



DEPARTAMENTO ADMINISTRATIVO NACIONAL DE
ESTADÍSTICA - DANE

NOTA METODOLÓGICA

**ESTIMACIÓN EN ÁREAS
PEQUEÑAS DE INDICADORES DE
BIENESTAR EN COLOMBIA 2024
A NIVEL MUNICIPAL**

DICIEMBRE 2025

B. Piedad Urdinola Contreras
Directora

Andrea Ramírez Pisco
Subdirectora

Álvaro Fernando Guzmán Lucero
Secretaria General

Jefes Oficinas Asesoras

Claudia Díaz Hernández
Oficina Asesora de Planeación

Andrea Catalina Zota Bernal
Oficina Asesora Jurídica

Luis Martín Barrera Pino
Oficina de Sistemas

Yuly Dayan Quiceno Russi
Oficina de Control Interno

Angela Patricia Santiago Enciso
Oficina de Control Interno Disciplinario

Directores Técnicos

Cesar Mauricio López Alfonso
Dirección de Metodología y Producción Estadística

Diana María Rojas Ordus
Dirección de Recolección y Acopio

Javier Plazas Echeverri
Dirección de Censos y Demografía

Juan Pablo Cardoso Torres
Dirección de Síntesis y Cuentas Nacionales

Carlos Alberto Hernández Lozano
Dirección de Regulación, Planeación, Estandarización y Normalización

Elkin Ernesto Ramírez Niño
Dirección de Geoestadística

Diana María Bohórquez Losada
Dirección de Difusión y Cultura Estadística

Grupo de trabajo

Subdirección

Natalia Ximena Arteaga Gutiérrez

Dirección de Metodología y Producción Estadística

Juan José Rubio Mesa

GIT – Metodología y Procesos Estadísticos

Dayana Alejandra Barrera Buitrago

Nicolás Ramírez Vargas

Dirección de Recolección y Acopio

José Alejandro Rojas Venegas

GIT – Registros Estadísticos

Hernán David Torres Cardona

Edición y diseño

Dirección de Difusión y Cultura Estadística

© DANE, 2025

Prohibida la reproducción total o parcial sin permiso o autorización del Departamento Administrativo Nacional de Estadística, Colombia.

Tabla de contenido

1	Introducción	2
2	Metodología general de estimación en áreas pequeñas (SAE)	2
2.1	Estimador directo	3
2.2	Función Generalizada de Varianza	4
2.3	Fuentes de información	4
2.4	Modelo Fay-Herriot	5
2.5	Validación de supuesto del modelo	6
2.6	Benchmark	6
3	Inseguridad alimentaria moderada o grave (FIES)	7
3.1	Definición del indicador y marco conceptual FAO	7
3.2	Estimaciones directas	8
3.3	Aplicación del modelo SAE para FIES	9
3.4	Validación de supuesto del modelo	11
3.5	Benchmark	17
3.6	Resultados	18
4	Pobreza monetaria	19
4.1	Definición del indicador y metodología oficial del DANE	19
4.2	Estimaciones directas	19
4.3	Aplicación del modelo SAE para Pobreza Monetaria 2024	20
4.4	Validación de supuesto del modelo	21
4.5	Benchmark	23
4.6	Resultados	25
5	Pobreza multidimensional (IPM)	26
5.1	Definición del indicador y metodología oficial del DANE	26
5.2	Estimaciones directas	27
5.3	Aplicación del modelo SAE para IPM	28
5.4	Validación de supuestos del modelo	29
5.5	Benchmark	31
5.6	Resultados	33
6	Bibliografía	34

1 Introducción

La medición del bienestar de los hogares colombianos constituye un elemento central para el diseño y la evaluación de políticas públicas orientadas al cierre de brechas y a la reducción de la pobreza. En este marco, el Departamento Administrativo Nacional de Estadística (DANE) produce regularmente indicadores esenciales como la prevalencia de inseguridad alimentaria, la pobreza monetaria y la pobreza multidimensional, cada uno derivado de distintas fuentes de información y metodologías especializadas.

La Encuesta Nacional de Calidad de Vida (ECV) aporta información detallada sobre las condiciones de vida de los hogares y, en particular, incluye la Escala de Experiencia de Inseguridad Alimentaria (FIES), desarrollada por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), que permite medir la prevalencia de inseguridad alimentaria moderada o grave en el país. Asimismo, la ECV permite estimar el Índice de Pobreza Multidimensional (IPM), que refleja las privaciones de los hogares en Colombia. La ECV cuenta con un diseño de muestreo probabilístico, estratificado y multietápico, que posibilita generar estimaciones confiables a nivel nacional, por zona (urbana y rural) y a nivel departamental.

Por su parte, la pobreza monetaria se estima a partir de la Gran Encuesta Integrada de Hogares (GEIH), que recopila información continua sobre ingresos, empleo y otras características socioeconómicas. Esta operación estadística constituye la base para el cálculo oficial del ingreso monetario per cápita y, en consecuencia, de la incidencia de pobreza monetaria y de pobreza extrema. La GEIH también cuenta con un diseño de muestreo probabilístico, estratificado y multietápico, que permite la desagregación de la pobreza monetaria a nivel nacional, por zona (urbana y rural), para 24 departamentos y sus ciudades principales.

Si bien estas operaciones permiten generar resultados robustos en los niveles de desagregación contemplados en sus diseños muestrales, no ofrecen estimaciones directas confiables a nivel municipal. No obstante, la creciente necesidad de disponer de información territorial más granular para la formulación de políticas públicas, la focalización de programas y el seguimiento de los Objetivos de Desarrollo Sostenible exige metodologías que permitan ampliar el detalle geográfico de manera estadísticamente válida.

En este contexto, el DANE, con el acompañamiento de expertos internacionales, adoptó la metodología de Estimación en Áreas Pequeñas (SAE, por sus siglas en inglés). Este enfoque combina la información de encuestas probabilísticas con variables auxiliares provenientes de censos, registros administrativos e información geoespacial, lo que permite mejorar la precisión de las estimaciones en dominios con tamaños de muestra reducidos o sin observaciones directas. En particular, se implementó el modelo de área de Fay-Herriot (1979), ampliamente utilizado para la estimación de indicadores sociales a escala subnacional.

La presente nota metodológica describe el proceso general aplicado para estimar, a nivel municipal, tres indicadores clave del bienestar en Colombia para los años 2022 y 2024:

- i) la prevalencia de inseguridad alimentaria moderada o grave (FIES),
- ii) la incidencia de pobreza monetaria, y
- iii) la incidencia de pobreza multidimensional (IPM).

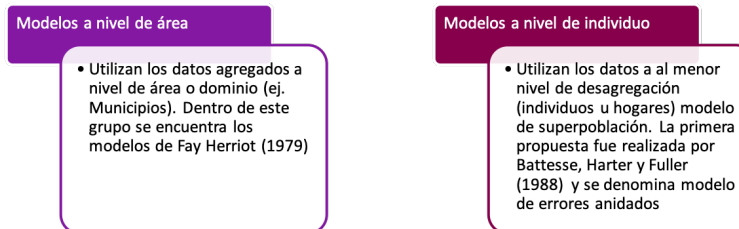
2 Metodología general de estimación en áreas pequeñas (SAE)

La metodología de Estimación en Áreas Pequeñas (SAE, por sus siglas en inglés) se utiliza cuando, mediante estimaciones directas —es decir, a partir de los datos recolectados en la encuesta y usando los factores de expansión— no es posible obtener resultados desagregados para los dominios de interés, debido a que el tamaño de muestra en dichos dominios es pequeño o incluso nulo. En general, los dominios se definen como áreas geográficas (departamentos, municipios, veredas) o como grupos poblacionales (grupos étnicos, población LGBTIQ+, víctimas, entre otros).

Para la aplicación de SAE se emplean estimadores indirectos, los cuales incorporan información de la encuesta, variables auxiliares provenientes de otras fuentes y la relación existente entre áreas o dominios. De acuerdo con la Comisión Económica para América Latina y el Caribe (CEPAL), la ganancia en eficiencia de los estimadores indirectos frente a los directos es mayor cuanto menor es el tamaño de muestra del dominio. No obstante, el uso de estimadores indirectos implica renunciar al insesgamiento propio de los estimadores basados exclusivamente en el diseño muestral, por lo que resulta fundamental seleccionar un modelo adecuado y verificar su ajuste a los datos.

Existen distintos modelos para la estimación en áreas pequeñas, que se agrupan en modelos de área y modelos de unidad, como se presenta en la siguiente ilustración.

Ilustración 1. Modelo conceptual FIES



2.1 Estimador directo

Las estimaciones directas son aquellas que se obtienen siguiendo el diseño de muestreo de la encuesta, sin incorporar información auxiliar. Para este tipo de estimaciones es necesario evaluar sus medidas de incertidumbre —como la varianza, los intervalos de confianza y los coeficientes de variación, entre otras— con el fin de determinar la calidad y confiabilidad de los resultados.

Uno de los estimadores más utilizados en encuestas probabilísticas es el estimador de Horvitz–Thompson (HT), el cual es insesgado respecto al diseño muestral y se define como:

$$\hat{t}_y = \sum_{k \in s} \frac{y_k}{\pi_k}$$

donde π_k representa la probabilidad de inclusión del elemento k en la muestra, determinada por el diseño de muestreo aplicado.

Toda estimación directa debe acompañarse de parámetros de calidad que permitan establecer su validez. Entre estos se encuentran el error estándar, el margen de error, el intervalo de confianza, el coeficiente de variación y el efecto de diseño (DEFF), entre otros. Estos parámetros se derivan de la estimación de la varianza del estimador. En el caso del estimador HT, una expresión general de la varianza está dada por:

$$\widehat{Var}_{\hat{t}_{y^k}} = \sum_{k \in s} \frac{\Delta_{kl}}{\pi_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

donde $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$, y π_{kl} es la probabilidad conjunta de inclusión de los elementos k y l en la muestra seleccionada.

En la desagregación de variables a partir de encuestas por muestreo se distinguen dos tipos de agrupaciones: (i) los estratos, que son subgrupos mutuamente excluyentes definidos desde el diseño muestral; por ello, el tamaño de muestra y el error muestral pueden controlarse desde el inicio de la operación estadística, lo que permite asegurar la calidad de las estimaciones a este nivel de desagregación; y (ii) los dominios,

que también corresponden a subgrupos poblacionales, pero no necesariamente se definen en el diseño de la encuesta. En consecuencia, su tamaño de muestra y error muestral dependen del comportamiento efectivo de la recolección, y la viabilidad de producir estimaciones a este nivel está condicionada a la calidad alcanzada por las estimaciones directas en cada dominio.

2.2 Función Generalizada de Varianza

La Función Generalizada de Varianza (FGV) es un modelo de regresión que asocia la varianza de las estimaciones con otros factores con el fin de realizar el suavizamiento ¹ a las estimaciones de las varianzas a nivel de los dominios de interés (municipios).

Para esta etapa se aplicó un modelo log-lineal usando el estimador de Mínimos Cuadrados Ordinarios, con la siguiente ecuación.

$$\log(\hat{V}_i) = \beta_0 + \beta_1 \frac{n_i}{N_i} + \beta_2 n_i + e_i$$

Donde, \hat{V}_i es la estimación de la varianza del estimador basado en el diseño y del modelo de Rasch.

