



DEPARTAMENTO ADMINISTRATIVO NACIONAL DE
ESTADÍSTICA - DANE

NOTA METODOLÓGICA

**ESTIMACIÓN DE LA
PREVALENCIA DE LA
INSEGURIDAD ALIMENTARIA
MODERADA O GRAVE EN
COLOMBIA DURANTE EL AÑO
2022 A NIVEL MUNICIPAL**

FEBRERO 2024

B. Piedad Urdinola Contreras
Directora

Leonardo Trujillo Oyola
Subdirector

Edna Patricia Rangel Barragán
Secretaria General

Jefes Oficinas Asesoras

Claudia Díaz Hernández
Oficina Asesora de Planeación

Maria Fernanda de la Ossa Archila
Oficina Asesora Jurídica

Luis Martín Barrera Pino
Oficina de Sistemas

Duvy Johanna Plazas Socha
Oficina de Control Interno

Angela Patricia Santiago Enciso
Oficina de Control Interno Disciplinario

Directores Técnicos

Andrea Ramírez Pisco
Dirección de Metodología y Producción Estadística

Andrés Felipe Ortiz Rico
Dirección de Recolección y Acopio

Javier Sebastián Ruiz Santacruz
Dirección de Censos y Demografía

Juan Pablo Cardoso Torres
Dirección de Síntesis y Cuentas Nacionales

Julieth Alejandra Solano Villa
Dirección de Regulación, Planeación,
Estandarización y Normalización

Elkin Ernesto Ramírez Niño
Dirección de Geoestadística

Álvaro Enrique Duque Soto
Dirección de Difusión y Cultura Estadística

Grupo de trabajo

Subdirección

Natalia Ximena Arteaga Gutiérrez
Hanwen Zhang
Juan José Rubio Mesa
Yury Vanessa Ochoa Montes

Dirección de Metodología y Producción Estadística

Carlos Alberto Hernández Lozano
Santiago Andrés Barragán López

Dirección de Geoestadística

Carlos Alberto Durán Gil
Gustavo Adolfo García Velez

Dirección de Recolección y Acopio

Gina Fernanda Otalora Peña

Edición y diseño

Dirección de Difusión y Cultura Estadística

© DANE, 2024

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

TABLA DE CONTENIDO

Tabla de contenido

1	Introducción	3
2	La Escala de Experiencia de Inseguridad Alimentaria (FIES)	3
3	Estimaciones directas	5
4	Estimación en áreas pequeñas	5
5	Metodología	6
5.1	Estimación directa - Experiencia de inseguridad alimentaria grave o moderada	7
5.2	Función Generalizada de Varianza	7
5.3	Fuentes de información	8
5.4	Modelo Fay-Herriot	9
5.5	Validación de supuesto del modelo	12
5.6	Consistencia	14
5.7	Benchmark	15
5.8	Validación y presentación de resultados	17
6	Bibliografía	18

1 Introducción

La Escala de Experiencia de Inseguridad Alimentaria (*Food Insecurity Experience Scale*, FIES) es un instrumento de evaluación desarrollado por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) con el propósito de medir el acceso a los alimentos en términos de cantidad y calidad. Esta herramienta emplea ocho preguntas para recopilar la experiencia de los hogares frente a la inseguridad alimentaria, brindando así una medición de la gravedad de esta situación. Estas preguntas fueron incluidas en la Encuesta Nacional de Calidad de Vida (ECV) 2022, permitiendo así una evaluación integral de la experiencia alimentaria en los hogares residentes del territorio nacional.

La ECV constituye una herramienta esencial para comprender y comparar las condiciones socioeconómicas de los hogares a nivel nacional. ECV es una encuesta probabilística estratificada en varias etapas que permite realizar inferencia de los indicadores medidos en este instrumento. Además, fue diseñada para la publicación de resultados a nivel nacional, urbano-rural y de los treinta y tres (33) departamentos del país.

Este informe presenta la metodología aplicada y resultados de la estimación de la prevalencia de la inseguridad alimentaria moderada o grave en Colombia durante el año 2022 a nivel municipal. Estos resultados son producto del trabajo conjunto que durante el año 2023 el DANE adelantó junto con el equipo técnico de la FAO (Food and Agriculture Organization, por sus siglas en inglés).

Actualmente el DANE realiza la estimación de FIES con el fin de reportar al indicador *2.1.2 Prevalencia de inseguridad alimentaria moderada o grave en la población, del ODS 2. Hambre cero* hasta el nivel departamental, teniendo en cuenta los dominios de desagregación definidos por el diseño muestral de la encuesta. Durante el proceso de análisis de los resultados se identificó la necesidad de obtener información más detallada a nivel municipal, lo cual no era posible a través de estimaciones directas de la encuesta, es decir, haciendo uso de inferencia basada en el diseño dada por el diseño de muestreo probabilístico implementado en la ECV.

