



III Congreso
Nacional de
Estadística



Instituto Nacional
de Estadística

Metodología de Estimación de la Pobreza a Nivel Municipal

Secretaría de Planificación y Programación de la Presidencia –SEGEPLAN–

Contenido

1. Base Legal
2. Antecedentes
3. Justificación técnica
4. Técnica SAE
5. Escenarios para la elección del modelo
6. Aplicación del modelo FH
7. Campos de aplicación
8. Ventajas / Desventajas de su aplicación
9. Consideraciones y desafíos a futuro



Base legal

**DECRETO
NÚMERO
42-2001**

**LEY DE
DESARROLLO
SOCIAL**

Artículo 21. “Mapa de Pobreza. El Estado, por medio de la Secretaría de Planificación y Programación de la Presidencia, en coordinación con el Ministerio de Economía y el Instituto Nacional de Estadística será la responsable de elaborar y mantener actualizado el mapa oficial de pobreza y pobreza extrema, así como los sistemas de información georreferenciados relacionados con las condiciones económicas y sociales de los hogares guatemaltecos, que permita formular estrategias orientadas a la reducción de la pobreza y a lograr las metas propuestas en el Programa de Desarrollo Social y Población”.

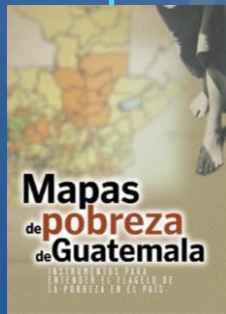


**III Congreso Nacional
de Estadística**



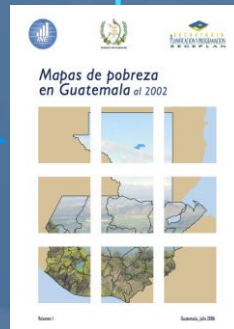
Antecedentes

2002



ENIGFAM 98/99 – Censo de Población y Habitación 1994

2006



ENCOVI 2000 – Censo de Población y Habitación 2002

2013



ENCOVI 2011 – Censos Rurales Municipales 2008-2011

2024



ENCovi 2023 – Censo de Población y Habitación 2018



Congreso Nacional de Estadística



Justificación técnica

Las encuestas de hogares suelen ser la mejor fuente de datos para proporcionar información detallada sobre múltiples indicadores de bienestar entre ellas la pobreza y desigualdad a nivel nacional.

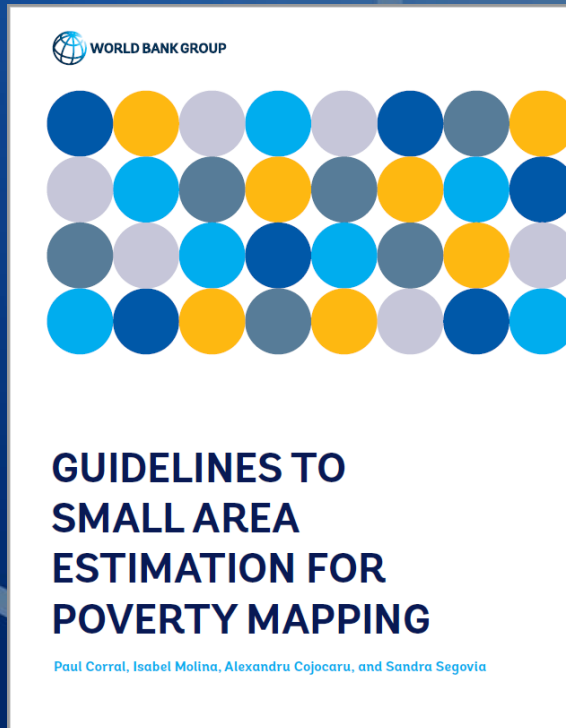
Se planifican con el fin de generar información precisa y confiable en los dominios de estudio que se han predefinido. Sin embargo, las muestras son demasiado pequeñas para ser representativas de subgrupos geográficos o poblacionales sobre los cuales se quisiera obtener estadísticas confiables y para los cuales la encuesta no abordó como parte de su diseño.

Existen dos soluciones a plantear:

- a) Aumentar el tamaño de muestra en los dominios de interés (aumento de los costos)
- b) Aplicar técnicas estadísticas que permitan estimaciones confiables en esos subgrupos de interés (métodos SAE).

Técnica Small Area Estimation –SAE –

SAE (Small Area Estimation) se refiere a las técnicas estadísticas utilizadas para producir estimaciones confiables para **áreas pequeñas geográficas o subpoblaciones** donde los métodos de encuesta tradicionales son insuficientes debido al tamaño de la muestra.