2.3 Fuentes de información

Para la construcción de las variables auxiliares que se utilizarán en la estimación del modelo de área, se revisaron las siguientes fuentes de información:

1. **Estadísticas Vitales (DANE)**: información sobre mortalidad, causas de defunción y análisis por ciclo de vida.
2. **Censo Nacional de Población y Vivienda (DANE)**: información sociodemográfica de hogares y viviendas a nivel nacional.
3. **Censo Nacional Agropecuario (DANE)**: información de producción y variables socioeconómicas de hogares, viviendas, Unidades de Producción Agropecuaria (UPA) y Unidades No Agropecuarias (UPNA).
4. **SIPSA – Sistema de Información de Precios (DANE)**: variables relacionadas con el abastecimiento en centros mayoristas y el municipio de origen.
5. **Educación formal (EDUC) (DANE)**: información de establecimientos educativos públicos y privados del país.
6. **Valor agregado municipal (DANE)**: información económica municipal por actividad económica.
7. **SISPRO – Ministerio de Salud y Protección Social**: Sistema Integrado de Información de la Protección Social, con indicadores municipales de cobertura, infraestructura y resultados.
8. **SIPRA (UPRA)**: sistema que integra información de distintas entidades sobre producción agropecuaria, EVA y recursos ambientales.
9. **TerriData (DNP)**: información territorial sobre planeación, finanzas, gestión del riesgo y desempeño municipal.
10. **ICFES**: promedios municipales de puntajes globales y por áreas en pruebas por nivel.

¹los métodos de suavizamiento consisten en la reducción del ruido y datos atípicos de un conjunto de datos

11. **Policía Nacional:** estadísticas de delitos denunciados por la ciudadanía, diferenciadas por tipo de delito.
12. **IDEAM:** promedios climáticos municipales.
13. **Google Earth Engine:** fuentes geoespaciales (p. ej., luces nocturnas), y variables derivadas como accesibilidad a servicios (tiempo a un hospital), pendiente y altitud.
14. Otras fuentes complementarias.

La construcción de variables auxiliares para el modelo de Estimación en Áreas Pequeñas (SAE) se realizó mediante la integración de múltiples fuentes estadísticas y registros administrativos con cobertura nacional. Dado que los indicadores analizados —prevalencia de inseguridad alimentaria moderada o grave (FIES), incidencia de pobreza monetaria e incidencia de pobreza multidimensional (IPM)— capturan dimensiones distintas del bienestar, las variables auxiliares se seleccionaron con un enfoque amplio, incorporando factores asociados al entorno socioeconómico, demográfico, productivo, educativo, sanitario y territorial de los municipios del país.

Para el caso de FIES, se consideraron además los ejes y determinantes de la seguridad alimentaria y nutricional definidos en la Política Nacional de Seguridad Alimentaria y Nutricional (CONPES 113 de 2008), tales como:

- disponibilidad de alimentos;
- acceso físico y económico;
- consumo y hábitos alimentarios;
- aprovechamiento biológico y estado nutricional;
- calidad e inocuidad de los alimentos.

Para los indicadores de pobreza monetaria e IPM, se incorporaron determinantes tradicionalmente asociados a la pobreza, como características del mercado laboral, educación, ingreso, acceso a servicios públicos, condiciones de la vivienda, demografía municipal y estructura productiva local, entre otros.

El objetivo fue construir un conjunto robusto de variables que, de forma directa o por aproximación, reflejara condiciones estructurales relacionadas con el bienestar y las privaciones de los hogares. Como resultado, se consolidó una base municipal con 208 variables auxiliares para los 1.122 municipios del país.

En los casos en que se identificó información faltante en alguna variable para menos del 10% de los municipios, se aplicó un proceso de imputación mediante el método de k vecinos más cercanos (k -NN), con el fin de preservar la completitud y consistencia de la base, y asegurar la implementación adecuada del modelo SAE.

2.4 Modelo Fay-Herriot

Como se ha señalado en este documento, la estimación directa de indicadores a nivel municipal presenta limitaciones importantes debido a los tamaños de muestra reducidos o nulos en muchos dominios. Con el fin de mejorar la precisión de las estimaciones y aprovechar información auxiliar disponible para todos los municipios del país, se recurre a la aplicación de un modelo de área dentro del marco de Estimación en Áreas Pequeñas. El modelo adoptado es el propuesto por Fay y Herriot (1979), ampliamente utilizado en la literatura internacional para la estimación de proporciones e indicadores sociales en dominios pequeños.

El modelo de Fay–Herriot corresponde a un modelo lineal mixto, que incorpora simultáneamente un componente determinístico basado en covariables auxiliares y un componente estocástico que captura la variabilidad no explicada entre las áreas. En este enfoque, la estimación directa del parámetro de interés para el área d , denorada por $\hat{\delta}_d$, se relaciona con el verdadero valor poblacional δ_d mediante la expresión:

$$\hat{\delta}_d = \delta_d + \epsilon_d, \quad d = 1, \dots, D$$

donde los ϵ_d son los residuales del modelo con $\epsilon_d \sim N(0, \sigma_d^2)$. A su vez, el parámetro del área d corresponde a:

$$\delta_d = X_d' \beta + u_d$$

Donde u_d es el efecto aleatorio del dominio d con $u_d \sim N(0, \sigma_u^2)$.

2.5 Validación de supuesto del modelo

Para garantizar la adecuación del modelo de Fay–Herriot y la confiabilidad de las estimaciones obtenidas, es necesario verificar el cumplimiento de cuatro supuestos fundamentales:

- **Normalidad de los residuales:** los errores del modelo deben seguir, de manera aproximada, una distribución normal. Este supuesto puede evaluarse mediante pruebas como Shapiro–Wilk y gráficos cuantil–cuantil (Q–Q).
- **Normalidad de los efectos aleatorios:** los efectos específicos de cada área deben comportarse como realizaciones de una distribución normal, lo que permite una estimación adecuada de la variabilidad entre áreas.
- **Media cero de los residuales:** los residuales deben estar centrados en cero, lo cual sugiere ausencia de sesgos sistemáticos en las predicciones del modelo.
- **Independencia de los residuales:** los residuales no deben presentar patrones ni autocorrelación; su independencia garantiza que los errores asociados a un área no dependan de los de otras áreas.

La validación conjunta de estos supuestos permite confirmar que el modelo es estadísticamente consistente y adecuado para producir estimaciones en áreas pequeñas.

2.6 Benchmark

Con el fin de garantizar la coherencia entre las estimaciones municipales generadas mediante el modelo de áreas pequeñas y las cifras oficiales publicadas por el DANE en niveles de desagregación superiores (departamental y nacional), se aplicó un procedimiento de benchmarking basado en técnicas de calibración utilizadas en encuestas por muestreo. Este enfoque permite ajustar las estimaciones modelo-dependientes de manera que, al agregarse, reproduzcan los valores oficiales obtenidos mediante estimaciones directas para cada indicador (FIES, pobreza monetaria e IPM).

El proceso de calibración, de acuerdo con Gutiérrez (2016), puede describirse mediante los siguientes pasos:

1. Suponga que se tiene acceso a un vector de información $\mathbf{x}_k = (x_{1k}, x_{2k}, \dots, x_{pk})'$ de p variables auxiliares y conocido para cada uno de los individuos de la muestra.

2. A través de registros administrativos o datos oficiales se conoce el valor del total del vector \mathbf{x}_k para el total de la población, $T_x = \sum_{k \in U} \mathbf{x}_k$.
3. El propósito es estimar el total de \mathbf{x}_k usando la información en la muestra, x_k $k \in s$.
4. Buscando que la estimación cumpla con la siguiente restricción, $\sum_{k \in s} w_k \mathbf{x}_k = T_x$.
5. La idea consiste en buscar pesos w_k tan cercanos como sea posible al inverso de la probabilidad de inclusión del k -ésimo elemento $d_k = 1/\pi_k$.

En el contexto de esta investigación, el objetivo del benchmark es asegurar que la agregación de las estimaciones municipales obtenidas mediante el modelo de Fay–Herriot coincida exactamente con los valores departamentales oficiales reportados por el DANE para cada indicador. Esto garantiza que las cifras municipales sean coherentes, comparables y estadísticamente consistentes dentro del sistema de indicadores oficiales del país.

3 Inseguridad alimentaria moderada o grave (FIES)

Este capítulo presenta la estimación de la prevalencia de inseguridad alimentaria moderada o grave en los hogares colombianos a partir de la Escala de Experiencia de Inseguridad Alimentaria (FIES). Se describen la definición del indicador, las estimaciones directas derivadas de la Encuesta Nacional de Calidad de Vida (ECV) y la aplicación de la metodología de áreas pequeñas para obtener resultados a nivel municipal, junto con su análisis territorial.

3.1 Definición del indicador y marco conceptual FAO

FIES, abreviatura de la Escala de Experiencia de Inseguridad Alimentaria en inglés, es una herramienta de medición para el acceso a los alimentos en términos de cantidad y calidad. A través de ocho preguntas, esta herramienta captura la experiencia de inseguridad alimentaria en hogares, ofreciendo una métrica precisa de la gravedad de esta problemática. La Tabla 1 muestra las preguntas usadas para el cálculo de FIES usando la ECV 2024.

Tabla 1. Preguntas usadas para la estimación de FIES

Durante los últimos 12 meses, DEBIDO A LA FALTA DE DINERO U OTROS RECURSOS hubo algún momento en que usted u otra persona del hogar:			
Fraseo del ítem	Opciones de respuesta		Experiencia
1. Se preocupó por no tener suficientes alimentos para comer	Sí	No No sabe/ no informa	Preocupado
2. No pudo comer alimentos saludables y nutritivos	Sí	No No sabe/ no informa	Saludable
3. Consumió poca variedad de alimentos	Sí	No No sabe/ no informa	Poca variedad
4. Tuvo que saltar una comida (desayuno, almuerzo, comida o cena)	Sí	No No sabe/ no informa	Saltó comidas
5. Comió menos de lo que pensaba que debía comer	Sí	No No sabe/ no informa	Comió menos
6. El hogar se quedó sin alimentos	Sí	No No sabe/ no informa	Sin alimentos
7. Tuvo hambre pero no comió	Sí	No No sabe/ no informa	Sintió hambre
8. No comió en un día entero	Sí	No No sabe/ no informa	Día entero sin comer

Fuente: DANE, adaptado de FAO.

3.2 Estimaciones directas

Investigaciones especializadas señalan que la vivencia de la inseguridad alimentaria sigue un proceso que comienza con la preocupación sobre la capacidad de obtener alimentos suficientes. A medida que esta inseguridad se intensifica, se observa una adaptación en la dieta, con una disminución en la calidad y variedad de los alimentos. En etapas más avanzadas, se evidencia una reducción en la cantidad de alimentos consumidos por los hogares.

Las ocho preguntas de la Escala FIES se fundamentan en un modelo conceptual que ilustra las diversas experiencias relacionadas con la inseguridad alimentaria, atribuidas a la falta de recursos, ya sea dinero u otros. Este modelo abarca desde la incertidumbre inicial o la preocupación acerca de la capacidad para obtener alimentos, pasando por la disminución en la calidad y variedad de los alimentos disponibles, hasta llegar a la reducción en las cantidades consumidas, incluyendo saltarse comidas principales o incluso enfrentar días enteros sin alimentos.

Ilustración 1. Modelo conceptual FIES



Fuente: FAO

La metodología para analizar los datos de las preguntas FIES (Tabla 1) se basa en la aplicación de un modelo estadístico de Teoría de Respuesta al Ítem (TRI). Los modelos TRI se usan para la medición de rasgos que no se pueden observar directamente (rasgos latentes), pero que no se pueden inferir de datos observables, como las respuestas a encuestas. En el caso del análisis FIES, se utiliza un modelo de *Rasch*, bajo el supuesto fundamental de que la situación de inseguridad alimentaria de un hogar o un individuo se puede analizar como rasgo latente. Encontrando que a mayor probabilidad de tener ítem o experiencias de mayor gravedad se encuentra una mayor gravedad de la experiencia de seguridad alimentaria del hogar.²

3.2 Estimaciones directas

La Encuesta Nacional de Calidad de Vida 2024 aplica un diseño de muestreo probabilístico estratificado (Departamentos - Zona: Urbano/Rural) en varias etapas (Municipios, Segmentos). Al realizar estimaciones directas a nivel municipal, se encuentran 636 de 1.121 municipios con por lo menos un segmento seleccionado por la muestra. Sin embargo, es importante señalar que aunque se haya podido construir una estimación directa para 636 municipios eso no implica que cumplan con los criterios de calidad para su publicación.

