



ESTIMACIÓN EN ÁREAS PEQUEÑAS DESNUTRICIÓN CRÓNICA INFANTIL A NIVEL CANTONAL

OCTUBRE 2024



ESTIMACIÓN EN ÁREAS PEQUEÑAS

DESNUTRICIÓN CRÓNICA INFANTIL A NIVEL CANTONAL

Resumen:

La creciente demanda de información desagregada enfrenta el desafío de los elevados costos asociados con la generación de información. En este contexto, las metodologías de estimación en áreas pequeñas emergen como una solución viable, permitiendo la generación de datos precisos y representativos de manera más eficiente y económica. En el presente estudio se utiliza la información de la Encuesta Nacional sobre Desnutrición Infantil (ENDI) 2022-2023, cuyo diseño muestral permite la estimación de la prevalencia de desnutrición crónica infantil (DCI) en menores de cinco años hasta el nivel provincial. Mediante el uso de metodologías de estimación en áreas pequeñas, específicamente con el modelo de Fay-Herriot, se mejoró la precisión de las estimaciones de la prevalencia de DCI a nivel cantonal, utilizando como fuente auxiliar el Censo de Población y Vivienda (CPV) 2022. Esto permitió obtener estimaciones para los 221 cantones del Ecuador, con coeficientes de variación inferiores al 30%.

Palabras clave: DCI; Mejor predictor lineal insesgado empírico; Error cuadrático medio; Modelo Fay-Herriot; Estimación en áreas pequeñas.

Abstract:

The growing demand for disaggregated information faces the challenge of high costs associated with data generation. In this context, small area estimation methodologies emerge as a viable solution, allowing for the production of accurate and representative data in a more efficient and cost-effective manner. In this study, we use data from the 2022-2023 National Child Malnutrition Survey (ENDI), whose sample design allows for estimating the prevalence of child chronic malnutrition (DCI) in children under five at the provincial level. By applying small area estimation methodologies, specifically through the Fay-Herriot model, we improved the precision of DCI prevalence estimates at the cantonal level, using the 2022 Census as an auxiliary data source. This approach provided estimates for all 221 cantons in Ecuador, achieving coefficients of variation below 30%.

Keywords: DCI; Empirical best linear unbiased predictor; Mean squared error; Fay-Herriot model; Small area estimation.

ESTIMACIÓN EN ÁREAS PEQUEÑAS

DESNUTRICIÓN CRÓNICA INFANTIL A NIVEL CANTONAL

Coordinación

Coordinación General Técnica de Innovación en Métricas y Análisis de la Información

Dirección/Departamento

Dirección de Innovación en Métricas y Metodologías

Elaborado por:

Danilo J. Vera-Ponce
Galo Egas G.

Revisado por:

Galo Egas G.
Director de Innovación en Métricas y Metodologías

Aprobado por:

Darío Vélez
Coordinador General de Innovación en Métricas y Análisis de la Información

Citar como:

INEC (2024). Estimación en áreas pequeñas. Desnutrición crónica infantil a nivel cantonal. Instituto Nacional de Estadística y Censos (2024). Quito, Ecuador.

AGRADECIMIENTO

El Instituto Nacional de Estadística y Censos expresa su más profunda gratitud a la Fundación REDNI y al Dr. Domingo Morales, catedrático de la Universidad Miguel Hernández de Elche, por su apoyo invaluable en la provisión de recursos y asesoría especializada. Su contribución fue fundamental para el fortalecimiento de las capacidades técnicas en metodologías de estimación en áreas pequeñas y en el éxito del ejercicio de estimación.

De igual manera, extendemos nuestro sincero reconocimiento a la Comisión Económica para América Latina y el Caribe (CEPAL) por sus valiosas contribuciones técnicas en el desarrollo de los modelos empleados. Su colaboración fue determinante para asegurar la calidad y precisión de las estimaciones, y su apoyo constante ha sido esencial para la excelencia de este trabajo.

Los autores desean también expresar su agradecimiento a Paul Corral Rodas, cuya visión y comentarios críticos enriquecieron el desarrollo de los modelos utilizados en este estudio. Su guía y experiencia fueron cruciales para enfrentar los desafíos metodológicos y asegurar la solidez de los resultados. Finalmente, nuestro reconocimiento se extiende a Guillermo Feijoó, quien realizó importantes aportaciones en la definición de los parámetros y en el análisis de los resultados, contribuyendo de manera significativa a la calidad del presente estudio.

Índice

GLOSARIO	6
1 INTRODUCCIÓN	7
2 METODOLOGÍA	9
2.1 Áreas pequeñas	9
2.2 Modelo Fay-Herriot	9
2.3 Error cuadrático medio del modelo FH	11
2.4 Estimación de $\hat{\sigma}_u^2$ por ML y REML	13
2.5 Transformaciones	14
2.6 Benchmarking	15
2.7 Fuentes de información	16
2.8 Dominios (áreas pequeñas)	17
2.8.1 Dominios "out-sample"	17
2.9 Selección de variables & proceso de estimación	18
2.9.1 Selección de variables	18
2.9.2 Proceso de estimación	18
2.10 R: Paquetes utilizados	19
3 RESULTADOS	21
3.1 Resultados del modelo y estimaciones	21
3.2 Coeficientes de variación	23
3.3 Mapa de prevalencia de Desnutrición Crónica Infantil cantonal	24
4 CONCLUSIONES Y CONSIDERACIONES	27
4.1 Conclusiones	27
4.2 Consideraciones	27
5 REFERENCIAS	29
ANEXOS	32
Anexo 1: Gráficos de Diagnóstico del modelo final	32
Anexo 2: Variables	33
Anexo 3: Estimaciones Cantonales	34

Índice de figuras

1	Normalidad en los efectos aleatorios y errores del modelo FH	22
2	Coeficiente de variación FH vs Directas, ordenado por tamaño de muestra . .	23
3	Mapa de la Estimación en Áreas Pequeñas de la DCI cantonal 2023	25

Índice de cuadros

1	Resultados modelo Fay-Herriot (FH) para la prevalencia de DCI	21
---	---	----

GLOSARIO

- **AIC:** Criterio de Información de Akaike
- **AdaENET:** Adaptive Elastic-NET (Red elástica adaptativa)
- **BIC:** Criterio de Información Bayesiano
- **CEPAL:** Comisión Económica para América Latina y el Caribe
- **CV:** Coeficiente de variación
- **DCI:** Desnutrición Crónica Infantil
- **BLUP:** Best Linear Unbiased Predictor (Mejor Predictor Lineal Insesgado)
- **EBLUP:** Empirical Best Linear Unbiased Predictor (Mejor Predictor Lineal Insesgado Empírico)
- **ENSANUT:** Encuesta Nacional de Salud y Nutrición
- **ENDI:** Encuesta Nacional sobre Desnutrición Infantil
- **FH:** Fay-Herriot
- **FGV:** Función Generalizada de Varianza
- **IC:** Intervalo de confianza
- **INEC:** Instituto Nacional de Estadística y Censos
- **LASSO:** Least Absolute Shrinkage and Selection Operator (Operador de selección y contracción mínima absoluta)
- **ML:** Maximum likelihood (Máxima verosimilitud)
- **ECM:** Error Cuadrático Medio
- **OMS:** Organización Mundial de la Salud
- **REML:** Restricted maximum likelihood (máxima verosimilitud restringida o residual)
- **SAE:** Small Area Estimation (Estimación en Áreas Pequeñas)
- **UNICEF:** Fondo de las Naciones Unidas para la Infancia

1 INTRODUCCIÓN

El diseño, implementación y evaluación de políticas públicas constituyen procesos complejos que demandan un enfoque altamente contextualizado, lo que implica la necesidad de contar con información desagregada a niveles territoriales específicos. Esta desagregación es crucial, ya que permite identificar y comprender las particularidades y dinámicas propias de cada comunidad, región o localidad. La diversidad sociocultural, económica y demográfica que caracteriza a los distintos territorios implica que las necesidades y prioridades de la población pueden variar significativamente, incluso dentro de un mismo territorio como una provincia.

Por su parte, la producción de información estadística se lleva a cabo, en la mayoría de los casos, mediante diseños muestrales que, al ser inherentemente costo-eficientes, ofrecen una alternativa viable a la exhaustividad de los censos tradicionales. Estos diseños muestrales pueden variar desde métodos simples, como el muestreo aleatorio, hasta enfoques más complejos que incluyen muestreos estratificados, por conglomerados o multietápico, cada uno de los cuales presenta ventajas y desventajas en términos de representatividad y precisión. Sin embargo, uno de los principales inconvenientes de los diseños muestrales radica en su limitada capacidad para garantizar la representatividad de los datos a dominios predeterminados, lo que puede resultar en una subrepresentación de subgrupos o áreas geográficas específicas, comprometiendo la validez externa de las inferencias estadísticas y limitando la capacidad de los responsables de formular políticas para responder de manera efectiva a necesidades particulares de la población.

En consecuencia, la implementación de estrategias de estimación en áreas pequeñas (SAE por sus siglas en inglés) se torna fundamental, ya que permite mejorar la precisión de las estimaciones mediante el uso de información auxiliar y modelos estadísticos avanzados, brindando así un enfoque robusto de generación de estadísticas para dominios subrepresentados y que permita la respuesta adaptada a las dinámicas locales o particulares. En el Ecuador, el uso de estas metodologías como parte de la estadística oficial se ha utilizado en la generación de mapas de pobreza y desigualdad por consumo a nivel de cantón y parroquia (Cabrera et al. 2016).

Por otro lado, La desnutrición crónica infantil es un indicador que hace referencia a un retraso en el crecimiento de un menor. Su medición se realiza contrastando la talla del niño con el estándar recomendado para su edad y sexo de acuerdo con las mediciones de crecimiento infantil de la Organización Mundial de la Salud – OMS (WHO 2006). La DCI indica una carencia de los nutrientes necesarios durante un tiempo prolongado, por lo que aumenta el riesgo de que contraiga enfermedades y afecta al desarrollo físico e intelectual del niño (UNICEF 2011). En el Ecuador, la DCI es un serio problema que se ha arrastrado por décadas; y de manera histórica, la información sobre DCI se obtenía a partir de: la Encuesta Demográfica Materno Infantil (ENDEMAIN) que 1986 incluyó un módulo de antropometría, la Encuesta de Condiciones de Vida (ECV) 2006 y 2014; y, la Encuesta Nacional de Salud y Nutrición (ENSANUT) 2012 y 2018. Entre el 2006 y el 2014 se observó una reducción estadísticamente significativa de 2 puntos; mientras que entre 2014 y 2018 se obtuvo una reducción estadísticamente no significativa. A partir del año 2022, el INEC comienza la recolección de la primera encuesta especializada en materia de nutrición infantil denominada: Encuesta Nacional sobre Desnutrición Infantil (ENDI), y cuyo diseño

muestral considera el levantamiento continuo en 12 meses en todo el territorio nacional, con representatividad hasta el nivel provincial y que investiga indicadores antropométricos, de calidad de agua, salud materno-infantil, acceso a servicios, entre otros. En el periodo 2022-2023. La DCI en menores de 5 años fue de 21.6% (INEC 2023a).

El presente estudio tiene como objetivo estimar la prevalencia de la DCI en menores de cinco años a nivel cantonal, utilizando como fuente de información principal la Encuesta Nacional sobre Desnutrición Infantil (ENDI) del periodo 2022 y 2023; y, como fuente auxiliar el Censo de Población y Vivienda de 2022. Para llevar a cabo esta estimación, se empleará la metodología de Estimación en Áreas Pequeñas mediante un modelo de Fay-Herriot a nivel de área. Para alcanzar este objetivo es esencial: (1) examinar los fundamentos teóricos que sustentan el modelo Fay-Herriot; (2) recopilar y procesar las fuentes de información primaria y auxiliar; (3) realizar un proceso de selección de variables y aplicar el modelo Fay-Herriot; y (4) validar los resultados obtenidos utilizando fuentes de información disponible.

2 METODOLOGÍA

Esta sección se centra en el enfoque metodológico requerido para generar estimaciones a nivel cantonal de la DCI. Se explora la conceptualización teórica del modelo FH, un método de estimación SAE que opera a nivel de área en contraste a los modelos de unidad que no se tratan en este documento. Además, se analizan aspectos relacionados a la selección de variables, la precisión de las estimaciones, el uso de transformaciones y el proceso de conformidad (benchmarking). Por otro lado, se detallan las fuentes de información, el procedimiento utilizado para obtener las estimaciones y las herramientas empleadas.

2.1 Áreas pequeñas

El planteamiento de modelos SAE considera una población finita U que consta de N unidades que se particionan en D dominios o áreas de tamaños específicos N_1, \dots, N_D . Entonces, es posible extraer una muestra aleatoria de tamaño n de U y particionadas en las D áreas con n_1, \dots, n_D observaciones por dominio, de manera que se cumpla:

$$u = \bigcup_{d=1}^D u_d \subset U = \bigcup_{d=1}^D U_d$$

Además, para todo $d, d' \in \{1, \dots, D\}$, con $d \neq d'$, $U_d \cap U_{d'} = \emptyset$ y $u_d \subset U_d$. Los conjuntos u_d se denominan **áreas pequeñas**, donde $\text{Card}(u_d) = n_d \geq 0$.

