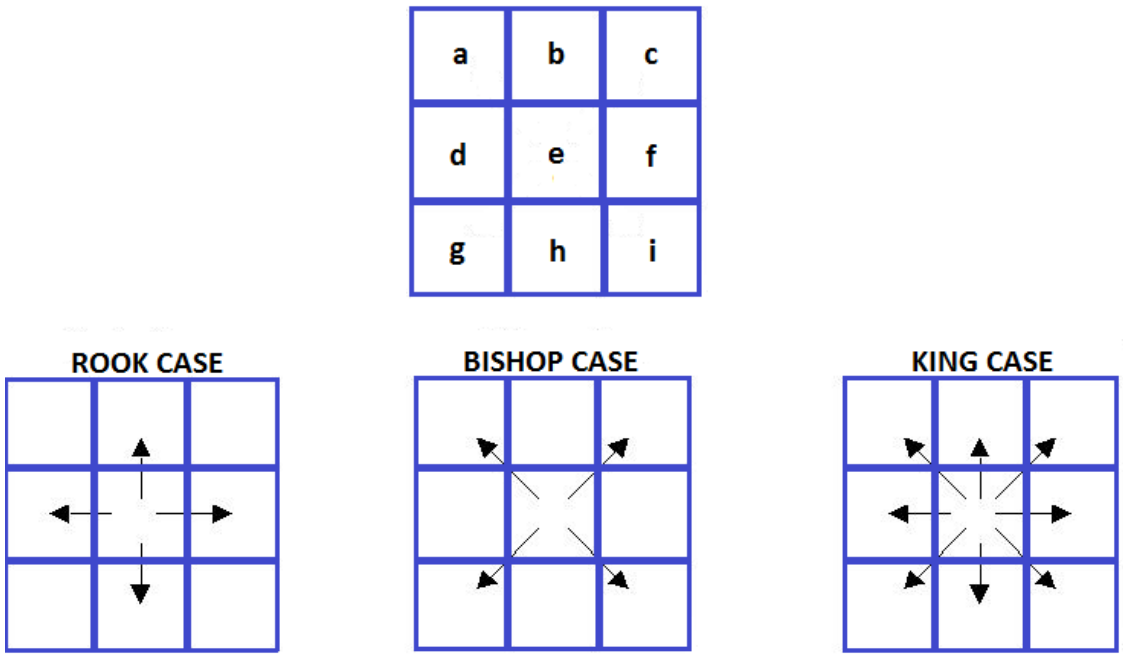
Por lo tanto, la autocorrelación espacial tiene que ver tanto con la localización geográfica como con los valores hallados de la variable de prevalencia que se esté estudiando. Para determinar si el patrón de distribución espacial dista de ser meramente aleatorio, debe utilizarse un índice de comparación y responder a un mismo principio: poner en relación las diferencias en 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} está 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 bien una medida de la proximidad de los valores a y *j*, en otra dimensión (por ejemplo, distancia euclídea, distancia esférica, distancia de Manhattan, etcétera).

Cuando la matriz P_{aj} representa una medición de la contigüidad, la matriz se compone de ceros y unos, según se considere la existencia de contigüidad o no, entre las localizaciones geográficas. En general, los datos geográficos se presentan en un mapa continuo, como en el ejemplo de la Ilustración 5 (donde hay sólo nueve localizaciones próximas). La contigüidad respecto de la localización central puede entonces definirse en forma general de las tres maneras expuestas en la misma ilustración.

Ilustración 5. Diferentes tipos de contigüidad.



Fuente: Spatial Autocorrelation and Spatial Regression Elizabeth Root.

Tomando como referencia a la localización “e”: en el caso Rook, se consideran adyacentes las localizaciones laterales, es este caso serían las localizaciones “b”, “d”, “f”, “h”; el caso Bishop, por el contrario, analiza las relaciones de proximidad diagonales, y se consideraría vecinas las localizaciones “a”, “c”, “g”, “i”; el criterio de King combina los dos anteriores.

Un método muy conocido para obtener una medida de autocorrelación espacial es el Índice de Moran (I de Moran).

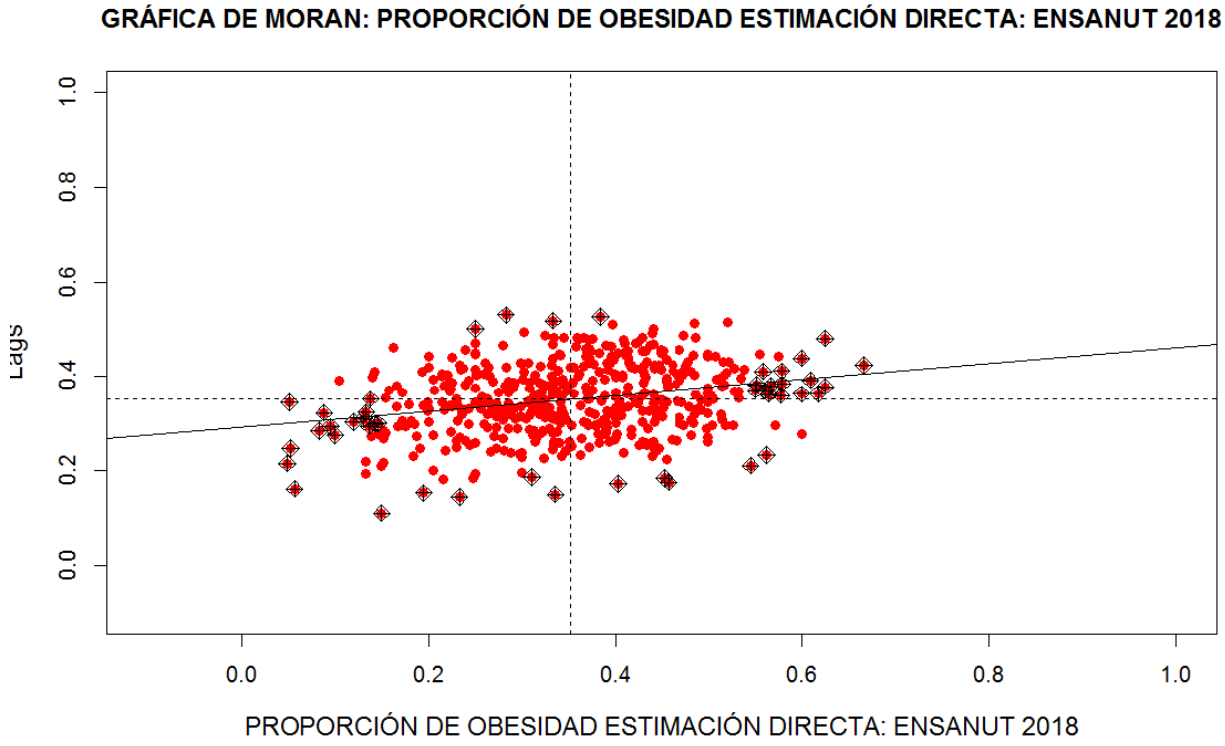
El índice calcula el valor medio y la varianza para la prevalencia de enfermedades que se evalúa. A continuación, resta el valor medio en cada valor del atributo, lo que crea una desviación del valor medio. Los valores de desviación para todos los municipios vecinos (los municipios dentro del criterio 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. 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) como se muestra en las siguientes expresiones:

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

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

La gráfica de dispersión de los datos de Obesidad y sus correspondientes "Lags" se muestra en la Gráfica 2:

Gráfica 2. Dispersión del índice de Moran para Obesidad.



Fuente: Elaboración propia.

Los "Lags" municipales se obtienen sumando el producto de los valores contenidos en la matriz W por el valor observado de la Obesidad, Hipertensión y Diabetes para cada municipio.

El valor del Índice de Moran es igual a:

Obesidad = 0.1684

Hipertensión = 0.1013

Diabetes = 0.0615

Mientras que la correlación de Pearson es igual a :

Obesidad = 0.2570

Hipertensión = 0.1718

Diabetes = 0.1222

El software R dispone de una prueba de inferencia estadística en donde la hipótesis nula se establece que los datos se distribuyen geográficamente de una forma aleatoria, mientras que la hipótesis alterna asume que la información se agrupa espacialmente. El resultado de esta prueba y su *p*-valor asociado para cada una de las tres variables se muestra a continuación:

Obesidad *p*-valor = 0.0000001804

Hipertensión *p*-valor = 0.00034

Diabetes *p*-valor = 0.0167

Si se toma un valor de significancia igual a 0.05, entonces se puede concluir que se rechaza la hipótesis nula y por consecuencia se concluye que existe correlación espacial en un sentido positivo, es decir, valores grandes en alguna enfermedad se agrupan en distancias grandes entre vecinos, mientras que valores bajos de alguna enfermedad se agrupan en distancias pequeñas entre vecinos.

8 Construcción del modelo

8.1 Software disponible y complementado con desarrollos propios

Se realizó un análisis para la selección del software de cómputo estadístico necesario para la implementación de EAP. Se estudiaron los que se describen a continuación.

Programas elaborados por EURAREA (2004) en la plataforma SAS para la EAP 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 ENSANUT y los registros administrativos disponibles. Se concluyó que lo desarrollado en esta plataforma ofrece resultados básicos y limitados para las necesidades del Instituto.

Librería “mme” del CRAN de R, desarrollada por E. Lopez-Vizcaino, M.J. Lombardia y D. Morales. Se replicaron las rutinas contenidas en esta librería, y se desarrollaron nuevas con el fin de complementar la salida de productos como las estimaciones ajustadas y los errores estándar para todos los municipios y alcaldías del país. Al concluir esta actividad, se determinó que el software ofrecía una gran ventaja con respecto a otros, al utilizar modelos *multinomiales* con tres categorías, en donde se encuadran las cifras para dar un total global, o bien por categorías; sin

embargo, 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.

Librería "sae" del CRAN de R, desarrollada por Isabel Molina y Yolanda Marhuenda. La primera investigadora, por cierto, es coautora junto a J.N.K. Rao del libro clásico "Small Area Estimation". Esta librería ha sido el soporte básico para la estimación de la prevalencia por enfermedad de los municipios de México; 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.

Una vez que se preparan los insumos provenientes de la ENSANUT como son las variables auxiliares y las distancias entre municipios es necesario correr un fichero de R (PROGRAMA ENSANUT 2018.r) para obtener las estimaciones a nivel municipal sobre la prevalencia de las enfermedades seleccionadas.

El fichero de R contiene un ajuste final de las estimaciones con las cifras de cada entidad federativa, en esta parte del proceso se utilizan dos tablas. La primera contiene los datos estimados a nivel municipal; la segunda contiene los totales por estado con las columnas Obesidad, Hipertensión y Diabetes de la ENSANUT.

Estas tablas son el insumo para la aplicación del método de Ajuste Proporcional Iterativo (IPF por sus siglas en inglés). Se toman dos pivotes, que para este caso son los totales por estado y el valor de Obesidad, Hipertensión y Diabetes para cada municipio. Este último se ajusta previamente para que la suma de los municipios sea igual al total estatal, lo cual se hace dividiendo el total estatal de la ENSANUT entre la suma de las estimaciones de cada enfermedad mencionada de los municipios del estado correspondiente, y multiplicando el factor resultante por el valor de la estimación de cada municipio del estado al que pertenece.

8.2 Variables auxiliares de los modelos

Después de diversas pruebas al conjunto de variables auxiliares, en que el método principal fue el denominado paso a paso (Forward stepwise regression), las seleccionadas por su poder predictivo, se describen en los siguientes puntos.

8.2.1 Para Obesidad

1. Relación de hombres entre mujeres para la población de 20 años y más (RelHomMuj20ymas).
2. Porcentaje de viviendas particulares habitadas que disponen de teléfono celular (VivCelular).
3. Porcentaje de defunciones por diabetes respecto a la población de 20 años y más (DefunDiabetes).

8.2.2 Para Hipertensión

1. Porcentaje de población de 65 años y más respecto a la población de 20 años y más (Pob65ymas).
2. Porcentaje de población de 18 a 24 años que asiste a la escuela (Pob18a24AsisEsc).
3. Porcentaje de viviendas con algún nivel de hacinamiento (VivHacinamiento).
4. Porcentaje de viviendas particulares habitadas que disponen de Internet (VivInternet).