²<https://www.dane.gov.co/files/operaciones/FIES/bol-FIES-2024.pdf>

3.3 Aplicación del modelo SAE para FIES

Como se ha mencionado para la estimación del nivel de inseguridad alimentaria moderada o grave se utilizó un modelo de *Rash*, el cuál se define como:

$$Prob(X_{i,j} = 1) = \frac{\exp(a_i - b_j)}{1 + \exp(a_i - b_j)}$$

Donde $Prob(X_{i,j} = 1)$ es la probabilidad del hogar i de responder positivamente el ítem j . a_i es el parámetro de severidad del hogar i y b_j es el parámetro de severidad del ítem, es decir, el nivel de la severidad de la experiencia de inseguridad alimentaria asociada al ítem j .

La estimación de la $Prob(X_{i,j} = 1)$ es basado en la aproximación de máxima verosimilitud. Al realizar la estimación para el nivel nacional y departamental se calculan las estimaciones directas basados en el diseño muestral haciendo uso de los correspondientes factores de expansión del diseño muestral. La varianza de las estimaciones está compuesta tanto por la varianza del estimador de muestreo como la varianza proveniente de la estimación del modelo y está dada como:

$$\text{Varianza total} = \text{Varianza de muestreo} + \text{Varianza de estimación del modelo}$$

Con el fin de considerar el error de medición del modelo de *Rasch* dentro del modelo SAE, se evaluaron los modelos de área en el marco de los modelos de TRI. Para lo cual se requiere definir las estimaciones directas y sus correspondientes estimaciones de varianza de estimadores que cumplan con ciertos criterios de calidad con el fin de no inducir a sesgos en las estimaciones municipales.

Para ello se definieron los siguientes criterios de calidad para la selección de estimaciones municipales:

1. Tamaño de muestra efectivo, es decir, $neff = n/defl$. Manteniendo aquellos municipios con un $neff \geq 40$.
2. Grados de libertad $gl = UPM - Estratos$. Manteniendo aquellos municipios con $gl \geq 2$.
3. Coeficiente de variación logaritmico, manteniendo aquellos municipios con $CVLOG \leq 30$.
4. Efecto de diseño $DEFF$, manteniendo aquellos municipios con $DEFF \geq 1$.
5. Excluir municipios en los que la muestra seleccionada haya sido recolectada únicamente en una zona: urbana o rural y que esta zona no represente la mayor proporción de la población del municipio.

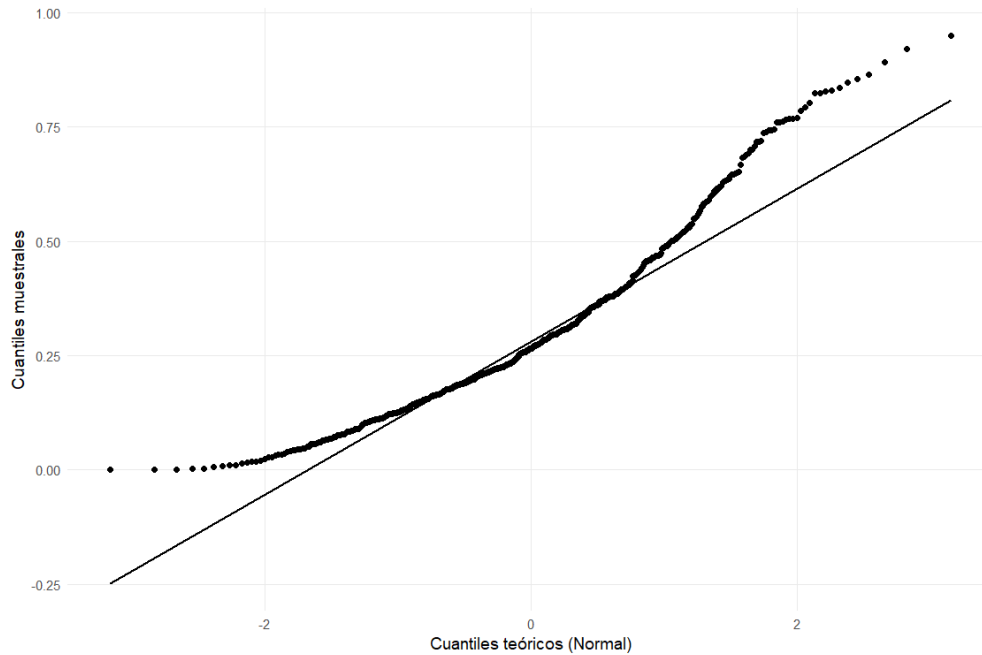
Como resultado de la aplicación de las 5 reglas se tiene un total de 344 municipios con estimaciones directas de calidad para ser usadas en el modelo de área.

3.3 Aplicación del modelo SAE para FIES

Para garantizar el cumplimiento de los supuestos del modelo de Fay–Herriot y asegurar que las estimaciones se mantuvieran dentro del soporte propio de una proporción (0 a 1), la variable de interés —la prevalencia de inseguridad alimentaria moderada o grave— fue transformada mediante la función arco-seno. Esta transformación permite aproximar la distribución de la variable a una forma más cercana a la normalidad, dado que las proporciones estimadas directamente no presentan un comportamiento normal, como se evidencia en la gráfica siguiente:

Gráfica 1. Normalidad inseguridad alimentaria moderada o grave municipal

3.3 Aplicación del modelo SAE para FIES



Fuente: DANE.

Una vez transformada la variable dependiente, y contando con el conjunto completo de variables auxiliares construidas para esta aplicación, se realizó un proceso de selección de covariables con el fin de identificar aquellas que mejoran la capacidad predictiva del modelo. Para ello se empleó el algoritmo stepwise con criterio de información AIC, implementado en la librería bigstep de R. El procedimiento permitió seleccionar un conjunto final de 52 variables auxiliares relevantes para la estimación de FIES a nivel municipal.

Con la variable transformada y las covariables seleccionadas, se estimó el modelo de área de Fay–Herriot utilizando la función fh de la librería emdi de R. El proceso se ejecutó mediante 500 iteraciones, garantizando estabilidad en la estimación de los parámetros y en la recuperación de los efectos aleatorios. Con un R^2 de 71%, se obtiene el siguiente modelo:

Tabla 2. Estimación modelo FH

Variable	Coefficiente (s.e.)
Intercepto	0,97421 (0,38621) *
Elevación promedio	-0,03012 (0,00868) ***
Número promedio de consultas - embarazos	0,00195 (0,00069) **
Departamento - Santander	-0,16333 (0,03858) ***
Desplazamiento forzado	0,0833 (0,04846) .
Departamento - Arauca	-0,08739 (0,05428)
Promedio UPAS dedicadas a pesca	0,18204 (0,07108) *
Mujeres afiliadas a salud	3,37004 (1,52142) *
Porcentaje hogares cohabitados	-0,59373 (0,15148) ***
Sedes educativas de primaria	0,06998 (0,0294) *
Precipitacion	-0,00048 (0,00014) ***
Departamento - Chocó	-0,1064 (0,04805) *
Departamento - Boyacá	-0,10022 (0,03939) *
Departamento - Caldas	-0,15317 (0,04996) **
Porcentaje de áreas construidas	-0,32995 (0,14258) *

3.4 Validación de supuesto del modelo

Departamento - Cundinamarca	-0,12826 (0,0402) **
Proyección población de número de mujeres	-1,9096 (0,56729) ***
Departamento - Córdoba	0,09693 (0,03704) **
Promedio de cultivos con eventos	0,17482 (0,06181) **
Tasa de amenazas	-0,00013 (0,0001)
Departamento - Caquetá	-0,12121 (0,05658) *
Porcentaje de desertores del sistema de educación	0,5502 (0,20576) **
Departamento - Guainía	-0,2085 (0,08506) *
Rural disperso	0,35607 (0,1014) ***
Sedes educativas en resguardos indígenas	0,00107 (0,0004) **
Porcentaje de personas pertenecientes al regimen subsidiado	0,90448 (0,28138) **
Porcentaje de jefes de hogar con bachillerato o superior	0,2293 (0,15984)
Departamento - Amazonas	-0,29157 (0,09131) **
Porcentaje de hogares que NO tienen servicio de recolección de basuras	0,10239 (0,07163)
Sedes de educación rural	-0,00063 (0,00033) .
temperatura	0,00604 (0,00282) *
Tasa de homicidios	-0,0009 (0,00032) **
Proporción de personas con educación superior	-0,86105 (0,28859) **
Proporción de hogares que no cuentan con un lugar adecuado para cocinar	-0,22522 (0,1414)
Porcentaje de población rural en el municipio	-0,22273 (0,07024) **
Área de UPAS	0,10129 (0,02847) ***
Cobertura bosques	-0,13604 (0,04665) **
Animales en las UPAS	0,19 (0,07383) *
Departamento - Risaralda	-0,14601 (0,05354) **
Promedio área cultivos por UPA	-0,13754 (0,09145)
Porcentaje de hogares donde el jefe del hogar es mujer	-0,72008 (0,30679) *
Eventos de minas antipersonales	-0,06207 (0,03205) .
Proposición de embarazo adolescente	-0,00004 (0,00001) *
área per UPA	0,09228 (0,0278) ***
Área agrícola	-0,03123 (0,00953) **
Cobertura educación primaria	0,11953 (0,05291) *
Departamento - Huila	-0,10665 (0,04025) **
Número de hombres afiliados a salud	-2,26447 (1,41473) **
Total de unidades ocupadas del CNUE 2021	-0,11607 (0,05766) *
Total de unidades económicas del sector comercio del CNUE 2021	0,07964 (0,05562)
Índice de Desempeño Fiscal	0,11772 (0,06813) .
Estudiantes promedio por sede	-0,00018 (0,00012)
Departamento - Meta	-0,02113 (0,04155)

	Skewness	Kurtosis	Shapiro_W	Shapiro_p
Standardized_Residuals	-0.0133877	2.998100	0.9982376	0.9786743
Random_effects	0.0911543	2.821954	0.9961933	0.5822047

Fuente: DANE.

3.4 Validación de supuesto del modelo

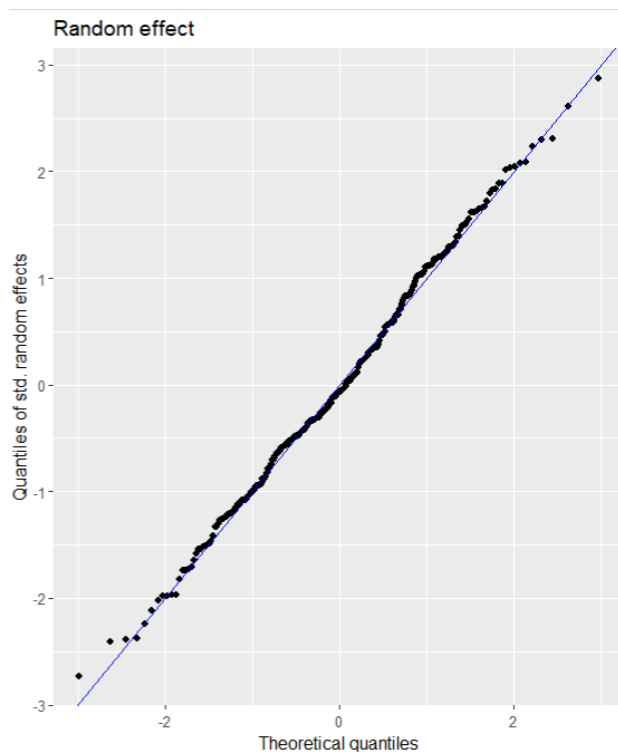
El modelo debe cumplir con los supuestos de:

1. Normalidad en los residuales
2. Normalidad en los efectos aleatorios
3. Media cero en los residuales
4. Independencia en los residuales

3.4.1 Normalidad en los residuales

De acuerdo con la segunda parte de la Tabla 2 se observa que el valor p de la prueba de hipótesis de Shapiro-Wilk para los residuales es de 0.978, es decir, que no hay evidencia estadística suficiente para rechazar la hipótesis de que los residuales se distribuyen bajo una distribución normal. Adicionalmente, la gráfica cuantil cuantil para la validación visual de la distribución confirma que la distribución normal es razonable para los residuales:

Gráfica 2. Gráfica cuantil cuantil para los residuales del modelo.

