

Aproximación espacial sobre la pobreza monetaria en Bogotá: una aplicación desde la estimación en áreas pequeñas

Carlos Alberto Durán Gil

Universidad Santo Tomás
Facultad de Estadística
División de Ciencias Económicas y Administrativas
Bogotá, D.C., Colombia
2024

Aproximación espacial sobre la pobreza monetaria en Bogotá: Una aplicación desde la estimación en áreas pequeñas

Carlos Alberto Durán Gil

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magíster en Estadística Aplicada

Directores:

Cristian Téllez Piñérez Ph.D.

Felipe Ortiz Rico M.Sc.

Línea de Investigación:

Muestreo y Marketing

Grupo de Investigación:

USTAdística

Universidad Santo Tomás

Facultad de Estadística

División de Ciencias Económicas y Administrativas

Bogotá, D.C., Colombia

2024

A mi papá Alberto, quien no está presente, mi mamá Amparo y mi abuela Rosalba, por inculcarme el interés, la disciplina y la nobleza del proceso de aprendizaje continuo.

A mi esposa Elizabeth, una dedicación especial por todo su apoyo, comprensión y paciencia en el tiempo que me tomó realizar este proceso. A mis hijas Sofía y Camila, espero que este esfuerzo en pro de una meta pueda ser visto como ejemplo para sus proyectos de vida.

A mis hermanos, Sandra y Juan David por su respaldo y motivación constante.

Lo que sabemos es una gota de agua, lo que ignoramos es el océano (Isaac Newton)

Agradecimientos

Expreso mi total agradecimiento a mis directores de trabajo de grado, Cristian Téllez Ph.D. y Felipe Ortiz M.Sc. por su respaldo y asesoría permanente en el desarrollo de este trabajo. Sin duda, fueron los guías más adecuados por su conocimiento y experiencia, para llevar a cabo este proceso de forma exitosa.

Al Departamento Administrativo Nacional de Estadística (DANE), por ser mi segundo hogar, la Entidad donde me he desarrollado profesionalmente, de la cual surgió tanto la idea como los insumos base necesarios para el desarrollo de este trabajo.

A mi compañero Rolando Javier Barajas, con quien desarrollé mis trabajos grupales durante la maestría, entre ellos, las primeras aproximaciones a la aplicación de modelos de estimación en áreas pequeñas con datos de pobreza.

Resumen

La incidencia de la pobreza monetaria es un indicador fundamental en la evaluación de las condiciones socioeconómicas de la población, cuyo seguimiento hace parte de la Agenda de los Objetivos de Desarrollo Sostenible (ODS). Ante la creciente necesidad de contar con información detallada para su monitoreo, este trabajo desarrolla una metodología enfocada en la estimación en áreas pequeñas (SAE) con el objetivo de lograr desagregaciones y mapas de la pobreza monetaria en los hogares a nivel de unidad de planeamiento zonal (UPZ) en la ciudad de Bogotá. Con base en los microdatos derivados de la Gran Encuesta Integrada de Hogares (GEIH) vigencia 2021, y el uso de 25 covariables obtenidas de datos geospaciales, se llevan a cabo modelos Fay-Herriot, con el fin de obtener los mejores estimadores lineales insesgados (EBLUP) junto a sus extensiones robustas espaciales (RSEBLUP), comparando sus precisiones a través de los errores marginales. Los resultados obtenidos reflejan que las covariables empleadas en los modelos son predictoras adecuadas de la pobreza monetaria, y que la adición de la componente espacial al modelo, aplicando procesos robustos, ofrece mejores precisiones en comparación con las estimaciones directas resultantes de la encuesta.

Palabras clave: Encuesta por muestreo, estimación de áreas pequeñas, mejor estimador lineal insesgado, modelo Fay-Herriot, pobreza monetaria.

Abstract

The incidence of monetary poverty is a fundamental indicator in assessing the socioeconomic conditions of the population, and its monitoring is part of the Agenda for the Sustainable Development Goals (SDG). In response to the growing need for detailed information for monitoring purposes, this work develops a methodology focused on small area estimation (SAE) with the aim of achieving disaggregations and maps of monetary poverty in households at the level of the zonal planning unit (UPZ for its acronym in Spanish) in the city of Bogotá. Based on the microdata derived from the Integrated Household Survey (GEIH for its acronym in Spanish) for the year 2021, and the use of 25 covariates obtained from geospatial data, Fay-Herriot models are carried out in order to obtain the best linear unbiased estimators (EBLUP) along with their robust spatial extensions (RSEBLUP), comparing their precisions through marginal errors. The results obtained show that the covariates used in the models are adequate predictors of monetary poverty, and that the addition of the spatial component to the model, applying robust processes, provides better precision compared to the direct estimates resulting from the survey.

Keywords: Sample survey, small-area estimation, best unbiased linear estimator, Fay-Herriot model, monetary poverty.

Lista de Figuras

3-1. Distribución de las UPZ en Bogotá.	13
3-2. Flujo de procesos	15
4-1. Histograma de muestras por UPZ	18
4-2. Margen de error estimación directa por UPZ	19
4-3. Matriz de correlaciones	20
4-4. Prueba Shapiro-Wilk normalidad	22
4-5. Prueba Breush-Pagan homocedasticidad	22
4-6. Normal Cuantil-Cuantil de los residuales y efectos aleatorios	22
4-7. Densidad residuales estandarizados	22
4-8. Densidad efectos aleatorios	22
4-9. Intervalos de confianza estimaciones FH con ED	23
4-10. Gráficos de caja CV	23
4-11. Incremento CV	23
4-12. Errores marginales ED y EBLUP	25
4-13. Resumen estadístico modelo FH robusto espacial	26
4-14. QQ-plot residuales estandarizados	27
4-15. QQ-plot efectos aleatorios	27
4-16. Intervalos de confianza estimaciones RSFH con ED	28
4-17. Margen de error EBLUP y RSEBLUP	28
4-18. Distribución EBLUP	29
4-19. Distribución RSEBLUP	29
4-20. LISA EBLUP	30
4-21. LISA RSEBLUP	30
A-1. Exploración de datos - información auxiliar	36

Lista de Tablas

4-1.	Medidas de diagnóstico de los modelos FH estándar	21
4-2.	UPZ con mayores diferencias en estimaciones	24
B-1.	Comparación estimación directa vs EBLUP	37
C-1.	Predicciones EBLUP en UPZ áreas pequeñas	39
D-1.	Comparación EBLUP vs RSEBLUP	41
E-1.	Predicciones RSEBLUP en UPZ áreas pequeñas	43

1. Introducción

En aras de promover un desarrollo armonizado con el crecimiento económico, la inclusión social y la sostenibilidad ambiental (Qiu et al., 2022), la Organización de las Naciones Unidas (ONU, por sus siglas en español) en el año 2015, propuso 17 Objetivos para el Desarrollo Sostenible (ODS, por sus siglas en español) en el marco de una Agenda Global con horizonte al año 2030, los cuales están conformados por una serie de metas e indicadores medibles. El lema principal de los ODS, "no dejar a nadie atrás", destaca la necesidad de producir información relevante, continua y desagregada, conforme a los principios fundamentales que rigen la producción de estadísticas oficiales (Organización de las Naciones Unidas, 2014). La Comisión de Estadística de la ONU establece que estos indicadores deben desagregarse por componentes temáticos o geográficos, lo cual es esencial para su monitoreo en el contexto social, económico y ambiental, y así mejorar la focalización de políticas a partir de un enfoque marcado por desigualdades espaciales (Organización de las Naciones Unidas, 2017; Qiu et al., 2022; Corral, 2023).

El ODS 1, enfocado en erradicar la pobreza en todas sus formas, requiere de la formulación de programas y políticas que determinen las actuaciones necesarias para su intervención. En ese sentido, está aumentando la necesidad de estadísticas con mayor desagregación territorial, como es el caso de los indicadores de pobreza (Franco, 2023; Pratesi et al., 2023). De esta manera, mejorar las estimaciones de indicadores socioeconómicos a través de información más granular permitirá mejorar el monitoreo, la evaluación, la cobertura y focalización de recursos, en torno a las intervenciones de política pública (Chandra et al., 2018; Asian Development Bank, 2020; Departamento para la Prosperidad Social - DPS, 2021; Newhouse, 2024).

Buena parte de los datos estadísticos se obtienen de forma recurrente por inferencia a través de encuestas, las cuales dependen de una estrategia adecuada que garantice estimaciones confiables para variables e indicadores de interés (Gutiérrez, 2021). En este sentido, la aplicación de encuestas permite obtener un conjunto de estadísticas con representatividad a niveles superiores de agregación, pero no así en subdivisiones más detalladas, donde carecen de precisión (Chandra et al., 2018; Molina, 2019; Comisión Estadística para América Latina y El Caribe - CEPAL, 2021). De hecho, la obtención de estimaciones desagregadas no siempre es viable, debido al tamaño limitado de las muestras que se aplican en las encuestas oficiales sobre los dominios de interés, por lo que aumentar su cobertura es algo que las oficinas na-

cionales de estadística no pueden sostener por costos en recursos administrativos, financieros y técnicos, ante lo cual, métodos para la estimación en áreas pequeñas (SAE, por sus siglas en inglés), permiten obtener valores en subdivisiones geográficas con tamaños de muestra pequeños o nulos (Guadarrama et al., 2016; Franco, 2023), a partir de modelos construidos con datos de la propia encuesta y otras variables auxiliares, con el fin de lograr estimaciones detalladas de adecuada precisión (Asian Development Bank, 2020; Gutiérrez, 2021; Corral, 2023). Por lo tanto, SAE se considera adecuado sobre indicadores que son costosos de recolectar directamente, y que pueden ser estimados usando datos auxiliares (Newhouse, 2024). Es así que, empleando información auxiliar derivada de censos, registros administrativos, datos geospaciales y otras encuestas, se logra reducir la incertidumbre en comparación con los estimadores directos, cumpliendo con los umbrales de calidad establecidos por las oficinas de estadística (Franco, 2023).

Los métodos SAE se han empleado para la estimación de medidas e indicadores en diferentes temáticas socioeconómicas, entre ellas la pobreza (Asian Development Bank, 2020; Harming et al., 2023). A nivel global, el Banco Mundial emplea el método de Elbers, Lanjouw y Lanjouw (ELL) para realizar estimaciones y mapas de pobreza o desigualdad (Molina, 2019). Asimismo, la Comisión Europea desarrolló el proyecto para las Estimaciones de Condiciones de Vida y Pobreza con Métodos de Áreas Pequeñas (SAMPLE, por sus siglas en inglés), usando datos de encuestas con la ayuda de registros administrativos (Molina and Morales, 2009). La Comisión Económica para América Latina y El Caribe (CEPAL por sus siglas en español) desarrolló un modelo para realizar estimaciones de los ingresos promedio en 18 países de América Latina (Comisión Estadística para América Latina y El Caribe - CEPAL, 2021). Respecto a las aplicaciones realizadas en países, Estados Unidos construyó un programa de estimaciones de ingresos y pobreza en áreas pequeñas, denominado SAIPE (por sus siglas en inglés), para producir datos oficiales en diversos grupos poblacionales y niveles geográficos (Franco, 2021, 2023). En Italia, se han aplicado métodos SAE para la estimación del índice de pobreza educacional por regiones geo-estadísticas (Pratesi et al., 2021), mientras que en España se han desarrollado modelos para estimar la incidencia y la brecha de pobreza con enfoque de género en las provincias (Molina, 2019). En Chile, se emplearon variables correlacionadas para estimar las tasas de pobreza comunales, las cuales fueron ajustadas por el método SAE Fay-Herriot (Ministerio de Desarrollo Social de Chile, 2013).

En Colombia, el Departamento de la Prosperidad Social (DPS, por sus siglas en español) desarrolló modelos SAE para la estimación de pobreza monetaria y pobreza extrema en los hogares (Departamento para la Prosperidad Social - DPS, 2021). Otros trabajos son reseñados por Ortiz et al. (2024), destacando las estimaciones sobre líneas de pobreza a nivel de municipios y localidades, empleando modelos Fay-Herriot a partir del uso de datos de la Encuesta Multipropósito. De esta manera, es posible obtener estimaciones en áreas peque-

ñas usando diferentes tipos de herramientas estadísticas (Ministerio de Desarrollo Social de Chile, 2013), con beneficios potenciales en torno a la reducción de la incertidumbre relativa a las estimaciones y obtenerlas, incluso, en áreas donde no se tuvo muestra para las encuestas (Franco, 2022).

Los datos geospaciales proporcionan una alternativa factible para la aplicación de modelos SAE, ya que se correlacionan con buena parte de las variables de interés a estimar (Newhouse, 2024). Con el fin de obtener mediciones granulares de fenómenos socioeconómicos, estos datos se convierten en una mejor opción para derivar variables auxiliares, en comparación con las encuestas o registros administrativos (Masaki et al., 2021; Merfeld et al., 2023), siendo una alternativa válida para proporcionar estimaciones actualizadas en áreas pequeñas hasta que se tengan nuevos datos detallados provenientes de un próximo censo (Edochie et al., 2024). Asimismo, es sabido que los datos de encuestas exhiben generalmente dependencia entre áreas (Sikov and Cerda, 2024) y que los fenómenos socioeconómicos como la pobreza presentan patrones espaciales provocados por causas naturales y antrópicas (Pusponegoro and Rachmawati, 2018; Pratesi et al., 2023), por lo que el enfoque geoespacial es crucial en muchas aplicaciones para el modelado en áreas pequeñas, ya que puede mejorar la eficiencia y efectividad de las estimaciones al incorporar los efectos de áreas correlacionadas espacialmente a través de un criterio de vecindad (Pratesi et al., 2023).

Adicionalmente, la representación de la pobreza a través de mapas, refleja la variación de grados de bienestar o desigualdad social en niveles detallados de desagregación territorial (Molina, 2019; Delatie, 2021), como se ha podido corroborar a través de la metodología que ha desarrollado CEPAL en los países de la región, entre ellos Colombia (Gutiérrez et al., 2022). De esta manera, la descripción de la distribución espacial de la pobreza dentro de un país se convierte en una herramienta útil para que los gobiernos y organizaciones tomen decisiones de política con capacidad de enfocar las acciones en sitios y grupos poblacionales de mayor interés (Artelaris and Kandyliis, 2014; Molina et al., 2014; Molina, 2019; Comisión Estadística para América Latina y El Caribe - CEPAL, 2021).

Los modelos SAE a nivel de área como el Fay-Herriot (FH) se han desarrollado como una forma eficiente para mejorar las estimaciones directas y obtener predicciones en áreas no muestreadas, tomando fuerza prestada en todas las áreas en consistencia con el diseño de la muestra, a partir de información auxiliar procedente de fuentes de datos alternativas, (Corral, 2023; Sikov and Cerda, 2024). Ahora bien, como el modelo básico FH supone que los efectos aleatorios son independientes entre las áreas, se han propuesto extensiones para dar cabida a las situaciones producidas por datos derivados de entornos complejos y realistas (Sikov and Cerda, 2024). De esta manera, extensiones del modelo FH que incorporan la correlación espacial se han aplicado para el mapeo de pobreza (Corral et al., 2022). En Indonesia se obtuvieron predicciones afectadas por procesos espaciales autorregresivos para

la estimación de la pobreza a nivel sub-distrital (Pusponegoro and Rachmawati, 2018). En España se aplicó un modelo SAE con efectos de correlación espacial y temporal para estimar la incidencia de la pobreza en las provincias (Marhuenda et al., 2013), mientras que en Italia se han logrado estimaciones de la pobreza a nivel regional, empleando modelos de área con efectos de correlación espacial (Giusti et al., 2017).

El presente estudio tiene como objetivo principal estimar la incidencia de pobreza monetaria en los hogares del año 2021, en las Unidades de Planeamiento Zonal (UPZ) en Bogotá, mediante técnicas de estimación en áreas pequeñas, con el fin de obtener información relevante y precisa a niveles intraurbanos, y así dar soporte a la producción estadística detallada requerida para la toma de decisiones informada en contextos locales. La aproximación espacial al proceso busca fortalecer la identificación de datos auxiliares y los modelos a emplear que, junto al análisis de los resultados, permitirá la evaluación de posibles patrones o tendencias que generen valor agregado a la comprensión y atención de este fenómeno.

2. Marco Teórico

2.1. Pobreza monetaria

La medición de la pobreza monetaria se basa en los ingresos, a través de la definición de un umbral mínimo, debajo del cual una persona u hogar no puede satisfacer los costos de sus necesidades básicas (Comisión Estadística para América Latina y El Caribe - CEPAL, 2021; Departamento Administrativo Nacional de Estadística - DANE, 2023a). Una persona se clasifica como pobre cuando el ingreso per cápita de su hogar es inferior a la línea de pobreza, es decir, no puede pagar el costo para satisfacer sus necesidades alimentarias definidas a través de la construcción de canastas básicas de alimentos (Gutiérrez et al., 2022).

En Colombia, la medición de la pobreza monetaria es un método indirecto que requiere de dos elementos a saber: 1) las líneas de pobreza definidas por valores asociados a una canasta de alimentos y/o bienes básicos, las cuales se extraen de la Encuesta Nacional de Ingresos y Gastos; 2) el ingreso per cápita por Unidad de Gasto, es decir, la persona que atiende sus propios gastos o los hogares que comparten viviendas para satisfacer necesidades en común, información que se obtiene a través de la Gran Encuesta Integrada de Hogares (GEIH) (Botello and Valderrama, nd; Consejo Nacional de Política Económica y Social - CONPES, 2012; Departamento Administrativo Nacional de Estadística - DANE, 2021, 2023a).