8.2.3 Para Diabetes

1. Porcentaje de población de 0 a 24 años (Pob0a24).
2. Grado promedio de escolaridad (GraPromEsc).
3. Porcentaje de defunciones por diabetes respecto a la población de 20 años y más (DefunDiabetes).

8.3 Obtención de los archivos base para la construcción de los modelos

8.3.1 Municipios con muestra

Para la ENSANUT 2018 se registraron municipios en muestra y su varianza respectiva: 506 para Obesidad, 640 para Hipertensión y 566 para Diabetes, de los 2457 del marco geo-estadístico en que se basó la encuesta. Los cálculos se realizaron para cada una de estas áreas pequeñas.

8.3.2 Predicciones

Con base en esos modelos se realizaron las predicciones para la prevalencia de Obesidad, Hipertensión y Diabetes; posteriormente fueron ajustadas a los montos registrados por la ENSANUT a nivel de entidad federativa, mediante el algoritmo de Ajuste Proporcional Iterativo para Tablas de dos Dimensiones (Hunsinger, 2008).

9 Procesamiento de la información

9.1 Estimación de los coeficientes β del modelo

En la Tabla 2 se observa que los valores de los coeficientes son estadísticamente diferentes de cero de acuerdo al *p*-valor obtenido.

Tabla 2. Valores de los coeficientes β del modelo.

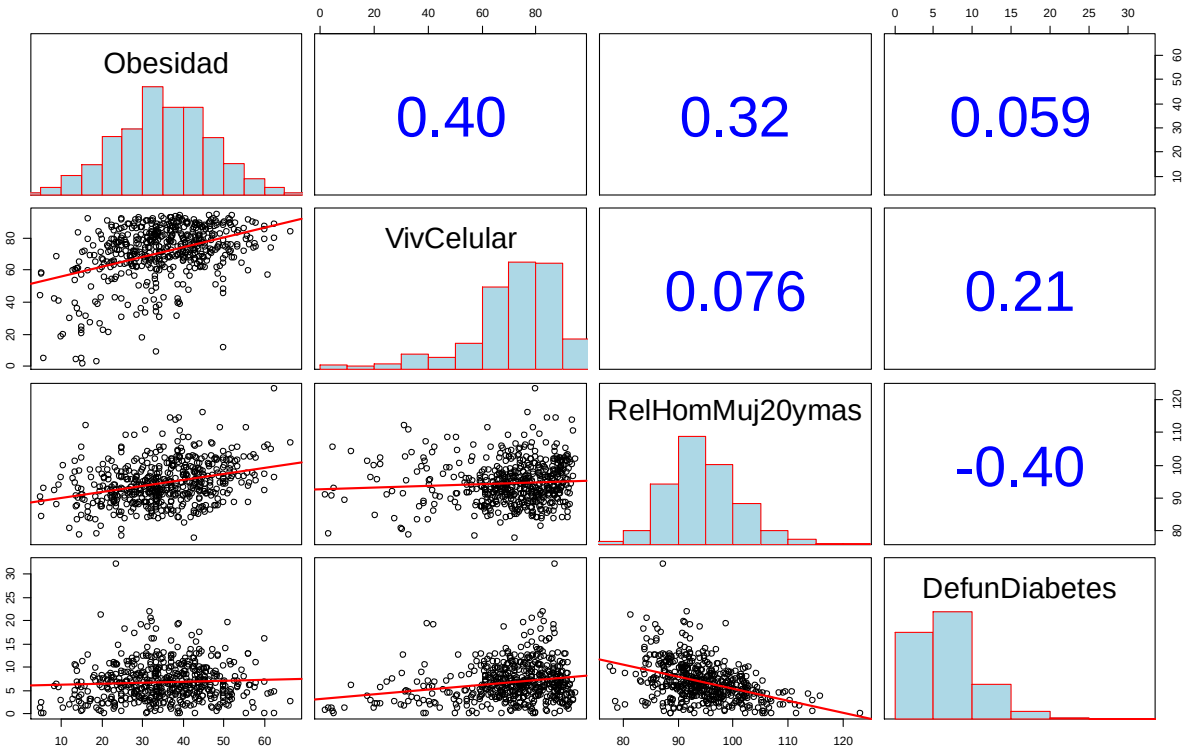
Enfermedad	Concepto	Valores			
		Beta(β)	Error Est.	Valor t	Valor p
Obesidad	(Intercepto)	-38.498	6.949	-5.540	3.018e-08
	VivCelular	0.209	0.025	8.302	1.023e-16
	RelHomMuj20ymas	0.605	0.071	8.544	1.294e-17
	DefunDiabetes	0.234	0.116	2.020	4.336e-02
Hipertensión	(Intercepto)	17.982	2.597	6.926	4.341e-12
	VivHacinamiento	-0.180	0.039	-4.556	5.214e-06
	Pob18a24AsisEsc	0.102	0.039	2.623	8.717e-03
	Pob65ymas	0.354	0.086	4.137	3.515e-05
	VivInternet	-0.071	0.033	-2.152	3.138e-02
Diabetes	(Intercepto)	23.984	2.606	9.203	3.492e-20
	Pob0a24	-0.228	0.041	-5.541	3.004e-08
	DefunDiabetes	0.260	0.041	6.265	3.729e-10
	GraPromEsc	-0.606	0.117	-5.164	2.417e-07

Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

9.2 Coeficiente de correlación, histograma de frecuencias y gráficas de dispersión

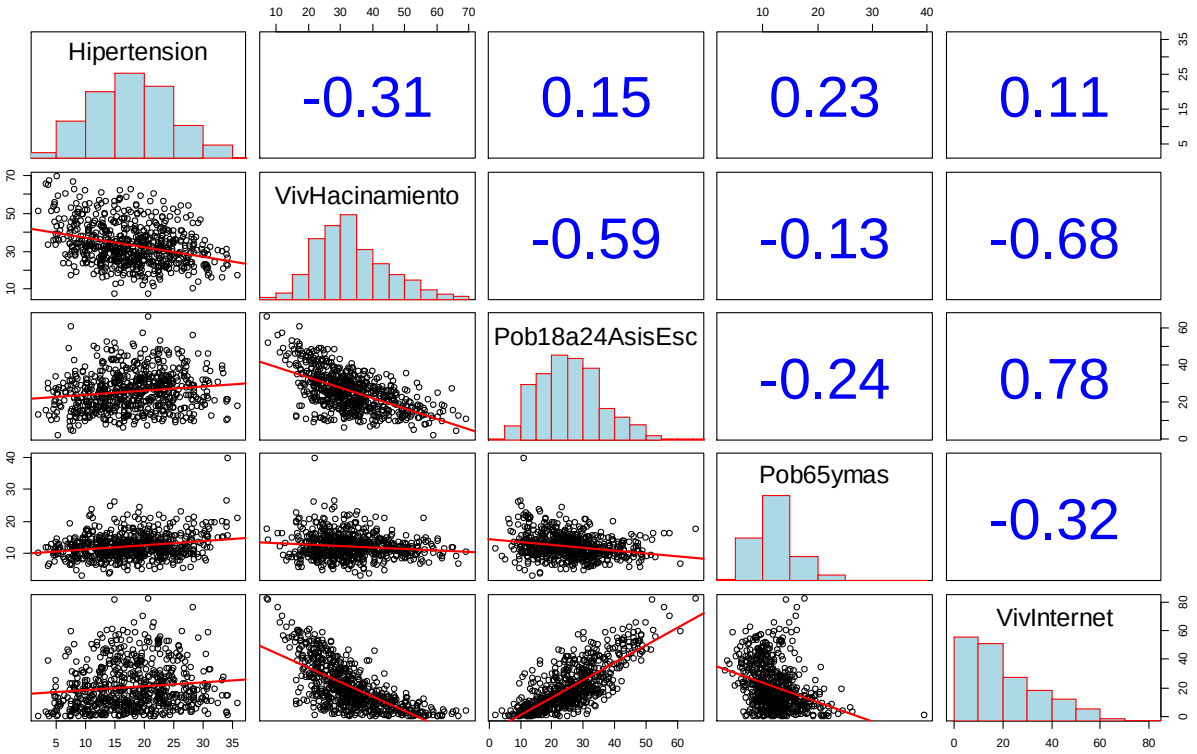
Las gráficas de panel 3, 4 y 5 son de utilidad para diagnosticar si existe una relación lineal entre las variables dependientes y las auxiliares. En efecto, se observa la existencia de correlación entre estas variables; no obstante, debe tomarse con reserva ya que la correlación proviene de unidades de observación (en este caso municipios) en donde algunas de ellas pueden tener muy poca muestra. De manera que las ilustraciones son sólo con fines de diagnóstico.

Gráfica 3. Obesidad y las variables auxiliares.



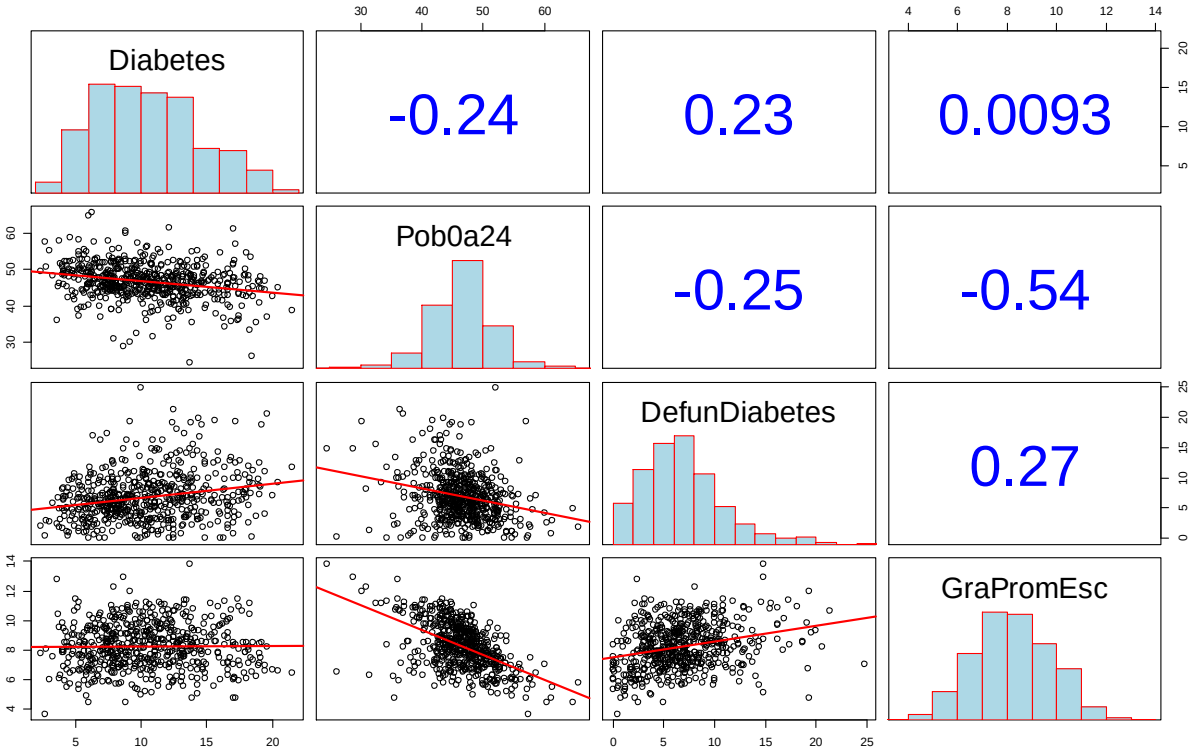
Fuente: Elaboración propia.

Gráfica 4. Hipertensión y las variables auxiliares.



Fuente: Elaboración propia.

Gráfica 5. Diabetes y las variables auxiliares.



Fuente: Elaboración propia.

9.3 Planteamiento de hipótesis

Sin pérdida de generalidad, para el caso de los residuales (R) o bien para los efectos aleatorios, el planteamiento de una prueba de hipótesis se formula bajo el siguiente esquema:

Hipótesis Nula H_0: *Los residuales (o efectos aleatorios) se distribuyen conforme a una distribución Normal*

vs.