2.2 Modelo Fay-Herriot

El modelo Fay-Herriot (FH) para D áreas es un modelo Bayesiano de dos niveles. El nivel 1 o muestral pretende explicar la variabilidad debida al muestreo (de una encuesta) del estimador de interés; y el nivel 2 o modelo de enlace relaciona el estimador de interés con p covariables. Considerados los dos niveles anteriores es posible plantear un **modelo lineal mixto**. Con este modelo Fay y Herriot (1979) capturaron los efectos específicos de cada área utilizando efectos aleatorios, algo que no es posible utilizando modelos de regresión clásicos (Novo 2018).

Al respecto de las covariables, es necesario contar con información auxiliar confiable la cual debe ser agregada al mismo nivel D . Comúnmente esta información proviene de censos o registros administrativos lo cual supone un nivel de confiabilidad muy elevado y un margen de error mínimo. Así, el modelo FH enlaza los estimadores directos de las áreas δ_d , $d = 1, \dots, D$ y asume que estos se relacionan linealmente con un vector de p covariables, \mathbf{x}_d .

El estimador directo de una variable de interés para cada área d puede ser estimado por el método de Horvitz-Thompson (Horvitz y Thompson 1952) con el uso de los pesos muestrales ($w = w_{d,j}$, $d = 1, \dots, D$; $j = 1, \dots, n_d$) o conocidos también como factores de expansión (f_{exp}) los cuales se derivan del diseño muestral de la encuesta. Para un indicador $\delta = \delta_d$, el

estimador directo es>

$$\hat{\delta}_d^{DIR} = \left(\sum_{j=1}^D w_{d,j} \right)^{-1} \sum_{j=1}^{n_d} w_{d,j} y_{d,j}.$$

Además, el estimador directo $\hat{\delta}_d^{DIR}$ de δ_d es insesgado bajo el diseño muestral (MDSF-CEPAL 2021).

Por un lado, el *modelo de muestreo* (nivel 1) se define como:

$$\hat{\delta}_d^{DIR} = \delta_d + e_d; \quad d = 1, \dots, D. \quad (1)$$

Por otro lado, el *modelo de enlace* que vincula linealmente el indicador δ con las variables auxiliares \mathbf{x}_d del área, está dado por:

$$\delta_d = \mathbf{x}_d' \beta + u_d \quad d = 1, \dots, D, \quad (2)$$

donde $u_d \stackrel{iid}{\sim} (0, \sigma_u^2)$ es un vector de efectos aleatorios y β un vector de parámetros de efectos fijos desconocidos.

Así, la combinación de los modelos de muestreo (1) y enlace (2) conduce a un modelo lineal mixto:

$$\hat{\delta}_d^{DIR} = \mathbf{x}_d' \beta + u_d + e_d, \quad d = 1 \dots, D. \quad (3)$$

De (3), se debe tomar en cuenta que:

- e_d es el error debido al muestreo, $e_d \stackrel{ind}{\sim} (0, \psi_d)$;
- la varianza $\psi_d = \text{var}_{\pi}(\hat{\delta}_d^{DIR} | \delta_d)$ se estima con los microdatos de la encuesta¹.

Cuando se conoce el valor de σ_u^2 , al minimizar el error cuadrático medio (EMC) se obtiene el mejor predictor lineal insesgado (BLUP por sus siglas en inglés) para $\delta_d = \mathbf{x}_d' \beta + u_d$. Por tanto, el BLUP bajo el modelo FH de δ_d viene dado por:

$$\tilde{\delta}_d^{FH} = \mathbf{x}_d' \tilde{\beta} + \tilde{u}_d, \quad (4)$$

donde

$$\tilde{\beta} = \left(\sum_{d=1}^D \left(\frac{1}{\sigma_u^2 + \psi_d} \right) \mathbf{x}_d \mathbf{x}_d' \right)^{-1} \sum_{d=1}^D \left(\frac{1}{\sigma_u^2 + \psi_d} \right) \mathbf{x}_d \hat{\delta}_d^{DIR}$$

y

$$\tilde{u}_d = \gamma_d (\hat{\delta}_d^{DIR} - \mathbf{x}_d' \tilde{\beta}),$$

¹Un insumo importante en el modelo FH es la varianza del estimador directo, a nivel cantonal, la cual no puede calcularse de ningún modo. En su lugar, este valor se estima con los datos recolectados. Sin embargo, en los cantones con un tamaño de muestra muy pequeño, estas estimaciones no tendrán un buen comportamiento, lo cual podría dificultar el proceso de modelado. En estos casos es pertinente analizar el uso de un modelo de suavizamiento de las varianzas (FGV - Función Generalizada de Varianza) para eliminar el ruido y la volatilidad de estas estimaciones (MDSF-CEPAL, 2021).

con

$$\gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \psi_d}. \quad (5)$$

Si se sustituye \tilde{u}_d en (4), el BLUP bajo el modelo FH, se obtiene como

$$\tilde{\delta}_d^{FH} = \gamma_d \hat{\delta}_d^{DIR} + (1 - \gamma_d) \mathbf{x}_d' \tilde{\beta}, \quad (6)$$

lo cual es una combinación lineal convexa del estimador directo y del estimador sintético a nivel de área.

De (5), se puede ver que, si la varianza muestral ψ_d es pequeña comparada con la heterogeneidad no explicada σ_u^2 , $\gamma_d \rightarrow 1$. Entonces, cuando el tamaño muestral del área es grande ($\psi_d \rightarrow 0$), el BLUP $\tilde{\delta}_d^{FH}$ se acerca al estimador directo. Por tanto, no se necesita saber si el área es pequeña para usar este estimador.

Habitualmente, no se conoce el verdadero valor de σ_u^2 de los efectos aleatorios u_d , por lo cual se recurre a $\hat{\sigma}_u^2$ un estimador consistente de σ_u^2 . A partir de esto, se obtiene el mejor predictor lineal insesgado empírico (EBLUP por sus siglas en inglés) de δ_d :

$$\hat{\delta}_d^{FH} = \hat{\gamma}_d \hat{\delta}_d^{DIR} + (1 - \hat{\gamma}_d) \mathbf{x}_d' \hat{\beta},$$

donde $\hat{\gamma}_d = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \psi_d)$ y

$$\hat{\beta} = \left(\sum_{d=1}^D \left(\frac{1}{\hat{\sigma}_u^2 + \psi_d} \right) \mathbf{x}_d \mathbf{x}_d' \right)^{-1} \sum_{d=1}^D \left(\frac{1}{\hat{\sigma}_u^2 + \psi_d} \right) \mathbf{x}_d \hat{\delta}_d^{DIR}.$$

En un área no muestreada, al tener $n_d \rightarrow 0$, la varianza del estimador directo $\psi_d \rightarrow \infty$ y por tanto $\gamma_d \rightarrow 0$. Tomando el valor límite $\gamma_d = 0$, se obtiene el estimador sintético de regresión. Es decir, los dominios para los que no existen resultados de estimación directa porque el tamaño de la muestra es cero o su tamaño de muestra es reducido y fueron excluidos (out-sample), la estimación se reduce al componente sintético de regresión:

$$\hat{\delta}_{d,out}^{FH} = \mathbf{x}_d' \hat{\beta}. \quad (7)$$

Finalmente, de (6) y (7), se tiene:

$$\hat{\delta}_d^{FH} = \begin{cases} \hat{\gamma}_d \hat{\delta}_d^{DIR} + (1 - \hat{\gamma}_d) \mathbf{x}_d' \hat{\beta} & , \text{ si } d \text{ es in-sample,} \\ \mathbf{x}_d' \hat{\beta} & , \text{ si } d \text{ es out-of-sample.} \end{cases} \quad (8)$$

2.3 Error cuadrático medio del modelo FH

Si se conocen los parámetros del modelo (4), β y σ_u^2 , el error cuadrático medio (ECM) del BLUP $\tilde{\delta}_d^{FH}$ viene dado por

$$\text{ECM}(\tilde{\delta}_d^{FH}) = \gamma_d^2 \psi_d \leq \psi_d = \text{var}_{\pi}(\hat{\delta}_d^{DIR} | \delta_d). \quad (9)$$

De (10), se puede concluir que el BLUP bajo el modelo FH no puede ser menos eficiente que el estimador directo. No obstante, en la práctica, no se dispone de estos valores, lo cual hace que el ECM se incremente, pero este crecimiento tiende a cero con un aumento en el número de áreas D .

Nuevamente, considerando $\hat{\sigma}_u^2$ un estimador consistente de σ_u^2 obtenido mediante máxima verosimilitud (ML por sus siglas en inglés) o máxima verosimilitud residual (REML por sus siglas en inglés). Un estimador insesgado de segundo orden del ECM, llamado el estimador Prasad-Rao (Prasad y Rao 1990) viene dado por:

$$\text{ECM}_{\text{PR}}(\hat{\delta}_d^{FH}) = g_{d,1}(\hat{\sigma}_u^2) + g_{d,2}(\hat{\sigma}_u^2) + 2g_{d,3}(\hat{\sigma}_u^2), \quad (10)$$

donde

$$\begin{aligned} - g_{d,1}(\hat{\sigma}_u^2) &= \gamma_d \psi_d, \\ - g_{d,2}(\hat{\sigma}_u^2) &= (1 - \gamma_d)^2 \mathbf{x}_d' \mathbf{Q} \mathbf{x}_d, \\ - g_{d,3}(\hat{\sigma}_u^2) &= (1 - \gamma_d)^2 (\sigma_u^2 + \psi_d)^{-1} \overline{\text{var}}(\hat{\sigma}_u^2), \end{aligned}$$

con

$$\begin{aligned} - \mathbf{Q} &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1}; \\ - \mathbf{X} &= \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_d \end{pmatrix} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,p} \\ x_{2,1} & \cdots & x_{2,p} \\ \vdots & \ddots & \vdots \\ x_{d,1} & \cdots & x_{d,p} \end{pmatrix}; \\ - \mathbf{V} &= \text{Diag}(\hat{\sigma}_u^2 + \psi_1, \dots, \hat{\sigma}_u^2 + \psi_d); \gamma \\ - \overline{\text{var}}(\hat{\sigma}_u^2) &= [\mathcal{J}(\hat{\sigma}_u^2)]^{-1} = 2 \left(\sum_{d=1}^D (\hat{\sigma}_u^2 + \psi_d)^{-2} \right)^{-1}, \text{ es la varianza asintótica } (D \rightarrow \infty) \\ &\text{para un estimador ML o REML e } \mathcal{J} \text{ es la información de Fisher.} \end{aligned}$$

Se debe tomar en cuenta que, $g_{d,2}(\hat{\sigma}_u^2)$ y $g_{d,3}(\hat{\sigma}_u^2)$ tienden a cero, cuando el número de áreas D es suficientemente grande.

Adicionalmente, de acuerdo con I. Molina y Rao (2015), una aproximación para el EMC de los estimadores FH en dominios out-sample está dado por:

$$\text{ECM}(\hat{\delta}_{d,out}^{FH}) = \hat{\sigma}_u^2 + \mathbf{x}_{d,out}' \mathbf{Q} \mathbf{x}_{d,out}. \quad (11)$$

Finalmente, de (10) y (11), se tiene que:

$$\text{ECM}(\hat{\delta}_d^{FH}) = \begin{cases} g_{d,1}(\hat{\sigma}_u^2) + g_{d,2}(\hat{\sigma}_u^2) + 2g_{d,3}(\hat{\sigma}_u^2) & , \text{ si } d \text{ es in-sample,} \\ \hat{\sigma}_u^2 + \mathbf{x}_d' \mathbf{Q} \mathbf{x}_d & , \text{ si } d \text{ es out-of-sample.} \end{cases} \quad (12)$$

Por último, de acuerdo con el Bando de Desarrollo de Asia (ADB 2020), el CV como una medida de confiabilidad de las estimaciones se puede calcular utilizando la fórmula:

$$CV(\hat{\delta}_d^{FH}) = \frac{\sqrt{ECM(\hat{\delta}_d^{FH})}}{\hat{\delta}_d^{FH}}.$$

De la misma manera, se puede construir un intervalo de confianza (IC) del 95% utilizando la ecuación:

$$IC(\hat{\delta}_d^{FH}) : \left(\hat{\delta}_d^{FH} - 1.96\sqrt{ECM(\hat{\delta}_d^{FH})}; \hat{\delta}_d^{FH} + 1.96\sqrt{ECM(\hat{\delta}_d^{FH})} \right).$$

2.4 Estimación de $\hat{\sigma}_u^2$ por ML y REML

De I. Molina y Rao (2015) y Isabel Molina y Marhuenda (2015) se tiene que la forma de estimar σ_u^2 por medio de máxima verosimilitud es usando el algoritmo iterativo de puntuación de Fisher:

$$\begin{aligned} \sigma_u^{2(0)} &= 0 \\ \sigma_u^{2(a+1)} &= \hat{\sigma}_u^{2(a)} + \left[J(\sigma_u^{2(a)}) \right]^{-1} S_{ML}(\sigma_u^{2(a)}), \quad a = 1, 2, 3, \dots, \end{aligned} \quad (13)$$

donde

$$S_{ML}(\sigma_u^2) = -\frac{1}{2} \sum_{d=1}^D \frac{1}{\sigma_u^2 + \psi_d} + \frac{1}{2} \sum_{d=1}^D \left(\frac{\hat{\delta}_d^{DIR} - \mathbf{x}'_d \tilde{\beta}}{\sigma_u^2 + \psi_d} \right)^2. \quad (14)$$

El estimador ML final $\hat{\sigma}_u^2 = \hat{\sigma}_{u,ML}^2$ es tomado como: $\hat{\sigma}_{u,ML}^2 = \max\{\sigma_{u,ML}^{2*}, 0\}$, donde $\sigma_{u,ML}^{2*}$ es la solución obtenida de la última iteración de (13) cuando se cumpla con un valor tolerable o se hayan alcanzado el número máximo de iteraciones. En (14), $\tilde{\beta} = \tilde{\beta}(\sigma_u^2)$ se obtiene por la siguiente ecuación (Isabel Molina y Marhuenda 2015):

$$\tilde{\beta}(\sigma_u^2) = \mathbf{Q}^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y},$$

con $\mathbf{y} = (\delta_d^{DIR})_{d=1}^D$ (vector columna $D \times 1$).