Uno de los índices más empleados para la medición de la pobreza es el conocido como FGT (Foster, Greer y Thorbecke), el cual, al ser una medida paramétrica, brinda a los analistas y tomadores de decisiones un instrumento para evaluar la pobreza bajo diferentes ópticas y sensibilidades en torno a la distribución de ingresos (Foster et al., 2010; Ministerio de Economía y Finanzas de Perú, nd). Este es un índice de carencias en el consumo que toma como referencia una línea de pobreza definida, otorgando pesos a la proporción de la población (Navarro and Chávez, 2001), cuya propiedad de descomposición aditiva la convierte en una medida de mayor aplicabilidad, ya que permite establecer diferentes niveles de pobreza que se relacionan con el de la población total (Foster et al., 2010). Su formulación es la siguiente (Departamento Administrativo Nacional de Estadística - DANE, 2024):

$$P_{\alpha}(y, z) = \frac{1}{n} \sum_{i=1}^q \left(\frac{g_i}{z} \right)^{\alpha} \quad (2-1)$$

donde, i es el hogar; z corresponde a la línea de pobreza; g_i el nivel de ingreso que le hace falta al hogar para alcanzar el nivel de bienestar; q el número de personas definidas como pobres; n el número total de personas; y α el parámetro de “aversión a la pobreza” (Foster et al., 2010), el cual determina las condiciones que cumple el índice. Con $\alpha = 0$ se obtiene la tasa o incidencia de la pobreza, es decir, la proporción de población que cae debajo del umbral de ingresos definido, mientras que con $\alpha = 1$ se deriva la brecha e intensidad de la pobreza. Asimismo, con valores α superiores, se puede obtener una medida de severidad de la pobreza, con base en la distribución per cápita entre quienes son catalogados como pobres (Ministerio de Economía y Finanzas de Perú, nd).

Con base en lo anterior, es posible afirmar que la pobreza, al tener relación con los ingresos y las condiciones de vida de las personas y los hogares, se considera una variable fundamental para el estudio de las desigualdades socioeconómicas en el territorio (Castañeda et al., 2019).

2.2. Estimación en áreas pequeñas

Las encuestas realizadas por entidades y organizaciones que pertenecen a un sistema estadístico nacional, están diseñadas para producir estimaciones en niveles geográficos o poblacionales definidos (Molina, 2019). En estimaciones de encuestas sobre áreas geográficas pequeñas, el bajo tamaño de la muestra puede derivar en resultados sin suficiente precisión, lo que hace necesario el empleo de modelos soportados en información auxiliar de variada complejidad (Ardilly et al., 2018). Justamente, la estimación en áreas pequeñas (SAE) es cada vez más importante para la inferencia basada en modelos, sobre todo en encuestas de estadísticas sociales y económicas (Marhuenda et al., 2014). Ahora bien, el desarrollo de metodologías estadísticas para predecir proporciones en áreas pequeñas requiere del entendimiento del fenómeno a estudiar (Boubeta et al., 2023).

Los estimadores basados en modelos surgen como alternativa para incrementar el “tamaño efectivo de la muestra”, empleando datos externos a los dominios objetivo, en aras de disminuir el error estándar de la estimación directa (Rao and Molina, 2015; Morales et al., 2021). En otras palabras, intentan tomar fortaleza prestada de otros dominios para lograr estimaciones en áreas no planificadas o cuyo tamaño de la muestra es pequeño, donde no es posible obtener estimaciones directas de forma confiable (Veijanen and Lehtonen, 2011).

Es así que los métodos SAE, basados en modelos, desarrollan procedimientos estadísticos para producir estimaciones eficientes de algún parámetro con su correspondiente error de predicción (Pratesi et al., 2023), a través de modelos mixtos que incluyen efectos fijos y aleatorios (Pusponegoro and Rachmawati, 2018), en donde el rol de la información auxiliar y su asociación con la variable de interés es clave para la confiabilidad de las estimaciones (Veijanen and Lehtonen, 2011; Téllez, 2020). Los modelos de efectos mixtos pueden ajustarse

a la variable de interés, y así producir medidas de incertidumbre específicas del dominio o área, junto a la verificación de supuestos a través de diagnósticos (Wakefield, 2022). En la parte de los efectos aleatorios, se tienen en cuenta las variaciones entre áreas más allá de las explicadas por las covariables del modelo (Chandra et al., 2010). Por consiguiente, en estos modelos se pueden incorporar efectos jerárquicos, correlaciones espaciales y temporales, debido a su flexibilidad y habilidad para combinar distintas fuentes de información, convirtiéndose en herramientas adecuadas para la estimación en áreas pequeñas (Morales et al., 2021; Wakefield, 2022).

Los estimadores SAE introducen los mejores predictores (BP por sus siglas en inglés) que minimizan el error medio cuadrático (MSE por sus siglas en inglés), medida común para evaluar la precisión de las estimaciones (Harmening et al., 2020). Ahora bien, la calidad de las predicciones dependerá en gran medida de los supuestos distribucionales (Warnholz, 2016). En efecto, los mejores predictores se sustituyen por estimadores apropiados en dominios pequeños, denominados BP empíricos (EBP por sus siglas en inglés) y, desde el supuesto de normalidad en la distribución de los datos, la estimación puede considerarse como el Mejor Predictor Lineal Empírico Insegado (EBLUP por sus siglas en inglés) (Téllez, 2020; Burgard et al., 2022). Asimismo, se han definido extensiones a través de la estructura de covarianza en los modelos que tienen correlaciones espaciales entre áreas, derivando en el EBLUP espacial (SEBLUP por sus siglas en inglés) (Puspongoro and Rachmawati, 2018; Pratesi et al., 2023).

Dentro de los modelos SAE se destaca el Fay-Herriot (FH), popular con datos disponibles de forma agregada, algo común en los estudios socioeconómicos de pobreza (Molina, 2022; Pratesi et al., 2023). Este se compone de dos niveles: i) el modelo de muestreo usado para representar el error del estimador directo; ii) el modelo de enlace con las características reales del área, las cuales varían linealmente con un número definido de variables auxiliares (Marhuenda et al., 2013; Novo, 2018). En suma, el modelo FH logra capturar los efectos específicos de cada área a través de la componente aleatoria, algo que no se puede obtener mediante los modelos de regresión clásicos (Novo, 2018). Su definición se presenta a continuación (Nandy, 2013):

$$y_d = \mathbf{x}_d^T \boldsymbol{\beta} + u_d + e_d \quad (2-2)$$

donde, y_d es el estimador insegado de la variable de interés; \mathbf{x}_d el vector de variables conocidas; $\boldsymbol{\beta}$ el vector de coeficientes de regresión desconocidos; u_d los efectos aleatorios independientes de área y; e_d los errores de muestreo con media cero y varianza conocida. No obstante, aunque este modelo asume que las varianzas son conocidas, en la práctica no lo son, por lo que se hace necesario estimarlas (Merfeld et al., 2023).

Con la teoría de los modelos mixtos es posible obtener estimaciones de los parámetros del dominio, siendo el EBLUP el promedio ponderado del estimador directo y la parte sintética

de regresión (Harmening et al., 2023). Además, el estimador del MSE se puede utilizar en predictores que son expresados como sumas ponderadas de los valores de la muestra (Chambers et al., 2009). De esta manera, el EBLUP le da un mayor peso al estimador de regresión cuando la muestra es demasiado pequeña en un área determinada (Molina et al., 2019). Su formulación es la siguiente (Schork, 2022):

$$\hat{\mu}_d^{FH} = \hat{\gamma}_d \hat{\mu}_d^{Dir} + (1 - \hat{\gamma}_d) \mathbf{x}_d^T \hat{\beta} \quad (2-3)$$

donde $\hat{\mu}_d^{FH}$ es el estimador FH del parámetro de interés; $\hat{\gamma}_d$ el factor de contracción; $\hat{\mu}_d^{Dir}$ el estimador directo del parámetro de interés; \mathbf{x}_d^T el vector transpuesto de variables auxiliares a nivel de área y; $\hat{\beta}$ el vector de efectos fijos para todos los dominios. El factor de contracción se sitúa entre 0 y 1, disminuye con la varianza del error, y asigna la ponderación al estimador directo (Corral et al., 2022). Se define de la siguiente forma:

$$\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_{ed}^2} \quad (2-4)$$

donde $\hat{\sigma}_{ed}^2$ es la varianza estimada de los errores de muestreo; y $\hat{\sigma}_u^2$ la varianza estimada de los efectos aleatorios. Para su estimación se emplean métodos de verosimilitud, ya sea Máxima (ML por sus siglas en inglés), o Restringida (REML por sus siglas en inglés), más eficientes en comparación a los métodos de momentos, aunque pueden producir varianzas negativas o iguales a cero (Harmening et al., 2023), algo que puede ocurrir en estimaciones de pobreza, en áreas donde la muestra se aproxime a valores cero o uno (Merfeld et al., 2023).

El modelo FH asume la independencia de los efectos aleatorios pero, en áreas geográficas, se hace necesario suponer que están correlacionados con alguna estructura de vecindad cuantificable (Harmening et al., 2023), ya que un área estará influenciada por aquellas circundantes, donde el supuesto de efectos aleatorios correlacionados parece ser más aceptable (Asfar et al., 2016; Warnholz, 2016). Por consiguiente, la estructura de autocorrelación espacial se incorpora a través de la vecindad (Tzavidis et al., 2018), bajo el supuesto que las relaciones espaciales se modelan por procesos autorregresivos simultáneos de orden uno (SAR(1) por sus siglas en inglés), particularmente en aplicaciones socioeconómicas que describen estructuras globales de correlación y dependencia espacial (Kubacki and Jędrzejczak, 2016; Warnholz, 2016; Harmening et al., 2023; Pratesi et al., 2023). De esta manera, el efecto aleatorio es una función de la matriz de pesos espaciales, mientras que los efectos fijos son invariantes espacialmente (Chandra et al., 2010; Puspongoro and Rachmawati, 2018; Pratesi et al., 2023).

Al incluir la dependencia espacial en el componente de error de los efectos aleatorios de área, el modelo sigue un proceso SAR, también conocido como Fay-Herriot espacial, lo que conduce a una mayor eficiencia en la estimación de los parámetros y a estimaciones en áreas pequeñas más eficientes, ya que se mejora la estructura de varianza-covarianza de los modelos que presentan correlaciones entre áreas (Marhuenda et al., 2013; Risal, 2021). A continuación,

se establece formalmente el modelo lineal mixto al cual se le incorpora la autocorrelación espacial entre áreas (Salvati, 2004; Kubacki and Jędrzejczak, 2016):

$$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} \quad (2-5)$$

donde \mathbf{X} es la matriz de variables auxiliares dependientes del área; $\boldsymbol{\beta}$ el vector de parámetros de regresión; \mathbf{Z} la matriz de constantes positivas conocidas y \mathbf{v} el vector de la variación de segundo orden. La dependencia espacial incluida en el componente del error del efecto aleatorio sigue un proceso SAR (Risal, 2021). Este modelo se define formalmente como se muestra a continuación (Kubacki and Jędrzejczak, 2016):

$$\mathbf{v} = \rho\mathbf{W}\mathbf{v} + \mathbf{u} \quad (2-6)$$

donde ρ es el parámetro de autorregresión espacial, \mathbf{W} la matriz de pesos espaciales y; \mathbf{u} el vector del término de error independiente con media cero y varianza constante. Por supuesto, si ρ es igual a cero, entonces el SEBLUP será igual al EBLUP (Crescenzi et al., 2015). Es importante señalar que la matriz de pesos espaciales, al ser un componente importante en el modelado de datos para la obtención de estimaciones precisas, requiere ser seleccionada conforme a las características y relaciones de las áreas geográficas, en donde a partir de simulaciones se ha concluido que las matrices más óptimas son aquellas establecidas a partir de criterios de contigüidad o por definición de vecinos más cercanos (Asfar et al., 2016).

De esta manera, en el SEBLUP se estima el coeficiente autorregresivo espacial, el cual, a un mayor valor absoluto, determinará una mayor fortaleza en la relación entre las áreas vecinas y donde los efectos aleatorios siguen un proceso SAR(1) en función de su rezago espacial o promedio ponderado de la variable en ubicaciones vecinas (Harmening et al., 2020). El estimador SEBLUP se define por medio de la siguiente fórmula (Kubacki and Jędrzejczak, 2016):

$$\tilde{\boldsymbol{\theta}}_i = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}} + \mathbf{b}_i^T \mathbf{G}\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \quad (2-7)$$

donde $\tilde{\boldsymbol{\theta}}_i$ es el estimador SEBLUP en la i -ésima posición; \mathbf{x}_i^T el estimador dependiente de la varianza y el parámetro de autorregresión en la i -ésima posición; $\tilde{\boldsymbol{\beta}}$ el estimador de mínimos cuadrados generalizados del parámetro de regresión; \mathbf{b}_i^T el vector con valor 1 en la posición i -ésima; \mathbf{G} la matriz de covarianza de los efectos aleatorios SAR; \mathbf{V} la matriz de covarianza del estimador del parámetro de interés; $\hat{\boldsymbol{\theta}}$ el estimador directo; \mathbf{X} la matriz de variables auxiliares dependientes del área.

El MSE del SEBLUP depende de dos parámetros, la componente de varianza desconocida y la autocorrelación espacial, a través de la siguiente formulación (Pratesi and Salvati, 2009):

$$MSE[SEBLUP] = g_{1i}(\sigma_u^2, \rho) + g_{2i}(\sigma_u^2, \rho) \quad (2-8)$$

donde σ_u^2 es la componente de varianza desconocida; ρ el parámetro de autocorrelación espacial; $\mathbf{g}_{1i}(\sigma_u^2, \rho)$ el término asociado a la estimación de los efectos aleatorios; y $\mathbf{g}_{2i}(\sigma_u^2, \rho)$ el término asociado a la estimación de los parámetros de regresión.

Aunque los supuestos del modelo respecto a la distribución de los efectos aleatorios se pueden cumplir en la mayoría de las observaciones, podría haber áreas que no se ajusten a estos supuestos u observaciones individuales que no estén descritas adecuadamente por el supuesto del modelo en torno a los términos de error, ante lo cual, su influencia en la estimación podría ser reducida al ajustar las ecuaciones del modelo (Gershunskaya and Lahiri, 2011).

De esta manera, cuando se tienen observaciones influyentes y los supuestos de normalidad no se cumplen, se desarrollan versiones robustas de los modelos FH estándar y espacial, entendidas como variaciones de la máxima verosimilitud (Tzavidis et al., 2018; Pratesi et al., 2023). La estimación de modelos lineales mixtos robustos se ha modificado para tener en cuenta los efectos aleatorios correlacionados, por lo que el EBLUP robusto (REBLUP) se define de la siguiente manera (Warnholz, 2016):

$$\hat{\theta}_i^{REBLUP} = \mathbf{x}_i^T \hat{\beta}^\psi + \mathbf{z}_i^T \hat{\mu}^\psi \quad (2-9)$$

donde: $\hat{\beta}^\psi$ y $\hat{\mu}^\psi$ son las soluciones robustas para los parámetros de varianza; \mathbf{x}_i denota el vector de información auxiliar; y \mathbf{z}_i el vector columna que contiene la fila i -ésima en la matriz identidad que depende del modelo específico. Aquí, ψ corresponde a la función de influencia acotada, la cual se aplica por separado tanto a los residuales como a los efectos de área estimados del modelo, haciendo que las ecuaciones de estimación sean robustas (Chambers et al., 2014).

En estos casos, los estimadores de MSE basados en pseudo-linearización son una alternativa a otros métodos basados en máxima verosimilitud y *bootstrap* (Chambers et al., 2009), ya que los estimadores SAE son pseudo-lineales en su estructura, con pesos que dependen de los valores muestrales (Chambers et al., 2011). Por consiguiente, esta es una solución analítica para el estimador robusto del MSE, que se adapta fácilmente a diversos estimadores, gracias a sus algoritmos orientados en encontrar soluciones para la variabilidad adicional asociada con la estimación de los parámetros de varianza, aunque se parte del supuesto subyacente que los pesos y valores muestreados son independientes (Warnholz, 2016).

Dado que los métodos SAE pueden diferir en tres aspectos clave, a saber: i) nivel al cual los modelos son propuestos; ii) método de estimación específico a usar; iii) el tipo de datos auxiliares a emplear (Merfeld et al., 2023), en su aplicación es fundamental evaluar los diagnósticos, con el fin de verificar tanto los supuestos subyacentes del modelo, como la fiabilidad y validez de las estimaciones (Chandra et al., 2018).

3. Metodología

Este trabajo abarca diferentes alcances en el marco de una investigación de tipo cuantitativo (Hernández et al., 2014), así: i) descriptivo, a partir de la definición y exploración de variables auxiliares que se emplearán en los modelos; ii) correlacional, al obtener estimaciones con base en relaciones entre variables; iii) explicativo, brindando una aproximación al entendimiento de la pobreza monetaria urbana.

Desde el enfoque espacial, el estudio se enmarca en la metodología PPDAC (Problema, Plan, Datos, Análisis y Conclusiones, por sus siglas en español), la cual es dinámica, flexible y empleada en problemas de contexto socioeconómico con enfoque geoespacial a diferentes escalas (de Smith et al., 2024). En este sentido, cada componente se describe así: i) el problema se delimita en obtener datos desagregados de pobreza monetaria a niveles intraurbanos; ii) el plan se lleva a cabo mediante métodos SAE con enfoque espacial; iii) los datos corresponden a la información de pobreza monetaria obtenida mediante encuesta por muestreo y las covariables definidas como información auxiliar; iv) el análisis se enfoca en la exploración de datos, modelos aplicados y estimaciones obtenidas; v) las conclusiones se enmarcan en el cumplimiento de objetivos y la discusión de los resultados.