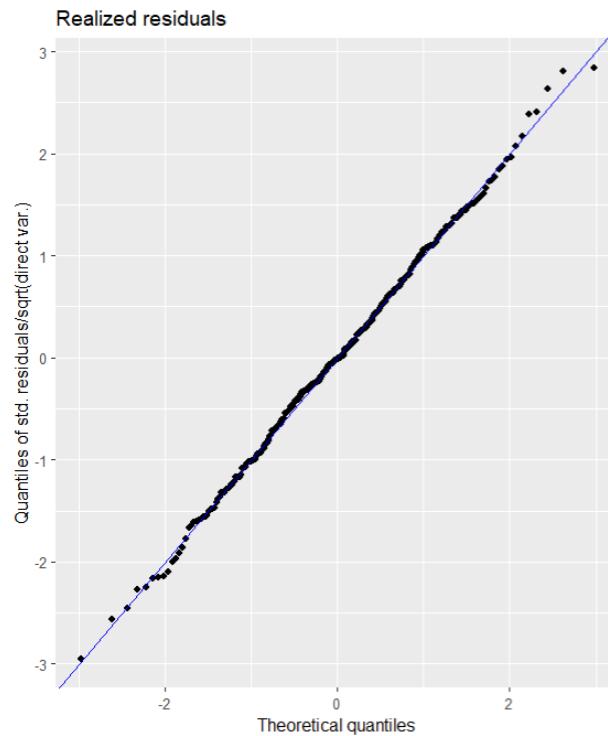


Fuente: DANE

3.4.2 Normalidad en los efectos aleatorios

Similar a los residuales, el valor p de la prueba de hipótesis de Shapiro-Wilk para los efectos aleatorios es de 0.582, es decir, que no hay evidencia estadística suficiente para rechazar la hipótesis de que los efectos aleatorios se distribuyen bajo una distribución normal, la misma conclusión se puede obtener observando la gráfica cuantil cuantil presentada a continuación:

Gráfica 3. Gráfica cuantil cuantil para los efectos aleatorios del modelo.

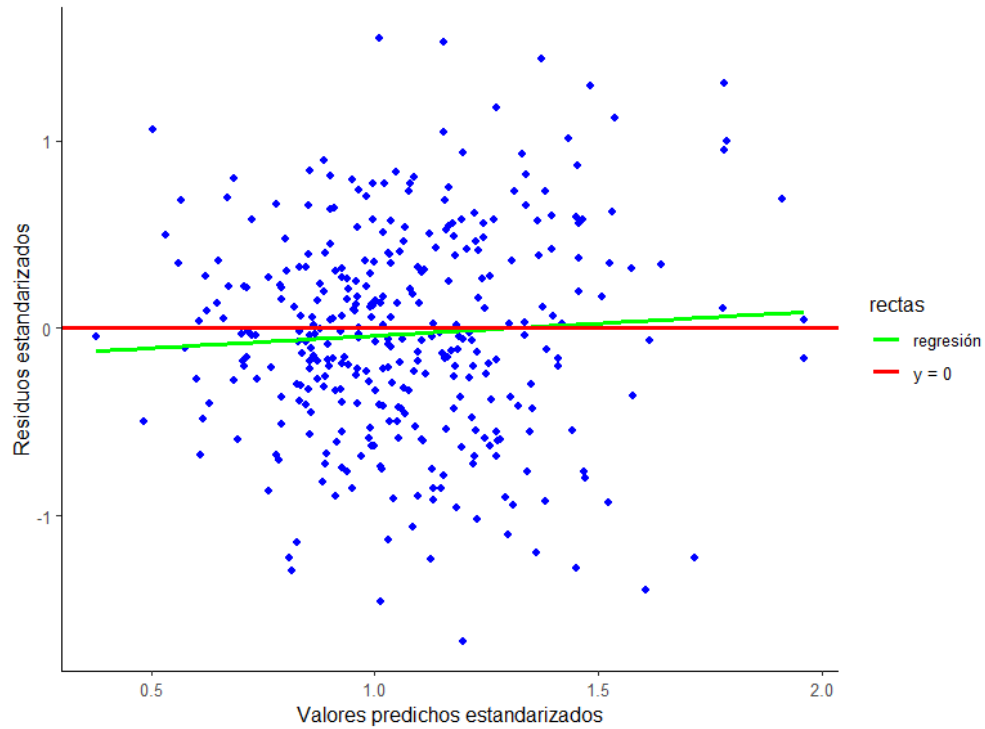


Fuente: DANE

3.4.3 Media cero en los residuales

En cuanto a la hipótesis de que los errores del modelo tenga media nula, de acuerdo con la siguiente gráfica se puede concluir que esta hipótesis es apropiada para los residuales.

Gráfica 4. Gráfica de dispersión de los residuales estandarizados y valores predichos estandarizados

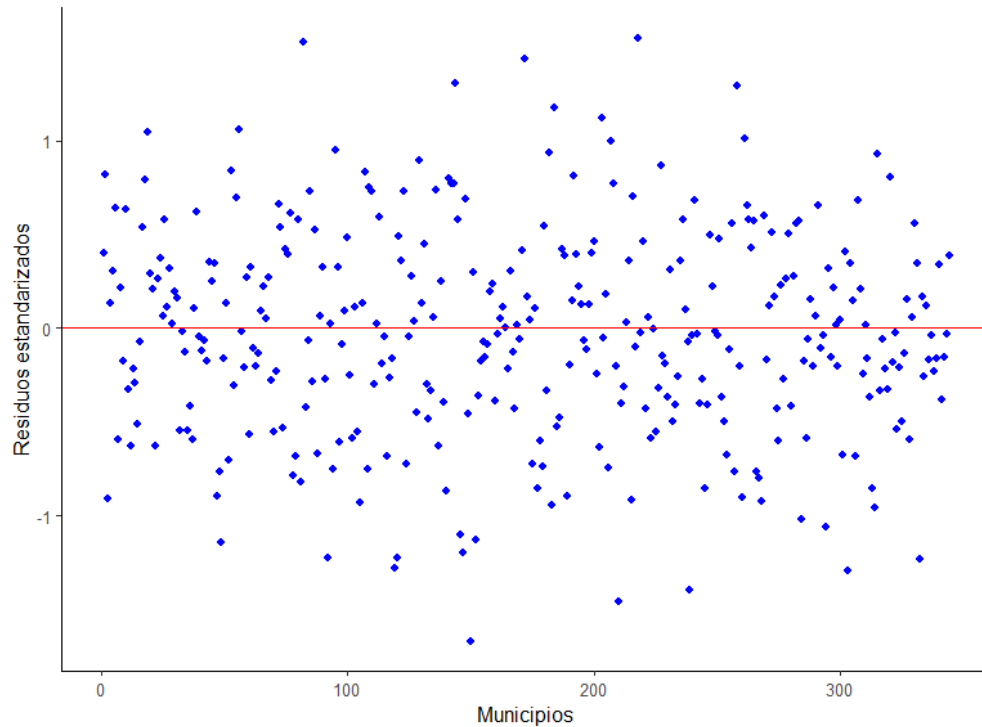


Fuente: DANE

3.4.4 Independencia en los residuales

De acuerdo con la siguiente gráfica se puede observar que no existe ningún patrón en el comportamiento de los residuales del modelo, por lo cual se puede asegurar la independencia.

Gráfica 5. Independencia de los residuales



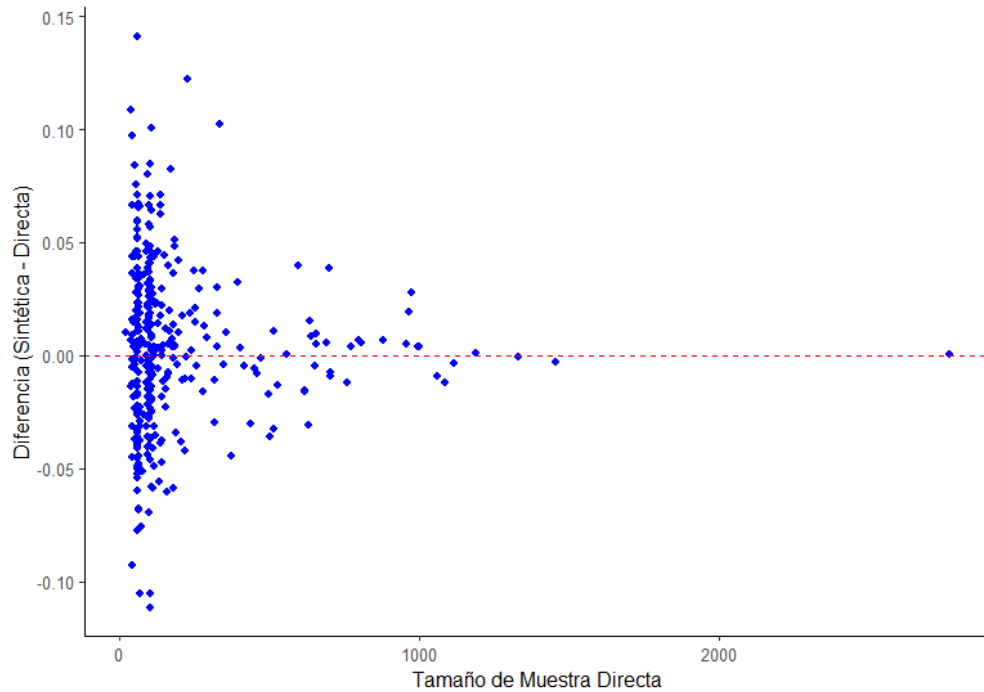
Fuente: DANE

3.4.5 Consistencia

Teniendo en cuenta que el modelo cumple con los supuestos requeridos, se realizan algunas pruebas de consistencia, como lo es validar que efectivamente para los municipios con tamaño de muestra grande en la encuesta mantienen una estimación similar o igual al de la estimación directa. En la siguiente gráfica se observa en el eje x el tamaño de muestra de la encuesta por cada municipio y en el eje y la diferencia entre la estimación del modelo FH y la estimación directa, encontrando que a mayor tamaño de muestra la diferencia de las estimaciones es aproximadamente cero.

Gráfica 6. Diferencia entre las estimaciones directas y de Fay-Herriot

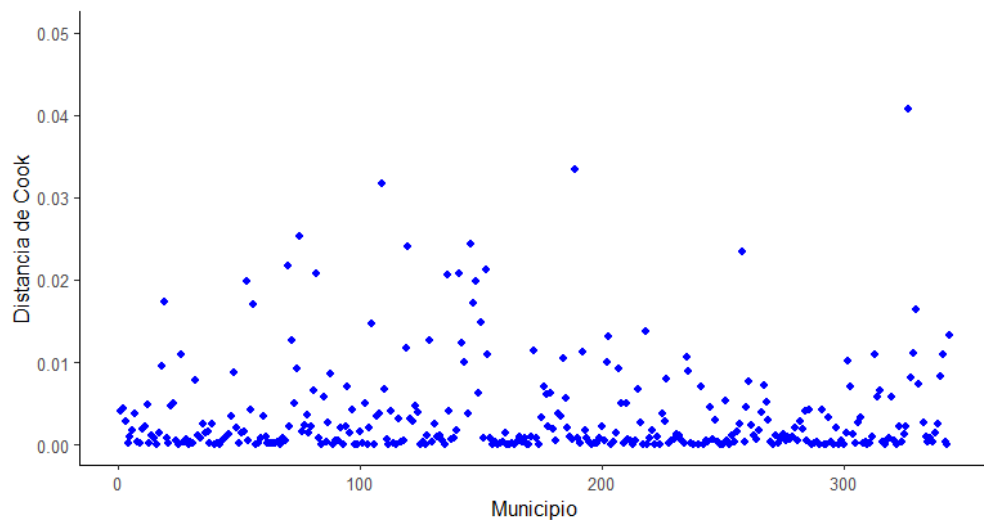
3.4 Validación de supuesto del modelo



Fuente: DANE

Adicionalmente, se calculó la distancia de Cook con el fin de identificar observaciones influyentes en el modelo que pudieran alterar las estimaciones, y se observa que no se encuentran municipios influyentes.