Para abordar esta necesidad, DANE y FAO de forma colaborativa optan por utilizar la técnica de Estimación en Áreas Pequeñas (SAE, por sus siglas en inglés). Esta técnica hace uso de estimadores basados en modelos, los cuales usan información auxiliar con el fin de mejorar la precisión de las estimaciones. De forma específica en esta aplicación se implementó un modelo de área definido por Fay y Herriot (1979), que hace uso de las estimaciones directas e información auxiliar al nivel de los dominios de interés. Para la estimación del modelo Fay-Herriot se utilizó información auxiliar de diferentes fuentes de información a nivel municipal provenientes de censos, registros administrativos e imágenes satelitales.

En la primera sección se ofrece una descripción detallada de la Escala FIES y su metodología de cálculo. Allí se destacan aspectos esenciales que influyeron en la construcción de la escala a nivel municipal. La segunda sección presenta conceptos sobre la desagregación de variables a través de estimadores directos. La tercera sección presenta conceptos sobre la metodología de estimación de áreas pequeñas SAE. En la cuarta sección se presenta el desarrollo metodológico del modelo implementado para la estimación de FIES a nivel municipal y la evaluación del modelo empleado. En la última sección, se presentan los resultados de la estimación municipal de la inseguridad alimentaria moderada o grave FIES.

2 La Escala de Experiencia de Inseguridad Alimentaria (FIES)

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 2022.

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.

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.¹

¹<https://www.dane.gov.co/files/operaciones/FIES/bol-FIES-2022.pdf>

3 Estimaciones directas

Las estimaciones directas son aquellas que se realizan siguiendo el diseño de muestreo propuesto para la encuesta sin usar otro tipo de información auxiliar y para las cuales es necesario validar las medidas de incertidumbre como varianzas, intervalos de confianza, coeficientes de variación (entre otros) con el fin de determinar la calidad de la estimación.

Uno de los estimadores más usados para las encuestas probabilísticas es el estimador de Horvitz-Thompson (HT), el cual es un estimador insesgado con respecto al diseño muestral definido, como:

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

Donde π_k representa las probabilidades de inclusión de cada uno de los elementos de la población a la muestra seleccionada a partir del diseño de muestreo aplicado.

La estimación directa debe estar acompañada de los parámetros de calidad con el fin de definir la validez de la estimación. Estos pueden ser definidos como: error estándar, margen de error, intervalo de confianza, coeficiente de variación, efecto de diseño (DEFF), entre otros. Todos estos derivados de la estimación de la varianza del estimador, que en el caso del estimador de HT, está definido, como:

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

Donde $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$, siendo π_{kl} la probabilidad conjunta de que el elemento k y l estén en la muestra seleccionada.

Para la desagregación de variables a través de encuestas por muestreo existen dos tipos de grupos: i) **los estratos** son subgrupos de la población, excluyentes entre sí, estos son definidos desde el diseño muestral por lo cual se controla el tamaño de muestra y error muestral de cada estrato desde el inicio de la operación, asegurando la calidad de las estimaciones a este nivel de desagregación, y ii) **los dominios** que son similares a los estratos y corresponden a subgrupos poblacionales, sin embargo, no están definidos desde el diseño de la encuesta por lo cual el tamaño de muestra y el error muestral depende del comportamiento de la recolección de información, en algunos casos será viable la estimación a este nivel de acuerdo a la calidad de las estimaciones directas para el dominio.

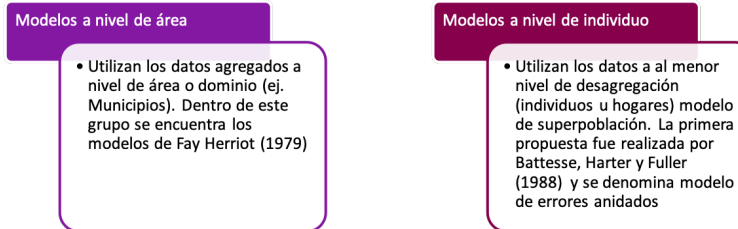
4 Estimación en áreas pequeñas

La metodología de estimación en áreas pequeñas es usada cuando a través de las estimaciones directas no es posible obtener resultados para desagregar a los dominios de interés dado que se tiene un tamaño de muestra pequeño o nulo. Generalmente, los dominios están definidos como áreas geográficas (departamentos, municipios, veredas) o grupos poblacionales (grupos étnicos, LGBTIQ+, víctimas, entre otros.).