Adicionalmente, el método SAE provee un marco analítico que permite mejorar la desagregación de información, integrando y aprovechando otras fuentes de información (Asian Development Bank, 2020).

3.1. Área de estudio

Corresponde a la zona urbana de Bogotá y su división por 112 Unidades de Planeamiento Zonal (UPZ), áreas contenidas en las localidades y que sirven para la planificación del desarrollo urbano (Cámara de Comercio de Bogotá, nd). Están conformadas por barrios, manteniendo unidad morfológica y funcional, lo que condiciona la aplicación de las políticas públicas (Unidad Administrativa Especial de Catastro Distrital, nd). En la Figura **3-1** se presenta el mapa con la distribución de las UPZ (Caballero, 2015).

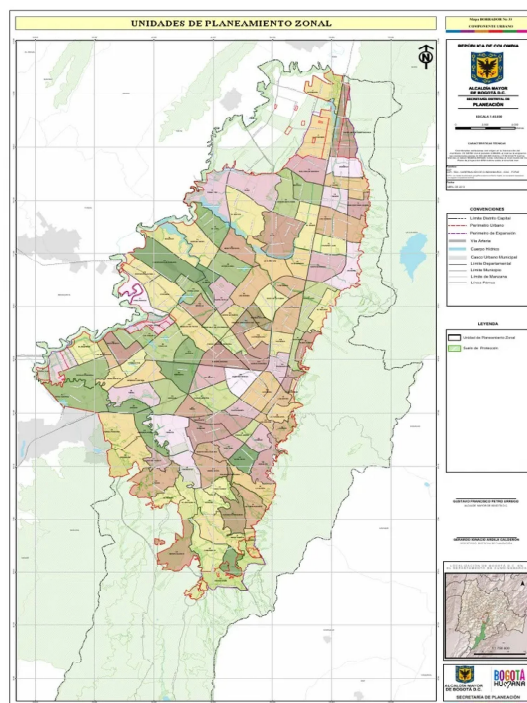


Figura 3-1.: Distribución de las UPZ en Bogotá.

3.2. Datos

El primer conjunto de datos corresponde a la base de los hogares respecto a ingresos, unidades de gasto y valor de las líneas de pobreza para la vigencia 2021 (Departamento Administrativo Nacional de Estadística - DANE, 2023a). Se extrajeron los datos para la ciudad de Bogotá, en donde se identifican aquellos hogares considerados pobres monetariamente a través de una variable binaria.

La segunda base de datos corresponde a la información auxiliar de 25 covariables a emplear en los modelos SAE, las cuales fueron derivadas a nivel de UPZ desde datos abiertos geoespaciales (Unidad Administrativa Especial de Catastro Distrital - UAECD, 2024; Open Street Map, 2024). Estas covariables con su nombre clave son: 1) densidad de población (V1_DensPob); 2) proporción de los delitos de alto impacto (V2_PrDelit); 3) proporción de Instituciones Prestadoras de Salud (V3_PrIPS); 4) sumatoria del valor de referencia del terreno sobre el área (V4_MedValT); 5) presencia de comisarías de familia (V5_ComFam); 6) proporción de jardines infantiles (V6_PrJInf); 7) proporción de pensionados activos (V7_PrPens); 8) proporción de víctimas del conflicto (V8_PrVict); 9) proporción de colegios (V9_PrColeg); 10) tasa de reprobación en hombres en colegios oficiales (V10_TRepH); 11) tasa de repro-

bación de mujeres en colegios oficiales (V10_TRepM); 12) tasa de deserción de hombres en colegios oficiales (V11a_TDesH); 13) tasa de deserción de mujeres en colegios oficiales (V11b_TDesM); 14) densidad de arbolado urbano (V12_DenArb) ¹; 15) proporción de ocupaciones ilegales (V13_PrOcup); 16) proporción de área en parques (V14_PrParq); 17) presencia de agregaciones de empresas (V15_ClusEc); 18) proporción de puntos críticos de arrojo clandestino de residuos (V16_PtArCl); 19) presencia de suelo en riesgo (V17_ARiesg); 20) distancia mínima a universidades (V18a_DtUnv); 21) distancia mínima a centros de salud (V18b_DtSal); 22) distancia mínima a centros de seguridad (V18c_DtSeg); 23) distancia mínima a centros comerciales (V18d_DtCC); 24) densidad de unidades económicas (V19_DensUE); 25) presencia de estratos socioeconómicos 1 y 2 (V20_Estrat).

3.3. Procesos

El flujo de procesos, representado de forma esquemática en la figura **3-2**, adapta el procedimiento analítico de estimación para modelos a nivel de área de Harmening et al. (2020), en donde se contemplan todas las etapas secuenciales necesarias, desde la estimación directa, la creación de los modelos, hasta sus correspondientes diagnósticos. La ejecución se desarrolló en R, gracias a un conjunto de librerías y funciones que permiten implementar un flujo de trabajo completo para diferentes modelos SAE (Merfeld et al., 2023).

A continuación, se describen los procesos del esquema, destacando que cada uno de ellos depende del anterior.

3.3.1. Base de datos por UPZ

Construcción de la base de datos con las variables temáticas y del diseño muestral requeridas para las estimaciones directas por UPZ, con información para 81 de estas unidades de un total de 112, equivalente al 72,3 %.

3.3.2. Estimación Directa - ED

Obtención de las estimaciones directas de la incidencia de la pobreza monetaria en los hogares para cada UPZ, las cuales dependen únicamente de los datos de la muestra (Harmening et al., 2020), junto a su margen de error (MoE por sus siglas en inglés o error marginal), que mide la precisión de una estimación a partir de la amplitud del intervalo de confianza (Departamento Administrativo Nacional de Estadística - DANE, 2023b).

¹Las zonas pobres presentan escasez de zonas verdes en comparación con otras de mayores ingresos (<https://climatica.coop/falta-zonas-verdes-barrios-mas-pobres-calor/>; <https://www.scientificamerican.com/article/trees-are-missing-in-low-income-neighborhoods/>)

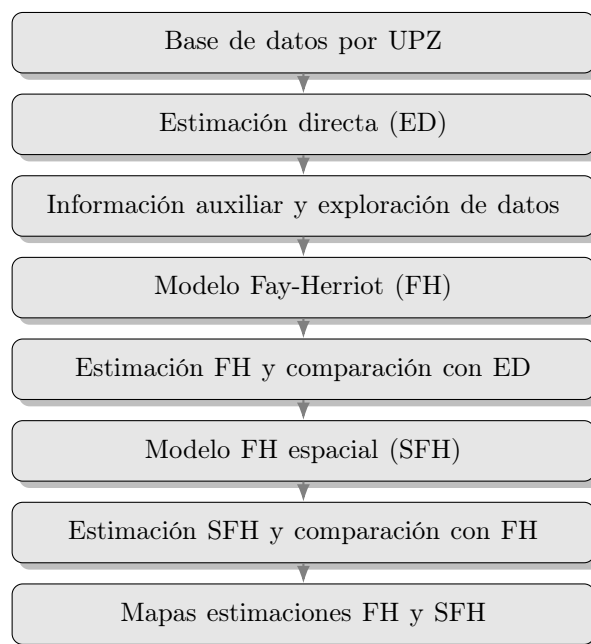


Figura 3-2.: Flujo de procesos

3.3.3. Información auxiliar y exploración de datos

Incorporación y exploración de datos de las 25 covariables seleccionadas como información auxiliar para desarrollar la estimación basada en modelos SAE, empleando resúmenes estadísticos y distribuciones, así como una matriz de correlaciones entre estas covariables y la estimación directa obtenida en el proceso anterior.

3.3.4. Modelo Fay-Herriot - FH

La selección de modelos lineales mixtos es diferente a la aplicada en modelos para regresión lineal (Müller et al., 2013), por lo que es importante evitar un modelo donde todas las variables auxiliares sean significativas, ya que se daría un efecto de sobre-suavizado que no lograría capturar de forma adecuada la variabilidad entre dominios (Morales et al., 2021). De esta manera, la elección de un modelo FH apropiado es un paso fundamental para la estimación de áreas pequeñas, por lo que se deben aplicar diferentes consideraciones conceptuales con el fin de determinar cuáles variables auxiliares deben hacer parte del modelo (Marhuenda et al., 2014; Harmening et al., 2020).

De acuerdo con Marhuenda et al.(2014), uno de los criterios de información más empleados, Akaike (AIC por sus siglas en inglés), es sumamente eficiente cuando el tamaño de la muestra es grande mientras que el número de parámetros desconocidos es pequeño; ahora bien, el criterio de divergencia simétrica de Kullback para la selección de modelos lineales se basa en la divergencia Kullback-Leibler, que a su vez se utiliza para desarrollar el AIC, logrando

capturar diferentes características del modelo, siendo una de sus variantes *bootstrap*, KICb2, la de mejor comportamiento en torno a la selección del modelo y el costo computacional.

Otro procedimiento para la selección de modelos lineales es a través de un operador de selección y contracción mínima absoluta, conocido como regresión LASSO (por sus siglas en inglés), el cual deriva un conjunto de covariables que ofrece las mejores predicciones a usar en un modelo mixto, a través de procesos de validación cruzada, evitando el sobreajuste de los datos al seleccionar modelos con valores similares del coeficiente de determinación (Merfeld et al., 2023; Edochie et al., 2024; Newhouse, 2024). Este método busca encontrar el valor óptimo del hiperparámetro de regularización que minimice el error medio cuadrático, obteniendo los coeficientes de las covariables representativas.

A diferencia de los métodos de selección anteriores, uno que sí considera los efectos aleatorios y que en esencia es similar a una eliminación paso a paso para modelos de regresión de efectos mixtos, corresponde a los mínimos cuadrados generalizados (GLS por sus siglas en inglés), en donde se remueven secuencialmente las covariables que no son significativas, siguiendo con aquellas que presentan multicolinealidad (Corral et al., 2022). Esta alternativa es efectiva para modelados sobre conjuntos de datos que exhiben heterocedasticidad y/o algún tipo de autocorrelación (timeseriesreasoning.com, nd).

Teniendo en cuenta que la estimación se basa en modelos, es importante incluir sus diagnósticos (Kreutzmann et al., 2019), en aras de elegir el modelo con mejores métricas.

3.3.5. Estimación FH y comparación con ED

Con el modelo seleccionado, se analizan los supuestos distribucionales y se comparan los EBLUP con las estimaciones directas, cuyas diferencias no deberían ser significativas (Harmening et al., 2020). Una medida de interés es la tasa de cobertura, que indica la proporción de áreas en la que el valor directo está dentro del intervalo de confianza estimado por el modelo, siendo una forma de evaluar la precisión de las medidas de incertidumbre (Merfeld et al., 2023).

Los resultados del modelo Fay-Herriot deben ser consistentes con las estimaciones directas en aquellos dominios muestreados y que tengan un MSE pequeño, aunque la precisión de las nuevas estimaciones debe ser mejor por el uso de información auxiliar (Harmening et al., 2020). Además, dentro de los beneficios de los métodos SAE, está el de proporcionar estimaciones para áreas sin muestra (Franco, 2023). Por consiguiente, con el modelo seleccionado, se realizan las predicciones de la incidencia de la pobreza monetaria en los hogares para aquellas UPZ no incluidas en la estimación directa. Es importante señalar que una de las máximas de la estimación de áreas pequeñas es que, predecir bien entre las áreas, es más

importante que estimar mejor dentro de una misma área (Newhouse, 2024).

3.3.6. Modelo FH espacial - SFH

El aspecto fundamental en este proceso es identificar las estructuras espaciales con la ayuda de una matriz de proximidad, la cual define la relación espacial entre cada par de dominios (Morales et al., 2021), más la aplicación de pruebas estadísticas de autocorrelación espacial global, cuyos resultados soportarán la incorporación de esta estructura en el modelo (Harmening et al., 2020). Teniendo en cuenta la configuración de las UPZ, las cuales se distribuyen de forma contigua en el territorio, se aplicó el criterio reina (*queen*) para la conformación de la matriz de pesos, que toma en cuenta la vecindad por límite compartido de un área a otra (Risal, 2021), siendo este uno de los enfoques recomendados cuando se manifiestan los efectos de correlación espacial (Asfar et al., 2016).

3.3.7. Estimación SFH y comparación con FH

En muchas aplicaciones de la vida real, las estimaciones del MSE vía máxima verosimilitud o *bootstrap* pueden arrojar valores negativos (Marhuenda et al., 2016). Adicionalmente, puede ocurrir que los SEBLUP mediante FH espacial y sus MSE estimados a través de ML, REML o *bootstrap* no logren mejoras en comparación con los EBLUP, por lo que las extensiones espaciales robustas del modelo estándar permiten abordar diversas situaciones que surgen en aplicaciones prácticas (Harmening et al., 2020). Por lo tanto, se realiza la aplicación de estos modelos con la obtención de MSE mediante formas pseudo-lineales, robustas a la heterocedasticidad (Chambers et al., 2009) y al sesgo, siendo sugeridas para la estimación de áreas pequeñas (Chambers et al., 2011).

Con el modelo robusto FH espacial evaluado se realizan las predicciones de la incidencia de la pobreza monetaria por hogares en las UPZ no incluidas en la estimación directa.

3.3.8. Mapas estimaciones FH y SFH

Los resultados de las estimaciones EBLUP y RSEBLUP para todas las UPZ se presentan en mapas temáticos para analizar su distribución espacial. Adicionalmente, con el fin de encontrar patrones estadísticamente significativos de la incidencia de la pobreza monetaria en los hogares por UPZ, se aplica el Indicador Local de Asociación Espacial (LISA por sus siglas en inglés) de Moran, el cual es una desagregación de su homónimo global, para determinar los patrones de agrupamiento dentro de la región de estudio (de Smith et al., 2024). En otras palabras, con este proceso se busca determinar con mayor certeza la presencia de patrones espaciales de pobreza monetaria en las UPZ, a fin de brindar resultados plausibles para la focalización de este fenómeno a niveles intraurbanos.

4. Resultados

4.1. Estimación directa

La base de datos a nivel de UPZ brinda información para 81 de estas unidades, de un total de 112, equivalente al 72,3 %. Ahora bien, se refleja una distribución asimétrica respecto a la cantidad de muestras (figura 4-1), en donde la mayor frecuencia oscila en el rango entre 50 y 100 muestras por UPZ. Se destacan las UPZ 028-El Rincón y 071-Tibabuyes, en la localidad de Suba, con más de 300 muestras cada una, mientras que las UPZ 033-Sosiego (localidad San Cristóbal) y 116-Álamos (localidad de Engativá), contienen menos de 30 muestras.

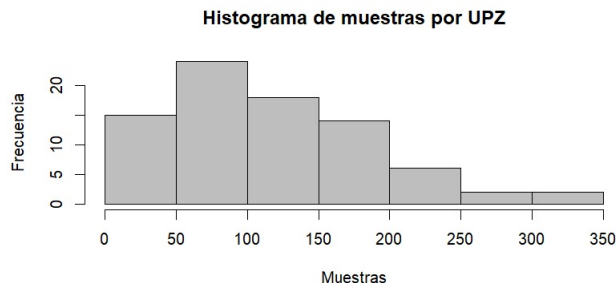


Figura 4-1.: Histograma de muestras por UPZ

En las estimaciones directas, se destacan las siguientes UPZ con mayor proporción de hogares en pobreza monetaria, por encima del 50 %: 052-La Flora (63,8 %) y 059-Alfonso López (58,1 %), ambas en la localidad de Usme; 053-Marco Fidel Suárez (57,3 %), localidad Rafael Uribe Uribe; 062-Tunjuelito (56,7 %) en la localidad del mismo nombre; 080-Corabastos (55,8 %) en la localidad de Kennedy; 095-Las Cruces (55,6 %) en la localidad Santa Fe; 068-El Tesoro (51,3 %) en la localidad de Ciudad Bolívar. De otro lado, las UPZ con menor proporción de hogares en pobreza monetaria, inferior al 2 %, son: 097-Chicó Lago y 013-Los Cedros, ubicadas en las localidades de Chapinero y Usaquén, respectivamente.

Al revisar los errores estándar de las estimaciones, se evidenció que en 15 UPZ sus valores eran cero, siendo estas las que tienen menor cantidad de muestras (inferior a 50). Al estimar proporciones, se necesitan tamaños de muestra más grandes (Corral et al., 2022), por lo que

dichas estimaciones no son confiables, ya que no es posible determinar la precisión de estos resultados a través del margen de error o el coeficiente de variación, provocando efectos no deseados dentro del modelado SAE. Por lo tanto, estas UPZ se catalogaron como áreas no muestreadas para que sus predicciones sean obtenidas vía modelos, en adición a las 31 que no contaban con muestra. De esta manera, con estimaciones directas para 66 UPZ, se evalúan los MoE según la cantidad de muestra en la figura 4-2.