Hipótesis Alterna H_a: *Los residuales (o efectos aleatorios) no se distribuyen conforme a una distribución Normal*

$$H_0 : R \sim \mathcal{N}$$

vs.

$$H_a : R \neq \mathcal{N}$$

El nivel de significancia, esto es, la probabilidad de rechazar la hipótesis nula aún siendo verdadera, se fija en 5.0 %; y para la prueba se utilizan los procedimientos de Shapiro-Wilk (W),

Kolmogórov-Smirnov (K-S) y Jarque-Bera (J-B). Ahora bien, dado que se trata de una prueba de bondad de ajuste, lo que interesa es no rechazar la hipótesis nula; de tal manera que si la probabilidad observada para la estadística calculada a partir de la muestra objeto de análisis, es mayor que el nivel de significancia, o riesgo prefijado, entonces no se rechaza la hipótesis nula; de lo contrario se rechaza.

Esto es, si la probabilidad asociada a las estadísticas de prueba $(\hat{\alpha})$ $(o\ p\text{-}valor) > \alpha$, entonces no se rechaza la hipótesis de normalidad. En suma, si se cumple la siguiente relación, entonces no se rechaza H_0:

$$p > \boldsymbol{\alpha}.$$

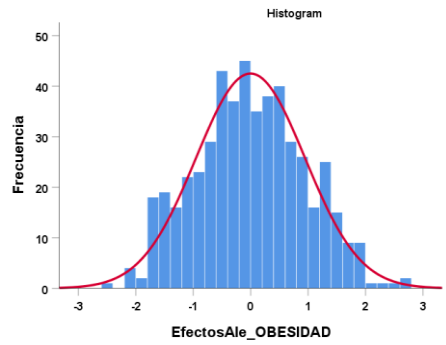
9.4 Verificación de supuestos

- Multicolinealidad: Se controla la correlación entre las variables auxiliares.
- Homocedasticidad: Se verifica que los residuales muestren igualdad de varianzas.
- Normalidad de efectos aleatorios: Se verifica que la distribución de los efectos aleatorios sea normal.
- Normalidad de los residuales: Se verifica que la distribución de los residuales sea normal.

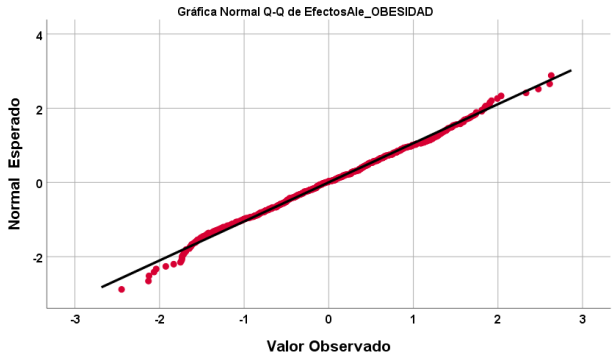
En efecto, de manera visual para el caso de la Obesidad, se tienen histogramas y gráficas cuantil-cuantil, donde según la Gráfica 6, se observa que los efectos aleatorios se distribuyen conforme a una normal; lo mismo sucede con los residuales. Para las otras dos variables objeto de estimación, se observó una situación muy similar.

Gráfica 6. Efectos aleatorios y residuales de Obesidad.

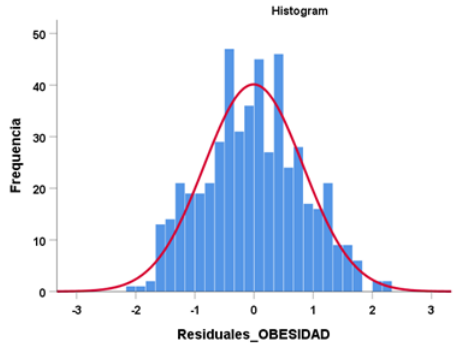
A) Obesidad. Distribución de los efectos aleatorios



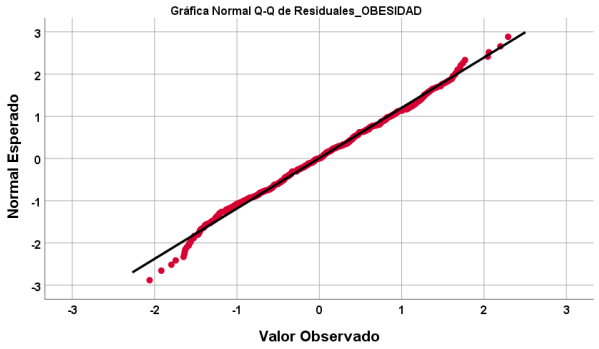
B) Obesidad. Gráfica Q-Q de los efectos aleatorios



C) Obesidad. Distribución de los residuales



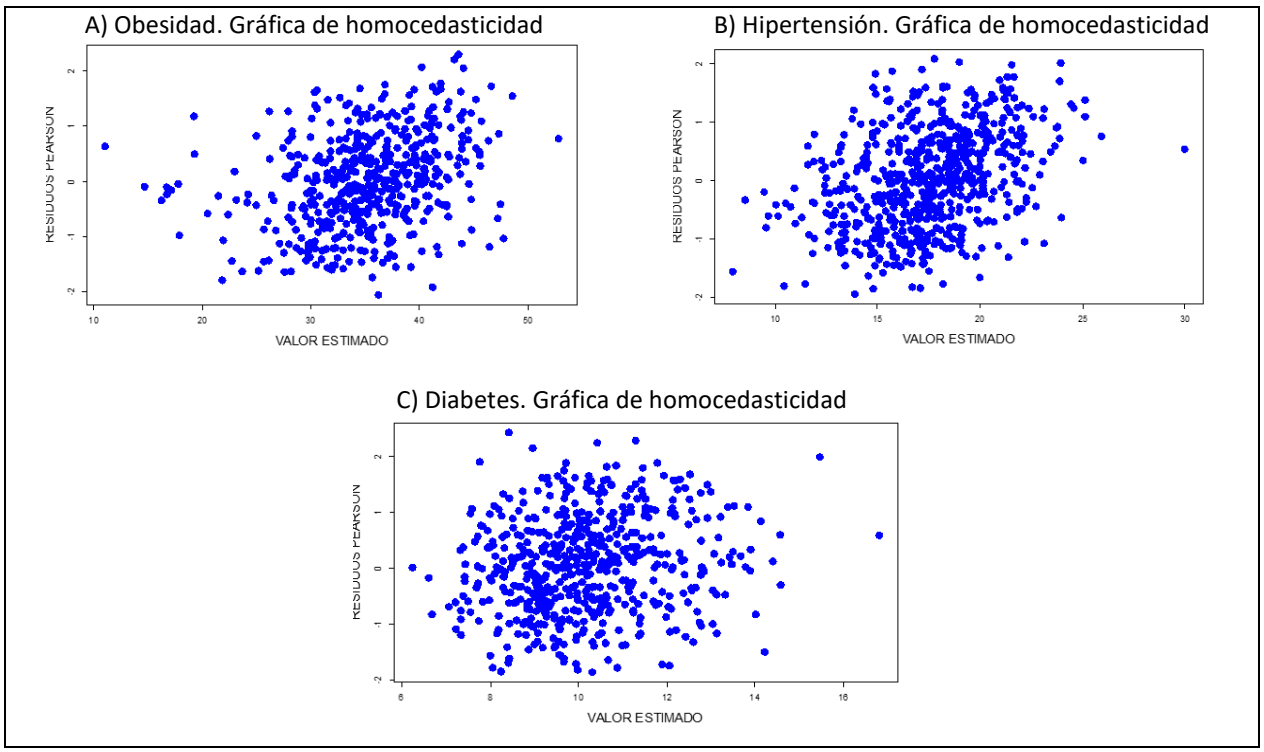
D) Obesidad. Gráfica Q–Q de los residuales



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

En cuanto la varianza constante de los residuales, para las tres variables consideradas, en la Gráfica 7 se muestra un comportamiento aproximadamente constante en el diagrama de dispersión, a lo largo del eje horizontal.

Gráfica 7. Residuales de Obesidad, Hipertensión y Diabetes.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

9.5 Pruebas estadísticas de normalidad, homocedasticidad y multicolinealidad

Para corroborar lo observado en las gráficas anteriores, es necesario realizar las pruebas numéricas de los supuestos. En la Tabla 3 se muestran los resultados obtenidos (W es la prueba de Shapiro Wilk, K-S es la prueba de Kolmogorov y J-B es la prueba de Jarque-Bera):

Tabla 3. Pruebas numéricas de los supuestos.

Variable de estimación	Número de casos (municipios)	I de Moran [-1,1]	K vecinos	Correlación espacial Rho [-1, 1]	p-valor (%)							Multicolinealidad Índice de información Kappa
					Efectos aleatorios			Residuales			Homocedasticidad	
					W	K-S	J-B	W	K-S	J-B	Breusch-Pagan	
Obesidad	506	0.168	4	-0.251	5.83	72.37	8.15	3.14	47.86	5.5	81.93	34.79
Hipertensión	640	0.101	4	0.503	0.12	0.33	1.25	0.05	2.18	0.8	98.57	20.4
Diabetes	566	0.062	6	0.554	0.42	17.77	2.6	0.58	5.04	2.7	24.08	27.78

Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

Como se observa, todos los *p*-valores obtenidos en las pruebas estadísticas son mayores que el nivel de significancia prefijado del 1.0 %, por lo que la hipótesis nula respectiva no se rechaza, cumpliéndose los supuestos de normalidad de los efectos aleatorios y de los residuales, así como el supuesto de igualdad de varianzas.

Para el caso del Índice de Moran, la hipótesis nula establece que no hay correlación, misma que según el *p*-valor, se rechaza para las tres variables de acuerdo a los siguientes resultados:

p-valor Obesidad = 0.0000018047

p-valor Hipertensión = 0.0003

p-valor Diabetes = 0.0167

9.6 Ajustes

9.6.1 Ajuste proporcional iterativo (IPF)

El IPF, que significa Ajuste Proporcional Iterativo y que en ocasiones es denominado “rastrillar”, es un procedimiento para ajustar una tabla de celda de datos de manera que se sumen a los totales seleccionados tanto para las columnas como para las filas (en el caso bidimensional) de la tabla. Las celdas de datos no ajustadas pueden denominarse celdas “semilla” y los totales seleccionados pueden denominarse totales “marginales”.

Procedimiento

Paso 1: cada fila de celdas semilla se ajusta proporcionalmente para igualar los totales de la fila marginal (específicamente, cada celda se divide por la suma real de la fila de celdas, luego se multiplica por el total de la fila marginal).

Paso 2: cada columna de celdas (ya ajustadas por filas) se ajusta proporcionalmente para igualar los totales de las columnas marginales. Este es el final de la primera 'iteración'.

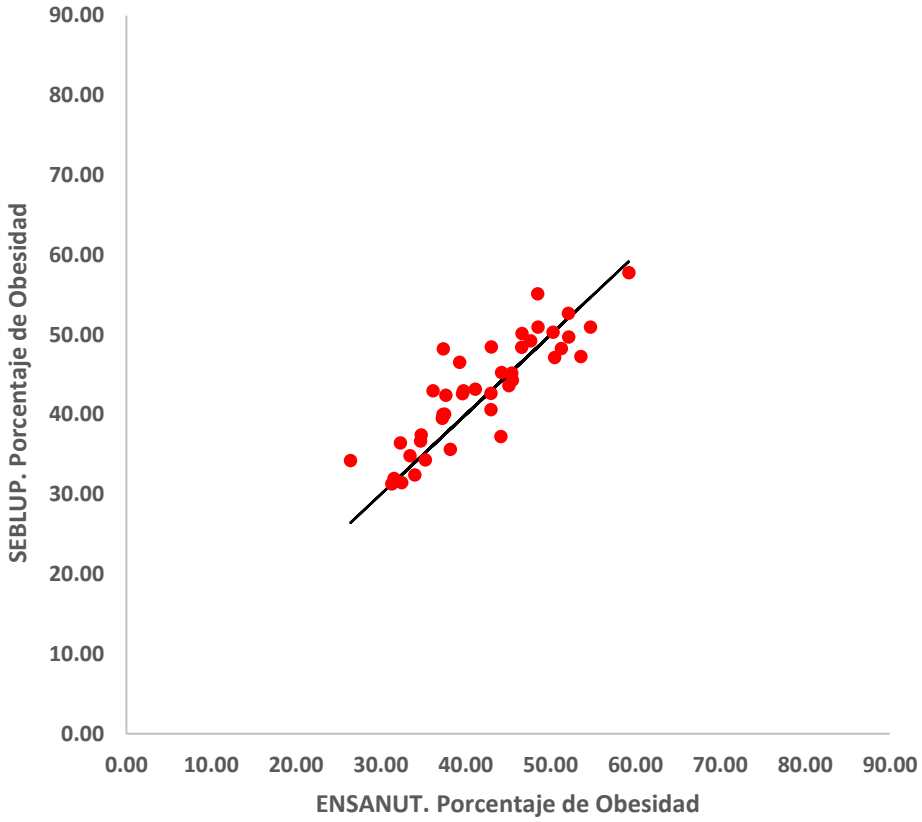
Pasos posteriores: los pasos anteriores se repiten hasta que se alcanza el nivel de convergencia seleccionada. Hunsinger, E. (2008) Ajuste Proporcional Iterativo para Tablas de dos Dimensiones.