No obstante, de acuerdo con lo descrito en I. Molina y Rao (2015), un inconveniente del estimador ML de σ_u^2 es que no tiene en cuenta la pérdida de grados de libertad que acarrea la estimación de $\tilde{\beta}$, punto que es tomado en cuenta en el método REML utilizando datos transformados, se hace uso de $\mathbf{y}^* = \mathbf{A}' \mathbf{y}$, donde \mathbf{A} es cualquier matriz de rango completo y $\mathbf{A}' \mathbf{y}$ no depende de β .

El algoritmo de puntuación del Fisher para la estimación REML de σ_u^2 se puede escribir como:

$$\begin{aligned} \sigma_u^{2(0)} &= 0 \\ \sigma_u^{2(a+1)} &= \sigma_u^{2(a)} + \left(\frac{1}{2} \text{tr}(\mathbf{P}\mathbf{P}) \right)^{-1} S_{REML}(\sigma_u^{2(a)}), \quad a = 1, 2, 3, \dots, \end{aligned} \quad (15)$$

donde $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{XQ}^{-1}\mathbf{X}'\mathbf{V}^{-1}$ y

$$S_{REML}(\sigma_u^{2(a)}) = -\frac{1}{2}\text{tr}(\mathbf{P}) + \frac{1}{2}(\mathbf{Py})'\mathbf{Py}.$$

El estimador REML final $\hat{\sigma}_u^2 = \hat{\sigma}_{u,REML}^2 = \max\{\sigma_{u,REML}^{2*}, 0\}$, donde $\sigma_{u,REML}^{2*}$ es la solución obtenida de (15) bajo los mismos criterios que el algoritmo ML. La convergencia de los procedimientos iterativos (13) y (15) es rápida y generalmente requiere menos de 10 iteraciones (I. Molina y Rao 2015).

2.5 Transformaciones

Una práctica habitual en el proceso de modelado es aplicar transformaciones, con el objetivo de optimizar las condiciones de las variables, mejorar las estimaciones y obtener resultados más efectivos. Esto se debe a que la variable original puede no presentar un comportamiento adecuado cuando se modela directamente. Siguiendo a Ghosh (2020), aplicar una transformación logarítmica² resulta ser estrategia propicia para mejorar el comportamiento del modelo y/o cumplir con los supuestos a los que este se encuentra sujeto. No obstante, es importante tomar en cuenta que cuando se realiza el análisis basado en los datos transformados, se debe transformar correctamente los resultados a la escala original (Ghosh 2020).

Al usar una transformación logarítmica, las estimaciones directas y las varianzas transformadas se determinan de la siguiente manera (Harmening et al. 2019; Neves, Silva, y Correa 2013):

$$\hat{\delta}_d^{DIR*log} = \log(\hat{\delta}_d^{DIR}), \quad (16)$$

y

$$\text{var}(\hat{\delta}_d^{DIR*log}) = (\hat{\delta}_d^{DIR})^{-2} \text{var}(\hat{\delta}_d^{DIR}), \quad (17)$$

donde la notación $*log$ representa la escala transformada logarítmica. Para obtener el estimador FH en la escala transformada se sustituye (16) y (17) en el modelo planteado en (3).

Dado que el logaritmo es una transformación no lineal, las estimaciones finales de FH en la escala original requieren una transformación inversa con corrección de sesgo. Se hace uso de un método *crude* (*bc_crude*) que tiene en cuenta las propiedades de la distribución log-normal (Slud y Maiti 2006). Así, se tiene que:

$$\hat{\delta}_d^{FH, crude} = \exp\left\{\hat{\delta}_d^{FH*log} + 0.5 \text{ECM}(\hat{\delta}_d^{FH*log})\right\} \quad (18)$$

y

$$\text{ECM}(\hat{\delta}_d^{FH, crude}) = (\hat{\delta}_d^{FH, crude})^2 \text{ECM}(\hat{\delta}_d^{FH*log}). \quad (19)$$

²Estudios como el de Moura, et al., (2020), Kreutzmann, et al., (2019) enfatizan el uso de la transformación logarítmica en datos asimétricos, no obstante, en los trabajos de Neves, et al. (2013) y Shiferaw (2024), se considera común el uso de la transformación logarítmica en datos económicos con la finalidad de capturar una relación no lineal entre las variables que intervienen en el modelo.

2.6 Benchmarking

Una vez obtenidas las estimaciones en áreas pequeñas (FH) resulta natural pensar en agregar estos resultados para obtener una estimación del dominio que las contiene. La idea del benchmarking es que las estimaciones obtenidas por medio del modelo Fay-Herriot, al momento de ser agregadas deben ser iguales a las estimaciones directas confiables de un nivel regional más alto (provincias) (Harmening et al. 2019).

En este sentido, cuando se agregan estas estimaciones cantonales es muy probable que no se obtenga las estimaciones provinciales. Sin embargo, se conoce que la estimación directa para una provincia, es confiable bajo el diseño muestral por ser un dominio planificado en una encuesta. En el diseño muestral de la ENDI, se menciona que "los dominios de diseño son a nivel urbano, rural y provincial con la agregación de los 12 meses de la muestra" (INEC 2023b). Por tanto, en caso de requerir una consistencia³ entre las estimaciones bajo el diseño muestral y las estimaciones SAE es necesario utilizar técnicas de conformidad (benchmarking).

Así, luego de obtener la estimación FH del área d , $\hat{\delta}_d^{FH}$, y contar con $w_d = N_d/N$, una proporción conocida de unidades en el área d . Para una provincia $prov \in \{01, 02, \dots, 24\}$ que se divide en d_p áreas pequeñas,

$$\hat{\delta}_{prov}^{FH} = \sum_{d=1}^{d_p} w_d \hat{\delta}_d^{FH}$$

es la estimación agregada obtenida a partir del modelo Fay-Herriot.

Además, se dispone de un estimador directo confiable $\hat{\delta}_{prov}^{DIR}$ proveniente del diseño muestral. Por lo tanto, lo dicho anteriormente señala que es deseable que los estimadores del área pequeña, cuando se agregan, concuerden con los estimadores directos confiables $\hat{\delta}_{prov}^{DIR}$. En este caso se dice que los estimadores FH satisfacen la propiedad de "benchmarking" para lo cual es necesario modificar la estimación $\hat{\delta}_d^{FH}$.

La propiedad de benchmarking por diferencias se obtiene sumando a cada $\hat{\delta}_d^{FH}$ un factor de ajuste común α_{p_d} de manera que

$$\hat{\delta}_d^{FH, bench} = \hat{\delta}_d^{FH} + \alpha_{p_d},$$

con

$$\alpha_{p_d} = \hat{\delta}_{prov}^{DIR} - \hat{\delta}_{prov}^{FH}, \quad (20)$$

donde p_d es el número de cantones que compone una provincia. Se puede ver que, el valor para alcanzar el benchmarking por diferencias será distinto para cada provincia.

De acuerdo con Ghosh (2020), se dispone de estimadores del ECM insesgados de segundo orden para el benchmarking por diferencias. Más precisamente, Ghosh y Steorts (2013) han

³Al respecto Corral, et al., (2022) mencionan que no se recomienda el uso de benchmarking a menos que los requisitos de la publicación incluyan que las estimaciones de los totales en un nivel de agregación inferior sumen el total estimado en un nivel superior.

demostrado que una aproximación razonable del ECM es:

$$\text{ECM}(\hat{\delta}_d^{FH, bench}) = \text{ECM}_{PR}(\hat{\delta}_d^{FH}) + g_4(\hat{\sigma}_u^2),$$

donde

$$g_4(\hat{\sigma}_u^2) = \sum_{i=1}^D w_i^2 (1 - \gamma_i)^2 (\hat{\sigma}_u^2 + \psi_i) - \sum_{i=1}^D \sum_{j=1}^D w_i w_j (1 - \gamma_i)(1 - \gamma_j) \mathbf{x}_i' \mathbf{Q} \mathbf{x}_j. \quad (21)$$

Adicionalmente, otra opción para tratar el tema de conformidad es el uso de benchmarking por ratio, en este caso el factor de ajuste común se lo obtiene como un cociente la estimación FH y la estimación agregada provincial de la encuesta ($\alpha_{prov} = \hat{\delta}_{prov}^{DIR} / \hat{\delta}_{prov}^{FH}$) y el ajuste se lo hace multiplicando este factor por la estimación obtenida a partir del modelo ($\hat{\delta}_d^{FH, bench} = \hat{\delta}_d^{FH} \times \alpha_{prov}$). No obstante, los estimadores del ECM insesgados de segundo orden no están disponibles para un ajuste con benchmarking por ratio (Ghosh 2020), y se requiere hacer uso de técnicas de remuestreo (bootstrap).

Finalmente, de acuerdo con Ghosh (2020) y I. Molina y Rao (2015), una crítica a las limitaciones del uso de benchmarking (por diferencias o por ratio) es que se aplica un ajuste común para todas las estimaciones, independientemente de su precisión. Además, se pierden las propiedades del diseño muestral que el modelo Fay-Herriot toma en cuenta en el planteamiento del modelo; puesto que $\hat{\delta}_d^{FH, bench}$ no es consistente con el diseño.

2.7 Fuentes de información

La fuente principal es la Encuesta Nacional sobre Desnutrición Infantil (ENDI 2023) la primera encuesta especializada en tema de desnutrición en el país. Esta encuesta contiene información antropométrica de menores de 5 años y a partir de la cual se ha estimado tasa de DCI del 17,5% en menores de 5 años (INEC 2023a).

Por otro lado, la fuente de información auxiliar es el CENSO 2022. El Censo de Población y Vivienda es el recuento completo de los habitantes y de las unidades de vivienda de un territorio definido, y de sus características más importantes; el cual requiere de un conjunto de operaciones consistentes en recoger, recopilar, evaluar, analizar y publicar o divulgar datos estadísticos de un país. El Censo 2022 recopiló información bajo la metodología de derecho durante 12 semanas, a través de dos modalidades de entrevistas: Censo en Línea, Censo Presencial.

La fuente de información principal permite generar una estimación directa de la DCI a nivel de dominio (cantón) así como la varianza⁴ del estimador. Esta estimación se ajusta por medio del modelo de Fay-Herriot usando como variables regresoras la información agregada a nivel cantonal proveniente del CENSO⁵. UNICEF (2011) señala que las causas de la desnutrición infantil incluyen factores sociales, económicos y políticos; por tanto, se analizan grupos de

⁴En el caso de que las varianzas estimadas no presenten buenas características para el modelo será necesario hacer uso de una FGV o algún otro método de ajuste para controlar la volatilidad.

⁵Uno de los supuestos fundamentales de los modelos SAE básicos es que las variables auxiliares se miden sin errores.

variables disponibles en el censo: variables geográficas, características del hogar/vivienda y características del jefe de hogar y cónyuge.

En este sentido, Morales et al. (2021) sostienen que los modelos a nivel de área describen el comportamiento de los estimadores directos considerando las relaciones con variables auxiliares agregadas, las interdependencias de los datos de diferentes dominios y las posibles estructuras de correlación temporal o espacial. Por tanto, las variables auxiliares que entran en un modelo deben tener una relación con la variable dependiente que pueda interpretarse razonablemente de acuerdo con el fenómeno de estudio.

2.8 Dominios (áreas pequeñas)

Ecuador se caracteriza por su división político-administrativa que organiza el país en provincias, cantones y parroquias. Las provincias, un total de 24, representan las principales divisiones políticas, cada una con un código de dos dígitos único que va desde 01 hasta 24. A su vez, los cantones constituyen subdivisiones más específicas y están codificados con cuatro dígitos, donde los primeros dos identifican la provincia y los siguientes dos asignan un código individual a cada cantón. Por ejemplo, el cantón Rumiñahui de la provincia de Pichincha (cuyo código de provincia es 17) tiene el código 1705. En total, Ecuador está compuesto por 221 cantones, y sus códigos varían desde el 0101 hasta el 2403.

Los dominios de interés para las estimaciones en áreas pequeñas son los cantones. Esto debido a que el diseño muestral de la ENDI, y en general de la mayor parte de las operaciones estadísticas de INEC se planifican para alcanzar representatividad provincial, haciendo que las estimaciones a nivel de cantón sean imprecisas, poco confiables y no publicables. Por ello la importancia de generar estimaciones cantonales confiables y puedan apoyar a las estadísticas oficiales.