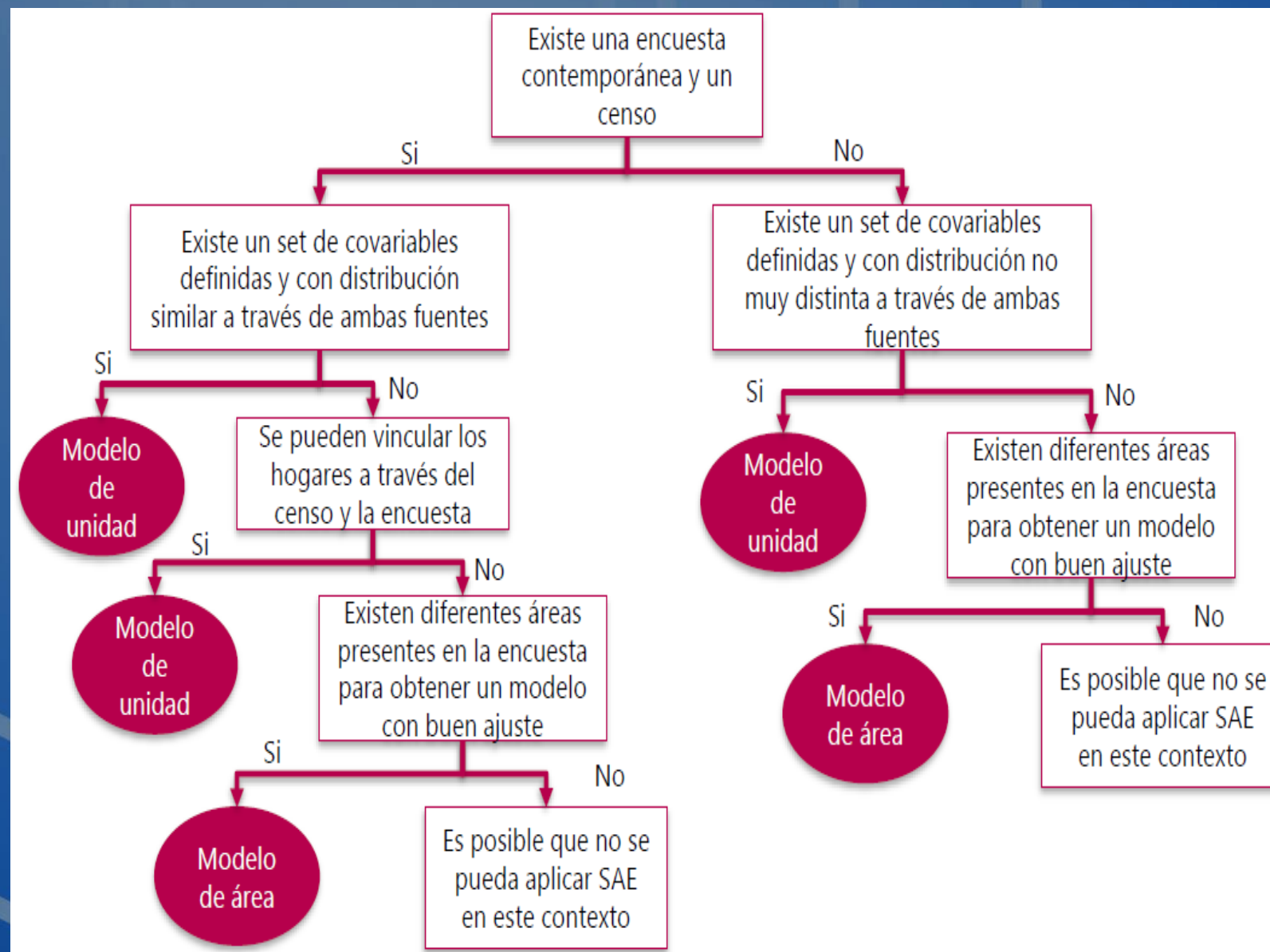


Para obtener estimaciones a nivel municipal, por ejemplo pobreza total, debemos combinar encuesta de hogares y otros datos auxiliares, como el censo de población, para explotar las fortalezas de cada fuente de información.

El Banco Mundial ha publicado las directrices para aplicar métodos estadísticos para estimar pobreza en áreas pequeñas. Entre éstas están: ELL, FH, entre otras.

El documento analiza las ventajas y desventajas de estos métodos en función de los distintos escenarios en cuanto a las fuentes de información disponibles.