Gráfica 7. Distancia de Cook para detectar municipios influyentes en el modelo



Fuente: DANE

3.5 Benchmark

El siguiente paso aplicado fue la comparación entre la estimación directa departamental y la agregación a partir del promedio ponderado por la proyección poblacional de la estimación a través del modelo FH, encontrando las siguientes diferencias:

Tabla 3. Diferencias entre las estimaciones directa y modelo FH a nivel departamental

Departamento	Estimación directa	Estimación FH	Diferencia (pp)
Antioquia	0.279	0.285	0.574
Atlántico	0.400	0.391	0.939
Bogotá, D.C.	0.139	0.139	0.003
Bolívar	0.358	0.404	4.567
Boyacá	0.186	0.192	0.580
Caldas	0.128	0.139	1.133
Caquetá	0.178	0.226	4.758
Cauca	0.278	0.294	1.606
Cesar	0.303	0.308	0.478
Córdoba	0.476	0.487	1.136
Cundinamarca	0.226	0.224	0.211
Chocó	0.363	0.360	0.286
Huila	0.173	0.189	1.568
La Guajira	0.524	0.509	1.452
Magdalena	0.321	0.316	0.481
Meta	0.182	0.230	4.784
Nariño	0.372	0.345	2.679
Norte de Santander	0.248	0.265	1.711
Quindío	0.200	0.217	1.742
Risaralda	0.205	0.212	0.651
Santander	0.162	0.155	0.682
Sucre	0.495	0.481	1.433
Tolima	0.256	0.287	3.149
Valle del Cauca	0.207	0.219	1.151
Arauca	0.218	0.242	2.386
Casanare	0.344	0.346	0.178
Putumayo	0.280	0.313	3.285
Archipiélago de San Andrés, Providencia y Santa Catalina	0.318	0.309	0.903
Amazonas	0.191	0.234	4.270
Guainía	0.375	0.365	1.002
Guaviare	0.336	0.335	0.137
Vaupés	0.461	0.457	0.383
Vichada	0.425	0.416	0.925

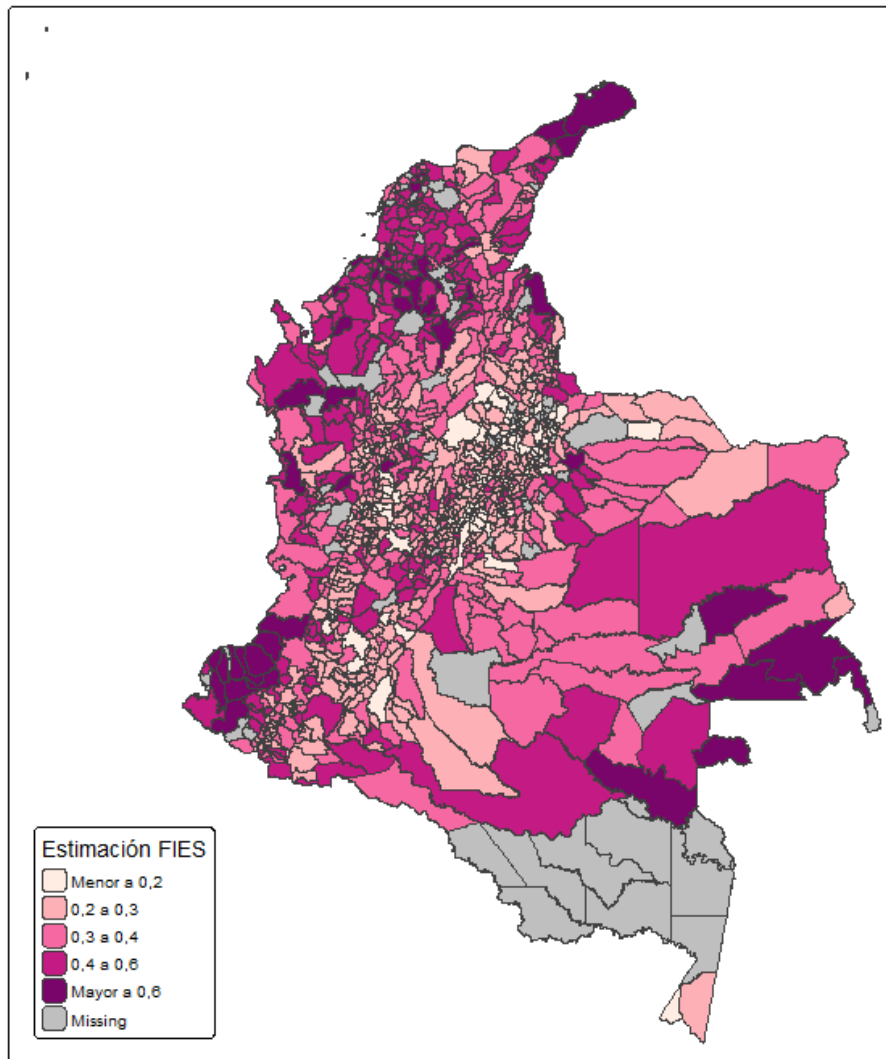
Fuente: DANE

Siendo el promedio de las diferencias a nivel departamental de 1.5 p.p. para la cual se propone un ejercicio de *benchmark* para eliminar las diferencias y así contar con la consistencia en las publicaciones de estadísticas oficiales en el DANE.

Al aplicar la calibración se garantiza que el promedio ponderado para los municipios por la proyección poblacional de 2024 de la estimación del modelo FH se obtiene el mismo valor reportado por la estimación directa para los 33 departamentos del país.

3.6 Resultados

Gráfica 8. Prevalencia de la inseguridad alimentaria moderada y grave en hogares (%)



Fuente: DANE.

4 Pobreza monetaria

Este capítulo presenta la estimación de la incidencia de pobreza monetaria en los hogares colombianos, a partir de la información recopilada por la Gran Encuesta Integrada de Hogares (GEIH). Se describe la definición oficial del indicador, las estimaciones directas obtenidas bajo el diseño probabilístico de la GEIH y la forma en que la metodología de áreas pequeñas permite producir estimaciones consistentes y comparables a nivel municipal.

4.1 Definición del indicador y metodología oficial del DANE

La pobreza monetaria es un indicador oficial del DANE que mide la proporción de personas cuyos ingresos per cápita del hogar son insuficientes para adquirir, en promedio, un conjunto mínimo de bienes y servicios esenciales. Para su cálculo se compara el ingreso per cápita del hogar con la Línea de Pobreza Monetaria (LPM), la cual representa el costo de dicha canasta y se actualiza periódicamente siguiendo los lineamientos metodológicos establecidos por la entidad.

El indicador se formula de la siguiente manera:

$$PobrezaMonetaria = \frac{\text{Número de personas en hogares con ingreso per cápita} < \text{LPM}}{\text{Total de personas}}$$

La estimación oficial se basa en la información recolectada por la Gran Encuesta Integrada de Hogares (GEIH), una operación estadística continua con diseño probabilístico, que permite generar resultados representativos a nivel nacional, urbano-rural y departamental, así como para dominios especiales definidos en el diseño de la encuesta.

El proceso metodológico incluye:

- la medición detallada de los ingresos corrientes monetarios del hogar,
- la imputación y ajuste de componentes de ingreso según estándares internacionales,
- la construcción del ingreso per cápita mediante el tamaño del hogar ajustado, y
- la comparación con la línea de pobreza definida por el DANE para el año de referencia.

Dado que la GEIH no está diseñada para producir estimaciones directas a nivel municipal, el cálculo de la pobreza monetaria en este nivel requiere complementar la información de encuesta con técnicas de estimación en áreas pequeñas (SAE), que permiten obtener estimaciones consistentes respetando la estructura del indicador oficial.

4.2 Estimaciones directas

La Gran Encuesta Integrada de Hogares (GEIH) aplica un diseño de muestreo probabilístico estratificado por departamentos y zona (urbano/rural), con selección en varias etapas (Municipios, Segmentos). Bajo este esquema, es posible construir estimaciones directas de pobreza monetaria para un número considerable de municipios del país. Sin embargo, es importante señalar que aun cuando se pueda generar una estimación directa para estos municipios, esto no implica que cumplan con los criterios de calidad necesarios para su publicación ni para su utilización en el modelo de Fay-Herriot.

4.3 Aplicación del modelo SAE para Pobreza Monetaria 2024

Las estimaciones directas se obtienen siguiendo el diseño muestral y utilizando los factores de expansión para representar adecuadamente la población objetivo. La precisión de estas estimaciones depende de la variabilidad del diseño, expresada mediante indicadores como el tamaño de muestra efectivo, el coeficiente de variación y el efecto de diseño.

Con el fin de asegurar que las estimaciones municipales utilizadas en el modelo SAE no introduzcan sesgos y cumplan con un nivel mínimo de precisión, se definieron los siguientes criterios de calidad para la selección de municipios:

1. Tamaño de muestra efectivo, definido como

$$neff = \frac{n}{DEFF}$$

Se mantienen aquellos municipios con $neff \geq 40$

2. Coeficiente de variación logarítmico (CVLOG). Se conservan los municipios con $CVLOG \leq 30$.
3. Efecto de diseño (DEFF). Se retienen únicamente los municipios con $DEFF \geq 1$.

Como resultado de la aplicación de estas tres reglas, se obtiene el conjunto final de municipios cuyas estimaciones directas presentan la calidad necesaria para ser utilizadas en el modelo de Fay–Herriot para la estimación de pobreza monetaria. De los 587 municipios iniciales, tras aplicar los filtros se conservan 392 municipios.

4.3 Aplicación del modelo SAE para Pobreza Monetaria 2024

Una vez construida la variable dependiente, la proporción de personas en condición de pobreza monetaria en cada municipio, y contando con el conjunto completo de variables auxiliares preparadas para esta aplicación, se llevó a cabo un proceso de selección de covariables con el fin de identificar aquellas que mejoran la capacidad predictiva del modelo. Para ello se empleó el algoritmo stepwise con criterio de información AIC, implementado en la librería bigstep de R. Este procedimiento permitió seleccionar un conjunto final de variables auxiliares relevantes para la estimación de la pobreza monetaria a nivel municipal.

Con la variable dependiente definida y las covariables seleccionadas, se estimó el modelo de área de Fay–Herriot utilizando la función fh de la librería emdi de R. El proceso se ejecutó con 500 iteraciones, garantizando estabilidad en la estimación de los parámetros y en las predicciones finales para cada municipio y obteniendo un R^2 de 88%.

Tabla 4. Estimación modelo FH – Pobreza Monetaria 2024

Variable	Coefficiente (s.e.)
Intercepto	0.881 (0.165)**
Régimen contributivo BDU A	-1.64 (0.255)**
Personas ocupadas	-0.39 (0.099)**
Departamento: Atlántico	-0.159 (0.03)**
Departamento: Caquetá	-0.156 (0.035)**
Régimen subsidiado BDU A	-1.021 (0.244)**
Cohabitación	0.46 (0.109)**
Departamento: Chocó	0.143 (0.033)**
Departamento: Magdalena	0.06 (0.032)
Cobertura de matorrales y arbustos	0.153 (0.048)*

4.4 Validación de supuesto del modelo

Mediana puntaje Global Saber 11	-0.001 (0)*
Proyección población urbana	0.263 (0.045)**
Hogares con recolección de basuras	0.234 (0.05)**
Departamento: Santander	-0.077 (0.022)**
Departamento: Meta	-0.093 (0.028)**
Departamento: Caldas	-0.102 (0.027)**
Precipitación	0 (0).
Derrumbes	0.736 (0.225)*
Tasa de secuestros	-0.004 (0.002).
Departamento: Risaralda	-0.05 (0.03)
Cobertura del suelo sin información	-4.507 (1.398)*
Víctimas del conflicto armado	0.089 (0.037).
Personas mayores de 65 años	0.759 (0.228)**
Hogares con energía eléctrica	0.181 (0.066)*
Proporción de personas pensionadas	-0.564 (0.164)**
Unidades económicas ocupadas	0.029 (0.006)**
Rural disperso	-0.035 (0.018)
Proporción de hogares en Nivel B	0.29 (0.179)
Número de sedes educativas primaria	0.047 (0.02).
Municipio PDET	-0.015 (0.018)
UPAS con problemas en los cultivos	-0.024 (0.034)

Fuente: DANE.