2.8.1 Dominios “out-sample”

Antes de iniciar el proceso de modelado, es crucial identificar los dominios que presentan características muestrales adecuadas para generar estimaciones directas de la tasa de DCI y sus varianzas, ya que estos datos son insumos principales para el modelo FH. En este caso, se considera pertinente excluir del modelo aquellos dominios que tienen al menos 10 observaciones: 0111, 0114, 0507, 0608, 0714, 0902, 0914, 0918, 0921, 0924, 1110, 1115, 1116, 1213, 1321, 1322, 1707 y 1708.

Además, es importante tratar con especial cuidado aquellos dominios cuyas estimaciones directas sean cero o estén muy cercanas a cero⁶; estas estimaciones pueden ser poco confiables y además, que resulte poco útil para los fines de las estimaciones SAE, sin importar el tamaño de la muestra en el dominio. Así se opta por excluir del modelo 5 cantones adicionales: 0306, 0923, 1108, 1113 y 2202. Esto ayuda a considerar directamente el uso de transformaciones logarítmicas para los modelos.

⁶Una consecuencia que se desprende de excluir dominios con estimaciones cercanas a cero es la posibilidad de utilizar la transformación logarítmica sin inconvenientes, ya que esta función tiende a $-\infty$ cuando se evalúa en valores cercanos a cero.

Adicionalmente, se tiene 20 cantones que no formaron parte del diseño muestral de la ENDI: 0107, 0110, 0112, 0113, 0609, 0705, 0909, 0913, 0916, 0920, 0925, 0927, 1105, 1318, 1408, 1411, 1804, 2001, 2002, 2003. Por último, se excluye el cantón 1409 debido a su comportamiento atípico en las estimaciones directas. Se espera obtener únicamente una estimación sintética que no esté influenciada por el comportamiento de la estimación directa.

Así, de los 221 dominios un total de 44 se excluyen del entrenamiento de los modelos, lo cual representa aproximadamente un 20% de los cantones. Estos cantones serán considerados como *dominios no muestreados* "out-of-sample domains"; por tanto, el análisis y los criterios de selección del modelo solo se calculan en función de los 177 *dominios muestreados* "in-sample domains" (Tabla 1).

2.9 Selección de variables & proceso de estimación

2.9.1 Selección de variables

La selección de variables es un proceso clave que ayuda a reducir el número de variables de entrada al construir un modelo, en este caso un modelo lineal mixto. Este proceso tiene como objetivos, entre otros, optimizar el costo computacional asociado al modelado y, al mismo tiempo, mejorar el rendimiento general del modelo. Al enfocarse en las variables más relevantes, se logra una mayor eficiencia y se potencia la capacidad del modelo para captar patrones significativos en los datos.

Los métodos de selección pueden basarse en estadísticas o en procesos automáticos que usan técnicas como la validación cruzada:

- Los métodos basados en estadísticas implican evaluar la relación entre cada variable de entrada y la variable objetivo utilizando estadísticas y seleccionando aquellas variables de entrada que tienen la relación más fuerte con la variable objetivo. Existen varios métodos uno de estos son los métodos *step-wise*.
- El método de selección automática considerado es el AdaENET. La red elástica adaptativa es una mezcla del LASSO adaptativo y la red elástica (ENET). El LASSO adaptativo logra la "oracle property" mientras que la red elástica soluciona el problema de multicolinealidad (Ogutu, Schulz-Streeck, y Piepho 2012).

2.9.2 Proceso de estimación

El proceso de estimación consistió en los siguientes pasos:

- (1) Procesamiento de las fuentes de información principal y auxiliar: generación de las estimaciones directas y sus varianzas⁷; y agregación de las covariables a nivel de cantón.

⁷El uso de la Función Generalizada de Varianza (FGV) para el modelado y suavizamiento de las varianzas estimadas a partir del diseño muestral cumple el propósito de mejorar el comportamiento de las varianzas sobretudo en dominios con tamaños de muestra reducidos. No obstante, este proceso generó resultados similares a los obtenidos sin el uso de la FGV hecho que pone en evidencia la robustez del modelo inicial sin FGV y el buen comportamiento inicial de las varianzas, dado que los coeficientes en signo y valor al igual que la significancia de las variables del modelo no se vieron afectadas.

- (2) Identificación y exclusión de los dominios out-of-sample.
- (3) Match de las estimación directa de la DCI a nivel cantonal con las p variables agregadas disponibles en el CENSO 2022.
- (4) Se aplica el método AdaENET para seleccionar variables con un $\alpha < 0.1$, esto con la finalidad de no eliminar una cantidad excesiva de variables. Se espera contar con un aproximado de entre 40 a 60 variables en este primer método; es decir, una cantidad $q < p$ de variables.
- (5) Identificadas las variables se hace uso del método step-wise a través de la función `step` del paquete `emdi`, que permite seleccionar el mejor modelo en función del AIC⁸ partiendo de las q variables. El proceso de stepwise permite obtener un conjunto de $q' < q$ variables que permitan minimizar el AIC.
- (6) Finalmente, se verifica si las q' variables obtenidas son significativas en el modelo. Este proceso debe ser repetido hasta que todas las variables del modelo sean significativas. Se espera que al finalizar completar el punto (5) la cantidad de variables resultantes en el modelo final sea relativamente baja concluyendo en una cantidad $r \leq p'$ variables finalizando del proceso de modelado.
- (7) Se analiza si alguna de las variables esté influyendo en el modelo, esto sucede si el valor de los coeficientes es gradualmente más alto (en valor absoluto) que el resto de las variables. Se hace un análisis de endogeneidad para estos casos. Se evalúa la exclusión de la variable y se repite el punto (5).
- (8) Se analiza la consistencia de los resultados. Analizando los errores y los efectos aleatorios del modelo, la consistencia en las estimaciones⁹, los coeficientes de variación e intervalos de confianza. Esto se lo realiza por medio de gráficos y estadísticos.
- (9) Se establece el modelo final, las estimaciones y se concluye con la generación del mapa.

2.10 R: Paquetes utilizados

El desarrollo del presente estudio requiere del uso de un software especializado en estadística mismo que permita generar desde las estimaciones directas, procesos de selección de variables y la implementación de métodos SAE. El software estadístico R cuenta con todos los componentes mencionados, además de una variedad de paquetes que ofrecen funciones similares. Sin embargo, para este análisis, se optó por utilizar las siguientes librerías específicas, que destacan por su capacidad y eficiencia en el manejo de datos y en la realización de análisis estadísticos:

⁸El criterio de información de Akaike (AIC) proporciona un medio para la selección de un modelo controlando el trade-off entre la bondad de ajuste del modelo y la complejidad del modelo https://es.wikipedia.org/wiki/Criterio_de_informaci%C3%B3n_de_Akaike

⁹En este punto, previo a generar las estimaciones finales, en caso de ser necesario, se podría hacer uso del proceso de benchmarking (conformidad), siempre que se asegure consistencia en las estimaciones cantonales conformes. En el presente estudio no se hizo uso de este recurso dado que se tratan de estimaciones de proporción y no de totales de acuerdo a lo sugerido por Corral et, al., (2022). Asimismo, se observó un comportamiento adecuado en todas las provincias sin la necesidad de este proceso.

- `survey`: provee estadísticas resumidas para muestras de encuestas estratificadas en varias etapas, muestreadas por conglomerados y ponderadas de manera desigual (Lumey 2023).
- `srvyr`: proporciona una nueva forma de calcular estadísticas resumidas sobre datos de encuestas, basada en el paquete `dplyr`¹⁰. El uso de las funciones `srvyr` consta de tres etapas: crear un objeto de encuesta, manipular los datos y calcular las estadísticas de la encuesta (Freedman 2024).
- `glmnet`: ofrece procedimientos eficientes para ajustar el camino de regularización de LASSO o red elástica para regresión lineal, modelos de regresión logística y multinomial, regresión de Poisson, modelo de Cox, gaussiano de respuesta múltiple y regresión multinomial agrupada (Friedman et al. 2023).
- `emdi`: (Estimating and Mapping Disaggregated Indicators) contiene funciones que soportan la estimación, evaluación y mapeo de indicadores desagregados. Los métodos de estimación incluyen al modelo a nivel de área Fay y Herriot (1979) y varias extensiones de este como: métodos de estimación de la varianza ajustada, transformación logarítmica y arcsen, modelos espaciales, robustos, entre otros (A.-K. Kreutzmann et al. 2019b).

Entre las funciones utilizadas para el proceso de modelado se encuentran:

- `cv.glmnet`: Realiza validación cruzada para `glmnet` con `k-folds` y devuelve un valor para `lambda`. La función se ejecuta `glmnet nfolds+1` veces; el primero para obtener la secuencia `lambda` y luego el resto para calcular el ajuste con cada uno de los pliegues omitidos. Se acumula el error y se calcula el error promedio y la desviación estándar sobre los pliegues.
- `step`: Esta función genérica permite seleccionar un modelo por diferentes criterios usando un algoritmo paso a paso (stepwise). Los criterios de información AIC y BIC¹¹ están especialmente desarrollados para los modelos de Fay-Herriot por Marhuenda, Morales, y Pardo (2014).
- `fh`: Estima indicadores utilizando el modelo Fay-Herriot. Se proporcionan EBLUPs y EMC. Además, están disponibles diferentes extensiones del modelo estándar de Fay-Herriot: Se ofrecen métodos de estimación ajustados para la varianza de los efectos aleatorios. Se pueden elegir transformaciones logarítmicas y arcsen para la variable dependiente y dos tipos de transformación inversa: una versión cruda y la introducida por Slud y Maiti (2006) para variables transformadas logarítmicas y una versión naive y con corrección de sesgo siguiendo a Hadam, Wuerz, y Kreutzmann (2020) para variables transformadas `arcsin`.

¹⁰Es una librería de R ampliamente utilizada para la manipulación de datos y pertenece al ecosistema de **tidyverse**. Facilita la realización de operaciones sobre dataframes como la selección, filtrado, agrupamiento y resumen de datos de forma eficiente y con una sintaxis comprensible. <https://cran.r-project.org/web/packages/dplyr/index.html>

¹¹El criterio de información bayesiano (BIC) permite la selección de modelos entre un conjunto finito de modelos. Se basa, en parte, de la función de probabilidad https://es.wikipedia.org/wiki/Criterio_de_informaci%C3%B3n_bayesiano

3 RESULTADOS

3.1 Resultados del modelo y estimaciones

El conjunto de variables independientes inicial partió de 215 variables (características del hogar, variables sociodemográficas del jefe de hogar y cónyuge). Luego, el proceso de selección de variables AdaENET redujo el conjunto a 40 variables (Anexo 2) que posteriormente fueron sometidas a un proceso step-wise del paquete `emd` el mismo que resulta en una cantidad de 15 variables, que minimizan el AIC, se descartaron aquellas variables no significativas, y el modelo resultante considera 12 variables. Finalmente, se realiza un análisis sobre el conjunto final de variables a fin de reducir problemas de endogeneidad y se concluye con 14 variables. Adicionalmente, con la finalidad de brindar información adicional al modelo sobre las provincias de Cañar, Loja y Zamora Chinchipe, se agrega una variable dummy para estas tres provincias logrando un mejor comportamiento en el modelo, con esto, el modelo final, considera 15 variables (Tabla 1).

Tabla 1: Resultados modelo Fay-Herriot (FH) para la prevalencia de DCI

Out-of-sample domains: 44					
In-sample domains: 177					
Variance and EMC estimation:					
Variance estimation method:	reml				
Estimated variance component(s):	0,008583				
EMC method:	prasad-rao				
Coefficients:					
	<i>coefficients</i>	<i>std.error</i>	<i>t.value</i>	<i>p.value</i>	<i>sig.</i>
(Intercept)	-1,4573	0,8587	-1,6972	0,0897	.
tipo_viv_6	15,8866	4,1112	3,8642	0,0001	***
techo_2	0,3385	0,1632	2,0740	0,0381	*
piso_1	1,7185	0,3260	5,2712	0,0000	***
p_agua_4	1,5495	0,3573	4,3367	0,0000	***
b_agua_3	0,8933	0,2160	4,1359	0,0000	***
mascotas_1	1,5832	0,3151	5,0244	0,0000	***
equip_3	-0,6754	0,3184	-2,1213	0,0339	*
cy_etnia_4	-6,7408	1,4680	-4,5918	0,0000	***
cy_educa_0	0,7305	0,4934	1,4804	0,1388	
fert_6	15,6238	3,4726	4,4992	0,0000	***
energia_1	-1,6498	0,7371	-2,2382	0,0252	*
jf_estadocyg_1	-1,9329	0,4489	-4,3061	0,0000	***
jf_seguro_1	0,9506	0,4471	2,1263	0,0335	*
f_etnia_3	1,5700	0,7154	2,1946	0,0282	*
dummy_clz	-0,1700	0,0633	-2,6860	0,0072	**
—					
Sig. codes: 0,000 ***; 0,001 **, 0,01 *, 0,05 .					
Explanatory measures:					
	loglike	AIC	BIC	AdjR2	FH_R2
	-43,5782	121,1565	175,151	0,4464	0,7208

Fuente: Elaboración propia

De acuerdo con la Tabla 1, el AIC del modelo final fue de 121,2 y el R^2 ajustado se encuentra en 0,4464. Esto sugiere que el modelo presenta un buen ajuste y explica una variabilidad moderada de los datos, siendo este el modelo que proporciona el mejor equilibrio entre capacidad de ajuste y complejidad. Adicionalmente, los gráficos de quantiles de la Figura