9.7 Diagnóstico

En las gráficas 8, 9 y 10 se muestra, por medio de diagramas de dispersión, qué tanto se ajusta el modelo de áreas pequeñas a las cifras obtenidas de la ENSANUT para aquellos municipios cuyo coeficiente de variación (CV) es menor a 20 %. Es decir, se toman los municipios cuyo tamaño de muestra es suficiente para dar estimaciones aceptables, según el CV registrado.

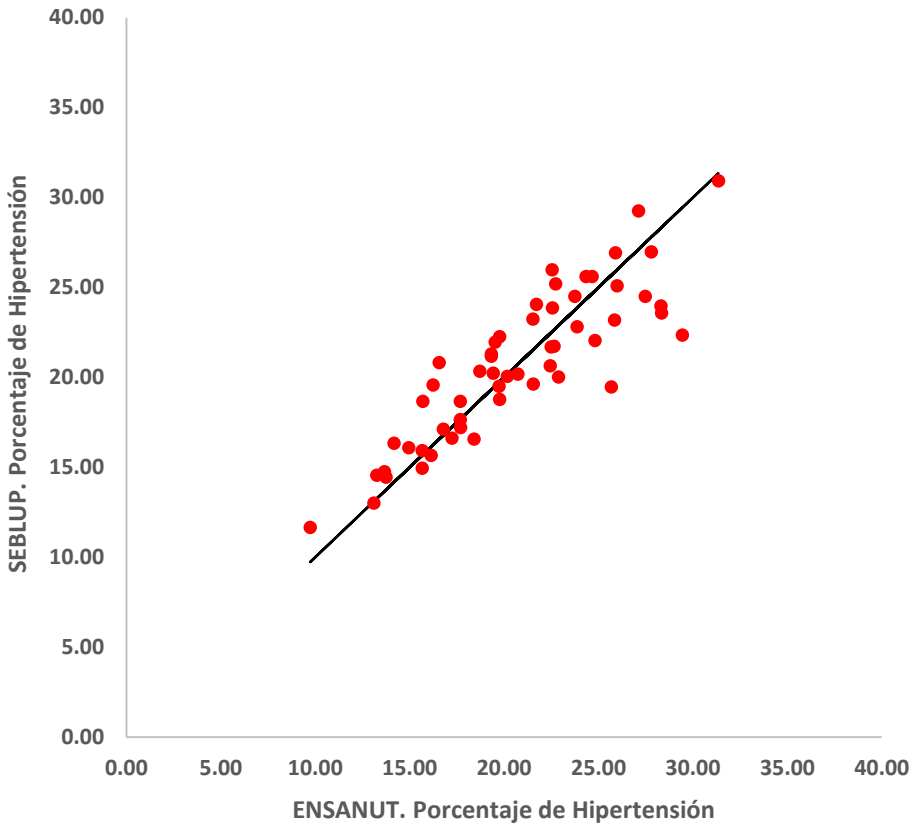
En todas estas gráficas se observa que las estimaciones obtenidas por SEBLUP se acercan a las obtenidas por la encuesta mencionada, aunque es mayor la diferencia para el caso de la enfermedad de Diabetes.

Gráfica 8. Obesidad. Comparativo entre estimaciones SEBLUP y ENSANUT para los casos en que CV (ENSANUT) < 20 %.



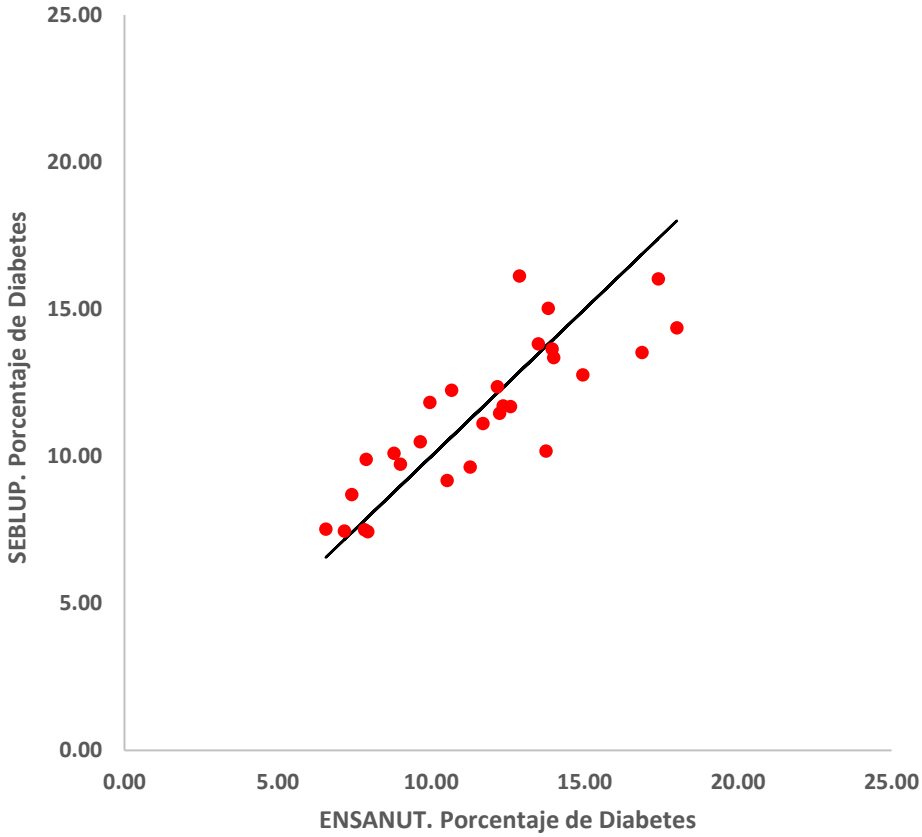
Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

Gráfica 9. Hipertensión. Comparativo entre estimaciones SEBLUP y ENSANUT para los casos en que CV (ENSANUT) < 20 %.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

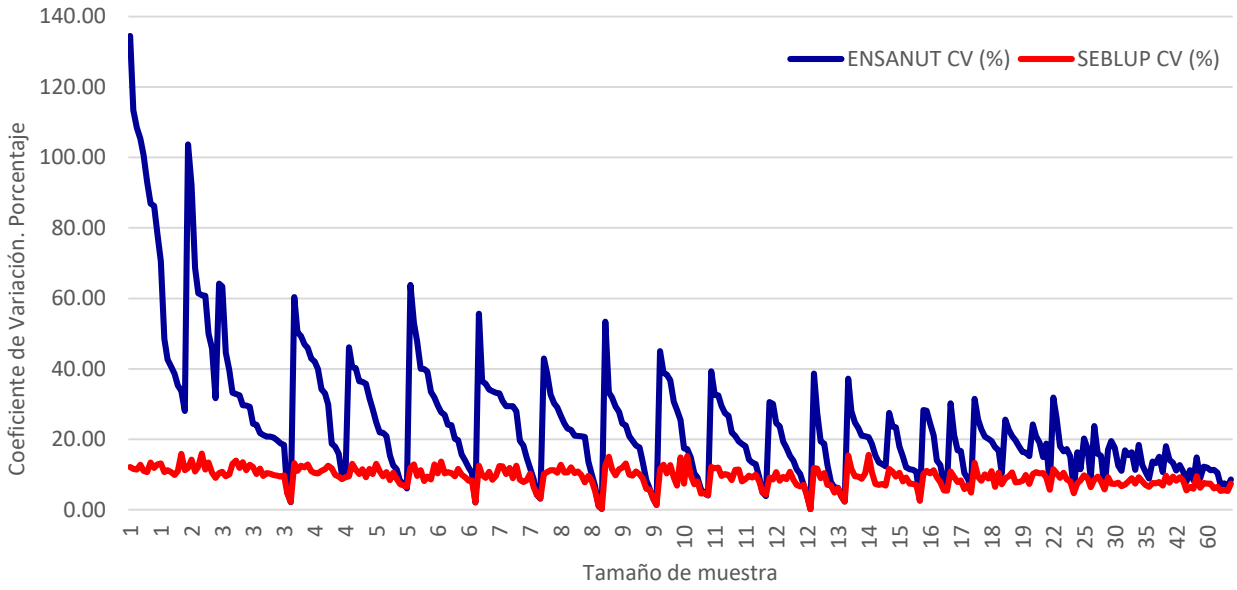
Gráfica 10. Diabetes. Comparativo entre estimaciones SEBLUP y ENSANUT para los casos en que CV (ENSANUT) < 20 %.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

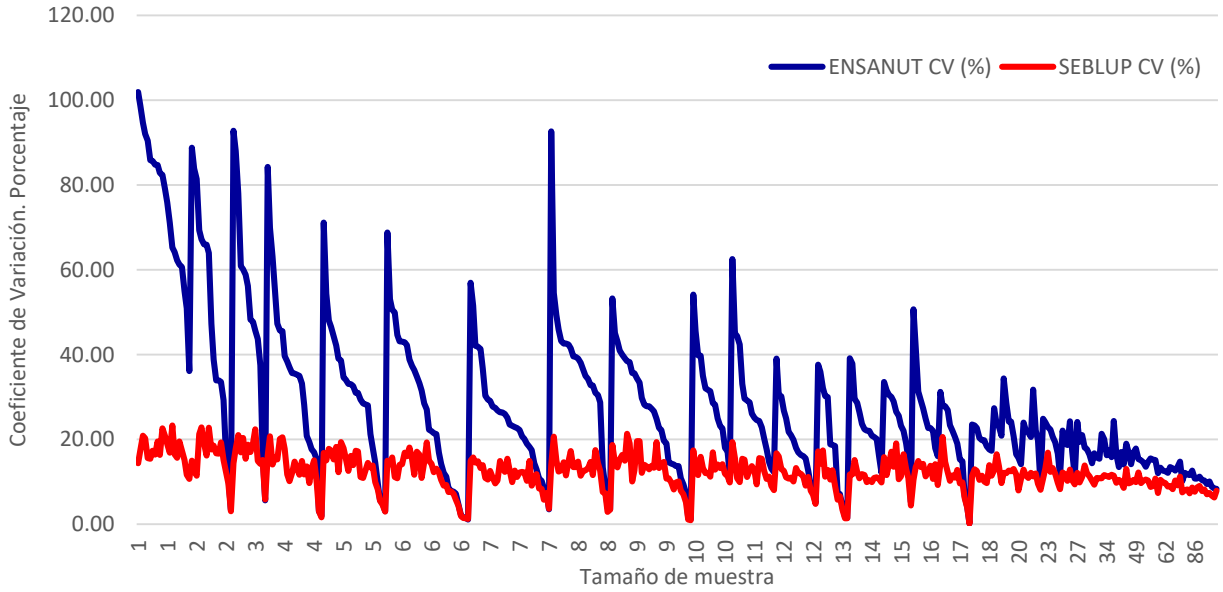
En cuanto a la precisión de los valores estimados por el modelo, las gráficas 11, 12 y 13 muestran los coeficientes de variación tanto del modelo de áreas pequeñas, como de los obtenidos por la ENSANUT. Las gráficas están ordenadas en el eje de las X por el tamaño de la muestra de los municipios; se aprecia claramente que los coeficientes de variación disminuyen a medida que se incrementa el tamaño de muestra, y los obtenidos por SEBLUP son notablemente más pequeños que los de la encuesta. Es importante recalcar que la ENSANUT no fue diseñada para obtener cifras a nivel municipal, por tal razón los CV son altos y dispersos.