	Skewness	Kurtosis	Shapiro_W	Shapiro_p
Standardized_Residuals	-0.00070	3.52246	0.99355	0.10815
Random_effects	-0.03547	3.43906	0.99670	0.63946

4.4 Validación de supuesto del modelo

Similar que en caso de FIES, el modelo debe cumplir con los siguientes supuestos:

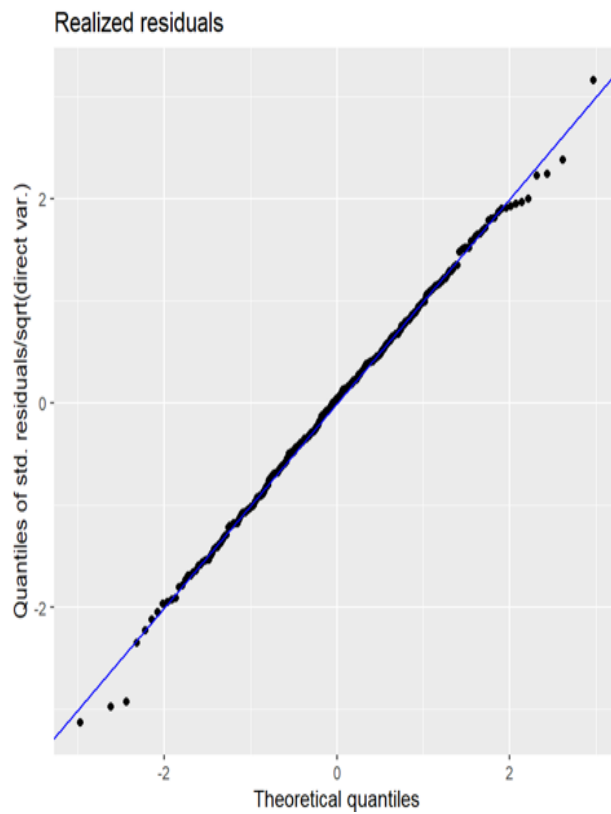
1. Normalidad en los residuales
2. Normalidad en los efectos aleatorios
3. Media cero en los residuales
4. Independencia en los residuales

Teniendo en cuenta que los supuestos 3 y 4 se cumplen y tienen un comportamiento similar a los mostrados anteriormente, se omiten las gráficas para su validación.

4.4.1 Normalidad en los residuales

De acuerdo con la segunda parte de la Tabla 4 se observa que el valor p de la prueba de hipótesis de Shapiro-Wilk para los residuales es de 0.108, es decir, que no hay evidencia estadística suficiente para rechazar la hipótesis de que los residuales se distribuyen bajo una distribución normal. Adicionalmente, la gráfica cuantil cuantil plot para la validación visual de la distribución confirma que la distribución normal es razonable para los residuales:

Gráfica 9. Gráfica cuantil cuantil para los residuales del modelo.

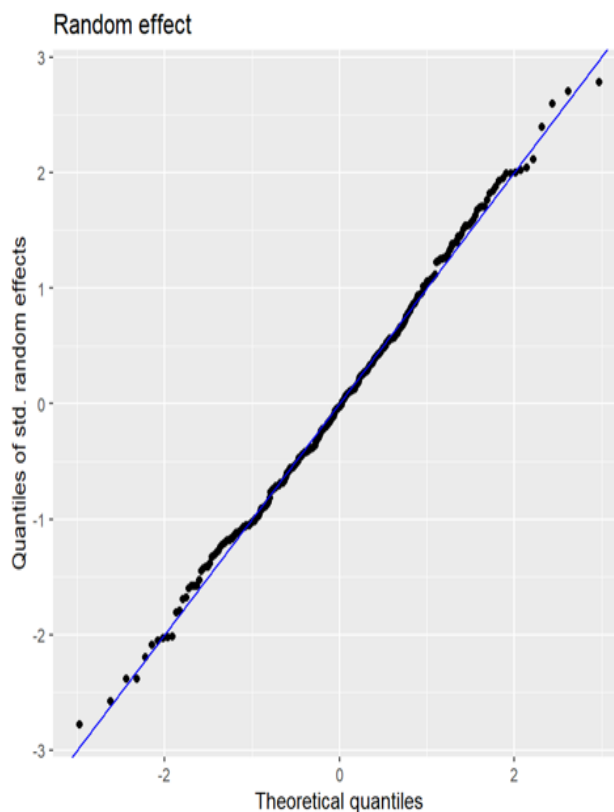


Fuente: DANE

4.4.2 Normalidad en los efectos aleatorios

Similar a los residuales, el valor p de la prueba de hipótesis de Shapiro-Wilk para los efectos aleatorios es de 0.639, es decir, que no hay evidencia estadística suficiente para rechazar la hipótesis de que los efectos aleatorios se distribuyen bajo una distribución normal, la misma conclusión se puede obtener observando la gráfica cuantil cuantil presentada a continuación:

Gráfica 10. Gráfica cuantil cuantil para los efectos aleatorios del modelo.



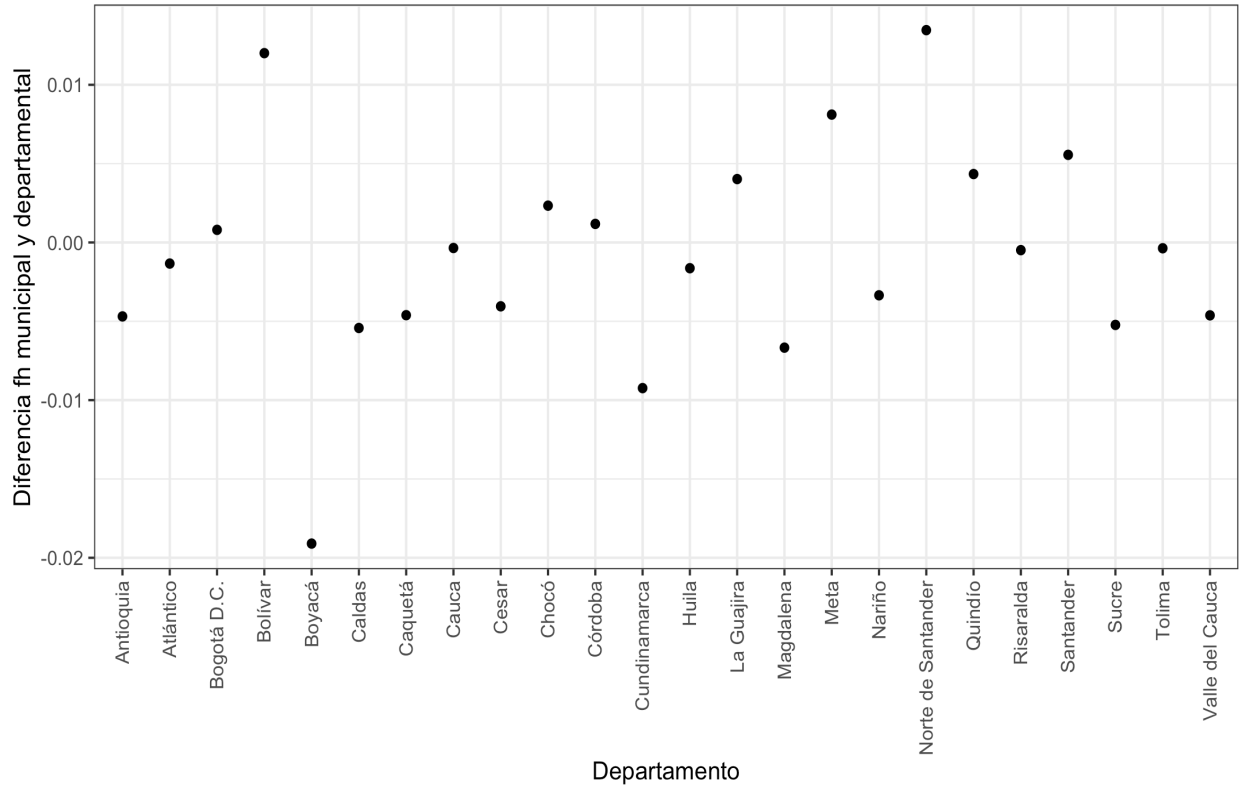
Fuente: DANE

4.5 Benchmark

Finalmente, con el fin de asegurar que las estimaciones municipales sean consistentes con las estimaciones departamentales, se realiza un proceso de ajuste (benchmarking) sobre las estimaciones obtenidas mediante el modelo de Fay–Herriot. Este procedimiento garantiza que, al agregar las estimaciones municipales ponderadas por el tamaño poblacional de cada municipio, se reproduzcan exactamente las estimaciones departamentales, de ciudades principales y nacional.

La Gráfica 11 presenta, para cada departamento, la diferencia entre la estimación departamental y la agregación de las estimaciones municipales posteriores al ajuste. Se observa que dichas diferencias son prácticamente nulas: **la mayor diferencia absoluta encontrada es de 0,01 puntos porcentuales que corresponde a Norte de Santander**, lo que evidencia una alta consistencia entre los niveles municipal y departamental.

Gráfica 11. Diferencia entre estimaciones municipales y departamentales después del benchmarking 2024

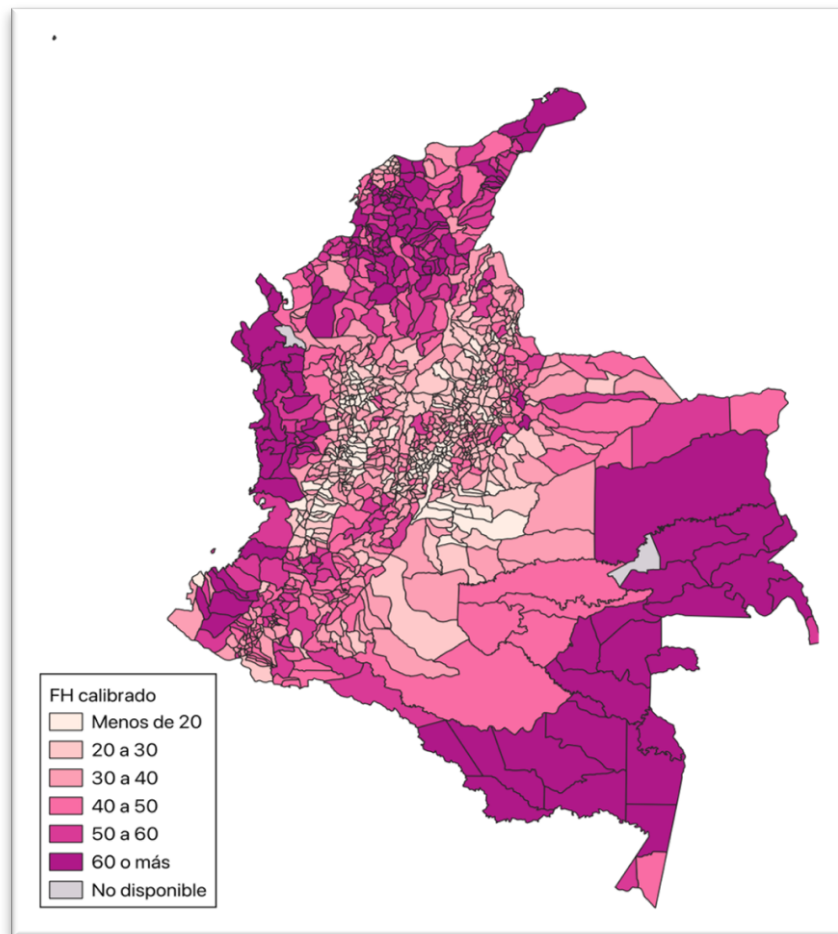


Fuente: DANE.