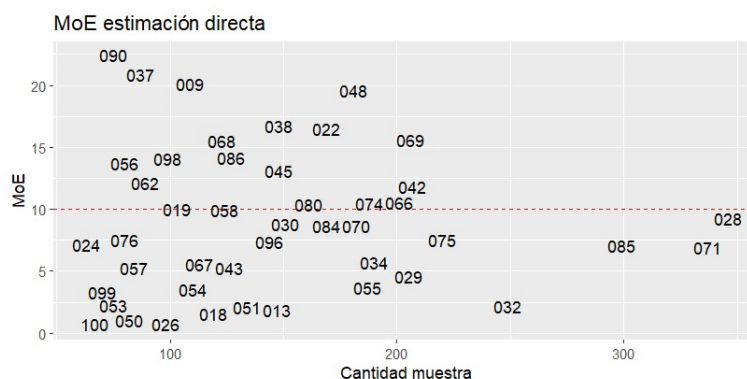


Figura 4-2.: Margen de error estimación directa por UPZ

De estas mediciones, se resalta que en 19 UPZ los errores marginales de las estimaciones directas son superiores al 10 %. Inicialmente, no se evidencian patrones o tendencias en torno a esta variabilidad dada por valores altos del MoE, ya sea por el tamaño de la muestra, los valores estimados o la ubicación de las UPZ.

4.2. Información auxiliar y exploración de datos

La exploración de las 25 covariables, a través de estadísticas descriptivas, se presenta en el Anexo A. Se realizó el escalamiento de los datos ya que las variables presentaban diferentes unidades de medida, sumado a distribuciones asimétricas. A continuación, se obtuvo la matriz de correlaciones de la figura 4-3.

Se destacan correlaciones directas significativas entre la incidencia de la pobreza monetaria con: porcentaje de jardines infantiles; porcentaje de víctimas; presencia de áreas en riesgo; y presencia de estratos socioeconómicos 1 y 2. Esto refleja que en las UPZ con mayor incidencia de pobreza en los hogares, se tendría mayor presencia de personas en situación de vulnerabilidad. Respecto a las correlaciones inversas, son significativas entre la pobreza y el porcentaje de IPS, así como con el valor de referencia del terreno, indicando que la valoración del suelo urbano y la prestación de servicios de salud se dan en UPZ con menor incidencia de pobreza en los hogares.

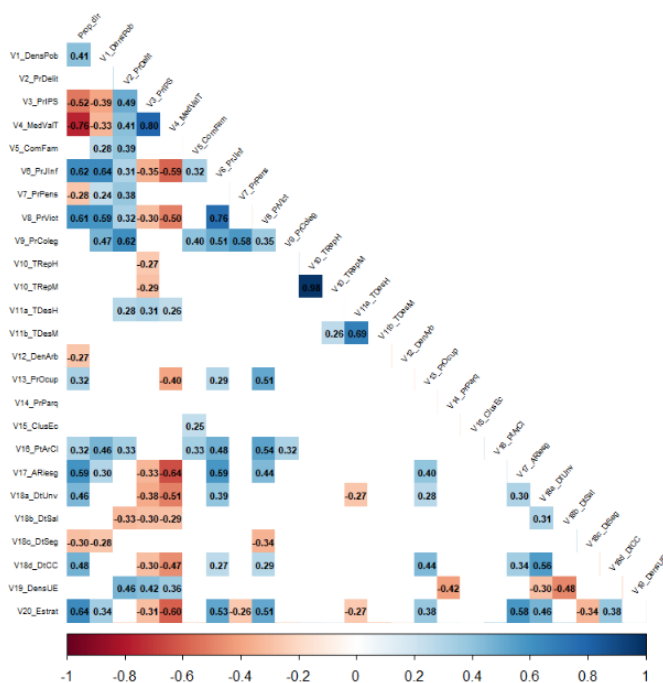


Figura 4-3.: Matriz de correlaciones

4.3. Modelo Fay-Herriot

Al aplicar los criterios de selección de covariables que harán parte del modelo Fay-Herriot, se obtuvieron los siguientes resultados.

- En el caso de la selección paso a paso *stepwise* bajo el criterio *KICb2* fueron seleccionadas las siguientes: V4_MedValT, V12_DenArb, V17_ARiesg y V19_DensUE.
- Para el procedimiento de selección a través de penalización LASSO, las covariables con mayor capacidad predictora son: V4_MedValT, V6_PrJInf, V7_PrPens, V8_PrVict, V12_DenArb, V17_ARiesg, V18d_DtCC, V19_DensUE y V20_Estrat.
- Referente al ajuste de modelos a través de mínimos cuadrados generalizados (GLS), las covariables a seleccionar son V4_MedValT, V6_PrJInf, V9_PrColeg, V12_DenArb, V18d_DtCC y V19_DensUE.

Para todos los casos, las covariables de mayor influencia respecto a la estimación del porcentaje de hogares en pobreza monetaria por UPZ son: el valor de referencia del terreno, la densidad de arbolado urbano, y la densidad de unidades económicas.

Los diagnósticos de los modelos ejecutados para su correspondiente comparación se efectuaron a través de las siguientes medidas:

- Componente(s) de varianza estimados (EVC), que determinan la varianza de los efectos aleatorios explicada por el modelo;
- Criterios *AIC* y *BIC*, que a valores más bajos indican mejor ajuste del modelo con menor pérdida de información;
- Criterio de divergencia simétrica de Kullback variante *bootstrap KICb2*, empleado para seleccionar modelos con covariables que maximizan la información;
- Bondad del ajuste a través del R^2 ajustado (AdjR2);
- p-valores de los estadísticos Shapiro-Wilk (SWp) y Breusch-Pagan (BPp), para la validación de supuestos de normalidad y homocedasticidad en los residuales de los efectos aleatorios, respectivamente.

Los valores obtenidos en las medidas de diagnóstico se presentan en la tabla 4-1.

Tabla 4-1.: Medidas de diagnóstico de los modelos FH estándar

Modelo	EVC	AIC	BIC	KICb2	Adj R2	SWp	BPp
FH - selección Stepwise KICb2	0,00581	-124,87	-111,73	-115,49	0,715	0,067	0,617
FH - selección LASSO	0,00454	-126,27	-102,18	-104,73	0,733	0,352	0,499
FH - selección GLS	0,00483	-130,51	-112,99	-117,32	0,746	0,113	0,613

En los tres modelos, los valores del criterio AIC son cercanos, indicando que las covariables seleccionadas explican una parte sustancial de la variabilidad entre las áreas (Tzavidis et al., 2018). Asimismo, los valores de EVC, BIC, KICb2 y R^2 ajustado también son próximos entre sí. Ahora bien, el modelo FH ajustado con las covariables seleccionadas por GLS presenta los valores más bajos en los tres criterios de información, así como un R^2 ajustado mayor, sumado al cumplimiento de los supuestos de normalidad y homocedasticidad en los efectos aleatorios, gracias a los p-valores que no permiten rechazar las hipótesis nulas de las pruebas estadísticas de Shapiro-Wilk y Breusch-Pagan, respectivamente.

De esta manera, se presenta el modelo Fay-Herriot seleccionado para realizar la estimación de la incidencia de pobreza monetaria en los hogares por UPZ:

$$\widehat{PobMon}_{UPZ} = 0,2756 - 0,0948(V4MedValt) + 0,0507(V6PrJInf) - 0,0281(V9PrColeg) - 0,0271(V12DenArb) + 0,0312(V18dDtCC) + 0,0422(V19DensUE) \quad (4-1)$$

Las pruebas estadísticas de normalidad y homocedasticidad de los residuales de los efectos aleatorios obtenidas en el modelo se presentan en las figuras 4-4 y 4-5 respectivamente.

```
Residual diagnostics:
Skewness Kurtosis Shapiro_W Shapiro_p
Standardized_Residuals 0.8633271 4.175852 0.9562744 0.02054976
Random_effects -0.1162008 4.575107 0.9702141 0.11304544
```

Figura 4-4.: Prueba Shapiro-Wilk normalidad

```
studentized Breusch-Pagan test
data: model_residuals2
BP = 4.4689, df = 6, p-value = 0.6135
```

Figura 4-5.: Prueba Breush-Pagan homocedasticidad

El comportamiento de los errores del modelo a través de los *Q-Q plots* y los gráficos de densidad se presentan en las figuras 4-6, 4-7 y 4-8.

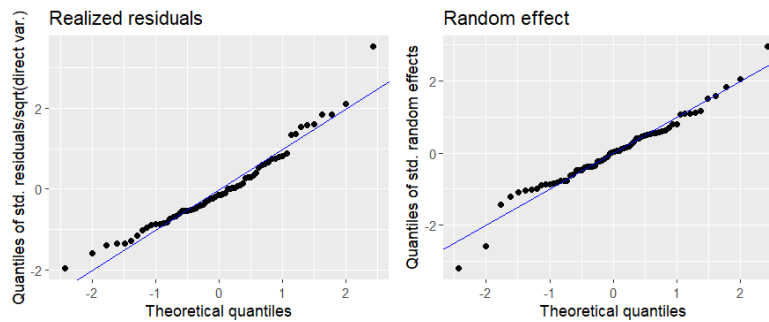


Figura 4-6.: Normal Cuantil-Cuantil de los residuales y efectos aleatorios

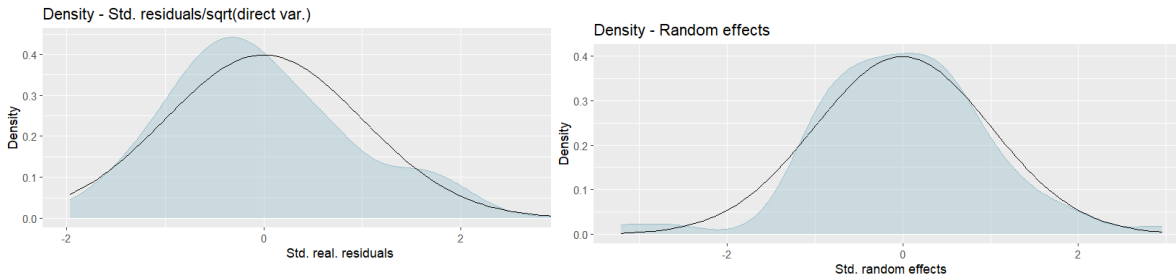


Figura 4-7.: Densidad residuales estandarizados

Figura 4-8.: Densidad efectos aleatorios

Se evidencia una aproximación razonable a la normalidad, principalmente en los efectos aleatorios, tanto en el gráfico normal cuantil-cuantil como en el gráfico de la densidad.

4.4. Estimación FH y comparación con ED

Se realizó la prueba de Brown, cuya hipótesis nula establece que los EBLUP no difieren significativamente de las estimaciones directas, arrojando que no hay suficiente evidencia estadística para rechazarla (estadístico Wald 7,37 y p-valor igual a 1). En la prueba estadística de diagnóstico de cobertura, no hay suficiente evidencia estadística para rechazar la

hipótesis inicial, es decir, los intervalos de confianza se superponen en al menos el 95 %, lo que se evidencia en la figura 4-9 que compara las ED con los intervalos de confianza de los EBLUP. Además, la correlación entre la parte sintética del modelo y el estimador directo es de 0,88.

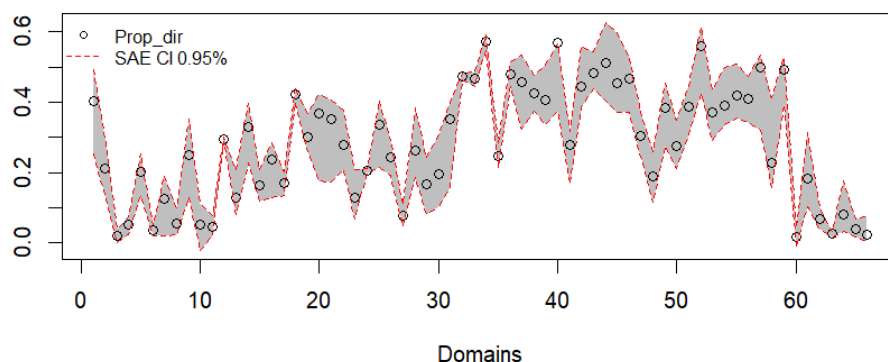


Figura 4-9.: Intervalos de confianza estimaciones FH con ED

Los gráficos de comparación de los coeficientes de variación derivados de las estimaciones directas y los obtenidos mediante modelos, se presentan en las figuras 4-10 y 4-11.

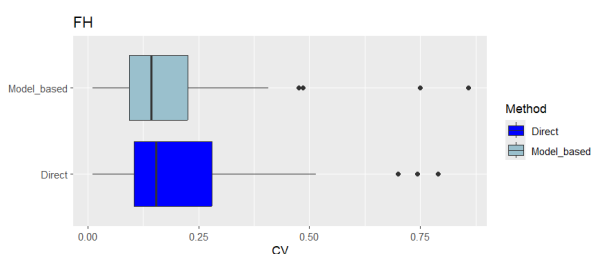


Figura 4-10.: Gráficos de caja CV

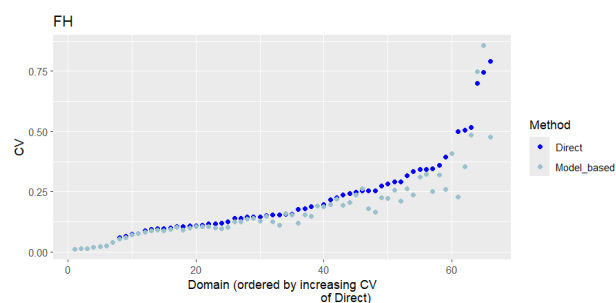


Figura 4-11.: Incremento CV

Al comparar los *boxplot*, la precisión de los resultados del modelo FH supera a las estimaciones directas, ya que el rango intercuartil de los coeficientes de variación es menor. Asimismo, el gráfico del CV incremental refleja que las estimaciones basadas en el modelo presentan reducciones de error respecto a las estimaciones directas, a medida que aumentan los coeficientes de variación. Esto supone un beneficio de los EBLUP, ya que al darle ponderación a la parte sintética de la regresión, se mejora la precisión de las estimaciones.

En el Anexo B se presentan los resultados de la estimación directa (ED), estimación Fay-Herriot (EBLUP), valor γ , margen de error de la estimación directa (ME_ED) y margen de error de la estimación FH (ME_FH) para cada UPZ con su tamaño de muestra (n). De acuerdo con el EBLUP, en tres UPZ se presentan incidencias de pobreza monetaria en los hogares por encima del 50 % (053-Marco Fidel Suárez, 080-Corabastos; 068-El Tesoro). Respecto al valor γ , en 55 UPZ tiene mayor peso la estimación directa (valores superiores a 50), mientras que en 11 lo tiene la estimación basada en el modelo, siendo estas: 009-Verbenal; 022-Doce de Octubre; 037-Santa Isabel; 038-Restrepo; 048-Timiza; 056-Danubio; 068-El Tesoro; 069-Ismael Perdomo; 086-El Porvenir; 090-Pardo Rubio; 098-Los Alcázares. En la tabla 4-2 se presentan las UPZ con mayores diferencias entre la ED y el EBLUP.

Tabla 4-2.: UPZ con mayores diferencias en estimaciones

UPZ	NOMBRE	ED	EBLUP	DIF
062	Tunjuelito	56,77	46,97	9,8
048	Timiza	35,23	27,79	7,44
086	El Porvenir	49,81	42,82	6,99
037	Santa Isabel	36,82	29,97	6,85
038	Restrepo	35,31	28,88	6,43
090	Pardo Rubio	22,88	28,54	5,66

Uno de los aspectos más relevantes del EBLUP es que en todas las UPZ se reduce el error marginal respecto a la estimación directa. En la figura 4-12 se manifiesta la tendencia a esta disminución a medida que los valores del error marginal son más altos para la ED.

Las predicciones obtenidas sobre UPZ no muestreadas o con escasa muestra extraídas de las estimaciones directas, se presentan en la tabla del Anexo C. Para validar su precisión se realizaron análisis de sensibilidad, a través de dos escenarios: i) variación de un 10 % en las seis covariables auxiliares, por encima y por debajo de sus valores de forma intercalada; ii) modificación de la varianza de la estimación directa, incrementándola un 20 %. La comparación de las predicciones originales con las obtenidas en estos dos escenarios, arroja que la media de las diferencias absolutas para el primer caso es del 1,06 %, y para el segundo del 0,1 %. Por lo tanto, las predicciones de la incidencia de la pobreza en los hogares, para estas UPZ, presentan precisiones adecuadas.

De estas predicciones, se resaltan tres UPZ con valores de incidencia de pobreza monetaria por encima del 50 %, a saber: 089-San Isidro Patios (57,81 %), localidad Chapinero; 052-La Flora (56,83 %) y 061-Ciudad Usme (55,18 %), ambas en la localidad Usme. Ahora bien, se obtuvieron dos estimaciones de valor negativo, en las UPZ 088-El Refugio (localidad

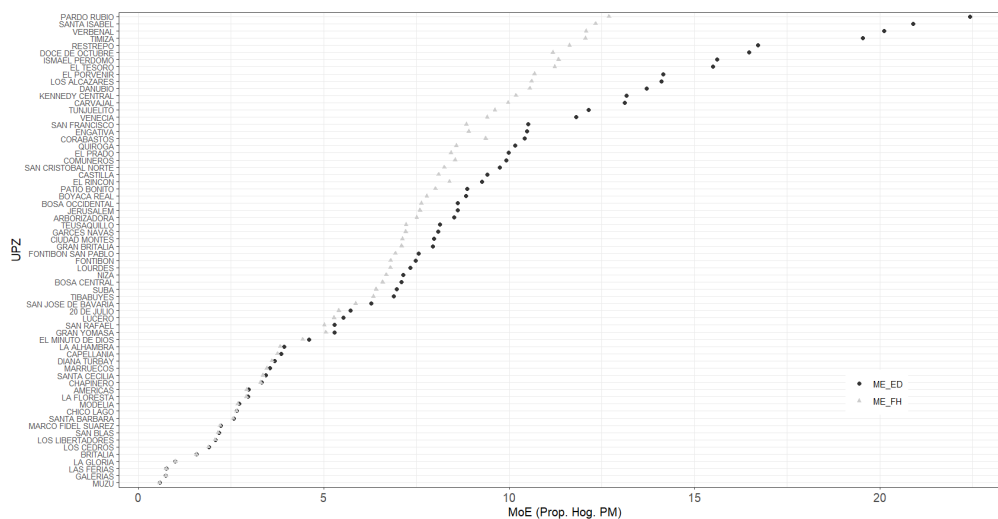


Figura 4-12.: Errores marginales ED y EBLUP

Chapinero) y 110-Ciudad Salitre Occidental (localidad Fontibón). De acuerdo con Corral et al.(2022), cuando no se aplican transformaciones en el modelo FH, estimaciones de tasas de pobreza podrían ser negativas para algunas áreas, por lo que se recomienda no usarlas para efectos de resultados. De esta manera, se extraen estas dos estimaciones de los resultados finales del EBLUP.