Escenarios para la elección del tipo de modelo



Modelo de unidad: se intenta modelar la distribución del bienestar del hogar y a partir de ahí estimar la pobreza (ELL).

Modelo de área: se intenta modelar la distribución espacial de la pobreza (FH).



Aplicación del modelo FH: requisitos de datos

1. Estimaciones directas de los indicadores de interés y su varianza muestral para las áreas consideradas (en la encuesta).

- ENCOVI 2023: 10.964 hogares.
- Estimaciones directas de la pobreza a nivel municipal derivadas de la ENCOVI.
- Tasa nacional de pobreza en ENCOVI 2023: 56.0%.
- Representatividad de la encuesta: nacional, rural/urbana y departamental.
- Conocer la estructura de la muestra de la encuesta (diseño muestral).

2. **Datos agregados:** a nivel de área pequeña (municipio) de todas las covariables disponibles

- Censo de Población y Vivienda 2018.
- Otras fuentes de información auxiliares. (registros administrativos, satelitales, altitud media, etc.)

3. **Variable de ubicación geográfica para vincular** el censo (o cualquier otro dato auxiliar) y la encuesta en ese nivel.

→ En este caso a **nivel municipal**



Aplicación del modelo FH: modelo supuesto -planteamiento

Primera etapa supone que la tasa verdadera de pobreza, y_m para todos los municipios $m = 1, \dots, M$, se relaciona linealmente con un conjunto de covariables a nivel de municipio, x_m , a través del siguiente modelo de vinculación:

$$y_m = x'_m \beta + u_m \quad (1)$$

- Errores aleatorios (efectos de área) y representan una heterogeneidad inexplicable entre áreas, que se supone que tiene una media cero y una varianza constante, σ_u^2 .
- El modelo que aquí se presenta no puede ser adecuado, ya que no se observan las verdaderas estimaciones de la pobreza a nivel de municipio y, en cambio, lo que se observa son las estimaciones directas de la pobreza a nivel de municipio basadas en encuestas, $\hat{y}_m^{dir}$.
- La ENCOVI es representativa a nivel departamental, y aunque las estimaciones a nivel de municipio son posibles, son bastante ruidosas debido al tamaño pequeño de las muestras. Sin embargo, a pesar de ser ruidosas, las estimaciones directas son en su mayoría insesgadas.

$$E(\hat{y}_m^{dir} - y_m) = 0$$

Segunda etapa Modela el error de muestreo asumiendo que los estimadores directos se centran en las verdaderas tasas de pobreza del municipio:

$$\hat{y}_m^{dir} = y_m + e_m \quad (2)$$

- Los errores, e_m , en la ecuación 2 se suponen ser **heterocedásticos**, donde $var[e_m | y_m] = \psi_m$.



Aplicación del modelo FH: Solución al planteamiento

- Tenga en cuenta que la varianza del error muestral (ψ_m) se supone que se conocen, pero en su lugar se estiman a partir de los datos. Esta es una advertencia de la metodología que se menciona a menudo. Ver Corral et al. (2022, Ch3).

El modelo real estimado reemplaza $\mathbf{y}_m$ en la ecuación 2 con el lado derecho de la ecuación 1. El modelo se ajusta a través de la máxima verosimilitud restringida (REML), pero hay alternativas disponibles.

- Las estimaciones obtenidas del modelo se basan en el mejor predictor lineal insesgado (BLUP), que da como resultado estimaciones que no están sesgadas según el modelo y son "mejores" en el sentido de que minimizan el error cuadrático medio (MSE).
 - No se basa en supuestos de normalidad en los errores del modelo o en los efectos de área, y no está sesgado en el modelo
- La estimación resultante se puede expresar como un promedio ponderado entre la estimación directa basada en la encuesta, $\hat{\mathbf{y}}_m^{dir}$, y el estimador sintético de regresión, $\mathbf{x}_m' \hat{\boldsymbol{\beta}}$.
 - Las $\hat{\boldsymbol{\beta}}$ incorporan las varianzas (σ_u^2, ψ_m)
- Los pesos son dados por $\gamma_m = \sigma_u^2 / (\sigma_u^2 + \psi_m)$.
 - El peso dado al estimador directo es mayor para los municipios donde el tamaño de la muestra es grande.
 - Para áreas no muestreadas el estimador Fay-Herriot colapsa al estimador sintético.