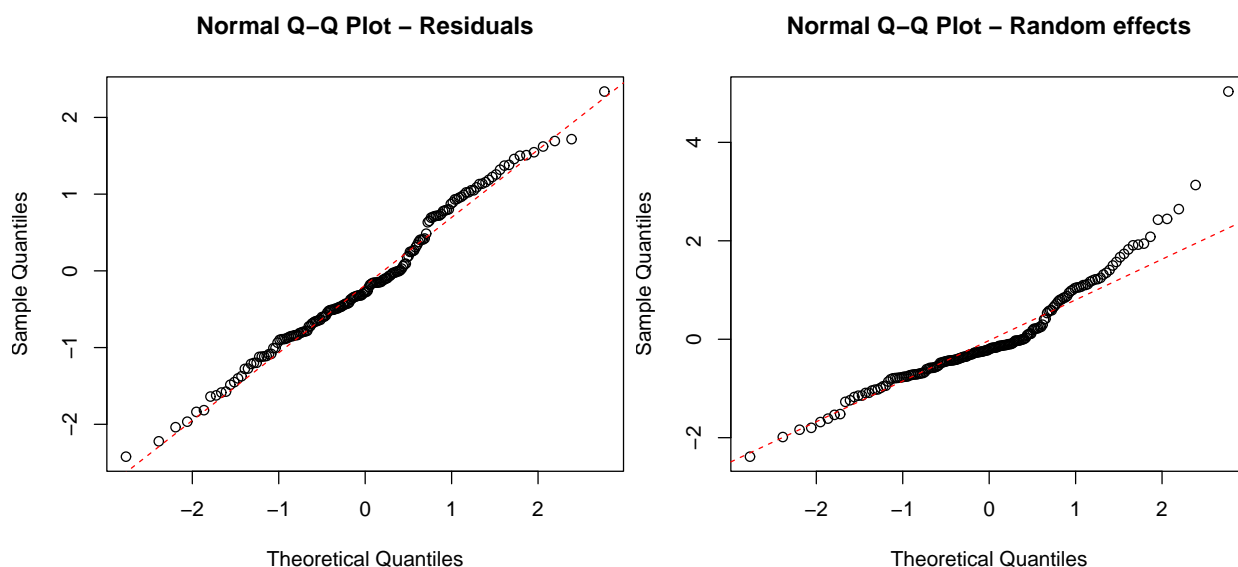


Figura 1: Normalidad en los efectos aleatorios y errores del modelo FH

1, muestran que tanto residuos como efectos aleatorios tienen un comportamiento normal, además de haber superado el test KS con un valor superior a 0.05.

Los resultados del modelo de Fay-Herriot incluyeron una transformación logarítmica de la tasa de DCI y consideraron variables significativas¹² a un nivel de 0.05. Con base en los resultados de la Tabla 1 se puede ver que la mayor parte de las variables del modelo resultantes muestran una asociación positiva con la prevalencia de DCI.

De las variables con asociación positiva con el fenómeno se destacan las variables (*tipo_viv_6*) y (*fert_6*) que corresponden al porcentaje de viviendas de tipo *covacha* y al porcentaje de mujeres en edad fértil que tienen 6 de hijos vivos, respectivamente, por ser los predictores más importantes de la DCI. También, es importante notar que las variables *piso_1*) y (*jf_seguro_1*) que corresponden al porcentaje de viviendas con piso de *duela*, *parquet*, *tablón* o *piso flotante* y al porcentaje de hogares cuyo jefe de hogar tiene afiliación al *IESS seguro general* podrían relacionarse de manera contraintuitiva con el fenómeno, sin embargo, se comprueba su significativa estadística y estas aportan con capacidad predictiva al modelo que es el objetivo final del ejercicio. Por último, en esta misma línea se destaca la capacidad predictiva de la variable (*mascotas_1*) que corresponde al el porcentaje de hogares que *tienen perros*.

Por otro lado, en cuanto a las variables que se relacionan de manera negativa con el fenómeno se resalta la variable de (*cy_etnia_4*) que corresponde al porcentaje de hogares donde la autoidentificación étnica del cónyuge del jefe de hogar es *mulata(o)* y que es el tercer predictor más importante del modelo. Asimismo, la variable de (*energia_1*) que

¹²La variable *cy_educ_0* no logró alcanzar el nivel de significancia de 0.05 después de incluir la variable *dummy_clz*. No obstante, se ha optado por mantener esta variable, puesto que era significativa antes de la inclusión de *dummy_clz*. El objetivo de este proceso era proporcionar información adicional sobre ciertos dominios específicos, sin buscar reducir el número de variables.

corresponde al porcentaje de hogares que tienen energía eléctrica proveniente de una *empresa pública* destaca en cuanto a su significancia y capacidad predictiva considerando que el promedio nacional de los territorios en el Ecuador es alto y con poca dispersión, esto implica que este predictor es particularmente útil solo en aquellas áreas con baja cobertura de servicio eléctrico.

3.2 Coeficientes de variación

La Figura 2 muestra los CVs de las estimaciones directas (azul) y de las estimaciones SAE-FH (rojo). Estos resultados permiten ver que a medida que el tamaño de muestra crece los CVs de las estimaciones SAE son similares a los CVs de las estimaciones directas, cumpliendo con la propiedad del modelo FH.

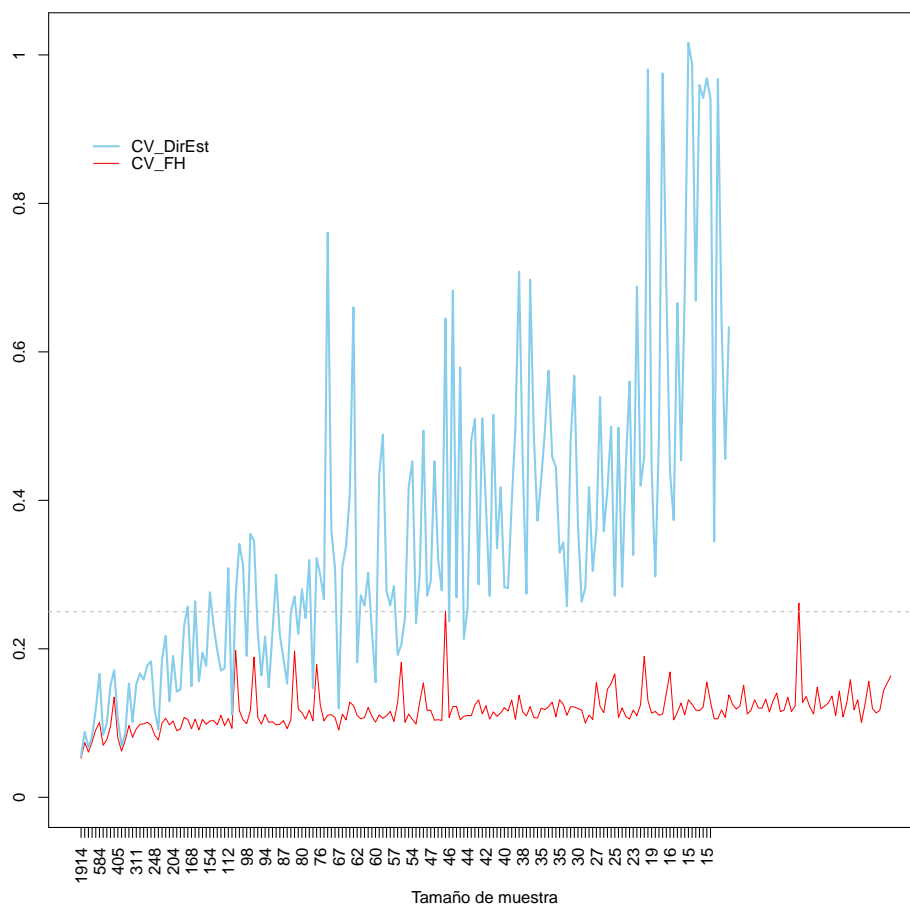


Figura 2: Coeficiente de variación FH vs Directas, ordenado por tamaño de muestra

Por otro lado, a medida que el tamaño de muestra aumenta, las diferencias van creciendo, y el valor de los CVs de las estimaciones directas crecen significativamente, y los CVs de las estimaciones SAE para los 221 cantones se mantiene por debajo del 30%. Al final se puede observar los 44 cantones out-sample que no también tienen un CV que se comporta bajo

los límites tolerables ($CV < 30\%$). Además, la distribución de coeficientes de variación de las estimaciones SAE concentran el 99.1% de las estimaciones en CV menores al 20% lo cual indica que los estimadores son confiables. Solamente el 0.9% de las estimaciones tienen un CV entre 20% y 30% y ninguna estimación tiene un CV superior al 30%.

3.3 Mapa de prevalencia de Desnutrición Crónica Infantil cantonal

Las estimaciones finales se generan una vez que se ha diagnosticado el modelo final y se han analizado adecuadamente los dominios out-of-sample. Con estos resultados se puede construir un mapa de prevalencia de DCI cantonal que se presenta en la figura 3. En este mapa resaltan los territorios de Taisha (en Morona Santiago) y Guamote (en Chimborazo) por tener las prevalencias más altas de DCI (por encima del 55%). De igual manera, los territorios con menor prevalencia de DCI son Camilo Ponce Enríquez (en Azuay) y Samborondón (en Guayas). Por último, las ciudades principales de Quito, Guayaquil y Cuenca tienen prevalencias de 19.6%, 12.7% y 22.7% respectivamente, prevalencias que son convergentes con las estimadas directamente a partir de la ENDI lo que agrega confiabilidad en las estimaciones del modelo.

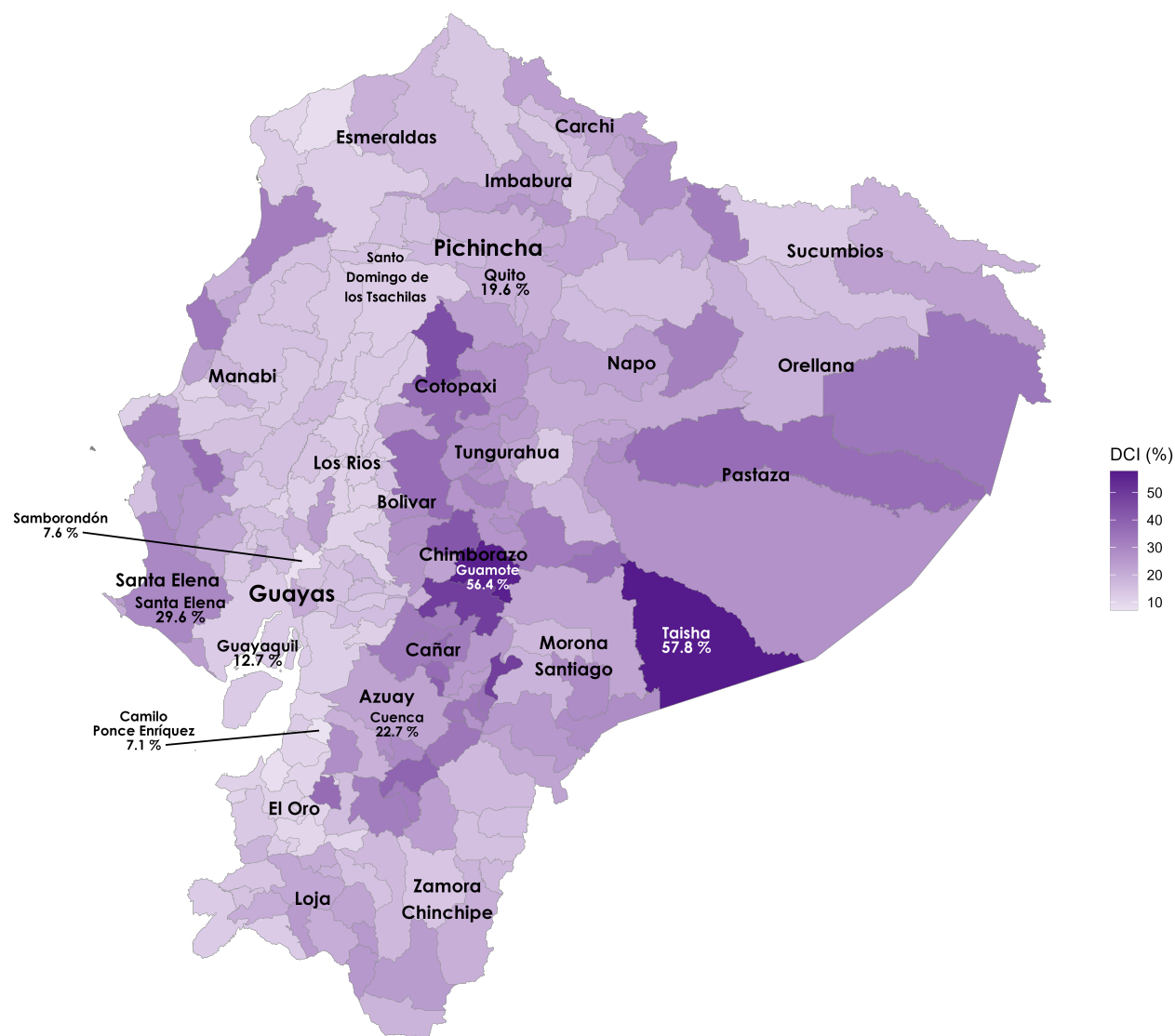


Figura 3: Mapa de la Estimación en Áreas Pequeñas de la DCI cantonal 2023

A finales del 2021, el Estado Ecuatoriano decidió seleccionar al cantón Limón Indanza, un territorio con altos niveles de DCI, para actualizar su prevalencia. Es así que, el INEC llevó a cabo en ese territorio el CENSO, en diciembre del 2021, y la ENDI, en enero del 2022. El levantamiento de información de la encuesta correspondió a una muestra de 355 viviendas, que identificó a 429 niños menores de 5 años, lo cual es mucho mayor que la información disponible en la encuesta ENSANUT para este cantón. El diseño muestral desarrollado para esta operación estadística fue probabilístico, y permitió obtener una estimación directa de la DCI de 21.7% (CV: 8,28%; IC: 18.3%-25,5%) para niños menores de 5 años. En el presente documento, para este cantón, la estimación basada en el modelo FH es de 25,45% con un CV de 10,90% y un IC: 20,01%- 30,89%. Esta estimación es estadísticamente igual a la obtenida

en el 2022, a pesar de que la muestra de la fuente principal es de tan solo 42 menores de 5 años. Asimismo, se puede realizar un análisis pormenorizado en los territorios que componen cada provincia en el país y se observa que todas las medias ponderadas se sitúan dentro de los intervalos de confianza de la estimación provincial obtenida con estimación directa. Estos resultados, además de los análisis de variabilidad y diagnóstico del modelo, constituyen un punto de referencia importante que arroja evidencia sobre la consistencia y confiabilidad de las estimaciones.