En este documento no se presentan los resultados del modelo de pobreza monetaria 2022, basado en que se aplico la misma metodología y pasos cumpliendo con los supuestos planteados similar que para el caso 2024.

4.6 Resultados

Gráfica 12. Pobreza Monetaria en hogares (%)



Fuente: DANE.

5 Pobreza multidimensional (IPM)

Este capítulo presenta la estimación de la Incidencia de Pobreza Multidimensional (IPM) en los hogares colombianos, un indicador que identifica múltiples privaciones simultáneas en dimensiones esenciales del bienestar, tales como educación, salud, trabajo, niñez y adolescencia y condiciones de la vivienda. Se describe la definición oficial del IPM utilizada por el DANE, las fuentes de información que sustentan su construcción, principalmente la Encuesta Nacional de Calidad de Vida (ECV) y las limitaciones de representatividad de sus estimaciones directas. Asimismo, se expone cómo la metodología de áreas pequeñas permite generar estimaciones consistentes y comparables a nivel municipal, complementando la información disponible con variables auxiliares provenientes de diversas fuentes administrativas y censales.

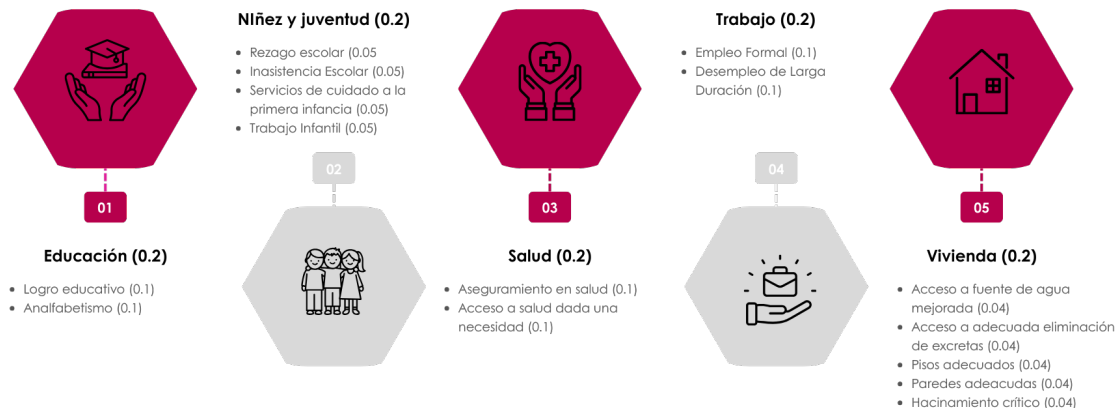
5.1 Definición del indicador y metodología oficial del DANE

El IPM es un indicador oficial del DANE que mide la pobreza a partir de la identificación de privaciones simultáneas en diversas dimensiones del bienestar de las personas y los hogares. A diferencia de la pobreza monetaria, que se basa exclusivamente en los ingresos, el IPM integra múltiples aspectos que afectan las oportunidades y el desarrollo humano. Su objetivo es capturar de manera más amplia y estructural las carencias que enfrentan los hogares colombianos, permitiendo una comprensión más integral de las condiciones que limitan su bienestar.

La metodología del IPM se basa en el enfoque Alkire–Foster, que identifica las privaciones a nivel de hogar, las pondera según su importancia relativa y determina si un hogar es pobre multidimensional cuando supera un umbral predeterminado de privaciones. Este método permite articular de manera clara cómo contribuye cada dimensión al indicador final.

A continuación, en la Ilustración 2 se presenta de manera sintética la estructura del IPM, mostrando las dimensiones que lo componen, los indicadores asociados y sus ponderaciones respectivas. El gráfico permite visualizar cómo se organiza el índice y cómo se distribuye el peso relativo de cada componente en el cálculo de la pobreza multidimensional en Colombia.

Ilustración 2. Variables y ponderaciones IPM



Una vez definidas las privaciones y sus ponderaciones, se calcula para cada hogar el nivel de privación ponderada, representado por M_i . Para ello, se suman las privaciones en las que incurre el hogar, multiplicadas por su peso correspondiente:

$$M_i = \sum_{j=1}^k w_j g_{ij}$$

Donde $g_{ij} = 1$ si el hogar está privado en el indicador j , y 0 en caso contrario; y w_j es el peso asignado a dicho indicador dentro del índice. Esta suma permite obtener una medida continua de la intensidad de privación del hogar.

Posteriormente, este valor se compara con un umbral de pobreza multidimensional establecido por la metodología oficial. Un hogar se clasifica como pobre multidimensional cuando:

$$M_i \geq k$$

es decir, cuando su nivel de privación ponderada alcanza o supera el umbral definido.

Finalmente, el IPM combina dos componentes fundamentales:

- H, la incidencia, que corresponde a la proporción de personas que viven en hogares clasificados como pobres multidimensionales;
- A, la intensidad, que es el promedio de privaciones ponderadas entre los hogares pobres.

El índice total se obtiene mediante la expresión:

$$IPM = H \times A$$

lo que permite capturar simultáneamente cuántas personas son pobres y qué tan pobres son, en términos de la acumulación de privaciones.

5.2 Estimaciones directas

Teniendo en cuenta que la estimación oficial del Índice de Pobreza Multidimensional (IPM) se realiza a partir de la información recolectada en la Encuesta de Calidad de Vida (ECV), se cuenta con estimaciones precisas y confiables a nivel nacional y departamental. Sin embargo, debido al diseño muestral de la encuesta, no es posible obtener estimaciones directas con suficiente precisión para la totalidad de los municipios del país.

Para avanzar hacia una medición municipal del IPM, se calcularon las estimaciones directas municipales empleando el estimador basado en el diseño probabilístico de la ECV, utilizando los factores de expansión y los componentes de varianza asociados al diseño estratificado y multietápico. Estas estimaciones constituyen el insumo inicial para la implementación del modelo SAE. Dado que en la mayoría de los municipios de la muestra de la ECV no es posible garantizar con los criterios de calidad que tiene el DANE para su publicación, la estimación municipal debe realizarse mediante un modelo de área.

Para garantizar que únicamente se incluyeran en el modelo de Fay–Herriot aquellas estimaciones municipales del IPM con precisión adecuada, se aplicaron los criterios de calidad definidos: tamaño de muestra efectivo mínimo, coeficiente de variación logarítmico acotado y efecto de diseño mayor o igual a uno. La aplicación de estos filtros permitió depurar el conjunto inicial de municipios con información recolectada.

Como resultado de ello, para la medición realizada en 2024, de los 617 municipios con estimación directa del IPM a un total de 191 municipios que serán incluidos dentro del proceso de modelamiento.

5.3 Aplicación del modelo SAE para IPM

Una vez construida la variable dependiente —el Índice de Pobreza Multidimensional (IPM) para cada municipio— y contando con el conjunto completo de variables auxiliares preparadas para esta aplicación, se llevó a cabo un proceso de selección de covariables con el fin de identificar aquellas que mejoran la capacidad predictiva del modelo. Para ello se empleó el algoritmo stepwise con criterio de información AIC, implementado en la librería bigstep de R. Este procedimiento permitió seleccionar un conjunto final de variables auxiliares relevantes para la estimación del IPM a nivel municipal.

Con la variable dependiente definida y las covariables seleccionadas, se estimó el modelo de área de Fay–Herriot utilizando la función fh de la librería emdi de R. El proceso se ejecutó con 500 iteraciones, garantizando estabilidad en la estimación de los parámetros y en las predicciones finales para cada uno de los municipios incluidos en el modelo.

Este proceso fue realizado de manera independiente para cada una de las vigencias y permitió obtener los siguientes resultados:

5.3.1 IPM 2024

A continuación, se presenta las covariables que permitieron obtener el modelo empleado en el proceso de estimación del IPM en la vigencia 2024, junto con los coeficientes estimados y su respectivo error estándar y significancia estadística, el cuál da como resultado un modelo con R^2 de 90%.

Tabla 7. Estimación modelo FH – IPM 2024

Variable	Coefficiente (s.e.)
Intercepto	-0.66 (0.27).
Hogares con jefe de hogar bachiller	-0.21 (0.09).
Departamento: Amazonas	-0.19 (0.05)**
Promedio de personas por hogar	0.11 (0.03)**
Hogares en vivienda tipo cuarto	0.27 (0.07)**
Privación cocina - déficit cualitativo de vivienda	0.68 (0.15)**
Hogares con recolección de basuras	0.15 (0.05)*
Departamento: Vichada	0.39 (0.04)**
Cobertura educación secundaria	-0.18 (0.05)**
Promedio de unidades económicas	0.03 (0.01)**
Eventos en cultivos UPAS	0.11 (0.05).
Área agrícola	-0.03 (0.01)*
Departamento: Córdoba	0.11 (0.03)**
Número de afiliados a salud	1.8 (0.35)**
Departamento: Guainía	0.21 (0.05)**
Departamento: Putumayo	-0.18 (0.03)**
Cobertura de árboles	0.11 (0.04)*
Eventos de minas antipersonales	-0.05 (0.02).
Hogares con servicio de gas	-0.2 (0.03)**
Presencia de grupos armados	0.03 (0.01).
Hogares en Nivel A de Sisben	1.33 (0.43)*
Sedes de educación media	-0.48 (0.22).
Departamento: Sucre	0.09 (0.03)*
Número de afiliados hombres	0.69 (0.33).
Área de nevados,	-5.49 (2.36).
Departamento: Cesar	-0.09 (0.04)*
Sedes educativas rurales	0 (0)**
Rural disperso	0.07 (0.02)**
Medición de Desempeño Municipal	-0.16 (0.06)*
Hogares con Sisben	-0.89 (0.35).

5.4 Validación de supuestos del modelo

Tasa de homicidios	0 (0)*
Departamento: Guaviare	0.06 (0.04)
Tasa de amenazas contra la población	0 (0).
Sedes educativas de secundaria	0.41 (0.14)*
Proyección personas entre 5 y 17 años	0.81 (0.29)*
Departamento: Valle del Cauca	-0.13 (0.03)**
Área construida	-0.26 (0.1).
Número de docentes en área rural	0 (0).
temperatura	0 (0)
Problemas estructurales en cultivos	0.18 (0.07).
Promedio de animales por UPA	-0.29 (0.06)**
Promedio de ganado por UPA	0.12 (0.05).
Promedio de área con pasto	0.27 (0.1)*
Promedio de problemáticas en cultivos	-0.05 (0.04)
Hogares en nivel C de Sisben	0.8 (0.82)

Fuente: DANE.

Respecto al análisis diagnóstico de los residuales, a continuación, se reporta la medición de los diferentes estadísticos de prueba para validar el comportamiento de los residuales estandarizados y los efectos aleatorios, obteniendo un valor p en el test de Shapiro del 0.91 y 0.4 respectivamente.

Tabla 8. Estimación modelo FH – IPM 2024

	Skewness	Kurtosis	Shapiro W	Shapiro p
Residuales estandarizados	-0.1895222	3.040634	0.9962113	0.9194771
Efectos aleatorios	-0.1587826	3.597044	0.9922763	0.4077175

Fuente: DANE.