Aplicación del modelo FH: Selección de modelo

- Las covariables se estandarizan (media = 0, desarrollo estándar = 1) antes de la selección del modelo.
- Las covariables no significativas se eliminan del modelo secuencialmente, comenzando con aquellas con los valores p más grandes.
- La etapa de selección del modelo utiliza la opción FH debido a los menores requisitos computacionales.
- Las covariables con un factor de inflación de varianza (VIF) superior a 5 se excluyen del modelo final.
- El modelo final emplea tantas covariables y un intercepto para explicar la pobreza en los 340 municipios.
- El R^2 ajustado del modelo final busca un poder explicativo significativo.
- Es necesario verificar los supuestos del modelo.



Aplicación del modelo FH: proceso de estimación



Aplicación del modelo FH: estimaciones obtenidas

Los modelos de Fay-Herriot son el enfoque tradicional para los casos en los que no es posible acceder a los microdatos o, cuando el censo y la encuesta no están alineados.

- El método consiste en modelar las tasas de pobreza (u otros indicadores) a nivel de área pequeña.
 - Ruidoso pero insesgado
 - Obtenemos la distribución conjunta de las características de la pobreza.
 - Los datos utilizados para el modelado son los mismos que los utilizados para la predicción.
- La estimación resultante es un promedio ponderado entre las estimaciones directas (las derivadas directamente de la encuesta) y las estimaciones basadas en modelos.
 - El peso dado a cada estimación en un área determinada depende del tamaño de la muestra para esa área y de la calidad del modelo.
 - Para las áreas que no están en la muestra, las estimaciones están basadas únicamente en los modelos.
- Debido a que el modelo solo se ajusta a las áreas muestreadas en lugar de a los hogares, las estimaciones obtenidas suelen ser menos eficientes que las obtenidas bajo modelos a nivel de unidad.



Aplicación del modelo FH: evaluación de las estimaciones

La comparación de las estimaciones " $\hat{\theta}$ " se basa en medidas estadísticas de Exactitud (o Sesgo) y Precisión (o Variabilidad)

→ **Exactitud** mide la proximidad de la estimación " $\hat{\theta}$ " al valor real del parámetro poblacional θ .

→ **Precisión** está determinado por la proximidad de los valores replicables de una estimación " $\hat{\theta}$ " al parámetro de población (MSE).

Fay-Herriot sacrifica sesgo para lograr más precisión que los estimadores directos

- El error cuadrático medio estimado (MSE) es una medida de precisión.
- El coeficiente de variación (CV) es una medida relativa del error, representada como $cv(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} / E(\hat{\theta})$, haciéndola más fácil de interpretar que el MSE, pero su interpretación requiere que el valor estimado sea mayor que cero y que los tamaños de muestra sean mayores.



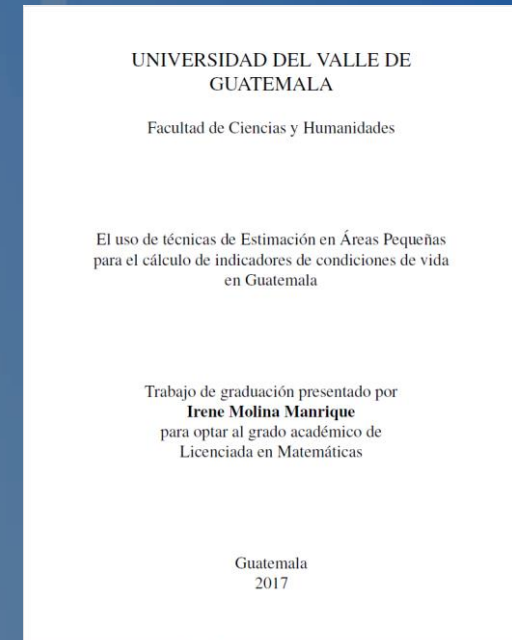
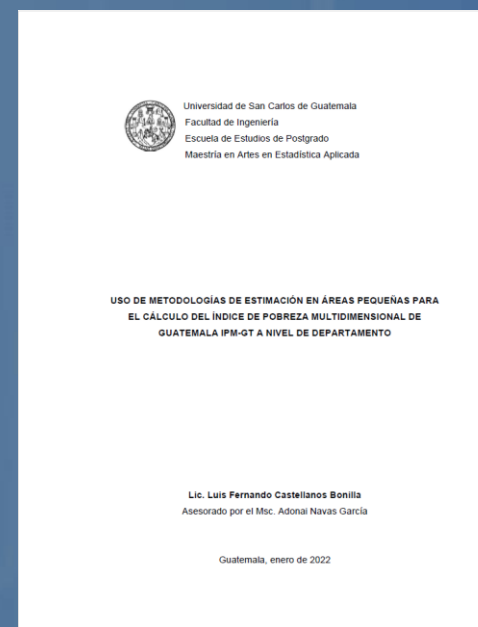
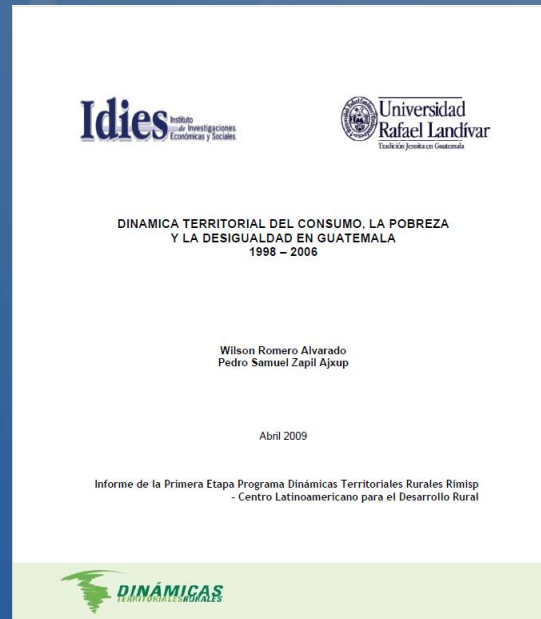
Limitaciones del modelo de Fay-Herriot