Gráfica 11. Obesidad. Coeficientes de variación de las estimaciones ENSANUT y SEBLUP para los municipios con muestra, ordenados según tamaño.



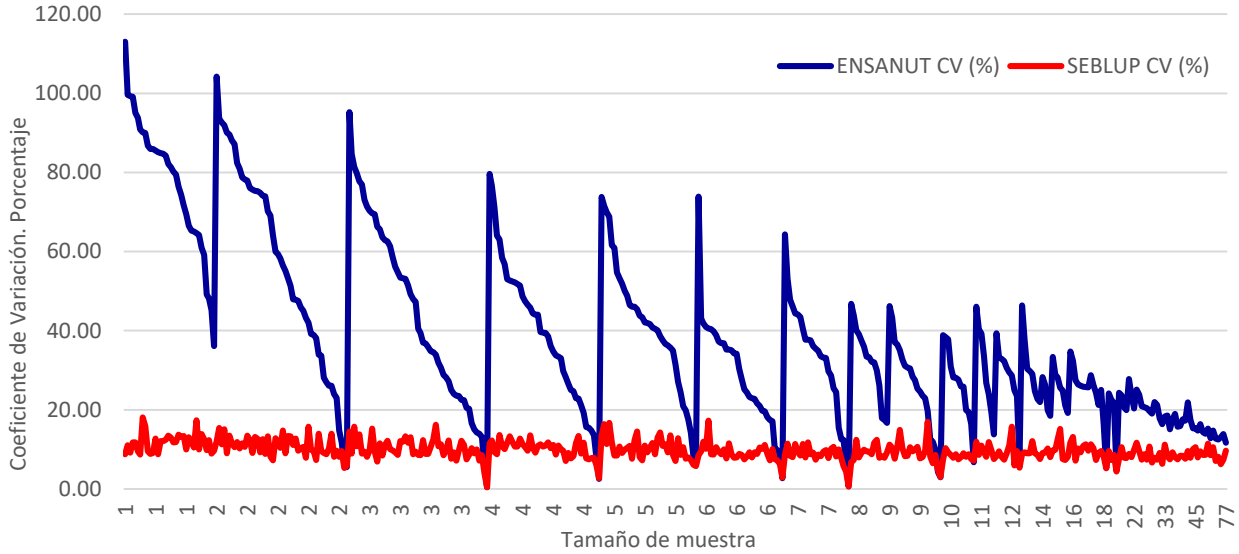
Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

Gráfica 12. Hipertensión. Coeficientes de variación de las estimaciones ENSANUT y SEBLUP para los municipios con muestra, ordenados según tamaño.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

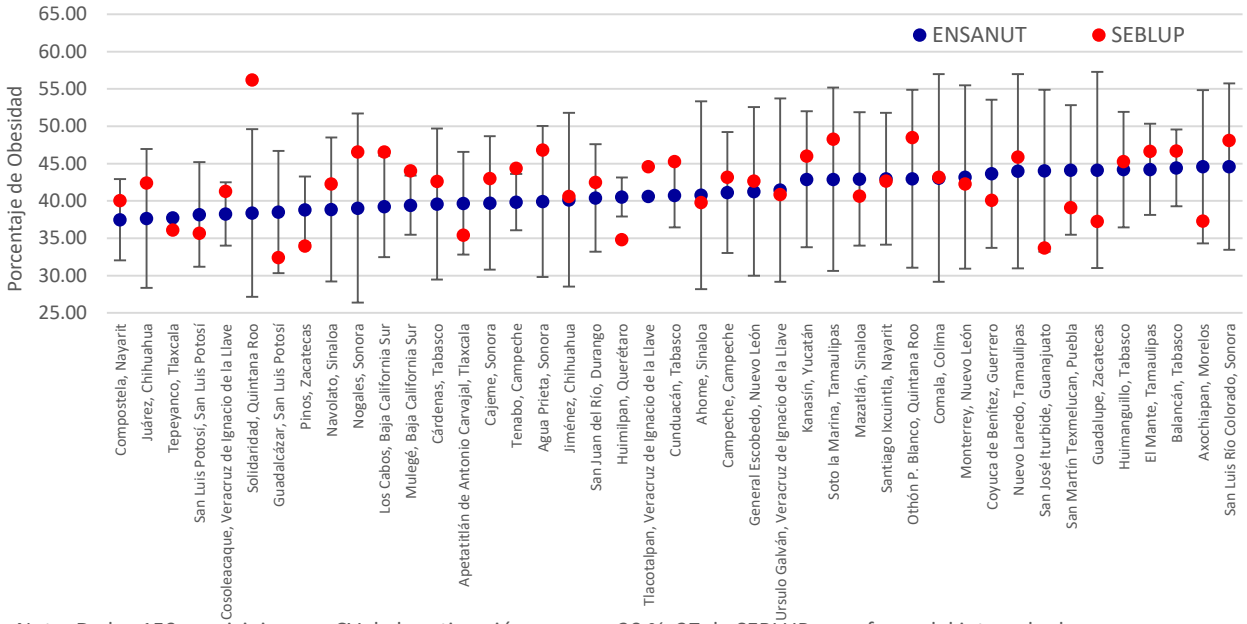
Gráfica 13. Diabetes. Coeficientes de variación de las estimaciones ENSANUT y SEBLUP para los municipios con muestra, ordenados según tamaño.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

Conforme a la Gráfica 14, para la Obesidad, de los 159 municipios de la encuesta con CV menor a 20 %, 37 de SEBLUP caen fuera del intervalo de confianza de las estimaciones. Por motivo de espacio sólo se muestra la parte central de las estimaciones.

Gráfica 14. Obesidad. Intervalos de Confianza de las estimaciones ENSANUT y SEBLUP para los municipios con CV < 20 %.

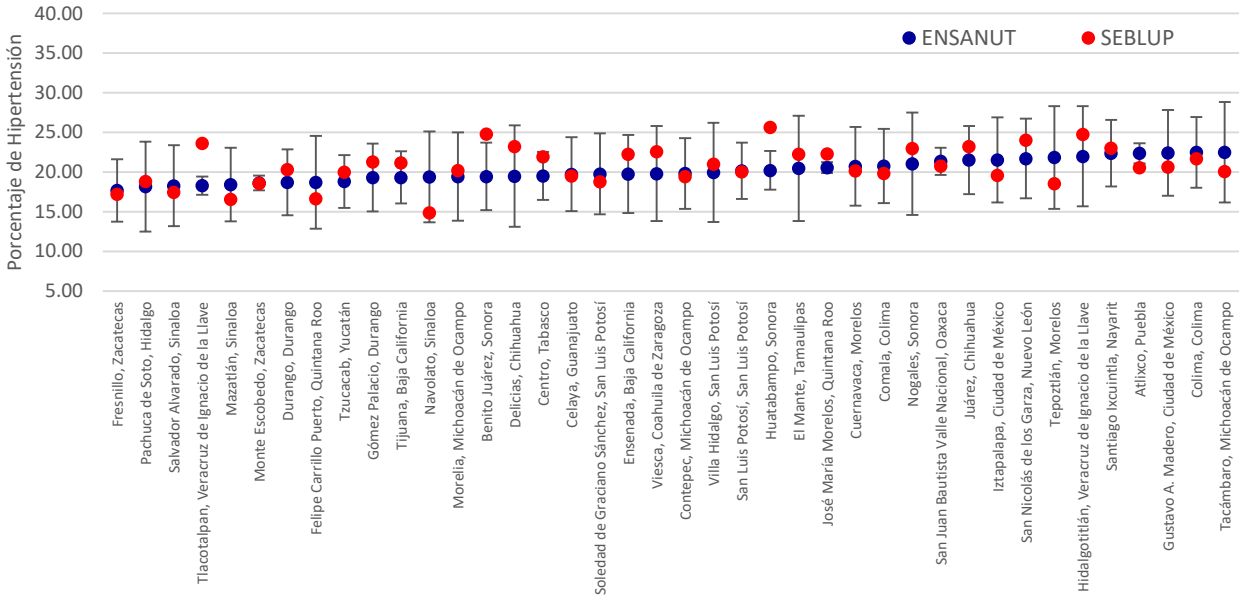


Nota: De los 159 municipios con CV de la estimación menor a 20 %, 37 de SEBLUP caen fuera del intervalo de confianza de las estimaciones. Por motivo de espacio solo se muestra la parte central de las estimaciones.

Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

Según la Gráfica 15, para la Hipertensión, de los 166 municipios de la encuesta con CV menor a 20 %, 36 de SEBLUP caen fuera del intervalo de confianza de las estimaciones. Por motivo de espacio sólo se muestra la parte central de las estimaciones.

Gráfica 15. Hipertensión. Intervalos de Confianza de las estimaciones ENSANUT y SEBLUP para los municipios con CV < 20 %.

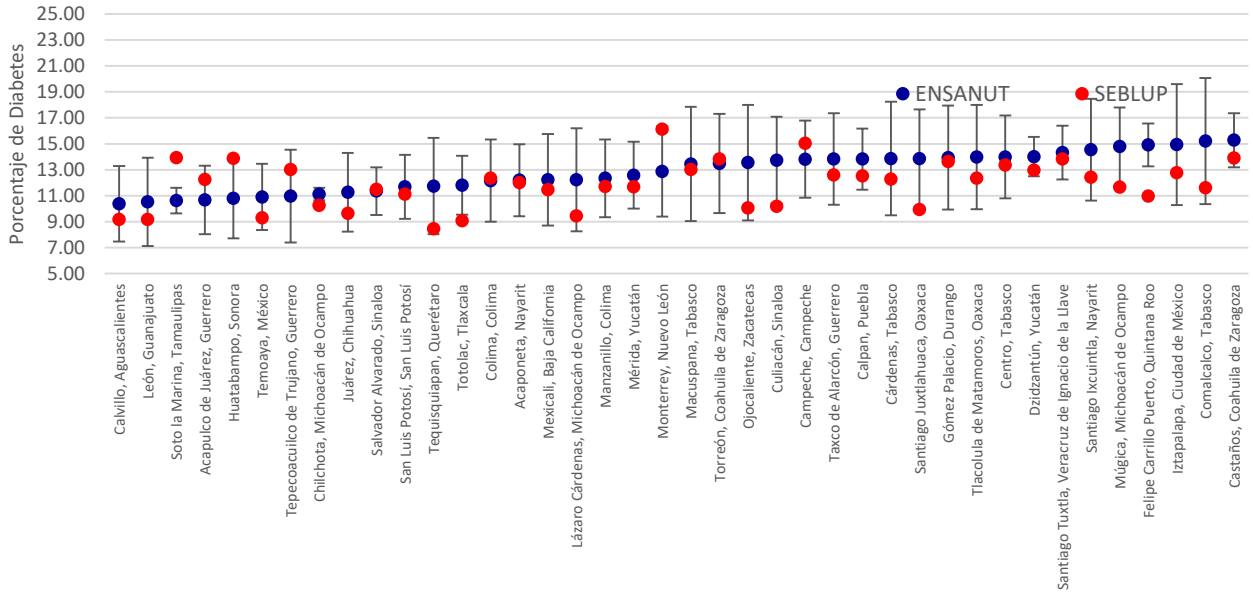


Nota: De los 166 municipios con CV de la estimación menor a 20 %, 36 de SEBLUP caen fuera del intervalo de confianza de las estimaciones. Por motivo de espacio solo se muestra la parte central de las estimaciones.

Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

Al observar la gráfica 16, Para la Diabetes, de los 88 municipios de la encuesta con CV menor a 20 %, 32 de SEBLUP caen fuera del intervalo de confianza de las estimaciones. Por motivo de espacio sólo se muestra la parte central de las estimaciones.

Gráfica 16. Diabetes. Intervalos de Confianza de las estimaciones ENSANUT y SEBLUP para los municipios con CV < 20 %.



Nota: De los 88 municipios con CV de la estimación menor a 20 %, 32 de SEBLUP caen fuera del intervalo de confianza de las estimaciones. Por motivo de espacio solo se muestra la parte central de las estimaciones.

Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

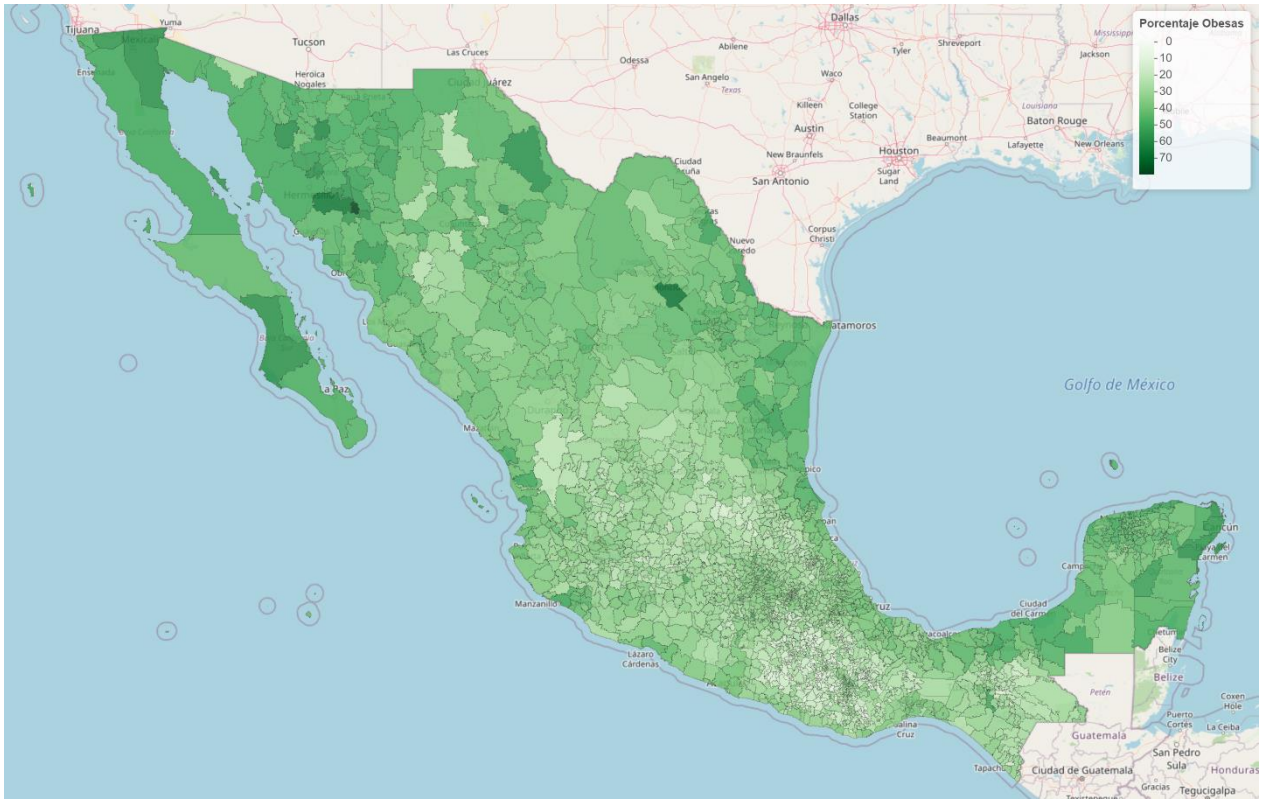
10 Resultados

A continuación se presentan los productos obtenidos mediante el proceso anteriormante descrito, en formato de mapas, gráficas y tabulados a nivel municipal.

En las ilustraciones 6, 7 y 8 se observa el mapa de la Republica Mexicana, por municipio, con las estimaciones de la prevalencia para cada una de las enfermedades.

Según se observa en la Ilustración 6, se tiene un grado importante de Obesidad en todo el país. No obstante, por un lado, sobresalen la zona norte y toda la zona costera de la República Mexicana, incluyendo la península de Yucatán, y por el otro, destacan los niveles bajos de Obesidad en zonas como la Sierra del Nayar, la Sierra Tarahumara (tanto en su parte alta y baja), así como algunos municipios pertenecientes a los estados de Oaxaca, Chiapas, Hidalgo y San Luis Potosí.

Ilustración 6. Prevalencia de Obesidad para los municipios de México, 2018.

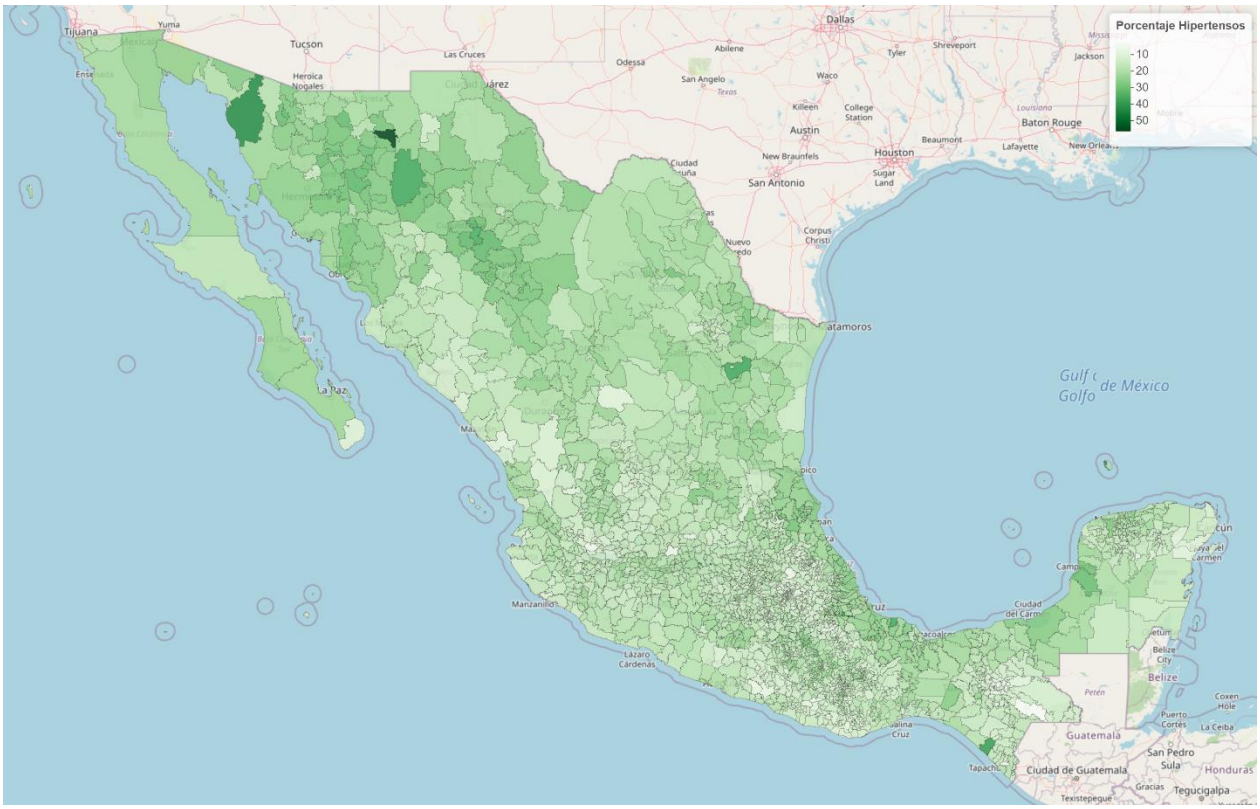


Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

En la Ilustración 7 se observa que la Hipertensión se manifiesta con valores altos en zonas del noroeste del país, particularmente en los estados de Baja California, Sonora y Chihuahua. Adicionalmente, la enfermedad se extiende en la parte media baja de la República; entidades

como Jalisco, Estado de México, Ciudad de México, Morelos y Veracruz tiene valores altos e intermedios de esta enfermedad. Por otra parte, resalta geográficamente un cinturón de valores bajos en la parte media alta de la República, desde el estado de Sinaloa en el Pacífico, así como en buena parte de la zona central hasta el territorio de Tamaulipas en el Golfo de México.

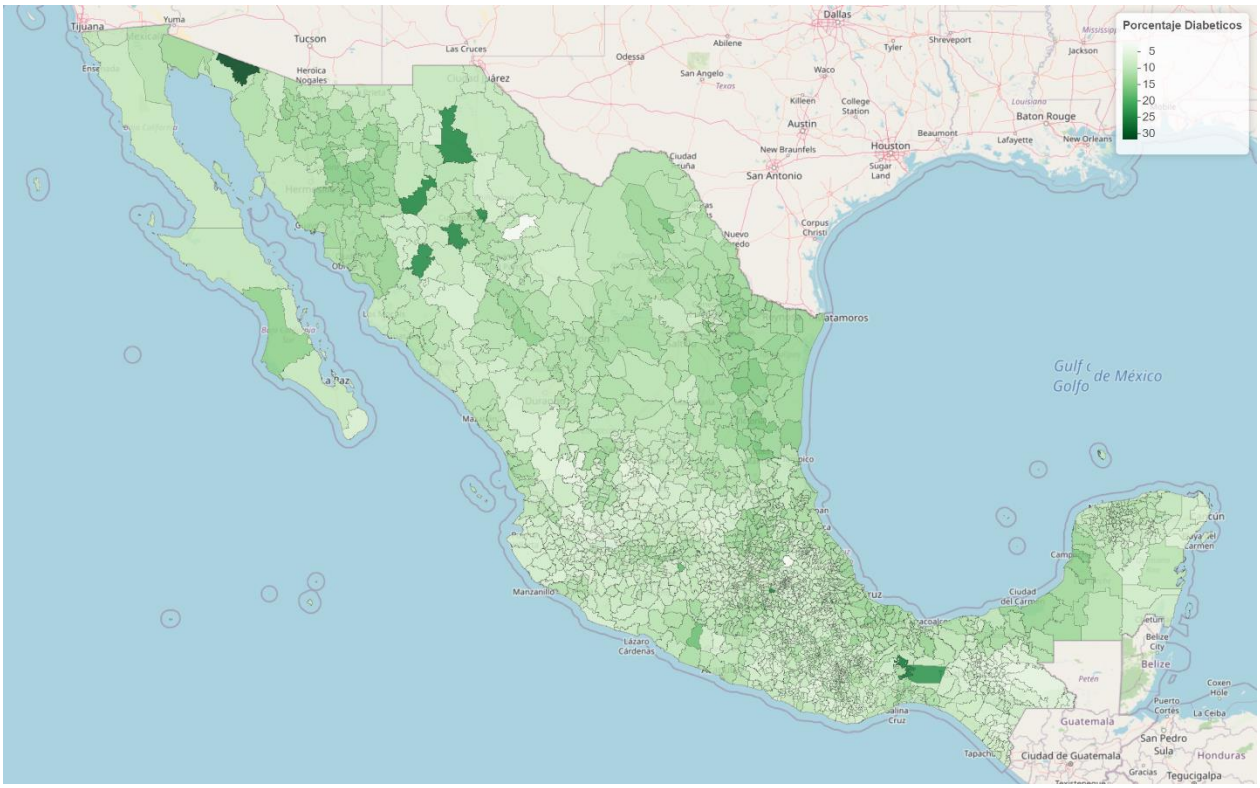
Ilustración 7. Prevalencia de Hipertensión para los municipios de México, 2018.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

En la Ilustración 8 se muestra que la Diabetes tiene un comportamiento más uniforme, sin embargo, se observan menores proporciones de la enfermedad en Puebla, Chihuahua, Chiapas y Oaxaca. En el otro extremo, las proporciones de Diabetes más altas se dan en los municipios del Noreste de los estados de Nuevo León y Tamaulipas, así como en municipios de Sonora.

Ilustración 8. Prevalencia de Diabetes para los Municipios de México, 2018.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

Se generó, como complemento a los mapas, un formato interactivo de Excel con la información de Obesidad, Hipertensión y Diabetes, a nivel municipal . De esta manera se puede accesar con facilidad, tanto de manera global como específica, a las proporciones estimadas y a sus precisiones estadísticas. A manera de ejemplo, en la Ilustración 9 se muestran los datos para los municipios de los estados de Aguascalientes y Zacatecas.