4.5. Modelo FH espacial

A partir de la matriz de pesos definida por criterio de contigüidad tipo reina, se aplicaron dos estadísticos espaciales globales, el Índice de Moran (valor 0,6363) y el Coeficiente de Geary (valor 0,3734), cuyas pruebas rechazan con suficiente significancia estadística (p-valores iguales a cero) la hipótesis nula de distribución aleatoria de la pobreza monetaria de los hogares en las UPZ, sustentando la aplicación del modelo FH espacial, el cual se ajusta con las mismas covariables del FH estándar seleccionado.

El ajuste del modelo Fay-Herriot Espacial empleando el método de máxima verosimilitud para la estimación de la varianza de los efectos aleatorios, no obtuvo mejores resultados en torno a las medidas de diagnóstico, en comparación con el modelo FH estándar, como se evidencia en los criterios de información AIC (-128,51 vs -130,51) y BIC (-108,8 vs -112,99), aunque el R^2 ajustado se mantiene igual (0,746), y se cumplen los supuestos de normalidad en los efectos aleatorios (p-valor prueba Shapiro-Wilk igual a 0,109) y homocedasticidad (p-valor prueba Breusch-Pagan igual a 0,606). Adicionalmente, los errores marginales de las estimaciones son superiores a los obtenidos mediante EBLUP en todas las 66 UPZ de la muestra.

Una siguiente aproximación del modelo se llevó a cabo aplicando máxima verosimilitud restringida para la estimación de la varianza, y el método de estimación *bootstrap* no paramétrico con corrección de sesgo para modelos espaciales Fay-Herriot para el MSE. Nuevamente, los criterios de información no evidencian mejoras en el ajuste respecto al modelo estándar (AIC -128,17 y BIC -106,28); además, no se cumple el supuesto de normalidad en los efectos aleatorios ya que se rechaza marginalmente la hipótesis nula (p-valor prueba Shapiro-Wilk igual a 0,031), aunque sí se presenta la homocedasticidad (p-valor prueba Breusch-Pagan igual a 0,465). Aunque las estimaciones directas están contenidas dentro de los intervalos de confianza del SEBLUP, el margen de error se reduce respecto al obtenido en el EBLUP en solo 12 de las 66 UPZ.

Por lo anterior, se ejecutó un modelo robusto Fay-Herriot con su extensión espacial, a partir de las covariables empleadas en el FH estándar, incorporando la matriz de pesos definida previamente. Para la estimación del MSE se definió su aproximación por pseudo-linearización, mientras que para la varianza se definió por máxima verosimilitud robusta con corrección de sesgo y 100 repeticiones *bootstrap*. Con base en lo anterior, se define formalmente el modelo robusto espacial Fay-Herriot con los coeficientes definidos para cada covariable.

$$\widehat{PobMon}_{UPZ} = 0,27364 - 0,09240(V4MedValt) + 0,04853(V6PrJInf) - 0,02807(V9PrColeg) - 0,03121(V12DenArb) + 0,03125(V18dDtCC) + 0,04229(V19DensUE) \quad (4-2)$$

El resumen del modelo aplicado, presentando los componentes de varianza, estadísticos de los efectos aleatorios y residuales, así como las soluciones a las ecuaciones de estimación robusta, se presenta en la figura 4-13.

```

Variance Components:
correlation      variance
-0.075165       0.003585

                Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
Random Effects  -2.005e-01  -3.264e-02  2.893e-03  1.457e-04  3.312e-02  1.797e-01
Residuals      -5.822e-02  -8.432e-03  -2.508e-05  7.292e-03  1.199e-02  1.502e-01

## Solutions to the Robust Estimation Equations:
(Intercept)    V4_MedValT    V6_PrJInf    V9_PrColeg    V12_DenArb    V18d_DtCC    V19_DensUE
-1.236e-03     2.375e-04     3.567e-04     3.914e-04     -6.296e-04     5.494e-04     5.279e-04
correlation    variance
-2.146e-11     -6.500e-02

Random Effects:
                Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
-1.164e-06     0.000e+00    0.000e+00    7.242e-05    1.340e-07    1.552e-03

## Iterations:
Allowed: 100 (10) -- 100
model parameter  random effects
                21                8

```

Figura 4-13.: Resumen estadístico modelo FH robusto espacial

Los gráficos *QQ-plot* para residuales estandarizados y efectos aleatorios se presentan en las figuras 4-14 y 4-15.

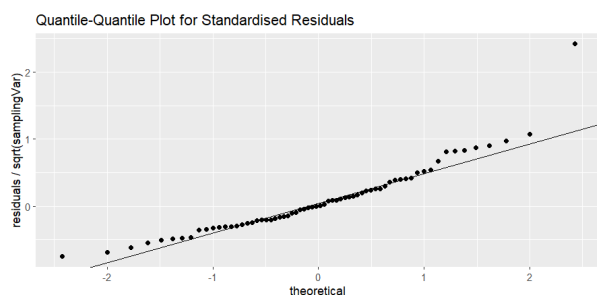


Figura 4-14.: QQ-plot residuales estandarizados

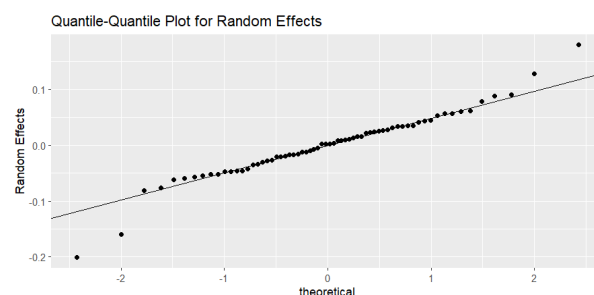


Figura 4-15.: QQ-plot efectos aleatorios

Los efectos aleatorios tienden a ajustarse en mayor medida sobre la recta de regresión, con ligera desviación en las colas, en comparación con los residuales estandarizados. Ahora bien, la prueba de normalidad Shapiro-Wilk en los efectos aleatorios se rechaza marginalmente (p -valor = 0,0129), aunque la prueba Breusch-Pagan valida la hipótesis de homocedasticidad (p -valor = 0,431).

4.6. Estimación SFH y comparación con FH

En las pruebas de diagnóstico de cobertura, no hay suficiente evidencia estadística para rechazar las hipótesis iniciales, por lo cual las estimaciones del modelo robusto espacial Fay-Herriot (RSEBLUP) no difieren significativamente de las estimaciones directas, y los intervalos de confianza se superponen en al menos el 95 %, lo que se evidencia en la figura 4-16.

En la tabla del Anexo D se presentan los resultados del EBLUP; RSEBLUP; *gamma* (GAM); los márgenes de error de EBLUP (ME FH) y del RSEBLUP (ME RS); con su tamaño de muestra (n). El valor *gamma* refleja que en 55 UPZ hay mayor ponderación para la estimación directa, mientras que en 11 habría mayor peso para la estimación basada en el modelo. Las tres UPZ donde se presentan las incidencias de pobreza monetaria en los hogares por encima del 50 % son: 053-Marco Fidel Suárez; 068-El Tesoro y 080-Corabastos. Asimismo, en 64 UPZ el error marginal del RSEBLUP es menor al obtenido en el EBLUP, por lo que en solo dos UPZ esta estimación tendría mejor precisión: 062-Tunjuelito y 086-El Porvenir. En la figura 4-17 se refleja la disminución del error marginal del RSEBLUP respecto al EBLUP a medida que estos aumentan, salvo las dos excepciones referidas previamente. Este error es superior al 10 % en ocho UPZ: 086-El Porvenir; 062-Tunjuelito; 090-Pardo Rubio; 037-Santa Isabel; 048-Timiza; 009-Verbenal; 038-Restrepo; 069-Ismael Perdomo.

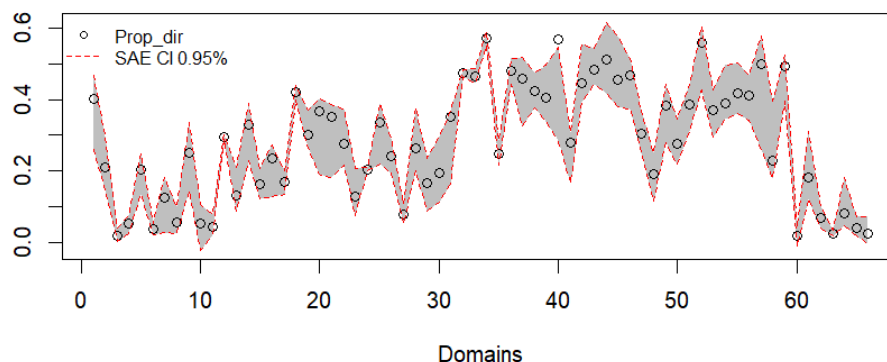


Figura 4-16.: Intervalos de confianza estimaciones RSFH con ED

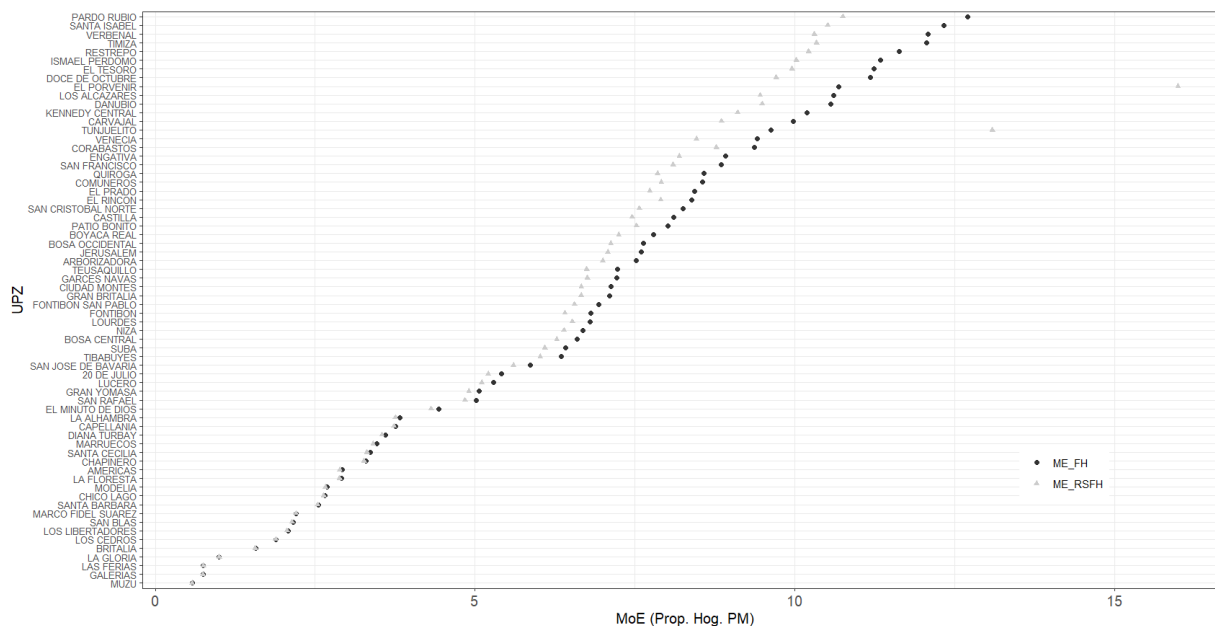


Figura 4-17.: Margen de error EBLUP y RSEBLUP

De manera similar a lo realizado en las predicciones sobre UPZ con escasa muestra o no muestreadas en el modelo anterior, se validó su precisión por medio de los dos escenarios de análisis de sensibilidad. Al comparar estos resultados con las predicciones originales del RSEBLUP, se evidenció lo siguiente: i) al aplicar variación del 10 % superior e inferior de forma intercalada sobre las seis covariables del modelo, la media de las diferencias absolutas es del 1,04 %; ii) aumentando en un 20 % la varianza de la estimación directa, la media de estas diferencias es del 0,1 %. Por lo tanto, las predicciones de la incidencia de la pobreza en los hogares por RSEBLUP para estas UPZ, presentan precisiones adecuadas.

De estas predicciones se destacan tres que presentan valores de incidencia de pobreza monetaria en los hogares por encima del 50 %: 089-San Isidro Patios (57,19 %); 052-La Flora (56,1 %); y 061-Ciudad Usme (54,95 %). De otro lado, se mantiene un RSEBLUP de valor negativo en la UPZ 110-Ciudad Salitre Occidental, por lo que se extraerá de los resultados finales. De esta manera, las predicciones completas se presentan en la tabla del Anexo E.

4.7. Mapas estimaciones FH y SFH

Con el fin de analizar la distribución espacial de la incidencia de la pobreza monetaria en las UPZ, obtenida por EBLUP (figura 4-18) y RSEBLUP (figura 4-19), se presentan los mapas temáticos clasificados por rangos a través de rompimientos naturales, los cuales destacan los valores extremos ante distribuciones con presencia de asimetrías.

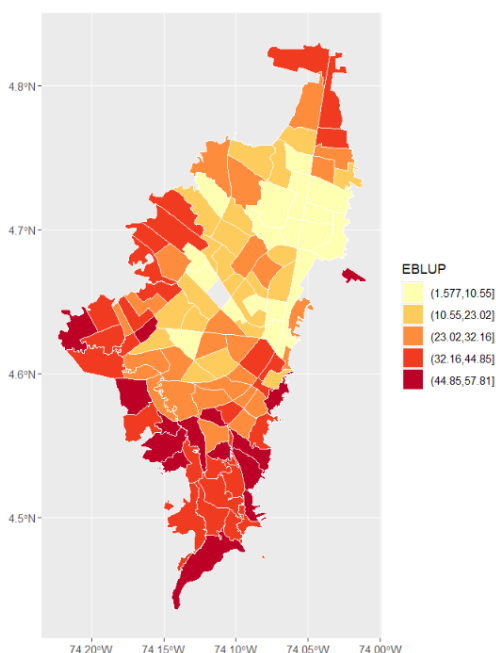


Figura 4-18.: Distribución EBLUP

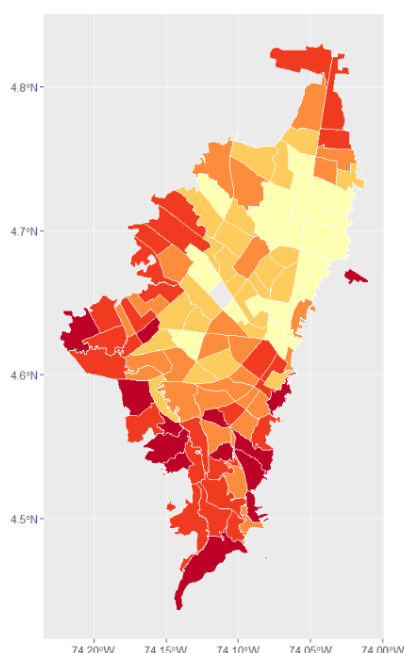


Figura 4-19.: Distribución RSEBLUP

Ambos mapas reflejan tendencias espaciales en la pobreza monetaria de los hogares, donde las UPZ ubicadas hacia el nororiente tienen una menor incidencia, mientras que hacia el sur, suroriente y suroccidente, la prevalencia es hacia mayores porcentajes. Este comportamiento diferencial se ha manifestado en otras representaciones de indicadores socioeconómicos ¹,

¹Tendencia espacial de las clases sociales (<https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-monetaria/pobreza-monetaria-2021>); pobreza multidimensional y transferencias monetarias (<https://razonpublica.com/la-pobreza-colombia-despues-la-covid-19/>)

ante lo cual se puede destacar que los resultados de las estimaciones son consistentes con la distribución espacial de este tipo de fenómenos en la ciudad de Bogotá.

Asimismo, la UPZ que fue descartada por valor negativo en sus predicciones, 110-Ciudad Salitre Occidental, no presenta color de alguno de los rangos. Aunque las diferencias entre los mapas son mínimas, en la distribución del RSEBLUP se presenta la UPZ 088-El Refugio al nororiente de la ciudad (posición aproximada 74.05°W entre 4.6° y 4.7°N), unidad que no presenta valor en el mapa del EBLUP.

Por último, se presentan los mapas de los indicadores locales de asociación espacial derivados de los EBLUP (figura 4-20) y los RSEBLUP (figura 4-21).