- Las estimaciones se basan en un modelo que debe verificarse minuciosamente.
- Las varianzas (ψ_m) se supone que se conocen, pero en su lugar se estiman a partir de los datos. Esta es una advertencia que se menciona a menudo en la metodología.
 - Tenga en cuenta que la obtención de estimaciones para ψ_m no siempre es factible, por ejemplo, cuando todos los habitantes de la zona son pobres o no son pobres o cuando la zona está representada por una sola fuente de alimentación
- El modelo se ajusta solo a las áreas muestreadas, que pueden ser un número muy pequeño del número total de áreas utilizando solo una observación por área (el estimador directo). Como consecuencia, los estimadores de parámetros de modelo serán mucho menos eficientes que los obtenidos bajo modelos a nivel de unidad y, por lo tanto, se espera que las ganancias en precisión sean menores que las de los estimadores basados en modelos a nivel de unidad.
- Cada indicador necesita su propio modelo.

Campos de aplicación

Ampliamente utilizado en diversos campos como la salud pública, la economía y las ciencias sociales.

Las aplicaciones incluyen la estimación de las tasas de desempleo en regiones pequeñas, la evaluación de las necesidades locales de salud y el **mapeo de la pobreza**, por ejemplo.



Ventajas

- Proporcionar datos detallados a nivel local para la formulación de políticas y planificación de programas sociales
- Asignar recursos públicos basados en evidencia estadística
- Establecer alianzas públicas-privadas que permitan el uso eficiente de los recursos disponibles
- Permitir el seguimiento de los indicadores de pobreza a diferentes niveles geográficos y demográficos, de acuerdo con la Agenda de Objetivos de Desarrollo Sostenible (ODS)
- Profundizar la investigación social y análisis del estudio de la pobreza, entre otros.



Desventajas

- Áreas que son demasiado pequeñas pueden presentar problemas de confidencialidad.
- Las estimaciones de área pequeña pueden diferir demasiado de las estadísticas basadas en el conocimiento local.
- Disponibilidad limitada de datos auxiliares correlacionados con la variable de interés.
- Cada indicador de interés a desagregar necesita su propio modelo.
- Tamaño de muestra debe ser suficientemente grande para permitir estimaciones confiables mediante el uso de este método.

Consideraciones a tomar en cuenta

- Todos los métodos SAE requieren datos auxiliares a nivel del área pequeña desde el cual toman fuerza prestada para mejorar las estimaciones.
- La efectividad de los métodos SAE depende del grado de asociación entre la variable de interés y los datos auxiliares.
- La búsqueda de buenas variables auxiliares es crítica, incluida la construcción conceptual de tales variables.
- Los datos auxiliares deben medirse de manera consistente a través de las áreas pequeñas.



Desafíos a futuro

- Aumento de la complejidad en los contenidos de los cuestionarios y por lo tanto la carga de respuesta.
- Aumento de las tasas de no respuesta.
- Aumento de la demanda de estimaciones para áreas pequeñas para grupos poblacionales como etnia entre otras.
- Aumento de la demanda de análisis secundarios de los datos.
- Uso público y archivos de datos de uso restringido.



¡Gracias!