Ilustración 9. Vista del formato interactivo con la información de Obesidad, Hipertensión y Diabetes.

Identificador único del municipio	Clave de entidad federativa	Entidad federativa	Clave de municipio o delegación	Municipio o delegación	Estimador	Porcentaje de población de 20 años y más con obesidad.	Porcentaje de población de 20 años y más con diagnóstico previo de hipertensión.	Porcentaje de población de 20 años y más con diagnóstico previo de diabetes.
01001	01	Aguascalientes	001	Aguascalientes	Valor	31.5	14.9	7.5
01001	01	Aguascalientes	001	Aguascalientes	Error estándar	2.3	1.2	0.7
01001	01	Aguascalientes	001	Aguascalientes	Límite inferior de confianza	27.6	13.0	6.3
01001	01	Aguascalientes	001	Aguascalientes	Límite superior de confianza	35.3	16.9	8.7
01001	01	Aguascalientes	001	Aguascalientes	Coeficiente de variación	7.4	7.9	9.7

⋮

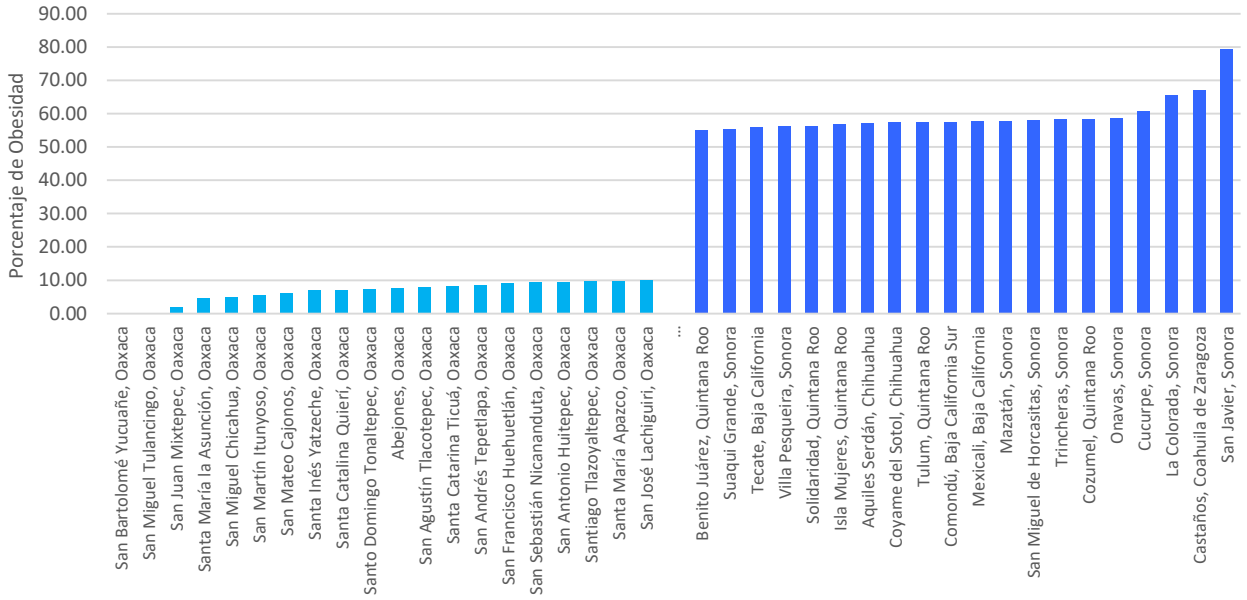
Identificador único del municipio	Clave de entidad federativa	Entidad federativa	Clave de municipio o delegación	Municipio o delegación	Estimador	Porcentaje de población de 20 años y más con obesidad.	Porcentaje de población de 20 años y más con diagnóstico previo de hipertensión.	Porcentaje de población de 20 años y más con diagnóstico previo de diabetes.
32056	32	Zacatecas	056	Zacatecas	Valor	36.8	20.0	11.8
32056	32	Zacatecas	056	Zacatecas	Error estándar	4.1	2.3	1.1
32056	32	Zacatecas	056	Zacatecas	Límite inferior de confianza	30.0	16.3	10.0
32056	32	Zacatecas	056	Zacatecas	Límite superior de confianza	43.6	23.7	13.7
32056	32	Zacatecas	056	Zacatecas	Coeficiente de variación	11.2	11.3	9.6

Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

10.1 Proporciones de Obesidad, Hipertensión y Diabetes

En la Gráfica 17 se muestra que los menores porcentajes de Obesidad, en su mayoría corresponden a municipios de Oaxaca, con proporciones menores al 10 %. En el otro extremo, las mayores proporciones se acercan al 80 %, concentrándose principalmente en municipios de Sonora.

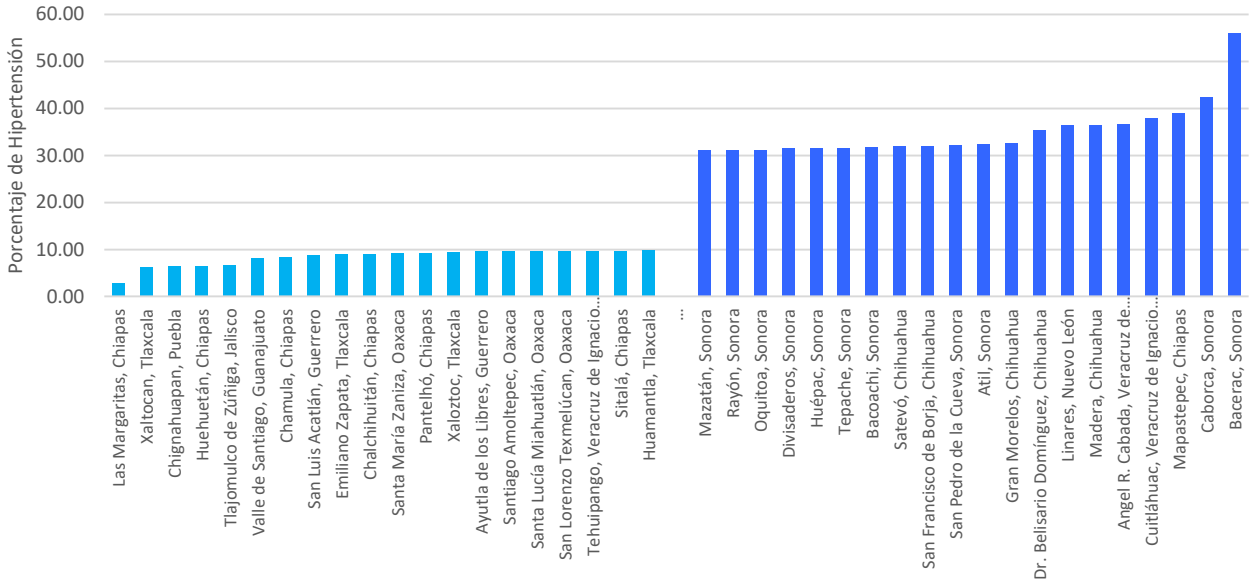
Gráfica 17. Obesidad. Estimaciones SEBLUP para los municipios de México. Ordenados de menor a mayor. Los 20 valores más bajos y los 20 más altos.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

En la Gráfica 18 se muestra que los menores porcentajes de Hipertensión se presentan en los municipios del sur del país, con proporciones menores al 10 por ciento. En el otro extremo, las mayores proporciones se acercan al 40 %, salvo un municipio de Sonora que rebasa el 50 %, y se concentran en municipios del norte del país.

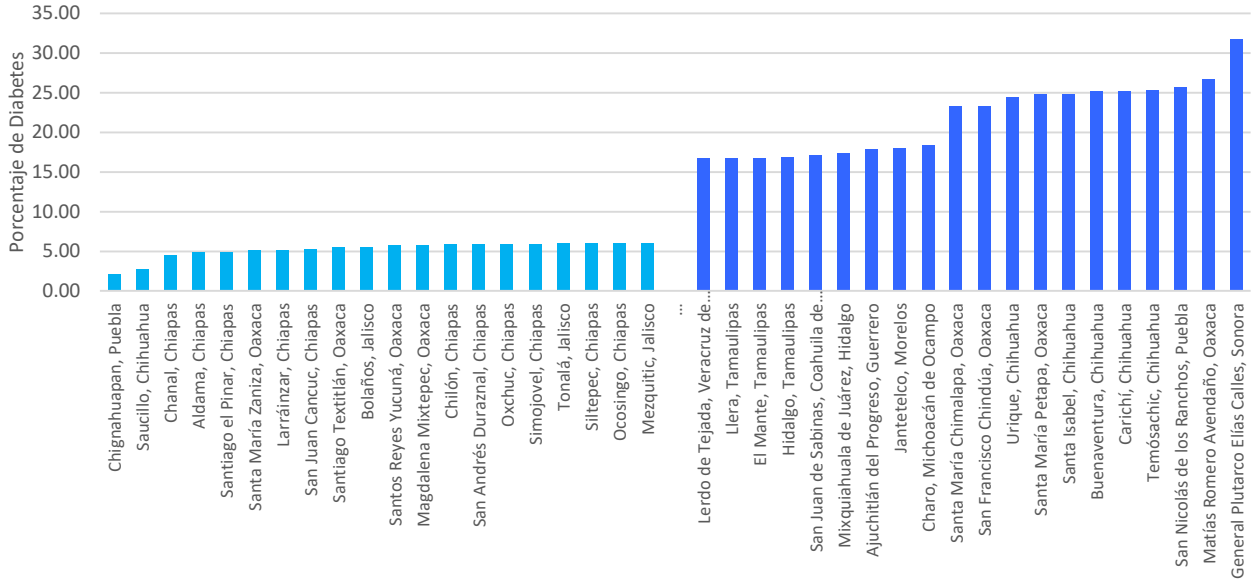
Gráfica 18. Hipertensión. Estimaciones SEBLUP para los municipios de México. Ordenados de menor a mayor. Los 20 valores más bajos y los 20 más altos.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

En la Gráfica 19 se observa que los municipios con menores porcentajes de Diabetes se localizan en Puebla, Chihuahua, Chiapas y Oaxaca; con proporciones menores al 10 por ciento. Al contrario, las mayores proporciones de esta enfermedad son menores al 35 %, y se concentran también en municipios de Oaxaca, Puebla y Chihuahua. En estos estados se encuentran, tanto los municipios con los más bajos porcentajes de Diabetes, como los municipios con los más alto porcentajes de la misma enfermedad.

Gráfica 19. Diabetes. Estimaciones SEBLUP para los municipios de México. Ordenados de menor a mayor. Los 20 valores más bajos y los 20 más altos.



Fuente: Cálculos por técnicas de Estimación para Áreas Pequeñas.

11 Bibliografía

Anselin, L., (1988). Spatial Econometrics. Methods and Models. Kluwer, Boston.

Avila C. A. (2007). Medigraphic. México. Recuperado de https://www.medigraphic.com/pdfs/revinvcli/nn-2007/nn074c.pdf

Barber S. (2017). Sciencedirect. Brasil. Sciencedirect. Recuperado de https://www.sciencedirect.com/science/article/abs/pii/S0277953617303489

Bhushan, B., Girja, S., Shukla, K. & Kundu D. (2005). Spatio-Temporal Models in Small Area Estimation. Canada.

Benedetti, R. Piersimoni, F. (2015). Sampling Spatial Units for Agricultural Surveys USA, SPRINGER 2015.

Borrego, J. (2018) Modelos de Regresión para Datos Espaciales. Trabajo de Fin de Grado. Universidad de Sevilla. Facultad De Matemáticas. Departamento de Estadística e Investigación Operativa. Documento no publicado.

Buford W. (2016). NCBI. EEUU. ncbi.nlm.nih.gov. Recuperado de https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4768730/

Chandra, H., Salvati, N, & Chambers. (2009) R. Small Area Estimation for Spatially Correlated Populations. USA.