Para la aplicación de SAE se hace uso de estimadores indirectos los cuales consideran la información de la encuesta, variables auxiliares de otras fuentes de información y la relación existente con otras áreas o dominios. De acuerdo con CEPAL la ganancia en eficiencia de los estimadores indirectos respecto de los directos es mayor cuanto menor es el tamaño de muestra del dominio. Al usar estimadores indirectos se está sacrificando el insesgamiento de los estimadores basados en el diseño y es necesario escoger un modelo adecuado que se ajuste correctamente a los datos.

Existen diferentes modelos para la estimación de áreas pequeñas, estos se agrupan en modelos de área y modelos de unidad, definidos en la siguiente ilustración.

Ilustración 2. Modelo conceptual FIES



Fuente: DANE

5 Metodología

Para la desagregación del nivel de inseguridad alimentaria moderada o grave a nivel municipal FIES aplicando la metodología de estimación en áreas pequeñas, se llevó a cabo en los siguientes pasos:

1. Estimaciones directas a nivel municipal, haciendo uso de la Encuesta de Calidad de Vida.
2. Suavizamiento de la estimación de la varianza del estimador directo
3. Construcción de variables auxiliares
4. Aplicación del modelo de área - Fay-Herriot
5. Análisis de consistencia de las estimaciones
6. Validación de supuestos del modelo
7. Benchmark - alineación con las estimaciones directas a nivel departamental
8. Presentación de resultados

Al aplicar los pasos listados, se obtuvieron los siguientes resultados para la aplicación de FIES:

Ilustración 3. Resumen resultados estimación modelo SAE



Fuente: DANE

5.1 Estimación directa - Experiencia de inseguridad alimentaria grave o moderada

5.1 Estimación directa - Experiencia de inseguridad alimentaria grave o moderada

La Encuesta Nacional de Calidad de Vida 2022 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 640 de 1.122 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 640 municipios eso no implica que cumplan con los criterios de calidad para su publicación.

Como se ha mencionado para la estimación del nivel de inseguridad alimentaria moderada o grave se utilizó un modelo de *Rash*, el cual 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, $neff = n/def$. 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 428 municipios con estimaciones directas de calidad para ser usadas en el modelo de área.

5.2 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 suavizamiento² a las estimaciones de las varianzas a nivel de los dominios de interés (municipios).

²los métodos de suavizamiento consisten en la reducción del ruido y datos atípicos de un conjunto de datos

5.3 Fuentes de información

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. En la siguiente tabla se encuentra los resultados de estimación del modelo log-lineal:

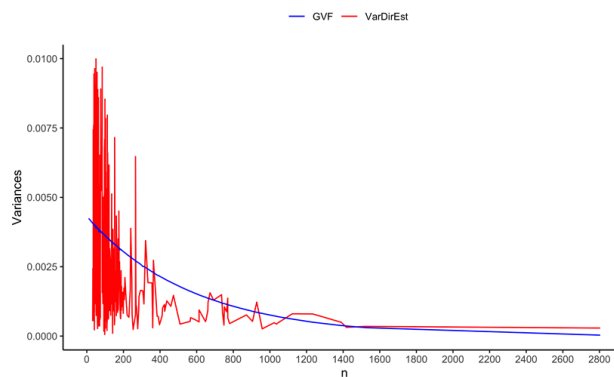
Tabla 2. Modelo función generalizada de varianza

Variable	Coefficiente
Intercepto	-5,800 (0,0457)***
Proporción de la muestra sobre la población	-0,3523 (5,0400)
Tamaño de muestra	-0,0017 (0,0002)***
R ²	0,17

Fuente: DANE.

En la siguiente gráfica se tiene en color rojo las estimaciones de la varianza del estimador directo para los 640 municipios en la muestra y el color azul el suavizamiento de la varianza, con el fin de comparar el comportamiento de las estimaciones de la varianza del estimador contra el tamaño de muestra de cada uno de los dominios en la muestra.

Gráfica 2. Función generalizada de Varianza



Fuente: DANE.

Encontrando una función de varianza relacionada con el tamaño de muestra, a mayor tamaño de muestra menor varianza para la estimación, evitando que el error estándar resultante a nivel municipal esté influenciado por el ruido inicial de las estimaciones directas municipales.