4 CONCLUSIONES Y CONSIDERACIONES

4.1 Conclusiones

Se estimó la tasa de prevalencia de DCI para el Ecuador a partir de la información de ENDI 2022-2023 y CPV 2022 con la utilización de un modelo lineal mixto de Fay-Herriot obteniendo 221 estimaciones. El cantón con mayor desnutrición del país es Taisha, ubicado en la provincia de Morona Santiago, lo que evidencia la importancia de realizar ejercicios de estimación en áreas pequeñas, puesto que, con este insumo se puede orientar con mayor precisión los recursos de política pública considerando que Morona Santiago no es la provincia con mayor prevalencia de DCI según los resultados provinciales, pero sí uno de sus cantones.

De las 221 estimaciones, el 99.1% presentan coeficiente de variación inferiores al 20%; esto evidencia la eficiencia de las estimaciones realizadas mediante modelos de estimación en áreas pequeñas. Asimismo, la diagnosis del modelo arrojó una suficiente capacidad predictiva del modelo y un comportamiento adecuado respecto a la normalidad en los residuos; aunque los efectos aleatorios no siguen una distribución normal pero se encuentran centrados alrededor del cero. Estos hallazgos no solo respaldan la viabilidad de utilizar estos modelos de estimación a nivel de área para la estimación de la desnutrición crónica infantil, sino que también subraya su capacidad para optimizar el gasto público en la generación de encuestas especializadas como la ENDI, al incrementar el alcance de los resultados.

Adicionalmente, de los 44 dominios out-of-sample del modelo, 20 cantones corresponden a territorios que no formaron parte del diseño muestral de la ENDI, incluyendo tres cantones de Galápagos cuyo diseño muestral y recolección de información se lo realizó de manera independiente. Esto evidencia otra ventaja de los modelos empleados a nivel de área empleados, que no solo muestran una mejor correlación con el fenómeno y mejor capacidad predictiva que modelos a nivel de unidad utilizados anteriormente; y, son no intensivos en recursos computacionales, sino que también permiten obtener información confiable y precisa sobre territorios no muestreados.

Los criterios de exclusión de los dominios, basados en el número de individuos y estimaciones, utilizado en este trabajo, no buscó restringir en gran medida el número de observaciones, debido a que la población objetivo compuesta por niños menores de 5 años representa una proporción relativamente pequeña dentro del análisis de las encuestas de hogares, lo que puede impactar la cantidad de datos disponibles y, por lo tanto, la capacidad para identificar modelos adecuados que generen estimaciones confiables. Además, es importante puntualizar que la varianza requerida para el modelo FH proviene de estos dominios, y su volatilidad puede influir en la calidad del modelo. Sin embargo, el modelo empleado presenta buenas condiciones con los 177 dominios incluidos, lo que se considera un resultado favorable, constatando que la varianza de las estimaciones in-sample presenta una propiedad consistente con el tamaño de la muestra.

4.2 Consideraciones

Es importante considerar que los resultados obtenidos a partir de la estadística basada en modelos, como los generados mediante métodos de estimación en áreas pequeñas,

presentan características distintivas frente a los derivados de diseños muestrales u otras técnicas tradicionales. A diferencia de estos últimos, la estimación basada en modelos depende de supuestos estadísticos específicos y de la adecuada selección de variables auxiliares que permitan captar las variaciones territoriales de interés. En este sentido y a pesar de toda la evidencia vertida en este documento sobre las propiedades deseables del modelo empleado para obtener las estimaciones, los resultados deben interpretarse con cautela, especialmente en contextos donde la precisión absoluta es crítica. Sin embargo, esto no resta utilidad a las estimaciones para el análisis y la toma de decisiones en política pública, especialmente en entornos de recursos limitados o cuando no es factible obtener datos a través de fuentes exhaustivas.

Por último, es importante considerar que la ENDI cuenta con un diseño muestral apropiado para la estimación de la prevalencia de DCI a niveles provinciales. En este sentido y en línea con la consideración anterior, se debe priorizar el uso de fuentes oficiales y métodos de estimación directa, en este caso provincial, a partir de la encuesta especializada y utilizar el insumo presentado en este estudio solamente en los niveles cantonales, es decir, no se recomienda la ponderación de tasas cantonales para obtener tasas provinciales en ningún escenario.

5 REFERENCIAS

- ADB. 2020. «Introduction to Small Area Estimation Techniques, A Practical Guide for National Statistics Offices». Asian Development Bank. <https://www.adb.org/sites/default/files/publication/609476/small-area-estimation-guide-nsos.pdf>.
- Albán, Andrés. 2022. «Estimación de la tasa de desnutrición crónica infantil cantonal en el Ecuador». Trabajo de Fin de Máster, Universidad de Cantabria, Facultad de Ciencias. <https://repositorio.unican.es/xmlui/bitstream/handle/10902/29818/AlbanFernandezAndresGregorio.pdf>.
- Cabrera, Esteban, Andrea Molina Vera, Mary Alexander Sharman, Lorena Moreno, y Facundo Cuevas. 2016. «Análisis Geográfico de la Pobreza y Desigualdad por Consumo en el Ecuador, Más allá del Nivel Provincial». En *Reporte de Pobreza por Consumo Ecuador 2006-2014*, editado por INEC-BM, 146-72. Quito, Ecuador: Instituto Nacional de Estadística y Censos (INEC), Banco Mundial (BM).
- Corral, P, I. Molina, A. Cojocar, y S. Segovia. 2022. «Guidelines to Small Area Estimation for Poverty Mapping». World Bank Group. <https://www.openknowledge.worldbank.org/server/api/core/bitstreams/1d1fcadc-43e3-541b-8949-fea45dd2a528/content>.
- Fay, R. E., y R. A. Herriot. 1979. «Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data». *Journal of the American Statistical Association*, n.º 85: 398-409. <https://doi.org/10.1080/01621459.1979.10482505>.
- Freedman, Greg. 2024. *Package «srvyr»*. Vienna, Austria: R Foundation for Statistical Computing. <https://cran.r-project.org/web/packages/srvyr/srvyr.pdf>.
- Friedman, Jerome, Trevor Hastie, Rob Tibshirani, Balasubramanian Narasimhan, Kenneth Tay, y Noah Simon. 2023. *Package «glmnet»*. Vienna, Austria: R Foundation for Statistical Computing. <https://cran.r-project.org/web/packages/glmnet/glmnet.pdf>.
- Ghosh, M. 2020. «Small area estimation: its evolution in five decades». *Statistics in Transition New Series*, 1-22.
- Ghosh, M., y R. Steorts. 2013. «Two-stage benchmarking as applied to small area estimation». *Test*, 1-18.
- González-Manteiga, W., M. J. Lombardía, I. Molina, D. Morales, y L. Santamaría. 2008. «Analytic and bootstrap approximations of prediction errors under a multivariate Fay–Herriot model». *Computational Statistics & Data Analysis* 52 (12): 5242-52. <https://doi.org/https://doi.org/10.1016/j.csda.2008.04.031>.
- Hadam, S., N. Wuerz, y A.-K. Kreutzmann. 2020. «Estimating regional unemployment with mobile network data for Functional Urban Areas in Germany». *Refubium - Freie Universitaet Berlin Repository*, 1-28.
- Harmening, S., A.-K. Kreutzmann, S. Schmidt, N. Salvati, y Schmid T. 2019. «A Framework for Producing Small Area Estimates Based on Area-Level Models in R». R-project. https://cran.r-project.org/web/packages/emdi/vignettes/vignette_fh.pdf.
- Horvitz, D. G., y D. J. Thompson. 1952. «A generalization of sampling without replacement from a finite universe». *Journal of the American Statistical Association*, 663-85.
- INEC. 2023b. «Diseño Muestral de la Encuesta Nacional sobre Desnutrición Infantil - ENDI 2022-2023». Instituto Nacional de Estadística y Censos. https://www.ecuadorencifras.gob.ec/documentos/web-inec/ENDI/Disenio_muestral_ENDI_R1.pdf.
- . 2023a. «ENDI». Instituto Nacional de Estadística y Censos. https://www.ecuadorencifras.gob.ec/encuesta_nacional_desnutricion_infantil/.

- Kreutzmann, A.-K., P. Marek, N. Salvati, y Schmid T. 2019a. «Estimating regional wealth in Germany: How different are east and west really?» *Deutsche Bundesbank Discussion Paper No. 35/2019*.
- Kreutzmann, Ann-Kristin, Sören Pannier, Natalia Rojas-Perilla, Timo Schmid, Matthias Templ, y Nikos Tzavidis. 2019b. «The R Package emdi for Estimating and Mapping Regionally Disaggregated Indicators». *Journal of Statistical Software* 91 (7): 1-33. <https://doi.org/10.18637/jss.v091.i07>.
- Lumey, Thomas. 2023. *Package «survey»*. Vienna, Austria: R Foundation for Statistical Computing. <https://cran.r-project.org/web/packages/survey/survey.pdf>.
- Marhuenda, Y., D. Morales, y M. C. Pardo. 2014. «Information criteria for Fay-Herriot model selection». *Computational Statistics and Data Analysis* 70: 268-80.
- MDSF-CEPAL. 2021. «Informe Metodológico. Estimaciones Comunales de Pobreza por ingresos en Chile Mediante Métodos de Estimación en Áreas Pequeñas». Ministerio de Desarrollo Social y de Familia y la Comisión Económica para América Latina y el Caribe (CEPAL). https://traspaso.digital.gob.cl/actas/otros/descarga/Informe_SAE_2020.pdf.
- Molina, I., y J. N. K. Rao. 2015. *Small Area Estimation*. Wiley Series in Survey Methodology.
- Molina, Isabel, y Yolanda Marhuenda. 2015. *R package sae: Methodology*. Vienna, Austria: R Foundation for Statistical Computing. https://cran.r-project.org/web/packages/sae/vignettes/sae_methodology.pdf.
- Molina, Isabel, y J. N. K. Rao. 2010. «Small area estimation of poverty indicators». *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 38 (3): 369-85. <http://www.jstor.org/stable/27896031>.
- Morales, Domingo, María Dolores Esteban, Agustín Pérez, y Tomáš Hobza. 2021. *A Course on Small Area Estimation and Mixed Models. Methods, Theory and Applications in R*. Springer.
- Moura, F., A. Neves, y D. Do N. Silva. 2017. «Small Area Models for Skewed Brazilian Business Survey Data». *Journal of the Royal Statistical Society Series A: Statistics in Society* 18: 1039-55.
- Neves, A., D. Silva, y S. Correa. 2013. «Small Domain Estimation for the Brazilian Service Sector Survey». *Estadística*, 13-37.
- Novo, M. 2018. «Selección de modelos para estimación de áreas pequeñas. Aplicación a datos socioeconómicos de la Comunidad de Galicia». UCS. http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1591.pdf.
- Ogutu, Joseph O, Torben Schulz-Streeck, y Hans-Peter Piepho. 2012. «Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions». *Proceedings of the 15th European workshop on QTL mapping and marker assisted selection (QTLMAS)* 6 (2). <https://doi.org/https://doi.org/10.1186/1753-6561-6-S2-S10>.
- Prasad, N. G. N., y J. N. K. Rao. 1990. «The estimation of the mean squared error of small-area estimators». *Journal of the American Statistical Association* 85 (409): 163-71. <https://doi.org/10.1080/01621459.1990.10475320>.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Shiferaw, Y. 2024. «A spatial analysis of the digital gender gap in South Africa: Are there any fundamentals differences?» *Techological Forecasting & Social Change* 204.
- Slud, E., y T Maiti. 2006. «Mean-squared error estimation in transformed Fay-Herriot models». *Journal of the Royal Statistical Society: Series B* 68 (2): 239-57.
- UNICEF. 2011. «La Desnutrición Infantil: Causas, consecuencias y estrategias para su prevención y tratamiento». UNICEF España. <https://www.salud.gob.ec/wp-content/>