5.4 Validación de supuestos del modelo

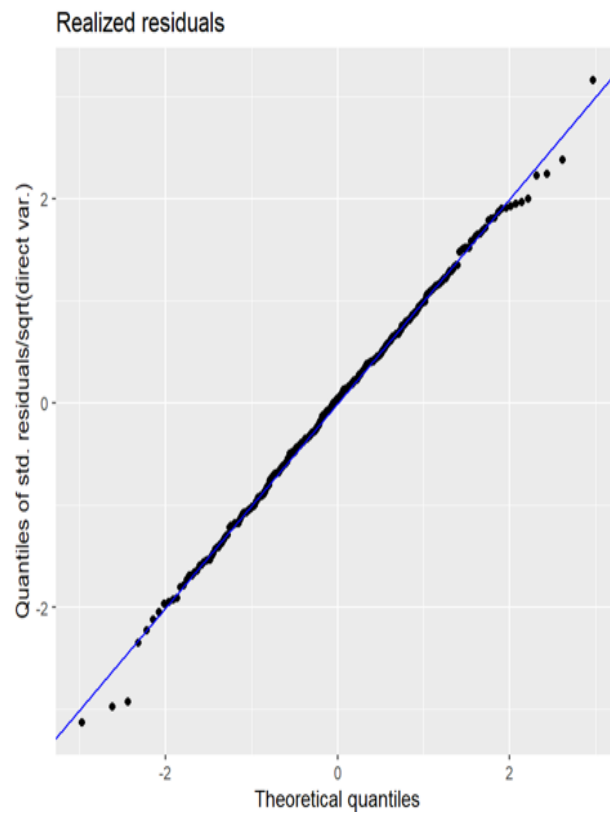
El modelo debe cumplir con los supuestos de:

1. Normalidad en los residuales
2. Normalidad en los efectos aleatorios
3. Media cero en los residuales
4. Independencia en los residuales

5.4.1 Normalidad en los residuales

De acuerdo con la segunda parte de la Tabla 5 se observa que el valor p de la prueba de hipótesis de Shapiro-Wilk para los residuales es de 0.91, es decir, que no hay evidencia estadística suficiente para rechazar la hipótesis de que los residuales se distribuyen bajo una distribución normal. Adicionalmente, la gráfica cuantil cuantil plot para la validación visual de la distribución confirma que la distribución normal es razonable para los residuales:

Gráfica 13. Gráfica cuantil cuantil para los residuales del modelo.

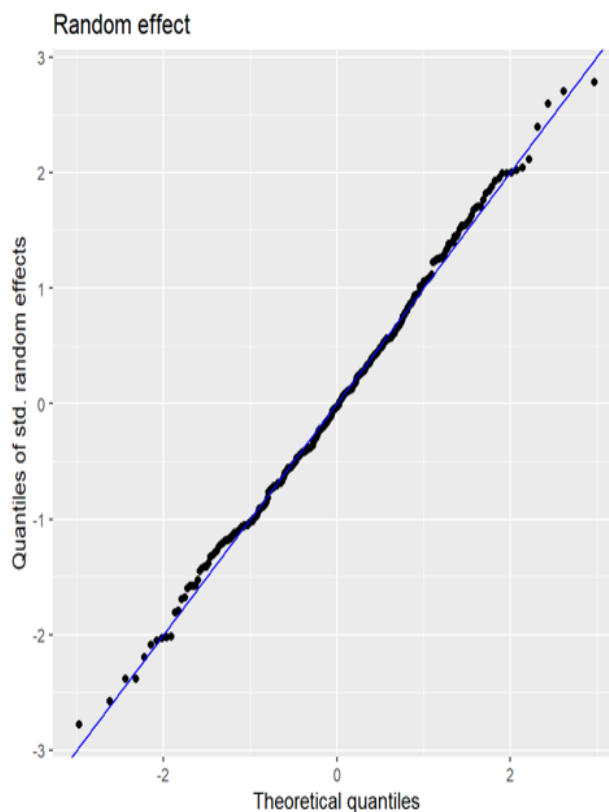


Fuente: DANE

5.4.2 Normalidad en los efectos aleatorios

Similar a los residuales, el valor p de la prueba de hipótesis de Shapiro-Wilk para los efectos aleatorios es de 0.4077, es decir, que no hay evidencia estadística suficiente para rechazar la hipótesis de que los efectos aleatorios se distribuyen bajo una distribución normal, la misma conclusión se puede obtener observando la gráfica cuantil cuantil presentada a continuación:

Gráfica 14. Gráfica cuantil cuantil para los efectos aleatorios del modelo.



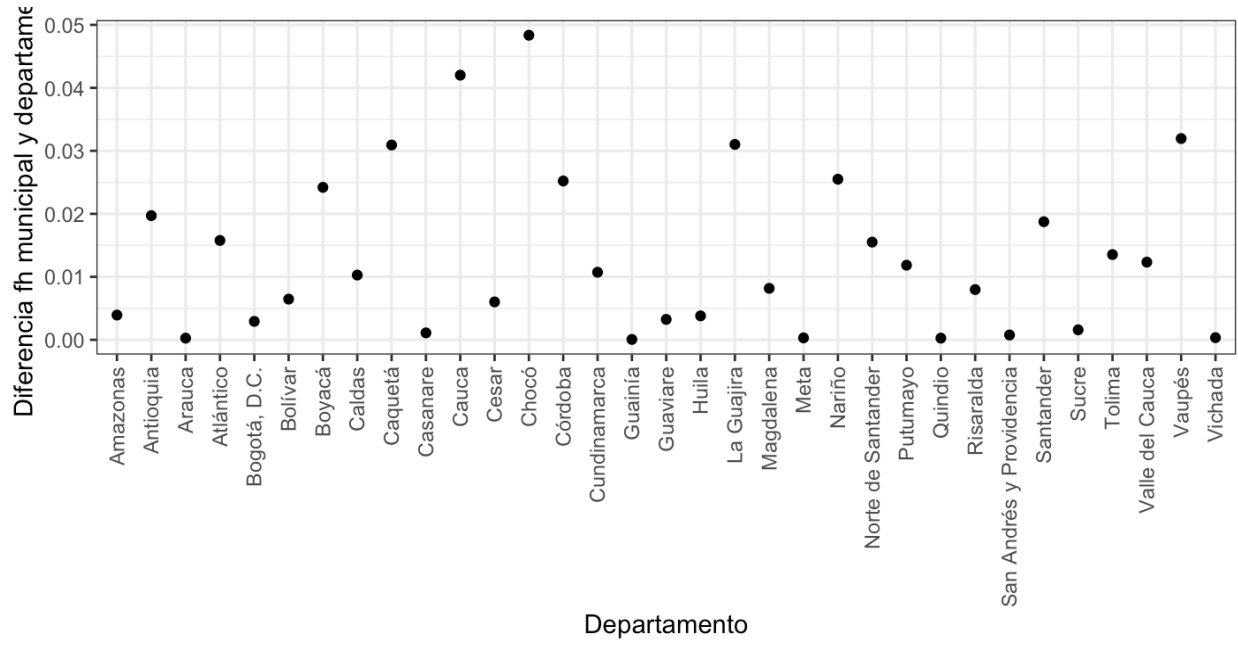
Fuente: DANE

5.5 Benchmark

Finalmente, con el fin de asegurar que las estimaciones municipales sean consistentes con las estimaciones departamentales, se realiza un proceso de ajuste (benchmarking) sobre las estimaciones obtenidas mediante el modelo de Fay–Herriot. Este procedimiento garantiza que, al agregar las estimaciones municipales ponderadas por el tamaño poblacional de cada municipio, se reproduzcan exactamente las estimaciones departamentales.

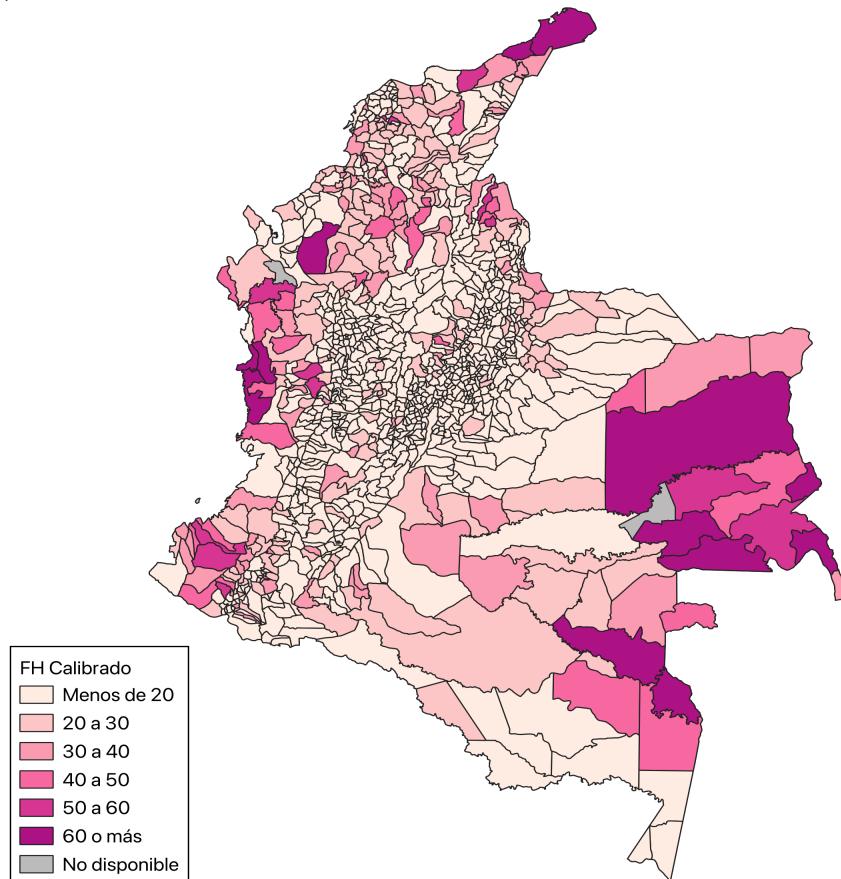
La siguiente gráfica presenta, para cada departamento, la diferencia entre la estimación departamental y la agregación de las estimaciones municipales posteriores al ajuste de la vigencia 2024. Se observa que dichas diferencias son prácticamente nulas: **la mayor diferencia absoluta encontrada es de 0,048 puntos porcentuales en el departamento de Chocó**, lo que evidencia una alta consistencia entre los niveles municipal y departamental.

Gráfica 15. Diferencia entre estimaciones municipales y departamentales después del benchmarking 2024



5.6 Resultados

Gráfica 17. IPM en hogares (%) 2024



Los resultados son presentados en POWER BI publicado en la página web del DANE para la vigencia 2022 y 2024.

6 Bibliografía

- DANE (2025), Boletín técnico Escala de experiencia de inseguridad alimentaria (FIES) 2024, <https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/escala-de-experiencia-de-inseguridad-alimentaria-fies-2024>
- Gutiérrez, Andrés (2016), Estrategias de muestreo, Ediciones de la U.
- Gutiérrez, Zhang y Rodríguez (2016), The Performance of Multivariate Calibration on Ratios, Means and Proportions, <https://repositorio.unal.edu.co/handle/unal/66515>.
- Khalil Clara Aida (2023), presentación de aplicación de *Small Area Estimation of Food Insecurity based on FAO's Food Insecurity Experience Scale*, <https://www.cepal.org/sites/default/files/presentations/sem-regional-metodologias-sae-fao-aidakhalil.pdf>
- Molina, Isabel (2019), Desagregación de datos en encuestas de hogares, CEPAL, - Khalil Clara Aida (2023), presentación de aplicación de *Small Area Estimation of Food Insecurity based on FAO's Food Insecurity Experience Scale*, <https://www.cepal.org/sites/default/files/presentations/sem-regional-metodologias-sae-fao-aidakhalil.pdf>
- Rao, J.N.K. y Molina (2015), Small area estimation, Second Ed., Hoboken, NJ: Wiley.
- You, Yong y Hidioglou, Mike (2023), Application of Sampling Variance Smoothing Methods for Small Area Proportion Estimation, Journal of Official Statistics, Vol. 39, No. 4, 2023, pp. 571–590 <https://intapi.sciendo.com/pdf/10.2478/jos-2023-0026>