5.3 Fuentes de información

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

5.4 Modelo Fay-Herriot

1. **Estadísticas Vitales - DANE**, esta fuente permite analizar tasas de mortalidad, causas de muerte y análisis por ciclo de vida.
2. **Censo Nacional de Población y Vivienda - DANE**, esta fuente contiene información sociodemográfica del total de hogares y vivienda en el territorio nacional.
3. **Censo Nacional Agropecuario - DANE**, esta fuente contiene información sobre la producción y variables socioeconómicas de los hogares, viviendas, Unidades de Producción Agropecuaria UPA y no agropecuarias UPNA
4. **SIPSA - Sistema de Información de Precios - DANE**, variables relacionadas con la información de abastecimiento de los centros mayoristas del país y su municipio de origen.
5. **Educación formal (EDUC) - DANE**, esta fuente de información contiene la información de establecimientos educativos públicos y privados del país.
6. **Valor agregado municipal - DANE** información económica por actividad económica a nivel municipal.
7. **SISPRO - Ministerio de Salud y Protección Social** es el Sistema Integrado de Información de la Protección Social con indicadores a nivel municipal de cobertura, infraestructura y resultado.
8. **SIPRA - Unidad de Planificación Rural Agropecuaria (UPRA)** este sistema de información recolecta información de distintas entidades sobre producción agropecuaria, EVAS y recursos ambientales.
9. **Terridata - Departamento Nacional de Planeación (DNP)** reportes de información a nivel de entidad territorial con información de planeación, finanzas, gestión del riesgo y desempeño municipal.

La construcción de variables auxiliares a través de las fuentes de información listadas, se hizo a partir de los ejes y determinantes de la seguridad alimentaria y nutricional contempladas en la política nacional de SAN (Conpes 113 de 2008):

- Disponibilidad
- Acceso
- Consumo
- Aprovechamiento y utilización
- Calidad e inocuidad

Es decir, buscando que cada variable hiciera parte de cada uno de los ejes o determinantes de la seguridad alimentaria y nutricional, ya sea de forma directa ya sea de forma indirecta, por aproximación o relación. Construyendo un total de 215 variables a nivel municipal para el total de los 1.122 municipios del país, en el caso de tener información faltante en alguna de las variables para un valor menor al 10% de los municipios, se imputo a través de un modelo de K-vecinos cercanos usando las variables de la base de datos.

5.4 Modelo Fay-Herriot

Como ya se ha mencionado en este documento para poder incluir la varianza del modelo Rasch en la aplicación de SAE, y teniendo en cuenta la disponibilidad de la información es necesario aplicar un modelo de área: el modelo de Fay-Herriot, el cual fue desarrollado por Robert E. Fay y Roger A. Herriot en 1979. Este modelo corresponde a un modelo lineal mixto que incluía efectos aleatorios a nivel de área, y asocia la estimación directa del parámetro de interés del área d , δ_d con el parámetro de la siguiente forma:

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

5.4 Modelo Fay-Herriot

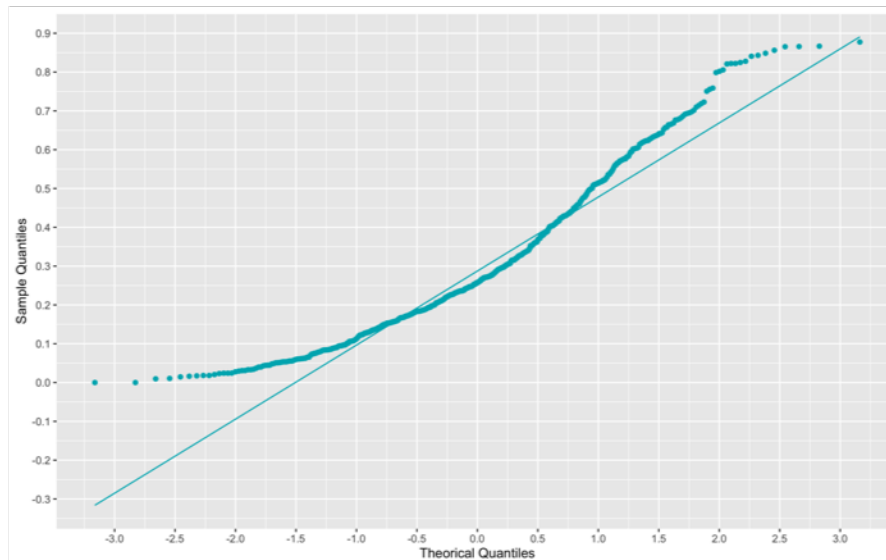
donde los ϵ_d son los residuales del modelo con $\epsilon_d \sim N(0, \sigma_d^2)$. A su vez, el parámetro del área d corresponde a:

$$\delta_d = X'_d \beta + u_d$$

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

Con el fin de cumplir con los supuestos del modelo sobre la distribución de los residuales y efectos aleatorios, y de garantizar que las estimaciones se encuentren en el soporte 0 y 1 se aplicó la transformación de arco seno a la variable de interés, es decir, la estimación de inseguridad alimentaria moderada o grave a nivel municipal. Esto teniendo en cuenta que la distribución de esta variable no sigue una distribución normal, como se puede ver en la siguiente gráfica:

Gráfica 3. Normalidad inseguridad alimentaria moderada o grave municipal



Fuente: DANE.