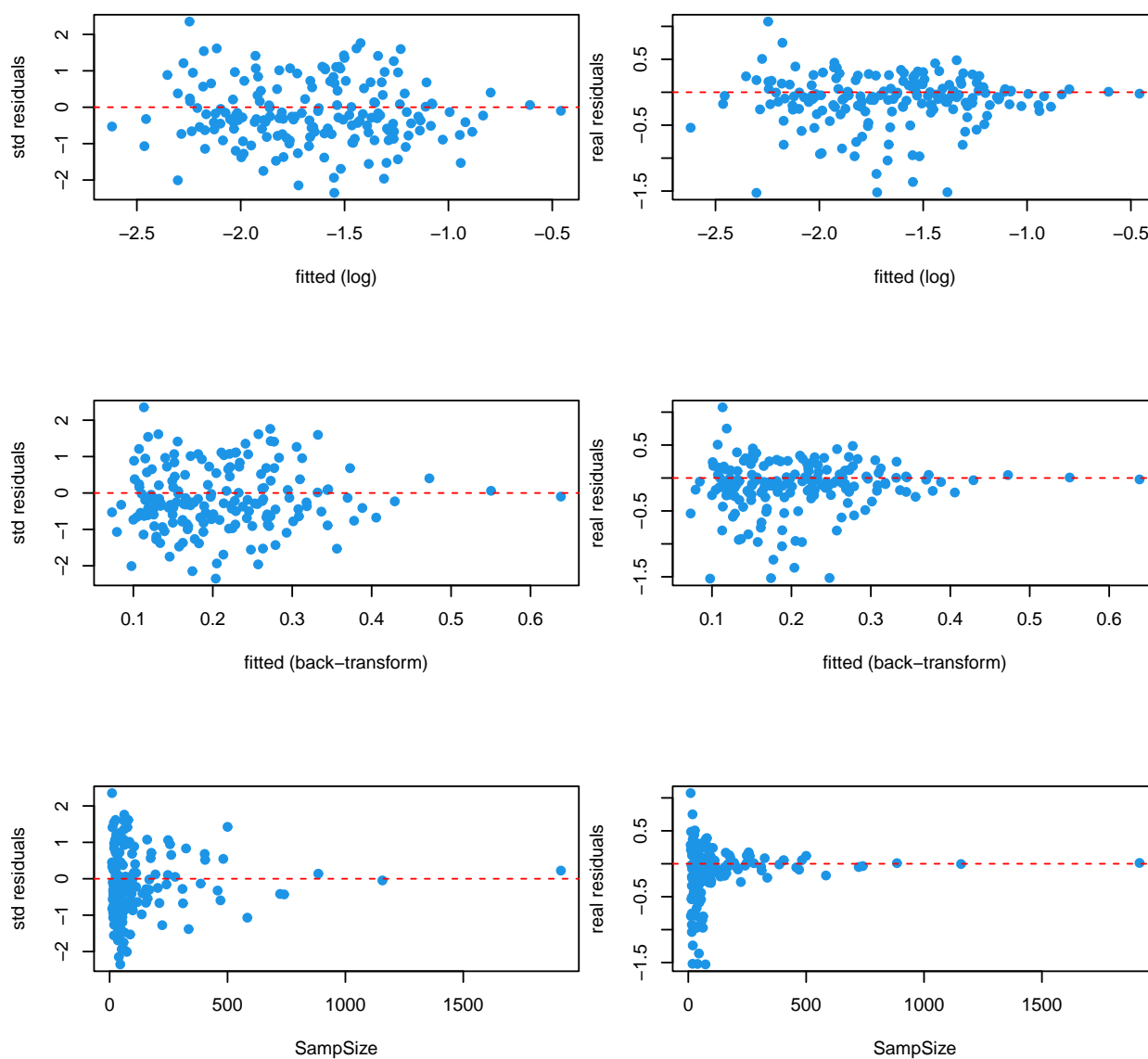
uploads/2016/09/Dossierdesnutricion.pdf.

Wallgren, A., y B. Wallgreen. 2012. «Estadísticas basadas en registros. Aprovechamiento estadístico de datos administrativos». INEGI.

WHO. 2006. *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development*. World Health Organization. https://iris.who.int/bitstream/handle/10665/43413/924154693X_eng.pdf.

ANEXOS

Anexo 1: Gráficos de Diagnosis del modelo final



Anexo 2: Variables

Variable	Descripción
tipo_viv_3	Tipo vivienda: Cuarto/s en casa de inquilinato
tipo_viv_6	Tipo vivienda: Covacha
techo_2	Material techo: Fibrocemento, asbesto (eternit, eurolit)
techo_5	Material techo: Palma, paja u hoja
pared_3	Material pared: Panel prefabricado (yeso, fibrocemento, etc.)
piso_1	Material piso: Duela, parquet, tablón o piso flotante
piso_3	Material piso: Mármol o marmetón
piso_6	Material piso: Caña sin tratar
p_agua_4	El agua proviene...: Carro o tanquero repartidor
ssh_6	Servicio higiénico: Descarga directa al mar, río, lago o quebrada
energia_1	Fuente de energía: Red pública
energia_2	Fuente de energía: Planta eléctrica (generador de luz)
energia_3	Fuente de energía: Energía solar (panel fotovoltaico)
energia_4	Fuente de energía: Energía eólica (a partir del viento)
energia_5	Fuente de energía: Otra fuente (desechos vegetales y animales)
basura_5	Eliminación basura: La entierran
ccocina_1	Combustible para cocinar: Gas (tanque o cilindro)
ccocina_4	Combustible para cocinar: Leña o carbon
b_agua_3	Agua para beber...: La hierven
b_agua_5	Agua para beber...: La filtran (colocan filtros en el grifo o usan purificadores)
mascotas_1	Tiene este hogar perros
propiedad_3	Propia (regalada, donada, heredada o por posesión)
equip_3	Equipamiento: TV por cable, satelital, digital
equip_7	Equipamiento: Máquina lavadora de ropa
equip_12	Equipamiento: Motocicleta de uso exclusivo del hogar
cy_edad_5	Cónyuge entre 25 y 35 años
cy_dfun_4	Cónyuge: Dificultad (funcional) para ver, aun usando lentes
cy_idioma_4	Cónyuge: Indígena o Castellano, español
cy_etnia_4	Cónyuge: Mulata/o
cy_educa_0	Cónyuge: Educación - Ninguna
cy_educa_2	Cónyuge: Educación - entre 6 y 9 años
cy_estadocyg_4	Cónyuge: Viuda/o
nin_edad_1	Niños/as de 0 año
nin_edad_2	Niños/as de 1 años
nin_edad_3	Niños/as de 2 años
nin_cdi	Niños en Centro de Desarrollo Infantil
fert_0	0 hijos vivos de mujeres entre 10-49 años
fert_2	2 hijos vivos de mujeres entre 10-49 años
fert_3	3 hijos vivos de mujeres entre 10-49 años
fert_6	6 hijos vivos de mujeres entre 10-49 años
empleo_2	2 empleados en el hogar
empleo_4	4 empleados en el hogar
empleo_5	5 empleados en el hogar
empleo_6	6 empleados en el hogar

Anexo 3: Estimaciones Cantonales

ID	PROVINCIA	ID	CANTÓN	DCI	EMC	CV
01	AZUAY	0101	CUENCA	0,227	0,0003	0,075
01	AZUAY	0102	GIRÓN	0,290	0,0012	0,121
01	AZUAY	0103	GUALACEO	0,344	0,0012	0,101
01	AZUAY	0104	NABÓN	0,390	0,0019	0,111
01	AZUAY	0105	PAUTE	0,244	0,0007	0,110
01	AZUAY	0106	PUCARÁ	0,280	0,0009	0,106
01	AZUAY	0107	SAN FERNANDO	0,282	0,0011	0,119
01	AZUAY	0108	SANTA ISABEL	0,194	0,0005	0,114
01	AZUAY	0109	SÍGSIG	0,343	0,0014	0,110
01	AZUAY	0110	OÑA	0,326	0,0016	0,123
01	AZUAY	0111	CHORDELEG	0,268	0,0013	0,135
01	AZUAY	0112	EL PAN	0,349	0,0020	0,127
01	AZUAY	0113	SEVILLA DE ORO	0,488	0,0044	0,137
01	AZUAY	0114	GUACHAPALA	0,263	0,0010	0,117
01	AZUAY	0115	CAMILO PONCE ENRÍQUEZ	0,071	0,0001	0,131
02	BOLÍVAR	0201	GUARANDA	0,370	0,0005	0,062
02	BOLÍVAR	0202	CHILLANES	0,262	0,0007	0,102
02	BOLÍVAR	0203	CHIMBO	0,282	0,0008	0,099
02	BOLÍVAR	0204	ECHEANDÍA	0,150	0,0003	0,106
02	BOLÍVAR	0205	SAN MIGUEL	0,275	0,0007	0,097
02	BOLÍVAR	0206	CALUMA	0,148	0,0003	0,112
02	BOLÍVAR	0207	LAS NAVES	0,121	0,0002	0,110
03	CAÑAR	0301	AZOGUES	0,254	0,0006	0,097
03	CAÑAR	0302	BIBLIÁN	0,369	0,0014	0,101
03	CAÑAR	0303	CAÑAR	0,327	0,0006	0,077
03	CAÑAR	0304	LA TRONCAL	0,118	0,0002	0,107
03	CAÑAR	0305	EL TAMBO	0,372	0,0019	0,117
03	CAÑAR	0306	DÉLEG	0,403	0,0024	0,122
03	CAÑAR	0307	SUSCAL	0,254	0,0018	0,169
04	CARCHI	0401	TULCÁN	0,237	0,0004	0,081
04	CARCHI	0402	BOLÍVAR	0,162	0,0004	0,120
04	CARCHI	0403	ESPEJO	0,228	0,0005	0,100
04	CARCHI	0404	MIRA	0,167	0,0004	0,112
04	CARCHI	0405	MONTÚFAR	0,224	0,0004	0,092
04	CARCHI	0406	SAN PEDRO DE HUACA	0,274	0,0011	0,123
05	COTOPAXI	0501	LATACUNGA	0,267	0,0006	0,090
05	COTOPAXI	0502	LA MANÁ	0,129	0,0002	0,105
05	COTOPAXI	0503	PANGUA	0,233	0,0006	0,105
05	COTOPAXI	0504	PUJILÍ	0,362	0,0014	0,103
05	COTOPAXI	0505	SALCEDO	0,266	0,0007	0,099
05	COTOPAXI	0506	SAQUISILÍ	0,332	0,0011	0,101
05	COTOPAXI	0507	SIGCHOS	0,446	0,0027	0,116
06	CHIMBORAZO	0601	RIOBAMBA	0,264	0,0004	0,080
06	CHIMBORAZO	0602	ALAUÍS	0,489	0,0021	0,093
06	CHIMBORAZO	0603	COLTA	0,424	0,0019	0,103
06	CHIMBORAZO	0604	CHAMBO	0,276	0,0008	0,104
06	CHIMBORAZO	0605	CHUNCHI	0,318	0,0012	0,109
06	CHIMBORAZO	0606	GUAMOTE	0,564	0,0026	0,091
06	CHIMBORAZO	0607	GUANO	0,313	0,0008	0,093
06	CHIMBORAZO	0608	PALLATANGA	0,230	0,0007	0,119
06	CHIMBORAZO	0609	PENIPE	0,273	0,0009	0,110
06	CHIMBORAZO	0610	CUMANDÁ	0,192	0,0004	0,106
07	EL ORO	0701	MACHALA	0,080	0,0001	0,101

Continúa...