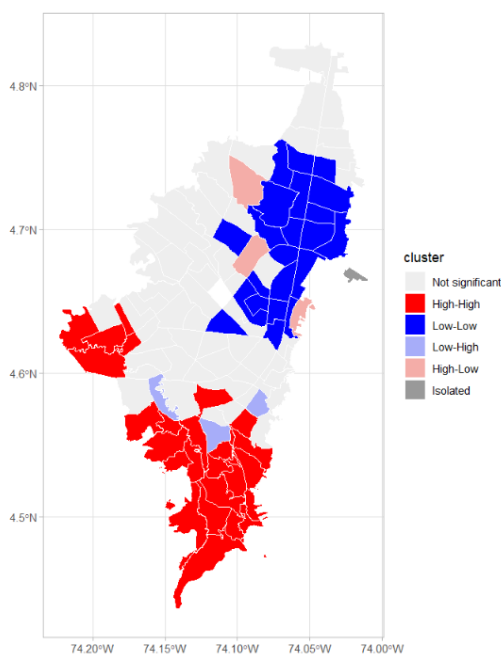


Figura 4-20.: LISA EBLUP

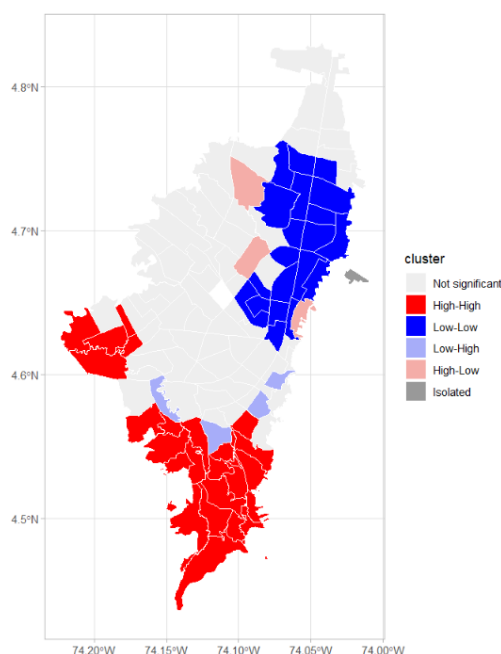


Figura 4-21.: LISA RSEBLUP

En estos mapas, la conformación de las agrupaciones espaciales estadísticamente significativas es prácticamente la misma, con escasas diferencias, particularmente en el RSEBLUP, donde los *clusters* de valores altos (UPZ de color rojo) y bajos (UPZ de color azul) de la incidencia de la pobreza monetaria en los hogares, tienden a una mayor consistencia por contigüidad, ya que en el EBLUP se adicionan algunas UPZ a estas agrupaciones, así: en las concentraciones de pobreza monetaria baja, las UPZ 030-Boyacá Real (localidad Engativá) y 111-Puente Aranda (localidad del mismo nombre); en las agrupaciones de pobreza monetaria alta, la UPZ 039-Quiroga (localidad Rafael Uribe Uribe).

Vale la pena señalar los atípicos espaciales alto-bajo (color rosa), es decir, UPZ con valores altos de pobreza monetaria en los hogares rodeadas de otras con valores bajos. En ambos casos son las mismas unidades, a saber: 028-El Rincón (localidad Suba); 026-Las Ferias (localidad Engativá); y 090-Pardo Rubio (localidad Chapinero). De otro lado, los atípicos bajo-alto (color lila), donde unidades con baja incidencia de pobreza monetaria en los hogares están rodeadas de otras con valores altos, se reflejan en las siguientes UPZ para ambos casos: 065-Arborizadora (localidad Ciudad Bolívar); 054-Marruecos (localidad Rafael Uribe Uribe); 033-Sosiego (localidad San Cristóbal). En el caso del RSEBLUP, se adiciona un atípico bajo-alto, correspondiente a la UPZ 094-La Candelaria (localidad del mismo nombre). Otro aspecto a mencionar es la UPZ que se considera aislada (color gris), ya que no presenta contigüidad con otras, siendo esta la 089-San Isidro Patios. En el resto de UPZ, el comportamiento espacial de la pobreza monetaria en los hogares sería aleatorio.

5. Conclusiones y recomendaciones

5.1. Conclusiones

En este trabajo se desarrolló una metodología enfocada en la estimación de áreas pequeñas (SAE) con enfoque espacial, para lograr estimaciones de la pobreza monetaria en los hogares a un nivel mayor de desagregación territorial, por unidades de planeamiento zonal (UPZ) en Bogotá, ya que la fuente para la estimación de este fenómeno, la Gran Encuesta Integrada de Hogares (GEIH), puede ofrecer resultados con representatividad en grandes ciudades en su conjunto. El enfoque planteado se sustenta en la necesidad de contar con estadísticas a un mayor nivel de desagregación, información cada vez más apetecida por tomadores de decisiones, académicos y usuarios en general, sumado a la importancia de analizar fenómenos complejos como la pobreza, a una mayor escala, para lograr una mejor comprensión y análisis de su distribución y formas de agrupación en el territorio.

La ejecución de procesos, soportada en planteamientos metodológicos de investigación cuantitativa y geoespacial, se desarrolló en una serie de etapas, iniciando por la obtención de estimaciones directas de la pobreza monetaria en las UPZ, tomando como base los microdatos de la pobreza monetaria vigencia 2021, junto a las variables del diseño muestral. Esto permitió obtener resultados iniciales para 66 de las 112 UPZ de la ciudad. La incorporación de información auxiliar a partir del uso de datos geoespaciales, derivó en un conjunto de 25 covariables para la obtención de estimaciones mediante modelos, con cobertura para cada UPZ y temporalidad igual o próxima a la vigencia de los datos de pobreza monetaria. Es importante resaltar que los datos auxiliares permitieron capturar características relevantes de cada UPZ, asociadas a la pobreza monetaria, ayudando a reflejar la heterogeneidad espacial de este tipo de fenómenos socioeconómicos en Bogotá.

La construcción rigurosa de los modelos a nivel de área Fay-Herriot, a través de la aplicación de tres técnicas diferentes (*stepwise KICb2*; penalización LASSO y modelado GLS), sumado a la evaluación de los diferentes diagnósticos, se convirtió en un paso esencial para la selección de los modelos más adecuados, que lograrían estimaciones de mejor precisión, así como el cumplimiento de los supuestos distribucionales requeridos en este tipo de modelos. Por lo tanto, se destaca la aplicación de diferentes métodos de selección para la obtención del modelo con mejor ajuste, que a su vez resultó en estimaciones de la pobreza monetaria con menor incertidumbre, en comparación con las obtenidas directamente de la encuesta.

Otro aspecto a destacar como ventaja de la estimación a través de modelos, concierne a las predicciones en las áreas con escasa o nula muestra, logrando completar, para el caso de este estudio, el mapa de la incidencia de la pobreza monetaria en los hogares por UPZ para Bogotá. Adicionalmente, el uso de estadísticas espaciales para soportar la extensión del modelo Fay-Herriot, validó la necesidad de incorporar la componente espacial en el modelado, con el uso de la matriz de pesos por contiguidad de UPZ. En este caso, se realizaron diferentes pruebas y análisis con el fin de validar que esta extensión del modelo en realidad fuese un avance respecto a la precisión de los EBLUP, lo que se logró al incorporar soluciones robustas definidas para los modelos Fay-Herriot, obtener los RSEBLUP y lograr la estimación del error medio cuadrático a través de métodos de pseudo-linearización, recomendados actualmente para la aplicación de enfoques SAE en contextos más reales, donde los fenómenos presentan una mayor heterogeneidad y variabilidad asociada a los efectos aleatorios de las áreas donde se presentan.

Se espera que estos resultados no solo sirvan de referencia para la comprensión del fenómeno de la pobreza monetaria en los hogares a escala intraurbana en la ciudad de Bogotá, sino que fomenten el uso de datos y modelos con enfoque geoespacial para la obtención de estadísticas socioeconómicas desagregadas territorialmente, con suficiente precisión, como insumo fundamental para la elaboración de políticas públicas focalizadas. De esta manera, los datos referidos en el espacio, no solo dan contexto a la interpretación de las estadísticas en general, sino que se convierten en potenciales factores explicativos de los fenómenos socioeconómicos, como la pobreza.

5.2. Recomendaciones

Este trabajo es tan solo una exploración de las potencialidades que ofrecen los datos y modelos con enfoque geoespacial para la obtención de información relevante en el ámbito socioeconómico. Ahora bien, sus resultados están acotados al área y unidades de análisis, es decir, las Unidades de Planeamiento Zonal de Bogotá ya que, en análisis espacial, los resultados son sensibles a los niveles de agregación y combinaciones de unidades que se realicen, ya sea por la escala o por las diferentes formas de agrupación espacial que se pueden generar (Sáenz, 2016). Por lo tanto, los resultados obtenidos deben ser comprendidos e interpretados exclusivamente en la escala en la cual fueron analizados y modelados.

El uso de información auxiliar logra mejoras en la precisión de las estimaciones basadas en modelos y, sobretudo, permite obtener predicciones en las áreas donde la encuesta original no seleccionó muestra suficiente para la inferencia de resultados, ya que se toma fortaleza prestada de estos datos asociados a las áreas de interés. Ahora bien, se debe resaltar que

el uso de covariables puede tener limitaciones debido a posibles causas, ya sea, la fuente de estos datos, o por las concepciones originales de los modelos usados. Por consiguiente, se pueden presentar barreras asociadas a determinar la dependencia requerida con las covariables adecuadas, más la sensibilidad a estructuras de correlación espacial definidas tanto de los datos como de los modelos, que posiblemente no logren capturar completamente la variabilidad en todas las áreas analizadas.

Aunque los modelos a nivel de área Fay-Herriot son de los más empleados en SAE, y se han desarrollado distintas extensiones para potencializar su uso, tienen limitaciones en cuanto a la necesidad de cumplir con los supuestos distribucionales, más la estimación de varianzas y del MSE, que en algunos casos puede ser negativa si la variable de interés, en la muestra de base, tiene respuestas únicas o de poca variabilidad en las áreas a estimar. De acá que en la actualidad el uso de modelos bayesianos y de aprendizaje automático o *Machine Learning* se contemplan como enfoques avanzados que pueden generar predicciones de mayor precisión con suficientes conjuntos de datos para entrenar y probar los modelos (Merfeld et al., 2023). También son importantes las aproximaciones basadas en estimar de arriba a abajo (*top-down*) o abajo a arriba (*bottom-up*) con los cuales se busca desagregar o completar información estadística, principalmente demográfica, en zonas donde la escasez de datos es preponderante (WorldPop, 2024).

Al margen de lo anterior, se hace necesario que las Oficinas Nacionales y Locales de Estadística y, en general, todas las entidades productoras de información dentro de un Sistema Estadístico Nacional, se adhieran a la producción de datos con mayor desagregación temática o territorial, ya que son el insumo clave para la toma de decisiones y la acción en política pública en el ámbito local. En este caso se sugiere que, a la par de la generación de estadísticas convencionales a niveles de agregación superior, se opte por enfoques basados en modelos para áreas más detalladas. Más aún, se recomienda considerar la adopción de estimaciones basadas en modelos con componentes espaciales para los estudios de variables clave en las ciudades con alta heterogeneidad socioeconómica, ya que esto permite determinar de forma más precisa las variaciones y patrones espaciales que estos fenómenos presentan a escalas intraurbanas, brindando mayor soporte para la toma de decisiones informada y la definición de políticas públicas focalizadas por parte de los gobiernos locales.

Por supuesto, todo lo anterior debe fortalecerse a partir de establecer un sistema de monitoreo periódico que permita actualizar las estimaciones de la incidencia de la pobreza a nivel de unidades intraurbanas de planeación socioterritorial, en sintonía con la producción anual de estadísticas oficiales a mayor agregación. Con esto se busca asegurar que las políticas y programas sociales se ajusten y focalicen según las tendencias actualizadas y manifiestas del fenómeno de la pobreza, en cumplimiento a las metas establecidas para los Objetivos de Desarrollo Sostenible. Adicionalmente, esta producción debe contar con la orientación reque-

rida para que todos los usuarios de información comprendan la disposición de estadísticas a través de diferentes métodos de estimación (Tzavidis et al., 2018).

A. Anexo: Exploración de datos - covariables información auxiliar

La figura A-1 presenta el resumen numérico y distribucional de las 25 covariables que conforman la información auxiliar.

skim_type	skim_variable	n_missing	complete_rate	numeric.mean	numeric.sd	numeric.p0	numeric.p25	numeric.p50	numeric.p75	numeric.p100	numeric.hist
numeric	Prop_dir	0	1	2.810768e-01	1.628317e-01	1.826216e-02	1.637497e-01	2.783901e-01	4.166562e-01	5.732417e-01	
numeric	V1_DensPob	0	1	2.387637e+02	8.874290e+01	7.034625e+01	1.837064e+02	2.268277e+02	2.880234e+02	4.647962e+02	
numeric	V2_PrDelit	0	1	1.142568e-02	5.680536e-03	4.098911e-03	7.211359e-03	1.061499e-02	1.479603e-02	3.090982e-02	
numeric	V3_PriPS	0	1	1.118413e-02	1.864298e-02	3.459011e-04	2.507783e-03	4.842615e-03	1.167416e-02	1.259080e-01	
numeric	V4_MedValT	0	1	2.085955e+06	1.015037e+06	5.800000e+05	1.400000e+06	2.090000e+06	2.692500e+06	6.500000e+06	
numeric	V5_ComFam	0	1	3.636364e-01	4.847319e-01	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00	
numeric	V6_PrJInf	0	1	1.312544e-02	1.188493e-02	0.000000e+00	2.906977e-03	1.162791e-02	1.744186e-02	4.651163e-02	
numeric	V7_PrPens	0	1	1.258937e-02	7.955433e-03	1.217532e-03	6.899351e-03	1.082251e-02	1.626759e-02	3.409091e-02	
numeric	V8_PrVict	0	1	1.273654e-02	1.250847e-02	5.048917e-04	4.214179e-03	7.675783e-03	1.576358e-02	4.836005e-02	
numeric	V9_PrColeg	0	1	1.270015e-02	7.563757e-03	4.302926e-04	8.175559e-03	1.161790e-02	1.753442e-02	4.388985e-02	
numeric	V10_TRepH	0	1	5.141171e+00	3.733231e+00	0.000000e+00	1.883420e+00	5.272723e+00	7.678877e+00	1.484594e+01	
numeric	V10_TRepM	0	1	3.988727e+00	2.977159e+00	0.000000e+00	1.536027e+00	4.314074e+00	5.853075e+00	1.264697e+01	
numeric	V11a_TDesH	0	1	5.109389e-01	4.239899e-01	0.000000e+00	2.269028e-01	4.091269e-01	6.467592e-01	2.016129e+00	
numeric	V11b_TDesM	0	1	3.928245e-01	3.231175e-01	0.000000e+00	1.464046e-01	2.963240e-01	5.878584e-01	1.260365e+00	
numeric	V12_DenArb	0	1	3.373203e+03	2.323727e+03	5.305556e+02	1.857040e+03	2.928103e+03	4.373979e+03	1.227912e+04	
numeric	V13_PrOcup	0	1	1.078327e-02	2.896872e-02	0.000000e+00	0.000000e+00	2.904203e-04	1.378459e-02	2.154504e-01	
numeric	V14_PrParq	0	1	7.627581e-02	3.847622e-02	5.761124e-03	5.434550e-02	6.870960e-02	9.508217e-02	2.060079e-01	
numeric	V15_ClusEc	0	1	6.818182e-01	4.693397e-01	0.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	
numeric	V16_PtArCl	0	1	1.171600e-02	1.356560e-02	0.000000e+00	1.937984e-03	7.751938e-03	1.550388e-02	8.139535e-02	
numeric	V17_ARiesg	0	1	3.939394e-01	4.923660e-01	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00	
numeric	V18a DtUnv	0	1	1.870917e+00	1.328780e+00	1.219670e-01	9.626088e-01	1.566801e+00	2.461679e+00	5.137613e+00	
numeric	V18b DtSal	0	1	1.263390e+00	7.337228e-01	1.675334e-01	7.504924e-01	1.110911e+00	1.761777e+00	3.281313e+00	
numeric	V18c DtSeg	0	1	7.458823e-01	5.018733e-01	7.498491e-02	3.886359e-01	6.572345e-01	9.957905e-01	2.703116e+00	
numeric	V18d DtCC	0	1	1.151208e+00	7.470894e-01	3.892876e-02	7.059718e-01	9.795882e-01	1.399933e+00	3.657770e+00	
numeric	V19_DensUE	0	1	1.336669e+03	8.158788e+02	3.535103e+02	8.145084e+02	1.093751e+03	1.608890e+03	4.323200e+03	
numeric	V20_Estrat	0	1	2.878788e-01	4.562439e-01	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	1.000000e+00	