Consejo Nacional de Población (CONAPO) (2013). Proyecciones de la Población de México y de las Entidades. Consultado en https://datos.gob.mx/busca/dataset/proyecciones-de-la-poblacion-de-mexico-y-de-las-entidades-federativas-2016-2050/resource/8844b275-b2bb-45d8-9ee3-b101a6306f19

Cuevas G. (2017). NCBI. EEUU. ncbi.nlm.nih.gov. Recuperado de https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5407387/

Dávila T. J. (2014). Revistamedica. México. Recuperado de http://revistamedica.imss.gob.mx/editorial/index.php/revista_medica/article/viewFile/21/54

Dorfman, A. (2018). Towards a Routine External Evaluation Protocol for Small Area Estimation. https://onlinelibrary.wiley.com/doi/full/10.1111/insr.12248

Federación Mexicana de Diabetes. (2014). Diabetes-en-mexico. México. Recuperado de http://fmdiabetes.org/diabetes-en-mexico/

Griffith, D. & Arbia, G. (2010) Detecting negative spatial autocorrelation in georeferenced random variables, International Journal of Geographical Information Science, 24:3, 417-437, DOI: 10.1080/13658810902832591.

Grifth, D. & ¬Paelinck, J. Morphisms for Quantitative. Spatial Analysis Advanced Studies in Theoretical and Applied Econometrics. Springer. ISBN 978-3-319-72552-9 ISBN 978-3-319-72553-6 (eBook). https://doi.org/10.1007/978-3-319-72553-6

Hawkley L. (2017). NCBI. EEUU. ncbi.nlm.nih.gov. Recuperado de https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2841310/

Hidiroglou, M. A., & Patak, Z. (2009). An Application of Small Area Estimation Techniques to the Canadian Labour Force Survey. Proceedings of the Survey Methods Section.

Hunsinger, E. (2008) Ajuste Proporcional Iterativo para Tablas de dos Dimensiones https://edyhsgr.github.io/eddieh/IPFDescription/AKDOLWDIPFTWOD.pdf

Instituto Canario de Estadística (2008). Estadísticas Laborales. Encuesta de Población Activa. Metodología para la estimación en pequeñas áreas de Canarias.

Instituto Nacional de Estadística y Geografía (INEGI). Encuesta Nacional de Salud y Nutrición (ENSANUT) 2018. Consultado en https://www.inegi.org.mx/programas/ensanut/2018/

Instituto Nacional de Estadística y Geografía (INEGI). Encuesta Intercensal 2015. Consultado en https://www.inegi.org.mx/programas/intercensal/2015/

Instituto Nacional de Estadística y Geografía (INEGI). Estadísticas de Salud en Establecimientos Particulares 2018. Consultado en https://www.inegi.org.mx/programas/salud/

Instituto Nacional de Estadística y Geografía (INEGI), Marco Geoestadístico (2018). Consultado en https://www.inegi.org.mx/temas/mg/default.html#Descargas

Instituto Nacional de Estadística y Geografía (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. Consultado en https://sc.inegi.org.mx/repositorioNormateca/O_05Sep18.pdf

Instituto Nacional de Estadística y Geografía (INEGI). Registros administrativos, defunciones 2018. Consultado en https://www.inegi.org.mx/programas/mortalidad/

INEGI. (2015). internet.contenidos.inegi. México. Recuperado de http://internet.contenidos.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825075019.pdf

Krapavickait, D., & Rudys, T. (2015). Small area estimates for the fraction of the unemployed. Lithuanian Mathematical Journal, Vol. 55, No. 2, April, 2015, pp. 243–254. 2015 Springer Science+Business Media New York. DOI 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. Disponible en:

https://cran.r-project.org/web/packages/mme/index.html.

Martini, A. (2016). Small Area Estimation of employment and unemployment for Local Labour

McEwin, M. & Elazar, D. (2006). Regional Statistics: Small Area Estimation in Official Statistics. Australian Bureau of Statistics (Room document for APEX 2, Daejon, Republic of Korea, 21-22 September, 2006.

Molina, I., Saei, A., & Lombardía, M. J. (2007). Small area estimates of labour force participation under a multinomial logit mixed model. J. R. Statist. Soc. A, 975-1000.

Molina I. (2018). Desagregación de datos en encuestas en hogares. Metodología de estimación en áreas pequeñas. En Serie Estudios Estadísticos. CEPAL. Naciones Unidas.

Molina, Isabel & Nandram, Balgobin & Rao, J. (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. The Annals of Applied Statistics.

https://www.researchgate.net/publication/264425103_Small_area_estimation_of_general_parameters_with_application_to_poverty_indicators_A_hierarchical_Bayes_approach

Nguyen T. N. (2011). Pubmed.gov. EE.UU. Recuperado de https://pubmed.ncbi.nlm.nih.gov/21128002/

Paradis, E. (2012). Moran´s Autocorrelation Coefficient in Comparative Methods.

Pérez. J., Renatas, K. & Luis, M. (2003). Autocorrelación. Proyecto e-Math, Universidad de Oberta de Cataluña. 2003.

Pratesi, M. & Salvati N. (2007). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Stat. Meth. & Appl. (2008) 17:113–141. DOI 10.1007/s10260-007-0061-9.

Ramos Q. R., Bucyibaruta G. y Rivera R. F. J., (2018). Tasas de Fecundidad a Nivel Municipal con base en la Encuesta Intercensal 2015. Centro de Investigación en Matemáticas. México. Reporte de proyecto del Fondo Sectorial Conacyt - INEGI. Documento no publicado.

Ramos Q. R. (2018). Estimación de Áreas Pequeñas: Modelos Básicos. Centro de Investigación en Matemáticas. México. Notas de curso en el marco del proyecto del Fondo Sectorial Conacyt - INEGI. Documento no publicado.

Ramos, R., Bucyibaruta, G. y Rivera, F. (2019). Tasas de Fecundidad a Nivel Municipal con Base en la Encuesta Intercensal 2015. Centro de Investigación en Matemáticas (CIMAT), Guanajuato, México. 2019.

Rao J. N. K. (2003). Small Area Estimation. Wiley Series in Survey Methodology. USA.

Rao, J. (2003). Small Area Estimation. Wiley, New York, (2003).

Rao J. & Molina, I. (2015). Small Area Estimation. Second Edition, WILEY, USA, 2015,

Rao J. & Mingyu, Y. (1994) Small Area Estimation by Combining Time Series and Cross-Sectional Data (paper). Carleton University. Canadian Journal of Statistics - Wiley Online Library

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. En Realidad Datos y Espacio. Revista internacional de Estadística y Geografía. Vol. 6. Núm. 3. Septiembre – diciembre 2015. INEGI. México.

Song, Soon (2011, June). Small area estimation of unemployment: From feasibility to implementation. Paper presented at the New Zealand Association of Economists Conference, Wellington, New Zealand.

The EURAREA Consortium (2004). Enchancing Estimation Techniques to meet European Needs: SAS Programs a Documentation.

The EURAREA Consortium. (2005). Enhancing Small Area Estimation Techniques to meet European Needs. Vol 1 pp C4-12 – C4-14, Vol 3 pp 3C-1 – 3C-5.

Torben, K. (2010). Spatial model selection and spatial knowledge spillovers: a regional view of Germany. ZEW Discussion Papers, No. 10-005, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim.}

Tzavidis N., Zhang L.-C., Rojas-Perilla N., Luna A. & Schmid T. (2018). From start to finish: a framework for the production of small area official statistics. En Journal of the Royal Statistical Society: Serie A (Statistics in Society), Read paper, 2018, 181, pp. 927-979.

U.S. BUREAU OF LABOR STATISTICS. (2018). Local Area Unemployment Statistics. Handbook of Methods.

WILLIAM W. D. (1988). NCBI. EEUU. ncbi.nlm.nih.gov. Recuperado de https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1350294/pdf/amjph00245-0108.pdf

WCRF International. (2018). World Cancer Research Fund International, Reino Unido. Recuperado de https://www.wcrf.org/int/blog/articles/2018/11/can-too-much-screen-time-affect-our-weight

FE DE ERRATAS

Página	Descripción	Decía	Dice	Fecha
14	Se modificó una palabra.	útimo	último	21 de julio, 2020
17	Se agregó un acento.	Preparacion	Preparación	21 de julio, 2020
20	Se eliminó parte de una oración.	El valor del Índice de Moran así como la correlación de Pearson es igual a:	El valor del Índice de Moran es igual a:	21 de julio, 2020
20	Se modificó un valor numérico.	Obesidad = 0.1222	Obesidad = 0.1684	21 de julio, 2020
20	Se modificó un valor numérico.	Diabetes = 0.1684	Diabetes = 0.0615	21 de julio, 2020
40	Se modificó parte de una oración.	… Distrito Federal …	… Ciudad de México …	21 de julio, 2020
41	Se modificó una letra minúscula por una mayúscula	león	León	21 de julio, 2020
42	Se modificó la ilustración.	*Ver la ilustración previa que se encuentra debajo de esta tabla*	*Ver la ilustración posterior que se encuentra debajo de esta tabla*	21 de julio, 2020

Ilustración previa

Identificador único del municipio	Clave de entidad federativa	Entidad federativa	Clave de municipio o delegación	Municipio o delegación	Estimador	Porcentaje de población de 20 años y más con obesidad.	Porcentaje de población de 20 años y más con diagnóstico previo de hipertensión.	Porcentaje de población de 20 años y más con diagnóstico previo de diabetes.
01001	01	Aguascalientes	001	Aguascalientes	Valor	31.0	14.9	7.5
01001	01	Aguascalientes	001	Aguascalientes	Error estándar	2.3	1.2	0.7
01001	01	Aguascalientes	001	Aguascalientes	Límite inferior de confianza	27.2	13.0	6.3
01001	01	Aguascalientes	001	Aguascalientes	Límite superior de confianza	34.9	16.9	8.7
01001	01	Aguascalientes	001	Aguascalientes	Coeficiente de variación	7.5	7.9	9.7

⋮

Identificador único del municipio	Clave de entidad federativa	Entidad federativa	Clave de municipio o delegación	Municipio o delegación	Estimador	Porcentaje de población de 20 años y más con obesidad.	Porcentaje de población de 20 años y más con diagnóstico previo de hipertensión.	Porcentaje de población de 20 años y más con diagnóstico previo de diabetes.
32056	32	Zacatecas	056	Zacatecas	Valor	36.5	20.0	11.8
32056	32	Zacatecas	056	Zacatecas	Error estándar	4.1	2.3	1.1
32056	32	Zacatecas	056	Zacatecas	Límite inferior de confianza	29.7	16.3	10.0
32056	32	Zacatecas	056	Zacatecas	Límite superior de confianza	43.3	23.7	13.7
32056	32	Zacatecas	056	Zacatecas	Coeficiente de variación	11.3	11.3	9.6

Ilustración posterior

Identificador único del municipio	Clave de entidad federativa	Entidad federativa	Clave de municipio o delegación	Municipio o delegación	Estimador	Porcentaje de población de 20 años y más con obesidad.	Porcentaje de población de 20 años y más con diagnóstico previo de hipertensión.	Porcentaje de población de 20 años y más con diagnóstico previo de diabetes.
01001	01	Aguascalientes	001	Aguascalientes	Valor	31.5	14.9	7.5
01001	01	Aguascalientes	001	Aguascalientes	Error estándar	2.3	1.2	0.7
01001	01	Aguascalientes	001	Aguascalientes	Límite inferior de confianza	27.6	13.0	6.3
01001	01	Aguascalientes	001	Aguascalientes	Límite superior de confianza	35.3	16.9	8.7
01001	01	Aguascalientes	001	Aguascalientes	Coeficiente de variación	7.4	7.9	9.7

⋮

Identificador único del municipio	Clave de entidad federativa	Entidad federativa	Clave de municipio o delegación	Municipio o delegación	Estimador	Porcentaje de población de 20 años y más con obesidad.	Porcentaje de población de 20 años y más con diagnóstico previo de hipertensión.	Porcentaje de población de 20 años y más con diagnóstico previo de diabetes.
32056	32	Zacatecas	056	Zacatecas	Valor	36.8	20.0	11.8
32056	32	Zacatecas	056	Zacatecas	Error estándar	4.1	2.3	1.1
32056	32	Zacatecas	056	Zacatecas	Límite inferior de confianza	30.0	16.3	10.0
32056	32	Zacatecas	056	Zacatecas	Límite superior de confianza	43.6	23.7	13.7
32056	32	Zacatecas	056	Zacatecas	Coeficiente de variación	11.2	11.3	9.6