ID	PROVINCIA	ID	CANTÓN	DCI	EMC	CV
07	EL ORO	0702	ARENILLAS	0,131	0,0002	0,104
07	EL ORO	0703	ATAHUALPA	0,106	0,0001	0,111
07	EL ORO	0704	BALSAS	0,114	0,0002	0,108
07	EL ORO	0705	CHILLA	0,367	0,0028	0,143
07	EL ORO	0706	EL GUABO	0,104	0,0001	0,106
07	EL ORO	0707	HUAQUILLAS	0,102	0,0001	0,117
07	EL ORO	0708	MARCABELÍ	0,127	0,0002	0,105
07	EL ORO	0709	PASAJE	0,106	0,0001	0,103
07	EL ORO	0710	PIÑAS	0,100	0,0001	0,111
07	EL ORO	0711	PORTOVELO	0,102	0,0001	0,107
07	EL ORO	0712	SANTA ROSA	0,113	0,0001	0,105
07	EL ORO	0713	ZARUMA	0,162	0,0003	0,098
07	EL ORO	0714	LAS LAJAS	0,184	0,0004	0,112
08	ESMERALDAS	0801	ESMERALDAS	0,084	0,0001	0,135
08	ESMERALDAS	0802	ELOY ALFARO	0,170	0,0010	0,189
08	ESMERALDAS	0803	MUISNE	0,123	0,0010	0,251
08	ESMERALDAS	0804	QUININDÉ	0,123	0,0002	0,101
08	ESMERALDAS	0805	SAN LORENZO	0,138	0,0007	0,197
08	ESMERALDAS	0806	ATACAMES	0,099	0,0004	0,198
08	ESMERALDAS	0807	RIOVERDE	0,193	0,0012	0,179
09	GUAYAS	0901	GUAYAQUIL	0,127	0,0001	0,091
09	GUAYAS	0902	ALF BAQUERIZO MORENO (JUAN)	0,149	0,0004	0,131
09	GUAYAS	0903	BALAO	0,106	0,0001	0,114
09	GUAYAS	0904	BALZAR	0,154	0,0004	0,122
09	GUAYAS	0905	COLIMES	0,136	0,0003	0,125
09	GUAYAS	0906	DAULE	0,155	0,0004	0,124
09	GUAYAS	0907	DURÁN	0,208	0,0010	0,154
09	GUAYAS	0908	EL EMPALME	0,166	0,0004	0,117
09	GUAYAS	0909	EL TRIUNFO	0,161	0,0003	0,108
09	GUAYAS	0910	MILAGRO	0,154	0,0003	0,103
09	GUAYAS	0911	NARANJAL	0,120	0,0002	0,111
09	GUAYAS	0912	NARANJITO	0,164	0,0004	0,117
09	GUAYAS	0913	PALESTINA	0,108	0,0002	0,127
09	GUAYAS	0914	PEDRO CARBO	0,213	0,0007	0,125
09	GUAYAS	0916	SAMBORONDÓN	0,076	0,0001	0,158
09	GUAYAS	0918	SANTA LUCÍA	0,155	0,0003	0,120
09	GUAYAS	0919	SALITRE (URBINA JADO)	0,193	0,0006	0,128
09	GUAYAS	0920	SAN JACINTO DE YAGUACHI	0,149	0,0003	0,118
09	GUAYAS	0921	PLAYAS	0,239	0,0008	0,120
09	GUAYAS	0922	SIMÓN BOLÍVAR	0,179	0,0005	0,130
09	GUAYAS	0923	CORONEL MARCELINO MARIDUEÑA	0,137	0,0002	0,112
09	GUAYAS	0924	LOMAS DE SARGENTILLO	0,135	0,0003	0,132
09	GUAYAS	0925	NOBOL	0,216	0,0008	0,131
09	GUAYAS	0927	GENERAL ANTONIO ELIZALDE	0,150	0,0002	0,101
09	GUAYAS	0928	ISIDRO AYORA	0,144	0,0003	0,113
10	IMBABURA	1001	IBARRA	0,138	0,0002	0,098
10	IMBABURA	1002	ANTONIO ANTE	0,225	0,0005	0,099
10	IMBABURA	1003	COTACACHI	0,233	0,0006	0,107
10	IMBABURA	1004	OTAVALO	0,259	0,0006	0,092
10	IMBABURA	1005	PIMAMPIRO	0,162	0,0004	0,118
10	IMBABURA	1006	SAN MIGUEL DE URCUQUÍ	0,206	0,0005	0,105
11	LOJA	1101	LOJA	0,172	0,0003	0,097
11	LOJA	1102	CALVAS	0,201	0,0005	0,116
11	LOJA	1103	CATAMAYO	0,152	0,0003	0,115
11	LOJA	1104	CELICA	0,200	0,0006	0,122
11	LOJA	1105	CHAGUARPAMBA	0,206	0,0007	0,126
11	LOJA	1106	ESPÍNDOLA	0,255	0,0009	0,116

Continúa...

ID	PROVINCIA	ID	CANTÓN	DCI	EMC	CV
11	LOJA	1107	GONZANAMÁ	0,223	0,0007	0,117
11	LOJA	1108	MACARÁ	0,126	0,0003	0,136
11	LOJA	1109	PALTAS	0,216	0,0006	0,116
11	LOJA	1110	PUYANGO	0,145	0,0003	0,115
11	LOJA	1111	SARAGURO	0,322	0,0012	0,107
11	LOJA	1112	SOZORANGA	0,249	0,0010	0,127
11	LOJA	1113	ZAPOTILLO	0,127	0,0003	0,128
11	LOJA	1114	PINDAL	0,148	0,0003	0,120
11	LOJA	1115	QUILANGA	0,251	0,0010	0,123
11	LOJA	1116	OLMEDO	0,193	0,0006	0,130
12	LOS RÍOS	1201	BABAHOYO	0,108	0,0001	0,105
12	LOS RÍOS	1202	BABA	0,250	0,0017	0,166
12	LOS RÍOS	1203	MONTALVO	0,133	0,0002	0,108
12	LOS RÍOS	1204	PUEBLOVIEJO	0,113	0,0001	0,105
12	LOS RÍOS	1205	QUEVEDO	0,113	0,0001	0,097
12	LOS RÍOS	1206	URDANETA	0,114	0,0002	0,110
12	LOS RÍOS	1207	VENTANAS	0,121	0,0002	0,111
12	LOS RÍOS	1208	VINCES	0,119	0,0002	0,111
12	LOS RÍOS	1209	PALENQUE	0,146	0,0004	0,140
12	LOS RÍOS	1210	BUENA FE	0,129	0,0002	0,105
12	LOS RÍOS	1211	VALENCIA	0,133	0,0003	0,121
12	LOS RÍOS	1212	MOCACHE	0,126	0,0002	0,125
12	LOS RÍOS	1213	QUINSALOMA	0,129	0,0003	0,123
13	MANABÍ	1301	PORTOVIEJO	0,132	0,0002	0,100
13	MANABÍ	1302	BOLÍVAR	0,136	0,0003	0,122
13	MANABÍ	1303	CHONE	0,153	0,0003	0,107
13	MANABÍ	1304	EL CARMEN	0,135	0,0002	0,117
13	MANABÍ	1305	FLAVIO ALFARO	0,146	0,0003	0,122
13	MANABÍ	1306	JIPIJAPA	0,279	0,0009	0,110
13	MANABÍ	1307	JUNÍN	0,183	0,0006	0,137
13	MANABÍ	1308	MANTA	0,113	0,0001	0,108
13	MANABÍ	1309	MONTECRISTI	0,303	0,0030	0,182
13	MANABÍ	1310	PAJÁN	0,265	0,0011	0,122
13	MANABÍ	1311	PICHINCHA	0,138	0,0004	0,146
13	MANABÍ	1312	ROCAFUERTE	0,110	0,0002	0,122
13	MANABÍ	1313	SANTA ANA	0,152	0,0004	0,125
13	MANABÍ	1314	SUCRE	0,234	0,0008	0,119
13	MANABÍ	1315	TOSAGUA	0,177	0,0005	0,128
13	MANABÍ	1316	24 DE MAYO	0,363	0,0032	0,155
13	MANABÍ	1317	PEDERNALES	0,323	0,0018	0,131
13	MANABÍ	1318	OLMEDO	0,214	0,0011	0,156
13	MANABÍ	1319	PUERTO LÓPEZ	0,155	0,0005	0,138
13	MANABÍ	1320	JAMA	0,191	0,0009	0,155
13	MANABÍ	1321	JARAMIJÓ	0,181	0,0006	0,140
13	MANABÍ	1322	SAN VICENTE	0,334	0,0025	0,151
14	MORONA SANTIAGO	1401	MORONA	0,221	0,0004	0,091
14	MORONA SANTIAGO	1402	GUALAQUIZA	0,169	0,0003	0,104
14	MORONA SANTIAGO	1403	LIMÓN INDANZA	0,255	0,0008	0,109
14	MORONA SANTIAGO	1404	PALORA	0,178	0,0004	0,108
14	MORONA SANTIAGO	1405	SANTIAGO	0,185	0,0004	0,107
14	MORONA SANTIAGO	1406	SUCÚA	0,177	0,0003	0,104
14	MORONA SANTIAGO	1407	HUAMBOYA	0,355	0,0018	0,120
14	MORONA SANTIAGO	1408	SAN JUAN BOSCO	0,225	0,0007	0,120
14	MORONA SANTIAGO	1409	TAISHA	0,578	0,0228	0,262
14	MORONA SANTIAGO	1410	LOGROÑO	0,285	0,0014	0,131

Continúa...

ID	PROVINCIA	ID	CANTÓN	DCI	EMC	CV
14	MORONA SANTIAGO	1411	PABLO SEXTO	0,330	0,0014	0,114
14	MORONA SANTIAGO	1412	TIWINTZA	0,284	0,0029	0,190
15	NAPO	1501	TENA	0,207	0,0003	0,078
15	NAPO	1503	ARCHIDONA	0,231	0,0005	0,098
15	NAPO	1504	EL CHACO	0,159	0,0003	0,117
15	NAPO	1507	QUIJOS	0,167	0,0004	0,124
15	NAPO	1509	CARLOS JULIO AROS TOLA	0,219	0,0006	0,111
16	PASTAZA	1601	PASTAZA	0,270	0,0004	0,070
16	PASTAZA	1602	MERA	0,227	0,0005	0,098
16	PASTAZA	1603	SANTA CLARA	0,282	0,0010	0,114
16	PASTAZA	1604	ARAJUNO	0,365	0,0022	0,127
17	PICHINCHA	1701	D M QUITO	0,196	0,0001	0,053
17	PICHINCHA	1702	CAYAMBE	0,226	0,0007	0,121
17	PICHINCHA	1703	MEJÍA	0,236	0,0006	0,105
17	PICHINCHA	1704	PEDRO MONCAYO	0,208	0,0010	0,153
17	PICHINCHA	1705	RUMIÑAHUI	0,189	0,0004	0,106
17	PICHINCHA	1707	SAN MIGUEL DE LOS BANCOS	0,175	0,0004	0,115
17	PICHINCHA	1708	PEDRO VICENTE MALDONADO	0,154	0,0003	0,118
17	PICHINCHA	1709	PUERTO QUITO	0,155	0,0003	0,118
18	TUNGURAHUA	1801	AMBATO	0,269	0,0003	0,061
18	TUNGURAHUA	1802	BAÑOS DE AGUA SANTA	0,135	0,0002	0,115
18	TUNGURAHUA	1803	CEVALLOS	0,252	0,0008	0,112
18	TUNGURAHUA	1804	MOCHA	0,285	0,0011	0,117
18	TUNGURAHUA	1805	PATATE	0,224	0,0005	0,103
18	TUNGURAHUA	1806	QUERO	0,270	0,0010	0,117
18	TUNGURAHUA	1807	SAN PEDRO DE PELILEO	0,220	0,0005	0,098
18	TUNGURAHUA	1808	SANTIAGO DE PÍLLARO	0,247	0,0008	0,112
18	TUNGURAHUA	1809	TISALEO	0,298	0,0011	0,110
19	ZAMORA CHINCHIPE	1901	ZAMORA	0,141	0,0002	0,099
19	ZAMORA CHINCHIPE	1902	CHINCHIPE	0,183	0,0004	0,108
19	ZAMORA CHINCHIPE	1903	NANGARITZA	0,197	0,0006	0,125
19	ZAMORA CHINCHIPE	1904	YACUAMBI	0,239	0,0008	0,117
19	ZAMORA CHINCHIPE	1905	YANTZAZA	0,155	0,0003	0,105
19	ZAMORA CHINCHIPE	1906	EL PANGUI	0,161	0,0003	0,114
19	ZAMORA CHINCHIPE	1907	CENTINELA DEL CÓNDOR	0,184	0,0004	0,112
19	ZAMORA CHINCHIPE	1908	PALANDA	0,244	0,0010	0,128
19	ZAMORA CHINCHIPE	1909	PAQUISHA	0,172	0,0004	0,122
20	GALÁPAGOS	2001	SAN CRISTÓBAL	0,101	0,0002	0,144
20	GALÁPAGOS	2002	ISABELA	0,104	0,0003	0,154
20	GALÁPAGOS	2003	SANTA CRUZ	0,086	0,0002	0,164
21	SUCUMBÍOS	2101	LAGO AGRIO	0,130	0,0002	0,095
21	SUCUMBÍOS	2102	GONZALO PIZARRO	0,212	0,0005	0,104
21	SUCUMBÍOS	2103	PUTUMAYO	0,191	0,0006	0,128
21	SUCUMBÍOS	2104	SHUSHUFINDI	0,156	0,0003	0,103
21	SUCUMBÍOS	2105	SUCUMBÍOS	0,277	0,0013	0,131
21	SUCUMBÍOS	2106	CASCALES	0,322	0,0012	0,109
21	SUCUMBÍOS	2107	CUYABENO	0,233	0,0007	0,109
22	ORELLANA	2201	FRANCISCO DE ORELLANA	0,189	0,0003	0,092
22	ORELLANA	2202	AGUARICO	0,341	0,0026	0,149
22	ORELLANA	2203	LA JOYA DE LOS SACHAS	0,153	0,0003	0,103
22	ORELLANA	2204	LORETO	0,316	0,0012	0,110
23	STO DOM DE LOS TSÁCHILAS	2301	SANTO DOMINGO	0,124	0,0001	0,074
23	STO DOM DE LOS TSÁCHILAS	2302	LA CONCORDIA	0,153	0,0003	0,104
24	SANTA ELENA	2401	SANTA ELENA	0,296	0,0005	0,077
24	SANTA ELENA	2402	LA LIBERTAD	0,242	0,0004	0,084
24	SANTA ELENA	2403	SALINAS	0,204	0,0004	0,102



@InecEcuador



@ecuadorencifras



@ecuadorencifras



INECEcuador