Figura A-1.: Exploración de datos - información auxiliar

B. Anexo: Estimación directa y Fay-Herriot estándar

Tabla B-1.: Comparación estimación directa vs EBLUP

UPZ	NOMBRE	n	ED	EBLUP	GAMMA	ME_ED	ME_FH
009	VERBENAL	109	40.298	37.296	31.454	20.123	12.084
011	S. CRISTOBAL NORTE	148	21.020	21.771	66.169	9.747	8.257
013	LOS CEDROS	147	1.884	1.988	98.087	1.904	1.892
016	SANTA BARBARA	63	5.209	4.966	96.543	2.579	2.556
017	SAN JOSE DE BAVARIA	84	20.332	19.319	82.493	6.280	5.861
018	BRITALIA	119	3.576	3.637	98.681	1.576	1.569
019	EL PRADO	103	12.539	10.548	65.047	9.992	8.435
020	LA ALHAMBRA	66	5.587	6.105	92.296	3.938	3.829
022	DOCE DE OCTUBRE	169	25.108	24.072	40.627	16.479	11.183
024	NIZA	63	5.209	4.555	78.466	7.141	6.687
025	LA FLORESTA	87	4.432	4.780	95.492	2.962	2.917
026	LAS FERIAS	98	29.506	29.478	99.695	0.754	0.753
027	SUBA	134	13.002	14.557	79.274	6.970	6.418
028	EL RINCON	346	32.830	31.287	68.360	9.274	8.392
029	EL MINUTO DE DIOS	205	16.302	16.382	89.723	4.613	4.434
030	BOYACA REAL	151	23.628	20.692	70.402	8.838	7.785
031	SANTA CECILIA	102	16.917	16.799	94.019	3.438	3.363
032	SAN BLAS	249	42.109	42.021	97.522	2.173	2.154
034	20 DE JULIO	190	29.983	31.501	85.039	5.717	5.409
037	SANTA ISABEL	87	36.825	29.971	29.829	20.907	12.335
038	RESTREPO	148	35.314	28.880	39.939	16.716	11.637
039	QUIROGA	156	27.740	29.336	64.246	10.169	8.580
040	CIUDAD MONTES	187	12.881	13.858	74.477	7.980	7.128
041	MUZU	84	20.496	20.495	99.820	0.579	0.578
042	VENECIA	207	33.491	31.026	57.088	11.818	9.416
043	SAN RAFAEL	126	24.260	24.201	86.890	5.295	5.021
044	AMERICAS	82	7.668	7.921	95.467	2.970	2.922
045	CARVAJAL	148	26.348	28.434	51.924	13.117	9.973

046	CASTILLA	182	16.595	16.105	67.693	9.417	8.103
047	KENNEDY CENTRAL	144	19.503	20.628	51.719	13.170	10.190
048	TIMIZA	181	35.234	27.794	32.709	19.551	12.065
050	LA GLORIA	82	47.354	47.375	99.462	1.002	1.001
051	LOS LIBERTADORES	134	46.644	46.643	97.708	2.088	2.073
053	MARCO FIDEL SUAREZ	75	57.324	56.850	97.404	2.225	2.205
054	MARRUECOS	110	24.650	25.007	93.650	3.549	3.468
055	DIANA TURBAY	187	48.022	48.008	93.196	3.683	3.604
056	DANUBIO	80	45.734	43.021	49.711	13.710	10.560
057	GRAN YOMASA	84	42.437	42.591	86.908	5.291	5.060
058	COMUNEROS	124	40.663	42.156	65.379	9.920	8.550
062	TUNJUELITO	89	56.777	46.976	55.734	12.148	9.625
065	ARBORIZADORA	82	27.938	24.203	71.922	8.517	7.516
066	SAN FRANCISCO	201	44.483	47.258	62.716	10.510	8.853
067	LUCERO	113	48.286	49.133	85.832	5.538	5.284
068	EL TESORO	123	51.285	51.438	43.595	15.505	11.238
069	ISMAEL PERDOMO	206	45.495	48.475	43.255	15.613	11.336
070	JERUSALEM	182	46.784	44.848	71.418	8.623	7.600
071	TIBABUYES	337	30.510	31.123	79.639	6.892	6.343
073	GARCES NAVAS	294	19.037	18.659	73.973	8.086	7.216
074	ENGATIVA	188	38.435	36.567	62.851	10.480	8.917
075	FONTIBON	220	27.566	27.989	76.826	7.487	6.806
076	FONTIBON SAN PABLO	80	38.631	37.709	76.491	7.557	6.936
080	CORABASTOS	161	55.836	51.975	63.111	10.422	9.372
081	GRAN BRITALIA	164	37.209	36.355	74.653	7.943	7.106
082	PATIO BONITO	179	38.932	41.745	70.278	8.865	8.017
084	BOSA OCCIDENTAL	169	41.850	43.282	71.413	8.625	7.636
085	BOSA CENTRAL	299	41.113	40.877	78.640	7.104	6.594
086	EL PORVENIR	127	49.811	42.820	48.110	14.157	10.686
090	PARDO RUBIO	75	22.884	28.547	26.964	22.435	12.697
096	LOURDES	144	49.339	46.132	77.508	7.343	6.803
097	CHICO LAGO	70	1.826	1.577	96.324	2.663	2.651
098	LOS ALCAZARES	99	18.296	20.795	48.286	14.107	10.606
099	CHAPINERO	70	6.842	7.123	94.371	3.329	3.300
100	GALERIAS	67	2.529	2.576	99.701	0.747	0.746
101	TEUSAQUILLO	74	8.220	10.453	73.750	8.132	7.225
114	MODELIA	73	4.055	4.252	96.160	2.724	2.688
115	CAPELLANIA	64	2.484	4.024	92.614	3.849	3.753

C. Anexo: Predicciones EBLUP en UPZ áreas pequeñas

Tabla C-1.: Predicciones EBLUP en UPZ áreas pequeñas

UPZ	NOMBRE	EBLUP
001	PASEO DE LOS LIBERTADORES	37.068
002	LA ACADEMIA	30.673
003	GUAYMARAL	38.309
010	LA URIBE	25.425
012	TOBERIN	27.422
014	USAQUEN	5.555
015	COUNTRY CLUB	6.963
021	LOS ANDES	21.323
023	CASA BLANCA SUBA	18.588
033	SOSIEGO	25.683
035	CIUDAD JARDIN	28.835
036	SAN JOSE	38.002
049	APOGEO	30.826
052	LA FLORA	56.836
059	ALFONSO LOPEZ	44.210
060	PARQUE ENTRENUBES	34.935
061	CIUDAD USME	55.186
063	EL MOCHUELO	38.834
064	MONTE BLANCO	42.449
072	BOLIVIA	7.093
077	ZONA FRANCA	36.081
078	TINTAL NORTE	37.445
079	CALANDAIMA	27.164
083	LAS MARGARITAS	28.870
087	TINTAL SUR	46.142
089	SAN ISIDRO-PATIOS	57.810
091	SAGRADO CORAZON	7.841
092	LA MACARENA	28.070

093	LAS NIEVES	40.798
094	LA CANDELARIA	21.865
095	LAS CRUCES	41.053
102	LA SABANA	37.701
103	PARQUE EL SALITRE	20.950
104	PARQUE SIMON BOLIVAR-CAN	16.094
105	JARDIN BOTANICO	23.022
106	LA ESMERALDA	9.591
107	QUINTA PAREDES	15.616
108	ZONA INDUSTRIAL	32.160
109	CIUDAD SALITRE ORIENTAL	9.033
111	PUENTE ARANDA	25.502
112	GRANJAS DE TECHO	21.588
113	BAVARIA	20.129
116	ALAMOS	19.932
117	AEROPUERTO EL DORADO	37.036

D. Anexo: Estimaciones Fay-Herriot estándar y robusta espacial

Tabla D-1.: Comparación EBLUP vs RSEBLUP

UPZ	NOMBRE	n	EBLUP	RSEBLUP	GAM	ME FH	ME RS
009	VERBENAL	109	37.30	36.37	31.45	12.08	10.31
011	SAN CRISTOBAL NORTE	148	21.77	22.25	66.17	8.26	7.57
013	LOS CEDROS	147	1.99	2.03	98.09	1.89	1.88
016	SANTA BARBARA	63	4.97	4.92	96.54	2.56	2.54
017	SAN JOSE DE BAVARIA	84	19.32	19.12	82.49	5.86	5.60
018	BRITALIA	119	3.64	3.66	98.68	1.57	1.56
019	EL PRADO	103	10.55	10.39	65.05	8.43	7.73
020	LA ALHAMBRA	66	6.11	6.23	92.30	3.83	3.75
022	DOCE DE OCTUBRE	169	24.07	23.95	40.63	11.18	9.71
024	NIZA	63	4.55	4.10	78.47	6.69	6.39
025	LA FLORESTA	87	4.78	4.90	95.49	2.92	2.88
026	LAS FERIAS	98	29.48	29.47	99.70	0.75	0.75
027	SUBA	134	14.56	14.78	79.27	6.42	6.09
028	EL RINCON	346	31.29	30.89	68.36	8.39	7.90
029	EL MINUTO DE DIOS	205	16.38	16.37	89.72	4.43	4.31
030	BOYACA REAL	151	20.69	20.01	70.40	7.79	7.25
031	SANTA CECILIA	102	16.80	16.77	94.02	3.36	3.31
032	SAN BLAS	249	42.02	41.96	97.52	2.15	2.14
034	20 DE JULIO	190	31.50	31.80	85.04	5.41	5.21
037	SANTA ISABEL	87	29.97	29.69	29.83	12.34	10.51
038	RESTREPO	148	28.88	28.28	39.94	11.64	10.22
039	QUIROGA	156	29.34	29.49	64.25	8.58	7.85
040	CIUDAD MONTES	187	13.86	14.00	74.48	7.13	6.66
041	MUZU	84	20.50	20.49	99.82	0.58	0.58
042	VENECIA	207	31.03	30.40	57.09	9.42	8.46
043	SAN RAFAEL	126	24.20	24.27	86.89	5.02	4.84
044	AMERICAS	82	7.92	8.00	95.47	2.92	2.88
045	CARVAJAL	148	28.43	28.76	51.92	9.97	8.85

046	CASTILLA	182	16.10	16.04	67.69	8.10	7.46
047	KENNEDY CENTRAL	144	20.63	20.50	51.72	10.19	9.11
048	TIMIZA	181	27.79	26.51	32.71	12.06	10.34
050	LA GLORIA	82	47.37	47.38	99.46	1.00	1.00
051	LOS LIBERTADORES	134	46.64	46.61	97.71	2.07	2.06
053	MARCO FIDEL SUAREZ	75	56.85	57.05	97.40	2.21	2.20
054	MARRUECOS	110	25.01	25.04	93.65	3.47	3.41
055	DIANA TURBAY	187	48.01	47.88	93.20	3.60	3.54
056	DANUBIO	80	43.02	42.28	49.71	10.56	9.49
057	GRAN YOMASA	84	42.59	42.49	86.91	5.06	4.90
058	COMUNEROS	124	42.16	41.86	65.38	8.55	7.91
062	TUNJUELITO	89	46.98	41.76	55.73	9.62	13.09
065	ARBORIZADORA	82	24.20	23.70	71.92	7.52	7.00
066	SAN FRANCISCO	201	47.26	47.47	62.72	8.85	8.10
067	LUCERO	113	49.13	49.26	85.83	5.28	5.11
068	EL TESORO	123	51.44	51.65	43.60	11.24	9.96
069	ISMAEL PERDOMO	206	48.48	47.87	43.26	11.34	10.03
070	JERUSALEM	182	44.85	44.41	71.42	7.60	7.08
071	TIBABUYES	337	31.12	31.09	79.64	6.34	6.02
073	GARCES NAVAS	294	18.66	18.41	73.97	7.22	6.76
074	ENGATIVA	188	36.57	36.25	62.85	8.92	8.19
075	FONTIBON	220	27.99	28.24	76.83	6.81	6.40
076	FONTIBON SAN PABLO	80	37.71	37.63	76.49	6.94	6.55
080	CORABASTOS	161	51.97	51.46	63.11	9.37	8.77
081	GRAN BRITALIA	164	36.36	36.16	74.65	7.11	6.66
082	PATIO BONITO	179	41.75	42.07	70.28	8.02	7.52
084	BOSA OCCIDENTAL	169	43.28	43.29	71.41	7.64	7.13
085	BOSA CENTRAL	299	40.88	40.62	78.64	6.59	6.28
086	EL PORVENIR	127	42.82	41.98	48.11	10.69	15.99
090	PARDO RUBIO	75	28.55	28.71	26.96	12.70	10.75
096	LOURDES	144	46.13	46.01	77.51	6.80	6.52
097	CHICO LAGO	70	1.58	1.56	96.32	2.65	2.63
098	LOS ALCAZARES	99	20.80	21.68	48.29	10.61	9.46
099	CHAPINERO	70	7.12	7.18	94.37	3.30	3.26
100	GALERIAS	67	2.58	2.56	99.70	0.75	0.75
101	TEUSAQUILLO	74	10.45	11.28	73.75	7.23	6.75
114	MODELIA	73	4.25	4.34	96.16	2.69	2.66
115	CAPELLANIA	64	4.02	3.38	92.61	3.75	3.73

E. Anexo: Predicciones RSEBLUP en UPZ áreas pequeñas

Tabla E-1.: Predicciones RSEBLUP en UPZ áreas pequeñas

UPZ	NOMBRE	RSEBLUP
UPZ	NOMBRE	RSEBLUP
001	PASEO DE LOS LIBERTADORES	36.650
002	LA ACADEMIA	30.657
003	GUAYMARAL	38.103
010	LA URIBE	25.354
012	TOBERIN	26.883
014	USAQUEN	5.789
015	COUNTRY CLUB	7.463
021	LOS ANDES	21.396
023	CASA BLANCA SUBA	18.207
033	SOSIEGO	25.254
035	CIUDAD JARDIN	28.547
036	SAN JOSE	37.011
049	APOGEO	30.442
052	LA FLORA	56.045
059	ALFONSO LOPEZ	43.918
060	PARQUE ENTRENUBES	33.774
061	CIUDAD USME	54.919
063	EL MOCHUELO	38.750
064	MONTE BLANCO	42.531
072	BOLIVIA	6.627
077	ZONA FRANCA	35.884
078	TINTAL NORTE	37.305
079	CALANDAIMA	26.907
083	LAS MARGARITAS	28.586
087	TINTAL SUR	45.522
088	EL REFUGIO	0.432
089	SAN ISIDRO-PATIOS	57.157

091	SAGRADO CORAZON	5.973
092	LA MACARENA	27.946
093	LAS NIEVES	41.121
094	LA CANDELARIA	21.820
095	LAS CRUCES	40.871
102	LA SABANA	37.556
103	PARQUE EL SALITRE	19.754
104	PARQUE SIMON BOLIVAR-CAN	15.970
105	JARDIN BOTANICO	23.309
106	LA ESMERALDA	9.416
107	QUINTA PAREDES	15.807
108	ZONA INDUSTRIAL	32.262
109	CIUDAD SALITRE ORIENTAL	9.243
111	PUENTE ARANDA	25.679
112	GRANJAS DE TECHO	21.673
113	BAVARIA	20.255
116	ALAMOS	20.023
117	AEROPUERTO EL DORADO	37.208

Bibliografía

- Ardilly, P., Bouche, P., and Zhu, W. (2018). Small areas and spatial correlation. In *Handbook of Spatial Analysis - Theory and Application with R*, number 131 in Insee Méthodes, pages 305–324. Insee - Eurostat, France, institut national de la statistique et des Études Économiques edition. <https://www.insee.fr/en/information/3635545>.
- Artelaris, P. and Kandylis, G. (2014). Mapping poverty at regional level in Greece. *Région et Développement*, (39):131–147. https://www.researchgate.net/publication/271441877_Mapping_poverty_at_regional_level_in_Greece.
- Asfar, K., Anang, K., and Sadik, K. (2016). Optimum Spatial Weighted in Small Area Estimation. *Global Journal of Pure and Applied Mathematics*, 12(5):3977–3989. https://www.ripublication.com/gjpam16/gjpamv12n5_10.pdf.
- Asian Development Bank (2020). *Introduction To Small Area Estimation Techniques: A Practical Guide for National Statistics Offices*. Asian Development Bank, Manila, Filipinas. <https://www.adb.org/sites/default/files/publication/609476/small-area-estimation-guide-nsos.pdf>.
- Botello, S. and Valderrama, D. (n.d.). Algoritmo para la construcción de líneas de pobreza y pobreza extrema a partir de la Encuesta Nacional de Ingresos y Gastos – ENIG 2006/2007. https://microdatos.dane.gov.co/index.php/catalog/545/download/9327/Algoritmo_para_la_construccion_de_las_lineas_de_pobreza.pdf.
- Boubeta, M., Lombardía, M., and Morales, D. (2023). Small area prediction of proportions and counts under a spatial Poisson mixed model. *Statistical Methods & Applications*. doi: 10.1007/s10260-023-00729-7.
- Burgard, J., Morales, D., and Wolwer, A. (2022). Small area estimation of socioeconomic indicators for sampled and unsampled domains. *AStA Advances in Statistical Analysis*, 106:287–314. doi: 10.1007/s10182-021-00426-4.
- Caballero, E. (2015). Division Upz Bogota. <https://es.scribd.com/doc/267439976/DIVISION-UPZ-BOGOTA>.
- Castañeda, J., Téllez, C., and Fúquene, J. (2019). Una alternativa para la estimación del ingreso promedio mediante métodos de estimación en áreas pequeñas. *arXiv*, 1(1907.05387):1–19. <https://arxiv.org/pdf/1907.05387>.