A partir de esta transformación y teniendo el total de variables construidas para esta aplicación se procedió a la selección del grupo de variables que mejoran la predicción de FIES municipal, para este caso se aplicó el algoritmo *stepwise* usando como criterio de selección *AIC* de la librería *bigstep* de R, a partir de un modelo de regresión lineal, obteniendo una selección de 30 variables, las cuales fueron usadas para la estimación municipal de la inseguridad alimentaria moderada o grave a partir de un modelo Fay-Herriot a partir de la función *fh* de la librería *emdi* de R. Este modelo se realizó a partir de 500 iteraciones usando una transformación de arco seno, de lo cual se obtienen las siguientes estimaciones:

Tabla 3. Estimación modelo FH

Variable	Coefficiente (s.e.)
Intercepto	-2.12769 (0.55236)***
Problemática en cultivos	0.43941 (0.15525)**
Departamento - Huila	-0.20736 (0.04154)***
Elevación del suelo	-0.23656 (0.0509)***
Coordenadas - longitudud	-0.03639 (0.00572)***

Nota metodológica

5.4 Modelo Fay-Herriot

Porcentaje de suelo desnudo	0.06347 (0.02343)**
Productividad otro cultivo	-0.02112 (0.01032)*
Cobertura de vacunación	0.13051 (0.07642).
Variación cultivos frutas y hortalizas	0.42473 (0.16164)**
Puntaje fiscal	-0.28928 (0.11866)*
Departamento - Amazonas	-0.1368 (0.07037).
Aseguramiento en salud	0.7746 (0.23373)***
Número de nacidos per cápita - Cuartil 5	-0.03702 (0.01606)*
Producción frutas y hortalizas	-0.07771 (0.01964)***
Provación paredes al cuadrado	1.28012 (0.34036)***
Departamento - Arauca	0.22431 (0.05286)***
Déficit cualitativo de vivienda	0.0015 (0.00057)**
Departamento - San Andrés	-0.8332 (0.16349)***
Producción cereales	0.03715 (0.02135).
Departamento - Norte de Santander	0.06945 (0.04021).
Departamento - Atlántico	0.18577 (0.04818)***
Índice de vulnerabilidad de cambio climático	-0.06771 (0.03039)*
Departamento-Caquetá	-0.12078 (0.04279)**
Privación de rezago escolar	0.35917 (0.20587).
Proporción de UPA con problemas de sequía	0.34356 (0.10498)**
Presencia de grupos armadas	-0.02965 (0.01637).
Proporción de hogares con nivel de Sisbén C	0.35 (0.20521).
Departamento - Sucre	0.16839 (0.04001)***
Municipio rural - Misión de Transformación del Campo	0.06632 (0.0208)**
Municipio rural disperso - Misión de Transformación del Campo	0.07985 (0.02624)**
Cuartil 2 - personas con discapacidad	0.03365 (0.01348)*
Departamento - La Guajira	0.15544 (0.04616)***
Tasa Mortalidad Infantil	-0.07992 (0.02803)**
Promedio de cultivos	0.41934 (0.10557)***
Promedio de pastos	0.75113 (0.2515)**
Cultivos para ventas	-0.18241 (0.06343)**
Privación inadecuada eliminación de excretas	0.17024 (0.07279)*
Departamento - Antioquia	-0.13924 (0.03478)***
proporción problemáticas cultivos	-0.49521 (0.15907)**
Departamento - Caldas	-0.17601 (0.03944)***
Privación material inadecuado de pisos	-0.2233 (0.07292)**
Departamento - Guainia	0.16301 (0.07065)*
Porcentaje de cobertura de suelo de cultivos	-0.02608 (0.00978)**
Departamento - Córdoba	0.11838 (0.04294)**
Departamento - Santander	-0.0855 (0.03441)*
Abastecimiento - tuberculos	0.02926 (0.01073)**
Unidades económicas	0.18451 (0.06229)**
Tasa de hurtos-2	-0.00501 (0.00242)*
Departamento - Tolima	0.08451 (0.03303)*
Proporción hogares - Servicio de internet	-0.64227 (0.17649)***
Proporción hogares estrato 4	-0.9646 (0.34476)**
Departamento - Quindío	0.11204 (0.0594).
IRCA	0.01241 (0.00585)*
Total de áreas de concentración económica rural	-0.02377 (0.00878)**
Tasa de homicidios	-0.02408 (0.01113)*
Luces nocturnas	-0.04672 (0.02215)*
Porcentaje del área correspondiente a títulos mineros	0.17631 (0.0966).
Medición de desempeño municipal	0.00582 (0.00364)
Longitud de vías	0.00836 (0.00469).

	Skewness	Kurtosis	Shapiro_W	Shapiro_p
Standardized_Residuals	-0.0910102	3.081794	0.9971304	0.6620053

5.5 Validación de supuesto del modelo

Random_effects	0.1916442	2.997624	0.9949270	0.1730914
----------------	-----------	----------	-----------	-----------

Fuente: DANE.

5.5 Validación de supuesto del modelo

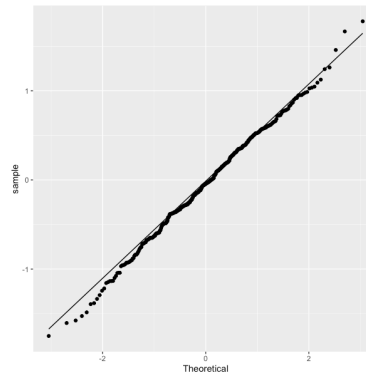
El supuesto debe cumplir con los supuestos de:

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

5.5.1 Normalidad en los residuales

De acuerdo con la segunda parte de la Tabla 3 se observa que el valor p de la prueba de hipótesis de Shapiro-Wilk para los residuales es de 0.173, 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 4. Gráfica cuantil cuantil para los residuales del modelo.



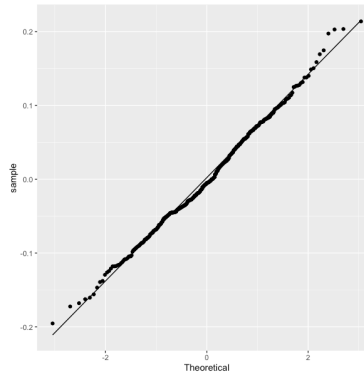
Fuente: DANE

5.5.2 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 0.662, 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 5. Gráfica cuantil cuantil para los efectos aleatorios del modelo.

5.5 Validación de supuesto del modelo

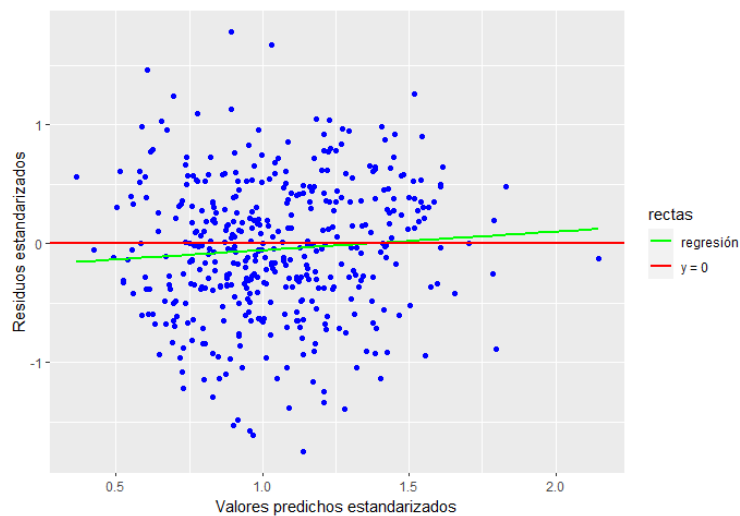


Fuente: DANE

5.5.3 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 6. Gráfica de dispersión de los residuales estandarizados y valores predichos estandarizados



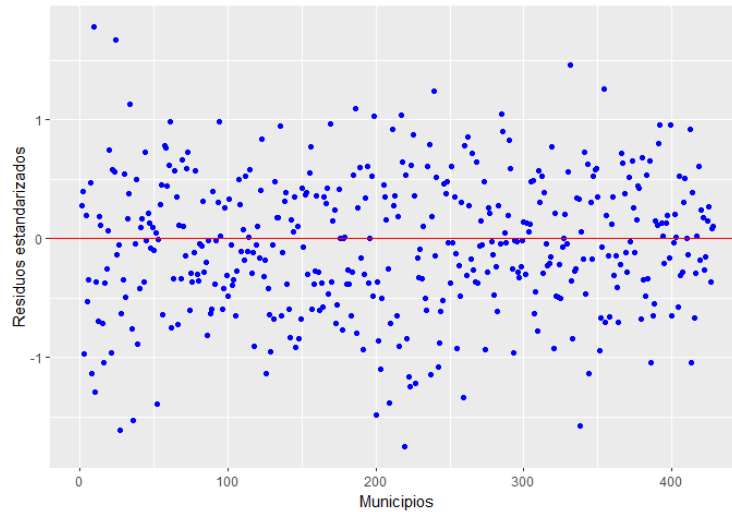
Fuente: DANE

5.5.4 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 7. Independencia de los residuales