- Chambers, R., Chandra, H., Salvati, N., and Tzavidis, N. (2009). Outlier Robust Small Area Estimation. Working Paper 16-09, Centre for Statistical & Survey Methodology. <https://ro.uow.edu.au/cssmwp/36>.
- Chambers, R., Chandra, H., Salvati, N., and Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society*, 76(1):47–69. https://ro.uow.edu.au/articles/journal_contribution/Outlier_robust_small_area_estimation/27790434?file=50560212.
- Chambers, R., Chandra, H., and Tzavidis, N. (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Methodology*, 37(2):153–170. <https://www150.statcan.gc.ca/n1/pub/12-001-x/2011002/article/11604-eng.pdf>.
- Chandra, H., Aditya, K., and Sud, U. C. (2018). Localised estimates and spatial mapping of poverty incidence in the state of Bihar in India - An application of small area estimation techniques. *PLoS ONE*, 13(6):1–14. doi: 10.1371/journal.pone.0198502.
- Chandra, H., Salvati, N., Chambers, R., and Tzavidis, N. (2010). Small Area Estimation under Spatial Nonstationarity. <https://ro.uow.edu.au/cssmwp/71/>.
- Comisión Estadística para América Latina y El Caribe - CEPAL (2021). Estimaciones subnacionales de la pobreza para América Latina. Technical Report 4, CEPAL. <https://repositorio.cepal.org/server/api/core/bitstreams/74483051-fd59-4953-9396-7b0e89c16b21/content>.
- Consejo Nacional de Política Económica y Social - CONPES (2012). Metodologías oficiales y arreglos institucionales para la medición de la pobreza en Colombia. <https://colaboracion.dnp.gov.co/cdt/conpes/social/150.pdf>.
- Corral, P. (2023). Dos décadas colocando la pobreza en el mapa: Estimación para áreas pequeñas en el Banco Mundial. <https://www.youtube.com/watch?v=hohZCYsyWk>.
- Corral, P., Molina, I., Cojocarú, A., and Segovia, S. (2022). Guidelines to Small Area Estimation for Poverty Mapping. <https://openknowledge.worldbank.org/server/api/core/bitstreams/1d1fcadc-43e3-541b-8949-fea45dd2a528/content>.
- Crescenzi, F., Betti, G., and Gagliardi, F. (2015). Comparing small area techniques for estimating poverty measures. *Quaderni del Dipartimento di Economia Politica e Statistica*, (721):1–25. <https://www.deps.unisi.it/sites/st02/files/allegatiparagrafo/16-12-2015/721.pdf>.
- Cámara de Comercio de Bogotá (n.d.). Unidades de Planificación Zonal / Definición. <https://recursos.ccb.org.co/ccb/pot/PC/files/3definicion.html>.

- de Smith, M., Goodchild, M., and Longley, P. (2024). *Geospatial Analysis - A comprehensive guide*. Number 2024-1. 7th edition. <https://spatialanalysisonline.com/>.
- Delatie, L. (2021). Poverty Mapping Using Small Area Estimation Techniques - The Jamaican Experience. <https://www.cepal.org/sites/default/files/presentations/poverty-mapping-using-small-area-estimation-jamaica-jul-2021.pdf>.
- Departamento Administrativo Nacional de Estadística - DANE (2021). Actualización metodológica para la construcción de las líneas de pobreza monetaria y pobreza monetaria extrema. https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2019/pobreza_monetaria_actualizacion_metodologica_2019-nuevas-lineas.pdf.
- Departamento Administrativo Nacional de Estadística - DANE (2023a). Medición de Pobreza Monetaria y Desigualdad 2021. <https://microdatos.dane.gov.co/index.php/catalog/733>.
- Departamento Administrativo Nacional de Estadística - DANE (2023b). Metodología General Gran Encuesta Integrada de Hogares GEIH. www.dane.gov.co/files/operaciones/GEIH/met-GEIH.pdf.
- Departamento Administrativo Nacional de Estadística - DANE (2024). Pobreza monetaria en Colombia Año 2023. Boletín Técnico, DANE, Bogotá. <https://www.dane.gov.co/files/operaciones/PM/bol-PM-2023.pdf>.
- Departamento para la Prosperidad Social - DPS (2021). Poverty Map - Colombia 2018 - 2019. <https://www.cepal.org/sites/default/files/presentations/poverty-mapping-prosperidad-social-colombia-jul-2021.pdf>.
- Edochie, I., Newhouse, D., Tzavidis, N., Schmid, T., Foster, E., Luna, A., Ouedraogo, A., Sanoh, A., and Savadogo, A. (2024). Small Area Estimation of Poverty in Four West African Countries by Integrating Survey and Geospatial Data. Policy Research Working Paper 10892, World Bank Group. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099158209042442519/idu170f4516f1b8a7146a619630191a7420b4631>.
- Foster, J., Greer, J., and Thorbecke, E. (2010). The Forster-Greer-Thorbecke (FGT) Poverty Measures: Twenty-Five Years Later. Working Paper IIEP-WP-2010-14, Institute for International Economic Policy Working Paper Series. https://www2.gwu.edu/~iiep/assets/docs/papers/Foster_IIEPWP2010-14.pdf.
- Franco, C. (2021). SAIPE: Poverty Mapping in the United States. <https://www.cepal.org/sites/default/files/presentations/saipe-poverty-mapping-united-states-jul-2021.pdf>.

- Franco, C. (2022). An Introduction to Small Area Estimation. <https://www.dropbox.com/scl/fi/p5bz1sc6xrow16gcu4tpx/SaeWebinarCIRSMay11.pdf?rlkey=bgf5pnajfn1xgaw56ezb9iop3&e=1&dl=0>.
- Franco, C. (2023). Aplicación de la Estimación de Áreas Pequeñas en las Estadísticas Oficiales de los Estados Unidos. <https://www.youtube.com/watch?v=hohZCYsyTWk>.
- Gershunskaya, J. and Lahiri, P. (2011). Robust Small Area Estimation Using a Mixture Model. In *Proceedings of the 58th World Statistics Congress*, pages 1132–1141, Dublin, Ireland. International Statistical Institute. <https://2011.isiproceedings.org/papers/450026.pdf>.
- Giusti, C., Masserini, L., and Pratesi, M. (2017). Local Comparisons of Small Area Estimates of Poverty: An Application Within the Tuscany Region in Italy. *Social Indicators Research*, 131:235–254. doi: 10.1007/s11205-015-1193-1.
- Guadarrama, M., Molina, I., and Rao, J. N. K. (2016). A Comparison Of Small Area Estimation Methods For Poverty Mapping. *Econstor*, 17(1):41–66. doi 10.21307/stattrans-2016-005.
- Gutiérrez, A. (2021). El enfoque de CEPAL en el mapeo de la pobreza. https://www.cepal.org/sites/default/files/presentations/eclac-approach-to-poverty-_mapping-jul-2021.pdf.
- Gutiérrez, A., Mancero, X., and Guerrero, S. (2022). Poverty mapping in Latin America: ECLAC experiences on small area estimation. *Statistical Journal of the IAOS*, 38(3):1021–1033. doi: 10.3233/SJI-220037.
- Harmening, S., Kreutzmann, A., Pannier, S., Salvati, N., and Schmid, T. (2020). emdi 2.0.1: A Framework for Producing Small Area Estimates based on Area-Level Models in R. https://r-project.ro/conference2020/presentations/emdi_2_0_1_A_Framework_for_Producing_Small_Area_Estimates_based_on_Area-Level_Models_in_R.pdf.
- Harmening, S., Kreutzmann, A.-K., Schmidt, S., Salvati, N., and Schmid, T. (2023). A framework for producing small area estimates based on area-level models in r. *The R Journal*, 15:316–341. doi: 10.32614/RJ-2023-039.
- Hernández, R., Fernández, C., and Baptista, M. (2014). *Metodología de la Investigación*. McGraw Hill, México D.F., sexta edition. <https://www.esup.edu.pe/wp-content/uploads/2020/12/2.%20Hernandez,%20Fernandez%20y%20Baptista-Metodolog%a%20Investigacion%20Cientifica%206ta%20ed.pdf>.
- Kreutzmann, A., Pannier, S., Rojas, N., Schmid, T., Templ, M., and Tzavidis, N. (2019). The R Package emdi for Estimating and Mapping Regionally Disaggregated Indicators. *Journal of Statistical Software*, 91(7):1–33. doi: 10.18637/jss.v091.i07.

- Kubacki, J. and Jędrzejczak, A. (2016). Small Area Estimation of Income Under Spatial SAR Model. *Statistics in Transition*, 17(3):365–390. doi: 10.21307/stattrans-2016-028.
- Marhuenda, Y., Molina, I., and Morales, D. (2013). Small area estimation with spatio-temporal Fay–Herriot models. *Computational Statistics and Data Analysis*, 58:308–325. doi: 10.1016/j.csda.2012.09.002.
- Marhuenda, Y., Morales, D., and Pardo, M. (2014). Information criteria for Fay–Herriot model selection. *Computational Statistics and Data Analysis*, 70:268–280. doi: 10.1016/j.csda.2013.09.016.
- Marhuenda, Y., Morales, D., and Pardo, M. (2016). Tests for the variance parameter in the Fay–Herriot model. *Statistics*, 50(1):27–42. doi: 10.1080/02331888.2015.1016026.
- Masaki, T., Newhouse, D., Silwal, A., Bedada, A., and Engstrom, R. (2021). Small Area Estimation with Geospatial Data. https://cega.berkeley.edu/wp-content/uploads/2020/04/Newhouse_MeasureDev2020_Slides.pdf.
- Merfeld, J., Chen, H., Lahiri, P., and Newhouse, D. (2023). Small Area Estimation with Geospatial Data: A Primer. <https://unstats.un.org/iswghs/documents/geospatial-data-for-SAE-outline.pdf>.
- Ministerio de Desarrollo Social de Chile (2013). Procedimiento de cálculo de la Tasa de Pobreza a nivel Comunal mediante la aplicación de Metodología de Estimación para Áreas Pequeñas (SAE). https://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/pobreza-comunal/2011/Procedimiento_de_calculo_de_la_Tasa_de_Pobreza_a_nivel_Comunal_11feb13.pdf.
- Ministerio de Economía y Finanzas de Perú (n.d.). Métodos para medir Pobreza. https://www.mef.gob.pe/es/?option=com_content&language=es-ES&Itemid=100412&lang=es-ES&view=article&id=370.
- Molina, I. (2019). *Desagregación de datos en encuestas de hogares - Metodologías de estimación en áreas pequeñas*. Number 97 in Estudios Estadísticos. Naciones Unidas, Santiago de Chile. <https://repositorio.cepal.org/server/api/core/bitstreams/5792f51b-c686-4624-9673-6bf6f6fa0d9d/content>.
- Molina, I. (2022). *Disaggregating data in household survey: Using small area estimation methodologies*. Number 97 in Statistics. ECLAC - UNFPA, Santiago de Chile. <https://repositorio.cepal.org/server/api/core/bitstreams/e4e38563-d196-4f01-b903-5f64faa6866b/content>.
- Molina, I., Marín, M., and Rao, J. (2019). Small Area Estimation. https://en.eustat.eus/sem19_curso_areas_pequenas_i.pdf.

- Molina, I. and Morales, D. (2009). Small area estimation of poverty indicators. *Boletín de Estadística e Investigación Operativa*, 25(3):218–225. https://www.researchgate.net/publication/28322424_Small_area_estimation_of_poverty_indicators.
- Molina, I., Nandram, B., and Rao, J. N. K. (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. doi: 10.1214/13-AOAS702.
- Morales, D., Esteban, M., Pérez, and Hobza, T. (2021). *A Course on Small Area Estimation and Mixed Models: Methods, Theory and Applications in R*. Statistics for Social and Behavioral Sciences. Springer, Switzerland.
- Müller, S., Scealy, J., and Welsh, A. (2013). Model Selection in Linear Mixed Models. *Statistical Science*, 28(2):135–167. doi: 10.1214/12-STS410.
- Nandy, A. (2013). An overview of Fay Herriot model with our package smallarea. <https://cran.r-project.org/web/packages/smallarea/vignettes/vignette.pdf>.
- Navarro, J. and Chávez, J. (2001). El Índice de Pobreza Foster Greer Thorbecke (FGT): Una Aplicación para Michoacán y sus Municipios, 1980-2000. *Economía y Sociedad*, 6(10):23–48. <https://dialnet.unirioja.es/servlet/articulo?codigo=5900498>.
- Newhouse, D. (2024). New Developments in Small Area Estimation: A Practitioner’s Perspective. <http://isi-iass.org/home/wp-content/uploads/Presentation-for-IASS-0226.pdf>.
- Novo, M. (2018). *Selección de modelos para estimación de áreas pequeñas. Aplicación a datos socioeconómicos de la Comunidad de Galicia*. Trabajo Fin de Máster, Universidad de Santiago de Compostela, Santiago de Compostela. http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1591.pdf.
- Open Street Map (2024). Colombia. <https://download.geofabrik.de/south-america/colombia.html>.
- Organización de las Naciones Unidas (2014). Principios Fundamentales de las Estadísticas Oficiales. https://unstats.un.org/unsd/dnss/hb/S-fundamental%20principles_A4-WEB.pdf.
- Organización de las Naciones Unidas (2017). Anexo: Marco de indicadores mundiales para los Objetivos de Desarrollo Sostenible y metas de la Agenda 2030 para el Desarrollo Sostenible. https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework_A.RES.71.313%20Annex.Spanish.pdf.

- Ortiz, F., Téllez, C., and Ramírez, N. (2024). A Review of the Use of Small Area Estimation in Colombia. *Revista Colombiana de Estadística - Applied Statistics*, 47(2):407–422. doi: 10.15446/rce.v47n2.112779.
- Pratesi, M., Marchetti, S., Giusti, C., and Salvati, N. (2023). The Use of Spatial Information in Area-level Models: An Evaluation Based on Auxiliary Data Availability. *Calcutta Statistical Association Bulletin*, 75(2):155–172. doi: 10.1177/00080683231198589.
- Pratesi, M., Quattrociochi, L., Bertarelli, G., Gemignani, A., and Giusti, C. (2021). Spatial Distribution of Multidimensional Educational Poverty in Italy using Small Area Estimation. *Social Indicators Research*, 156:563–586. doi: 10.1007/s11205-020-02328-5.
- Pratesi, M. and Salvati, N. (2009). Small Area Estimation in the Presence of Correlated Random Area Effects. *Journal of Official Statistics*, 25(1):37–53. <https://www.scb.se/contentassets/f6bcee6f397c4fd68db6452fc9643e68/small-area-estimation-in-the-presence-of-correlated-random-area-effects.pdf>.
- Pusponegoro, N. and Rachmawati, R. (2018). Spatial Empirical Best Linear Unbiased Prediction in Small Area Estimation of Poverty. In *Procedia Computer Science*, volume 135, pages 712–718. Elsevier. doi: 10.1016/j.procs.2018.08.214.
- Qiu, Y., Zhao, X., Fan, D., Li, S., and Zhao, Y. (2022). Disaggregating population data for assessing progress of SDGs: methods and applications. *International Journal of Digital Earth*, 15(1):2–29. doi: 10.1080/17538947.2021.2013553.
- Rao, J. and Molina, I. (2015). *Small Area Estimation*. Wiley, New Jersey, United States, second edition.
- Risal, A. (2021). Expenditure Per Capita Model With Spatial Small Area Estimation. *Parameter: Journal of Statistics*, 1(2):38–47. doi: 10.22487/27765660.2021.v1.i2.15502.
- Salvati, N. (2004). Small Area Estimation by Spatial Models: the Spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP). https://local.disia.unifi.it/pubblicazioni_DS/wp/2004/wp2004_03.pdf.
- Schork, J. (2022). Small Area Estimation Techniques | Analysis of Income, Poverty & Health. <https://statisticsglobe.com/small-area-estimation>.
- Sikov, A. and Cerda, J. (2024). Prediction in non-sampled areas under spatial small area models. *Statistical Methods & Applications*. doi: 10.1007/s10260-024-00754-0.
- Sáenz, H. (2016). Revisando los métodos de agregación de unidades espaciales: MAUP, algoritmos y un breve ejemplo. *Estudios demográficos y urbanos*, 31(2):385–411. <https://www.scielo.org.mx/pdf/educm/v31n2/0186-7210-educm-31-02-00385.pdf>.

- timeseriesreasoning.com (n.d.). A Tutorial On Generalized Least Squares Estimation Using Python And Statsmodels. <https://timeseriesreasoning.com/contents/generalized-least-squares-tutorial/>.
- Tzavidis, N., Zhang, L., Luna, A., Schmid, T., and Rojas, N. (2018). From Start to Finish: A Framework for the Production of Small Area Official Statistics. *Journal of the Royal Statistical Society*, 181(4):927–979. doi: 10.1111/rssa.12364.
- Téllez, C. (2020). *Estimación de áreas pequeñas utilizando imputación múltiple en modelos logísticos de tres parámetros*. Tesis de grado para optar al título de Doctor en Ciencias - Estadística, Universidad Nacional de Colombia, Bogotá D.C. <https://repositorio.unal.edu.co/handle/unal/78956>.
- Unidad Administrativa Especial de Catastro Distrital (n.d.). Unidades de Planeamiento Zonal -UPZ. <https://www.catastrobogota.gov.co/glosario-catastral/unidades-de-planeamiento-zonal-upz>.
- Unidad Administrativa Especial de Catastro Distrital - UAECD (2024). Datos Abiertos Bogotá. <https://datosabiertos.bogota.gov.co/>.
- Veijanen, A. and Lehtonen, R. (2011). Deliverable 2.2: Small Area Estimation of Indicators on Poverty and Social Exclusion. Technical report, European Commission. https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Ameli_Delivrables/AMELI-WP2-D2.2-supplement-20110402.pdf.
- Wakefield, J. (2022). Lecture 2: Area-Level Modeling. <http://faculty.washington.edu/jonno/SAEmaterial/2022-SISCER-SAE-area-level.pdf>.
- Warnholz, S. (2016). *Small Area Estimation Using Robust Extensions to Area Level Models: Theory, implementation and simulation studies*. Doctor Degree, Universitat Berlin, Berlín, Alemania. <https://refubium.fu-berlin.de/bitstream/handle/fub188/9706/main.pdf;jsessionid=B312ECC4C90A919DB948EA9442D3AF4AA?sequence=1>.
- WorldPop (2024). WorldPop gridded population estimate datasets and tools. How are they different and which should I use? <https://www.worldpop.org/methods/populations/>.