5.6 Consistencia

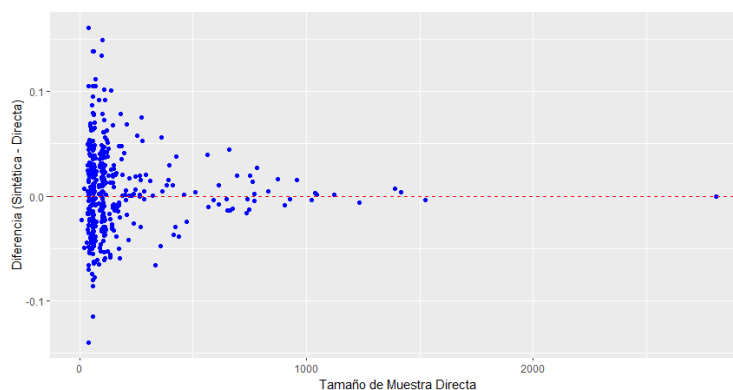


Fuente: DANE

5.6 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 8. Diferencia entre las estimaciones directas y de Fay-Herriot

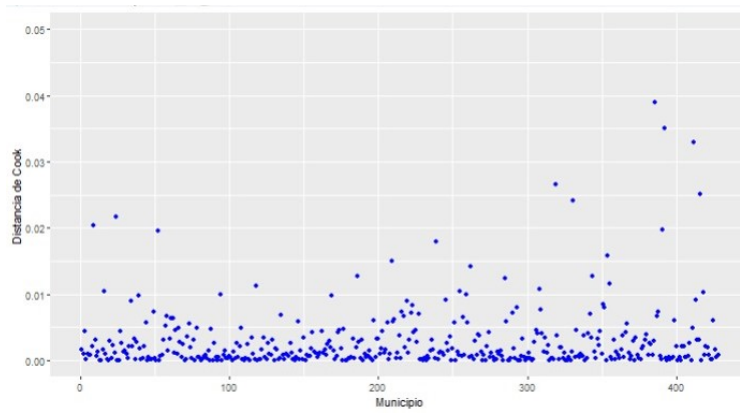


Fuente: DANE

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 9. Distancia de Cook para detectar municipios incluyentes en el modelo

5.7 Benchmark



Fuente: DANE

5.7 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 4. Diferencias entre las estimaciones directa y modelo FH a nivel departamental

Departamento	Estimación directa	Estimación FH	Diferencia
Antioquia	0.209	0.225	0.016
Atlántico	0.461	0.459	0.002
Bogotá, D.C.	0.254	0.254	0.000
Bolívar	0.326	0.345	0.019
Boyacá	0.207	0.224	0.017
Caldas	0.146	0.146	0.000
Caquetá	0.272	0.272	0.000
Cauca	0.223	0.250	0.027
Cesar	0.411	0.416	0.005
Córdoba	0.389	0.431	0.042
Cundinamarca	0.230	0.237	0.007
Chocó	0.432	0.446	0.014
Huila	0.190	0.197	0.007
La Guajira	0.597	0.600	0.003
Magdalena	0.453	0.424	0.029
Meta	0.255	0.270	0.015
Nariño	0.371	0.381	0.010
Norte de Santander	0.258	0.303	0.045
Quindio	0.173	0.190	0.017
Risaralda	0.175	0.206	0.031
Santander	0.205	0.206	0.001
Sucre	0.479	0.481	0.002
Tolima	0.274	0.303	0.029
Valle del Cauca	0.280	0.285	0.005

5.7 Benchmark

Arauca	0.393	0.391	0.002
Casanare	0.282	0.323	0.041
Putumayo	0.244	0.259	0.015
Archipiélago de San Andrés y Providencia	0.172	0.155	0.017
Amazonas	0.186	0.173	0.013
Guainía	0.298	0.292	0.006
Guaviare	0.304	0.309	0.005
Vaupés	0.365	0.375	0.010
Vichada	0.307	0.324	0.017

Fuente: DANE

Siendo el promedio de las diferencias a nivel departamental de 1.4 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.

El objetivo con el procedimiento *benchmark* es llevar las agregaciones de municipios a la cifra oficial reportada para el nivel nacional y departamental, con dos propósitos: i) mejorar las estimaciones municipales con base en las estimaciones oficiales y ii) permitir que el mapa de FIES construido tenga comparabilidad con las cifras reportadas por el DANE.

Para aplicar *benchmark* en nuestras estimaciones, se usó el método de calibración usado frecuentemente en el análisis de encuestas por muestreo. En muestreo esta técnica es usada cuando se tiene presencia de sesgo o poca precisión en las estimaciones a causa de la no respuesta o bajos tamaños de muestra, este método requiere de variables auxiliares para las que se conoce sus estimaciones o parámetros poblacionales a partir de otra encuesta o registro administrativo, con el fin de ajustar los factores de expansión de la encuesta teniendo en cuenta una combinación lineal que permita llegar al valor total de la o las variables auxiliares sobre las que se conoce esta información previamente al levantamiento de la encuesta.

Este proceso, de acuerdo con Gutiérrez (2016), consiste en 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 $d_k = 1/\pi_k$

Para este caso en particular no se tiene un total sino una razón/proporción de la variable para calibrar, en este caso Gutiérrez, Zhang y Rodríguez (2016), indican cómo hacer la calibración cuando contamos con información del total de la población para razones, que se expresan como $R_q = t_{yq}/t_{xq}$, donde t_{yq} y t_{xq} son conocidos para el total de la población, en este caso se va a buscar el vector de w_k que cumplen con la siguiente restricción:

$$\frac{\sum_{k \in s} w_k y_{qk}}{\sum_{k \in s} w_k x_{qk}} = R_q$$

Donde, $q = 1, \dots, Q$ y representan el total de ecuaciones de calibración.

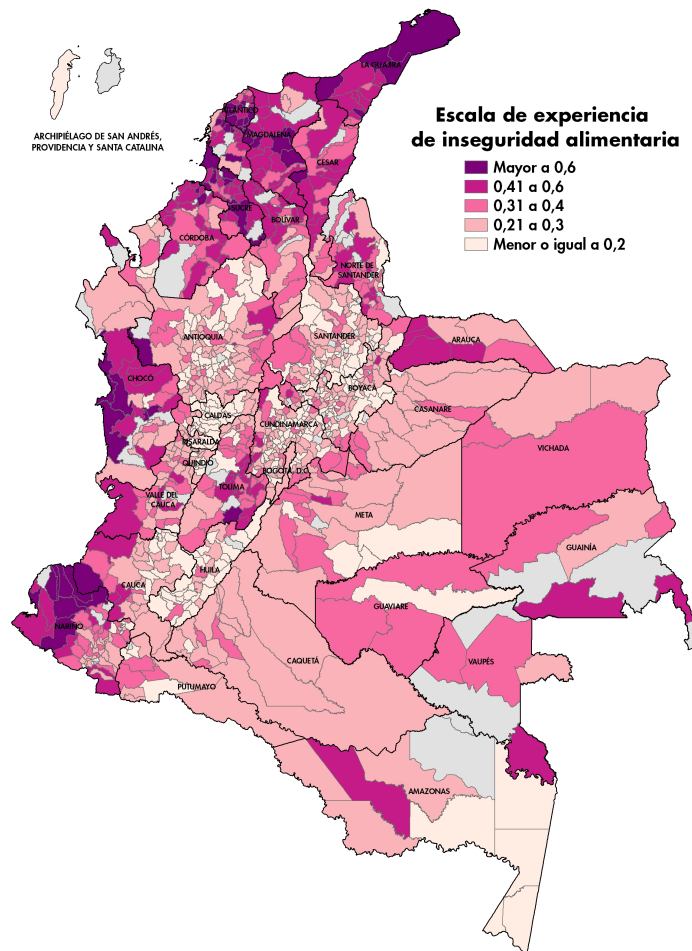
5.8 Validación y presentación de resultados

Al aplicar la calibración se garantiza que el promedio ponderado para los municipios por la proyección poblacional de 2022 de la estimación del modelo FH se obtiene el mismo valor reportado por la estimación directa para los 33 departamentos del país.

5.8 Validación y presentación de resultados

Finalmente, se realizó la validación y consistencia de los resultados por parte de un grupo de expertos de FAO: **Clara Aida Khalil**; **Stefano DiCandia**; **Santiago Mazo**; **María Rodríguez de España**; **Cristobal Ferhmann** y el experto de CEPAL: **Andrés Gutiérrez**, a quienes les agradecemos por sus acompañamiento constante (técnico y conceptual) durante el 2023 para la estimación de la prevalencia de inseguridad alimentaria moderada o grave a nivel municipal.

Gráfica 10. Prevalencia de la inseguridad alimentaria moderada y grave en hogares (%)



Fuente: DANE.

6 Bibliografía

- DANE (2023), Boletín técnico Escala de experiencia de inseguridad alimentaria (FIES) 2022, <https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/escala-de-experiencia-de-inseguridad-alimentaria-fies-2022>
 - 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 Hidirolou, 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>
-